ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

Ph.D. Thesis

RENÁTA LEVENDOVICS

Endoscopic image and kinematic data-based autonomous technical and non-technical skill assessment in Robot-Assisted Minimally Invasive Surgery

Robotsebészeti technikai és nem-technikai készségek automatizált mérése endoszkópos kameraképek és kinematikai adatok alapján

Supervisor:
Prof. Dr. Tamás Haidegger

DOCTORAL SCHOOL OF APPLIED INFORMATICS AND APPLIED MATHEMATICS

Budapest, 2023

Members of the Defense Committee:

Chair of the Committee:
Prof. Dr. Miklós Kozlovszky (Óbuda University)

Opponents:
Prof. Dr. Miklós Kozlovszky (Óbuda University)
Dr. Tamás Ungi (Queen's University)
Prof. Dr. Márta Takács (Óbuda University)

Secretary:
Dr. József Kuti (Óbuda University)

Members:
Dr. Szilveszter Pletl (University of Szeged)
Dr. Károly Széll (Óbuda University)
Prof. Dr. József Sándor (Semmelweis University)

Replacement Members of the Defense Committee:

Chair of the Committee:
Prof. Dr. Aurél Galántai (Óbuda University)

Secretary:
Prof. Dr. Péter Galambos (Óbuda University)

Opponent:
Dr. Róbert Lovas (Óbuda University)

Member:
Prof. Dr. Ferenc Bari (University of Szeged)

Members of the Comprehensive Examination Committee:

Chair:
Prof. Dr. Aurél Galántai (Óbuda University)

Members:
Prof. Dr. Levente Kovács (Óbuda University)
Prof. Dr. József Kázmér Tar (Óbuda University)
Prof. Dr. Márta Takács (Óbuda University)

Date of the Defense:

December, 2023.

DECLARATION

Undersigned, Renáta Nagyné Elek, hereby I state that this Ph.D. thesis is my own work, wherein I only used the sources listed in the references. All parts taken from other works, either as word for word citation or rewritten keeping the original meaning, have been unambiguously marked, and reference to the source was included.

NYILATKOZAT

Alulírott Nagyné Elek Renáta kijelentem, hogy ezt a doktori értekezést önállóan készítettem, és abban csak az irodalmi hivatkozások listájában szereplő forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Budapest, November 4, 2022

Renáta Nagyné Elek

DECLARATION

Undersigned, Renáta Nagyné Elek, hereby I ask that my thesis titled Endoscopic image and kinematic data-based autonomous technical and non-technical skill assessment in Robot-Assisted Minimally Invasive Surgery be published on the internet

- With no restraint;
- With availability only from Hungarian address;
- 2 years after awarding the degree, with no restraint;
- 2 years after awarding the degree, with availability only from Hungarian address

NYILATKOZAT

Alulírott Nagyné Elek Renáta kérem, hogy Robotsebészeti technikai és nem-technikai készségek automatizált mérése endoszkópos kameraképek és kinematikai adatok alapján című doktori értekezésem interneten történő nyilvánosságra hozatala

- Korlátozás nélkül;
- Elérhetőség csak magyarországi címről;
- A fokozat odaítélését követően 2 év múlva, korlátozás nélkül;
- A fokozat odaítélését követően 2 év múlva, csak magyarországi címről történjen meg.

Budapest, November 4, 2022

Renáta Nagyné Elek

ABSTRACT

In the case of Minimally Invasive Surgery (MIS), surgeons reach the human organs through small skin incisions. This approach – compared to open-access surgery – results in less trauma, smaller scars and quicker recovery to the patient. On the other hand, MIS requires extensive training, because handling the MIS surgical (laparoscopic) tools is not trivial, the operating area is visualized by an endoscopic camera, the surgeon has limited view, and long surgeries can cause great fatigue. Robot-Assisted Minimally Invasive Surgery (RAMIS) can provide help with these MIS challenges. The da Vinci Surgical System (dVSS) is at the moment the market leading RAMIS system with more than 7000 clinical systems all over the world. DVSS can help the surgeons with 3D vision, ergonomy, intuitive tool handling, tremor filtering and motion scaling. DVSS is a teleoperational system, the surgeon remotely operates the patient-side robotic arms from a master console. With dVSS, MIS can be more precise, and it can decrease the workload on the surgeon. MIS was a revolution in medicine 30 years ago, and it is now part of the everyday clinical routine. Training is crucial in the case of MIS, since it requires extensive practical skills as well. Nevertheless, MIS surgical skill assessment is not part of the clinical practice. During their studies, surgical residents have to take practical exams, but the assessment of these is usually done manually by an expert surgeon. The goal is to assess the skills of the surgeon autonomously, with available or additional sensors. With autonomous approaches, un-biased, objective surgical skill assessment can be achieved, furthermore, it does not require resources from expert surgeons. With the dVSS, these sensory information are more trivial to access, the kinematic and video data of the surgeon is recordable, and the motion of them can be examined. There are technical and non-technical skills in surgery; technical skills include instrument handling, knowledge of equipment, knowledge of procedure, and indirect indicators (used forces, elapsed time), etc.; non-technical skills are such as leadership, situation awareness, decision making, dealing with stress, etc. Both of the skillsets make a good surgeon, and with frequent assessment personalized training and better patient outcome is achievable. Since kinematic data-based skill classification in RAMIS can achieve an almost perfect accuracy, but endoscopic images are available in traditional MIS and in training videos, in this work, RAMIS technical skill assessment was examined through endoscopic images, and their correlation with kinematic data was studied. In my first thesis group, I proved the applicability of image-based surgical tool pose estimation in RAMIS technical skill assessment, with proposing an articulated tool pose estimation methodology for robotic surgery training videos and showing its correlation with kinematic data. I also suggested a semantic segmentation method for surgical tools in a skill-annotated database to validate image-based surgical skill assessment. In my second thesis group, I proved the hypothesis that surgical non-technical skills can be accurately estimated with objective parameters (image and force). In this thesis work, I presented a training environment and workflow for laparoscopic cholecystectomy training, simulating

a stressful surgical environment. I showed that Artificial Intelligence methods are powerful tools not just in technical, but in non-technical skill assessment as well. In my third thesis group, I proposed a complete framework for skill assessment in the case of surgical automation. Automating the motion of the camera holder arm is already the part of certain commercialized laparoscopic systems, and surgical skill assessment is essential in this cases for safety reasons. In this thesis group, I proved that Optical Flow is an image feature, which correlates with the surgical technical skills. I also proposed a camera motion automation method for the dVSS. Finally, I validated an Optical Flow ego-motion compensation method to extract the surgical tool motions only, and to exclude camera motion from the Optical Flow vector field. These theses can widely support automation of RAMIS technical and non-technical skill assessment based on sensory data, which can be the next step towards the deeper understanding of surgical data.

KIVONAT

Minimál Invazív Sebészet (MIS) esetén a sebészek kis bemetszéseken keresztül érik el a belső szerveket. Ez a megközelítés – a nyílt műtéthez képest – kisebb traumát és hegeket, gyorsabb felépülést eredményez a betegnek. A MIS azonban rendszeres gyakorlást igényel; a MIS sebészeti (laparoszkópos) eszközök kezelése nem triviális, a műtéti területet endoszkópos kamera által vizualizált, a sebésznek korlátozott rálátása van az operálandó területre, továbbá a hosszú műtétek nagy fizikai és mentális terhelést okozhatnak. A Robottal Támogatott Minimál Invazív Sebészet (Robot-Assisted Minimally Invasive Surgery, RAMIS) támogatást nyújthat az MIS kihívásaihoz. A da Vinci Sebészeti Robotrendszer (da Vinci Surgical System, dVSS) jelenleg a piacvezető RAMIS eszköz, több mint 7000 klinikai rendszerrel a világon. A dVSS segítséget nyújthat a sebészeknek a 3D látás, az ergonómia, az intuitív eszközkezelés, a kézremegés szűrés és átskálázott mozgás révén. A dVSS egy teleoperációs rendszer, a sebész távolról irányítja a beteg oldali robotkarokat a sebész oldali konzollal, segítségével a MIS beavatkozás pontosabb lehet, és csökkentheti a sebész terhelését. A MIS forradalmat jelentett az orvostudományban, azonban ma már a mindennapi klinikai rutin része. A képzés döntő fontosságú az MIS esetében, mivel széleskörű gyakorlati ismereteket igényel. Ezzel szemben az MIS sebészeti készségek felmérése nem része a klinikai gyakorlatnak. Tanulmányaik során a sebészeknek gyakorlati vizsgákat kell tenniük, de ennek kiértékelését általában egy sebész szakorvos manuálisan végzi. A cél a sebész képességeinek automatizált mérése, rendelkezésre álló vagy integrált szenzorokkal. Automatizált megoldásokkal megvalósítható az objektív sebészeti készségfelmérés, továbbá nem igényli gyakorlott sebész bevonását. A dVSS segítségével ezek a szenzoros információk elérhetőek, a sebész kinematikai és videó adatai rögzíthetők, mozgásuk vizsgálható. Léteznek ún. technikai és nem-technikai készségek a sebészetben; a technikai készségek, mint például: eszközkezelés és -ismeret, eljárások ismerete, indirekt indikátorok, pl. felhasznált erők, idő, stb.; a nem-technikai készségek, mint a vezetői készségek, a helyzeti tudatosság, a döntéshozás, a stressz kezelése, stb. Mindkét készség-típus fontos a sebészetben, készségfelméréssel pedig személyre szabott képzés és kedvezőbb műtéti kimenet érhető el. Mivel RAMIS esetében a kinematikai adatokon alapuló képességosztályozás majdnem tökéletes pontosságot érhet el, és az endoszkópos képek a hagyományos MIS és oktatóvideók esetében is elérhetők, ebben a dolgozatban a RAMIS technikai készségfelmérést endoszkópos kameraképeken keresztül vizsgáltam, illetve a képi adatok összefüggését a kinematikai adatokkal is bemutattam. Bizonyítottam a kép alapú sebészeti eszközök pozícióbecslésének alkalmazhatóságát a RAMIS technikai készségfelmérésben egy sebészeti eszköz pozícióbecslés módszertan javaslatával robotsebészeti oktatóvideókhoz és bemutatom a korrelációját a kinematikai adatokkal. Javasoltam továbbá egy szemantikai szegmentálási módszert a sebészeti eszközökhöz egy készség-annotált adatbázisban a képalapú sebészeti készségfelmérés validálására. Második tézisemben azt a hipotézist bizonyítottam, hogy a sebé-

szeti nem-technikai készségek objektív paraméterekkel (kép és erő) becsülhetők. Javasoltam a laparoszkópos cholecystectomia tréninghez képzési környezetet és munkafolyamatot, egy stresszes műtéti környezetet szimulálva. A dolgozatban bemutattam, hogy a mesterséges intelligencia alapú módszerek nemcsak a technikai, hanem a nem-technikai készségfelmérésben is hatékony eszközök. Harmadik téziscsoportomban egy keretrendszert javasoltam a készségfelméréshez sebészeti folyamatok automatizálásához. A kameratartó kar mozgásának automatizálása már része bizonyos kereskedelmi forgalomba hozott laparoszkópos rendszereknek, és biztonsági okokból ilyenkor elengedhetetlen a sebészi készségfelmérés. Dolgozatomban bizonyítottam, hogy az optikai áramlás egy olyan képjellemző, amely korrelál a sebész technikai készségeivel. Javasoltam egy kameramozgás automatizálási módszert továbbá a dVSS-hez. Végül validáltam egy optikai áramlási sajátmozgás kompenzációs módszert, amely által elérhetőek a sebészeti eszközök mozgása, és kizárja a kamera mozgását az optikai áramlási vektormezőből. Téziseim széleskörben támogatják a RAMIS szenzoros adat alapú technikai és nem-technikai készségfelmérés automatizálását, amely a következő lépés lehet a sebészeti adatok mélyebb megértése felé.

Contents

Acknowledgment

I thankfully acknowledge the professional and financial support from the Doctoral School of Applied Informatics and Applied Mathematics and the New National Excellence Program of the Ministry of Culture and Innovation.

I am indebted to my supervisor, Prof. Dr. Tamás Haidegger, for his continued guidance and endless support. Without his dedication, this work would not have been possible. His scientific thinking is an inexhaustible inspiration.

I would like to thank the colleagues of the Antal Bejczy Center for Intelligent Robotics for their constant help. I am grateful to be a part of this research group, and I would like to thank the director of the laboratory, Prof. Dr. Péter Galambos, who made it possible.

The help of the laboratory's surgical and industrial robotics research groups was essential in this work. I thank the hardware and software support in the da Vinci Surgical System for Tamás D. Nagy, and the valuable comments for Kristóf Takács. I thank Dr. Kristóf Móga for providing medical background as a surgeon for my research. I am grateful to work with Artúr I. Károly, Sándor Tarsoly and Dr. József Kuti in projects. Without the hardware knowledge of Alexander Ládi, László Szűcs and Tivadar Garamvölgyi this work would not have been possible. I thankfully acknowledge the work of my students: Gábor Lajkó, Dóra Papp, Cecília Molnár, Dávid El-Saig, Pálma Inczeffy, Berta Mach and Eszter Lukács.

I thankfully acknowledge the work of the Defense Committee and the Opponents of the thesis: Prof. Dr. Miklós Kozlovszky, Dr. Tamás Ungi and Prof. Dr. Márta Takács.

I am grateful to the acting dean of John von Neumann Faculty of Informatics, Dr. György Eigner for his support and faith in my research and teaching capabilities.

I would like to express my sincere gratitude to Dr. Gernot Kronreif, the Chief Scientific Officer at the Austrian Center for Medical Innovation and Technology (ACMIT), for his invaluable clinical insights and guidance that have shaped the direction of my future research.

I would like to thank my friends and my brother for their kindness and understanding. I would like to express my deepest gratitude to my loving husband for his unwavering support throughout my academic journey. I am especially grateful to my daughter, who is yet to be born but has already brought so much joy and strength into my life.

Structure of the Thesis

The thesis is consists of six chapters.

Chapter 1 gives an overview on Robot-Assisted Minimally Invasive Surgery and surgical skill assessment. Surgical skill assessment is divided into two main research domains: technical and non-technical approaches.

Chapter 2 collects the challenges in the main topical areas of the dissertation, highlighting why these problems require a scientific solution utilizing novel approaches. The problems stated in this chapter are related, but not restricted to technical and non-technical surgical skill assessment. The aim of my work is to propose a solution to the challenges of these domains.

Chapter 3, 4, 5 are covering the topics of the three major thesis groups, introducing the core research of my Ph.D. work. The chapters independently address the problems stated in Chapter 2. All of these chapters start with theoretical background, followed by methodology, development and experimental validation. The results and the evaluation of the findings are discussed at the end of each chapter.

Finally, Chapter 6 gives a structured summary of the key results of my research, providing an outlook on the current and future efforts that can utilize the findings of this work.

Numbering of equations, tables and figures is following the structure of the chapters. The independent references are numbered as [1],[2],..., thesis-related own publications are denoted as [RNE-1],[RNE-2],..., while the own publications that are not related to this thesis are numbered as [RNENR-I],[RNENR-II],... The language of the dissertation is English, following the U.S. English grammar and spelling rules.

Frequently used Notations and Symbols

TABLE 1: Common abbreviations and notations

AI	Artificial Intelligence
AR	Augmented Reality
CG	Control Group
CIS	Computer-Integrated Surgery
ConvAuto	Convolutional Autoencoder
CNN	Convolutional Neural Network
CSRT	Channel and Spatial Reliability Tracker
DCT	Discrete Cosine Transformation
DFT	Discrete Fourier Transformation
DoA	Degree of Autonomy
DoF	Degrees of Freedom
DNN	Deep Neural Network
DVRK	Da Vinci Research Kit
dVSS	da Vinci Surgical System
ECM	Endoscope Camera Manipulator
FCN	Fully Convolutional Neural Network
FDA	United States Food and Drug Administration
FLS	Fundamentals of Laparoscopic Surgery
FPS	Frames per Seconds
FN	False Negative
FP	False Positive
FRS	Fundamentals of Robotic Surgery
GEARS	Global Evaluative Assessment of Robotic Skills
HMM	Hidden Markov Model
HR	Heart Rate
ICARS	Interpersonal and Cognitive Assessment for Robotic Surgery
iDT	Improved Dense Trajectory
IEC	International Electrotechnical Commission
IoU	Intersection over Union
ISO	International Organization for Standardization
JIGSAWS	JHU–ISI Gesture and Skill Assessment Working Set
KT	Knot-tying
LC	Laparoscopic Cholecystectomy
LDA	Linear Discriminant Analysis
LDLJ	Natural log of Dimensionless Jerk
LOOCV	Leave One Out Cross-Validation
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MES	Medical Electrical System
MICCAI	The Medical Image Computing and Computer Assisted Intervention Society
MIS	Minimally Invasive Surgery

MP	Medical Professionals
MSE	Mean Squared Error
NASA-TLX	NASA Task Load Index
NN	Neural Network
NP	Needle-passing
NTS	Non-Technical Skills
NOTSS	Non-Technical Skills for Surgeons
OF	Optical Flow
OR	Operating Room
OSATS	Objective Structured Assessment of Technical Skill
PCA	Principal Component Analysis
PnP	Perspective n Point Transformation
PSM	Patient Side Manipulator
RAMIS	Robot-Assisted Minimally Invasive Surgery
ResNet	Residual Neural Network
ROI	Region of Interest
ROS	Robot Operating System
SURG-TLX	Surgical Task Load Index
SA	Situation Awareness
SDS	Surgical Data Science
SPARC	Spectral Arc Length
ST	Suturing
STIP	Space Temporal Interest Points
SVM	Support Vector Machines
TCP	Tool Center Point
TN	True Negative
TP	True Positive
VR	Virtual Reality
2/3D	Two/Three Dimensional

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 A brief introduction to Computer-Integrated Surgery

The history of surgical robotics started in the early 1960s – the US National Aeronautics and Space Administration (NASA) and the US Department of Defense, Defense Advanced Research Projects Agency (DARPA) laid the foundations of telerobotic surgical systems. Originally, their aim was to provide medical assistance for astronauts or wounded soldiers during their remote missions. For this, teleoperated robots would have been used, operated from the Earth. Mainly because of costs,at the end of the cold war, the technology was released for commercial purposes, the attention from telesurgery in space shifted to shorter distance telesurgery solutions, and soon, the first surgical robot prototypes received their Food and Drug Administration (FDA) clearance, and entered the U.S. market in the year 2000 [18].

The basic concept of Computer-Integrated Surgery (CIS) can be fitted to the Computer-Aided Design/Computer-Aided Manufacturing (CAD/CAM) paradigm known from the manufacturing industry, which involves the data–model–plan–execution–evaluation cycle, where surgical robotics takes the most important role in the execution step [1, 19] (Fig. 1.1). The evolution of advanced information sources mainly images, (such as endoscopic, Computed Tomography, Magnetic Resonance Imaging, etc.) and the development of robotic devices led to the concept of robot-assisted or robot-executed surgeries (Fig. 1.2). However, against the everyday terminology, surgical robotics does not necessarily mean high level of autonomy: there are devices, which only performs low-level, assistance-based automation, such as the market leading da Vinci Surgical System (dVSS, Intuitive Surgical Inc., Sunnyvale, CA). Advanced image-guided surgical systems, such as CyberKnife (Accuray Inc., Sunnyvale, CA) can perform high-level autonomy [20].

Pre-operative and intra-operative surgical planning in CIS (in most of the cases) is based on the human operator, however, with autonomous image segmentation, registration, classification and diagnosis techniques, the accuracy can be increased, and it can also decrease the workload of the surgeon [21] (Fig. 1.1). In the case of execution, autonomy, Augmented Reality (AR)/Virtual Reality (VR), tracking, navigation and advanced robot control can be an added value in the future. For the future CIS concepts, the intra-operative evaluation might include decision support, intelligent re-planning, data acquisition and image processing steps. The post-operative analysis can be extended with autonomy assessment, motion analysis and autonomous surgical skill assessment. With advanced algorith-

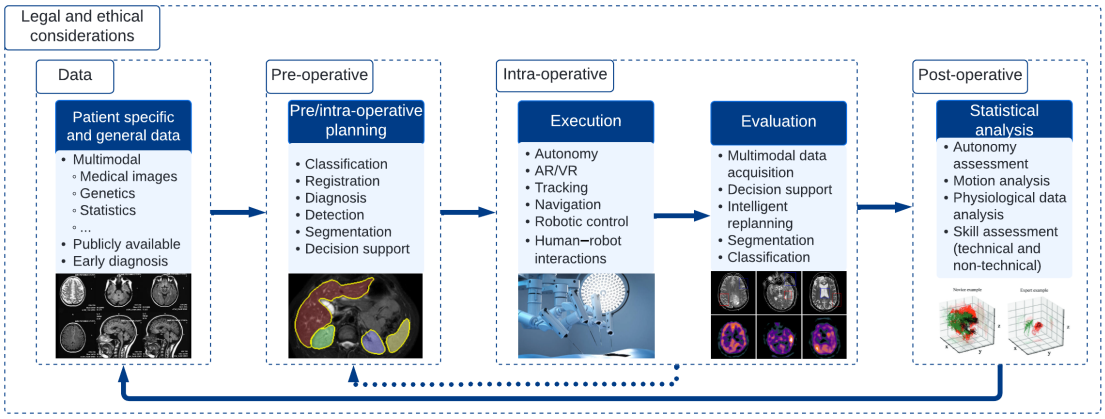

Fig. 1.1. The future concept of robotic surgery with advanced algorithm and AI-based support presented with the surgical CAD/CAM model, based on the original model by Taylor et al. [1].

mic and Artificial Intelligence (AI) approaches, higher accuracy and automated surgical skill ealuation can be achieved in surgical robotics, which can improve patient safety [22].

1.1.1 Robot-Assisted Minimally Invasive Surgery

Minimally Invasive Surgery (MIS) induced a paradigm change in medicine; however, it presented new challenges for surgeons [23, 24]. In the case of MIS – against traditional, open-access surgery – inside organs are reached through small skin incisions with laparoscopic instruments, and the operating area is visualized with an endoscopic camera. During MIS, the operator (surgeon) has to work in a team as a leader, s/he gives instructions to a camera handler assistant and the other operating room members, while s/he has to constantly monitor the operating area typically on a 2D screen in an uncomfortable position. Thus, despite the clear benefits of MIS, including the smaller scars and faster recovery time, there are drawbacks for the physicians, such as the limited motion space, complicated instrument control, not ergonomic environment and the 2D endoscopic camera image.

Robot-Assisted Minimally Invasive Surgery (RAMIS) was the next step in the evolution of MIS: it provided an improved vision system, more accurate and intuitive instrument control and an ergonomic master console [20, 25]. The most successful RAMIS system is the dVSS, which is a teleoperated, master-slave type surgical robot. In the case of the dVSS, the surgeon sits at an ergonomic master console, where he can operate with intuitively-controlled master arms. The surgeon can use pedals for clutch, and to control the endoscopic arm, thus camera control is only in the hands of the surgeon. At the patient side of the da Vinci, there are the remotely controlled (“slave”) arms, which accomplish the interventions minimally invasively with a motion mechanism called “Remote Center of Motion” (RCM), which can add to patient safety. The assistant crew works at the patient side of the da Vinci, where they can help the surgeon and support the intervention, such as changing the surgical instruments during the operation. At the patient side, there is a 3D endoscopic camera, through which images are visualized in the screens placed in the master console; thus, the surgeon can see a magnified 3D image of the operating

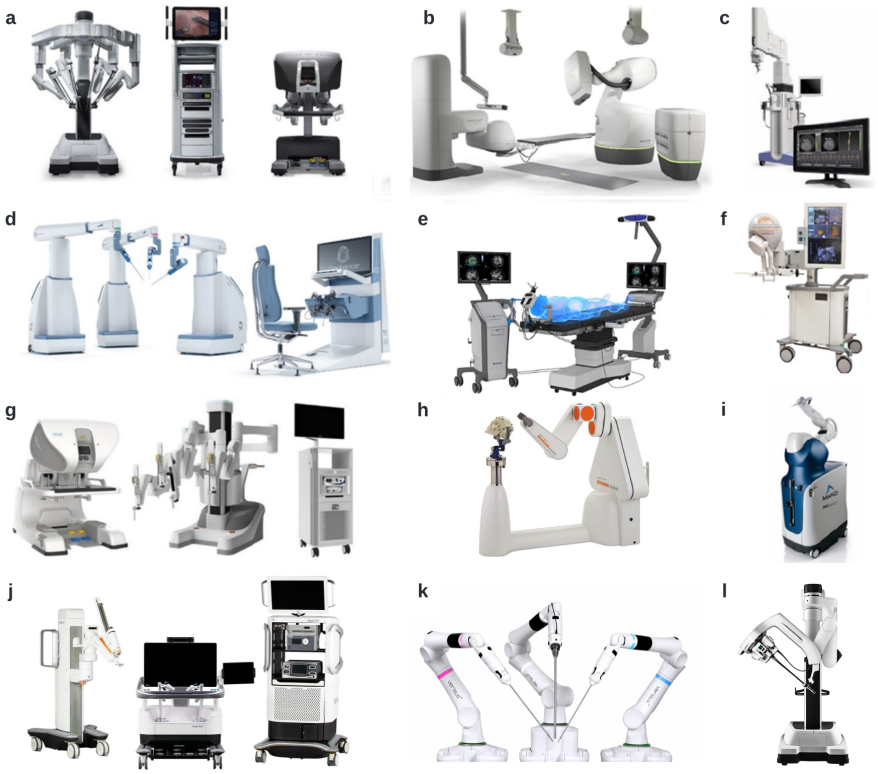

Fig. 1.2. Commercialized, ready-to-launch and research robotic surgical systems along the three main control types (teleoperation, cooperative, image-guided). A) Da Vinci Xi (Intuitive Surgical Inc.), b) CyberKnife (Accuray Inc.), c) TSolution One (THINK Surgical Inc.), d) Senhance Surgical Robotic System (Asensus Surgical), e) iSYS/Micromate/Stealth autoguide robot (Medtronic), f), Eigen ARTEMIS (Eigen Health) g), Revo-i (Meerecompany), h) Neuromate (Renishaw), i) Mako (Stryker Corporation), j) Hugo RAS system (Medtronic), k) Versius (CMR Surgical), l) da Vinci SP (Intuitive Surgical Inc.).

area. The motion of the surgeon can be re-scaled on the patient side of the da Vinci, which can provide more accurate motion. However, the original idea of remote surgery was to operate over long distances; for safety reasons, at the moment it is not part of the clinical practice. The dVSS does not present automation or decision making, the only very low-level automation in the dVSS is tremor and abrupt motion filtering. A steep learning curve has been identified with the da Vinci [26, 25]. Thus, despite the fact that RAMIS can decrease the mental workload of the surgeon as shown through by studies, RAMIS remains a challenging operation to perform not just physically, but mentally as well, because of the constant communication, teamwork, leadership, decision making and workload conditions [27, 28, 29].

1.2 RAMIS skill training and assessment

It may be important to evaluate surgical skills for quality assurance reasons, when that becomes part of the hospital's quality management system. Most commonly, only the proof of participation at theoretical and practical training is required. Arguably, objective feedback could assist trainees and practicing surgeons as well in improving their skills

SKILL LEVEL / Mental Function	NOVICE	COMPETENT	PROFICIENT	EXPERT	MASTER
Recollection	**Non-situational** →	**Situational** →	Situational	Situational	Situational
Recognition	Decomposed	**Decomposed** →	**Holistic** →	Holistic	Holistic
Decision	Analytical	Analytical	**Analytical** →	**Intuitive** →	Intuitive
Awareness	Monitoring	Monitoring	Monitoring	**Monitoring** →	**Absorbed**

Fig. 1.3. The Dreyfus model of skill acquisition. It defines 5 expertise levels and shows the differences between their qualities [2].

along the carrier. The fundamental challenge with skill assessment is that traditionally, the patient outcome used to be the only objective metric, and given the amazing variety and individual characteristic of each procedure, it has been really hard to derive distinguishing skill parameters. The subjective evaluation provided by other experts did not make it easy to compare results and metrics, therefore more generally agreed, standardized evaluation practices and training platforms had to be developed. A good example for this is the Fundamentals of Laparoscopic Surgery (FLS), a training and assessment method developed by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) in 1997, and widely adapted: it measures the manual skills and dexterity of an MIS surgeon, and provides a comparable scoring [30]. A similar metric for RAMIS surgeons was recently introduced, called Fundamentals of Robotic Surgery (FRS) [31].

In general, to understand the notions of 'skill' and 'skill assessment', the Dreyfus model [32] can be considered. The Dreyfus model refers to the evolution of the learning process, and it describes the typical features of the expertise levels (Fig. 1.3). For example, a novice (in general) can only follow simple instructions, but an expert can better react to previously unseen situations. In the literature, other skill models can be found, such as the classic Rasmussen model, which was created for modeling skill-, rule- and knowledge-based performance levels [33]. An other approach for modeling skills was introduced by Azari et al., which is specifically created for modeling surgical performance domains (Fig.1.4) [3]. RAMIS provides a unique platform to measure parameters which can help us in defining these skill levels objectively, since it makes low level motion data and spatial information available. Finding the proper parameters and algorithms that define the surgical skills is crucial in assessment [34].

There are several console training methods for RAMIS, which can provide the required practice for the surgeon [35]:

- virtual reality simulators;
- dry lab training;
- wet lab training;
- training in the operating room with a mentor.

Each has their own advantages and disadvantages, but from the clinical applicability point of view, the most important question is how fairly do these represent surgical skills.

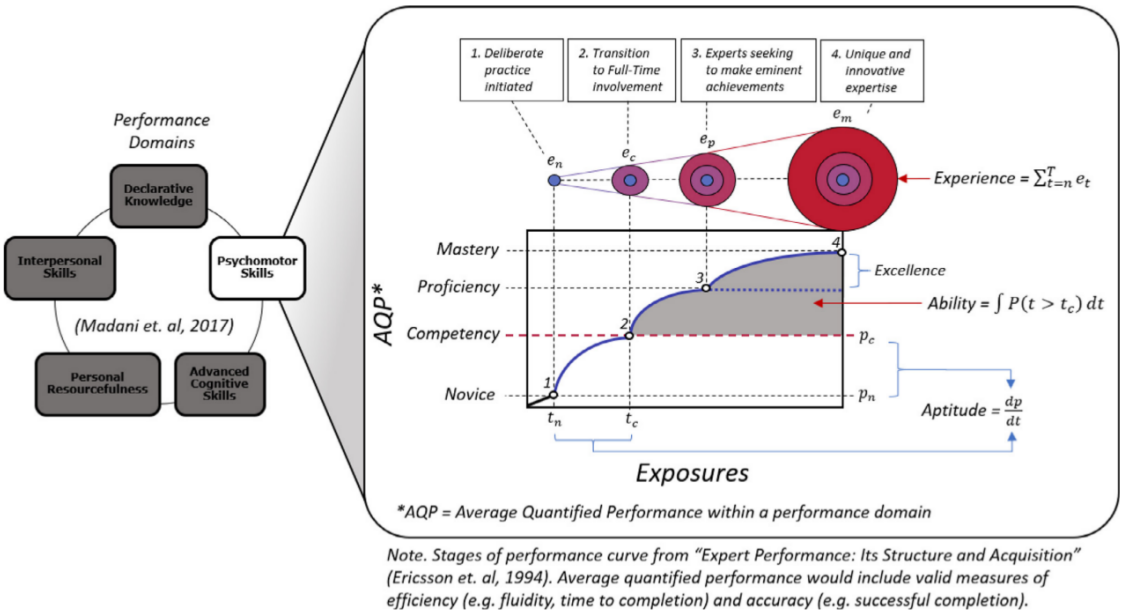

Fig. 1.4. Quantified performance model for surgical skill performance. The model describes the terms of 'skill': experience, excellence, ability and aptitude. "Ability" can be defined as the cumulative result of all actions that a surgeon has previously shown competence in, essentially represented as the area under continuous performance curves. On the contrary, "aptitude" refers to the speed or slope of this curve from the moment deliberate practice starts to when a surgeon consistently demonstrates competence. Meanwhile, "experience" can be visualized as an ever-expanding domain and depth of familiarity, encompassing the total number of distinct encounters and deliberate practice instances throughout a surgeon's professional journey [3].

Nowadays, there is still no objective surgical skill assessment method used in the operating room (OR) beyond board examination more experienced surgeons may provide some feedback, but rarely quantify the skills of their colleagues.

1.2.1 RAMIS technical and non-technical skill assessment

The improvements of RAMIS can help the surgeon, however, RAMIS is still a hard task to master; continuous training and feedback about the performance is crucial. Technical skills in RAMIS are related to the basic skills of the surgeon (knowing the instruments, using the right tools, etc.), the control of the robot and MIS tools (bimanual dexterity, endoscopic camera handling, clutch handling, instruments kept in view, etc.) and tissue handling (force sensitivity). Nevertheless, non-technical skill assessment is less exact. Despite the fact that RAMIS can decrease the mental workload of the surgeon as shown through studies, RAMIS remains a challenging operation to perform not just physically, but mentally as well, because of the constant communication, teamwork, leadership, decision making and workload conditions (Fig. 1.5) [27, 28, 29].

The workload on the surgeon – which represents the effort to perform a task – can be high in several segments of a procedure: there are mental, physical and temporal demands; furthermore, task complexity (including multitasking, task novelty), situational stress and distractions can influence the outcome of the surgery [14, 36] (Fig. 1.6). Naturally, the

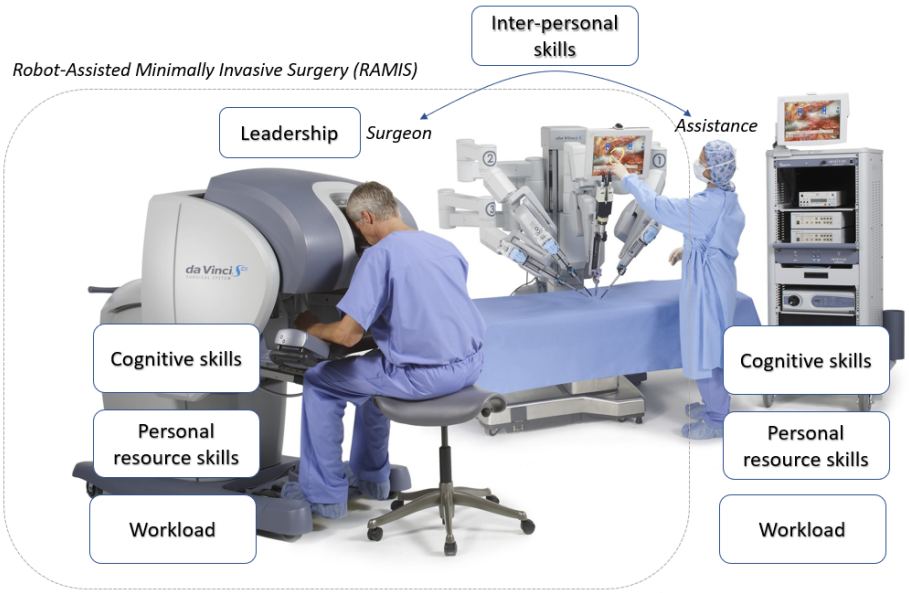

Fig. 1.5. The da Vinci S with the identified non-technical skills and workload. The surgeon operates at the master side of the system, while the assistance helps their work at the patient side. The patient side arms are controlled by the surgeon with the master arms. Robot-Assisted Minimally Invasive Surgery requires not just technical skills, but non-technical skills as well from the operating crew, specifically inter-personal skills, leadership, cognitive skills and personal resource skills, while they have to deal with the workload. Original image credit: Intuitive Surgical Inc. [4].

same task can cause different workload to different operators. Non-technical skills (NTS) related to the workload on the surgeon, furthermore, it can directly affect surgical outcome. NTS include communication, teamwork, task management, leadership, decision making, situational awareness and cope with stress, fatigue and distractions based on validated metrics, such as Non-Technical Skills for Surgeons (NOTSS) and Interpersonal and Cognitive Assessment for Robotic Surgery (ICARS) [16, 17] (Fig. 1.6). While it is straightforward that technical skills are crucial for better surgical outcomes, non-technical surgical skills can be as important as technical skills. Clinical failures in the operating room may come from low NTS of the surgeon than the lack of technical skills [37, 38, 39].

In the literature, three approaches for surgical performance assessment can be identified [40, 41]:

- self-rating questionnaires;
- expert-based scoring and
- automated skill assessment.

Questionnaires are filled by the operator; thus, it is easy to implement and subjective. Objective scoring is done by an expert panel, based on a standardized method [42]. Expert ratings are supposedly objective, yet may be biased for personal reasons, furthermore, they can be hard to implement, being human resource intensive. Automated skill assessment is based on objectively measurable parameters (such as applied forces, movement velocity, etc.), however, in most cases, it is technically not easy to implement. Robotic surgical systems can provide a unique platform for objective skill assessment due to the recordable kinematic and video data [20]. The mentioned surgical skill assessment approaches can be

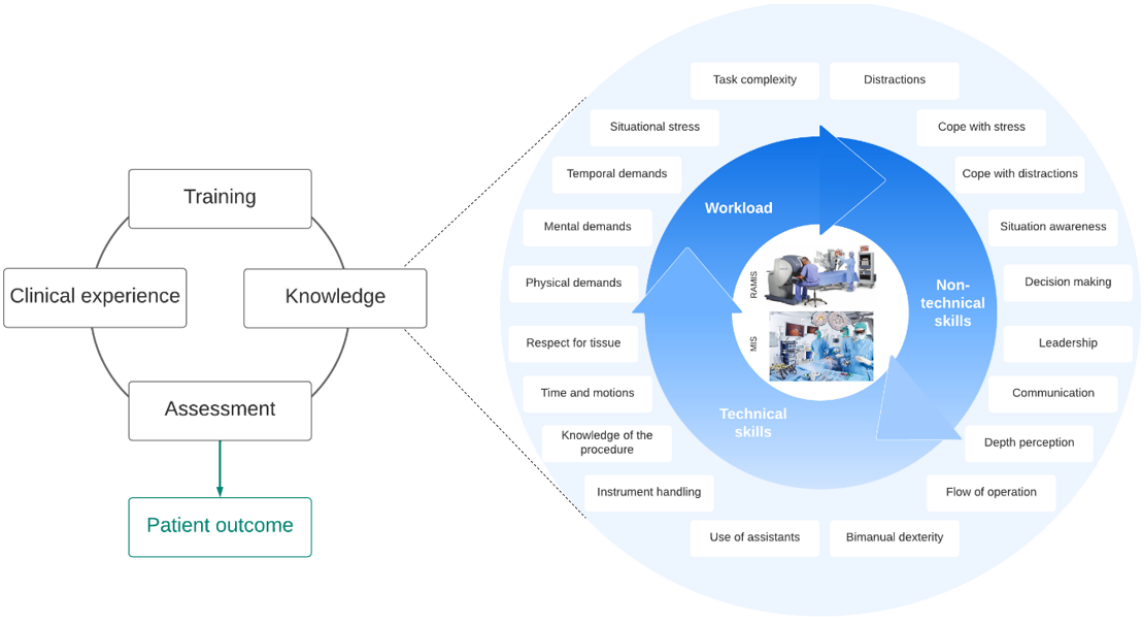

Fig. 1.6. Surgical skills are built of technical skills, non-technical skills and workload. Surgical skill training and assessment can directly affect the skills of the surgeon, and through patient outcome.

found in technical skill, non-technical skill and mental workload assessment as well. For mental workload assessment, questionnaires and automated solutions can be useful tools, and for technical and non-technical skill assessment all of the methods (questionnaires, expert-rating and automated techniques) can be utilized.

The difference between traditional MIS and RAMIS mental workload was examined in some studies [43, 44], demonstrating lower mental workload in the case of RAMIS. However, questionnaires created for traditional MIS were used in these studies, the main workload parameters in RAMIS are not yet defined. For RAMIS, non-technical skill assessment expert-rating methods originally created for traditional MIS can be found [45, 46]. There is one metric specifically created for RAMIS non-technical expert-rating assessment (ICARS, [17]), which collects the most important NTS in RAMIS (Fig. 1.6). NTS are naturally hard to be measured automatically. The possibilities for automated RAMIS non-technical skill assessment are similar to traditional MIS, such as relying on physiological signals measured by additional sensors [47].

The goal of technical and non-technical skill assessment is to employ automated and objective methods to measure the skills of the surgeon; thus avoiding biased assessment and the need for human resources. The built-in sensors of RAMIS can significantly ease automated skill assessment, since there are recordable kinematic and video parameters of the surgery (such as tool trajectory, orientation, velocity, etc.), which can provide input for skill assessment algorithms (statistical analysis or AI methods), towards manual MIS, where these data are only available with additional sensors. The clinical dVSS is a closed system, therefore to analyze surgical data, external recorders are necessary, such as the da Vinci Research Kit (DVRK, developed by a consortium led by Johns Hopkins University and Worcester Polytechnic Institute), which can provide open-source hardware and software elements with complete read and write access to the first generation da Vinci arms [48].

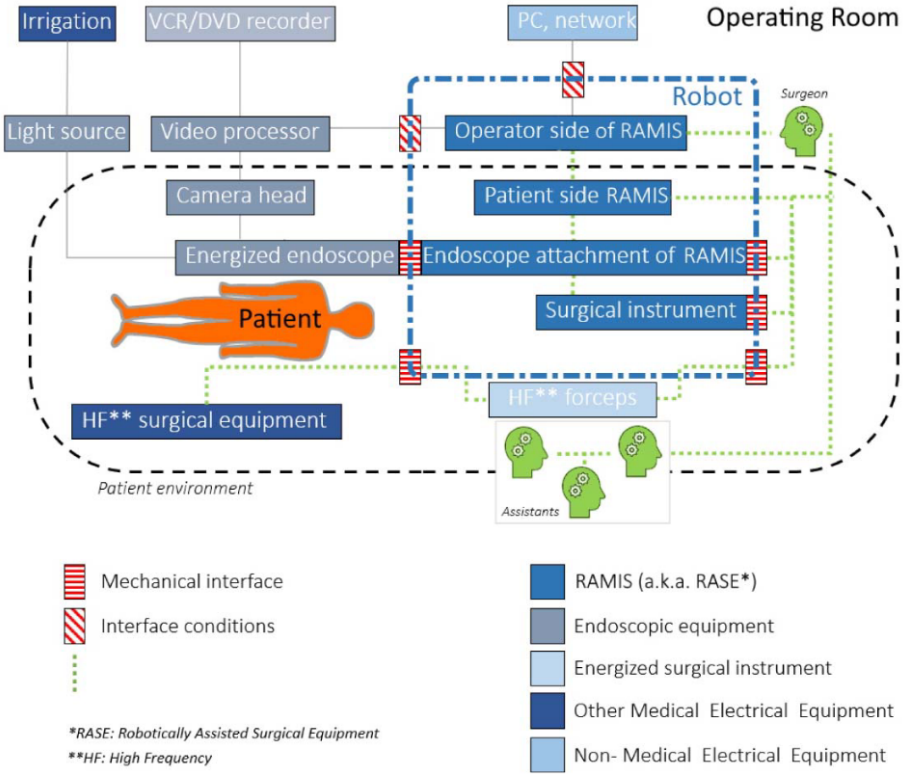

Fig. 1.7. Robot-Assisted Minimally Invasive Surgical systems architecture and typical layout diagram with the most important components in the case of non-technical skill assessment and mental load evaluation based on the IEC 80601-2-77 robotic surgery safety standard [5, 6].

To understand where NTS can be identified in the case of RAMIS, high priority (interaction and communication) channels and interfaces have to be identified and analyzed. The International Electrotechnical Commission (IEC) and the International Organization for Standardization (ISO) published a safety standard for surgical robots, the IEC 80601-2-77. In the standard, the components of RAMIS are defined, and a basic diagram of RAMIS is introduced [49, 5]. Based on the proposed working diagram, the most important components in non-technical skill assessment (Fig. 1.7) are highlighted. For this, the following definitions were used from IEC 80601-2-77, following the taxonomy of the IEC 60601-1 medical device core standard:

- **Robotically Assisted (or Robot-Assisted) Surgical Equipment – RASE**: '*Medical electrical equipment that incorporates programmable electrical medical system actuated mechanism intended to facilitate the placement or manipulation of robotic surgical instrument*'
- **Robotic surgical instrument**: '*Invasive device with applied part, intended to be manipulated by RASE to perform tasks in surgery*'
- **Interface conditions**: *Conditions that shall be fulfilled to achieve basic safety for any functional connection between RAMIS and other medical electrical equipment or non-Medical electrical equipment in the robotic surgery configuration*
- **Mechanical interface**: *Mounting surface on RAMIS that allows for attachment of detachable accessories, components, or parts that are mechanically manipulated by the RAMIS*

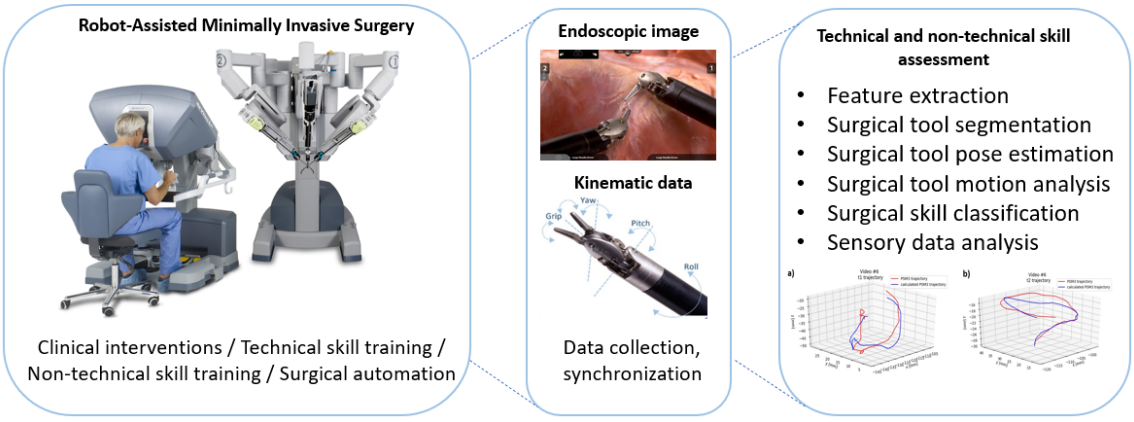

Fig. 1.8. Autonomous technical and non-technical surgical skill assessment input (endoscopic image and kinematic data) and steps. All steps under skill assessment are covered in this PhD thesis.

- **Endoscopic equipment**: '*Energized endoscope together with its supply unit(s), as required for its intended use*' [5, 49].

It is worth mentioning that the terminology of the ISO/IEC standard with respect to RASE slightly differs from RAMIS, mostly due to the fact that in the ISO sense, the term "robot" is defined with a much narrower meaning [50].

In Fig. 1.7, the components of RAMIS and the most important components in non-technical skill assessment are shown. Based on the literature findings, non-technical skill and workload can be assessed with the communication channel between the surgeon and the assistants, and with the cognitive and personal resource skills of the operating room crew, such as based on physiological signals or questionnaires, as it can be seen on the image, the surgeon's decisions are inseparable from the control loop of RAMIS systems. It suggests that NTS and workload might be shown in objectively measurable parameters, which means, non-technical skill assessment is not necessarily different from technical skill assessment in terms of the implementing approaches [51]. This may ease objective, automated non-technical surgical skill assessment in RAMIS. However, in the case of RAMIS, not many studies examined this correlation.

1.2.2 System design

In my PhD work, I have been focusing on autonomous technical and non-technical surgical skill assessment based on endoscopic image and kinematic data (Fig. 1.8). Sensory data in RAMIS can be originated from clinical interventions, technical or non-technical skill training, or even surgical subtask automation. The data should be properly collected and synchronized. Under autonomous skill assessment, image feature extraction, surgical tool segmentation, surgical tool pose estimation, motion analysis and skill classification techniques were considered.

Chapter 2

RESEARCH PROBLEM STATEMENT

Training and skill evaluation is crucial in the case of MIS, since it requires extensive practical skills as well, while surgical skill assessment is not yet part of the clinical practice. With autonomous technology-supported approaches, non-biased, objective surgical skill assessment can be achieved, furthermore, it does not require resources from human evaluations.

While kinematic data-based skill classification in RAMIS can achieve close to 100 % accuracy, but endoscopic images are more generally available in traditional MIS and in training videos, therefore there is a need for RAMIS technical skill assessment examined through endoscopic images. However, the accuracy of image-based surgical skill assessment is still below the kinematic data-based solutions. Surgical skill assessment has to be validated on an annotated database, but for the widely-used JHU–ISI Gesture and Skill Assessment Working Set (JIGSAWS), image annotation is not available, thus it is hard to validate image-based skill assessment.

- *Problem 1:* Image-based skill assessment's accuracy should be improved, because endoscopic camera images are the only widely accessible data. Semantic segmentation of the surgical tools on training videos should be available.

Non-technical surgical skill assessment is not a widely studied research domain, while clinical failures in the OR may just as commonly originate from low non-technical skills of the surgeon than the lack of technical skills.

- *Problem 2:* Non-technical skill needs training and evaluation. Non-technical surgical skills should be examined with autonomous techniques as well.

Automating the motion of the camera holder arm can decrease the cognitive workload on the surgeon, while surgical skill assessment is essential in this cases for safety reasons. However, image-based skill assessment can be complex if the camera is moving as well.

- *Problem 3:* Image-based skill assessment is necessary in the case of automated camera motion for safety reasons. A solution should be provided to filter the camera motion from the visual scene.

The problems identified above summarize three areas covering an important set of interconnected issues, and address scientific problems relevant to the clinical practice.

Chapter 3

IMAGE-BASED AUTOMATED TECHNICAL SKILL ASSESSMENT IN RAMIS

3.1 Methods for technical skill assessment in RAMIS

3.1.1 Manual assessment

In the case of manual RAMIS skill assessment, just like with traditional MIS, a team of expert surgeons in the OR evaluates the execution of the intervention based on their knowledge, the specific OR workflow and the expected outcome. This approach is easy to implement, yet, very costly (in terms of human resource and effort). It may be accurate averaged over multiple reviewers, but each individual assessment is quite subjective across boards, and it may be heavily distorted by personal opinions, and influenced by the level of expertise of that particular domain. The types of objective manual surgical skill evaluation in the case of RAMIS are *generic, procedure-specific* and *error-based* [52]. The simplest approach is the error-based manual assessment, because it only requires a typical error detection during the procedures. Procedure specific techniques examine the skills what needed in specific interventions. Generic manual skill assessment is the most complex, which evaluate the global skills of the surgeons.

A typical approach of manual RAMIS skill assessment is not to quantify the overall skills, just to evaluate particular skills needed in specific procedures, or only measure the errors made during the execution. In many cases, procedure-specific assessment is required, where the assessment metric is created for a specific surgical procedure (such as cholecystectomy, radical prostatectomy, etc.). Prostatectomy Assessment and Competence Evaluation (PACE) scoring is created for robot-assisted radical prostatectomy skill assessment. PACE metric includes the following evaluation points [53]:

- bladder drop;
- preparation of the prostate;
- bladder neck dissection;
- dissection of the seminal vesicles;
- preparation of the neurovascular bundle;
- apical dissection, anastomosis.

Cystectomy Assessment and Surgical Evaluation (CASE) is for robot-assisted radical cystectomy procedures. CASE evaluates the skills based on eight main domains [54]:

- pelvic lymph node dissection;
- development of the peri-ureteral space;
- lateral pelvic space;
- anterior rectal space;
- control of the vascular pedicle;
- anterior vesical space;
- control of the dorsal venous complex;
- apical dissection.

In the case of PACE and CASE, surgical proficiency was represented in every domain on a 5-point Likert scale, where 1 means the lowest and 5 means the highest performance (the score meaning is defined in every domain, such as injuries). Beyond these two specific methods, further scoring metrics for other interventions can be found in the literature [55, 56].

In most of the cases, any damage caused reflects the skills of the surgeons retrospectively: such as blood loss, tissue damage, etc. Generic Error Rating Tool (GERT) is a framework to measure technical errors during MIS; it was specifically created for gynecologic laparoscopy [57]. The validation tests showed promising results for the usability of GERT for objective skill assessment (its correlation to OSATS was examined) [58].

Generic manual assessment techniques evaluate the skills, based on the whole procedure/training technique, considering several points of the surgery, but not considering a specific technique. Global Evaluative Assessment of Robotic Skills (GEARS) was particularly created for robotic surgery, where expert surgeons assess the operator's robotic surgical skills manually. GEARS metric involves the assessment of the followings [35]:

- depth perception (from overshooting target to accurate directions to the right plane);
- bimanual dexterity (one from hand usage to using both hands in a complementary way);
- efficiency (from inefficient efforts to fluid and efficient progression);
- force sensitivity (from injuring nearby structures to negligible injuries);
- robotic control skills (based on camera and hand positions).

The surgical experts score the performance on a five scale score system. GEARS is a well-studied metric: validity tests and comparisons with GEARS can be found in the literature [35, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69]. GEARS showed results for the clinical usability and construct validity as well.

There are several modifications to the basic scoring skill assessment techniques, such as specified GEARS for endoluminal surgical platforms, called 'Global Evaluative Assessment of Robotic Skills in Endoscopy' (GEARS-E) [70]. GEARS-E is similar to GEARS, it measures depth perception, bimanual dexterity, efficiency, tissue handling, autonomy and endoscope control, but it was created for Master and Slave Transluminal Endoscopic Robot (MASTER) surgeries. GEARS-E is not yet widespread, because it is a relatively new technique, but the pilot study showed correlations to surgical expertise when using the MASTER.

Objective Structured Assessment of Technical Skills (OSATS) was originally created for evaluating traditional MIS skills along with FLS in 1997. OSATS involves the following evaluation points [71, 72]:

- respect for tissue (used forces, caused damage);
- time and motion (efficiency of time and motion);
- instrument handling (movements fluidity);
- knowledge of instruments (types and names);
- flow of operations (stops frequency);
- use of assistants (proper strategy);
- knowledge of specific procedure (familiarity of the aspect of the operation).

OSATS has an adaptation to robotic surgery: the Robotic Objective Structured Assessments of Technical Skills (R-OSATS) [73, 74]. The R-OSATS metric evaluates the skills of the surgeon based on the depth perception/accuracy, force/tissue handling, dexterity and efficiency. R-OSATS was tested typically with gynecology students, it has construct validity, and in the tests, both the interrater and intrarater reliability were high [72].

3.1.2 Virtual Reality simulators

VR surgical robot simulators primarily support training, they can also be a great tool to measure surgical skills objectively in a well-defined environment, since all motions, contacts, errors, etc. can be computed in the VR environment. A typical RAMIS simulator involves a master side construction and the virtual surgical task simulation. The master side is responsible for studying the usage of a teleoperation system (master arm handling, foot pedals, etc.), and to test the ergonomy. The simulation of the surgical task in the case of a surgical robot simulator has to looking life-like and be clinically relevant. During the training, the VR simulators often estimate the skills based on manual skill assessment techniques (such as OSATS), but in an automated way.

Since the dVSS dominates the global market, VR simulators are also focusing on da Vinci surgery. There are more than 5000 da Vinci simulators at the customer sites around the globe [75]. At the moment, there are six different commercially available da Vinci surgical robot simulators: the da Vinci Skills Simulator (Intuitive Surgical Inc.), dV-Trainer (Mimic Technologies Inc., Seattle, WA), Robotic Surgery Simulator (RoSS, Simulated Surgical Sciences LLC, Buffalo, NY), SEP Robot (SimSurgery, Norway), Robotix Mentor (3D systems (formerly Symbionix), Israel) and the Actaeon Robotic Surgery Training Console (BBZ Srl, University of Verona [76]). A novel surgical simulation program is the SimNow by da Vinci (Intuitive Surgical Inc.) [77]. SimNow involves surgical training using virtual instruments, guided and freehand procedure simulations and tracking skills and optimizing learning with management tools. In this section, the three most common types of VR simulators are reviewed: the DVSS, the dV-Trainer and the RoSS (Fig. 3.1).

DVSS can be attached to an actual da Vinci (da Vinci Xi, X or Si), with the main benefit that the surgeon can train on the actual robotic hardware, yet, it poses logistical problems, since while a trainee uses the simulator, the robot cannot be used for surgery. The dVSS contains the following surgical training categories [7]:

- EndoWrist manipulation;
- camera and clutching;
- energy and dissection;
- needle control;
- needle driving;

- suturing;
- additional games.

The dVSS measures the skills based on the economy of motion, time to complete, instrument collisions, master workspace range, critical errors, instruments out of view, excessive force applied, missed targets drops, misapplied energy time. The simulator costs about $85,000 [7, 35, 78, 79, 80, 81].

The dV-Trainer emulates the da Vinci master console, thus it operates separated from the actual da Vinci robot. It contains additional training exercises to the dVSS [7]:

- troubleshooting;
- Research Training Network (VR exercises to match physical devices in use by the research training network);
- Maestro AR (AR exercises that allow 3D interactions).

The dV-Trainer assesses skill with a very similar metric to the dVSS. In newer dV-Trainer versions, an alternative scoring system is available, called "Proficiency Based System", which based on expert surgeon data, and the interpretation of the data is different, furthermore the user can customize the protocol. The dV-Trainer costs about $96,000.

RoSS (as the dV-Trainer) is a stand-alone da Vinci simulator, involving numerous modules [7]:

- orientation module;
- motor skills module;
- basic surgical skills module;
- intermediate surgical skills module;
- blunt dissection and vessel dissection;
- hands-on surgical training module.

RoSS assesses the skills of the surgeon based on the camera usage, the number of left and right tool grasps, the distance while the left and right tool was out of view, the number of errors (collision or drop), the time to complete the task, the collisions of tools and tissue damage. RoSS costs about $126,000.

In the literature, most papers dealing with surgical robot simulators are focused on the curriculum and the technical layout, yet, in this work, the skill assessment and scoring part is crucial.

3.1.3 Automated assessment

Surgical robotics provides a unique platform to evaluate surgical skills automatically. RAMIS automated skill assessment does not need additional sensors to examine the surgeon's movements, camera handling, focusing on the image, etc., because these events, errors or movements can be recorded straight through the robotic control system. Automated assessment can be a powerful tool to evaluate surgical skills due to its objectivity, furthermore it does not require human resources, however, in some cases, it can be hard to implement these.

Two main types of automated skill assessment methods can be recognizable in the literature: *global information-based* and *language model-based* skill assessment. Global

Fig. 3.1. Virtual reality simulators for the da Vinci Surgical System [7, 8, 9]. A) da Vinci Skills Simulator (Intuitive Surgical Inc.), b) dV-Trainer (Mimic Technologies Inc.), c) Robotic Surgery Simulator (Simulated Surgical Sciences LLC), d) Robotix Mentor (3D systems), e) SEP Robot (SimSurgery), f) Actaeon Robotic Surgery Training Console (BBZ Srl).

information-based automated skill assessment means that the surgical skill is evaulated based on the whole procedure, based on the data of the endoscopic video, kinematic data, or other additional sensor data. The other approach is to evaluate skills on the subtask level, called language-model based skill assessment. Here, the first challenge is to recognize the surgical subtasks (often called ”surgemes”), then create a model for the procedure, and compare the models for skill assessment. Global skill assessment is easier to implement compared to language model-based techniques, but language models can be more accurate, and they are closer to the natural training (an expert will teach to the novice what was wrong on the subtask level, such as the way to hold the needle in a suturing task).

Data collection for automated assessment

The development of automated RAMIS skill assessment methods requires solutions for surgical data collection. The data – which correlates with the surgical skills – can be kinematic, video or additional sensor-based (e.g., force sensor). It is not trivial to access even training session data from RAMIS platforms. The dVSS has a read-only research

Fig. 3.2. JIGSAWS surgical tasks: knot-tying, suturing and needle passing (captured from the video dataset).

API (da Vinci Application Programmer's Interface, Intuitive Surgical Inc.), but it is only accessible to a very few chosen research groups. The da Vinci API provides a robust motion data set and it can stream the motion vectors, including joint angles, Cartesian position and velocity, gripper angle, joint velocity and torque data from the master side of the dVSS, furthermore events such as instrument changes [82].

To collect kinematic and sensory data from the da Vinci for research usage, the da Vinci Research Kit (DVRK) is a more accessible tool. DVRK is a research platform containing a set of open source software and hardware elements, providing complete read and write access to the first generation da Vinci [83]. DVRK is programmable via Robot Operating System (ROS) open source library [84]. The DVRK community is relatively small, but growing with around 40 DVRK sites currently [85].

While most of the da Vinci's have remote access and data storing enabled, due to legal and liability causes, clinical datasets are not available widely. In this case, annotated databases can provide input to RAMIS skill evaluation research. JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS) (developed by the LCSR lab at JHU and Intuitive) is an annotated database for surgical skill assessment, collected over training sessions [86]. JIGSAWS contains kinematic data (Cartesian positions, orientations, velocities, angular velocities and gripper angle of the manipulators) and stereoscopic video data captured during dry lab training (suturing ST, knot-tying NT and needle-passing NP). The dataset recorded on a dVSS involving surgeons with different expertise level (based on a manual evaluation technique). Beyond the manual skill annotations, JIGSAWS also includes annotations about the gestures ('surgemes').

Another approach is to capture surgical data with an additional data collecting device. A novel approach for da Vinci data collection, the dVLogger was developed in 2018 by Intuitive Surgical Inc. The dVLogger directly captures surgeons motion data on the dVSS. DVLogger can be easily connected to the da Vinci's vision tower with ethernet connection, and it records the data at 50 Hz. DVLogger provides the following informations from the robot [10]:

- kinematic data (such as instrument travel time, path length, velocity);
- system events (frequency of master controller clutch use, camera movements, third arm swap, energy use);
- endoscopic video data.

DVLogger can be a powerful tool in surgical skill assessment studies, due to its easy usage enabling the data collection for everyone, during live surgeries as well, however, it is a novel recording device, thus it is not widely recognized yet.

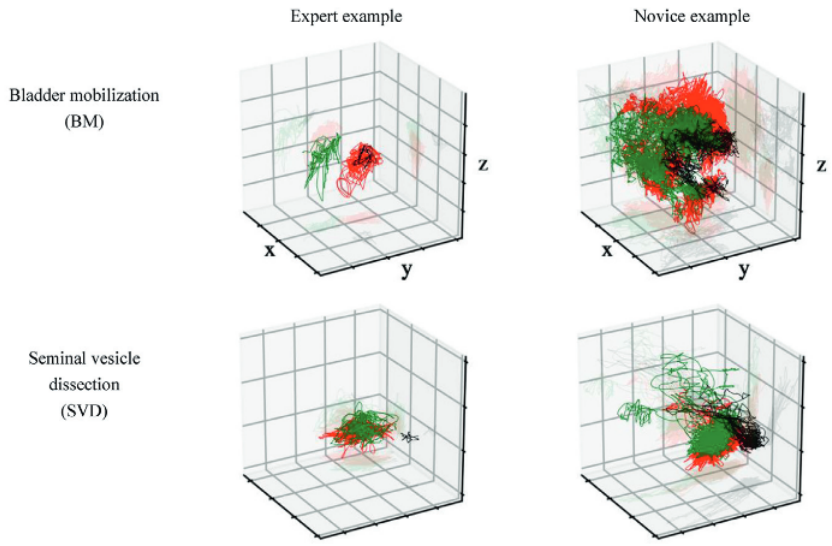

Fig. 3.3. A visual example of robot trajectories in case of a novice and an expert surgeon during robot-assisted radical prostatectomy (red: dominant instrument, green: non-dominant instrument, black: camera) [10].

SurgTrak (created by the University of Minnesota and University of Washington) is an additional hardware and software set which can be used for the da Vinci as well [87]. With SurgTrak, the endoscopic data can be captured from the DVI output of the da Vinci master side with an Epiphan DVI2USB device. The surgical instruments' position and orientation can be recorded with a 3D Guidance trakSTAR magnetic tracking system. Furthermore, grasper and wrist position is achievable with SurgTrak.

The above data collection techniques are useful for capturing kinematic and video data, but in some cases, other devices/sensors are needed to evaluate surgical skills with specific algorithms. Force sensors are often used in the field of surgical skill assessment. It is possible to estimate the used forces during the training based on the motor currents, but due to the construction of the da Vinci, data can be very noisy. A more popular approach is when an additional force sensor is used, such as developed at the University of Pennsylvania [88]. In this case, accelerometers were placed on the da Vinci master arms (which measured instrument vibrations), and a training board with a force sensor, which measured the forces during different types of training. They showed correlation between the measured data and the skill level.

Global information-based skill assessment

One approach for automated RAMIS skill assessment is to examine the whole procedure based on kinematic/video/additional sensor data. These methods are easier to implement than language model-based techniques, because they do not require the segmentation of the whole procedure (see details below). While global information-based methods are not sensitive to the performance quality of specific gestures, they can be as effective as language model-based techniques. There is an obvious correlation between the surgical skills and the kinematic data (Fig. 3.3), thus this is the most well-studied area in global information-based skill assessment [89, 90, 91, 92, 93, 11, 94, 95, 96, 97], but can find

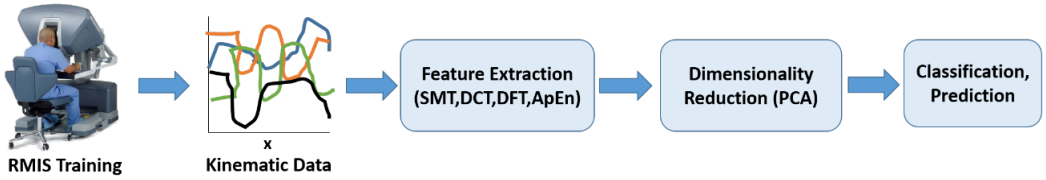

Fig. 3.4. Flow diagram for automated surgical skill assessment proposed by Zia et al. [11].

video, additional sensor-based [88, 98, 99], and the comparisons of several inputs [82, 100] automated techniques can be found as well. Global information-based skill assessment is not as deeply studied as language model-based methods, in general.

For the global methods, the classification of the input data is needed. A summary of these can be found in [11] (Fig. 3.4). The raw data (which can be any kind of data: endoscopic image, force, kinematic, etc. – in the figure you can find a specific example for kinematic-data based assessment) have to be processed with some kind of feature extraction technique, and in some cases, dimensionality reduction is needed as well. The processed data can be classified, and the skill can be predicted based on the extracted features from the data.

In [11], a motion-based automated skill assessment approach can be found. Their input was the JIGSAWS dataset. They used 4 types of kinematic holistic features: sequential motion texture, Discrete Fourier Transform, discrete cosine transform and approximate entropy. After the feature extraction and dimensionality reduction, they classified the data and predicted the skill score. The skill scoring was performed with a weighted holistic feature combination technique, which means that different prediction models were used to produce a final skill score. With this method a modified-OSATS score and a Global Rating Score was estimated. The results showed more accuracy than Hidden Markov Model-based solutions [11]. For more approaches, see Table 3.1.

Language model-based skill assessment

A surgical procedure model can be built up with different motion granularity. A surgical procedure (such as *Laparoscopic cholecystectomy*) is built from tasks (e.g., *exposing Calot's triangle*), which are built from subtasks (*e.g., blunt dissection*), which are built from surgemes (*grasp*), which can be translated to the components in the robotic domain (*motion primitives*) (Fig. 3.5). Global skill assessment methods approach the skill evaluation from the highest procedure/task level, thus not adverting the fact that surgical tasks are built from several, sometimes very different surgemes. These surgemes are not equally easy or complicated to execute, and even if a clinician believed to have intermediate skills based on a global skill assessment technique, they can be excellent/poor in just one, but very important surgeme and vice versa. Language model-based surgical skill assessment aims to assess surgical skills on the surgeme level, thus it requires three main steps: task segmentation, gesture recognition and gesture-based skill assessment. This approach has the further advantage that with the models defined, the transitions between the surgemes can be studied, and benchmark those as well. Language model-based skill assessment has been considered to be a cornerstone of the emerging field of Surgical Data Science (SDS) [101].

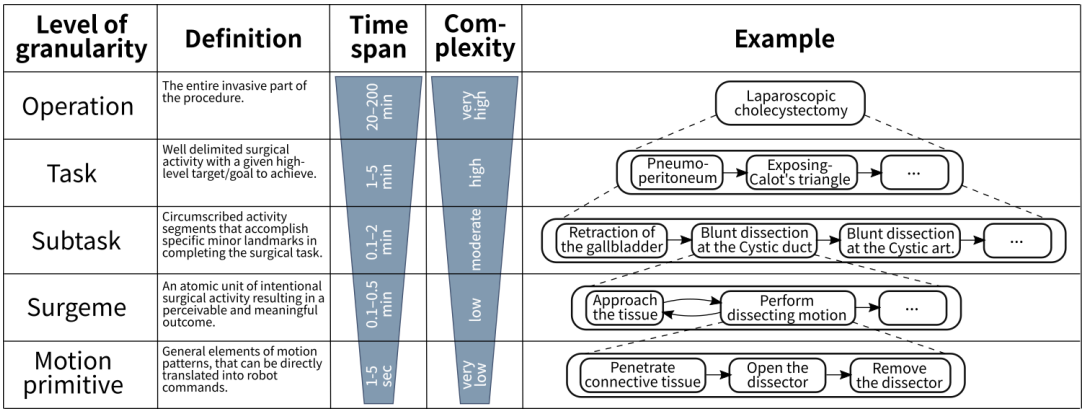

Fig. 3.5. A surgical procedure built from different levels [12]. Language model-based RAMIS skill assessment techniques typically evaluate the skills on the surgeme level.

It was the Johns Hopkins University Laboratory for Computational Sensing and Robotics (JHU LCSR) group who first proposed surgeme-based skill assessment [102], discrete Hidden Markov Models (HMM) were built for task and for surgeme level as well to assess skill. In practice, skill evaluation was based on a model built from annotated data (known expertise level), and this model tested against the new user. To create a model for user motions, they had to identified the surgemes with feature extraction, dimensionality reduction and classifier representation techniques. After that, the two models were compared. To train the discrete HMMs, they used vector quantization. Their method worked with 100 % accuracy using task level models and known gesture segmentation, at 95 % with task level models and unknown gesture segmentation, and at 100 % with the surgeme level models in correctly identifying the skill level.

The input of language model-based skill assessment methods can be kinematic data [103, 104, 102, 105, 106, 107, 108, 109, 110, 111], video data [112] or both [113, 114, 115, 116, 117]. In the literature, surgical activity/workflow segmentation can be found as well [118, 119, 120, 121, 122, 123, 124, 125]. For the details of the state of the art, see Table 3.1.

TABLE 3.1: Automated surgical skill assessment techniques in RAMIS. Used abbreviations: API – Application Programming Interface, HMM – Hidden Markov Model, LDA – Linear Discriminant Analysis, GMM – Gaussian Mixture Model, PCA – Principal Component Analysis, SVM – Support Vector Machines, LDS – Linear Dynamical System, NN – Neural Network.

Aim	Input data	Data collection	Training task	Technique	Ref.
kinematic data-based skill assessment	completion time, total distance traveled, speed, curvature, relative phase	da Vinci API	dry lab (bimanual carrying, needle passing, suture tying)	dependent and independent t-tests	[89]
framework for skill assessment of RAMIS training	stereo instrument video, hand and instrument motion, buttons and pedal events	da Vinci API	dry lab (manipulation, suturing, transection, dissection)	PCA, SVM	[82]
examine the effect of teleoperation and expertise on kinematic aspects of simple movements	position, velocity, acceleration, time, initial jerk, peak speed, peak acceleration, deceleration	magnetic pose tracker	dry lab (reach, reversal)	2-way ANOVA	[90]
longitudinal study tracking robotic surgery trainees	basic kinematic data, torque data, events from pedals, buttons and arms, video data	da Vinci API	dry lab (suturing, manipulation, transection, dissection)	SVM	[100]
generate an objective score for assessing skill in gestures	basic kinematic and video data	JIGSAWS	dry lab (suturing, knot-tying)	SVM	[115]
discriminate expert and novice surgeons based on kinematic data	completion time, path length, depth perception, speed, smoothness, curvature	da Vinci API	dry lab (suturing)	logistic regression, SVM	[91]
instrument vibrations-based skill assessment	completion time, instrument vibrations, applied forces	da Vinci API	dry lab (peg transfer, needle pass, intracorporeal suturing)	stepwise regression	[88]
automatic skill evaluation based on the contact force	contact forces, robot arm accelerations, time	da Vinci and Smart Task Board	peg transfer	regression and classification	[98]
skill assessment based on instrument orientation	time, path length, angular displacement, rate of orientation change	da Vinci Research Kit	dry lab (needle driving)	2-way ANOVA	[92]
discriminate expert and novice surgeons based on kinematic data	completion time, path length, depth perception, speed, smoothness, curvature, turning angle, tortuosity	da Vinci API	dry lab (suturing, knot-tying)	k-Nearest Neighbor, logistic regression, SVM	[93]
skill score prediction	sequential motion texture, discrete Fourier transform, discrete cosine transform and approximate entropy	JIGSAWS	dry lab (suturing, knot-tying, needle passing)	nearest neighbor classifier, support vector regression	[11]
objective skill level assessment based on metrics associated with stylistic behavior	basic kinematic and physiological data	limb inertial measurments unit, electromagnetic joint position tracker, EMG, GSR, IMU, cameras	da Vinci Skills Simulator tasks (ring and rail, suture sponge)	crowd sourced analysis	[99]
characterization of open and teleoperated suturing movement	speed, curvature, torsion of movement trajectories	da Vinci Research Kit, JIGSAWS	dry lab (suturing)	fitting the one-sixth power law, types of ANOVA	[95]
assess expertise and recognize surgical training activity	basic kinematic data	JIGSAWS	dry lab (suturing, knot-tying, needle-passing)	multi-output deep neural network architecture	[94]

Aim	Input data	Data collection	Training task	Technique	Ref.
evaluate skills based on kinematic data	time, errors, movement speed, jerkiness, trajectory redundancy, target scoring, trajectory volatility, max deviation	MicroHand S, magnetic sensor	dry lab (pick and place, ring threading)	one-way ANOVA	[96]
evaluate skills based on a deep learning model	basic kinematic data	JIGSAWS	dry lab (suturing, knot-tying, needle passing)	deep convolutional NN	[97]
gesture classification	basic kinematic data	da Vinci API	dry lab (suturing)	local feature extraction, LDA, Bayes classifier	[103]
gesture classification and recognition	basic kinematic data	da Vinci API	dry lab (suturing)	LDA, strawman GMM, 3-state HMM	[104]
compare task versus subtask	basic kinematic data	da Vinci API	dry lab (suturing)	vector quantization, HMM	[102]
gesture classification	basic kinematic data and video contextual cues (suture line deformations)	da Vinci API	dry lab (suturing)	HMM, high-order polynomial fitting to the extracted suturing line	[113]
gesture classification and recognition	basic kinematic data	da Vinci API	dry lab (suturing)	LDA, HMM	[105]
gesture classification	basic kinematic data	JIGSAWS	dry lab (suturing, knot-tying, needle passing)	sparse HMM	[106]
gesture classification	video features (image intensities, image gradients, optical flow)	JIGSAWS	dry lab (suturing, needle passing, knot-tying)	LDS, bag-of-features, multiple kernel learning	[112]
gesture classification	basic kinematic data and video features (Space-Time Interest Points)	JIGSAWS	dry lab (suturing, needle passing, knot-tying)	LDS, bag of features, multiple kernel learning	[114]
gesture classification	basic kinematic data	JIGSAWS, da Vinci Surgical System	dry lab (suturing)	descriptive curve coding, common string model, SVM	[107]
gesture classification	basic kinematic data	JIGSAWS	dry lab (suturing, needle passing, knot-tying)	Shared Discriminative Sparse Dictionary Learning, SVM, HMM	[108]
providing individualized feedback to surgical trainees	basic kinematic data	n/a	dry lab (suturing, knot-tying)	automatic identification of motifs in the tool motion signal	[119]
segmentation of surgical tasks into smaller phases	basic kinematic and video data	JIGSAWS	dry lab (suturing, knot-tying)	binary classifier, crowd-sourced segment ratings	[120]
unsupervised segmentation of surgical tasks into smaller phases	basic kinematic (position) and video (object grasp events and surface penetration) data	da Vinci Research Kit	dry lab (pattern cutting, suturing, needle passing)	milestone learning with Dirichlet Process Mixture Models	[121]
recognizing surgical activities	basic kinematic data	JIGSAWS	dry lab (suturing)	Recurrent Neural Network-based supervised learning	[123]

Aim	Input data	Data collection	Training task	Technique	Ref.
gesture classification and recognition	basic kinematic data	Raven-II, Sigma 7	peg transfer	unsupervised trajectory segmentation, k-Nearest Neighbors, SVM	[109]
describe differences in task flow	basic kinematic and video data	da Vinci API	dry lab (suturing, knot-tying)	hierarchical semantic vocabulary	[124]
gesture classification	basic kinematic and video data	JIGSAWS	dry lab (suturing, knot-tying, needle passing)	HMM, Sparse HMM, Markov semi-Markov Conditional Random Field, Skip-Chain CRF, Bag of spatiotemporal Features, LDS	[116]
temporal clustering of surgical activities	basic kinematic and video data	n/a	live surgery (two-handed robotic suturing, uterine horn dissection, suspensary ligament dissection, running robotic suturing, rectal artery skeletonization and clipping)	Hierarchical Aligned Cluster Analysis, Aligned Cluster Analysis, Spectral Cluster- ing, GMM	[125]
gesture classification and recognition	basic kinematic and video data	da Vinci SKILLS Simulator, SIMIMotion motion capture system	simulated tasks (peg transfer, pick and place)	Decision Tree Algorithm Model	[117]
gesture classification	basic kinematic data	JIGSAWS	dry lab (suturing, needle passing)	Transition State Clustering, uses hierarchical Dirichlet Process GMM	[110]
gesture classification	basic kinematic data	JIGSAWS/RAVEN-II, Sigma.7, leap motion device/dataset of micro-surgical suturing tasks captured using a dedicated robot	dry lab (suturing, needle passing, knot-tying/peg transfers/micro-surgical suturing)	Symbolic Aggregate approXimation, Bag of Words, vector space model	[111]
action segmentation and recognition	kinematic (end effector positions, velocity, gripper state, skip-length features) and video (distance to the closest object part from each tool, relative position of each tool to the closest object part) data	JIGSAWS	dry lab (suturing, needle passing, knot-tying)	Skip-Chain Conditional Random Field, Deformable Part Model	[118]

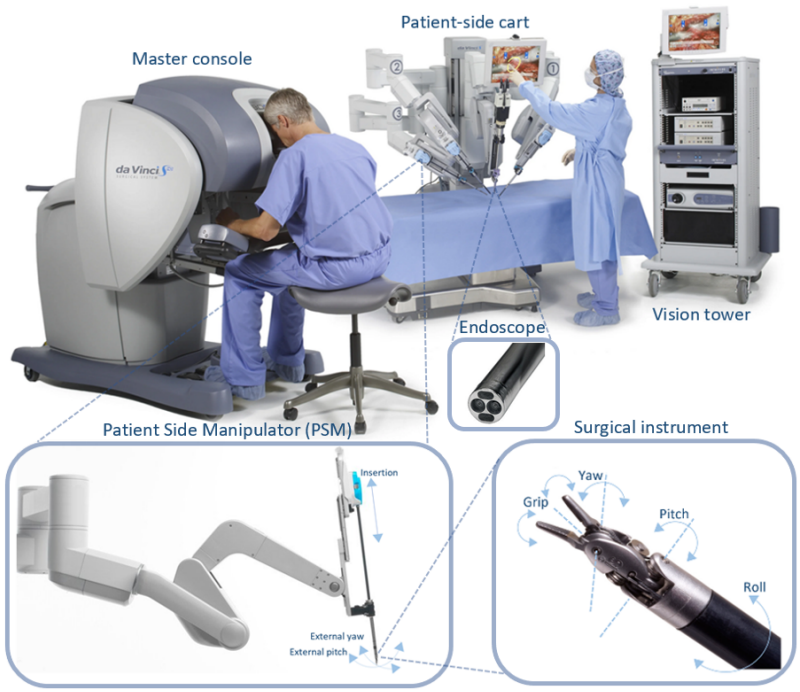

Fig. 3.6. The teleoperational dVSS with the surgeon-side master console, the vision tower and the patient-side cart with the articulated tools of the Patient-Side Manipulators (PSMs). The articulated tool has 6+1 Degrees of Freedom (DoF): insertion, external yaw, external pitch, roll, pitch, yaw and grip. The tools are visualized by an endoscopic camera.

3.2 Articulated tool pose estimation for autonomous image-based surgical skill assessment

In the case of classical MIS, or the widely available training and surgical videos, only 2D endoscopic image is accessible. In this Thesis, a tool pose estimation method for technical surgical skill assessment is introduced. The main reason for using generated image-based positions, is that kinematic data-based solutions typically provide much higher accuracy (up to 100 %) for skill classification, compared to endoscopic image-based solutions. The idea behind this study is that if accurate tool pose can be generated from the images, it can be a good alternative of kinematic data-based solutions.

The da Vinci Surgical System's tools have 6+1 Degrees of Freedom (DoF): insertion, external yaw, external pitch, internal yaw, internal pitch, roll and grip (Fig. 3.6). On the endoscopic image, the articulated surgical tool's 3 segments are visualized. The PSM's Tool Center Point (TCP) is at the gripper frame (where the yaw axis crosses the middle segment), and the orientation refers to the tooltip's orientation (in case of the tool is open, the orientation is the virtual middle point between the tooltips). For optimal tool pose estimation then, the TCP and the tooltip(s) have to be detected on the endoscopic image. The codes of this work can be found at https://github.com/ABC-iRobotics/end2kin.

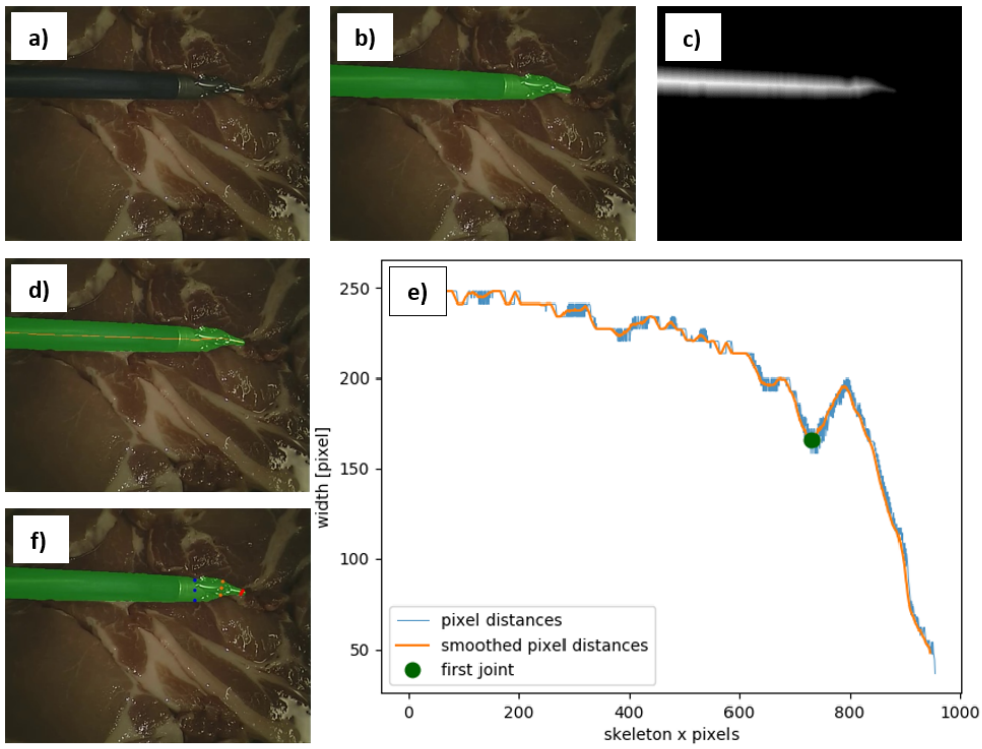

Fig. 3.7. Autonomous joint detection of a RAMIS tool (EndoWrist Large Needle Driver) on 2D endoscopic images. A) Original RGB image from the Synthetic MICCAI dataset; b) Ground truth; c) Distance transformed image from the instrument ground truth data; d) Skeleton of the surgical intrument; e) Width values along the skeleton's x axis, based on the distance transformed image. Blue line represents the original width data, orange line is the smoothed width data, and the green dot represents the first joint of the surgical tool based on prominence-based peak detection; f) Dots represents the joints and the tip points of the surgical instruments. From these 9 points, 6 DOF pose of the tool can be estimated.

3.2.1 Joint detection on RAMIS endoscopic images

At the moment, there are two main options to detect the TCP and the tooltip on an image: with classical image processing methods and Artificial Intelligence-based approaches (especially supervised Neural Network-based solutions). In the case of the first category, important features of the images are pre-defined, on the other hand, supervised NNs can learn these features from labeled samples. Since robotic surgical tools in most of the cases have well-defined shape features, the proposed RAMIS tool joint detection approach is based on classical image processing methods. Furthermore, in this thesis, tool joint detection and pose estimation are handled by a separate problem from surgical tool segmentation, thus all of the proposed method were tested on the Synthetic MICCAI ground truth data (Section 3.2.3). With a proper tool segmentation method, the proposed pose estimation approach can be done.The steps of the joint detection algorithm are the followings: first, on the captured ground truth data the distance transform of the mask can be calculated (Fig. 3.7). The distance transform is a derived representation of the image, where the pixel values represent the Euclidean distance from the nearest zero pixel. The skeleton of the tool can be calculated from morphological operations. By the product of the distance

transformed image and the skeleton, the widths of the surgical tool can be calculated along the skeleton. The assumption is that at the joints of the tools this product is less compared to other segments of the surgical tool due to its morphological features. However, this is a strong shape feature in most of the cases, for the first frame, the joints have to be visible, furthermore, there are tools for the dVSS where this shape feature is not applicable. Nevertheless, the thinner joint widths appear in most of the dVSS tools (blades, scissors, forceps, graspers, etc.). The first joint can be detected then based on width, with prominence-based peak detection after a smoothing filter. The second joint position can be detected based on the width as well. This can be done realtime, and in parallel the tools are tracked with sparse Optical Flow (Kanade–Lucas–Tomasi method). Optical Flow represents the real-world motion in the image scene. The pose estimation method requires 6 points on the image what the model can be fitted – it means, the two joint middle point is not enough for proper pose estimation. For this, the contour of the tools was calculated, and the nearest contour points from the joint middle points were found based on the Euclidean distances. From this, nine points on the two tool segments were detected; based on which the two segments' pose can be calculated. For all image frames (except the first), with Optical Flow the points' position can be validated. If more PSMs are visualized on the image, with connected components they can be handled separately.

3.2.2 Pose estimation

For tool pose estimation, due to robustness, the aim was to use a solution which does not require knowledge about the kinematic data, or a complete 3D model. Because of this, perspective *n* point (PnP) transformation was used for surgical tool pose estimation. For enough accuracy PnP requires at least six points physical positions on a model, and if these points can be properly found on the 2D image, this physical model can be fitted and the pose can be calculated. For this, camera calibration and distortion coefficients are necessary as well (however, PnP can provide acceptable results without it). PnP is responsible for calculating the pose (rotation and translation) between the world frame and the camera frame from N ($N \geq 3$) points. PnP problem can be formulated as using the pinhole camera model:

$$\begin{bmatrix} p_i \\ \hline 1 \end{bmatrix} = C[R, t] \begin{bmatrix} P_i \\ \hline 1 \end{bmatrix}, \tag{3.1}$$

where C is the camera calibration matrix, R is the rotation matrix, t is the translational vector, p_i is the 2D pixel coordinates and P_i is the corresponding 3D points. Based on the correspondence between p_i and P_i, t and R can be calculated [126]. For this, in this work, iterative PnP solving method was used, where Levenberg–Marquardt minimization scheme was employed.

In this work, the 6 point model was generated for the EndoWrist Large Needle Driver for both segments, what was used in Synthetic MICCAI dataset. This means 9 points: 3 for the first joint, 3 for the second joint and 3 for the tooltip (Fig. 3.7). It was necessary, because the first joint detection is very stable, but for orientation, the points of the second joint and the tooltip are necessary. These 3D points can be found on the 2D image with the proposed joint detection method. Further, translation and rotation of the tool can be calculated with PnP iterative method.

For the validation, joint angle representation of the Synthetic MICCAI dataset had to be converted to Cartesian coordinates, which was done with the DVRK package [12]. Based on the generated 3D points, with 3D rigid transformation the camera frame can be transformated to the robot's base frame. Since PSM TCP is not exactly on the second joint detected by the image processing method (due to the shape features of the tool), an offset should be added to the calculated position. With this, the generated and the original pose can be compared.

3.2.3 Synthetic MICCAI dataset

To test and validate the autonomous joint detection and pose estimation, a RAMIS video dataset was necessary, which is annotated not only with the segmentation ground truth, but the robot kinematic data as well. At the moment, Synthetic MICCAI dataset is the only RAMIS dataset annotated with respect to these two data [127]. The Synthetic MICCAI dataset (created by University College London) was recorded with the DVRK. In Synthetic MICCAI, EndoWrist Large Needle Drivers were used to perform a surgical movement. Kinematic data is represented with joint angles, 7xN matrix data for Patient Side Manipulators (PSMs) and 4xN for Endoscope Camera Manipulator (ECM). The rows refer to the joints and the columns represent consecutive time stamp. Synthetic MICCAI contains 15 scenario with different tool movements, each scenario contains 300 video frames. Videos were created with 30 Frame per Seconds (FPS).

3.2.4 JIGSAWS

Since Synthetic MICCAI dataset contains video and kinematic data of RAMIS scenes, furthermore it is annotated with the segmentation ground truth, in the meanwhile it does not contain the skills of the surgeons who performed the surgical motions, it is necessary to use a dataset where surgical skill annotations are available. JIGSAWS is a complex RAMIS skill assessment dataset, widely used for testing endoscopic image-based and kinematic data-based surgical skill assessment methods [86], see Section 3.1.3. JIGSAWS was captured using the dVSS, with 8 surgeons at different skill levels (expert, intermediate and novice), while performing well-known surgical training tasks (knot-tying, needle passing and suturing) with EndoWrist Large Needle Drivers. JIGSAWS not only contains the kinematic and stereo video data, but the skill level of the surgeons and the gesture annotations as well. While JIGSAWS's video data is available, it is not annotated with the instrument segmentation ground truth beyond the image frame resolution and the general video quality is poor. Because of these reasons, the kinematic data of JIGSAWS was only used in this work modified with the detected accuracy, to simulate image-based skill assessment.

3.2.5 Natural log of Dimensionless Jerk and Spectral Arc Length

Motion smoothness, what correlates with the precision and control of the motion, is a common feature of human motion studies. In surgical skill assessment, motion smoothness is shown to be an effective and sensitive parameter to classify the skills of the surgeons [128]. The goal of this work is to provide an endoscopic image-based generated tool pose, which can be utilized for automated image-based surgical skill assessment—and since motion

smoothness is an often-used parameter, the error of motion smoothness should not be significant.

If the motion is smooth, its acceleration does not have any discontinuities, which can be determined by derivative of the acceleration—it is called jerk, which is often examined in human motion studies. To eliminate jerk variations' (pure, dimensionless) inconsistency and wide variability, natural log of dimensionless jerk (LDLJ) was proposed [129].

For LDLJ, elapsed time (T) has to be calculated, where t_{end} and t_{entry} are the end and entry times, respectively:

$$T = t_{end} - t_{entry}. \tag{3.2}$$

Velocity profile (3.3) is based on the x, y and z directions of the movement velocity:

$$X = \sqrt{x^2 + y^2 + z^2}. \tag{3.3}$$

Path length (PL) (3.4) is calculated from the sum of Euclidean distances between points traversed:

$$PL = \int_{t_{entry}}^{t_{end}} \frac{dX}{dt}. \tag{3.4}$$

Then LDLJ can be calculated, which is the natural log of jerk integrated and squared:

$$LDLJ = -ln \left| \frac{T^5}{PL^2} \int_{t_{entry}}^{t_{end}} \left(\frac{d^3X}{dt^3} \right)^2 \right|. \tag{3.5}$$

The other motion smoothness parameter, which is robust to noise is proposed, called Spectral Arc Length (SPARC) [130]. SPARC is aimed to measure smoothness based on the fact smooth motion yields small magnitudes of low frequency profiles and vice versa. SPARC is derived from the arc length of the amplitude of the frequency-normalized Fourier magnitude spectrum of the velocity profile. SPARC is more robust regarding noise, compared to LDLJ.

$$SPARC = -\int_0^{\omega_c} \left[\left(\frac{1}{\omega_c} \right)^2 + \left(\frac{d\hat{A}(\omega)}{d\omega} \right)^2 \right]^{\frac{1}{2}} d\omega, \tag{3.6}$$

$$\hat{A}(\omega) = \frac{A(\omega)}{A(0)}, \tag{3.7}$$

where $A(\omega)$ is the Fourier magnitude spectrum of the acceleration signal $a(t)$, $\hat{A}(\omega)$ is the normalized magnitude spectrum.

3.2.6 Results

Tool pose estimation and motion smoothness

To test the accuracy of the pose estimation method, Mean Squared Error (MSE) between the PSM3 (left PSM) kinematic data and the estimated image-based kinematic data was calculated. Furthermore, the generated 3D position trajectories' smoothness was calculated with the natural log of Dimensionless Jerk (LDLJ) and Spectral Arc Length (SPARC)

based on the velocity profile. Since Synthetic MICCAI dataset contains not only PSM3 movement videos, but PSM1 and PSM3+PSM1 data as well, in this work, for simplicity, only PSM3 motion was considered. Furthermore, the data were cut when PSM3 was covered by PSM1, or did not show the TCP well at the beginning of the video. Based on this assumptions, 8 videos were chosen to be examined, wherein 1906 frames were tested for accuracy.

The MSE for translation x, y, z and rotation x and y were calculated for all scenarios (Table 3.2, Fig. 3.9). The minimum translation MSE (which is the maximum accuracy) in the case of x direction was 1.92 mm in the case of video #1, in y direction minimum error was 1.52 (video #1), and 1.19 mm for z direction (video #2). The maximum error was not greater than 6.44 mm, 7.60 mm and 6.35 mm for x, y and z directions, respectively (Fig. 3.8). Due to the fact that the proposed silhouette-based approach does not provide information of roll direction of the tool's rotation, this error was not calculated. However, this can significantly decrease the rotation accuracy along x and y axes (ϕ, θ) as well: the best-case ϕ error was 42.38° and 7.54° for θ, but there were MSE observations up to 125.76° as well. Based on these findings, 3D position estimation can provide accurate results, which can be a good input for endoscopic image-based skill assessment methods, but rotation estimation is not accurate enough to use it for such assessments. Because of this, only the 3D position used for further examinations.

TABLE 3.2
TRANSLATION AND ROTATION MEAN SQUARED ERROR OF THE ESTIMATED 5 DOF POSE ON THE EXAMINED VIDEOS.

Video #	**Mean Squared Error**				
	T. x [mm]	***T. y [mm]***	***T. z [mm]***	***R. ϕ [°]***	***R. θ [°]***
1	1.92	1.52	2.21	84.46	**7.54**
2	4.04	1.47	1.19	49.02	64.40
6	4.12	2.65	2.33	**41.38**	13.36
7	4.74	5.11	4.22	123.34	49.95
9	3.51	3.99	3.86	59.15	47.55
11	5.83	5.02	6.15	88.38	31.05
12	3.15	7.60	5.28	125.76	63.45
14	6.44	6.46	6.35	118.92	44.35

To analyze motion smoothness, the generated image-based and the original kinematic data were segmented to smaller portions of motion data. Since the Synthetic MICCAI dataset does not contain typical surgical tasks (such as suturing), decomposition of the videos for smaller surgical motions (such as surgemes) was not applicable, thus every motion what performed in 25 frames was considered. The MSE (Table 3.3) for LDLJ was 0.00 in the best case and 2.16 for the worst. In the case of SPARC, the best MSE result was 0.00 and 1.51 the worst. Since these results (besides 0.00) are not easy to understand, Wilcoxon signed rank tests were ran on robotic and image-based LDLJ and SPARC motion segments, which compared the original and generated trajectories. The results showed no significant difference between the original and the calculated LDLJ and SPARC results, and since skill classification studies showed significantly different LDLJ

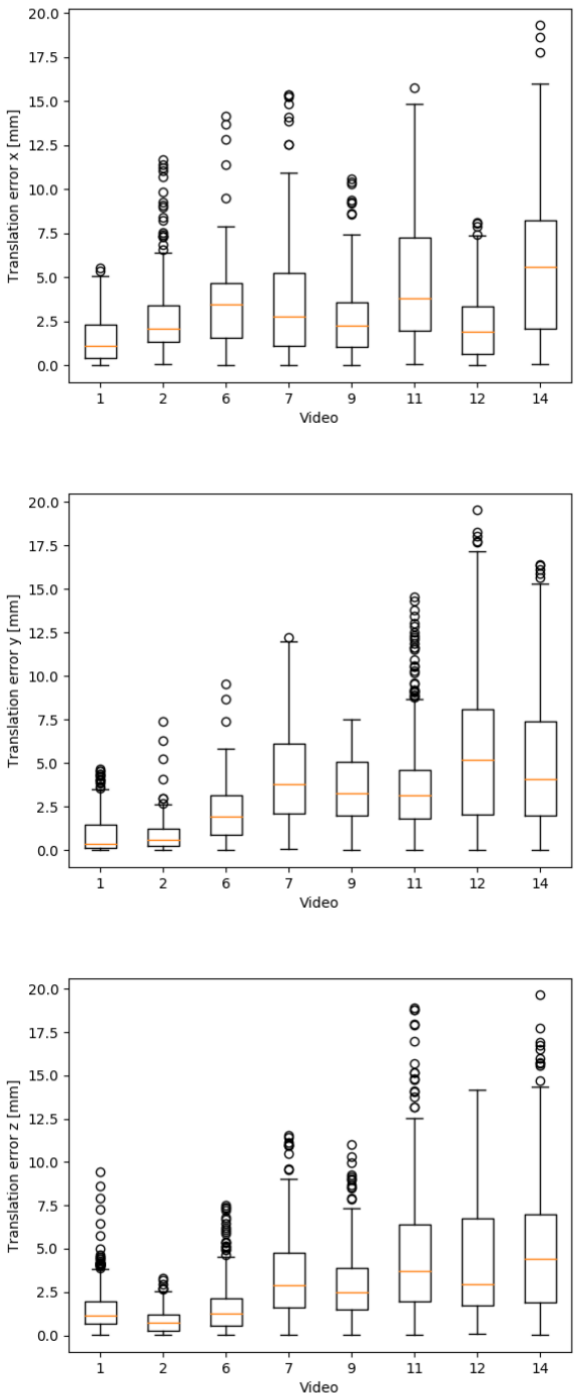

Fig. 3.8. Absolute translation error x, y and z components for all test video sequences involved in the assessment from the Synthetic MICCAI dataset.

and SPARC results, it can be concluded that these results are good enough to use the generated data for skill assessment.

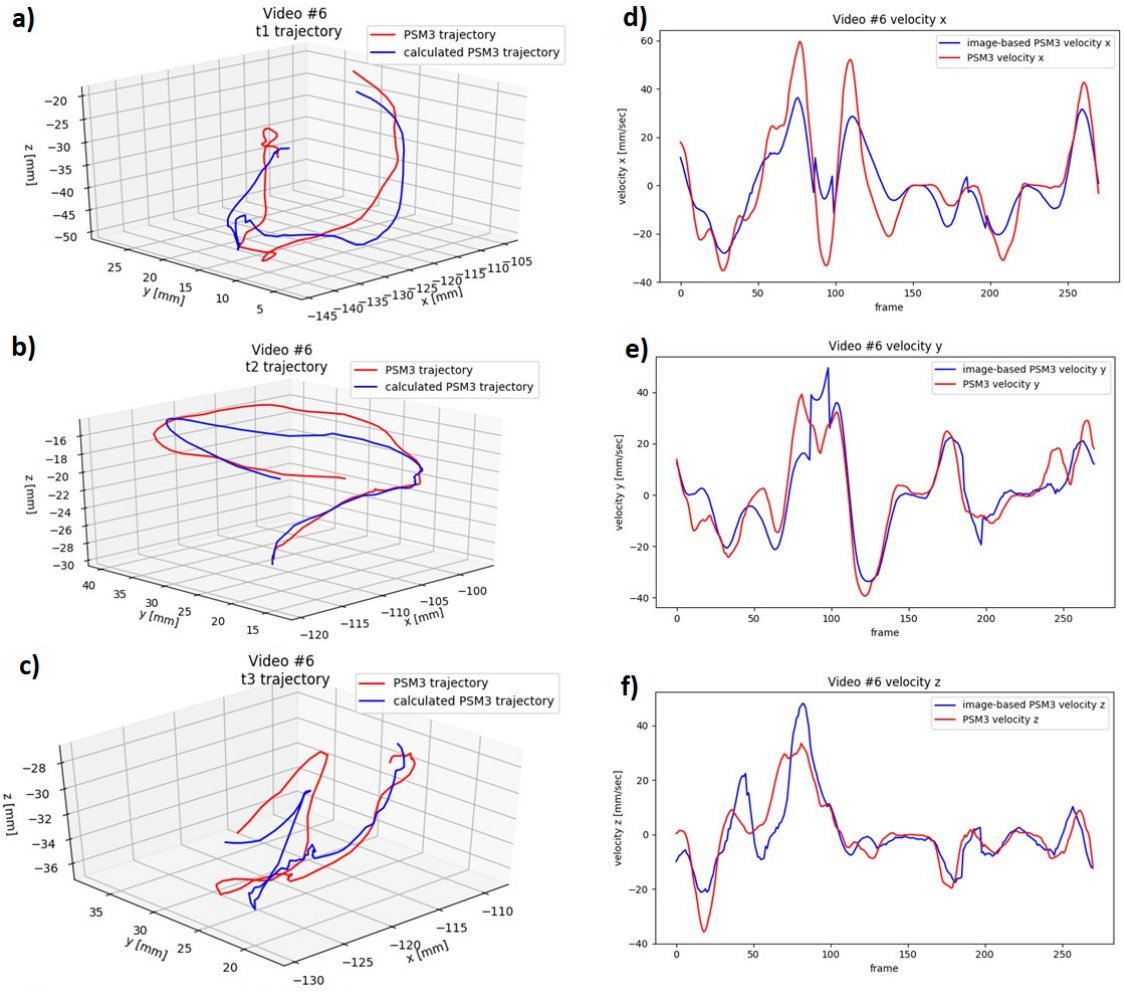

Fig. 3.9. Generated trajectories by image-based surgical tool pose estimation and velocity components of it. First row: one test video's three trajectory segments (a – c); blue: image-based positions, red: kinematic data-based positions; Second row: velocity x, y and z components (d – f) of the same motion. The similarity of the trajectories were proven with 3D position accuracy and velocity-based motion smoothness metrics.

Skill classification on the JIGSAWS dataset

As it was mentioned earlier, JIGSAWS endoscopic data does not have enough image quality or ground truth data for this proposed tool pose estimation method, however, my work was focusing on ground truth data and surgical tool segmentation for the JIGSAWS in the following sections. After the tool pose estimation accuracy was calculated, similar noise was added to JIGSAWS kinematic data for suturing, knot-tying and needle passing surgical tasks. In JIGSAWS, novice and expert surgeon's data were used for skill classification. Because the estimated 3D position data was more precise than the 6 DoF pose estimation, it was only used for skill classification. From JIGSAWS, orientation and other kinematic data (velocity, acceleration, etc.) were excluded. Because only one PSM's data was examined here, from JIGSAWS only PSM3's 3D position data was used for skill classification [131].

For skill classification of the 3D position data, a Random Forest-based time series supervised classifier was used. Time Series Forest (TSF) classifier is an interval-based

TABLE 3.3
NATURAL LOG DIMENSIONLESS JERK (LDLJ) AND SPECTRAL ARC LENGTH (SPARC) MEAN SQUARED ERROR (MSE).

Video #	Mean Squared Error	
	LDLJ	***SPARC***
1	1.98	1.63
2	0.00	0.00
6	2.05	0.76
7	2.11	0.87
9	1.83	0.44
11	1.28	1.51
12	1.13	0.86
14	2.16	0.89

classifier, which splits the series into intervals and then extract mean, standard deviation and slope features. The decision tree then trained to these features, and repeat the steps until the expected trees have been built [132]. 75 % of the data was randomly selected for training and 25 % for testing. Random noise was added to 3D position data with 0 mm mean and 4.22 mm, 4.23 mm and 3.95 mm standard deviation for x, y and z directions, respectively, which was the mean error of the proposed surgical tool pose estimation method. From sliding window cross validation, mean accuracy was 89.33 %, 76.66 % and 75.00 % for knot-tying, suturing and needle passing, respectively. These three surgical tasks-based skill classification accuracies are commonly different from each other (needle passing shows typically the best classification performance), in this case, probably because of the only one tool-based classification, knot-tying showed the highest accuracy. While it outperformed the solution proposed in Chapter 5 in knot-tying (Table 3.4), it provided lower accuracy for suturing and needle-passing.

3.2.7 Conclusion of surgical tool pose estimation

In this section, a surgical tool pose estimation technique was introduced for the da Vinci Surgical System's articulated tools, targeting an algorithmic support for autonomous technical skill assessment. The tool pose estimation was performed on 2D endoscopic images, based mainly on shape features and iterative Perspective n Point transformation method. The introduced technique was validated on the Synthetic MICCAI dataset, where it provided 4.22 mm, 4.23 mm and 3.95 mm x, y, z RMSE average accuracy of 8 videos, which contained 1906 frames. The smoothness of the generated trajectories were examined with the natural log of Dimensionless Jerk and Spectral Arc Length, where there were no significant differences between the image-based and the original trajectories. The experienced noise was then added to JIGSAWS kinematic data, and skill classification was done with time series forest classifier. The results were 76.66 %, 75 % and 89 % accuracy for suturing, needle-passing and knot-tying, respectively. Based on these findings, image-based surgical skill assessment can be an acceptable alternative to kinematic data-based skill assessment in the case of MIS.

TABLE 3.4

COMPARISON OF THE RESULTS TO THE STATE OF THE ART. METHODS LISTED HERE USED JIGSAWS 2D ENDOSCOPIC IMAGE DATA AS INPUT, EXCEPT THE PROPOSED SOLUTION (NAGYNÉ ELEK ET AL.). BLUE DENOTES THE BEST RESULTS. STIP: SPACE TEMPORAL INTEREST POINTS, IDT: IMPROVED DENSE TRAJECTORY, RESNET: RESIDUAL NEURAL NETWORK.

Author (Year)	Method	Suturing	Needle-Passing	Knot-Tying
Ming et al. (2021)	STIP[133]	79.29 %	87.01 %	72.57 %
Ming et al. (2021)	iDT[133]	76.79 %	83.81 %	76.65 %
Lajkó et al. (2021)	ResNet[134]	81.89 %	84.23 %	83.54 %
Nagyné Elek et al. (2022)	hybrid	76.66 %	75.00 %	89.33 %

3.3 Surgical tool segmentation on the JIGSAWS dataset

As it was mentioned in Section 3.2.4, JIGSAWS dataset is the most widely used dataset for surgical skill assessment, yet, there has been no solution for semantic segmentation of the surgical instruments in it. In this Section, different Neural Networks were examined and applied to segment surgical tools in the JIGSAWS dataset aiming to lead image-based surgical skill assessment. Image processing techniques have some advantages over NN-based methods, since the latter needs a huge amount of data for training. Moreover, NNs are highly dependent on the data on which they were trained. The codes of this work are available at https://github.com/ABC-iRobotics/SurgToolSegJIGSAWS.

3.3.1 Ground truth generation

Optical Flow is a frequently used algorithm for object tracking and movement detection [135]. It creates a 2D vector space, where each vector represents the movement of a pixel between two frames. The main assumption of the algorithm is that the intensities of an object's pixels do not change significantly in between consecutive frames. Assuming that $I(x, y, t)$ is a pixel of a 2D image frame:

$$I(x, y, t) = I(x + dx, y + dy, t + dt), \tag{3.8}$$

where the dx, dy is the distance of a pixel's movement during dt time. The following equation can be created with Taylor series approximation, which is also called the Optical Flow equation:

$$f_x u + f_y v + f_t = 0, \tag{3.9}$$

where $f_x = \frac{\partial f}{\partial x}$, $f_y = \frac{\partial f}{\partial y}$, $u = \frac{dx}{dt}$, $v = \frac{dy}{dt}$ and f_x and f_y are image gradients. The u and v are the unknowns in this equation. Since this equation has two unknowns, it can not be solved directly, it needs further assumptions. There are different methods for solving it, one of them is Lucas–Kanade method, and the other is Gunner Farneback's algorithm [135]. The main difference between them that Farneback's algorithm (dense Optical Flow) make a motion vector for every pixel, while Lucas–Kanade (sparse Optical Flow) only calculates (3.8) for some pixels (with good features). Lucas–Kanade method takes a 3 x 3 patch around the selected pixels, thus there are 9 equations with two unknowns, which can be solved. With this algorithm, the selected pixels in consecutive frames can be tracked.

Fig. 3.10. Masked frames with the semi-automatic ground truth generation method on the JIGSAWS dataset (example image frames for Knot-Tying, Needle-Passing, Suturing).

Using the Optical Flow algorithm [135], ground truth generation for semantic segmentation can be done in a semi-automatic way. Based on the performance of the Optical Flow tracking in several trials, it can be concluded empirically that the algorithm was able to track the labelled pixels through 30 frames. Above 30 frames, the Optical Flow could not track the selected pixels. Therefore, only every 30^{th} frame needed to be segmented manually, which is not as time-consuming as hand-labelling every frame. Then, the labelled pixels were tracked with the Optical Flow algorithm. This method gave acceptable results, because after every 30^{th} frame there is again a hand-labelled frame, so if the algorithm lost the pixels, then with the hand-labelled frame it was fixed. After that, some classical post-processing steps were applied. Small objects which did not belong to the tools were removed and also morphological closing was applied. Then the holes in the surgical tools were filled (which were created by the tracking, since the Optical Flow tend to track only the edges). At last, median blurring was applied on the frames to smooth the edges. The masked frames can be found in Fig. 3.10.

With this method a ground truth dataset was created which contains semantic labelling of 9 videos, 3-3 from each type of surgical tasks (Knot-Tying, Needle-Passing, Suturing). These segmentations are not always perfect, the edges of the masks are sometimes not accurate. But these masks are still good enough to consider that it is much faster than to segment the tools manually.

3.3.2 Benchmark methods

Currently, in surgical tool segmentation, the best results are achieved with NN-based methods. As these problems are similar to semantic segmentation on the JIGSAWS dataset, the most frequently used neural network architectures were examined in tool segmentation. Data augmentation is a powerful technique in deep learning that leverages various transformations leading to better model performance and generalization. In surgical tool segmentation, a series of data augmentations commonly used in computer vision tasks were defined to enhance the robustness and diversity of training data for deep learning models. These augmentations include resizing and cropping, as well as horizontal and vertical flips. First, it pads the input images to ensure they meet a minimum height and width requirement, followed by a random crop to the specified training crop size. Random cropping helps introduce variability into the training data. Additionally, horizontal and vertical flips with 50% probability are applied to simulate different viewing angles and orientations.

These augmentations help the neural network learn to recognize objects in various positions, orientations, and scales. Finally, normalization is performed to standardize the pixel values, aiding in the convergence and stability of the training process. These augmentations collectively contribute to improving the model's ability to generalize and perform well on a wide range of real-world images.

UNet

U-Net type architectures have a contracting path, which follows a typical convolutional network architecture and an expansive path which uses upsamplings. In the contracting path, there are 3 x 3 convolutions, which are followed by a rectified linear unit (ReLu) and 2 x 2 max pooling units for down-sampling. In the expansive path, the feature maps are up-sampled by up-convolutions (2 x 2), and after that, there are always 3 x 3 convolutions, which are followed by ReLU. There is a 1 x 1 convolution in the final layer to map the feature vectors to the given number of classes. In this architecture, there are also skip-connections between the contracting and expansive paths [136, 137].

LinkNet-34

LinkNet is also a part of the U-Net family, it has a similar architecture. The only difference is that it uses a ResNet-type architecture as an encoder. In LinkNet34, a pre-trained ResNet34 is used. The contracting path starts with 7 x 7 convolution with stride 2, and then a max-pooling layer. After that there are residual blocks, which contain two 3 x 3 convolutions, the first one with stride 2, the second one with stride 1. In the expansive path, there are decoder blocks which are connected to the corresponding encoder blocks. The decoder blocks consist of 1 x 1 convolution layer, followed by batch normalization and transposed convolution. LinkNet has an advantages that it contains less parameters, thus it is faster than other neural network architectures [138].

TernausNet-11

TernausNet is also a U-Net like architecture, which uses pre-trained VGG11 encoders. VGG11 has seven convolutional layers which are followed by ReLu activation function and five max-pooling layers, which reduce the feature maps by 2. All of the convolutions are 3 x 3. This architecture also worked well on aerial images of urban settlements [139, 140].

TernausNet-16

The difference between UNet11 and UNet16 is that they use different pre-trained encoders UNet16 uses VGG16 [141, 142].

3.3.3 Segmentation with Pre-Trained Networks

Pre-trained models on the JIGSAWS dataset were tested, which were trained on dataset from the MICCAI 2017 Endoscopic Vision SubChallenge (Robotic Instrument Segmentation) [136, 143]. These pre-trained networks did not work acceptable on JIGSAWS dataset

based on empirical preliminary studies, but with pre-processing steps, the segmentation masks could be improved. The best pre-processing step was L*a*b colorspace conversion, which could magnify the difference between the background and the surgical tools, therefore the pre-trained networks could provide a better segmentation.

3.3.4 Training with the JIGSAWS dataset

With the ground truth of the JIGSAWS dataset, four different neural network architectures were trained, namely UNet, TernausNet-11, TernausNet-16 and Linknet-34. The same loss function for training was used as for the MICCAI dataset training [136]. This loss function is a combination of the Jaccard index and binary cross entropy:

$$L = H - log(J), \tag{3.10}$$

where H is the binary cross entropy and J is the Jaccard-index. The Jaccard-index also known as Intersection over Union (IoU) is a similarity measure for sample sets (in this case, two images). For discrete objects, it can be calculated in the following way:

$$J = \frac{1}{n}\sum_{n=1}^{n}(\frac{y_i\hat{y}_i}{y_i + \hat{y}_i - y_i\hat{y}_i}), \tag{3.11}$$

where y_i is a binary value (label) and $\hat{y}$ is the predicted values for i pixel. The minimization of this function means the maximization the intersection between the ground truth and the predicted mask and maximize the probabilities of right predictions at the same time [136, 137].

The output of these architectures is a probability for each pixel, which indicates that they belong to the surgical instruments or not. A 0.8 threshold was used, which means that above this value, the pixels are considered as a part of the surgical tools. This threshold value was chosen because it is quite high, so it means that these pixels have a high probability that they belong to the surgical instrument according to the trained network.

3.3.5 Results

Results of the different methods on different surgical tasks can be seen in Figure 3.11. The predicted masks were evaluated with the following metrics: Jaccard-index (or Intersection over Union), Sørensen–Dice coefficient (DSC) and accuracy. Jaccard-index was also used for the training as a part of the loss function, now the same metric is used for the evaluation of the results.
Accuracy can be calculated:

$$\text{ACC} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}, \tag{3.12}$$

where TP refers to true positives, TN is true negatives, FP is false positives and FN is false negatives (in this case, TP is the number of pixels, which belong to surgical tools, both in the ground truth and the segmented image, TN is the pixels which belong to the background both ground truth the segmented image, FP which are segmented as tool but

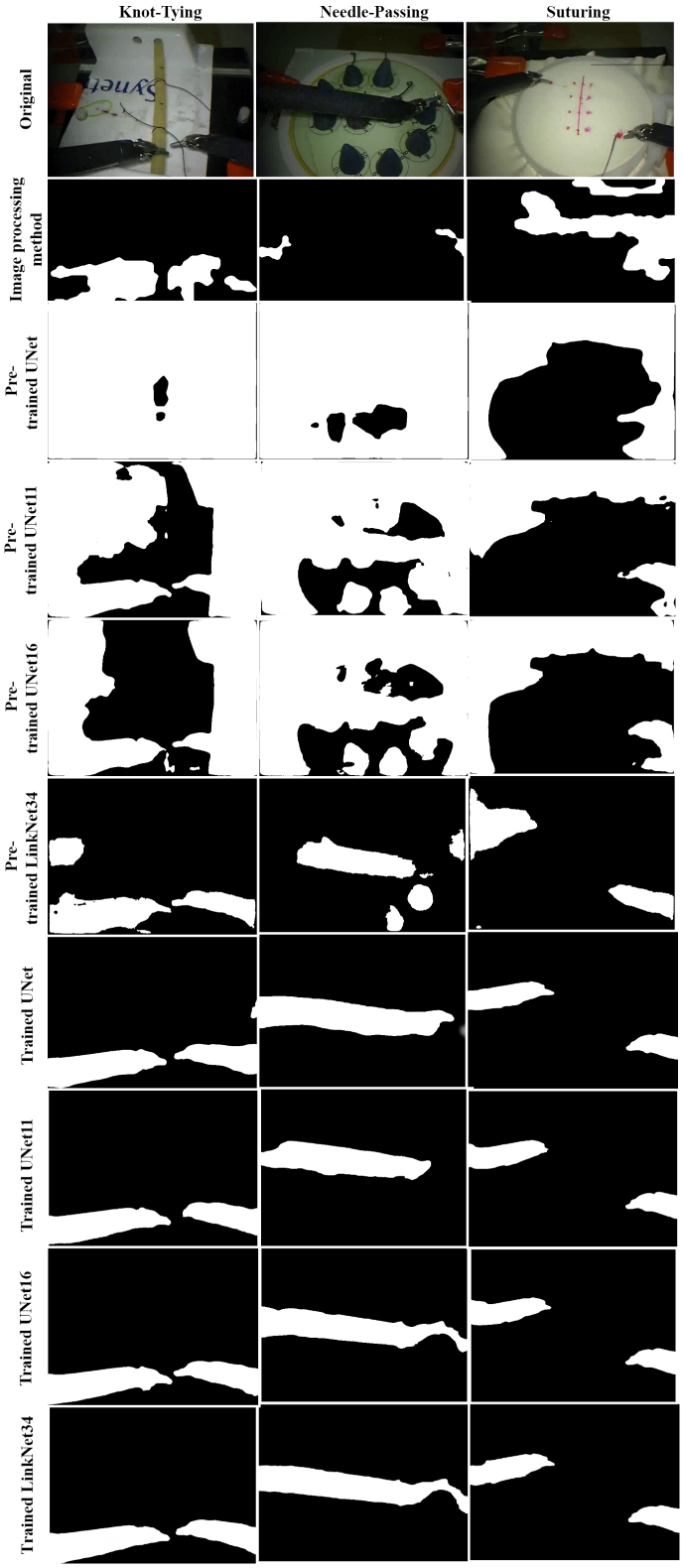

Fig. 3.11. Visual examples of the outcomes: the derived masks of the same frame for surgical tool segmentations (one example for each surgical tasks) by the examined methods.

do not belong to it, and FN are pixels, which were segmented as background, but belong to the objects). The Sørensen–Dice coefficient can be calculated with the following equation:

$$\text{DSC} = \frac{2|X \cap Y|}{|X| + |Y|}, \tag{3.13}$$

where X is set of the ground truth mask' pixels and Y is the set of segmented mask's pixels. So, in the numerator, the overlapping area is calculated and, in the denominator, the number of the pixels in both images is calculated.

TABLE 3.5
RESULTS OF DIFFERENT METHODS (PRE-TRAINED UNET, PRE-TRAINED UNET11, PRE-TRAINED UNET16, PRE-TRAINED LINKNET34, TRAINED UNET, TRAINED UNET11, TRAINED UNET16, TRAINED LINKNET34). EVALUATION METRICS: JACCARD-INDEX (IOU), SØRENSEN–DICE COEFFICIENT (DSC), ACCURACY IN PERCENTAGES.

Method	**Results**		
	IoU	***Dice coeff.***	***Acc.***
Pre-trained UNet	13.02	23.13	29.39
Pre-trained TernausNet11	14.78	27.74	46.61
Pre-trained TernausNet16	16.69	28.23	49.35
Pre-trained LinkNet34	37.47	50.41	89.02
Trained UNet	69.63	79.16	96.18
Trained TernausNet11	70.96	79.91	97.38
Trained TernausNet16	59.53	66.64	95.44
Trained LinkNet34	70.73	79.14	96.21

The results of the image processing method were evaluated quantitatively, and the segmentation with a pre-trained and trained models on nine videos, since the ground truth annotations were done on nine videos on the JIGSAWS dataset. The trained neural networks with the JIGSAWS were evaluated with 3-fold-cross-validation since there were three videos from each task [144]. The networks were trained with six videos and evaluated on three videos.

It can be seen in Table 3.5 and in Figure 3.12 that the best results with trained UNet11 architecture were provided: *IoU = 70.96* and *Dice = 79.91*. The results were slightly worse when LinkNet34 architecture was used for training: *IoU = 70.73* and *Dice = 79.14*. Simple UNet architecture could generate quite good results as well with *IoU = 69.63* and *Dice = 79.16*. UNet16 architecture overall was not so good with *IoU = 59.53* and *Dice = 66.64*. With pre-trained networks smaller accuracies were obtained which is because these networks were trained on the MICCAI dataset, which differs a lot from the JIGSAWS dataset. Still, with the pre-trained LinkNet34 *IoU = 37.47* and *Dice = 50.41* can be obtained, which could be increased by some post-processing steps. The pre-trained UNet, Unet11 and UNet16 performed worse.

If each surgical task is considered separately, on Knot-tying videos every method worked better than on Suturing or Needle Passing videos (Table 3.6). It could be because in the Knot-Tying videos the contrast between the background and foreground is higher. On Needle-passing videos objects in the background is really similar to the surgical tools. Trained LinkNet34 architecture could achieve high results on Knot-tying videos with *IoU*

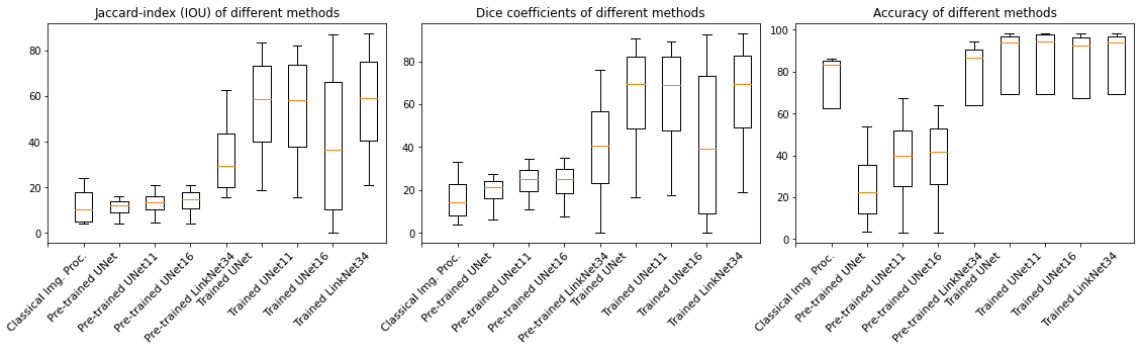

Fig. 3.12. Jaccard-index, Dice-Coefficient and Accuracy of different methods (Classical Image Processing, Pre-trained Unet, Pre-trained UNet11, Pre-trained UNet16, Pre-trained LinkNet34, Trained UNet, Trained UNet11, Trained UNet16, Trained LinkNet34). As the plots illustrate, the best results were achieved with the trained architectures. From the pre-trained architectures, the LinkNet34 gave the best results, the others did not work well on the JIGSAWS. Also, Classical Image Prcessing technique failed on the JIGSAWS dataset.

= *80.01* and *Dice = 88.41*. While on Needle-Passing and Suturing videos, UNet performed better. On Needle-Passing videos Unet achieved *IoU = 66.10* and *Dice = 75.09* and on Suturing videos it performed with *IoU = 69.48* and *Dice = 78.61*. Trained LinkNet34 preformed slightly less on these videos in the case of Needle-Passing *IoU = 62.88* and *Dice = 70.94* and *IoU = 69.31 Dice = 78.07* on Suturing. Also, Unet11 and UNet16 performed on Knot-tying videos better than on Suturing and Needle-Passing videos. This tendency also occured with pre-trained architectures, and pre-trained LinkNet34 gave the best results for all of the tasks. In the case of classical image processing method, also the best results were achieved on the Knot-tying videos. The resulted masks by different methods can be seen in Figure 3.11.

TABLE 3.6

THE RESULTS BY DIFFERENT TRAINING TASKS (KNOT-TYING, NEEDLE-PASSING, SUTURING). EVALUATION METRICS: JACCARD INDEX (IOU), DICE COEFFICIENT, ACCURACY IN PERCENTAGES.

Methods	**Different tasks**								
	Knot-Tying			**Needle-Passing**			**Suturing**		
	IoU	**Dice**	**Acc.**	**IoU**	**Dice**	**Acc.**	**IoU**	**Dice**	**Acc.**
Pre-trained UNet	11.91	22.46	15.80	11.39	19.87	18.85	15.76	27.08	53.52
Pre-trained TernausNet11	15.01	28.36	42.09	13.20	22.60	33.16	16.14	32.27	64.59
Pre-trained TernausNet16	19.06	32.07	51.65	13.36	22.87	34.12	17.65	29.77	62.28
Pre-trained LinkNet34	56.90	71.06	93.25	24.18	35.32	85.37	31.33	44.85	88.45
Trained UNet	73.32	83.89	96.01	66.10	75.09	96.48	69.48	78.61	96.05
Trained TernausNet11	77.58	87.74	96.77	65.91	73.68	96.35	69.39	78.31	96.04
Trained TernausNet16	76.97	88.33	97.01	47.33	52.33	94.13	54.31	59.26	95.19
Trained LinkNet34	80.01	88.41	97.24	62.88	70.94	95.45	69.31	78.07	95.94

3.3.6 Conclusion of surgical tool segmentation on the JIGSAWS dataset

Surgical tool segmentation based on visual data can be an important step toward the automation of image-based skill assessment. However, the JIGSAWS dataset, which is a widely used dataset for skill assessment, did not have a labelling for surgical tools. Therefore, the creation of a method for surgical instrument segmentation on the JIGSAWS dataset was a great challenge. This section analyzed different methods for surgical instrument segmentation on the JIGSAWS dataset. Furthermore, a semi-automatic method for ground truth generation to JIGSAWS was introduced, which was not available before. The examined NNs were the following: UNet, TernausNet-11, TernausNet-16, Linknet-34. First, pre-trained type of these architectures were considered and then training on the JIGSAWS dataset was done. The best overall result was achieved by TernausNet-11 (the one that was trained on the JIGSAWS). Moreover, LinkNet-34 and UNet architectures performed well in some specific surgical tasks. The results suggest that this work can be a good base for creating more efficient surgical tool segmentation methods on the JIGSAWS dataset, which can help the automation of skill assessment by denying critical control parameters in real time. With different NN architectures (e.g., with Recurrent Neural Networks) these results might be improved as well. The segmentation of the surgical tools could be also a good first step for the segmentation of different parts of the surgical instruments.

3.4 Summary of the Thesis group

Surgical technical skill assessment is crucial in MIS for safety and personalized training reasons, yet, it is not part of the clinical routine. Endoscopic image-based skill assessment can provide a solution for the problem, because in the case of MIS and training videos of MIS and RAMIS, only this sensory data is available. Since image-based skill assessment can not achieve the accuracy of kinematic-based solutions, my goal was to propose solutions which can increase the accuracy of image-based surgical skill assessment in RAMIS. In this thesis, a surgical tool pose estimation technique and semantic segmentation was introduced for the da Vinci Surgical System's articulated tools targeting autonomous technical skill assessment. My surgical tool pose estimation method – which was performed on 2D endoscopic images, based on shape features and iterative Perspective n Point transformation method – provides a general solution, since it does not require kinematic data or complete model of the tool. The method's accuracy was proven in surgical skill assessment, where it outperformed the state of the art in knot-tying videos. Ground truth generation for semantic segmentation for the JIGSAWS dataset was also introduced, which was not available before. Using deep learning technologies, I provided an accurate solution for semantic segmentation of the surgical tools for the JIGSAWS dataset, which was also not available while it is essential in image-based skill assessment validation. Based on my results, image-based surgical technical skill assessment can be a good alternative to kinematic data-based skill assessment in the case of RAMIS.
Related publications: [RNE1, RNE2, RNE3, RNE4, RNE5]

Chapter 4

AUTOMATED NON-TECHNICAL SKILL ASSESSMENT AND WORKLOAD EVALUATION IN MIS

4.1 Methods for non-technical skill and mental workload assessment in RAMIS

4.1.1 Non-technical skill an workload assessment – self-rating techniques

Performing a surgical procedure can be very stressful to the whole crew of the OR. Fatigue (mental and physical) can naturally influence the outcome of the surgery, furthermore, time limits can take serious stress and cognitive load on the surgeon, and working in a team can be disturbing in some cases. Workload is a term that represents the psychological cost to perform a task; it is human-specific, however, there are typical situations which can put a serious amount of mental workload on every operator. Workload can be defined with self-rating techniques, where a subject fills a questionnaire about his/her personal experience about the task is burden, yet it is biased and subjective [52]. Furthermore, there are works in the literature which studied both subjective and objective non-technical skill assessment metrics [45, 145], or objective physiological parameters [47, 146, 147, 148, 149, 150, 43, 151]. Workload measurements do not only help to assess the personal workload index, but also to define the main stressors and disturbing factors in surgery in general. Furthermore, it provides personal training assistance for novices [52].

The widely employed NASA Task Load Index (NASA-TLX, created by NASA's Ames Research Center in 1988) is a workload self-rate metric, originally created for assessing pilots' workload in aviation [152, 14]. NASA-TLX measures the workload on a subject with questions related to mental, physical and temporal demand, effort, performance and frustration level. The subject (which can be one person or all team members) has to answer the questions on a 100-point-scale with 5-point steps (Table 4.1). NASA-TLX is a generally used technique for workload measurement in aviation, military and healthcare. NASA-TLX can also be found in traditional MIS mental workload estimation [153, 154, 155, 156, 157], and employed in the case of surgical robotics workload

TABLE 4.1
NASA-TLX MENTAL WORKLOAD SELF-RATING QUESTIONNAIRE [14].

Title	Endpoint	Description
Mental demands	low/high	How much mental activity was required?
Physical demands	low/high	How much physical activity was required?
Temporal demands	low/high	How much time pressure did you feel?
Effort	low/high	How hard did you have to work?
Performance	good/poor	How stressful do you think you were?
Frustration level	low/high	How frustrated did you feel?

assessment as well [28, 45, 47, 145, 146, 148, 150, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181]. There are additional mental workload assessment techniques that were not created originally for surgery, and used in workload assessment for RAMIS. Such examples are:

- Multiple Resources Questionnaire (MRQ) [182, 44, 165, 183, 167];
- Dundee Stress State Questionnaire (DSSQ) [184, 183, 44];
- Rating Scale for Mental Effort (RSME) [185, 147];
- Psychometric Testing of Interpersonal Communication Skills Questionnaire (PTICSQ) [186];
- Safety Attitudes Questionnaire (SAQ) [187, 186];
- Wisconsin Card Sorting Test (WCST) [188, 162];
- Coping Inventory of Task Stress (CITS) [189, 44];
- Subjective Mental Effort Questionnaire (SMEQ) [190];
- Local Experienced Discomfort (LED) [190];
- Short Stress State Questionnaire (SSSQ) [191, 167].

MRQ estimates workload by 17 items, and it is specifically useful for multitasking workload measurements [182]. SSSQ is based on DSSQ, and both target stress measurement [191], such as CITS [189]. RSME and SMEQ estimate mental effort on a 9-point scale from extreme effort to absolutely no effort. RSME is validated in healthcare as well [185]. LED examines physical discomfort during a task [190]. For team communication quality estimation PTICSQ was created [186]. SAQ developed for healthcare, which examines employees' satisfaction with the job, teamwork, management, safety, stress and working conditions [187]. WCST is a neuropsychological tool, which was originally created for cognitive strategy adaptation measurements [188].

Surgery Task Load Index (SURG-TLX) (created by the cooperation of the University of Hong Kong, University of Exeter and the Department of Urology, Royal Devon and Exeter Hospital in 2011) is a modified NASA-TLX metric for surgical workload measurements [15]. SURG-TLX estimates the workload based on mental demands, physical demands, temporal demands, task complexity, situational stress and distractions (Table 4.2,

TABLE 4.2
THE COMPOSITION OF THE SURG-TLX MENTAL WORKLOAD SELF-RATING QUESTIONNAIRE [15].

Title	Endpoint	Description
Mental demands	low/high	How mentally fatiguing was the procedure?
Physical demands	low/high	How physically fatiguing was the procedure?
Temporal demands	low/high	How hurried or rushed was the pace of the procedure?
Task complexity	low/high	How complex was the procedure?
Situational stress	low/high	How anxious did you feel while performing the procedure?
Distractions	low/high	How distracting was the operating environment?

Fig. 1.6). SURG-TLX was tested on the Fundamentals of Laparoscopic Surgery (FLS) peg transfer task under stress, such as fatigue, multitasking, distraction and task novelty. While this metric was validated for surgery, only a few RAMIS publications were found on this topic [159, 147, 149]. Nevertheless, this topic is well-studied in traditional MIS [192, 193, 194, 195, 196], to the best of the authors' knowledge there is no workload self-rating measurement metric specifically created for RAMIS.

Self-rating techniques are easy to implement, and they do not require external human resources, however, they can definitely be biased. It is necessary to consider the usage of self-rating techniques in automated or expert-rating focused NTS and workload assessment studies, because these questionnaires can provide an easy validation tool for correlation examinations. With self-rating tools, the real stressors of the surgery can be observed, and other approaches have to fit to the clinical relevance. Self-rating studies can be found in Table 4.5 with the following references: [43, 44, 45, 47, 145, 146, 147, 148, 149, 155, 161, 162, 164, 165, 166, 167, 168, 169, 171, 173, 174, 175, 176, 177, 180, 181, 186, 190, 197, 198, 199].

4.1.2 Non-technical skill assessment – expert rating

In surgical skill assessment, expert rating techniques are widely used, not just in the case of technical skill assessment, but for non-technical skill assessment as well. Therein, an expert panel (usually 8–10 expert surgeons) assesses the skills of the practicing surgeon, based on a video recording of the procedure/training session, based on a validated set of requirements. Expert rating assessment is relatively easy to complete (compared to automated techniques), more objective than self-assessment, but it definitely requires significant effort and human resources, and it can still be biased for personal reasons. At the moment, expert rating technique is the gold standard for automated skill assessment.

In the case of non-technical skill assessment, there are several different expert-rating metrics for traditional MIS, such as NOTECHS, OTAS and NOTSS (Table 4.3). A few publications were identified which studied NOTSS in the case of RAMIS [45, 46, 200]. For surgical robotics, there is one metric which specifically measures the NTS of robotic surgeons [17]; the Interpersonal and Cognitive Assessment for Robotic Surgery (ICARS), developed by Raison et al. in 2017. It was created by 16 expert surgeons with the Delphi

TABLE 4.3

BEHAVIORAL RATING SYSTEMS IN TRADITIONAL SURGERY COMPARED TO ICARS, THE ONLY EXISTING NON-TECHNICAL SKILL ASSESSMENT METRIC FOR RAMIS [16, 17]. N/A: NOT APPLICABLE

	revised NOTECHS	NOTSS	OTAS	ICARS
Date	2008	2006	2006	2017
Reference	[204]	[205]	[206]	[17]
Non-technical skills	• Communication & interaction • Situational awareness • Team skills • Leadership & management • Decision making	• Situational awareness • Decision making • Task management • Leadership • Communication • Teamwork	• Task checklist • Shared monitoring • Communication • Cooperation • Coordination • Shared leadership	• Communication & teamwork • Leadership • Decision making • Situational awareness • Cope with stress and distractors
Content validity		✓	✓	✓
Construct validity			✓	
Inter-rater reliability	✓	✓	✓	✓
Sensitivity	n/a	not acceptable in some categories	n/a	n/a
Feasibility	✓(especially for self-assessment)	✓	limited to certain procedures	✓

method [201]. In ICARS, there were 28 NTS identified (Fig. 1.6), in 3 main non-technical skill categoriesTable 4.4, namely:

- Interpersonal skills (communication/teamwork and leadership);
- Cognitive skills (decision making and situational awareness);
- Personal resource skills (cope with stress and distractions.

However, only one clinical study was found which used ICARS for non-technical surgical skill assessment [200]. Despite the disadvantages of expert-rating techniques (need for expert surgeon resources, time, bias), they can be an objective tool for automated technique validation, furthermore, they can provide a model for NTS assessment through the critical NTS categories and the given points. Expert-rating studies can be found in Table 4.5 with the following references: [17, 28, 45, 145, 147, 150, 158, 159, 160, 162, 163, 169, 170, 172, 178, 179, 200, 202, 203].

TABLE 4.4
INTERPERSONAL AND COGNITIVE ASSESSMENT FOR ROBOTIC SURGERY (ICARS) EXPERT RATING METRICS [17]

NTS category	NTS group	NTS
Interpersonal skills	Communication and teamwork	Effective verbal communication
		Appropriate interaction with bedside surgeon
		Appropriate interaction with operating room staff
		Engages/initiates in confirmatory feedback with OR staff
	Leadership	Appropriate and polite instructions
		Effective workload management
		Coordination of the team from the console
		Coordination of the team at the bedside
		Delegating tasks to team members
		Maintenance of professional standards
Cognitive skills	Decision making	Appropriate decision making in case of equipment failure
		Appropriate decision making at the bedside
		Quick diagnosis of unexpected patient events
		Quick decision making in case of emergency
		Generation, selection and implementation of solutions
		Outcome review of decision
	Situation awareness	Awareness of patient status
		Ability to deal with patient at the bedside
		Ability of quick adaptation to problems
		Anticipation of potential problems
		Role awareness of surrounding team members at the console
Personal resource skills	Cope with stress and distractors	Understands personal limitations and asks for help (if necessary)
		Identification of stressor
		Maintenance of cognitive skills
		Maintenance of technical skills
		Professional and appropriate choice of resolution

4.1.3 Automated non-technical skill and mental workload assessment in RAMIS

Establishing the correlation between physiological signals, kinematic data or other objectively measurable features and NTS or mental workload can lead to autonomous non-technical skill assessment in RAMIS.

A common approach is to assess the NTS of the surgeon is through the measurement of physiological signals. However, this has limitations: the physiological signals are often linked to a particular non-technical skill, such as stress level, but they do not show other important factors (situational awareness, teamwork, etc.). In the literature, physiological measurements related to the stress level can be found, such as:

- brain activity [207];
- skin temperature [208, 209];
- nose temperature [210];
- heart rate (HR) [211];
- skin conductance [211];
- blood pressure [211];
- respiratory period [211];
- tremor [212];
- eye movement [213].

While these physiological signals are proven to be related to stress, they naturally have limitations in the usage of NTS and cognitive load assessment. Such example is skin conductance, which can be a useful technique to estimate workload [43], but it can be influenced by other physiological factors. Brain activity, HR, and eye movement are the most studied signals in RAMIS, which can refer to more complex underlying behavior, such as technical skills [214], but the correlation between these signals and NTS is harder to be established.

In the literature, there are examples of the usage of an electroencephalogram (EEG) [47, 174, 199, 148, 150, 215, 175, 177], given the fact that EEG measures the electrical activity of the brain [216]. While EEG is the most trivial physiological signal measurement technique for non-technical surgical skill assessment, the proven correlation between the measurable brain activity and NTS is limited. Another approach for physiological signal-based mental workload assessment is the measurement of the HR [27, 147, 149, 190, 177]. However, the accuracy of HR measurements for cognitive load assessment was not enough in some cases, because there is no scale for maximum tolerated workload levels, and their related effects on the surgeon's health [27]. The following forms of HR can be found in the non-technical skill assessment literature, however, the usage of them can also be cumbersome [216, 217]:

- simple HR;
- Heart Rate Variability (HRV);

- mean square of successive differences between consecutive heartbeats (MSSD);
- average heart rate (HRA).

Another objective method for non-technical skill or mental workload assessment is Functional Near-Infrared Spectroscopy (fNIRS) [149, 218, 219]. FNIRS is a functional neuroimaging technique to track the brain activity by monitoring the blood flow in the prefrontal lobe [220]. FNIRS shows a strong correlation with PET and fMRI data, yet it has better temporal resolution than fMRI, but limited compared to EEG; spatial resolution is more limited compared to fMRI, but better compared to EEG [221, 222]. Furthermore, time of isovolumetric contraction (PEP) [223], electromyography and electrodermal [177] can also be used in mental workload assessment [190], however, these signals can be influenced by the surgeon's general health. Pupillary response is also studied in relation to workload [151].

To summarize, the following sensors/imaging techniques were studied in NTS and workload assessment in RAMIS (detailed in Table 4.5):

- magnetic pose trackers;
- EEG;
- ECG;
- fNIRS;
- skin conductance sensor;
- electromyograph;
- eye-gaze tracker;
- heart rate monitor.

Sensors in RAMIS on the technical side are not external only, but there are built-in robot sensors, which can help NTS skill assessment and proven tools for technical skill assessment in RAMIS:

- position sensors (encoder);
- 3D endoscopic camera.

In RAMIS research, there are developed/used sensors which are not strongly related to NTS and workload assessment, but in most of the cases these modalities showed correlations with technical skills [224]. These sensors including but not limited to the following devices [225, 226, 227, 228]:

- force sensors (strain gauges, capacitive sensors, piezoelectric sensors, optical sensors);
- tool position sensing (optical, electromagnetic);
- master/surgeon arm position sensing (external);
- wearable eyeglasses (Oculus Rift, Google Glass);

- tool thermal sensor;
- pressure sensors;
- camera (RGB-D, external);
- communication (RF sensors);
- speech (microspeaker);
- sound (microphones).

Automated NTS and workload assessment can be a key to an objective, reproducible measure regarding the surgeon's skills without bias and the procedure's need of human resources. However, these techniques are hard to implement, and the usage of additional sensors can be a problem in a clinical environment. Nevertheless, NTS and workload might be demonstrable in objective, technical skills, as suggested in [199], which means these sensors can provide an option for NTS assessment as well, however, this research field is not studied widely yet. Automated studies can be found in Table 4.5 based on the following references: [27, 43, 47, 146, 147, 148, 149, 150, 151, 174, 175, 177, 190, 199, 215, 218, 219].

TABLE 4.5: Non-technical skill and mental workload assessment in surgical robotics. Used abbreviations: S. – number of subjects, RAMIS – Robot-Assisted Minimally Invasive Surgery, OR – Operating Room, VR – Virtual Reality, EEG – electroencephalogram, NASA-TLX – NASA Task Load Index, SURG-TLX – Surgery Task Load Index, NOTSS – Non-Technical Skills for Surgeons, MRQ – Multiple Resources Questionnaire, DSSQ – Dundee Stress State Questionnaire, ECG – electrocardiogram, HR – heart rate, HRV – heart rate variability, RSME – Rating Scale for Mental Effort, PTICSQ – Psychometric Testing of Interpersonal Communication Skills Questionnaire, SAQ – Safety Attitudes Questionnaire, fNIRS – Functional Near-Infrared Spectroscopy, PVT – Psychomotor Vigilance Test, WCST – Wisconsin Card Sorting Test, CITS – Coping Inventory of Task Stress, MSSD – mean square of successive differences between consecutive heartbeats, PEP – time of isovolumetric contraction, HRA – average heart rate, SMEQ – Subjective Mental Effort Questionnaire, LED – Local Experienced Discomfort, SSSQ – Short Stress State Questionnaire, p. – procedures (where no subject data was available), QoE – Quality of Evidence, mod.: moderate.

Ref.	S.	Environment	Input	Feature/ NTS	Conclusion	QoE
[43]	10	Dry lab	Skin conductance Self-rating (custom)	Workload Stress	Stress is less in the case of RAMIS compared to traditional MIS.	mod.
[163	5	VR simulator	NASA-TLX	Workload	Workload can be increased in proportion to delay time with the proposed simulators.	low
[44]	15	Dry lab	DSSQ MRQ CITS	Workload Stress	Stress is less, workload and stress coping strategies are the same in the case of RAMIS compared to traditional MIS.	low
[170	20	VR simulator	NASA-TLX	Workload	Mimic dV-Trainer shows reasonable workload results.	low
[165]	15	Dry lab	NASA-TLX MRQ	Workload	The usage of the da Vinci 3D view causes less workload compared to the 2D view in some cases.	low
[172	6	VR simulator	NASA-TLX	Workload	Time delay in teleoperation can significantly increase the workload.	low
[190]	16	Dry lab	MSSD PEP HRA SMEQ LED	Workload Stress	RAMIS causes less cognitive workload compared to traditional MIS.	low
[28]	34	Live porcine	NASA-TLX	Workload	RAMIS poses less mental workload compared to traditional MIS.	mod.
[178]	3	VR simulator	NASA-TLX	Workload	Workload is not improved under delays of 300 ms and 400 ms in the simulated environment.	low
[219	21	VR simulator	fNIRS	Workload	FNIRS can show the cognitive burden during training.	high
[183]	15	Dry lab	MRQ DSSQ	Workload Stress	Novices have less stress when working with the da Vinci compared to traditional MIS.	low
[179	12	Dry lab	NASA-TLX	Workload	After the proposed training, mental workload is similar between novices and experts.	low
[218]	21	VR simulator	fNIRS	Cortical activity	There is a significant difference between expert and non-expert subjects with Gaze-Contingent Motor Channeling.	mod.
[27]	2	OR	HR HRV	Stress	RAMIS poses less mental workload compared to traditional MIS. Workload measurement with HRV is cumbersome.	mod.
[160]	28	Dry lab	NASA-TLX	Workload	RAMIS poses significantly better workload perception compared to traditional MIS.	low
[158	52	VR simulator	NASA-TLX	Workload	Urethrovesical anastomosis VR training improves technical skill acquisition with cognitive demand.	mod.

Ref.	S.	Environment	Input	Feature/ NTS	Conclusion	QoE
[168]	28	VR simulator	NASA-TLX	Workload	Xperience Team Trainer emphasizes the importance of teamwork.	mod.
[199	10	Dry lab	EEG	Cognitive engagement Mental workload Mental state	Cognitive assessment can define the expertise levels.	high
[147]	32	Dry lab	SURG-TLX RSME Heart rate monitor	Workload HRV	RAMIS poses less mental workload compared to traditional MIS.	mod.
[203	6	Simulated OR	Expert rating (custom)	Communication Leadership	Repeated simulations and increased leadership mean faster and less flawed conversions in the OR.	mod.
[169]	24	Image display	NASA-TLX	Workload	Increasing the level of cognitive load is significantly increasing the inattention blindness.	mod.
[150	1	OR	EEG NASA-TLX	Workload Distractions Mental state	Expert surgeons use different mental resources based on their needs.	mod.
[202]	89	OR	Expert rating (custom)	Communication Decision making	RAMIS increases communication requirements for the team of the OR.	mod.
[186]	32	OR	PTICSQ SAQ	Communication	There is a significant correlation between team communication and surgical outcome.	mod.
[148	1	OR	EEG NASA-TLX	Workload	A surgical expert during mentoring concerned while he was observed the surgery.	low
[145]	89	OR	Expert rating (custom) NASA-TLX	Communication Workload	The proposed method is capable of capturing team activities during RAMIS.	mod.
[161	21	Live porcine VR simulator	NASA-TLX	Workload	Live animal and VR simulator training provide a comparable workload.	low
[175]	8	VR simulator	EEG NASA-TLX	Procedural memory Attention level Workload	EEG can show the learning progress in the case of RAMIS.	high
[164	55	OR	NASA-TLX	Workload	The study proposes a workload variety analysis with different members of the OR.	mod.
[197]	25 p.	OR	NASA-TLX	Workload	NASA-TLX is a useful tool for determining the appropriate staff member mix for RAMIS procedures.	mod.
[198	10	OR	SURG-TLX	Workload	Mental demands are higher for surgeons at the console than are assisting.	mod.
[171]	24	Live porcine	NASA-TLX	Workload	Single-site access surgery can significantly reduce the workload.	mod.
[47]	27	VR simulator	EEG NASA-TLX	Cognitive features Mental workload Engagement Asymmetry index Brain functional features Communication Integration Recruitment Workload	EEG features can be used for objective non-technical skill assessment.	high
[166]	27	OR	OR efficiency (custom) NASA-TLX	Communication Workload	Anticipation causes shorter operating time. Team familiarity causes less inconveniences. Less anticipation causes less cognitive load.	mod.
[167	32	VR simulator	NASA-TLX SSSQ MRQ	Workload Stress	Training with a VR simulator can decrease the workload and stress.	mod.
[146]	8	VR simulator	NASA-TLX Eye movements	Workload	Eye movements correlate with the workload.	high
[180	5	OR	NASA-TLX	Workload	Workload is less in the case of robot-assisted submucosal dissection compared to the traditional case.	low

Ref.	S.	Environment	Input	Feature/ NTS	Conclusion	QoE
[45]	20	OR	NOTSS NASA-TLX	Situational awareness Decision making Leadership Communication Teamwork Workload	Non-technical skills are associated with team efficiency, surgical flow disruptions and self-perceived performance.	high
[151	26	Dry lab	Task-evoked pupillary response	Workload	Under high cognitive workload, there can be a divergence in robotic movement profiles between expertise levels.	high
[46]	62	Dry lab Simulated OR	NOTSS	Situational awareness Decision making Leadership Communication Teamwork	Motor imaginary training technique is not effective in non-technical skill training.	mod.
[181	7	OR	NASA-TLX	Workload	RAMIS requires less mental demand and effort compared to open access surgery and traditional MIS.	mod.
[174]	4	OR	EEG NASA-TLX	Cognitive features Functional features Mental workload Mental load Engagement Situation awareness Blink rate Asymmetry index Completion time Communication	During a simple surgical task, functional brain features are sufficient to classify mentor-trainee trust.	high
[149	8	Dry lab	fNIRS SURG-TLX HRV	Prefrontal activation Workload Stress response	RAMIS improves performance during high workload conditions.	high
[215]	32	VR simulator	EEG	Electrocortical activity in temporoparietal and left frontal regions	There are significant differences in electrocortical activity between novices and experts.	high
[177	12	VR simulator	HRV NASA-TLX Wrist motion Electromyography Electrodermal EEG	Workload Expertise	The proposed skill and workload evaluation framework is accurate.	high
[173]	31	VR simulator	NASA-TLX	Workload	Specific self-directed robotic simulation curriculum was introduced, which can significantly decrease the workload.	mod.
[176	264 p.	OR	NASA-TLX	Workload	Mental workload is similar in the case of RAMIS, traditional MIS, hand-assisted MIS and open surgery.	mod.
[162]	30	Wet lab	NASA-TLX PVT WCST	Workload Concentration Cognitive function	Robotic assistance does not provide less mental workload with novices. Robotic assistance may be mentally taxing for robotic novices.	mod.
[200	n/a	OR	OTAS NOTSS ICARS NOTECHS II	Situation awareness Decision making Communication Teamwork Leadership Stress	The study proposed a structured approach to the analysis of non-technical skill using extracorporeal videos of both open radical cystectomy and RAMIS radical cystectomy	mod.
[159]	13	Dry lab	NASA-TLX	Workload	Physiological and cognitive ergonomics with robotic surgery are significantly less challenging compared to traditional MIS.	low

4.2 Autonomous non-technical surgical skill assessment and workload analysis in laparoscopic cholecystectomy training

RAMIS skill assessment is a relatively young research field, and the strong need for NTS and workload assessment has not found solid bases in the literature yet. A few publications suggested objective, sensor-based non-technical skill and mental load evaluation in RAMIS, while this approach can provide bias-free, reproducible solution in the clinical environment. Furthermore, during surgical education, personalized skill training would provide a more effective learning procedure, which can be achieved more easily with objective metrics. Nevertheless, objective metrics are hard to implement, additional sensor usage can be problematic in a clinical environment, and at the moment, there are not any validated objective and automated metrics in NTS assessment. On the other hand, there are close relations in manual MIS and RAMIS, and in manual MIS, it is already suggested to approach NTS assessment with technical skill assessment metrics [51, 229], which is a much more studied area in RAMIS. It suggests that technical and NTS are not different in RAMIS, thus the connections of these two seemingly diverse research approach should be studied more thoroughly . With validated manual techniques, it can be done by relatively simple statistical analysis, but in the case of automated techniques, appropriate test environment, amount of data, sensor usage, feature extraction and classification techniques should be examined carefully. For technical skill assessment, there are accurate results with kinematic [11] and video data [230] as well. However, these studies only focus on the surgeon and not the whole crew of the operating room, with external sensors (such as cameras) workflow and NTS (such as communication and teamwork) the correlation can be studied. However, the first step of these studies is to examine the different sensor outputs, which can correlate with technical and non-technical metrics. RAMIS built-in sensors (3D endoscopic camera and kinematic sensors) can significantly ease NTS skill assessment and workload examination, and with them correlations with other sensor outputs and/or self/expert-rating results can be studied.

In this thesis group, an automated, image and force sensor-based non-technical surgical skill classification solution is presented, aiming to understand the correlations behind tool motions, used forces and situation awareness, dealing with stress and distractions. To achieve this, a MIS setup was used because, at that time, there were no experienced console surgeons in Hungary. For this, a laparoscopic training platform was created, simulating certain parts of laparoscopic cholecystectomy: clipping the cystic artery and the dissection of the parietal peritoneal layer. The training task requires situation awareness, dealing with stress and distractions, and decision making. The analysis of the workload was done with a self-rating questionnaire (Surgery Task Load Index, SURG-TLX) [15]. Tool motion was processed with an object tracking computer vision algorithm (CSRT [231]). The time series data (force and image) were classified with a Fully Convolutional Neural Network-based classifier [13].

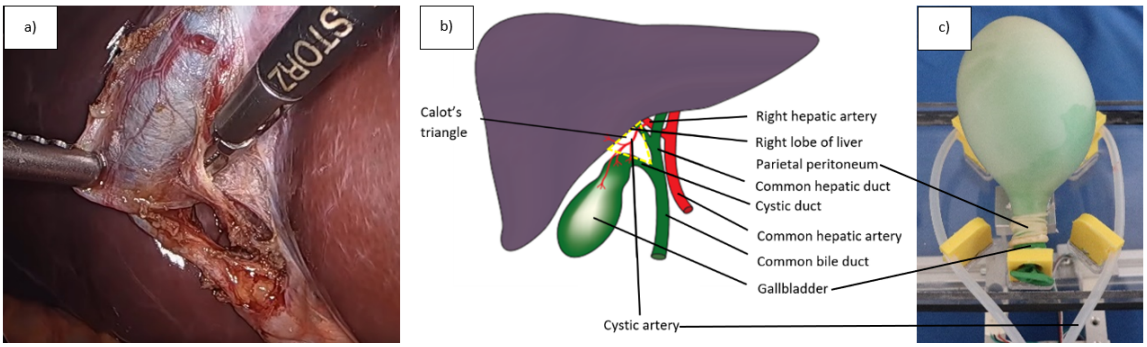

Fig. 4.1. Laparoscopic cholecystectomy surgical scene and anatomy. A) Laparoscopic cholecystectomy (LC) surgical scene, after exposing the Calot's triangle; b) Anatomy of the cyst and its environment; c) Surgical phantom created for LC with the peritoneum, cyst and the cystic artery.

4.2.1 Medical background—Laparoscopic Cholecystectomy

Cholecystectomy is one of the most common MIS interventions in the 21st century of medicine, which requires advanced non-technical skills from the surgeon. Main indications of the operation are gallstones (cholelithiasis), inflammation of gallbladder (cholecystitis), or less frequently, polyp or neoplasm of the cyst itself. Stones could lead to acute inflammation, decades of chronic ulceration of the cyst wall and tissue inflammation could lead to metaplasia, dysplasia and malignancy. According to Kenneth et al., approximately 300,000 cholecystectomies are performed in the US annually [232]. The gold standard approach of the procedure is MIS laparoscopic way, especially in elective cases. While it is a really frequent type of procedure, performing it requires a well-trained and skilled surgeon in the domain of hepato-biliary tract, to maintain patient safety, as any accidental damage could lead to serious complications. Manipulation with laparoscopic tools has a slow learning curve, but has enormous benefit to the patient, including less post-operative pain, faster recovery and less hospital stay, effect on overall health economy is unquestionable. The procedure itself starting with CO_2 inflation of the generally anesthetized patient's abdominal cavity with secure Veress-needle to gather room to manoeuvre the laparoscopic tools. Safety trocars are inserted through the abdominal wall, and the working instruments are applied in them. Clear vision of the operating area and tools are required for step-by-step appropriate decision making. Exploration of Calot's Triangle's structures at the porta hepatis is the main aspect of the whole surgery (Fig. 4.1), including the following steps done on the cystic artery and cystic bile duct (maintaining attention on avoidance of choledochal duct and duodenum itself):

- Localization;
- Dissection;
- Clipping;
- Transecting.

Failure in any of these steps can occur in severe intra-operative or post-operative bleeding, injury or damage of nearby organs or bile leakage with necessity of re-operation and High Dependency Intensive Care, caused by general inflammation of peritoneum and sys-

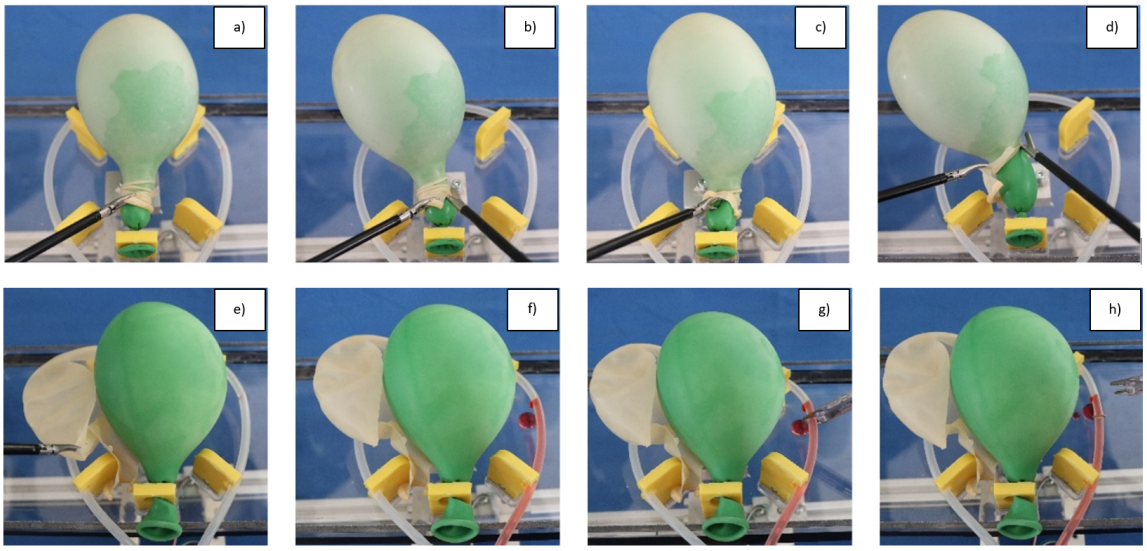

Fig. 4.2. Experiment's workflow. A) Grasping the outer layer (peritoneum) with a dissector; b) Cutting the outer layer with a pair of scissors; c) Blunt dissection; d) Cutting; e) Removing the covering layer; f) Abrupt bleeding; g) Localize the bleeding source, change the tools to a clipper; h) Clipping the blood vessel.

tematic septic condition. The finishing steps of LC involve the dissection and cauterisation of the parietal peritoneal layer to remove the cyst from its "nest". After removal of the gallbladder through the epigastric port, revision of the operating site is required for re-assuring the deficit of any complication, in certain cases drain/s is/are applied, finishing with desufflation of abdominal cavity and skin closure.

4.2.2 Surgical phantom, experimental environment and workflow

Surgical anatomical phantoms and training platforms are created to simulate real surgical environments and situations, and to train the proper surgemes to surgeons. The usage of phantoms and validated training tasks are extremely important for training in MIS, since the motions can be difficult to perform precisely with laparoscopic tools. In this part of my thesis work, a simple phantom for laparoscopic cholecystectomy was created based on a cyst dissection phantom (Laparoscopyboxx, Nijmegen, NL) [233]. In this phantom, a balloon was placed inside another balloon, which was then blown. It resulted that the outer balloon could be dissected without causing damage to the inner balloon, which represented the dissection of the peritoneum to remove the cyst (Fig. 4.1). In this work, a damaged tube was added to the model, which simulated the cystic artery clipping failure aiming to create a stressful situation which required quick problem solving. For the experiments, a laparoscopic dissector and a pair of scissors were used, and for handling the bleeding, 10 mm single-use clipper was utilized.

The workflow of the experiment started with a general introduction of the task to perform (with a video illustration), and for the novices, the proper usage of the laparoscopic tools was shown. The experiments were done in a laparoscopic trainer box, which represented the MIS environment (the tools were inserted to small incisions). For the experiments, only two ports were used for the two tasks: dissecting the outer balloon with the dissector and the scissors and for clipping the blood vessel. The operating area was visu-

TABLE 4.6
SURGICAL PHANTOM AND TRAINING ENVIRONMENT VALIDATION QUESTIONNAIRE.

Title	Endpoint	Description
Experiment's applicability	not appropriate/excellent (0–5)	How appropriate was the experiment to teach laparoscopy during the modeled surgery?
Movement similarity	not similar/same (0–5)	How similar are the movements to those required during surgery?
Anatomical similarity	not similar/same (0–5)	How realistic is the anatomical phantom designed to model the surgical area?
Simulation of stress	not appropriate/excellent (0–5)	How suitable is the anatomical phantom to simulate stressful surgical situations?

alized by an Intel RealSense D435i depth camera. On a screen, the two-dimensional RGB images were shown to the operator. The screen was placed beside the subjects, where they could only see the camera image in an uncomfortable posture, simulating the limited ergonomy in the OR. The technique of the dissection was not pre-defined. The operators could perform blunt dissection, cutting and grasping with the tools during the dissection task (Fig. 4.2). The first task ended with the proper dissection of the outer balloon (without damaging the inner balloon) or with a damaged inner balloon (which determines if the procedure was successful or not). After that, the second task was to localize the bleeding, and to use a clipper to handle it. For this, the subject had to remove at least one laparoscopic tool, and change to the clipper. The experiment ended with clipping the blood vessel. The subjects had to perform these tasks three times, and following each task he/she had to fill a questionnaire about the workload. For medical professionals, after the tests a questionnaire about the phantom, it had to be filled out. Surgical and simulator/video game usage was also surveyed.

4.2.3 Questionnaires

After every experiment, the subjects had to fill a questionnaire about the experienced workload. SURG-TLX is a modified NASA-TLX metric for surgical workload assessment [15]. SURG-TLX estimates the workload based on mental demands, physical demands, temporal demands, task complexity, situational stress and distractions (Table 4.2). SURG-TLX was tested on the FLS peg transfer task under stress, such as fatigue, multitasking, distraction and task novelty. It is necessary to consider the usage of self-rating techniques in automated or expert-rating focused NTS and workload assessment studies, because these questionnaires can provide an easy validation tool for correlation examinations. With self-rating tools, the real stressors of the surgery can be observed based on the subjects' opinion, and other approaches have to fit to the clinical relevance.

After the three trials, medical professionals had to fill a questionnaire about the surgical phantom and the experiment (Table 4.6). This included a question about the experiment's applicability for MIS training, the surgical tool movements similarity to real surgery, the anatomical similarity, and the experiment's capability to simulate stressful surgical simulations.

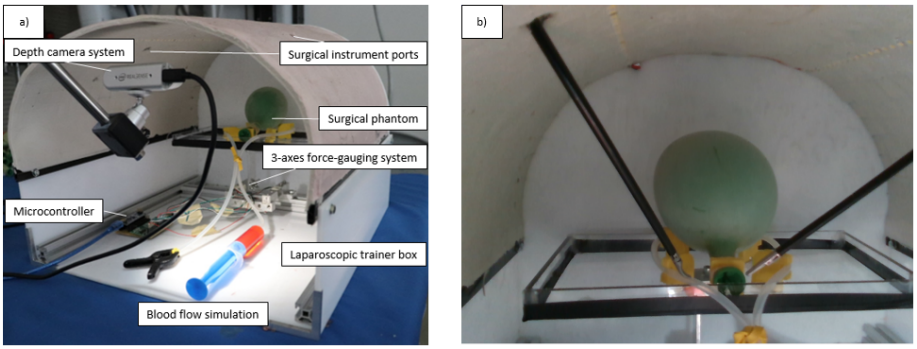

Fig. 4.3. Experimental environment. A) Surgical trainer box with the components of the experiment (phantom and sensors); b) Camera image streamed for the operators during the training and recorded for data processing.

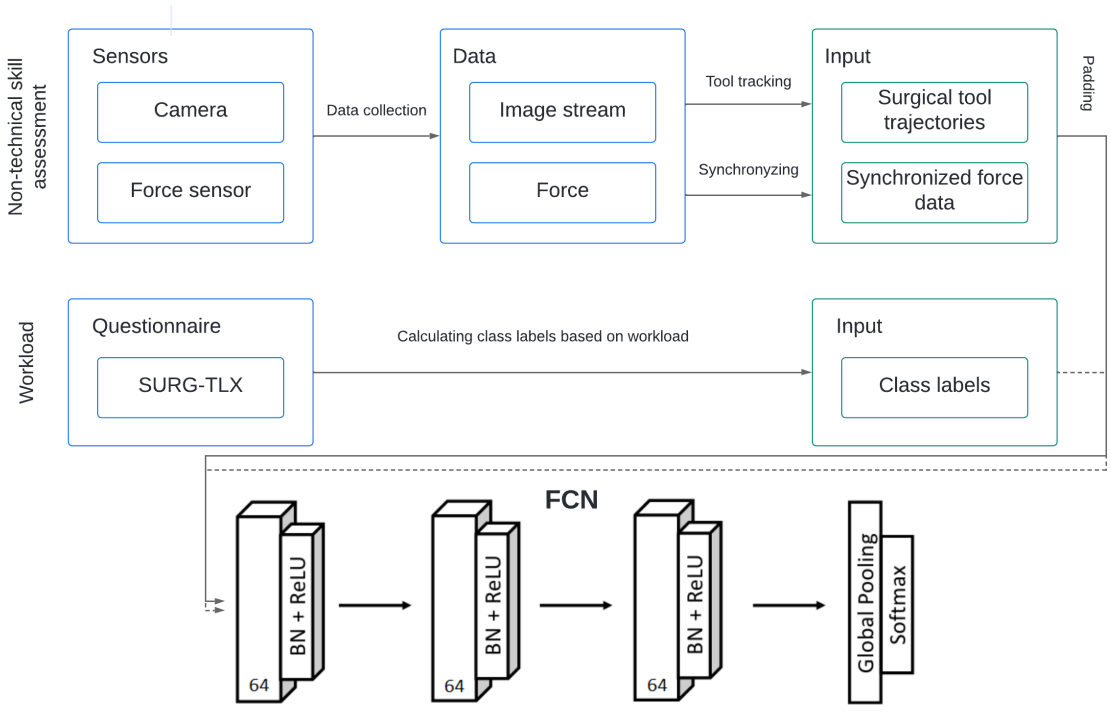

Fig. 4.4. Autonomous non-technical surgical skill assessment method workflow based on sensory data with FCN multivariate time series classification. FCN architecture for non-technical surgical skill classification originally proposed by Wang et al. [13]. The network is built of a convolutional layer followed by a batch normalization layer and a ReLU activation layer. The features are fed into a global average pooling layer, and the final label is produced by a softmax layer.

4.2.4 Statistical analysis

Statistical differences were tested between the medical professionals (MP) and the control group (CG) with Mann–Whitney U-tests. The Mann-Whitney U-test, also known as the Wilcoxon rank-sum test, is a non-parametric statistical test used to compare the distributions of two independent groups to determine if they come from populations with different medians. It assesses whether one group's values tend to be consistently higher or lower than those of the other group. It is suitable for comparing groups when the assumptions of parametric tests (such as the t-test) are not met or when dealing with ordinal or non-normally distributed data. Learning curve was tested along SURG-TLX criteria as well based on Wilcoxon signed-rank tests. The Wilcoxon signed-rank test is a non-parametric statistical test used to compare the medians of paired or matched data sets. It assesses whether there is a significant difference between two related groups when the data does not meet the assumptions for a parametric test like the paired t-test. The test evaluates whether the ranks of the differences between paired observations are significantly different from zero, indicating a shift in the medians of the two groups. It is often used when dealing with ordinal or non-normally distributed data or when the paired nature of the data is important [234].

4.2.5 Hardware and software environment

An Intel Realsense D435i camera's RGB data was streamed and saved at 30 FPS with 640x480 pixels resolution (Fig. 4.3). The used forces were measured with a force-gauging system, that measured the resultant external forces acting on the surgical phantom along three perpendicular axes, which connected to and Arduino Nano (Arduino Srl) microcontroller for low-level signal processing. The microcontroller and the camera were connected to a PC, where the data were saved. The main program (which handled the proper streaming of RGB images, synchronized the force data, calculated the two-dimensional position of the surgical tools, saved the data into a DNN-readable format and performed the classification) was written in Python 3 programming language, extended with OpenCV 4, Keras and TensorFlow 2 libraries.

4.2.6 Surgical tool tracking

The surgical tools (scissors, dissector and clipper) were manually selected on the first frame of the RGB stream with bounding boxes. For tracking the instruments, Channel and Spatial Reliability Tracker (CSRT) was used proposed by Lukezic et al. [231]. In the case of CSRT, an automatically estimated spatial reliability map restricts the discriminative correlation filter to the parts of an object suitable for tracking with improving the search range and performance for irregularly shaped objects. Channel reliability weights calculated in the constrained optimization step of the correlation filter learning reduce the noise of the weight-averaged filter response [231]. CSRT worked reliable on the surgical training videos, only in case of abrupt motions of the tools required the re-selection of the region of interest. All surgical tools' x and y camera coordinates were saved.

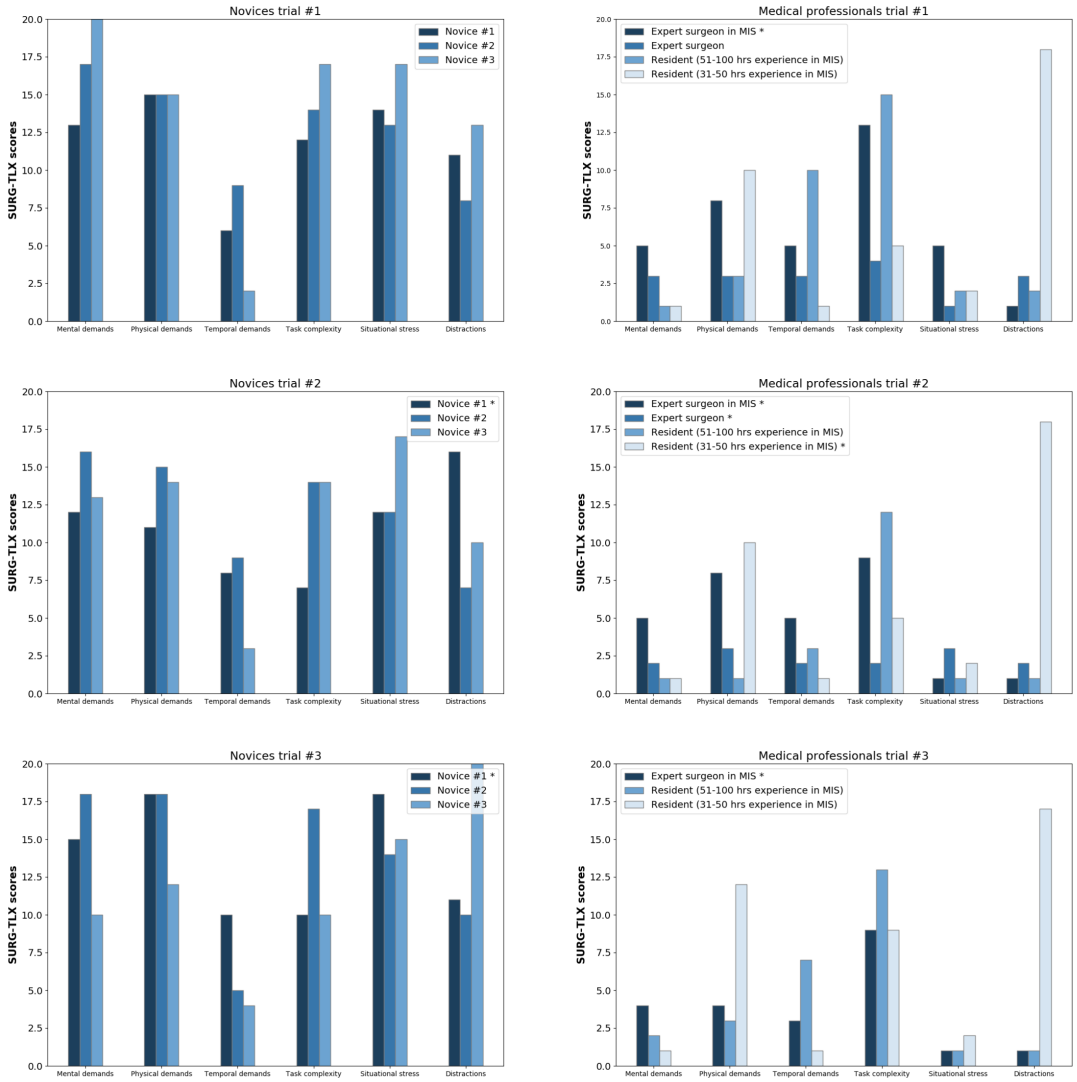

Fig. 4.5. SURG-TLX values (0–20) of medical professionals first to third trials (first row) and novices (second row). Asterisk (*) notes if the cyst dissection was successful.

TABLE 4.7

STATISTICAL COMPARISONS BETWEEN MEDICAL PROFESSIONALS (MP) AND THE CONTROL GROUP (CG). MEAN AND STANDARD DEVIATION (SD) VALUES ARE REPORTED FOR BOTH GROUPS BASED ON SURG-TLX CRITERIA. P VALUES ARE SIGNED AS MP/CG WHERE MP AND CG GROUPS WERE TESTED. *$0.001<p<0.05$ (STATISTICALLY SIGNIFICANT); **$p<0.001$ (STATISTICALLY EXTREMELY SIGNIFICANT); NOT SIGNED: STATISTICALLY NOT SIGNIFICANT.

SURG-TLX	MP mean	MP SD	CG mean	CG SD	p value (MP/CG)
Mental demands	2.36	1.63	14.89	3.18	0.0001**
Physical demands	7.41	3.89	14.78	2.33	0.0002**
Temporal demands	3.73	2.83	6.22	2.91	0.0072
Task complexity	8.73	4.27	12.78	3.35	0.0359*
Situational stress	1.91	1.22	14.67	2.24	0.0002**
Distractions	5.91	7.58	11.78	4.06	0.0552

4.2.7 Skill classification

The final dataset contained 9 parameters (dissector x and y coordinates, scissors x and y coordinates, clipper x and y coordinates, force data x, y and z directions) for each (20) human cases. Force data was synchronized with the image data through the recorded timestamps. Each measurement was built of 94068 frames recorded with 30 FPS, because the longest measurement was such long, and due to the classifier only accepts input data with the same length, other measurements were padded with zeros. In this work, multivariate time series classification was done with a Fully Convolutional Neural Network (FCN). FCN is known of its efficiency for semantic segmentation on images, and it shows very good performance for time series classification as well. The used FCN architecture is built of a convolutional layer followed by a batch normalization layer (for speed and generalization reasons) and a Rectified Linear Unit (ReLU) activation layer based on the work by Wang et al. (Fig. 4.4) [13]. The convolution operation is done by three 1D kernels. The convolution block is built of a convolutional layer:

$$y = W \otimes x + b, \tag{4.1}$$

where y is the output of the convolutional layer, calculated from the weights (W) and bias (b) with a convolution operation ($\otimes$). Y will be the input of the batch normalization layer (BN):

$$s = \mathrm{BN}(y), \tag{4.2}$$

and based on the output of it (s), the output of the ReLU activation layer (h) can be calculated:

$$h = \mathrm{ReLU}(s). \tag{4.3}$$

The NN was built by stacking three convolution blocks with the filter sizes (64, 64, 64) in each block. The features are then fed into a global average pooling layer, and the final label is produced by a softmax layer.

TABLE 4.8
LEARNING CURVE WERE ASSESSMENT BETWEEN 1ST TO 2ND, 2ND TO 3RD AND 1ST TO 3RD TRIALS. *0.001<P<0.05 (STATISTICALLY SIGNIFICANT); **P<0.001 (STATISTICALLY EXTREMELY SIGNIFICANT); NOT SIGNED: STATISTICALLY NOT SIGNIFICANT.

SURG-TLX	p value (1st to 2nd)	p value (2nd to 3rd)	p value (1st to 3rd)
Mental demands	>0.05	>0.05	>0.05
Physical demands	>0.05	>0.05	>0.05
Temporal demands	>0.05	>0.05	>0.05
Task complexity	0.0426*	>0.05	>0.05
Situational stress	>0.05	>0.05	>0.05
Distractions	>0.05	>0.05	>0.05

4.2.8 Subjects

In my research, 20 set of experiments were conducted by 7 subjects: 3 non-medical novices as a control group, 1 very experienced resident with 51–100 hours in MIS, 1 experienced resident with 31–50 hours in MIS, an expert surgeon and an expert surgeon in MIS. All subjects repeated the experiments several times (3 for each except 2 for the expert surgeon). The groups contained both male and female subjects, every subject was right-handed. The ages of the subjects were 25–61 years. None of the participants had experience in robotic surgery. Only the residents had experience in laparoscopic training in a trainer box (between 2 and 50 hours). Participants had very few or no experience in video games.

4.2.9 Results

Outcome

From 20 tests, 7 were successful (dissecting the outer layer without damaging the cyst); the expert MIS surgeon was successful for all 3 times, the expert surgeon was successful at the second trial, one resident could do the task for the second time, and one novice was successful for both the second and the third times. The clipping task was performed successfully by every subjects.

Workload analysis

Statistical differences were tested between the medical professionals (MP) and the control group (CG) with Mann-Whitney U-tests (4.7,4.8, 4.5). It was very high significance between MP and CG in the case of mental demands, physical demands, and situational stress (p value was less than 0.0001, 95 % confidence interval), and a significant difference in the case of task complexity ($p<0.05$). There were no significant differences in temporal demands and distractions. Learning curve was tested along SURG-TLX criteria as well based on Wilcoxon signed-rank tests. Statistical differences were studied for all cases (and 1st to 2nd, 2nd to 3rd and 1st to 3rd). In this case, only task complexity showed significant differences for the subjects between the first and the second trial ($p<0.05$).

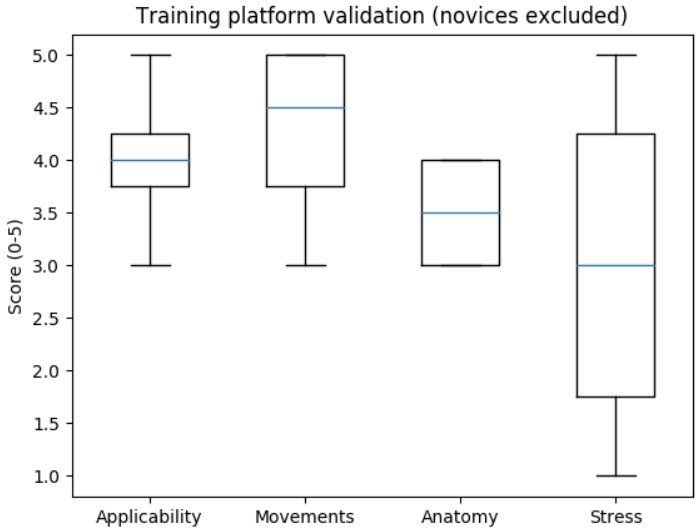

Fig. 4.6. Surgical training environment validation (applicability for laparoscopy training, similarity of motions, anatomical similarity, simulation of stress) based on only medical professionals. Results showed good assessment for applicability, very good for movement similarity, acceptable for anatomical similarity and moderate for stress simulation.

Phantom validation

For phantom validation reasons, all MP participants had to fill a questionnaire about the phantom's applicability, the movements' similarity, the anatomical similarity and stress simulation. Results showed good assessment for applicability (mean value: 4.0/5), very good for movement similarity (mean value: 4.5/5), acceptable for anatomical similarity (mean value: 3.5/5) and moderate for stress simulation (mean value: 3.0/5) (Fig. 4.6).

Autonomous non-technical skill assessment based on sensory data

FCN classified the recorded sensory data based on the class labels defined by SURG-TLX. All data (force_x, force_y, force_z, dissector_{xy}, scissors_{xy}, clipper_{xy}) were tested and the combinations of them (force data and all instruments' path, dissector with scissors path and force data, dissector with scissors path with force_z). Class labels were calculated from SURG-TLX results: every answer below the average received 0, and answers which were higher than the average got 1 as a class label. Based on this method, in the case of experience, mental demands and situational stress class labels were the same. Physical demands, temporal demands, task complexity and outcome determined different class label sets. Skill classification with FCN was validated with Leave One Out Cross-Validation (LOOCV), where each observation is considered as the validation set and the rest of the observations are considered as the training set. The accuracy averages of the 20 iterations can be found in Table 4.9 and Table 4.10. The best classification was done on temporal demands based on the z component of the used forces (85 % accuracy). The classification of temporal demands showed higher overall accuracy compared to other metrics based on the path of the dissector and the scissors with the z component of the used forces (70 %

TABLE 4.9

FCN-BASED AUTONOMOUS SKILL CLASSIFICATION ACCURACY RESULTS ALONG THE CLASSES DEFINED BASED BY SURG-TLX, EXPERIENCE AND EXPERIMENT OUTCOME (D_{xy}: DISSECTOR PATH; S_{xy}: SCISSORS PATH; C_{xy}: CLIPPER PATH) WITH LOOCV VALIDATION. BLUE DENOTES THE BEST RESULTS.

Class labels	**force**$_x$	**force**$_y$	**force**$_z$	**force**$_{xyz}$	**D**$_{xy}$	**S**$_{xy}$	**C**$_{xy}$
Experience/Mental demands/Situational stress	0.75	0.5	0.45	0.45	0.5	0.4	0.25
Physical demands	0.35	0.5	0.55	0.4	0.6	0.6	0.6
Temporal demands	0.55	0.65	0.85	0.6	0.55	0.45	0.55
Task complexity	0.5	0.5	0.55	0.55	0.65	0.4	0.45
Distractions	0.3	0.5	0.35	0.4	0.5	0.5	0.45
Outcome	0.45	0.55	0.55	0.35	0.55	0.55	0.5

TABLE 4.10

FCN-BASED AUTONOMOUS SKILL CLASSIFICATION ACCURACY RESULTS ALONG THE CLASSES DEFINED BASED BY SURG-TLX, EXPERIENCE AND EXPERIMENT OUTCOME WITH COMBINED INPUTS (D_{xy}: DISSECTOR PATH; S_{xy}: SCISSORS PATH; C_{xy}: CLIPPER PATH) WITH LOOCV VALIDATION. BLUE DENOTES THE BEST RESULTS.

Class labels	**D**$_{xy}$**, S**$_{xy}$**, C**$_{xy}$**, force**$_{xyz}$	**D**$_{xy}$**, S**$_{xy}$**, force**$_{xyz}$	**D**$_{xy}$**, S**$_{xy}$**, force**$_z$
Experience/Mental demands/Situational stress	0.5	0.4	0.65
Physical demands	0.25	0.35	0.35
Temporal demands	0.65	0.6	0.7
Task complexity	0.4	0.55	0.5
Distractions	0.45	0.6	0.5
Outcome	0.5	0.4	0.35

accuracy) or other combinations. This correlates with the subjective opinion of subjects as well: according to an intermediate skill level resident the experiment was very rushed, and while based on the experience/mental demands/situational stress the resident got a class label 1, but in the case of temporal demands, the class label was 0 based on SURG-TLX. It suggests that experience can help to cope with stress, but it is not necessarily related to temporal demands, which can be autonomously measured with high accuracy based on the tool paths and the z component of the used forces. Another higher accuracy (75 %) was resulted in the case of classifying mental demands/situational stress with the x component of the used forces. For the other cases, moderate or low accuracy were seen in the classification.

Limitations

While this research has highlighted the promise of using tool motion and force data to assess non-technical skills within surgical training, it is essential to recognize inherent limitations in the approach. One noteworthy constraint is the possibility that students, driven by the goal of achieving high scores within the training system, might prioritize minimizing applied forces over the thorough mastery of proper surgical techniques. This raises the concern that learners could develop a 'score-driven' mindset, potentially compromising the educational objectives. To address this limitation, future research and practical implementations must carefully navigate the balance between performance metrics and the acquisition of comprehensive surgical expertise. It is crucial to consider how the training system can offer constructive feedback and guidance that encourages both efficient execution and the cultivation of genuine surgical skills.

The statistical analysis and machine learning method relied on a relatively small dataset, which may result in potentially erroneous conclusions, despite the appropriate selection of methods. Additionally, the confidence interval in the statistical analysis was set at 95%, indicating a 5% probability of reaching an inaccurate conclusion. Future research projects should prioritize conducting similar studies with larger datasets to enhance the robustness of findings.

4.2.10 Conclusion of automated non-technical skill assessment in laparoscopic training

In this section, autonomous non-technical surgical skill assessment was presented in laparoscopic cholecystectomy training. The training was studied with workload assessment (SURG-TLX) and an autonomous non-technical skill assessment approach, based on an FCN with image-based and force input data. Laparoscopic phantom training and workflow were introduced to simulate stressful situations during surgery (bleeding, time-critical reaction, distractions, physical demands). 20 trials were recorded, performed by 7 subjects with different surgical experience. Statistical tests showed significant differences between the two groups (medical professionals and control group) in the case of mental demands, physical demands and situational stress ($p<0.0001$, 95 % CI), and also in the case of task complexity ($p<0.05$). There were no significant differences at temporal demands and distraction levels. Learning curve in workload was also studied with Wilcoxon signed-rank tests; only task complexity resulted significant difference between the first and the second

trials. Autonomous non-technical skill assessment was done based on image data with tracked instruments based on CSRT tracker and force data. FCN classification showed high accuracy on temporal demands classification based on the z component of the used forces (85 %) and 75 % accuracy for classifying mental demands/situational stress with the x component of the used forces with LOOCV validation. In this thesis group, the connection between sensory data and specific NTS was proven. Furthermore, since the control theory of RAMIS is the same as that of MIS, the methods can be transferred to RAMIS as well.

4.3 Summary of the Thesis group

Non-technical surgical skill assessment is as important as technical skill assessment in MIS, while it is not widely studied in the clinical practice, and it is not involved during training. In this thesis group, I proposed a methodology for autonomous non-technical skill assessment and workload evaluation in laparoscopic cholecystectomy training. It has been shown that there are NTS of the surgeon which can be classified based on sensory data (image and force) during the training with high accuracy. This finding can lead to objective and autonomous surgical non-technical skill assessment, and the proposed training environment can be suitable for personalized training as well.
Related publications: [RNE6, RNE7, RNE8, RNE9]

Chapter 5

A FRAMEWORK FOR SKILL ASSESSMENT IN THE CASE OF AUTONOMOUS CAMERA MOTION

5.1 Skill assessment and automation

Since RAMIS operates on soft tissues, automation in this domain can be extremely difficult due to the constantly changing environment. At the moment, on the market it is rare to see autonomy over Level of Autonomy (LoA) 1 in RAMIS [20]. RAMIS automation research is mainly focusing on surgical subtask automation, which belongs to partial/conditional automation, where the automation of surgemes and motion primitives are necessary as well. However, the workflow of RAMIS contains subtask elements, where choosing the proper subtask can be extremely hard, since it can be critical with respect to the patient outcome. These subtasks can be monotonous and time-consuming, thus automation of them could decrease the workload on the surgeon. Skill assessment in the case of automation can be crucial for safety reasons. In this thesis group, a complete framework for Optical Flow-based surgical skill assessment in the case of autonomous endoscope motion is introduced to assess surgical skills while automation is employed.

5.2 Visual servoing for the da Vinci Surgical System

Surgeons are facing numerous challenges in the operating room. The endoscopic camera mounted on the Endoscopic Camera Manipulator (ECM) is controlled by the surgeon through foot pedals by default, which presents a cognitive load [224]. It still poses a major risk when the surgical tools are not in the field-of-view of the camera, the tissues can easily be damaged by the tools. The camera is controlled by the surgeon through foot pedals by default. A possible solution for reducing the risk associated with the manual control is the autonomous positioning of the endoscopic camera. Detecting and following the movements of the Patient Side Manipulators (PSMs) can be achieved by eye-in-hand visual servoing algorithms using the frames of the endoscopic camera [235].

In this section, an open-source software solution is presented for visual servoing for camera control of a da Vinci classic with the DVRK [83]. The technique of visual ser-

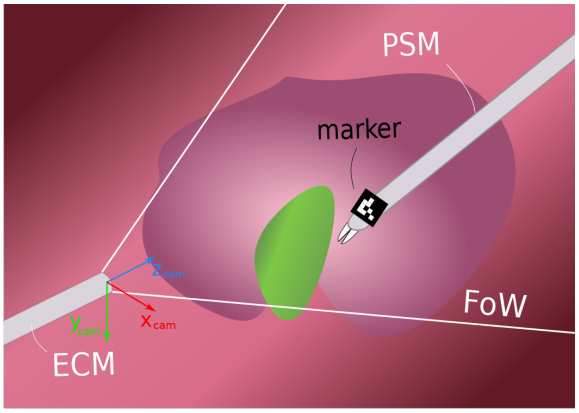

Fig. 5.1. The setup of the visual servoing with the dVSS: The eye-in-hand visual servoing can be executed if the PSMs (detected with markers) are in the field-of-view (FoW) of the ECM.

voing uses visual feedback to control the motion of a robot. The automation of the endoscope's motion can by achieved utilizing visual servoing, by the means of the ECM follows the movements of the PSMs, therefore the surgical tools are always in the field-of view of the camera. Developing the open-source software solution for visual servoing for camera control of a da Vinci classic means creating a more autonomous system for da Vinci, which is considered to be at LoA 2 [20]. At LoA 2, the system is trusted to complete certain tasks or sub-tasks in an autonomous manner, such as suturing or blunt dissection. The visual servoing feature of the proposed framework can assist through various parts of the whole surgery, and its tasks do not involve decision making. In this recent work, the focus is on the visual servoing robot control, thus a marker-based tool detection was employed. The codes of this work are available at https://github.com/ABC-iRobotics/irob-saf/tree/visual_servoing as a part of the iRob Surgical Automation Framework (irob-saf) [12].

5.2.1 Software frameworks

The proposed method uses eye-in-hand visual servoing approach [235]; it allows the autonomous movements of the camera by receiving an excepted goal function (expected view). It relies on stereoscopic images as an input, the positioning of the ECM operates in 3D. The tool detection was achieved via ArUco codes (marker-based detection) placed on the PSM [236]. The schematic figure of the setup is presented in Fig. 5.1.

Robot Operating System

The Robot Operating System[1] (ROS) is a flexible framework for the modular development of robot software [237]. It is a collection of functions, libraries and conventions, aiming to simplify the development of complex and robust robot applications across a wide variety of platforms. It consists of nodes, which are responsible for specific functions. Those

[1]https://www.ros.org/

communicate with each other via messages though channels called topics. The messages are specific data structures serving various purposes.

IROB Surgical Automation Framework The purpose of the `irob-saf`[2] is to facilitate automation of surgical subtasks for the dVSS with the DVRK, offering a modular architecture with built-in functionalities, like parameterizable surgemes, interface to the DVRK or computer vision algorithms and image pipeline from the endoscope [12]. In this this part of the thesis work, the goal was to extend the `irob-saf` with depth-inclusive visual servoing for the ECM.

5.2.2 Extraction of the Instrument Position

As mentioned above, ROS offers extensive stereo vision support, from stereo camera calibration to the calculation of the 3D point cloud of the scene. Utilizing this infrastructure the 3D position of the ArUco marker is calculated as follows. The image coordinates of the corners of the detected ArUco marker is received alongside the 3D point cloud calculated by ROS. The image coordinates of the marker's center is then calculated from the received corners; these coordinates are used to extract the marker's 3D position from the point cloud using the Point Cloud Library (PCL)[3]. The 3D position of the marker is published to the subtask level node.

Robot Control with Visual Servoing

To avoid tissue damage, the purpose of the developed visual servoing algorithm is to keep the tracked surgical instrument within the bounds of the endoscopic camera image, while avoiding unnecessary movements of the ECM, as it might disturb the surgeon. Thus, instead of implementing a controller that would keep the instrument in the center of the image by constantly adjusting the ECM pose, the following approach is proposed. The desired position of the instrument (D) and a distance threshold (t) are to be defined in the camera frame (Fig. 5.1) by the user; the position of the instrument (currently and ArUco marker, M) is tracked by the stereo camera of the ECM (Fig. 5.2). If the distance of the instrument and the desired position exceeds the threshold t in any direction, the ECM is moved to a new pose (TCP'), where the instrument's position in the camera frame is the same as the desired position; it is in the center of the image again. The pose TCP' is calculated as follows.

The motion of the ECM is restricted in a way so that at the point of insertion (inside the trocar) lateral motion is not possible; it would potentially harm the patient. This point is the so-called RCM, which is also the origin of the $base$ frame of the ECM (Fig. 5.2). Due to this lateral restriction, the ECM has 4 DoF: pivot around the RCM, insertion and rotation along the shaft.

The position of points D and M is defined/measured in the coordinate frame of the camera (cam, Fig. 5.1). As the marker is intended to keep near the center of the camera image, in distance q – defined by the surgeon or the assistance – from the camera (Fig. 5.2), the homogeneous coordinates of desired position of the marker is written as follows:

[2]https://github.com/ABC-iRobotics/irob-saf

[3]https://pointclouds.org/

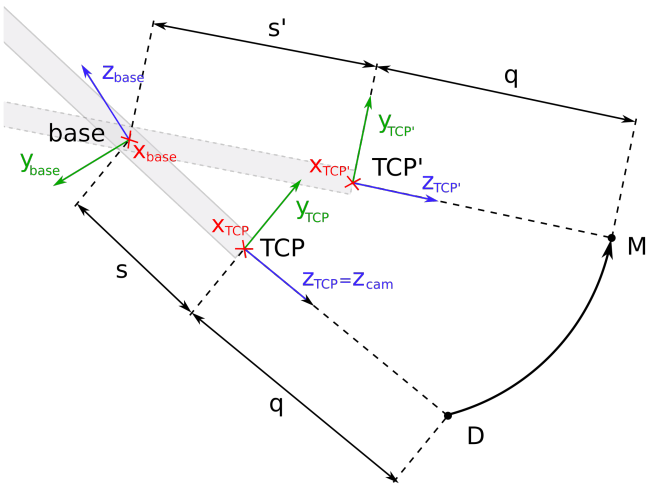

Fig. 5.2. The concept of following the instrument using visual servoing on the da Vinci surgical System. The TCP' pose of the Endoscopic Camera Manipulator (ECM) to be calculated, so the instrument is seen at the desired position in the camera image.

$$d_{cam} = \begin{bmatrix} 0 & 0 & q & 1 \end{bmatrix}^T. \tag{5.1}$$

To calculate the new TCP' pose, the coordinates of those points need to be converted to the TCP frame (Fig. 5.2). In the case of a 0 degree endoscope, this – by neglecting the displacement of the the camera along axis x – is simply a π rotation around axis x, with homogeneous transformation:

$$T_{TCP,cam} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.2}$$

Then, the coordinates of the points D and M in the $base$ frame are calculated using the position and orientation of TCP, received from the DVRK, converted to homogeneous transformation $\mathbf{T}_{TCP,base}$.

Thanks to the ECM's restricted form of motion, the calculation of the pose TCP' gravely simplifies by conversion to a spherical coordinate system around the origin of the $base$ frame. Defining the spherical coordinates of the points D and M in the $base$ frame would easily result in *inclination* π, as those points are typically located in the vicinity of the elongation of axis z_{base} (Fig. 5.2). If the inclination is 0 or π, the *azimuthal angle* is arbitrary; that ambiguity is to be avoided in this application. Thus, the frame of the spherical coordinate system (sph) is rotated by $\pi/2$ along axis x:

$$T_{sph,base} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.3}$$

The homogeneous coordinates of points D and M in the frame sph are then calculated by

$$d_{sph} = T_{sph,base} \cdot T_{TCP,base}^{-1} \cdot T_{TCP,cam} \cdot d_{cam} \tag{5.4}$$

and

$$m_{sph} = T_{sph,base} \cdot T_{TCP,base}^{-1} \cdot T_{TCP,cam} \cdot m_{cam} \tag{5.5}$$

consecutively, where d_{cam} and m_{cam} are the homogeneous coordinates in the camera frame. The spherical coordinates r (radial distance), θ (inclination) and ϕ (azimuthal angle) of those points are calculated by the following formulae:

$$r = \sqrt{x^2 + y^2 + z^2}, \tag{5.6}$$

$$\phi = \arctan2(y, x), \tag{5.7}$$

$$\theta = \arccos \frac{z}{r}. \tag{5.8}$$

By knowing the spherical coordinates of the desired and the current tool locations, the difference of those coordinates can be used to get the transformation from TCP to TCP'. The angular differences:

$$\Delta\phi = \phi_M - \phi_D, \tag{5.9}$$

$$\Delta\theta = \theta_D - \theta_M, \tag{5.10}$$

corresponding to the rotation from TCP to TCP', and distance

$$\Delta r = r_M - r_D \tag{5.11}$$

corresponding to the insertion of the endoscope, are defined. Important to note that the $\Delta\theta$ is intentionally defined with opposite sign, as the the direction of the inclination angle is opposite to the rotation defined by axis x.

From the spherical coordinates it is easy to see that the required length of the ECM's insertion that ensures the q distance from the tracked instrument is $s' = s + \Delta r$, where s is the current length of insertion (Fig. 5.2). The vector for this translational movement is calculated as follows:

$$t_{TCP} = \begin{bmatrix} 0 & 0 & \Delta r \end{bmatrix}^T, \tag{5.12}$$

$$t_{base} = R_{TCP,base} \cdot t_{z,TCP}, \tag{5.13}$$

where $R_{TCP,base}$ is the rotational part of $T_{TCP,base}$. The homogeneous matrix for this translation is:

$$T_r = \begin{bmatrix} I & tbase \\ 0 & 1 \end{bmatrix}. \tag{5.14}$$

As the azimuthal angle defines a rotation around axis z, and the inclination a rotation around axis x, the required rotations from the differences are written:

$$T_{az} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\Delta\theta) & -\sin(\Delta\theta) & 0 \\ 0 & \sin(\Delta\theta) & \cos(\Delta\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.15}$$

Similarly, as inclination defines the rotation along axis z:

$$T_{inc} = \begin{bmatrix} \cos(\Delta\phi) & -\sin(\Delta\phi) & 0 & 0 \\ \sin(\Delta\phi) & \cos(\Delta\phi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.16}$$

Finally, using these transformations the desired pose of the ECM TCP' is calculated:

$$T_{TCP',base} = T_{sph,base}^{-1} \cdot T_{inc} \cdot T_{az} \cdot T_{sph,base} \cdot T_r \cdot T_{TCP,base}. \tag{5.17}$$

Using the desired pose of the ECM, a Cartesian trajectory is generated and being sent to the controller to move the ECM into that pose; positions are generated by linear interpolation, the orientation by spherical linear interpolation (Slerp), offered by the Eigen C++ library[4].

5.2.3 Results of visual servoing

To validate the proposed visual servoing method, experimental scenarios were set up, where the measurement of the accuracy were tested. The tests on the dVSS were done in DVRK teleoperated mode with a phantom representing a gallbladder and its environment. During the test setups, the arms were controlled from the master side by an operator, except the ECM, which was moved autonomously by the proposed method. The anatomically correct surgical phantom (originally created for modeling laparoscopic cholecystectomy) provided the background for the tests, and an ArUco code was placed on the moving PSM (Fig. 5.3).

To determine the accuracy of the system, the ArUco marked PSM was moved in x, y and z directions by the operator. The Cartesian position of the ECM alongside the joint angles related to *yaw*, *pitch* and *insertion* of the endoscope were tracked, then shown in the graphs to easily determine if the created system functions properly. In the experiments, the endoscopic camera of the da Vinci was used with the original 640 × 480 pixels resolution camera of the da Vinci Classic.

First, an experiment of one instrument, with an attached ArUco marker moving in the x axis of the coordinate frame of the camera was performed. The results of the distance from the desired position in the x direction, the *yaw* angle and x component of the ECM's Cartesian position are shown in Fig. 5.4.

Similarly, in the second experiment, the marked instrument was moved along the y axis of the coordinate frame of the camera, while additionally to the marker's distance from

[4]http://eigen.tuxfamily.org

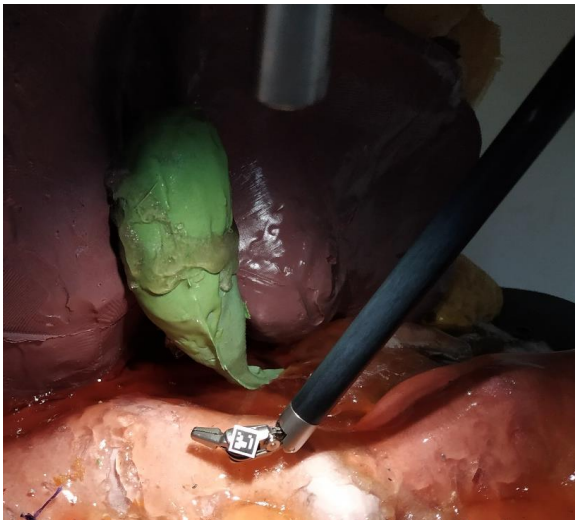

Fig. 5.3. Testing the proposed visual servoing method on the da Vinci Research Kit. An ArUco code was fixed on a surgical tool for instrument tracking, and the endoscope followed the tool's displacements. The tests were done in a surgical phantom environment.

the desired pose in this direction the *pitch* and the y component of the ECM's Cartesian position were recorded (Fig. 5.4).

In the third experiment, the instrument was moved along the z axis of the coordinate frame of the camera; apart from the z component of the distance of the marker from the desired position and z component of the position of the ECM, the values of the *insertion* joint were captured (Fig. 5.4).

In Fig. 5.4, the distance threshold value of $\pm 20\ mm$ is marked with red horizontal lines. The two threshold lines determine a limited area corresponding to the ideal point, where the ECM does not move. Outside this area, the ECM performs its corrections to reach the desired point determined by the thresholds as shown in those figures. The desired values along both axes x and y were 0 (corresponding to the center of the image), and $30\ mm$ along z axis (corresponding to $30\ mm$ distance from the camera).

The data of three graphs above are shown together, alongside the Euclidean distance from the desired position in Fig. 5.4, where the relationship between the directions of the marker's displacement and the changing of corresponding joint coordinates can be observed. In Fig. 5.4a, the absolute value of the instrument displacement in x, y, z directions are shown with the Euclidean distance of the marker from the desired point. In Fig. 5.4b, corresponding joint coordinates of the ECM *(yaw, pitch, insertion)* are shown. If the instrument displacement in x direction is higher than the set threshold, the corresponding joint coordinate of the ECM, the yaw will alter. Similarly, the pitch corrects the instrument displacement in y direction, and the instertion corrects the instrument displacement in z direction. If the movements in the experiments were complex, two or more joints were moving in the given interval.

5.2.4 Limitations

While the results have shown very promising outcomes in correcting distances in all directions, but it is important to note that this solution relies on marker-based tracking, which

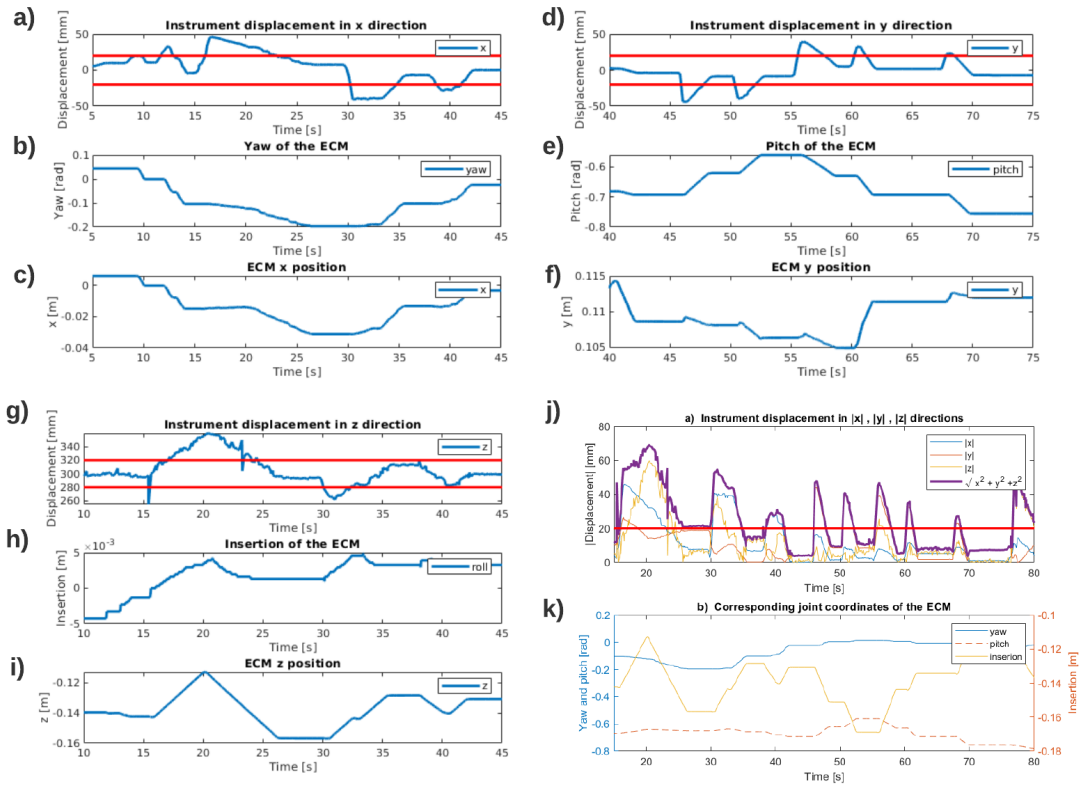

Fig. 5.4. Results of one-marker setup for x (a–c), y (d–f) and z (g–i) axes experiments. The red horizontal lines are the given threshold value for the displacement. J–k: results of one-marker setup for all directions shown in one graph. The red horizontal line is the given absolute threshold value for the displacement length. The purple line is the Euclidean distance of the marker from the desired point.

proves highly reliable in several scenarios (considering factors such as tool velocity). On the other hand, it cannot be applied in a real clinical setting. This problem and a potential solution are introduced in Thesis group 1. Since the proposed visual servoing method was tested on the dVSS with the DVRK, the clinical teleoperation mode was not available, only the research mode was accessible. In this case, teleoperation is sensitive to abrupt motions and larger rotations. Consequently, studies on human experiences may not be directly applicable. Nevertheless, the general feedback from 10 medical students has been positive. Addressing surgeon needs holds significance in the field of automation; however, it remains an underexplored area of study. Based on feedback from surgeons, it becomes evident that delving into this domain could yield substantial benefits.

5.2.5 Conclusion of visual servoing

In this section, a marker-based visual servoing method for RAMIS was proposed to automate camera moving, which can help the surgeon focusing only on the instrument control in DVRK enhanced research mode. The usage of visual markers ensured the elimination of a significant fraction of vision-related errors, thus during the development and validation

the main focus could remain on the robot control aspect of the visual servoing problem. The outcome showed that the framework can be considered a reliable base for future work.

5.3 Autonomous image-based surgical skill assessment based on Optical Flow

Image-based surgical skill assessment can be widely used across surgical domains, since training videos are available. In this section, a two dimensional endoscopic image data-based surgical skill assessment method is proposed, which can be used in any available surgical training videos, if the surgical skills are annotated. For skill assessment, Optical Flow was chosen as an objective video feature, because of its robustness and good computational time performance. If the camera is automated, with proper optical flow ego-motion filtering (which is also discussed in this Thesis group in Section 5.4), the skills of the surgeon can be measured with the proposed method. The codes of this work are available at https://github.com/ABC-iRobotics/VisDataSurgicalSkill.

5.3.1 Data generation

For 2D image-based skill assessment, the JIGSAWS dataset (Section 3.1.3) was used, where Optical Flow was applied as a skill feature (Section 3.3.1).

As a first step, a suitable initial frame is found, where both surgical tools are visible, then user-selected Region of Interest (ROI) are preprocessed and saved. The frames are first turned gray-scale, then blurred by a median filter, and ran through a binary filter using adaptive thresholds, in order to denoise the frames, and enable the better detection of features, using the Shi–Tomasi detector on the respective ROIs.

The resulting output constitutes the initial features to be tracked. Each video is traversed, set to the initial frame, the features of which are then tracked by the Kanade–Lucas OF. The features are extracted from each frame of the video, and collected in a list, with the dimensions: $frame\_number \times sample\_size \times 2 \times 2$. This list is then iterated through, and for each frame's data a row of 240 features – made up of the respective OF and Positional information of each tool – is added to the output. Given the ROI data and the generated output, grouping the data according to the surgical tasks and the expertise level of users constitutes the final output. This is accomplished using the sliding window preprocessing method implemented by Anh et al. for their benchmark [238], to process the multivariate time series and separate chunks of the data into uniformly sized local windows, thereby enabling the evaluation of the data with the same networks originally designed for the kinematic data of JIGSAWS. The workflow can be found in Fig. 5.5.

5.3.2 Classification Methods

To counteract overfitting – i.e., a higher model complexity than that of the data itself, learning outliers and noise as part of the pattern, thereby losing on generalisation-capability – the LOOCV cross-validation technique was used, along with L2 regularization for each of

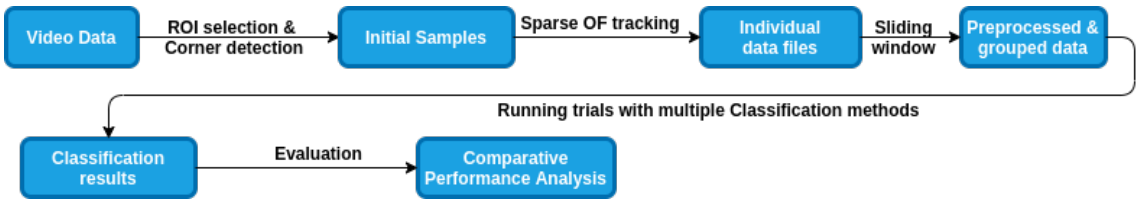

Fig. 5.5. The proposed workflow: First we compute and save the initial samples from each video input using user-selected Regions of Interest, then by tracking them with the sparse Lucas–Kanade Of – tracking the movement of both surgical tools independently – creating data files, that are further processed by a sliding window method, outputting the final input data. Using several different classification methods, it is possible to determine the users' expertise.

the methods introduced below. Same padding is used in most places, to keep the dimensions of the output.

$$output\_width = \frac{W - F_w + 2P}{S_w} + 1, \tag{5.18}$$

$$output\_height = \frac{H - F_h + 2P}{S_h} + 1, \tag{5.19}$$

where W and H are the width and height of input, F are filter dimensions and P is the padding size (i.e., the number of rows or columns to be padded). In case of same padding, the following stands:

$$output\_height = ceil\left(\frac{H}{S_h}\right), \tag{5.20}$$

$$output\_width = ceil\left(\frac{W}{S_w}\right). \tag{5.21}$$

Convolutional Neural Network

Commonly used for classification, segmentation and image processing, Convolutional Neural Networks (CNNs) ensure translation invariance and parameter sharing through convolution (essentially specialised sliding filter operations) [239, 240]. They are based on the assumption that nearby data points are more closely correlated than further ones. In the used data (see Section 5.3.1), the two surgical tools are separated, therefore closer datapoints are more likely to belong to the same tool. CNN relies on such local dependencies, and in this case this is further supported by the the arrangement in which the Optical Flow displacement values – describing speed in relation to the framerate – and the Position pixel-coordinates are side-by-side, with the rows corresponding to timestamps/frames.

The network uses two one-dimensional convolutional layers – with the Rectified Linear Unit (ReLU) activation functions, introducing non-linearity to the system with the equation $ReLu(z) = max(0, z)$ followed by batch normalization layers. Then a one-dimensional Global Average Pooling (GAP) is applied to reduce data complexity and avoid overfitting by reducing the total number of parameters in the models. Ultimately a Dense layer with softmax activation ($softmax(z_i) = \frac{exp(z_i)}{\Sigma_j exp(z_j)}$) is used to create the output as a vector of three, corresponding to the three classes of the classification.

Long Short-term Memory

Long Short-term Memory networks (LSTMs) are specialised Neural Networks by design, ideal for time series data analysis [241]. Using the combination of three types of gates (input, output and forget) and a dedicated memory cell – storing the internal representation of the learned information – they are able to record long term data representations. Their use is beneficial in surgical skill assessment, where the sequentiality of actions and precision movements is relevant for the successful execution of surgical subtasks. Suitable subtask-segmentation may be required, in order to avoid the confusion of patterns belonging to different sequences.

The input gate takes the input from the current time-stamp, $x_t \in R^N$, the output from the previous LSTM-unit, h_{t-1}, and the previous memory cell, c_{t-1}. LSTMs make use of the sigmoid activation function ($\sigma(z) = \frac{1}{1+e^{-z}}$, bounded between $[0, 1]$) to process the input information:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i). \tag{5.22}$$

Forget gates serve as a sort of selection method. This one layer neural network determines whether the given data should be kept or removed from the current internal state:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f. \tag{5.23}$$

The learned knowledge is stored in memory cells. These cells are updated by combining ($\odot$) the memory cell with the new information, where ($\odot$) denotes the Hadamard product:

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c). \tag{5.24}$$

And finally, the output gate controls the information passed onto the next LSTM block:

$$h_t = o_t \odot tanh(c_t), \qquad o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o). \tag{5.25}$$

CNN+LSTM

A straightforward combination of the previous two models: two convolutional layers – batch optimization and ReLu activation functions – followed by two LSTM blocks, CNN+LSTM is a slightly more complex Neural Network architecture. Among others, Li et al. [242] have demonstrated high accuracy predictions using the combination of these methods. The temporal information of the convolutional layers' outputs is processed by the LSTM blocks, in order to learn contextually, but from an already processed information source.

Residual Neural Network

Primarily used for classification tasks, using so-called skip connections, as shortcuts to solve the degradation problem [243] – essentially short circuiting shallow layers to deep layers (Residual Neural Networks (ResNets) enables the creation of deeper networks without loss of performance). They are reliable techniques even within smaller networks. The model of the benchmark also consists of only 3 blocks, meaning that it does not utilize its strength to the fullest (given, that there are not many layers to skip), but it still can perform accurate classification.

Convolutional Autoencoder

Traditionally used for data compression, dimensionality reduction and the denoising of data without significant information loss, an autoencoder – often symmetric, built up of two blocks: an encoder and a decoder – is an unsupervised machine learning model [244]. The encoder aims to create a copy of its input as an output, reverse engineering the problem by trying to find the right filter. The autoencoder created by Ahn et al. [238] is built using convolutional layers. It is also common to use fully-connected layers. The encoder compresses the input time series into a latent space representation, then the network tries restructuring it into the original input data in the decoder. For classification, after the network has been fully trained, the encoder's output is fed to an SVM classifier.

Frequency domain transformations

The Discrete Fourier Transformation (DFT) and the Discrete Cosine Transformation (DCT) are traditionally used to transform time series data from the time domain to the frequency domain. The use of frequency features in surgical skill assessment has been proven to perform well by Zia et al. [245].

The main assumption of these techniques is that the more competent a user is in the given skill, the smoother and more predictable the time series representation will be. The initial $X \in \mathbb{R}^{N \times L}$ time series, where N is the feature size (240) and L is the sample size (the given videos' frame number) is initially separated into univariate time series, the frequency coefficients of which are calculated and concatenated with a frequency matrix $Y \in \mathbb{R}^{N \times L}$.

After computing matrix Y, iterating through each row, the highest peaks are kept, others discarded. With $F < L$ being the number of peaks kept, the reduced matrix can be calculated $\hat{Y} \in \mathbb{R}^{N \times F}$. The F highest peaks represent the most significant frequencies of the input segment. Classification is achieved by converting $\hat{Y}$ into a one-dimensional matrix.

The two methods only differ essentially in the frequency matrix used:

$$\text{DFT:} \quad Y_k = \Sigma_{t=0}^{L-1} x_l \times e^{-\frac{i2\pi kl}{L}}, \tag{5.26}$$

$$\text{DCT:} \quad Y_k = \frac{1}{2}x_0 + \Sigma_{l=1}^{L-1} x_l \times cos[\frac{\pi}{L} l(k + \frac{1}{2})]. \tag{5.27}$$

The periodicity of DFT breaks the continuity, while DCT is fully continuous.

5.3.3 Results of optical flow-based surgical skill assessment

Each method were run 5 times for each generated input file. Within each run there were 5 trials, using the LOOCV cross-validation method, then the mean accuracy was calculated.

The intermediate class is prone to misclassification [246]. Funke et al.'s hybrid 3D network has misclassified every intermediate surgeon into either expert or novice, using Optical Flow data for knot-tying [247]. This may partially be due to data disparity, as intermediate and expert subjects are underrepresented in comparison to novices [248]. Anh et al.'s benchmark also faced issues with the classification of intermediate users [238].

TABLE 5.1
THE PERFORMANCE OVERVIEW OF EACH METHOD IN THE CASE OF 2 OR 3 CLASSES, RESPPECTIVELY.

Method	Eval.	class num.	Min. Run (%)	Max. Run (%)	Min. Trial (%)	Max. Trial (%)
CNN	Model	2	73.7	79.12	54.34	93.67
		3	49.6	66.83	35.69	75.44
CNN	SVM	2	53.97	66.53	29.45	94.68
		3	28.15	56.8	43.65	70.59
LSTM	Model	2	51.16	79.62	24.44	90.64
		3	40.48	62.19	15.25	69.63
LSTM	SVM	2	40.95	80.44	22.22	91.94
		3	30.2	60.87	15.25	70.86
CNN+ LSTM	Model	2	74.69	83.19	56.42	93.65
		3	56.12	73.09	46.37	82.65
CNN+ LSTM	SVM	2	58.987	73.44	28.27	93.23
		3	33.12	68.57	5.15	82.94
ResNet	Model	2	73.75	83.54	54.55	95.74
		3	47.93	70.25	33.46	80.0
ResNet	SVM	2	54.21	73.64	23.3	93.04
		3	25.92	61.5	7.15	79.17
convAuto	SVM	2	58.58	75.52	33.33	93.21
		3	30.82	52.57	16.14	77.36

Given this, intermediate subjects were excluded from the main evaluation, and only comparatively analysed 3-class classification, the result of which is presented in Table 5.1.

The following results have been obtained without the data of intermediate subjects, which means binary classification.

CNN

With a minimum standard deviation of 1.43 % and best mean accuracy of 80.72 %, CNN has responded well to the data.

LSTM

Although its best mean accuracy of 80.44 % is promising, its standard deviation ranging from 5.97 % to 15.49 %, as well as a high number of trials with zero true positives for experts show that LSTM by itself is not complex enough to fit to the data.

CNN+LSTM

Able to counteract the shortcomings of the simple LSTM, it improved on the results of both models, with 83.19 % maximum mean accuracy and consistently less than 2 % standard deviation. Its highest mean accuracy was 79.19 %, even outperforming the more complex ResNet model (78.65 %).

ResNet

ResNet modelled the data the best, with 84.23 % highest mean accuracy and 1.36 % standard deviation. Having the highest number of layers, it suggests that the data scales in performance with model complexity.

convAuto

The only unsupervised method in the benchmark, convAuto reached a maximum of 79.77 %, with an average of 5 % standard deviation. Its highest mean of 67.78 % is among the lower ones, but given that it relies on the SVM classifier, its efficacy could be improved by tuning its hyperparameters, or employing a kernel trick.

5.3.4 Results by surgical tasks

Knot-Tying

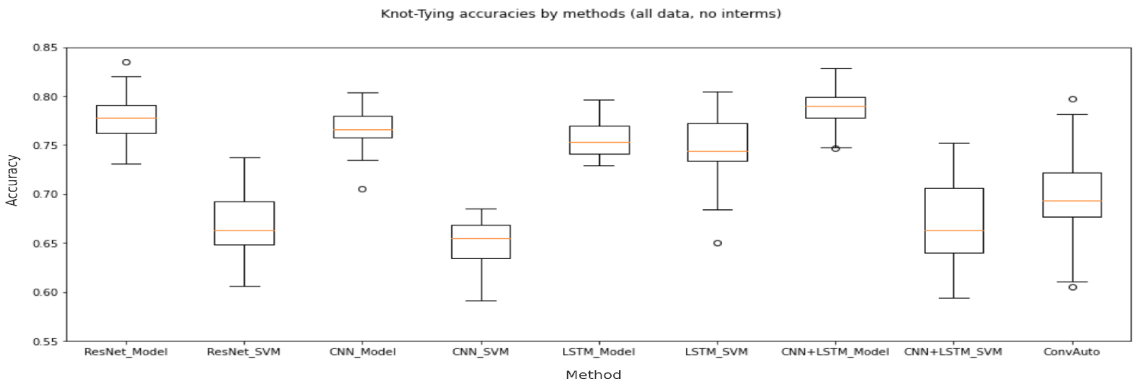

Fig. 5.6. Knot-Tying accuracies without intermediates. The best perfoming methods were: ResNet and CNN and LSTM.

Ming et al. found Knot-Tying to be the easiest surgical task to assess with both STIP and iDT [133]. The proposed method is similar in principle to their STIP method, as it also tracks the movement of interest points/features over time.

When it comes to the model evaluation, even the worst average accuracy (produced by LSTM) was at 74.75 %. Regarding SVM evaluation the same value was 62.1 %. Coincidentally the highest SVM evaluated accuracy (80.43 %) and the highest average SVM accuracy (76.07 %) were also by LSTM. Given that model-evaluated LSTM results had many outliers, SVM could improve the recall and precision. For the best performing configuration of LSTM, 8 out of 25 trials (32 %) had 0 expert true positives. To measure the efficacy specifically for expert classification, the Recall metric needs to be used:

$$\text{ExpertRecall} = \frac{\text{ExpertTruePositive}}{\text{NumOfExpertsInTrial}}. \tag{5.28}$$

The highest individual expert recall of LSTM was 89.7 %, but given the 8 cases where the value is 0, its mean is 24.36 %.

Overall ResNet, CNN+LSTM and convAuto all performed well for the Knot-Tying task. LSTM has done well, but its low mean expert recall leaves the need for further investigation before it could be deemed as reliable. CNN – considered in general to be a top-performer [246] – has fallen behind. Its highest accuracy was 80.42 %, and highest mean accuracy 78.38 %. SVM predictions dropped its accuracy below 70 %. ResNet's model-based evaluation resulted in the highest accuracy (83.54 %), while the highest mean accuracy (80.11 %) came from the model-predictions of CNN+LSTM. Figure 5.6 shows the ranges of accuracies for each used method with Knot-Tying skill data.

Suturing

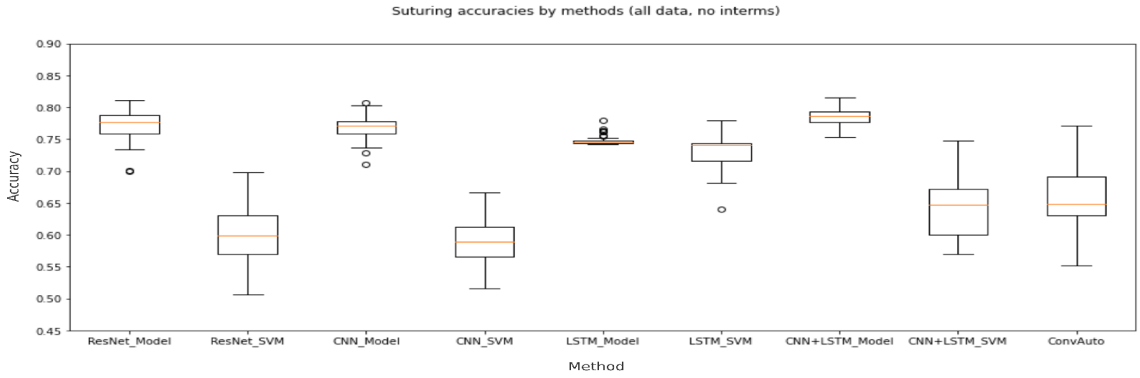

Fig. 5.7. Suturing accuracies without intermediates. The best performing methods are ResNet, CNN and CNN and LSTM.

The same observations apply to Suturing as to Knot-Tying: ResNet, CNN+LSTM and convAuto performed well, LSTM has shown high results in some cases, accompanied by confusion matrix anomalies (the maximum Expert Recall only being 3.33 %, with an overall average of 0.22 %), and CNN seemingly performed the worst, even though it still did so above 75 % on Model average accuracies. Even though Ming et al. [133] also found that Suturing is harder to classify than Knot-Tying, CNN has been found to be one of the most reliable methods by the review of Yanik et al. [246]. It is possible that the CNN model of the benchmark is too small, and it would perform better with higher complexity and more layers. Figure 5.7 illustrates the accuracy of each applied method, given the Suturing task.

Needle-Passing

Ming et al. claimed that Needle-Passing was the hardest skill to perform classification for, because they did not find significant differences between the trajectories of expert and novice users' left hand movements [133]. With this data generation method, only LSTM dropped significantly in efficacy in comparison with its performance on the other skills.

The highest Model average precision (79.74 %) and the highest SVM accuracy (71.58 %) were both achieved by CNN+LSTM, making it the overall best for the skill of Needle-Passing. The range of accuracies given all the Needle-Passing data for each method is illustrated in Figure 5.8.

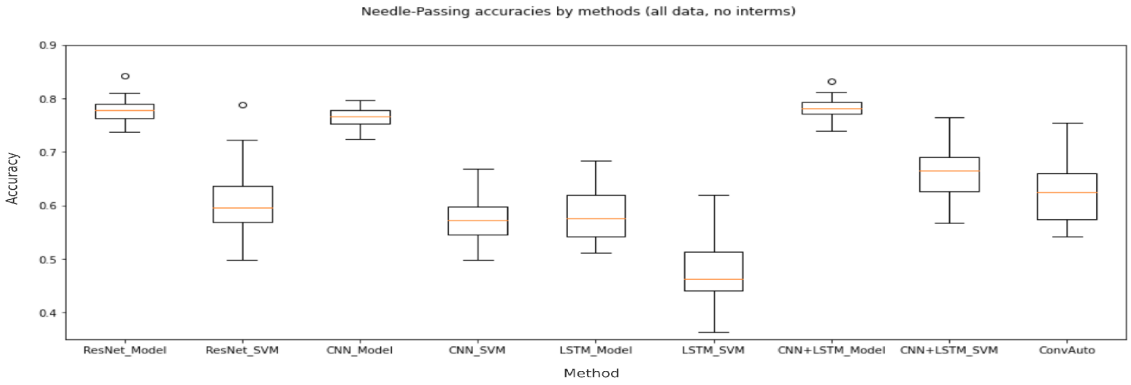

Fig. 5.8. Needle-Passing accuracies without intermediates. ResNet, CNN and CNN and LSTM outperform the other methods.

5.3.5 Performance analysis

The generated data combined with the benchmark of Anh et al. [238] has successfully outperformed the solutions of Ming et al. [133] in both Suturing and Knot-Tying, and only slightly fell short of their results in Needle-Passing. The goal was to find a method that can achieve similar results to the state of the art in the field of surgical skill assessment, while keeping generalisation and practicality in mind, in order to keep it relevant for manual MIS training, where 3D and pose information is not available. Table 5.2 presents the detailed comparison of these methods.

Residual Neural Network performed the best, closely followed by the combined model of CNN and LSTM. DFT and DCT have performed too poorly to be able to rely on them for classification. Even though LSTM's accuracy is high, its confusion matrices showed it to be unreliable. It has been observed that although the SVM classification decreases the overall accuracy, it performs more consistently overall. This is likely due to the fact that SVMs are suspectible to outliers [249]. They perform well in high-dimensional feature spaces, and although there are 240 features, the number of rows greatly outnumber this.

TABLE 5.2
THE COMPARISON OF THE RESULTS TO THAT OF THE STATE OF THE ART. EVERY METHOD LISTED HERE USED THE JIGSAWS DATASET AS VISUAL INPUT DATA SOURCE.

Author (Year)	Method	ST	NP	KT
Funke et al. (2019)	3D ConvNet +TSN [247]	100 %	96.4 %	95.8 %
Ming et al. (2021)	STIP[133]	79.29 %	87.01 %	72.57 %
Ming et al. (2021)	iDT[133]	76.79 %	83.81 %	76.65 %
Lajkó et al.	CNN[238]	80.72 %	79.66 %	80.41 %
Lajkó et al.	CNN+LSTM[238]	81.58 %	83.19 %	82.82 %
Lajkó et al.	ResNet[238]	81.89 %	84.23 %	83.54 %

DFT

DFT and DCT are often evaluated side-by-side, and even though their similarities seem to outweigh their differences at first sight, it is customary to investigate both, to see whether these seemingly small differences add up to significant differences in performance. The

DFT relies on a Support Vector Machine to classify the data based on the feature labels extracted with the help of the transformation. The results indicate that this method is not suited to evaluate OF-based data in this form, because even though its highest mean accuracy (73.83 %) indicates that it has the potential to classify well about two thirds of the cases, overall it is closer to 50 %, which means that it unable to make reliable predictions. DFT and DCT work with the assumption that the more experienced someone is, the smoother their data's time series is going to be represented in the frequency domain. This assumption holds well in the realm of kinematics, but OF may lack the sufficient features, or dimensions for this to hold true. Given an analysis it would seem that these transformations depend on the quality, rather than the quantity of data.

DCT

The Discrete Cosine Transformation works on similar principles to the DFT, and also uses a Support Vector Machine for classification. The highest mean accuracy reached with this method was 78.61 %, but similarly to DFT, it falls short in overall consistent accuracy. When comparing the two transformations, DCT has a clear edge over DFT – most likely due to its periodicity, which contrary to DFT does not introduce discontinuity. DCT outperformed DFT in all three metrics, and it gets close to being efficient enough, but it is still not suitable for reliable classification.

5.3.6 Conclusion of optical flow-based skill assessment

This part of my thesis work aimed to create a practical, generally applicable solution that can enable the creation of visual-based surgical skill assessment methods, and can potentially lead to the inclusion of automated skill assessment in the curriculum of minimally invasive surgical training, introducing benefits such as objectivity, reproducibility, and the fact that it would not require human expertise. The proposed method outperformed the state of the art in 2D visual-based skill assessment, with more than 80 % accuracy for all two from three surgical subtasks available in JIGSAWS (Knot-Tying and Suturing). By introducing new visual features – such as image-based orientation and image-based collision detection – or from the evaluation side: utilizing other SVM kernel methods, tuning the hyperparameters, or using boosted tree algorithms instead, classification accuracy can be further improved.

5.4 Optical Flow ego-motion filtering

In this thesis group, Optical Flow was proven to be an effective image feature for surgical skill assessment. However, optical flow-based skill assessment can be a much more complex problem, when the viewpoint is moving as well; Optical Flow techniques cannot make a difference between motion originated from moving objects in the space and from the self-motion of a moving camera. The motion of the viewpoint – called "self-motion", or "ego-motion" – has to be extracted from the Optical Flow vector field to detect moving objects in the space. In this section, an Optical Flow ego-motion filtering method for background motion compensation in the case of a moving viewpoint, with access to the

robot's state of motion and depth information is introduced. The result is the Optical Flow vector field without the flow vectors originated from the movement of the viewpoint. The proposed method can provide the velocity of the moving object in the space. The codes of this work are available at https://github.com/ABC-iRobotics/egomotion-filter.

5.4.1 Optical Flow ego-motion compensation method

In the following equations, the method is presented through only one pixel. For ego-motion filtering, it is necessary to take these steps for the whole image.
Based on the Optical Flow algorithm, the pixel displacements can be calculated; then the current pixel locations can be calculated:

$$dx = x_i - x_{i-1}, \tag{5.29}$$

$$dy = y_i - y_{i-1}, \tag{5.30}$$

where x_i, y_i is the current pixel location, and x_{i-1}, y_{i-1} is the previous pixel location. The pixel to camera coordinate method can be calculated from the camera coordinate to pixel perspective projection equation:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{s_x} & 0 & o_x & 0 \\ 0 & \frac{f}{s_y} & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}, \tag{5.31}$$

where x_i, y_i are the pixel coordinates, the intrinsic parameters are the focal length (f), principal point (o_x, o_y), and pixel size (s_x, s_y). X_i, Y_i, Z_i are the camera coordinates. Expand the perspective projection equation the following equations is derived:

$$x_i = \frac{1}{s_x} f \frac{X_i}{Z_i} + o_x, \tag{5.32}$$

$$y_i = \frac{1}{s_y} f \frac{Y_i}{Z_i} + o_y. \tag{5.33}$$

Since the depth information is available (Z_i), the point coordinates can be calculated:

$$X_i = \frac{s_x}{f} Z_i (x_i - o_x), \tag{5.34}$$

$$Y_i = \frac{s_y}{f} Z_i (y_i - o_v). \tag{5.35}$$

The x_{i-1}, y_{i-1} previous pixels have to be deprojected with the same method (Eq. 5.34, 5.35). From the current ($X_i, Y_i\ Z_i$) and previous camera coordinates $(X_{i-1}, Y_{i_1}, Z_{i-1})$ and the elapsed time (dt) the current optical velocity can be calculated ($v_{i,opt}$):

$$v_{i,opt} = \frac{(X_i, Y_i, Z_i) - (X_{i-1}, Y_{i-1}, Z_{i-1})}{dt}. \tag{5.36}$$

To compensate the ego-motion, the reference velocity has to be subtracted from the optical velocity. For this, transformation the camera coordinates with the transformation matrix is necessary originated from the robot's translation and rotation to get the actual camera coordinates:

$$\begin{bmatrix} X_{i,self} \\ Y_{i,self} \\ Z_{i,self} \\ 1 \end{bmatrix} = \begin{bmatrix} R_{cam_{2,1}} & t_{cam_{2,1}} \\ 0\,0\,0 & 1 \end{bmatrix} \begin{bmatrix} X_{i-1} \\ Y_{i-1} \\ Z_{i-1} \\ 1 \end{bmatrix}, \tag{5.37}$$

where $(X_{i,self}, Y_{i,self}, Z_{i,self})$ are the current camera coordinates originated from the viewpoint's motion. After that, the reference velocity can be calculated:

$$v_{i,self} = \frac{(X_{i,self}, Y_{i,self}, Z_{i,self}) - (X_{i-1}, Y_{i_1}, Z_{i-1})}{dt}. \tag{5.38}$$

From $v_{i,opt}$ and the reference velocity $vi, self$ the filtered Optical Flow can be estimated:

$$v_{i,filtered} = v_{i,opt} - v_{i,self}, \tag{5.39}$$

where $v_{i,filtered}$ is the 3D ego-motion filtered optical flow.

5.4.2 Results of Optical Flow ego-motion filtering

To test the accuracy of the implemented Optical Flow ego-motion filtering method, two Universal Robots UR5 manipulators were used [250] because of their accuracy, which was crucial to test the exact performance of the algorithm. The test scenarios were set up under the following conditions:

- The first robot arm holds the camera attached to the last link (known transformation to the robot's base coordinate system);
- The second robot holds a test object attached to the last link (known transformation to the robot's base coordinate system);
- Known transformation between the two robots' base frame.

Both of the robot arms moved on predefined trajectories with synchronized logging of their state of motion and the camera frames. The standard Hough transform was employed for the test object segmentation [251].

For testing purposes, six different scenarios were set up, where the object holder and the camera holder arm moved under different conditions. The details of the scenarios can be found in Table 5.3, where every value is shown in the camera coordinate system. The camera coordinate system is right-handed with the y axis pointing down, x axis pointing right, and $z$axis pointing away from the camera. During the motions, the velocities were constant. In Table 5.3 $v_{cam_{linear}}$ is the linear velocity of the camera, $v_{cam_{angular}}$ is the angular velocity of the camera, and $v_{obj_{linear}}$ is the test object's linear velocity. "Distance" is the distance between the camera and the moving object at the start point of the recording. In Scenarios 1, 2 and 3 there were only translational movements performed by the camera

TABLE 5.3

SCENARIO SETTINGS FOR TESTING THE PROPOSED METHOD; $v_{cam_{linear}}$: LINEAR VELOCITY OF THE CAMERA, $v_{cam_{angular}}$: ANGULAR VELOCITY OF THE CAMERA, $v_{obj_{linear}}$: TEST OBJECT'S LINEAR VELOCITY, "DISTANCE": THE DISTANCE BETWEEN THE CAMERA AND THE MOVING OBJECT AT THE START POINT OF THE RECORDING.

Scen.#	$v_{cam_{linear}}$ ((x,y,z), [m/s])	$v_{cam_{angular}}$ ((x,y,z), [rad/s])	$v_{obj_{linear}}$ ((x,y,z), [m/s])	Distance ([m])
1	(0.072, 0, 0)	(0, 0, 0)	(-0.072, 0, 0)	0.33
2	(0.072, 0, 0)	(0, 0, 0)	(-0.069, 0.012, 0)	0.33
3	(0.021, 0.018, 0.015)	(0, 0, 0)	(-0.033, 0, 0)	0.36
4	(0, 0, 0)	(0, 0, 0.5445)	(0.057, 0, 0)	0.23
5	(0, 0, 0)	(1.617, 0, 0)	(0, -0.057, 0)	0.24
6	(0, 0, 0)	(1.617, 0, 0)	(0, -0.057, 0)	0.35

TABLE 5.4

RESULTS OF OPTICAL FLOW EGO-MOTION FILTERING AND MOVING OBJECT STATE OF MOTION ESTIMATION; STD: STANDARD DEVIATION, MAE: MEAN ABSOLUTE ERROR.

Scen.#	Mean $v_{obj_{linear}}$ ((x,y,z), [m/s])	Std. $v_{obj_{linear}}$ ((x,y,z), [m/s])	MAE ((x,y,z), [m/s])
1	(-0.071, -0.001, 0)	(0.009, 0.002, 0.006)	(0.001, -0.001, 0)
2	(-0.068, 0.002, 0.004)	(0.013, 0002, 0.009)	(0.001, -0.01, 0.004)
3	(-0.016, 0.015, 0.020)	(0.029, 0.004, 0.06)	(0.017, 0.015, 0.020)
4	(0.058, 0, 0.027)	(0.007, 0.002, 0.005)	(0.001, 0, 0.027)
5	(-0.001, 0.074, -0.02)	(0.004, 0.016, 0.002)	(-0.001, 0.131, -0.02)
6	(0, 0.078, -0.006)	(0.0023, 0.009, 0.011)	(0, 0.135, -0.006)

and the test object. In Scenario 4, 5 and 6, the camera performed rotational movement. The accuracy was tested with different velocities and different distances.

The main goal of this research was to filter the robot's motion from the Optical Flow vector field. To test the accuracy of the background filtering, the ratio of the number of filtered pixels can be calculated and the number of moving pixels before the filtering in the background. The background was extracted based on the depth information. These calculations showed promising results for all of the experiments (best case scenario showed 99.6 % accuracy; the worst case was 89.3 %, Table 5.5). Based on these findings, the proposed method shows high accuracy results of ego-motion background filtering. Another

TABLE 5.5

RESULTS OF OPTICAL FLOW EGO-MOTION FILTERING ACCURACY.

Scen.#	Background filter accuracy [%]
1	98.0
2	98.5
3	99.6
4	89.3
5	93.9
6	90.0

approach to measuring the accuracy of ego-motion filtering is to compare the segmented moving object's state of motion after the filtering to the object's reference state of motion. For this, the test object was segmented on the pre-filtered image and compared the calculated velocity with the test object holder arm's velocity. The results can be found in Table 5.4, where the notions are the same as in Table 5.3. Mean, standard deviation (Std)

and Mean Absolute Error (MAE) were calculated from a set of frames. Very accurate results were gotten in the case of Scenario 1, 2, and 4, and low accuracy in the case of Scenario 2, 5, and 6 (Fig. 5.9). The results suggest that movements without depth changing (in this case translation in X and Y direction and rotation around Z-axis) can be easier to filter to the algorithm, but movements including depth changing (in this case rotation around Y-axis and translation in X, Y and Z directions) can be more complex to filter. Higher velocity differences between objects of interest and the camera can provide more accurate results. On the other hand, the relative movement direction between the object and the camera can be significant as well: if they are moving in the same direction, it is harder to extract ego-motion (Scenarios 5 and 6). Based on these findings, the proposed Optical Flow ego-motion filter solution an ideal case is where the robot's and the moving object's state of motion is significantly differs on the projected image plane. It is important to note that the performance of the method highly depends on the accuracy of the depth information provided by a depth sensor.

5.4.3 Conclusion of Optical Flow ego-motion compensation

In this section, an Optical Flow ego-motion compensation algorithm was intruduced. It is based on two-dimensional Farneback dense Optical Flow and image depth information. The camera's translational and rotational movement reference frame is known in this approach. The accuracy was tested with a moving test object, whose state of motion is also known. The background filter results showed very high accuracy: 94.88 % on average, in the different test scenarios.

5.5 Summary of the Thesis group

Surgical automation is a next step in the evolution of RAMIS. In this thesis group, a complete framework for OF-based surgical skill assessment in the case of autonomous endoscope motion was introduced to assess surgical skills while automation is employed. I proposed an autonomous endoscopic camera motion algorithm with visual servoing, tested on the da Vinci Surgical System. For surgical skill assessment, an accurate OF-based method was presented; it outperformed the state of the art, and I have proved its applicability in RAMIS with a robotic surgery skill-annotated database. Finally, an OF ego-motion compensation method was proposed, to extract only surgical tool motions in the visual scene. This framework can be employed in future safety examinations, where the motion of the endoscope is automated and the skills of the surgeon is assessed.
Related publications: [RNE10, RNE11, RNE12, RNE13, RNE14, RNE15]

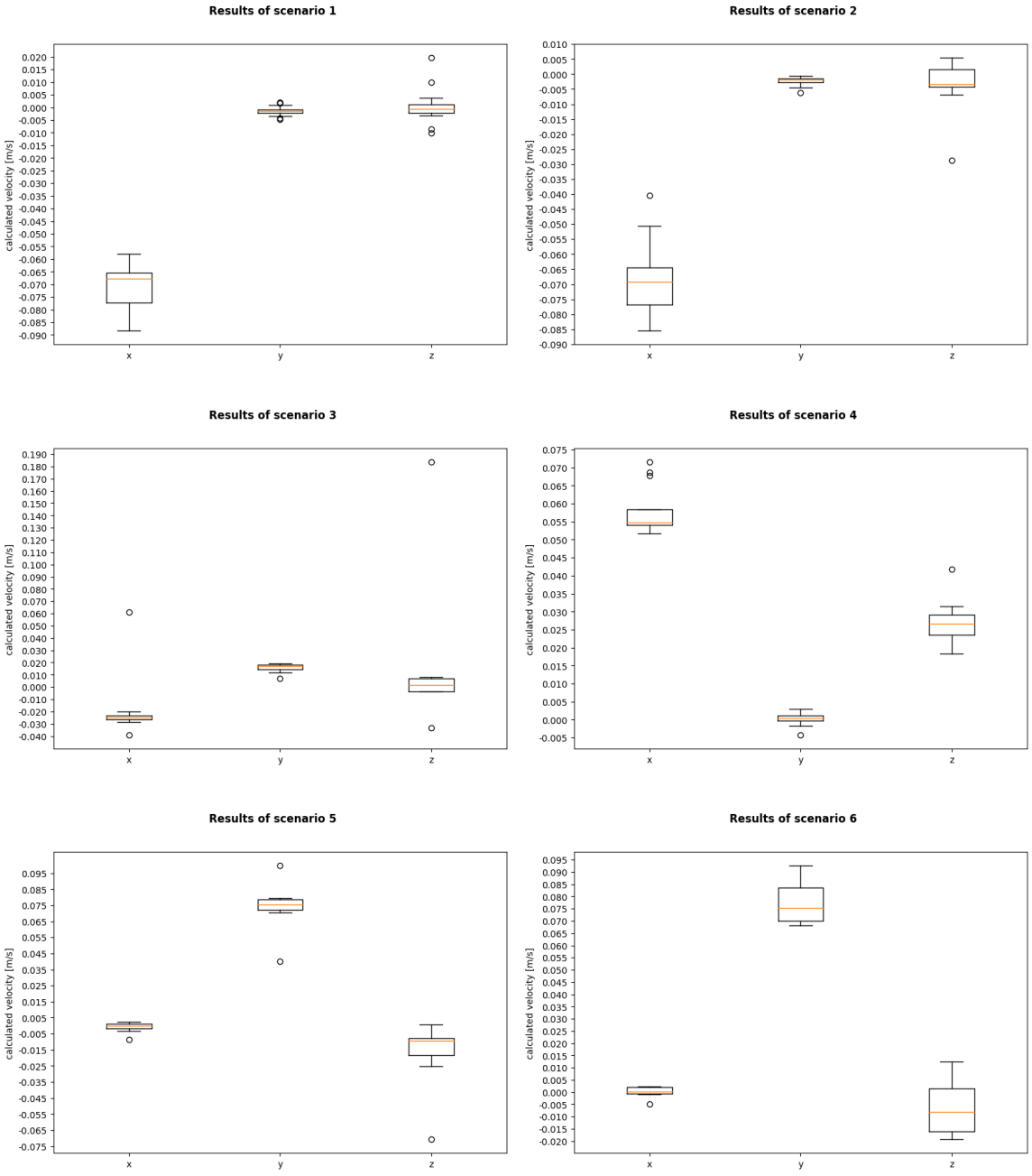

Fig. 5.9. State of motion estimation boxplot results by frames after Optical Flow ego-motion filtering in the different scenarios. Very accurate results were found in the case of Scenario 1, 2, and 4, and lower accuracy in the case of Scenario 2, 5, and 6.

Chapter 6

CONCLUSION

6.1 Summary of contributions and future work

In this thesis work, automated skill assessment in manual and Robot-Assisted Minimally Invasive Surgery (RAMIS) was studied. Endoscopic image data-based surgical technical skill assessment was shown, additional sensor and motion analysis-based non-technical surgical skill assessment was proposed, and finally, a framework for image-based skill assessment in the case of autonomous camera motion was introduced.

Technical robotic surgical skills were discussed, and the possibilities to autonomously measure these skills. For this, a 2D endoscopic image data-based skill classification method was introduced, where kinematic data was generated based on the image features. Its applicability in skill classification was examined as well. Non-technical surgical skills are also possible to estimate objectively with kinematic, endoscopic or additional sensor-based autonomous approaches. I identified the most important non-technical skills in Robot-Assisted Minimally Invasive Surgery, and the possibilities to assess these skills manually and autonomously. I showed that with Artificial Intelligence-based methods, certain non-technical skills can be classified autonomously. Finally, a method for autonomous camera motion for the da Vinci Surgical System was introduced, and I proposed an Optical Flow-based method for surgical skill assessment, which can be utilized when automation is employed. Furthermore, I proposed an Optical Flow ego-motion filter algorithm to extract only the surgical tool'smotion.

My thesis work has the potential to contribute significantly to the field of autonomous skill assessment and training for surgical residents. Following my doctoral research, my focus shifted towards assessing both technical and non-technical skills among surgeon residents and console surgeons using the da Vinci Xi surgical system, incorporating physiological signal analysis into the evaluation process. This collaborative effort involved Semmelweis University (Budapest, Hungary) and Queen's University (Kingston, Canada). Surgical skill assessment plays a vital role in the emerging field of Surgical Data Science (SDS), and I have been applying my expertise to SDS studies at the Austrian Center for Medical Innovation and Technology (ACMIT). My research findings have also been integrated into various projects, such as "Robot-enhanced skill improvement and assessment in minimally invasive surgery" (OTKA PD116121, 2015-2017) and "Research on Localization and Object Detection Based on Heterogeneous Sensors" (GINOP-2.2.1-15-2017-00097, 2018-2021). Additionally, my work in the domain of DVRK-related research

has been incorporated into `irob-saf`, a framework for robotic surgery automation. In my role as an assistant lecturer at the John von Neumann Faculty of Informatics, Óbuda University, I have had the privilege of sharing my research insights with BSc and MSc students. This experience has allowed me to collaborate closely with talented students from institutions such as Óbuda University, Queen's University, Eötvös Loránd University, Budapest University of Technology and Economics, and Pázmány Péter Catholic University. Looking ahead, my future projects involve non-technical skill assessment with Hungarian console surgeons in clinical settings and the development of an autonomous Surgical Process Modeling system to support clinical decision-making.

6.2 New Scientific Results

Thesis group 1

In my first thesis group, I proposed solutions for robotic surgery technical skill assessment, based on endoscopic camera data.

Thesis 1/I: I proposed a surgical tool pose estimation technique and semantic segmentation algorithm for the da Vinci Surgical System's articulated tools, applicable to autonomous technical skill assessment. My surgical tool pose estimation method provides a generic solution, since it does not require kinematic data or a complete model of the tool. It performed adequately on 2D endoscopic images, based on shape features and iterative Perspective n Point transformation method. The method's accuracy was proven in surgical skill assessment, where it outperformed the state of the art on knot-tying videos (89.33 % accuracy).

Thesis 1/II: Ground truth generation for semantic segmentation for the JIGSAWS dataset was introduced, which was not available before. Using DNN-based methods, I provided an accurate solution for semantic segmentation of the surgical tools for the JIGSAWS dataset, for image-based skill assessment (97.38 % accuracy, 79.91 % dice score) with TernausNet. Based on my results, image-based surgical technical skill assessment can be a good alternative to kinematic data-based skill assessment for RAMIS-type systems.

Related publications: [RNE1, RNE2, RNE3, RNE4, RNE5]

Thesis group 2

In my second thesis group, I proved the correlation between objectively measured data and surgical non-technical skills in robotic surgery.

Thesis 2/I: I proposed a methodology for autonomous non-technical skill assessment and workload evaluation in laparoscopic cholecystectomy training. The training was studied with workload assessment (SURG-TLX). Laparoscopic phantom training and workflow were introduced to simulate stressful situations during surgery (bleeding, time-critical reaction, distractions, physical demands). Statistical tests showed significant differences between the two groups (medical professionals and a non-expert control group) in the case of mental demands, physical demands and situational stress ($p<0.001$, 95 % CI). Learning curve in task complexity resulted significant difference between the first and the second trials ($p<0.05$, 95 % CI).

Thesis 2/II: It has been shown that there are non-technical skills of the surgeon which can be classified based on sensory data (image and force) during the surgeon's training, based on a Fully Convolutional Neural Network (with 85 % accuracy). This finding can lead to objective and autonomous surgical non-technical skill assessment, and the proposed training environment can be suitable for personalized training as well.

Related publications: [RNE6, RNE7, RNE8, RNE9]

Thesis group 3

I proposed a framework for Optical Flow-based surgical skill assessment in the case of autonomous endoscope motion to assess surgical skills, while ECM automation is employed.

Thesis 3/I: An accurate OF-based surgical skill assessment algorithm was presented; it outperformed the state of the art, and I have proved its applicability in RAMIS skill assessment. The results were tested on the JIGSAWS dataset: in suturing, needle-passing, and knot-tying tasks it provided 81.89 %, 84.23 % and 83.54 % accuracy, respectively, with ResNet classification.

Thesis 3/II: I proposed an autonomous endoscopic camera motion algorithm with visual servoing, tested on the dVSS. The method could successfully compensate the distances between PSMs and the ECM relying on a marker-based approach.

Thesis 3/III: An OF ego-motion compensation method was developed, to extract only surgical tool motions in the visual scene. This framework can be employed with future safety examinations, where the motion of the endoscope is automated, while the skills of the surgeon are being assessed.

Related publications: [RNE10, RNE11, RNE12, RNE13, RNE14, RNE15]

Further publications related to the Ph.D. Thesis and the accompanying research work: [RNENR1, RNENR2, RNENR3, RNENR4, RNENR5, RNENR6, RNENR7]

REFERENCES

[1] R. H. Taylor, A. Menciassi, G. Fichtinger, P. Fiorini, and P. Dario, "Medical Robotics and Computer-Integrated Surgery," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Cham: Springer International Publishing, 2016, pp. 1657–1684.

[2] J. T. O'Donovan, B. Kang, and T. Höllerer, "Competence Modeling in Twitter: Mapping Theory to Practice," in *Proceedings of the ASE Bigdata/Socialcom/Cybersecurity Conference*, 2015, pp. 1–15.

[3] D. Azari, C. Greenberg, C. Pugh, D. Wiegmann, and R. Radwin, "In search of characterizing surgical skill," *Journal of surgical education*, vol. 76, no. 5, pp. 1348–1363, 2019.

[4] "Intuitive robotic-assisted surgery da vinci surgical system," https://www.intuitive.com/en-us, Access date: 2023.10.27.

[5] "IEC 80601-2-77:2019," https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/84/68473.html, Access date: 2023.10.27.

[6] T. Haidegger, S. Speidel, D. Stoyanov, and R. M. Satava, "Robot-assisted minimally invasive surgery—surgical robotics in the data age," *Proceedings of the IEEE*, vol. 110, no. 7, pp. 835–846, 2022.

[7] D. Julian, A. Tanaka, P. Mattingly, M. Truong, M. Perez, and R. Smith, "A comparative analysis and guide to virtual reality robotic surgical simulators," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 14, no. 1, 2018.

[8] "BBZ - Medical Technologies," http://www.bbzsrl.com/index.html, access date: 2023.10.27.

[9] "SEP robot trainer," http://surgrob.blogspot.com/2013/10/sep-robot-trainer.html, access date: 2023.10.27.

[10] A. J. Hung, J. Chen, A. Jarc, D. Hatcher, H. Djaladat, and I. S. Gill, "Development and Validation of Objective Performance Metrics for Robot-Assisted Radical Prostatectomy: A Pilot Study," *J. Urol.*, vol. 199, no. 1, pp. 296–304, 2018.

[11] A. Zia and I. Essa, "Automated surgical skill assessment in RMIS training," *Int J Comput Assist Radiol Surg*, vol. 13, no. 5, pp. 731–739, 2018.

[12] T. D. Nagy and T. Haidegger, "A dvrk-based framework for surgical subtask automation," *Acta Polytechnica Hungarica*, pp. 61–78, 2019.

[13] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 1578–1585.

[14] S. G. Hart, "NASA-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006.

[15] M. R. Wilson, J. M. Poolton, N. Malhotra, K. Ngo, E. Bright, and R. S. W. Masters, "Development and validation of a surgical workload measure: The surgery task load index (SURG-TLX)," *World Journal of Surgery*, vol. 35, no. 9, pp. 1961–1969, Sep. 2011.

[16] B. Sharma, A. Mishra, R. Aggarwal, and T. P. Grantcharov, "Non-technical skills assessment in surgery," *Surgical Oncology*, vol. 20, no. 3, pp. 169–177, Sep. 2011.

[17] N. Raison, T. Wood, O. Brunckhorst, T. Abe, T. Ross, B. Challacombe, M. S. Khan, G. Novara, N. Buffi, H. Van Der Poel, C. McIlhenny, P. Dasgupta, and K. Ahmed, "Development and validation of a tool for non-technical skills evaluation in robotic surgery-the ICARS system," *Surg Endosc*, vol. 31, no. 12, pp. 5403–5410, 2017.

[18] A. Takács, D. A. Nagy, I. Rudas, and T. Haidegger, "Origins of Surgical Robotics: From Space to the Operating Room," *Acta Polytechnica Hungarica*, vol. 13, pp. 13–30, 2016.

[19] G. Fichtinger, J. Troccaz, and T. Haidegger, "Image-guided interventional robotics: Lost in translation?" *Proceedings of the IEEE*, vol. 110, no. 7, pp. 932–950, 2022.

[20] T. Haidegger, "Autonomy for Surgical Robots: Concepts and Paradigms," *IEEE Transactions on Medical Robotics and Bionics*, vol. 1, no. 2, pp. 65–76, May 2019.

[21] E.-Y. A. Tsui, "Application of Artificial Intelligence (AI) in Surgery," https://www.imperial.ac.uk/news/200673/application-artificial-intelligence-ai-surgery/, 2020, [Online; Access date: 2023.10.27.].

[22] T. Habuza, A. N. Navaz, F. Hashim, F. Alnajjar, N. Zaki, M. A. Serhani, and Y. Statsenko, "AI applications in robotics, diagnostic image analysis and precision medicine: Current limitations, future trends, guidelines on CAD systems for medicine," *Informatics in Medicine Unlocked*, vol. 24, pp. 1–8, Jan. 2021.

[23] T. Haidegger, J. Sándor, and Z. Benyó, "Surgery in space: The future of robotic telesurgery," *Surgical Endoscopy*, vol. 25, no. 3, pp. 681–690, Mar. 2011.

[24] J. L. Ochsner, "Minimally Invasive Surgical Procedures," *The Ochsner Journal*, vol. 2, no. 3, pp. 135–136, Jul. 2000.

[25] J. D. Hernandez, S. D. Bann, Y. Munz, K. Moorthy, V. Datta, S. Martin, A. Dosis, F. Bello, A. Darzi, and T. Rockall, "Qualitative and quantitative analysis of the learning curve of a simulated surgical task on the da Vinci system," *Surgical Endoscopy And Other Interventional Techniques*, vol. 18, no. 3, pp. 372–378, Mar. 2004.

[26] P. Gomes, "Surgical robotics: Reviewing the past, analysing the present, imagining the future," *Robotics and Computer-Integrated Manufacturing*, vol. 27, no. 2, pp. 261–266, Apr. 2011.

[27] J. Heemskerk, H. R. Zandbergen, S. W. M. Keet, I. Martijnse, G. van Montfort, R. J. A. Peters, V. Svircevic, R. A. Bouwman, C. G. M. I. Baeten, and N. D. Bouvy, "Relax, it's just laparoscopy! A prospective randomized trial on heart rate variability of the surgeon in robot-assisted versus conventional laparoscopic cholecystectomy," *Digestive Surgery*, vol. 31, no. 3, pp. 225–232, 2014.

[28] D. Stefanidis, F. Wang, J. R. Korndorffer, J. B. Dunne, and D. J. Scott, "Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload," *Surgical Endoscopy*, vol. 24, no. 2, pp. 377–382, Feb. 2010.

[29] K. Cleary and C. Nguyen, "State of the Art in Surgical Robotics: Clinical Applications and Technology Challenges," *Computer Aided Surgery*, vol. 6, no. 6, pp. 312–328, Jan. 2001.

[30] J. Sándor, B. Lengyel, T. Haidegger, G. Saftics, G. Papp, A. Nagy, and G. Wéber, "Minimally invasive surgical technologies: Challenges in education and training," *Asian Journal of Endoscopic Surgery*, vol. 3, no. 3, pp. 101–108, 2010.

[31] R. Smith, V. Patel, and R. Satava, "Fundamentals of robotic surgery: A course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development," *The international journal of medical robotics + computer assisted surgery: MRCAS*, vol. 10, no. 3, pp. 379–384, Sep. 2014.

[32] A. Peña, "The dreyfus model of clinical problem-solving skills acquisition: a critical perspective," *Medical education online*, vol. 15, no. 1, pp. 1–11, 2010.

[33] J. Rasmussen, "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 3, pp. 257–266, May 1983.

[34] T. M. Kowalewski and T. S. Lendvay, "Performance Assessment," in *Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties*, ser. Comprehensive Healthcare Simulation, D. Stefanidis, J. R. Korndorffer Jr., and R. Sweet, Eds. Cham: Springer International Publishing, 2019, pp. 89–105.

[35] A. N. Sridhar, T. P. Briggs, J. D. Kelly, and S. Nathan, "Training in Robotic Surgery—an Overview," *Curr Urol Rep*, vol. 18, no. 8, 2017.

[36] B. Gibaud, G. Forestier, C. Feldmann, G. Ferrigno, P. Gonçalves, T. Haidegger, C. Julliard, D. Katić, H. Kenngott, L. Maier-Hein, K. März, E. de Momi, D. Á. Nagy, H. Nakawala, J. Neumann, T. Neumuth, J. Rojas Balderrama, S. Speidel, M. Wagner, and P. Jannin, "Toward a standard ontology of surgical process models," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 9, pp. 1397–1408, Sep. 2018.

[37] S. Yule and S. Paterson-Brown, "Surgeons' Non-technical Skills," *Surgical Clinics*, vol. 92, no. 1, pp. 37–50, Feb. 2012.

[38] L. Hull, S. Arora, R. Aggarwal, A. Darzi, C. Vincent, and N. Sevdalis, "The Impact of Nontechnical Skills on Technical Performance in Surgery: A Systematic Review," *Journal of the American College of Surgeons*, vol. 214, no. 2, pp. 214–230, Feb. 2012.

[39] J. W. Collins, P. Dell'Oglio, A. J. Hung, and N. R. Brook, "The Importance of Technical and Non-technical Skills in Robotic Surgery Training," *European Urology Focus*, vol. 4, no. 5, pp. 674–676, Sep. 2018.

[40] C. Urban, P. Galambos, G. Györök, and T. Haidegger, "Simulated Medical Ultrasound Trainers A Review of Solutions and Applications," *Acta Polytechnica Hungarica*, vol. 15, no. 7, pp. 111–131, 2018.

[41] J. Collins and P. Wisz, "Training in robotic surgery, replicating the airline industry. How far have we come?" *World Journal of Urology*, vol. 38, no. 7, pp. 1645–1651, 2020.

[42] R. Sánchez, O. Rodríguez, J. Rosciano, L. Vegas, V. Bond, A. Rojas, and A. Sanchez-Ismayel, "Robotic surgery training: Construct validity of Global Evaluative Assessment of Robotic Skills (GEARS)," *Journal of Robotic Surgery*, vol. 10, no. 3, pp. 227–231, Sep. 2016.

[43] R. Berguer and W. Smith, "An ergonomic comparison of robotic and laparoscopic technique: The influence of surgeon experience and task complexity," *The Journal of Surgical Research*, vol. 134, no. 1, pp. 87–92, Jul. 2006.

[44] M. I. Klein, J. S. Warm, M. A. Riley, G. Matthews, K. Gaitonde, J. F. Donovan, and C. R. Doarn, "Performance, Stress, Workload, and Coping Profiles in 1st Year Medical Students' Interaction with the Endoscopic/Laparoscopic and Robot-Assisted Surgical Techniques," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 52, no. 12, pp. 885–889, Sep. 2008.

[45] Y. Ahmed, Z. Lone, A. A. Hussein, Y. Feng, H. Khan, S. Broad, R. Kannappan, A. Skowronski, A. Cole, D. Wang, K. Stone, A. Hasasneh, K. Sexton, A. Gotsch, T. Ali, J. Braun, S. Khan, A. Durrani, M. Durrani, and K. A. Guru, "Do surgeon non-technical skills correlate with teamwork-related outcomes during robot-assisted surgery?" *BMJ Leader*, vol. 3, no. 3, pp. 69–74, Sep. 2019.

[46] N. Raison, K. Ahmed, T. Abe, O. Brunckhorst, G. Novara, N. Buffi, C. McIlhenny, H. van der Poel, M. van Hemelrijck, A. Gavazzi, and P. Dasgupta, "Cognitive training for technical and non-technical skills in robotic surgery: A randomised controlled trial," *BJU International*, vol. 122, no. 6, pp. 1075–1081, Dec. 2018.

[47] S. B. Shafiei, A. A. Hussein, and K. A. Guru, "Dynamic changes of brain functional states during surgical skill acquisition," *PloS One*, vol. 13, no. 10, p. e0204836, 2018.

[48] "The da Vinci Research Kit (dVRK) – Intuitive Foundation," http://www.intuitive-foundation.org/dvrk/, Access date: 2023.10.27.

[49] K. Chinzei, "Safety of surgical robots and iec 80601-2-77: The first international standard for surgical robots," *Acta Polytechnica Hungarica*, vol. 16, pp. 171–184, 2019.

[50] T. Haidegger, "Taxonomy and standards in robotics," *In Marcelo H. Ang, Oussama Khatib, and Bruno Siciliano (eds), Encyclopedia of Robotics, Springer Nature*, pp. 1–12, 2021.

[51] N. Riem, S. Boet, M. D. Bould, W. Tavares, and V. N. Naik, "Do technical skills correlate with non-technical skills in crisis resource management: A simulation study," *British Journal of Anaesthesia*, vol. 109, no. 5, pp. 723–728, Nov. 2012.

[52] J. Chen, N. Cheng, G. Cacciamani, P. Oh, M. Lin-Brande, D. Remulla, I. S. Gill, and A. J. Hung, "Objective assessment of robotic surgical technical skill: a systematic review," *The Journal of urology*, vol. 201, no. 3, pp. 461–469, 2019.

[53] A. A. Hussein, K. R. Ghani, J. Peabody, R. Sarle, R. Abaza, D. Eun, J. Hu, M. Fumo, B. Lane, J. S. Montgomery, N. Hinata, D. Rooney, B. Comstock, H. K. Chan, S. S. Mane, J. L. Mohler, G. Wilding, D. Miller, K. A. Guru, and Michigan Urological Surgery Improvement Collaborative and Applied Technology Laboratory for Advanced Surgery Program, "Development and Validation of an Objective Scoring Tool for Robot-Assisted Radical Prostatectomy: Prostatectomy Assessment and Competency Evaluation," *J. Urol.*, vol. 197, no. 5, pp. 1237–1244, 2017.

[54] A. A. Hussein, K. J. Sexton, P. R. May, M. V. Meng, A. Hosseini, D. D. Eun, S. Daneshmand, B. H. Bochner, J. O. Peabody, R. Abaza, E. C. Skinner, R. E. Hautmann, and K. A. Guru, "Development and validation of surgical training tool: Cystectomy assessment and surgical evaluation (CASE) for robot-assisted radical cystectomy for men," *Surg Endosc*, vol. 32, no. 11, pp. 4458–4464, 2018.

[55] A. A. Hussein, N. Hinata, S. Dibaj, P. R. May, J. D. Kozlowski, H. Abol-Enein, R. Abaza, D. Eun, M. S. Khan, J. L. Mohler, P. Agarwal, K. Pohar, R. Sarle, R. Boris, S. S. Mane, A. Hutson, and K. A. Guru, "Development, validation and clinical application of Pelvic Lymphadenectomy Assessment and Completion Evaluation: Intraoperative assessment of lymph node dissection after robot-assisted radical cystectomy for bladder cancer," *BJU Int.*, vol. 119, no. 6, pp. 879–884, 2017.

[56] A. A. Hussein, R. Abaza, C. Rogers, R. Boris, J. Porter, M. Allaf, K. Badani, M. Stifelman, J. Kaouk, T. Terakawa, Y. Ahmed, E. Kauffman, Q. Li, K. Guru, and D. Eun, "Development and validation of an objective scoring tool for Minimally Invasive Partial Nephrectomy:: scoring for partial nephrectomy," *The Journal of Urology*, vol. 199, no. 4, pp. 159–160, 2018.

[57] H. Husslein, L. Shirreff, E. M. Shore, G. G. Lefebvre, and T. P. Grantcharov, "The Generic Error Rating Tool: A Novel Approach to Assessment of Performance and Surgical Education in Gynecologic Laparoscopy," *J Surg Educ*, vol. 72, no. 6, pp. 1259–1265, 2015.

[58] H. Husslein, E. Bonrath, T. Grantcharov, and G. Lefebvre, "Validation of the Generic Error Rating Tool (GERT) in Gynecologic Laparoscopy (Preliminary Data)," *Journal of Minimally Invasive Gynecology*, vol. 20, no. 6, p. S106, 2013.

[59] P. Ramos, J. Montez, A. Tripp, C. K. Ng, I. S. Gill, and A. J. Hung, "Face, content, construct and concurrent validity of dry laboratory exercises for robotic training using a global assessment tool," *BJU Int.*, vol. 113, no. 5, pp. 836–842, 2014.

[60] A. C. Goh, D. W. Goldfarb, J. C. Sander, B. J. Miles, and B. J. Dunkin, "Global evaluative assessment of robotic skills: Validation of a clinical assessment tool to measure robotic surgical skills," *J. Urol.*, vol. 187, no. 1, pp. 247–252, 2012.

[61] R. Sánchez, O. Rodríguez, J. Rosciano, L. Vegas, V. Bond, A. Rojas, and A. Sanchez-Ismayel, "Robotic surgery training: Construct validity of Global Evaluative Assessment of Robotic Skills (GEARS)," *J Robot Surg*, vol. 10, no. 3, pp. 227–231, 2016.

[62] M. A. Aghazadeh, I. S. Jayaratna, A. J. Hung, M. M. Pan, M. M. Desai, I. S. Gill, and A. C. Goh, "External validation of Global Evaluative Assessment of Robotic Skills (GEARS)," *Surg Endosc*, vol. 29, no. 11, pp. 3261–3266, 2015.

[63] M. Liu, S. Purohit, J. Mazanetz, W. Allen, U. S. Kreaden, and M. Curet, "Assessment of Robotic Console Skills (ARCS): Construct validity of a novel global rating scale for technical skills in robotically assisted surgery," *Surg Endosc*, vol. 32, no. 1, pp. 526–535, 2018.

[64] K. R. Ghani, D. C. Miller, S. Linsell, A. Brachulis, B. Lane, R. Sarle, D. Dalela, M. Menon, B. Comstock, T. S. Lendvay, J. Montie, J. O. Peabody, and Michigan Urological Surgery Improvement Collaborative, "Measuring to Improve: Peer and Crowd-sourced Assessments of Technical Skill with Robot-assisted Radical Prostatectomy," *Eur. Urol.*, vol. 69, no. 4, pp. 547–550, 2016.

[65] A. Guni, N. Raison, B. Challacombe, S. Khan, P. Dasgupta, and K. Ahmed, "Development of a technical checklist for the assessment of suturing in robotic surgery," *Surg Endosc*, vol. 32, no. 11, pp. 4402–4407, 2018.

[66] Q. Ballouhey, P. Clermidi, J. Cros, C. Grosos, C. Rosa-Arsène, C. Bahans, F. Caire, B. Longis, R. Compagnon, and L. Fourcade, "Comparison of 8 and 5 mm robotic instruments in small cavities : 5 or 8 mm robotic instruments for small cavities?" *Surg Endosc*, vol. 32, no. 2, pp. 1027–1034, 2018.

[67] S. L. Vernez, V. Huynh, K. Osann, Z. Okhunov, J. Landman, and R. V. Clayman, "C-SATS: Assessing Surgical Skills Among Urology Residency Applicants," *J. Endourol.*, vol. 31, no. S1, pp. S95–S100, 2017.

[68] A. J. Hung, T. Bottyan, T. G. Clifford, S. Serang, Z. K. Nakhoda, S. H. Shah, H. Yokoi, M. Aron, and I. S. Gill, "Structured learning for robotic surgery utilizing a proficiency score: A pilot study," *World J Urol*, vol. 35, no. 1, pp. 27–34, 2017.

[69] A. Volpe, K. Ahmed, P. Dasgupta, V. Ficarra, G. Novara, H. van der Poel, and A. Mottrie, "Pilot Validation Study of the European Association of Urology Robotic Training Curriculum," *Eur. Urol.*, vol. 68, no. 2, pp. 292–299, 2015.

[70] N. Takeshita, S. J. Phee, P. W. Chiu, and K. Y. Ho, "Global Evaluative Assessment of Robotic Skills in Endoscopy (GEARS-E): Objective assessment tool for master and slave transluminal endoscopic robot," *Endosc Int Open*, vol. 6, no. 8, pp. E1065–E1069, 2018.

[71] H. Niitsu, N. Hirabayashi, M. Yoshimitsu, T. Mimura, J. Taomoto, Y. Sugiyama, S. Murakami, S. Saeki, H. Mukaida, and W. Takiyama, "Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room," *Surg Today*, vol. 43, no. 3, pp. 271–275, 2013.

[72] N. Y. Siddiqui, M. L. Galloway, E. J. Geller, I. C. Green, H.-C. Hur, K. Langston, M. C. Pitter, M. E. Tarr, and M. A. Martino, "Validity and reliability of the robotic Objective Structured Assessment of Technical Skills," *Obstet Gynecol*, vol. 123, no. 6, pp. 1193–1199, 2014.

[73] M. R. Polin, N. Y. Siddiqui, B. A. Comstock, H. Hesham, C. Brown, T. S. Lendvay, and M. A. Martino, "Crowdsourcing: A valid alternative to expert evaluation of robotic surgery skills," *Am. J. Obstet. Gynecol.*, vol. 215, no. 5, pp. 644.e1–644.e7, 2016.

[74] M. E. Tarr, C. Rivard, A. E. Petzel, S. Summers, E. R. Mueller, L. M. Rickey, M. A. Denman, R. Harders, R. Durazo-Arvizu, and K. Kenton, "Robotic objective structured assessment of technical skills: A randomized multicenter dry laboratory training pilot study," *Female Pelvic Med Reconstr Surg*, vol. 20, no. 4, pp. 228–236, 2014 Jul-Aug.

[75] "Intuitive Surgical Investor Presentation 021218 — Surgery — Cardiothoracic Surgery," https://www.scribd.com/document/376731845/Intuitive-Surgical-Investor-Presentation-021218, access date: 2023.10.27.

[76] F. Bovo, G. De Rossi, and F. Visentin, "Surgical robot simulation with BBZ console," *J Vis Surg*, vol. 3, 2017.

[77] "Intuitive products services education training," https://www.intuitive.com/en/products-and-services/da-vinci/education, access date: 2023.10.27.

[78] "Intuitive Surgical da Vinci Si Surgical System Skills Simulator," https://www.intuitivesurgical.com/products/skills_simulator/, access date: 2023.10.27.

[79] A. Tanaka, C. Graddy, K. Simpson, M. Perez, M. Truong, and R. Smith, "Robotic surgery simulation validity and usability comparative analysis," *Surg Endosc*, vol. 30, no. 9, pp. 3720–3729, 2016.

[80] H. Schreuder, R. Wolswijk, R. Zweemer, M. Schijven, and R. Verheijen, "Training and learning robotic surgery, time for a more structured approach: A systematic review: Training and learning robotic surgery," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 119, no. 2, pp. 137–149, 2012.

[81] R. Smith, M. Truong, and M. Perez, "Comparative analysis of the functionality of simulators of the da Vinci surgical robot," *Surg Endosc*, vol. 29, no. 4, pp. 972–983, 2015.

[82] R. Kumar, A. Jog, B. Vagvolgyi, H. Nguyen, G. Hager, C. C. G. Chen, and D. Yuh, "Objective measures for longitudinal assessment of robotic surgery training," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 143, no. 3, pp. 528–534, 2012.

[83] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da Vinci® Surgical System," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 6434–6439.

[84] "ROS.org Powering the world's robots," http://www.ros.org/, access date: 2023.10.27.

[85] "CISST/SAW stack for the da Vinci Research Kit. Contribute to jhu-dvrk/sawIntuitiveResearchKit," https://github.com/jhu-dvrk/sawIntuitiveResearchKit, access date: 2023.10.27.

[86] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh *et al.*, "JHU-ISI gesture and skill assessment working set (JIGSAWS): A surgical activity dataset for human motion modeling," in *MICCAI workshop: M2cai*, vol. 3, no. 3, 2014, pp. 1–10.

[87] K. Ruda, D. Beekman, L. W. White, T. S. Lendvay, and T. M. Kowalewski, "SurgTrak — A Universal Platform for Quantitative Surgical Data Capture," *Journal of Medical Devices*, vol. 7, no. 3, Jul. 2013.

[88] E. D. Gomez, R. Aggarwal, W. McMahan, K. Bark, and K. J. Kuchenbecker, "Objective assessment of robotic surgical skill using instrument contact vibrations," *Surg Endosc*, vol. 30, no. 4, pp. 1419–1431, 2016.

[89] T. N. Judkins, D. Oleynikov, and N. Stergiou, "Objective evaluation of expert and novice performance during robotic surgical training tasks," *Surgical endoscopy*, vol. 23, pp. 590–597, 2009.

[90] I. Nisky, M. H. Hsieh, and A. M. Okamura, "The effect of a robot-assisted surgical system on the kinematics of user movements," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, pp. 6257–6260, 2013.

[91] M. J. Fard, S. Ameri, R. B. Chinnam, A. K. Pandya, M. D. Klein, and R. D. Ellis, "Machine learning approach for skill evaluation in robotic-assisted surgery," *In Proceedings of the World Congress on Engineering and Computer Science, vol. I*, pp. 1–8, 2016.

[92] Y. Sharon, T. S. Lendvay, and I. Nisky, "Instrument Orientation-Based Metrics for Surgical Skill Evaluation in Robot-Assisted and Open Needle Driving," *arXiv:1709.09452 [cs]*, pp. 1–12, 2017.

[93] M. J. Fard, S. Ameri, R. D. Ellis, R. B. Chinnam, A. K. Pandya, and M. D. Klein, "Automated robot-assisted surgical skill evaluation: Predictive analytics approach," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 14, no. 1, p. e1850.

[94] Z. Wang and A. M. Fey, "Satr-dl: improving surgical skill assessment and task recognition in robot-assisted surgery with deep neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 1793–1796.

[95] Y. Sharon and I. Nisky, "What Can Spatiotemporal Characteristics of Movements in RAMIS Tell Us?" *Journal of Medical Robotics Research*, pp. 1–14, 2018.

[96] K. Liang, Y. Xing, J. Li, S. Wang, A. Li, and J. Li, "Motion control skill assessment based on kinematic analysis of robotic end-effector movements," *Int J Med Robot*, vol. 14, no. 1, pp. 1–9, 2018.

[97] Z. Wang and A. Majewicz Fey, "Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery," *International journal of computer assisted radiology and surgery*, vol. 13, pp. 1959–1970, 2018.

[98] J. D. Brown, C. E. O Brien, S. C. Leung, K. R. Dumon, D. I. Lee, and K. J. Kuchenbecker, "Using Contact Forces and Robot Arm Accelerations to Automatically Rate Surgeon Skill at Peg Transfer," *IEEE Trans Biomed Eng*, vol. 64, no. 9, pp. 2263–2275, 2017.

[99] M. Ershad, R. Rege, and A. M. Fey, "Meaningful Assessment of Robotic Surgical Style using the Wisdom of Crowds," *Int J Comput Assist Radiol Surg*, vol. 13, no. 7, pp. 1037–1048, 2018.

[100] R. Kumar, A. Jog, A. Malpani, B. Vagvolgyi, D. Yuh, H. Nguyen, G. Hager, and C. C. Grace Chen, "Assessing system operation skills in robotic surgery trainees," *Int J Med Robot*, vol. 8, no. 1, pp. 118–124, 2012.

[101] L. Maier-Hein, S. Vedula, S. Speidel, N. Navab, R. Kikinis, A. Park, M. Eisenmann, H. Feussner, G. Forestier, S. Giannarou, M. Hashizume, D. Katic, H. Kenngott, M. Kranzfelder, A. Malpani, K. März, T. Neumuth, N. Padoy, C. Pugh, N. Schoch, D. Stoyanov, R. Taylor, M. Wagner, G. D. Hager, and P. Jannin, "Surgical Data Science: Enabling Next-Generation Surgery," *Nature Biomedical Engineering*, vol. 1, no. 9, pp. 691–696, 2017.

[102] C. E. Reiley and G. D. Hager, "Task versus subtask surgical skill evaluation of robotic minimally invasive surgery," *Med Image Comput Comput Assist Interv*, vol. 12, no. Pt 1, pp. 435–442, 2009.

[103] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," *Computer Aided Surgery*, vol. 11, no. 5, pp. 220–230, 2006.

[104] C. E. Reiley, H. C. Lin, B. Varadarajan, B. Vagvolgyi, S. Khudanpur, D. D. Yuh, and G. D. Hager, "Automatic recognition of surgical motions using statistical modeling for capturing variability," in *Studies in Health Technology and Informatics*, 2008, pp. 396–401.

[105] B. Varadarajan, C. Reiley, H. Lin, S. Khudanpur, and G. Hager, "Data-Derived Models for Segmentation with Application to Surgical Assessment and Training," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, ser. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2009, pp. 426–434.

[106] L. Tao, E. Elhamifar, S. Khudanpur, G. D. Hager, and R. Vidal, "Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation," in *Information Processing in Computer-Assisted Interventions*, ser. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2012, pp. 167–177.

[107] N. Ahmidi, Y. Gao, B. Béjar, S. S. Vedula, S. Khudanpur, R. Vidal, and G. D. Hager, "String motif-based description of tool motion for detecting skill and gestures in robotic surgery," *Med Image Comput Comput Assist Interv*, vol. 16, no. Pt 1, pp. 26–33, 2013.

[108] S. Sefati, N. Cowan, and R. Vidal, "Learning Shared, Discriminative Dictionaries for Surgical Gesture Segmentation and Classification," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, vol. 4, 2015, pp. 1–10.

[109] F. Despinoy, D. Bouget, G. Forestier, C. Penet, N. Zemiti, P. Poignet, and P. Jannin, "Unsupervised Trajectory Segmentation for Surgical Gesture Recognition in Robotic Training," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 6, pp. 1280–1291, 2016.

[110] S. Krishnan, A. Garg, S. Patil, C. Lea, G. Hager, P. Abbeel, and K. Goldberg, "Transition State Clustering: Unsupervised Surgical Trajectory Segmentation for Robot Learning," in *Robotics Research: Volume 2*, ser. Springer Proceedings in Advanced Robotics, A. Bicchi and W. Burgard, Eds. Cham: Springer International Publishing, 2018, pp. 91–110.

[111] G. Forestier, F. Petitjean, P. Senin, F. Despinoy, A. Huaulmé, H. I. Fawaz, J. Weber, L. Idoumghar, P.-A. Muller, and P. Jannin, "Surgical motion analysis using discriminative interpretable patterns," *Artif Intell Med*, no. 91, pp. 3–11, 2018.

[112] B. B. Haro, L. Zappella, and R. Vidal, "Surgical gesture classification from video data," *Med Image Comput Comput Assist Interv*, vol. 15, no. Pt 1, pp. 34–41, 2012.

[113] H. C. Lin and G. Hager, "User-Independent Models of Manipulation Using Video Contextual Cues," *Workshop on Modeling and Monitoring of Computer Assisted Interventions*, pp. 1–8, 2009.

[114] L. Zappella, B. Béjar, G. Hager, and R. Vidal, "Surgical gesture classification from video and kinematic data," *Medical Image Analysis*, vol. 17, no. 7, pp. 732–745, 2013.

[115] A. Malpani, S. S. Vedula, C. C. G. Chen, and G. D. Hager, "Pairwise Comparison-Based Objective Score for Automated Skill Assessment of Segments in a Surgical Task," in *Information Processing in Computer-Assisted Interventions*, ser. Lecture Notes in Computer Science, D. Stoyanov, D. L. Collins, I. Sakuma, P. Abolmaesumi, and P. Jannin, Eds. Springer International Publishing, 2014, pp. 138–147.

[116] N. Ahmidi, L. Tao, S. Sefati, Y. Gao, C. Lea, B. B. Haro, L. Zappella, S. Khudanpur, R. Vidal, and G. D. Hager, "A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2025–2041, 2017.

[117] S. Jun, M. S. Narayanan, P. Agarwal, A. Eddib, P. Singhal, S. Garimella, and V. Krovi, "Robotic Minimally Invasive Surgical skill assessment based on automated video-analysis motion studies," in *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012, pp. 25–31.

[118] C. Lea, G. D. Hager, and R. Vidal, "An Improved Model for Segmentation and Recognition of Fine-Grained Activities with Application to Surgical Training Tasks," in *2015 IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 1123–1129.

[119] "Automated skill assessment for individualized training in robotic surgery-Science of Learning," http://scienceoflearning.jhu.edu/research/automated-skill-assessment-for-individualized-training-in-robotic-surgery, access date: 2023.10.27.

[120] A. Malpani, S. S. Vedula, C. C. G. Chen, and G. D. Hager, "A study of crowdsourced segment-level surgical skill assessment using pairwise rankings," *Int J CARS*, vol. 10, no. 9, pp. 1435–1447, 2015.

[121] S. Krishnan, A. Garg, S. Patil, C. Lea, G. D. Hager, P. Abbeel, and K. Goldberg, "Unsupervised Surgical Task Segmentation with Milestone Learning," in *Proc. Intl Symp. on Robotics Research (ISRR)*, 2015, pp. 1–16.

[122] C. Lea, A. Reiter, R. Vidal, and G. D. Hager, "Segmental Spatiotemporal CNNs for Fine-Grained Action Segmentation," in *Computer Vision – ECCV 2016*, ser. Lecture Notes in Computer Science, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Springer International Publishing, 2016, pp. 36–52.

[123] R. DiPietro, C. Lea, A. Malpani, N. Ahmidi, S. S. Vedula, G. I. Lee, M. R. Lee, and G. D. Hager, "Recognizing Surgical Activities with Recurrent Neural Networks," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, ser. Lecture Notes in Computer Science, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Springer International Publishing, 2016, pp. 551–558.

[124] S. S. Vedula, A. O. Malpani, L. Tao, G. Chen, Y. Gao, P. Poddar, N. Ahmidi, C. Paxton, R. Vidal, S. Khudanpur, G. D. Hager, and C. C. G. Chen, "Analysis of the Structure of Surgical Activity for a Suturing and Knot-Tying Task," *PLoS ONE*, vol. 11, no. 3, p. e0149174, 2016.

[125] A. Zia, C. Zhang, X. Xiong, and A. M. Jarc, "Temporal clustering of surgical activities in robot-assisted surgery," *Int J Comput Assist Radiol Surg*, vol. 12, no. 7, pp. 1171–1178, 2017.

[126] L. Zhou and M. Kaess, "An Efficient and Accurate Algorithm for the Perspecitve-n-Point Problem," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019, pp. 6245–6252.

[127] 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.

[128] M. J. Fard, S. Ameri, R. Darin Ellis, R. B. Chinnam, A. K. Pandya, and M. D. Klein, "Automated robot-assisted surgical skill evaluation: Predictive analytics approach," *Int J Med Robot*, vol. 14, no. 1, 2018.

[129] Y. Beck, T. Herman, M. Brozgol, N. Giladi, A. Mirelman, and J. M. Hausdorff, "SPARC: a new approach to quantifying gait smoothness in patients with parkinson's disease," *Journal of neuroengineering and rehabilitation*, vol. 15, no. 1, pp. 1–9, 2018.

[130] S. Singh, J. Bible, Z. Liu, Z. Zhang, and R. Singapogu, "Motion Smoothness Metrics for Cannulation Skill Assessment: What Factors Matter?" *Frontiers in Robotics and AI*, vol. 8, 2021.

[131] D. Liu, Q. Li, T. Jiang, Y. Wang, R. Miao, F. Shan, and Z. Li, "Towards Unified Surgical Skill Assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9522–9531.

[132] M. Petković, S. Džeroski, and D. Kocev, "Feature Ranking for Hierarchical Multi-Label Classification with Tree Ensemble Methods," *Acta Polytechnica Hungarica*, vol. 17, pp. 129–148, Jan. 2020.

[133] Y. Ming, Y. Cheng, Y. Jing, L. Liangzhe, Y. Pengcheng, Z. Guang, and C. Feng, "Surgical skills assessment from robot assisted surgery video data," in *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, Jan. 2021, pp. 392–396.

[134] G. Lajkó, R. N. Elek, and T. Haidegger, "Surgical Skill Assessment Automation Based on Sparse Optical Flow Data," in *2021 IEEE 25th International Conference on Intelligent Engineering Systems (INES)*, Jul. 2021, pp. 201–208.

[135] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[136] A. A. Shvets, A. Rakhlin, A. A. Kalinin, and V. I. Iglovikov, "Automatic Instrument Segmentation in Robot-Assisted Surgery using Deep Learning," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2018, pp. 624–628.

[137] V. Iglovikov, S. Mushinskiy, and V. Osin, "Satellite imagery feature detection using deep convolutional neural network: A kaggle competition," *arXiv preprint arXiv:1706.06169*, 2017.

[138] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 2017, pp. 1–4.

[139] V. Iglovikov and A. A. Shvets, "TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation," *arXiv: Computer Vision and Pattern Recognition*, pp. 1–5, Jan. 2018.

[140] Z. Zhao, Y. Jin, and P.-A. Heng, "TraSeTR: track-to-segment transformer with contrastive query for instance-level instrument segmentation in robotic surgery," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 11 186–11 193.

[141] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *3rd International Conference on Learning Representations, ICLR, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[142] M. Dobeš, R. Andoga, and L. Fozo, "Sensory integration in deep neural networks," *Acta Polytechnica Hungarica*, vol. 18, pp. 245–254, 2021.

[143] "EndoVisSub2017-RoboticInstrumentSegmentation – Grand Challenge," https://endovis.grand-challenge.org/, access date: 2023.10.27.

[144] M. W. Browne, "Cross-validation methods," *Journal of mathematical psychology*, vol. 44, no. 1, pp. 108–132, 2000.

[145] J. Tiferes, A. A. Hussein, A. Bisantz, J. D. Kozlowski, M. A. Sharif, N. M. Winder, N. Ahmad, J. Allers, L. Cavuoto, and K. A. Guru, "The Loud Surgeon Behind the Console: Understanding Team Activities During Robot-Assisted Surgery," *Journal of Surgical Education*, vol. 73, no. 3, pp. 504–512, 2016.

[146] C. Wu, J. Cha, J. Sulek, T. Zhou, C. P. Sundaram, J. Wachs, and D. Yu, "Eye-tracking metrics predict perceived workload in robotic surgical skills training," *Human factors*, vol. 62, no. 8, pp. 1365–1386, 2020.

[147] L. J. Moore, M. R. Wilson, J. S. McGrath, E. Waine, R. S. W. Masters, and S. J. Vine, "Surgeons' display reduced mental effort and workload while performing robotically assisted surgical tasks, when compared to conventional laparoscopy," *Surgical Endoscopy*, vol. 29, no. 9, pp. 2553–2560, Sep. 2015.

[148] A. A. Hussein, S. B. Shafiei, M. Sharif, E. Esfahani, B. Ahmad, J. D. Kozlowski, Z. Hashmi, and K. A. Guru, "Technical mentorship during robot-assisted surgery: A cognitive analysis," *BJU international*, vol. 118, no. 3, pp. 429–436, Sep. 2016.

[149] H. Singh, H. N. Modi, S. Ranjan, J. W. R. Dilley, D. Airantzis, G.-Z. Yang, A. Darzi, and D. R. Leff, "Robotic Surgery Improves Technical Performance and Enhances Prefrontal Activation During High Temporal Demand," *Annals of Biomedical Engineering*, vol. 46, no. 10, pp. 1621–1636, Oct. 2018.

[150] K. A. Guru, S. B. Shafiei, A. Khan, A. A. Hussein, M. Sharif, and E. T. Esfahani, "Understanding Cognitive Performance During Robot-Assisted Surgery," *Urology*, vol. 86, no. 4, pp. 751–757, Oct. 2015.

[151] J. Nguyen, J. Chen, S. Marshall, S. Ghodoussipour, A. Chen, I. Gill, and A. Hung, "Using objective robotic automated performance metrics and task-evoked pupillary response to distinguish surgeon expertise," *World Journal of Urology*, vol. 38, no. 7, pp. 1599–1605, 2020.

[152] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Human Mental Workload*, ser. Advances in Psychology, 52. Oxford, England: North-Holland, 1988, pp. 139–183.

[153] C. M. Carswell, D. Clarke, and W. B. Seales, "Assessing mental workload during laparoscopic surgery," *Surgical Innovation*, vol. 12, no. 1, pp. 80–90, Mar. 2005.

[154] R. K. Wadhera, S. H. Parker, H. M. Burkhart, K. L. Greason, J. R. Neal, K. M. Levenick, D. A. Wiegmann, and T. M. Sundt, "Is the "sterile cockpit" concept applicable to cardiovascular surgery critical intervals or critical events? The impact of protocol-driven communication during cardiopulmonary bypass," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 139, no. 2, pp. 312–319, Feb. 2010.

[155] M. I. Klein, M. A. Riley, J. S. Warm, and G. Matthews, "Perceived Mental Workload in an Endocopic Surgery Simulator," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 11, pp. 1014–1018, Sep. 2005.

[156] J. F. Ruiz-Rabelo, E. Navarro-Rodriguez, L. L. Di-Stasi, N. Diaz-Jimenez, J. Cabrera-Bermon, C. Diaz-Iglesias, M. Gomez-Alvarez, and J. Briceño-Delgado, "Validation of the NASA-TLX Score in Ongoing Assessment of Mental Workload During a Laparoscopic Learning Curve in Bariatric Surgery," *Obesity Surgery*, vol. 25, no. 12, pp. 2451–2456, Dec. 2015.

[157] B. Zheng, X. Jiang, G. Tien, A. Meneghetti, O. N. M. Panton, and M. S. Atkins, "Workload assessment of surgeons: Correlation between NASA TLX and blinks," *Surgical Endoscopy*, vol. 26, no. 10, pp. 2746–2750, Oct. 2012.

[158] A. Chowriappa, S. J. Raza, A. Fazili, E. Field, C. Malito, D. Samarasekera, Y. Shi, K. Ahmed, G. Wilding, J. Kaouk, D. D. Eun, A. Ghazi, J. O. Peabody, T. Kesavadas, J. L. Mohler, and K. A. Guru, "Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: A multi-institutional randomised controlled trial," *BJU international*, vol. 115, no. 2, pp. 336–345, Feb. 2015.

[159] G. I. Lee, M. R. Lee, T. Clanton, T. Clanton, E. Sutton, A. E. Park, and M. R. Marohn, "Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries," *Surgical Endoscopy*, vol. 28, no. 2, pp. 456–465, Feb. 2014.

[160] L. Panait, S. Shetty, P. A. Shewokis, and J. A. Sanchez, "Do laparoscopic skills transfer to robotic surgery?" *The Journal of Surgical Research*, vol. 187, no. 1, pp. 53–58, Mar. 2014.

[161] V. Mouraviev, M. Klein, E. Schommer, D. D. Thiel, S. Samavedi, A. Kumar, R. J. Leveillee, R. Thomas, J. M. Pow-Sang, L.-M. Su, E. Mui, R. Smith, and V. Patel, "Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules," *Journal of Robotic Surgery*, vol. 10, no. 1, pp. 49–56, Mar. 2016.

[162] E. Lau, N. A. Alkhamesi, and C. M. Schlachta, "Impact of robotic assistance on mental workload and cognitive performance of surgical trainees performing a complex minimally invasive suturing task," *Surgical endoscopy*, vol. 34, no. 6, pp. 2551–2559, 2020.

[163] K. Kawamura, Y. Kobayashi, and M. G. Fujie, "Development of Real-Time Simulation for Workload Quantization in Robotic Tele-surgery," in *2006 IEEE International Conference on Robotics and Biomimetics*, Dec. 2006, pp. 1420–1425.

[164] L. A. Cavuoto, A. A. Hussein, V. Vasan, Y. Ahmed, A. Durrani, S. Khan, A. Cole, D. Wang, J. Kozlowski, B. Ahmad, and K. A. Guru, "Improving Teamwork: Evaluating Workload of Surgical Team During Robot-assisted Surgery," *Urology*, vol. 107, pp. 120–125, Sep. 2017.

[165] M. I. Klein, C. H. Lio, R. Grant, C. M. Carswell, and S. Strup, "A Mental Workload Study on the 2d and 3d Viewing Conditions of the da Vinci Surgical Robot," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 53, no. 18, pp. 1186–1190, Oct. 2009.

[166] K. Sexton, A. Johnson, A. Gotsch, A. A. Hussein, L. Cavuoto, and K. A. Guru, "Anticipation, teamwork and cognitive load: Chasing efficiency during robot-assisted surgery," *BMJ Quality & Safety*, vol. 27, no. 2, pp. 148–154, Feb. 2018.

[167] G. I. Lee and M. R. Lee, "Can a virtual reality surgical simulation training provide a self-driven and mentor-free skills learning? Investigation of the practical influence of the performance metrics from the virtual reality robotic surgery simulator on the skill learning and associated cognitive workloads," *Surgical Endoscopy*, vol. 32, no. 1, pp. 62–72, Jan. 2018.

[168] S. Xu, M. Perez, C. Perrenot, N. Hubert, and J. Hubert, "Face, content, construct, and concurrent validity of a novel robotic surgery patient-side simulator: The Xperience™ Team Trainer," *Surgical Endoscopy*, vol. 30, no. 8, pp. 3334–3344, Aug. 2016.

[169] A. Hughes-Hallett, E. K. Mayer, H. J. Marcus, P. Pratt, S. Mason, A. W. Darzi, and J. A. Vale, "Inattention blindness in surgery," *Surgical Endoscopy*, vol. 29, no. 11, pp. 3184–3189, Nov. 2015.

[170] A. S. Sethi, W. J. Peine, Y. Mohammadi, and C. P. Sundaram, "Validation of a novel virtual reality robotic simulator," *Journal of Endourology*, vol. 23, no. 3, pp. 503–508, Mar. 2009.

[171] A. Y. B. Teoh, S. M. Chan, H. C. Yip, V. W. Y. Wong, P. W. Y. Chiu, and E. K. W. Ng, "Randomized controlled trial of EndoWrist-enabled robotic versus human laparoendoscopic single-site access surgery (LESS) in the porcine model," *Surgical Endoscopy*, vol. 32, no. 3, pp. 1273–1279, Mar. 2018.

[172] K. Kawamura, Y. Kobayashi, and M. G. Fujie, "Operability evaluation using an simulation system for gripping motion in robotic tele-surgery," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2009, pp. 5106–5109, 2009.

[173] W. Gerull, A. Zihni, and M. Awad, "Operative performance outcomes of a simulator-based robotic surgical skills curriculum," *Surgical endoscopy*, vol. 34, no. 10, pp. 4543–4548, 2020.

[174] S. B. Shafiei, A. A. Hussein, S. F. Muldoon, and K. A. Guru, "Functional Brain States Measure Mentor-Trainee Trust during Robot-Assisted Surgery," *Scientific Reports*, vol. 8, no. 1, p. 3667, Feb. 2018.

[175] G. Borghini, P. Arico, G. Di Flumeri, A. Colosimo, S. F. Storti, G. Menegaz, P. Fiorini, and F. Babiloni, "Neurophysiological measures for users' training objective assessment during simulated robot-assisted laparoscopic surgery," *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2016, pp. 981–984, Aug. 2016.

[176] B. Marçon, W. Sime, F. Guillemin, N. Hubert, F. Lagrange, C. Huselstein, and J. Hubert, "An ergonomic assessment of four different donor nephrectomy approaches for the surgeons and their assistants," *Research and Reports in Urology*, vol. 11, pp. 261–268, 2019.

[177] T. Zhou, J. Cha, G. Gonzalez, J. Wachs, C. Sundaram, and D. Yu, "Joint Surgeon Attributes Estimation in Robot-Assisted Surgery," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 285–286.

[178] K. Kawamura, Y. Kobayashi, and M. G. Fujie, "Pilot study on verification of effectiveness on operability of assistance system for robotic tele-surgery using simulation," *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2010, pp. 2308–2312, 2010.

[179] G. Dulan, R. V. Rege, D. C. Hogg, K. M. Gilberg-Fisher, N. A. Arain, S. T. Tesfay, and D. J. Scott, "Proficiency-based training for robotic surgery: Construct validity, workload, and expert levels for nine inanimate exercises," *Surgical Endoscopy*, vol. 26, no. 6, pp. 1516–1521, Jun. 2012.

[180] D. Turiani Hourneaux de Moura, H. Aihara, P. Jirapinyo, G. Farias, K. Hathorn, A. Bazarbashi, A. Sachdev, and C. Thompson, "Robot-assisted endoscopic submucosal dissection versus conventional ESD for colorectal lesions: Outcomes of a randomized pilot study in endoscopists without prior ESD experience (with video)," *Gastrointestinal Endoscopy*, vol. 90, no. 2, pp. 290–298, 2019.

[181] K. E. Law, B. R. Lowndes, S. R. Kelley, R. C. Blocker, D. W. Larson, M. S. Hallbeck, and H. Nelson, "NASA-Task Load Index Differentiates Surgical Approach: Opportunities for Improvement in Colon and Rectal Surgery," *Annals of Surgery*, vol. 271, no. 5, pp. 906–912, May 2020.

[182] D. B. Boles and L. P. Adair, "The multiple resources questionnaire (MRQ)," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 45, no. 25. SAGE Publications Sage CA: Los Angeles, CA, 2001, pp. 1790–1794.

[183] M. I. Klein, J. S. Warm, M. A. Riley, G. Matthews, C. Doarn, J. F. Donovan, and K. Gaitonde, "Mental workload and stress perceived by novice operators in the laparoscopic and robotic minimally invasive surgical interfaces," *Journal of Endourology*, vol. 26, no. 8, pp. 1089–1094, Aug. 2012.

[184] G. Matthews, "Multidimensional Profiling of Task Stress States for Human Factors: A Brief Review," *Human Factors*, vol. 58, no. 6, pp. 801–813, Sep. 2016.

[185] A. Ghanbary Sartang, M. Ashnagar, E. Habibi, and S. Sadeghi, "Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses," *Journal of Occupational Health and Epidemiology*, vol. 5, no. 4, pp. 211–217, Oct. 2016.

[186] L. Schiff, Z. Tsafrir, J. Aoun, A. Taylor, E. Theoharis, and D. Eisenstein, "Quality of Communication in Robotic Surgery and Surgical Outcomes," *JSLS : Journal of the Society of Laparoendoscopic Surgeons*, vol. 20, no. 3, 2016.

[187] J. B. Sexton, R. L. Helmreich, T. B. Neilands, K. Rowan, K. Vella, J. Boyden, P. R. Roberts, and E. J. Thomas, "The Safety Attitudes Questionnaire: Psychometric properties, benchmarking data, and emerging research," *BMC Health Services Research*, vol. 6, no. 1, p. 44, Apr. 2006.

[188] A. Kohli and M. Kaur, "Wisconsin Card Sorting Test: Normative data and experience," *Indian Journal of Psychiatry*, vol. 48, no. 3, pp. 181–184, 2006.

[189] C. Li, Q. Liu, T. Hu, and X. Jin, "Adapting the short form of the Coping Inventory for Stressful Situations into Chinese," *Neuropsychiatric Disease and Treatment*, vol. 13, pp. 1669–1675, Jun. 2017.

[190] R. H. van der Schatte Olivier, C. D. P. Van't Hullenaar, J. P. Ruurda, and I. a. M. J. Broeders, "Ergonomics, user comfort, and performance in standard and robot-assisted laparoscopic surgery," *Surgical Endoscopy*, vol. 23, no. 6, pp. 1365–1371, Jun. 2009.

[191] W. S. Helton, "Validation of a short stress state questionnaire," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 11. Sage Publications Sage CA: Los Angeles, CA, 2004, pp. 1238–1242.

[192] N. Malhotra, J. M. Poolton, M. R. Wilson, J. K. M. Fan, and R. S. W. Masters, "Conscious motor processing and movement self-consciousness: Two dimensions of personality that influence laparoscopic training," *Journal of Surgical Education*, vol. 71, no. 6, pp. 798–804, 2014 Nov-Dec.

[193] A. M. Abdelrahman, J. Bingener, D. Yu, B. R. Lowndes, A. Mohamed, A. L. McConico, and M. S. Hallbeck, "Impact of single-incision laparoscopic cholecystectomy (SILC) versus conventional laparoscopic cholecystectomy (CLC) procedures on surgeon stress and workload: A randomized controlled trial," *Surgical Endoscopy*, vol. 30, no. 3, pp. 1205–1211, Mar. 2016.

[194] M. Weigl, P. Stefan, K. Abhari, P. Wucherer, P. Fallavollita, M. Lazarovici, S. Weidert, E. Euler, and K. Catchpole, "Intra-operative disruptions, surgeon's mental workload, and technical performance in a full-scale simulated procedure," *Surgical Endoscopy*, vol. 30, no. 2, pp. 559–566, Feb. 2016.

[195] R. J. Berg, K. Inaba, M. Sullivan, O. Okoye, S. Siboni, M. Minneti, P. G. Teixeira, and D. Demetriades, "The impact of heat stress on operative performance and cognitive function during simulated laparoscopic operative tasks," *Surgery*, vol. 157, no. 1, pp. 87–95, Jan. 2015.

[196] M. Weigl, S. Antoniadis, C. Chiapponi, C. Bruns, and N. Sevdalis, "The impact of intra-operative interruptions on surgeons' perceived workload: An observational study in elective general and orthopedic surgery," *Surgical Endoscopy*, vol. 29, no. 1, pp. 145–153, Jan. 2015.

[197] C. Walters and P. J. Webb, "Maximizing Efficiency and Reducing Robotic Surgery Costs Using the NASA Task Load Index," *AORN journal*, vol. 106, no. 4, pp. 283–294, Oct. 2017.

[198] D. Yu, C. Dural, M. M. B. Morrow, L. Yang, J. W. Collins, S. Hallbeck, M. Kjellman, and M. Forsman, "Intraoperative workload in robotic surgery assessed by wearable motion tracking sensors and questionnaires," *Surgical Endoscopy*, vol. 31, no. 2, pp. 877–886, Feb. 2017.

[199] K. A. Guru, E. T. Esfahani, S. J. Raza, R. Bhat, K. Wang, Y. Hammond, G. Wilding, J. O. Peabody, and A. J. Chowriappa, "Cognitive skills assessment during robot-assisted surgery: Separating the wheat from the chaff," *BJU International*, vol. 115, no. 1, pp. 166–174, 2015.

[200] A. Beulens, W. Brinkman, E. Koldewijn, A. Hendrikx, J. van Basten, J. van Merriënboer, H. Van der Poel, C. Bangma, and C. Wagner, "A Prospective, Observational, Multicentre Study Concerning Nontechnical Skills in Robot-assisted

Radical Cystectomy Versus Open Radical Cystectomy," *European Urology Open Science*, vol. 19, pp. 37–44, 2020.

[201] S. Thangaratinam and C. W. Redman, "The Delphi technique," *The Obstetrician & Gynaecologist*, vol. 7, no. 2, pp. 120–125, 2005.

[202] K. Catchpole, C. Perkins, C. Bresee, M. J. Solnik, B. Sherman, J. Fritch, B. Gross, S. Jagannathan, N. Hakami-Majd, R. Avenido, and J. T. Anger, "Safety, efficiency and learning curves in robotic surgery: A human factors analysis," *Surgical Endoscopy*, vol. 30, no. 9, pp. 3749–3761, Sep. 2016.

[203] F. Zattoni, A. Guttilla, A. Crestani, A. De Gobbi, F. Cattaneo, M. Moschini, F. Vianello, C. Valotto, F. Dal Moro, and F. Zattoni, "The Value of Open Conversion Simulations During Robot-Assisted Radical Prostatectomy: Implications for Robotic Training Curricula," *Journal of Endourology*, vol. 29, no. 11, pp. 1282–1288, Nov. 2015.

[204] N. Sevdalis, R. Davis, M. Koutantji, S. Undre, A. Darzi, and C. A. Vincent, "Reliability of a revised NOTECHS scale for use in surgical teams," *American Journal of Surgery*, vol. 196, no. 2, pp. 184–190, Aug. 2008.

[205] S. Yule, R. Flin, S. Paterson-Brown, and N. Maran, "Non-technical skills for surgeons in the operating room: A review of the literature," *Surgery*, vol. 139, no. 2, pp. 140–149, Feb. 2006.

[206] L. Hull, S. Arora, E. Kassab, R. Kneebone, and N. Sevdalis, "Observational teamwork assessment for surgery: Content validation and tool refinement," *Journal of the American College of Surgeons*, vol. 212, no. 2, pp. 234–243, Feb. 2011.

[207] J. W. Ahn, Y. Ku, and H. C. Kim, "A Novel Wearable EEG and ECG Recording System for Stress Assessment," *Sensors (Basel, Switzerland)*, vol. 19, no. 9, pp. 1–14, Apr. 2019.

[208] K. A. Herborn, J. L. Graves, P. Jerem, N. P. Evans, R. Nager, D. J. McCafferty, and D. E. McKeegan, "Skin temperature reveals the intensity of acute stress," *Physiol Behav*, vol. 152, no. Pt A, pp. 225–230, 2015.

[209] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddharaju, R. Joseph, A. Mandapati, B. Dunkin, and B. Bass, "Fast by Nature - How Stress Patterns Define Human Experience and Performance in Dexterous Tasks," *Scientific Reports*, vol. 2, pp. 1–9, 2012.

[210] "How the temperature of your nose shows how much strain you are under – The University of Nottingham," https://www.nottingham.ac.uk/news/pressreleases/2018/january/how-the-temperature-of-your-nose-shows-how-much-strain-you-are-under.aspx, access date: 2023.10.27.

[211] C. L. etitia Lisetti and F. Nasoz, "Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals," *EURASIP J. Appl. Signal Process.*, vol. 2004, pp. 1672–1687, 2004.

[212] G. G. Youngson, "Nontechnical skills in pediatric surgery: Factors influencing operative performance," *Journal of Pediatric Surgery*, vol. 51, no. 2, pp. 226–230, 2016.

[213] L. L. Di Stasi, M. B. McCamy, S. L. Macknik, J. A. Mankin, N. Hooft, A. Catena, and S. Martinez-Conde, "Saccadic eye movement metrics reflect surgical residents' fatigue," *Annals of Surgery*, vol. 259, no. 4, pp. 824–829, Apr. 2014.

[214] B. Law, M. S. Atkins, A. E. Kirkpatrick, and A. J. Lomax, "Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment," in *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, ser. ETRA '04. San Antonio, Texas: Association for Computing Machinery, Mar. 2004, pp. 41–48.

[215] M. Shahbazi, B. Poursartip, K. Siroen, C. M. Schlachta, and R. V. Patel, "Robotics-Assisted Surgical Skills Evaluation based on Electrocortical Activity," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2018, pp. 3673–3676, Jul. 2018.

[216] J. C. Sigl and N. G. Chamoun, "An introduction to bispectral analysis for the electroencephalogram," *Journal of Clinical Monitoring*, vol. 10, no. 6, pp. 392–404, Nov. 1994.

[217] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, p. 258, 2017.

[218] G. P. Mylonas, K.-W. Kwok, D. R. C. James, D. Leff, F. Orihuela-Espina, A. Darzi, and G.-Z. Yang, "Gaze-Contingent Motor Channelling, haptic constraints and associated cognitive demand for robotic MIS," *Medical Image Analysis*, vol. 16, no. 3, pp. 612–631, Apr. 2012.

[219] D. R. C. James, F. Orihuela-Espina, D. R. Leff, G. P. Mylonas, K.-W. Kwok, A. W. Darzi, and G.-Z. Yang, "Cognitive burden estimation for visuomotor learning with fNIRS," *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 13, no. Pt 3, pp. 319–326, 2010.

[220] S. Bunce, M. Izzetoglu, K. Izzetoglu, B. Onaral, and K. Pourrezaei, "Functional near-infrared spectroscopy," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 4, pp. 54–62, Jul. 2006.

[221] S. Venclove, A. Daktariunas, and O. Ruksenas, "Functional near-infrared spectroscopy: A continuous wave type based system for human frontal lobe studies," *EXCLI Journal*, vol. 14, pp. 1145–1152, Oct. 2015.

[222] T. Wilcox and M. Biondi, "fNIRS in the developmental sciences," *Wiley interdisciplinary reviews. Cognitive science*, vol. 6, no. 3, pp. 263–283, 2015.

[223] M. N. Frank and W. B. Kinlaw, "Indirect measurement of isovolumetric contraction time and tension period in normal subjects," *The American Journal of Cardiology*, vol. 10, pp. 800–806, Dec. 1962.

[224] R. Nagyné Elek and T. Haidegger, "Robot-Assisted Minimally Invasive Surgical Skill Assessment—Manual and Automated Platforms," *Acta Polytechnica Hungarica*, vol. 16, no. 8, pp. 141–169, 2019.

[225] "How will sensor technology impact the medical robots of the future? — MediSens Conference," https://medisens-conference.com/2018/01/22/how-sensor-technologies-will-transform-the-surgical-robots-of-the-future/ Access date: 2023.10.27.

[226] G. Fu, E. Azimi, and P. Kazanzides, "Mobile teleoperation: Feasibility of wireless wearable sensing of the operator's arm motion," pp. 4238–4243, 2021.

[227] S. Abeywardena, Q. Yuan, A. Tzemanaki, E. Psomopoulou, L. Droukas, C. Melhuish, and S. Dogramadzi, "Estimation of Tool-Tissue Forces in Robot-Assisted Minimally Invasive Surgery Using Neural Networks," *Frontiers in Robotics and AI*, vol. 6, 2019.

[228] M. T. Thai, P. T. Phan, T. T. Hoang, S. Wong, N. H. Lovell, and T. N. Do, "Advanced Intelligent Systems for Surgical Robotics," *Advanced Intelligent Systems*, vol. 2, no. 8, p. 1900138, 2020.

[229] A. Mishra, K. Catchpole, T. Dale, and P. McCulloch, "The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy," *Surgical Endoscopy*, vol. 22, no. 1, pp. 68–73, Jan. 2008.

[230] I. Funke, S. T. Mees, J. Weitz, and S. Speidel, "Video-based surgical skill assessment using 3D convolutional neural networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 7, pp. 1217–1225, Jul. 2019.

[231] A. Lukezic, T. Vojir, L. ˇCehovin Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6309–6318.

[232] K. R. Hassler, J. T. Collins, K. Philip, and M. W. Jones, "Laparoscopic Cholecystectomy," in *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2022.

[233] "Laparoscopyboxx Laparoscopic training boxes & instruments," https://laparoscopyboxx.com/, available: 10.05.2022.

[234] P. Sprent and N. C. Smeeton, *Applied nonparametric statistical methods*. CRC press, 2016.

[235] S. Hutchinson, G. Hager, and P. Corke, "A tutorial on visual servo control," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, Oct. 1996.

[236] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, Jun. 2014.

[237] Quigley, Morgan, "ROS: an open-source Robot Operating System - Google Scholar," *ICRA workshop on open source software*, vol. 3, no. 3.2, p. 5, 2009.

[238] N. X. Anh, R. M. Nataraja, and S. Chauhan, "Towards near real-time assessment of surgical skills: A comparison of feature extraction techniques," *Computer methods and programs in biomedicine*, vol. 187, p. 105234, 2020.

[239] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018.

[240] N. Yusupova, G. Shakhmametova, and R. Zulkarneev, "Complex analysis of medical data with data mining usage," *Acta Polytechnica Hungarica*, vol. 17, pp. 75–93, 2020.

[241] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[242] F. Li, K. Shirahama, M. Nisar, L. Köping, and M. Grzegorzek, "Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors," *Sensors*, vol. 18, no. 3, p. 679, Feb. 2018.

[243] R. P. Monti, S. Tootoonian, and R. Cao, "Avoiding Degradation in Deep Feed-Forward Networks by Phasing Out Skip-Connections," in *Artificial Neural Networks and Machine Learning – ICANN 2018*, V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis, Eds. Cham: Springer International Publishing, 2018, vol. 11141, pp. 447–456, series Title: Lecture Notes in Computer Science.

[244] Q. V. Le *et al.*, "A tutorial on deep learning part 2: Autoencoders, convolutional neural networks and recurrent neural networks," *Google Brain*, vol. 20, pp. 1–20, 2015.

[245] A. Zia, Y. Sharma, V. Bettadapura, E. L. Sarin, T. Ploetz, M. A. Clements, and I. Essa, "Automated video-based assessment of surgical skills for training and evaluation in medical schools," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 9, pp. 1623–1636, Sep. 2016.

[246] E. Yanik, X. Intes, U. Kruger, P. Yan, D. Diller, B. Van Voorst, B. Makled, J. Norfleet, and S. De, "Deep neural networks for the assessment of surgical skills: A systematic review," *The Journal of Defense Modeling and Simulation*, vol. 19, no. 2, pp. 159–171, 2022.

[247] I. Funke, S. T. Mees, J. Weitz, and S. Speidel, "Video-based surgical skill assessment using 3D convolutional neural networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 7, pp. 1217–1225, Jul. 2019.

[248] A. K. Lefor, K. Harada, A. Dosis, and M. Mitsuishi, "Motion analysis of the JHU-ISI Gesture and Skill Assessment Working Set using Robotics Video and Motion Assessment Software," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 12, pp. 2017–2025, Dec. 2020. [Online]. Available:

[249] T. Kanamori, S. Fujiwara, and A. Takeda, "Breakdown Point of Robust Support Vector Machines," *Entropy*, vol. 19, no. 2, p. 83, Feb. 2017.

[250] P. M. Kebria, S. Al-wais, H. Abdi, and S. Nahavandi, "Kinematic and dynamic modelling of UR5 manipulator," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2016, pp. 4229–4234.

[251] P. Mukhopadhyay and B. B. Chaudhuri, “A survey of Hough Transform,” *Pattern Recognition*, vol. 48, no. 3, pp. 993–1010, Mar. 2015.

PUBLICATIONS RELATED TO THE THESIS

[RNE1] R. Nagyné Elek and T. Haidegger, "Robot-assisted minimally invasive surgical skill assessment—manual and automated platforms," *Acta Polytechnica Hungarica*, vol. 16, no. 8, pp. 141–169, 2019.

[RNE2] D. El-Saig, R. Nagyné Elek, and T. Haidegger, "A Graphical Tool for Parsing and Inspecting Surgical Robotic Datasets," in *2018 IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2018, pp. 131–136.

[RNE3] R. Nagyné Elek and T. Haidegger, "Endoszkópos kameraképek feldolgozásának lehetőségei a robotsebészetben," *IME: Informatika és Menedzsment az Egészségügyben*, vol. 17, no. 4, pp. 49–54, 2018.

[RNE4] D. Papp, R. Nagyné Elek, and T. Haidegger, "Surgical tool segmentation on the jigsaws dataset for autonomous image-based skill assessment," *IEEE 10th Jubilee International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC)*, pp. 49–56, 2022.

[RNE5] R. Nagyné Elek and T. Haidegger, "Towards autonomous endoscopic image-based surgical skill assessment: Articulated tool pose estimation," *IEEE 10th Jubilee International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC)*, pp. 35–42, 2022.

[RNE6] R. Nagyné Elek and T. Haidegger, "Non-technical skill assessment and mental load evaluation in robot-assisted minimally invasive surgery," *Sensors*, vol. 21, no. 8, p. 2666, 2021.

[RNE7] R. Nagyné Elek, B. Mach, K. Móga, A. Ládi, and T. Haidegger, "Autonomous non-technical surgical skill assessment and workload analysis in laparoscopic cholecystectomy training," *2021 IEEE 25th International Conference on Intelligent Engineering Systems (INES)*, pp. 33–39, 2022.

[RNE8] R. Nagyné Elek and T. Haidegger, "Future Trend: AI x Robot," *Artificial Intelligence in Surgery – Recent Advances and Future (Springer Nature), editor: Ken Masamune*, p. 22 (in press), 2022.

[RNE9] R. Nagyné Elek and T. Haidegger, "Next in surgical data science: Autonomous non-technical skill assessment in minimally invasive surgery training," *Journal of Clinical Medicine*, vol. 11, no. 24, p. 7533, 2022.

[RNE10] R. Elek, T. Dániel Nagy, D. Á. Nagy, G. Kronreif, I. J. Rudas, and T. Haidegger, "Recent trends in automating robotic surgery," in *2016 IEEE 20th Jubilee International Conference on Intelligent Engineering Systems (INES)*, 2016, pp. 27–32.

[RNE11] C. Molnár, T. D. Nagy, R. Nagyné Elek, and T. Haidegger, "Visual servoing-based camera control for the da Vinci Surgical System," in *2020 IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY)*, 2020, pp. 107–112.

[RNE12] C. Molnár, R. Nagyné Elek, T. D. Nagy, R. Kiss, and T. Haidegger, "Vizuális kameravezérlés da Vinci sebészeti robotrendszerre: Visual servoing for da Vinci surgical system," *Nemzetközi Gépészeti Konferencia – OGÉT*, pp. 329–332, 2020.

[RNE13] G. Lajkó, R. Nagyné Elek, and T. Haidegger, "Surgical Skill Assessment Automation Based on Sparse Optical Flow Data," in *2021 IEEE 25th International Conference on Intelligent Engineering Systems (INES)*, 2021, pp. 201–208.

[RNE14] G. Lajkó, R. Nagyné Elek, and T. Haidegger, "Endoscopic Image-Based Skill Assessment in Robot-Assisted Minimally Invasive Surgery," *Sensors*, vol. 21, no. 16, p. 5412, 2021.

[RNE15] R. Elek, A. Károly, T. Haidegger, and P. Galambos, "Towards Optical Flow Ego-motion Compensation for Moving Object Segmentation," in *International Conference on Robotics, Computer Vision and Intelligent Systems*, 2022, pp. 114–120.

FURTHER PUBLICATIONS

[RNENR1] R. Elek, T. D. Nagy, D. Á. Nagy, T. Garamvölgyi, B. Takács, P. Galambos, J. K. Tar, I. J. Rudas, and T. Haidegger, "Towards surgical subtask automation — Blunt dissection," in *2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES)*, 2017, pp. 253–258.

[RNENR2] R. Elek, T. D. Nagy, D. Á. Nagy, B. Takács, P. Galambos, I. Rudas, and T. Haidegger, "Robotic platforms for ultrasound diagnostics and treatment," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 1752–1757.

[RNENR3] A. I. Károly, R. N. Elek, T. Haidegger, K. Széll, and P. Galambos, "Optical flow-based segmentation of moving objects for mobile robot navigation using pre-trained deep learning models*," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 3080–3086.

[RNENR4] A. I. Károly, R. Nagyné Elek, T. Haidegger, and P. Galambos, "Moving Obstacle Segmentation with an Optical Flow-based DNN: An Implementation Case Study," in *2021 IEEE 25th International Conference on Intelligent Engineering Systems (INES)*, 2021, pp. 189–194.

[RNENR5] D. Á. Nagy, T. D. Nagy, R. Elek, I. Rudas, and T. Haidegger, "Ontology Based Surgical Subtask Automation," in *International Conference on Robotics and Automation 2017*, Singapore, 2017.

[RNENR6] D. Á. Nagy, T. D. Nagy, R. Elek, I. J. Rudas, and T. Haidegger, "Ontology-Based Surgical Subtask Automation, Automating Blunt Dissection," *Journal of Medical Robotics Research*, vol. 03, no. 03n04, p. 1841005, 2018.

[RNENR7] K. Takács, R. N. Elek, and T. Haidegger, "Image processing-based methods to improve the robustness of robotic gripping," in *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo)*. IEEE, 2022, pp. 345–350.