ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

Ph.D. Thesis

TAMÁS DÁNIEL LEVENDOVICS
(formerly TAMÁS DÁNIEL NAGY)

Methodological Approach for Subtask Automation in Robot-Assisted Minimally Invasive Surgery

Részfeladatok Automatizálásának Módszertani Megközelítése a Robottal Támogatott Minimál Invazív Sebészetben

Supervisor:
Prof. Dr. Tamás Haidegger

DOCTORAL SCHOOL OF APPLIED INFORMATICS AND APPLIED MATHEMATICS

Budapest, 2025

Members of the Defense Committee:

Chair of the Committee:
Prof. Dr. László Pokorádi (Óbuda University)

Opponents:
Dr. Szilveszter Pletl (University of Szeged)
Dr. Ákos Odry (University of Dunaújváros)
Dr. Károly Széll (Óbuda University)

Secretary:
Dr. József Kuti (Óbuda University)

Members:
Dr. György Cserey (Pázmány Péter Catholic University)
Dr. Gábor Kertész (Óbuda University)
Dr. Ákos Szlávecz (Budapest University of Technology)

Replacement Members of the Defense Committee:

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

Secretary:
Dr. habil. Sándor Szénási (Óbuda University)

Opponent:
Dr. Adrienn Dineva (Óbuda University)

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

Members of the Complex 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: 2025

ABSTRACT

The technical revolution of Minimally Invasive Surgical (MIS) procedures had a significant influence on the manual practice, opening the way to laparoscopic surgery, then evolving into robotics surgery. Along with the benefits for patients, such as less trauma, lower risk of complications, or faster recovery, this "keyhole" technique also presents significant new challenges to the surgeons. The interventions often require weary body posture, the range of motion is limited for the tools and the surgical instruments are cumbersome to manage. Teleoperated master–slave robots, such as the da Vinci Surgical System, offer a general solution to these, enabling the surgeons to operate in a more ergonomic, seated position at a master console, while their hand motions are copied onto a robotic instrument inside the patient. Due to its positive aspects, Robot-Assisted Minimally Invasive Surgery (RAMIS) has become a standard of care in the past few decades, having performed over 2.5 million procedures per year worldwide. The next major step in the evolution of surgery could be the introduction of automation. Partial and sequential introduction and increase of autonomous capabilities could provide a safe way towards the concept of *Surgery 4.0*, which integrates advanced robotics, digital technologies, and surgical data science to improve precision, efficiency, and patient outcomes. The workflow of RAMIS procedures frequently contains monotonous and time-consuming elements; automation of such subtasks would ease the cognitive load on the surgeons, allowing them to pay more attention on the critical parts of the intervention. Unfortunately, autonomy in the given environment, consisting mostly of soft organs, suffers from grave difficulties. Unlike working on hard tissues, where exact registration of imaging data to the instrument and the robot is possible, soft tissues are are highly deformable, permanently in motion, thus no pre-computed tool trajectories can be used efficiently. Another challenge of surgical automation is undoubtedly the implementation of perception algorithms usable in the complex human environment. Computer vision suffers from reflections and features of different organs being highly similar, yet it is still the gold standard. Due to the complexity of the environment, the equipment, and the workflow, the field of surgical subtask automation is found to be quite challenging. Although serious research efforts have been invested to this area across the globe, the objective definition and assessment of autonomous functions, standard evaluation metrics, or benchmarking techniques have still not been formed. Another important question is to what extent the surgeons are able to work together with those autonomous functions, whether they are able to concentrate long enough while the subtask is performed autonomously in order to supervise the execution. In this thesis, the above-mentioned aspects of subtask automation in surgery are to be presented, introducing the recent advances in surgical robot motion planning, perception, and human–machine interaction, along with the limitations of task-level autonomy. A characterization model for surgical automation, and a method for performance evaluation and comparison of automated surgical subtasks will be also shown. Finally, the effect of automation of the surgeons' performance is studied, providing insights into the safe integration of autonomous functions into surgical practice.

KIVONAT

A Minimálisan Invazív Sebészet (MIS) elterjedése jelentős hatással volt a sebészeti gyakorlatra, és lehetővé tette a laparoszkópos, majd a robotsebészeti technikák alkalmazásának elterjedését. A betegek számára biztosított előnyök, mint például a kisebb szöveti trauma, vagy a gyorsabb felépülés ellenére a MIS jelentős kihívást is jelent a sebészek számára. A humán operatőr a beavatkozásokat gyakran fárasztó testhelyzetben kényszerül végezni; a MIS eszközök kezelése nehézkes. A teleoperációs mester–szolga típusú robotok, mint a da Vinci Sebészeti Rendszer, megoldást kínálnak ezekre a problémákra, lehetővé téve a sebész számára, hogy ergonomikusabb testhelyzetben, egy teleoperációs konzolnál dolgozzon, miközben kézmozdulatait a páciensen belül elhelyezkedő eszközök lekövetik. Előnyös tulajdonságainak köszönhetően a robottal támogatott minimál invazív sebészet (Robot-Assisted Minimally Invasive Surgery, RAMIS) az elmúlt néhány évtizedben standarddá vált a sebészeti gyakorlatban, évente több mint két és fél millió RAMIS beavatkozást hajtanak végre világszerte. A sebészet fejlődésében a következő nagy lépés az automatizálás bevezetése lehet. Az önvezető autózással analóg módon az autonóm készségek részleges és fokozatos bevezetése a sebészetben biztonságos utat jelenthet a *Sebészet 4.0* koncepció felé, amely fejlett robotikai és digitális technológiákat, valamint sebészeti adattudományt és adatalapú döntéstámogatást integrál a pontosság, hatékonyság növelésének érdekében. A RAMIS eljárások munkafolyamata gyakran tartalmaz monoton és időigényes elemeket; az ilyen részfeladatok automatizálása csökkentené a sebész kognitív terhelését, lehetővé téve, hogy nagyobb figyelmet fordíthasson a beavatkozás kritikus részeire. A többnyire lágyszövetekből álló műtéti környezetben az autonóm rendszerek fejlesztése komoly kihívást jelent. Ellentétben például a csontszöveteken végzett beavatkozásokkal, ahol lehetséges a pontos regisztráció az eszközök, a robot és az anatómia között, a lágyrészek folyamatosan mozgásban vannak, erősen deformálódnak, így nem igazán használhatóak előre generált robot trajektóriák. A komplex környezet, a használt eszközök és a munkafolyamat összetettsége miatt a sebészeti részfeladat-automatizálás területe meglehetősen nagy kihívást jelent. Bár világszerte komoly kutatási projektek folynak ezen a területen, az autonóm funkciók objektív karakterizálása és validációja, standard validációs metrikák vagy benchmarking technikák még nem alakultak ki. További fontos kérdés, hogy a sebészek hogyan tudnak együtt dolgozni ezekkel az autonóm rendszerekkel, például képesek-e a koncentrációt kellő ideig fenntartani a részfeladatok robot általi végrehajtása közben az autonóm rendszer felügyeletéhez. A disszertációban a sebészeti részfeladat-automatizálás fent említett szempontjait vizsgáltam. Sebészeti részfeladatok automatizálására kidolgozott, emberi mozgásmintákon alapuló módszertant, illetve azt támogató szoftver keretrendszert hoztam létre. Szintén megalkottam egy, az autonóm sebészeti rendszereket karakterizáló modellt, valamint az automatizált sebészeti részfeladatok teljesítményértékelésének, összehasonlításának és validációjának módszertanát. Végül vizsgáltam az automatizálás sebészekre gyakorolt hatását, és rámutattam az autonóm funkciók sebészeti gyakorlatba történő biztonságos integrálásának lehetőségeire.

DECLARATION

I, the undersigned, Tamás Dániel Levendovics, hereby declare that this Ph.D. thesis is my own original work. All sources used are listed in the references. All parts taken from other works, either as word for word citation or rewritten keeping the original meaning, have been clearly indicated and properly cited.

NYILATKOZAT

Alulírott Levendovics Tamás Dániel kijelentem, hogy ezt a doktori értekezést magam 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, Április 9, 2025

.....

Tamás Dániel Levendovics

Contents

Acknowledgment

This challenging endeavor would not have been possible without my mentor and supervisor, Prof. Dr. Tamás Haidegger. I would like to express my gratitude for his continuous guidance and support. His enthusiasm towards science and critical thinking had a great effect on me. Also, without him, and the international support team, the essential research platform of my work, the da Vinci Research Kit would not be available in the laboratory.

I would like to extend my sincere thanks to the Doctoral School of Applied Informatics and Applied Mathematics, especially to Prof. Dr. József Tar, Prof. Dr. Gyula Simon, Prof. Dr. Aurél Galántai, and Prof. Dr. László Horváth for all the support I received during the doctoral program. I would like to thank to Zsuzsanna Bácskai for helping me with the paperwork whenever I asked. I also could not have undertaken this journey without my defense committee, who generously provided knowledge and expertise.

I would like to express my appreciation to the colleagues of Antal Bejczy Center for Intelligent Robotics and Prof. Dr. Péter Galambos, the leader of the laboratory. I am also thankful to Dr. Árpád Takács and Dr. Dániel Drexler for advising me numerous times. I would like to extend my sincere thanks to Klára Haidegger, who kindly reviewed and corrected my work, providing invaluable assistance throughout the process. I would like to thank my students, especially Nikita Ukhrenkov for their help in this research work.

I am grateful to Dr. György Eigner, the dean of John von Neumann Faculty of Informatics, for the opportunity to start my teaching career at Óbuda University.

I would also like to thank to the colleagues of the Noise Research Group of University of Szeged for starting me on my scientific career.

I thankfully acknowledge the financial support from the New National Excellence Program of the Ministry for Innovation and Technology.

I would like to express my deepest gratitude to my loving wife for being my partner in this journey. She was with me during ups and downs, encouraging me always, despite that she was also working on her Ph.D. during these years. I am also grateful to my little daughter, who doesn't even know yet how much she has supported me throughout the process of earning this degree. I am also thankful to my friends and my family for their never-ending support.

Abbreviations and Notations

ADAS	Advanced Driver Assistance Systems
ALFUS	Autonomy Levels for Unmanned Systems
AI	Artificial Intelligence
AMBF	Asynchronous Multi-Body Framework
CAC	Contextual Autonomous Capability
DoA	Degree of Autonomy
DoF	Degrees of Freedom
DVRK	da Vinci Research Kit
ESS	European Society Of Surgery
FLS	Fundamentals of Laparoscopic Surgery
FRS	Fundamentals of Robotic Surgery
FDA	United States Food and Drug Administration
GDPR	General Data Protection Regulation
GEARS-E	Global Evaluative Assessment of Robotic Skills in Endoscopy
HISIC	Hazard Identification and Safety Insurance Control
HMI	Human–Machine Interface
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IROB	Antal Bejczy Center for Intelligent Robotics
ISO	International Organization for Standardization
IQR	Interquartile Range
JIGSAWS	JHU—ISI Gesture and Skill Assessment Working Set
LC	Laparoscopic Cholecystectomy
LLM	Large Language Model
LoA	Level of Autonomy
LoCR	Level of Clinical Realism
LoEC	Level of Environmental Complexity
LoTC	Level of Task Complexity

LSPB	Linear Segments with Parabolic Blends
MAUDE	Manufacturer and User Facility Device Experience
MDR	Medical Devices Regulation
MEE	Medical Electrical Equipment
MES	Medical Electrical System
MIS	Minimally Invasive Surgery
MTM	Master Tool Manipulators
NASA-TLX	NASA Task Load Index
QLA	Quad Linear Amplifier
RAMIS	Robot-Assisted Minimally Invasive Surgery
RASE	Robotically Assisted Surgical Equipment
RMSE	Root-Mean-Square Error
ROI	Region of Interest
ROS	Robot Operating System
RPC	Remote Procedure Calls
R-OSATS	Robotic Objective Structured Assessments of Technical Skills
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SAGE	Strategic Advisory Group of Experts
SART	Situational Awareness Rating Technique
SCAC	Surgical Contextual Autonomous Capability
SLERP	Spherical Linear Interpolation
SNAP	Suture Needle Angular Positioner
SD	Standard Deviation
STAR	Smart Tissue Autonomous Robot
TC	Technical Committee
TCP	Tool Center Point
TR	Technical Report
UMS	Unmanned Systems
WHO	World Health Organization

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Robot-Assisted Minimally Invasive Surgery

Minimally Invasive Surgery (MIS) induced a revolution in the surgical practice over the last three decades. Contrary to the traditional manual technique operating through large incisions, MIS is performed through few-centimeter-wide ports—incisions like keyholes—using laparoscopic instruments, the area of operation is observed via an endoscopic camera. Smaller incisions offer benefits both for the patient and the hospital, such as lower risk of complications, faster recovery, and thus shorter hospital stay. On the other hand, MIS presents new challenges to the surgeons, such as the limited range of motion with less Degrees of Freedom (DoF) and also fatigue from weary body postures.

Robot-Assisted Minimally Invasive Surgery (RAMIS) was introduced to ease these difficulties. The idea of teleoperated master–slave RAMIS systems originates from space research: the intervention was to be performed on the patient—in this case an astronaut—by a teleoperated device, controlled by a human surgeon through a master device on Earth [1, 2]. The slave-side robot arms are equipped with laparoscopic instruments and an endoscopic camera, and copy the movement of the surgeon at the remote site, while at the master console, the surgeon observes the operation via the endoscopic camera stream.

However, real remote teleoperation has not become a daily practice, and stalled at the state of research, mainly due to the issues caused by time delay. It has been demonstrated that teleoperation itself can present a number of benefits. The communication latency can be reduced to a level that is insignificant for the surgeon by placing the master and the slave devices close to each other; in the case of commercial RAMIS systems, the master- and the slave-side devices are in the same room. This technology can still reduce the fatigue of the surgeon, being able to operate in a more ergonomic, seated position. Furthermore, the movement of the surgeon can be scaled on the slave side—the most delicate movements can be controlled by relatively large hand movements, and hand tremors can also be filtered. Moreover, the advanced endoscopic system and robotic surgical instruments of RAMIS enable the execution of more advanced workflows, potentially improving the outcome of the interventions. One, and the best-known of those is the nerve-sparing technique in radical prostatectomy, increasing the rate of preserving the patient's erectile function [3, 4].

Undoubtedly, the most successful RAMIS device is the da Vinci Surgical System (Intuitive Surgical Inc., Sunnyvale, CA), with over 9200 da Vinci units installed worldwide,

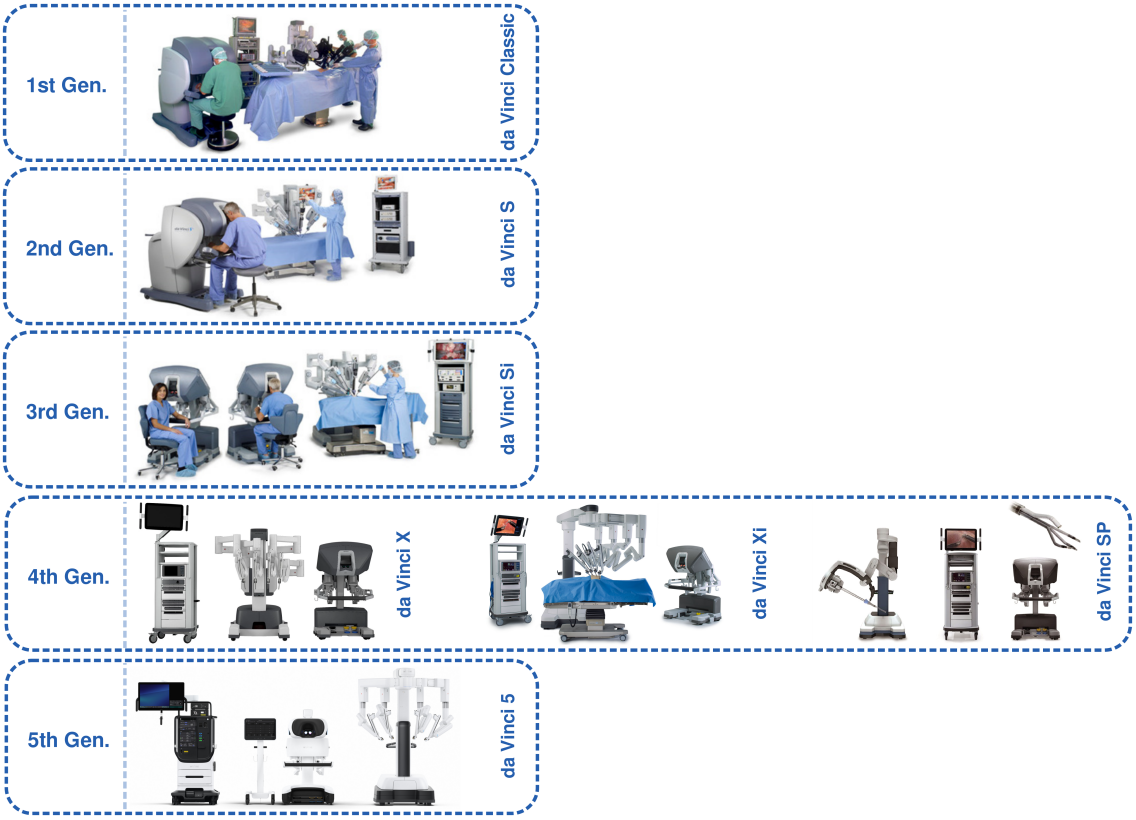

Fig. 1.1. The five generations of the da Vinci Surgical System: da Vinci Classic, launched in 2000 (1st generation); da Vinci S, launched in 2006 (2nd generation); da Vinci Si, launched in 2009 (3rd generation); da Vinci X, da Vinci Xi, and da Vinci SP, launched from 2014 to 2018 (4th generation); da Vinci 5, launched in 2024 (5th generation).

that performed more than 15 million procedures to date. It has been 20 years since the 1st generation was cleared by the U. S. Food and Drug Administration, and today, the 5th generation is available (Fig. 1.1.), along with the research-enhanced version of the original system, the da Vinci Research Kit (DVRK) [5, 6].

1.2 Partial Automation in Surgery

Many believe that the next step in the advancement of surgery will be subtask-level automation. Automating monotonous and time-consuming subtasks may decrease the cognitive load on the surgeon, who then could better focus on the critical steps of the operation [7]. Currently, many research groups are working on this problem [LT5][LTNR1]; some groups chose to work in ex vivo, in vivo [8], or realistic phantom environments [9], but simplified silicone phantoms are utilized mostly [10, 11, 12, 13][LT1, LT2][LTNR5].

Unfortunately for the researchers, autonomy in the surgical environment affecting mostly soft tissues, presents grave difficulties. Unlike working on rigid-tissues, where exact registration to the instrument is possible, soft tissues are permanently in motion, and highly deformable, thus no pre-computed tool trajectories can be used. Another challenge of surgical automation is undoubtedly the implementation of perception algorithms usable

in this complex environment. Computer vision suffers from reflections and the fact that the visual features of different organs being very similar, yet it is still the gold standard. New methods relying on palpation emerged in the last years, where force sensors can be utilized e.g., to find the location of tumors or other anatomies/pathologies [12, 14, 15][LTNR9].

Finally, irrespective of the nature of automation, the supervision of the surgeon is crucial and safety critical during the execution, that requires effective human–machine interaction. The surgeon should be able to parameterize and launch autonomous execution. They also must have the ability to observe the area of operation during autonomous execution, and to gain back manual control anytime if necessary, or the autonomous algorithm encounters events it cannot solve [16]. This requirement closely parallels the domain of self-driving cars, where human intervention remains essential in cases of sensor failures, unpredictable environmental changes, or ethical decision-making dilemmas [17][LT9][LTNR6]. In both fields, autonomous decision-making relies on real-time data interpretation, risk assessment, and predefined safety protocols; the challenge in both domains lies in creating robust algorithms that ensure safety and reliability while adapting to unpredictable conditions.

1.3 Research Goals

RAMIS often involves time-consuming and monotonous subtasks. Automating these subtasks can reduce the surgeons' cognitive load, allowing them to focus more on critical aspects of the procedure. Recent technological advancements, like deep learning or smart mechatronics, offer an increased capability in the automation of surgical subtasks; and consequently, it became a prevailing topic in the research community. Meanwhile, several challenges remain, including operating in a continuously changing soft tissue environment, difficulties in anatomical and pathological recognition due to visual limitations, such as glare and the complexity of both the procedures and the instruments used.

To address these challenges, my research aimed to:

1. **Develop a standardized methodology to support the automation of surgical subtasks in RAMIS.** The field of RAMIS automation is highly fragmented, with most systems being far from clinical translation. A unified, transparent framework is necessary to support the development of autonomous RAMIS systems.
2. **Establish robust validation metrics for clinical applicability.** Many developed systems lack proper validation and adjacent metrics, making comparisons difficult. A standardized methodology is required to assess and compare the clinical applicability of autonomous surgical systems.
3. **Monitor and quantify situational awareness in surgical automation.** Increased automation may reduce the operator's SA, potentially affecting performance. It is essential to monitor and quantify this effect to ensure safe and effective human–machine interaction.

Chapter 2

RELATED WORK

2.1 Degree of Autonomy of Surgical Equipment

Just over a decade ago, the joint International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) Technical Committee (TC) group analyzed the status of surgical robot standardization. Only one major gap was found: the Degree of Autonomy (DoA)—employed in *ISO 8373:2012 – Robots and robotic devices – Vocabulary*—was not defined properly. Understanding the fact that the proper definition of autonomy and its conjugated forms "autonomous", "automation", or related definitions can be unambiguous, the ISO/IEC joint working group decided to extend the scope of their work to all Medical Electrical Equipment (MEE) or Medical Electrical System (MES) with a DoA (other than zero). The discussion on the topic was conceived in a new Technical Report (TR) *IEC/TR 60601-4-1: Medical electrical equipment – Part 4-1: Guidance and interpretation – Medical electrical equipment and medical electrical systems employing a degree of autonomy*. The TR recommended to omit such words like "automation" or "automatic" within this robotic standard; DoA was defined instead as "taxonomy based on the properties and capabilities of the MEE or MES related to autonomy".

Derived from the field of industrial automation [18] and service robotics [19, 20], the TR recommends the parametrization of DoA along four cognition-related functions of a system, which are affecting options of an MES:

- **Generate an option:** to formulate possible options, based on the result of the monitoring task for achieving predefined goals;
- **Execute an option:** to carry out the selected option. Robots can typically be active or passive supporters of a surgical task execution;
- **Monitor an option:** to collect necessary information to perceive the status of MEE or MES, patient, operator, or environment. Therefore, signals beyond the internal (proprioceptive) control signals of the robot;
- **Select an option:** to decide on a particular option from the pool of generated.

$$DoA = F\{G|E|M|S\} \tag{2.1}$$

where the overall DoA metric is normed sum of the four functions of the system assessed on a linear scale, 0 meaning fully manual and 1 fully autonomous.

TABLE 2.1: Descriptive classification of Degree of Autonomy adapted from [18] based on *IEC/TR 60601-4-1*. H: the human operator performs the given function. C: the computer-driven system performs the given function.

DoA	Description	Monitor	Generate	Select	Execute
1.	**Full manual (FM):** No autonomy involved. The operator performs all tasks including monitoring the state of the system, generating performance options, selecting the option to perform (decision making) and executing the decision made, i.e., physically implementing it.	H	H	H	H
2.	**Teleoperation (TO):** The equipment assists the operator with the execution of the selected action, although continuous operator control is required. The operator performs all tasks, including monitoring the state of the equipment, generating options, selecting the desired option and execution of it. (Master–Slave teleoperation.) Note: traditional robotics standards consider teleoperation as zero DoA.	H/C	H	H	H/C
3.	**Pre-programmed execution (PE):** The operator generates and selects the options to be performed without any analysis or selection by the equipment. Note: traditional robotic standards considered this as "autonomous" or "automatic" operation.	H/C	H	H	C
4.	**Shared decision (SD):** Both the operator and the equipment generate possible decision options. The operator retains full control over the selection of which option to execute. Both the operator and the equipment participate in the execution.	H/C	H/C	H	H/C

5.	**Decision support (DS):** The equipment generates a list of decision options, which the operator can select from, or the operator may generate alternative options. Once the human has selected an option, it is turned over to the equipment to execute it.	H/C	H/C	H	C
6.	**Blended decision (BD):** The equipment generates a list of decision options, which it selects from and executes if the operator consents. The operator may also generate and select an alternative option; the equipment will then execute the selected action. BD represents a high-level decision support system that is capable of selecting among alternatives as well as executing the selected option.	H/C	H/C	H/C	C
7.	**Guided decision (GD):** The equipment presents a set of actions to the operator. The operator's role is to select from this set, he/she cannot generate any other additional option. The equipment will fully executes the selected action.	H/C	C	H	C
8.	**Autonomous decision (AD):** The equipment selects the best option and executes it, based upon a list of alternatives it has generated (this list can be augmented by alternatives suggested by the operator).	H/C	H/C	C	C
9.	**Operator monitoring (OM):** The equipment generates options, selects the option to implement and executes it. The operator monitors the equipment and intervenes if necessary. Intervention places the human in the role of making a different option selection. During the procedure there may be decision making points that will be decided by the equipment.	H/C	C	C	C

10.	**Full autonomy (FA):** The equipment carries out all actions. The operator does not intervene except to e-stop the equipment (which is a general requirement).	C	C	C	C

The objective assessment of the DoA of a system can be utilized along the four described functions, each can be performed by a human or by a computer. The DoA of the system is defined on a scale from 0 to 1; DoA = 0 means "no autonomy", and the highest DoA represents "full autonomy". The low-level electronic and computational functions of MEE or MES, like communication or motor control, are excluded from this definition, as the term "no autonomy" is meant on the system level. Also, a classification of DoA can be given at different levels of granularity, depending on the level where those functions are implemented.

Autonomous execution is already present in surgery (especially in image-guided surgery), however decision making (selecting) is mostly done by human experts. On the other hand, computer systems are more capable in monitoring tasks compared to humans, hence most critical processes happen at a faster scale humans can perceive; this is the safety concept of Situation Awareness (SA).

The TR offers various alternatives for DoA assessment; the most applicable one, based on industrial automation, defines 10 levels of DoA (Table 2.1). Practically, during the risk management of a surgical robot (most commonly according to ISO 14971 – Application of risk management to medical devices), DoA should also be taken into account; DoA does not necessarily correlates with the level of risk, but it may impact risk management gravely. In the case of an error or malfunction, the take-over of the human operator may be necessary. Due to loss of SA the operator may not be able to control the situation properly, thus the DoA determines the handling of the hazards. At lower DoA, the responsibility can be shared between the human and the robot; at higher DoA malfunctions presents critical hazards.

2.1.1 Basic Safety and Essential Performance of Surgical Robots

From the aspect of the user (and also of the manufacturer), avoiding and managing any kind of failure (software, hardware, communication, system-level) is critical. In the past three decades, 33 documented casualties were caused by industrial robots, which is still only giving 0.0005% of all work-related deaths [21]. None of the surgical robotic cases are included in this statistics. Alemzadeh et al. conducted a study in the field of robotic surgery [22], using the United Sates Food and Drug Administration (FDA) Manufacturer and User Facility Device Experience (MAUDE) database, covering reports from 2000 to 2013. According to their findings, out of 1.74 million robotic procedures, 10,624 adverse events were reported, including 1,391 injuries and 144 deaths. A high number of injuries were caused by electrical arcing, which led to injuries such as burned tissues. Overall, 9 patient deaths were reported in conjunction with device malfunctions, such as unintended operation of instruments or detachment of broken pieces, fell into the patient's body. In the

case of surgical robotics—especially in RAMIS—the human surgeon takes full liability for the outcome of the intervention (as ruled in all juridical cases up to now).

With the increase of DoA, risk mitigation and management become crucial. Methodologies to support the safety of design and development of robotic devices were published by various groups, like the generic Hazard Identification and Safety Insurance Control (HISIC) policy that has been applied to several robotic systems [23].

In 2015, the ISO/IEC TC 62/SC 62D joint committee started a discussion on the minimum requirements for a practical degree of safety for surgical robots; the results were published in 2019 as *IEC 80601-2-77: Particular requirements for the basic safety and essential performance of Robotically Assisted Surgical Equipment (RASE)*. *IEC 80601-2-77* is tailored specifically for the invasiveness of surgical robots, making it distinct from non-medical robotic applications. The standard collects all relevant thermal and mechanical hazards, along with the required usability trials. It also defines the basic types of surgical robots and their relevant components. In terms of RAMIS, the focal points of the standard are:

- robotic surgical instruments;
- the patient-side part of the robot;
- the operator-side part of the robot;
- the endoscope holder (if any).

The same committee also finalized a standard, focusing on the hazards related to loss of SA, namely *IEC 80601-2-78: Particular requirements for the basic safety and essential performance of medical robots for rehabilitation, compensation or alleviation of disease, injury or disability*. This standard defines SA as "the operator's perception, comprehension, and prediction of a robot's behavior in its environment". SA is a key factor in tasks where human supervision or interaction with the robot is necessary to reduce risk. The standard states that the manufacturers will have to include necessary SA information for their future medical robotic systems. The quality of the Human–Machine Interface (HMI) is essential to keep SA; if the critical information is forwarded to the human operator through an adequate way, e.g., sound alerts, that may highly increase SA.

By the introduction of Artificial Intelligence (AI) methods to robotic surgery, the system may also offer decision support to handle complex situations [16]. With the increasing role of AI-driven decision support in surgical robotics, *ISO/IEC 25059: Quality models for AI systems* provides a standardized quality model for AI systems. This standard outlines metrics for evaluating AI performance, ensuring that AI components used in surgery meet rigorous accuracy, robustness, and transparency criteria. From August 1, 2024, the European Union enacted the *EU AI Act*, introducing a risk-based regulatory framework for artificial intelligence, categorizing AI systems into different risk levels. High-risk AI systems, including those that could negatively impact human health, safety, or fundamental rights, are permitted but must comply with strict requirements and obligations before entering the EU market. These requirements include transparency, human oversight, data governance, robustness, and accuracy to ensure AI systems operate safely and ethically. The *EU AI Act* aims to balance innovation with public safety, ensuring that AI technologies enhance healthcare outcomes without compromising patient well-being or ethical standards.

It is believed in the community, that upcoming standards and regulations would focus more on the safety of the patients and the improvement of the treatments rather than technical metrics, and willing to continue work aiming that goal.

2.2 Level of Autonomy for Surgical Robots

Although the standards, mentioned above are fundamental for the assessment of the capabilities of surgical robotic systems, they are not definite enough to present taxonomy to generally assess the development phases of surgical robotics, or to perform benchmarking. A structured classification system is necessary to categorize surgical robotic systems based on their advancement relative to the field. A gradual mapping was presented in [24], to classify the autonomous capabilities of surgical robots. Some earlier work suggested to put the HMI into the center of the classification, defining a 0–7 scale [25]. Similar concepts are also presented in the field of self-driving; in [26] a 6-grade scale was introduced for autonomous vehicles. At higher levels of autonomy, the role of the robot is still argued; the robot is not only a medical device anymore, but it also practices medicine, which is entirely different from the viewpoint of regulatory. The FDA, for example, regulates medical devices, but not the practice of medicine.

The mapping of [24] has one fundamental problem in the middle ranges Level of Autonomy (LoA), where the most of the current autonomous capabilities would fall into: this mapping offers no metric to determine the level of human supervision required. The role of SA may be crucial to distinguish the cognitive level up to which the human may be able and shall be allowed to perform take-over; described as human-on-the-loop control [27].

Any autonomy classification must account for human supervision dependency. Since human sensory and cognitive processing capabilities are limited, the loss of SA can directly impact a surgeon's ability to intervene when required. Whether in partial autonomy—where the system assists but requires human oversight—or conditional autonomy—where the system operates independently under certain predefined conditions—the distinction lies in how and when human intervention is expected. The key factor is the cognitive time horizon available for human response, which determines the feasibility of safe supervisory control in an autonomous surgical system.

Coherent to the current standardization efforts, yet fitting to the commonly used terms, the following scale of LoA is suggested [LT5] (Fig. 2.1):

- **LoA 0 — No autonomy:** all system-level functions (generating, selecting, executing, and monitoring actions) are performed by the human operator. Technically it means that during the surgery no active robotic equipment is used, thus it may be considered identical to a non-robotic case.
- **LoA 1 — Robot assistance:** the surgical robot performs specific, low level functions only. E.g., teleoperated systems, tremor filtering, minor safety features.
- **LoA 2 — Task-level autonomy:** the system is trusted to complete certain tasks or sub-tasks in an autonomous manner. E.g., image-guided bone drilling, wound closure. It may only happen for a short instance.
- **LoA 3 — Supervised autonomy:** the system can autonomously complete large sections of a surgical procedure, while making low-level cognitive decisions. All actions are performed under human supervision, assuming the operator's full SA.

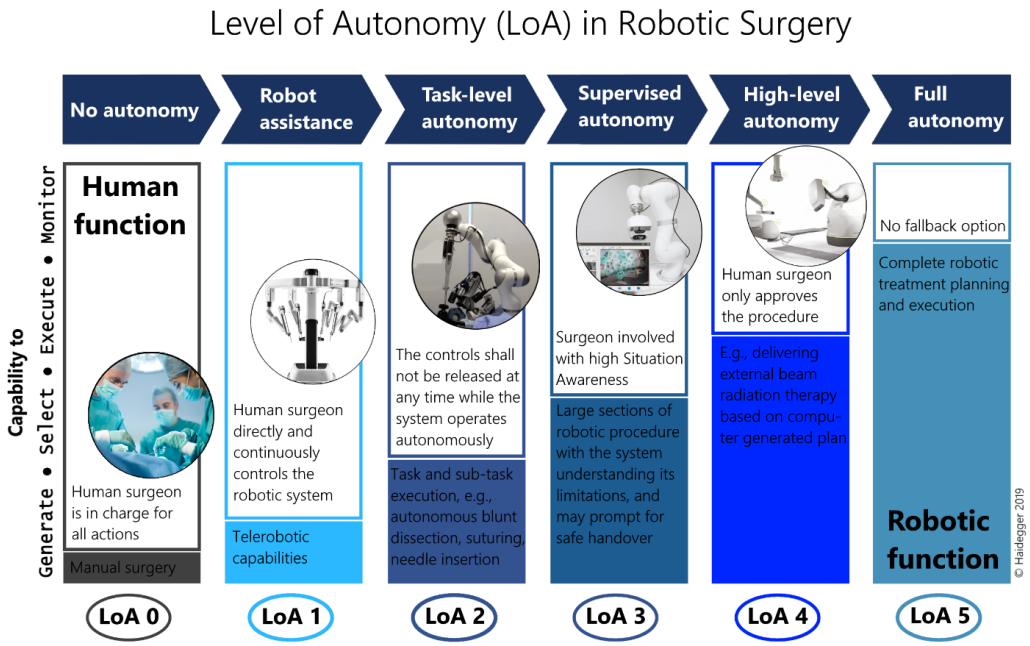

Fig. 2.1. The 6-grade classification for assessing the autonomous capabilities of surgical robots. The concept of Level of Autonomy follows the ISO/IEC standardization framework, determining LoA based on the human versus robotic functions of the system [LT5].

- **LoA 4 — High-level autonomy:** the robotic system executes complete procedures based on human-approved surgical plans, while the human only has the capability to emergency stop (e-stop) the procedure. The robot shall be able to complete the task even if the human fails to respond appropriately to a request to intervene.
- **LoA 5 — Full autonomy:** the robotic system operates autonomously at all times, managing both routine and adverse conditions without requiring human intervention. The system succeeds in scenarios where even the best human operator would fail, therefore there will be no need for a human fallback option.

Unlike DoA, this LoA definition is empirical, focusing on the key enabling robot capabilities of a system. Full autonomy of surgical robots still belongs to the domain of science fiction, however, several techniques based on AI and machine learning are being under intensive research by various research groups [28, 29]. Many believe that, similarly to domain of self-driving vehicles, the market will suddenly get interested in autonomy, as LoA 3 (Supervised Autonomy) becomes reliable and widely used.

2.3 Recent Trends in Automating Surgery

The first papers on RAMIS subtask automation appeared in the middle of '00s, with focus on knot-tying and suturing [LTNR1]. Currently, partial (or conditional) automation is the most intensively researched domain of surgical automation. The workflow of RAMIS interventions often contains subtask elements, that are time-consuming and repetitive, such as blunt dissection or grasping and holding tissues [30]. Partial automation—the

automation of such subtasks—may reduce the cognitive load and fatigue on the human surgeon, making possible them to pay more attention on the critical subtasks of the operation [LTNR5]. The technological advancements of the past few years in the domain of deep learning or mechatronics offer a rising potential on the research of surgical subtask automation [31].

In the most recent years, the automation of simple surgical training exercises on rigid [41, 43, 48, 49, 50, 51] or deformable [32, 52] phantoms tends to get into the focus of attention. Among all the training exercises, the automation of different versions of *peg transfer* is presented in a significant number of studies [41, 43, 48, 49, 50, 51], probably due to its simplicity, enabling to elaborate the basic principles and best algorithms for automation. A number of further subtasks are currently under active research, such as different aspects of suturing, soft tissue cutting, debridement, palpation, or blunt dissection [32, 38, 39, 11, 12, 8, 40, 10, 28, 37]. These works are compiled into Table 2.2.

One of the first successful projects in the domain—the work of UC Berkeley AUTOLAB and Center for Automation and Learning for Medical Robotics (CAL-MR)—was presented [32], with not one, but two surgical subtasks completed autonomously in phantom environment, using the DVRK: multilateral (using at least two arms) debridement (Fig. 2.2a–b) and shape cutting (Fig. 2.2c). In this work, the learning by observation approach was used: human motion patterns were recorded and segmented, and then those patterns were used to generate robot trajectories during autonomous execution. The motion segments were, e.g., in the case of debridement: motion, penetration, grasping, retraction, and cutting. In order to autonomously execute the motion segments, a state machine was compiled for each subtask. The state machine required parameters for the motion segments for execution, e.g., the height of lifting motion in the case of retraction. The parameters were determined empirically, using binary search methodology. After each

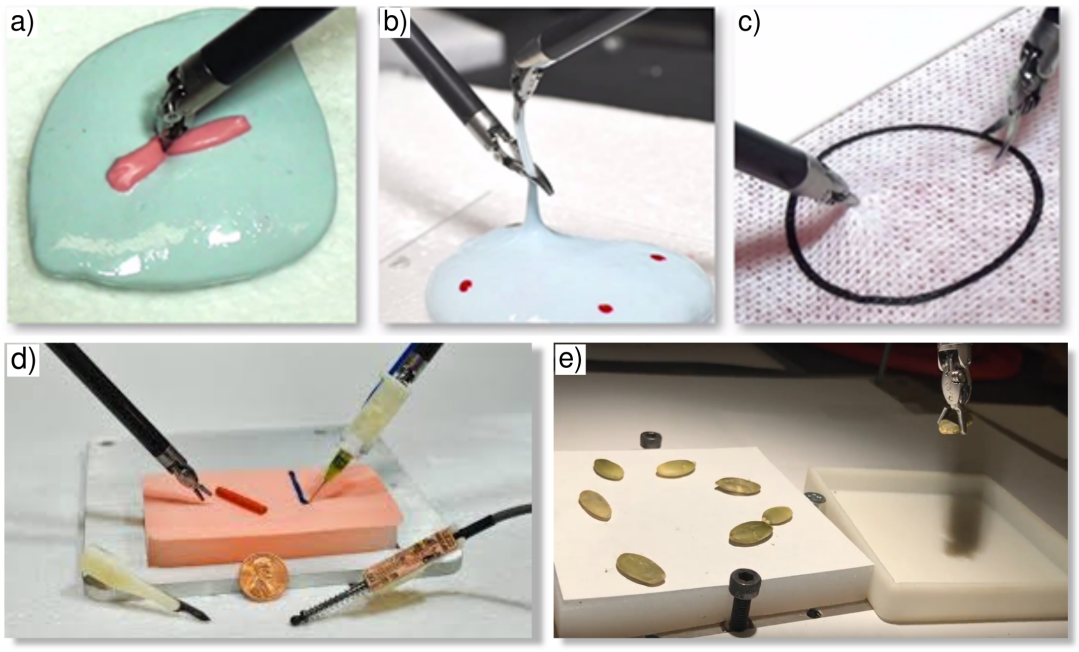

Fig. 2.2. Recently automated surgical subtasks. a-c) Multilateral cutting [32]; d) tumor palpation [39]; and e) resection, debridement [10].

TABLE 2.2

LIST OF SURGICAL SUBTASKS FROM THE ASPECT OF SUITABILITY FOR PARTIAL AUTOMATION.

Subtask	**Platform**	**Sensory input**	**Experimental environment**	**Complexity**	**Clinical relevance**	**Ref.**
shape cutting	DVRK	stereo camera	gauze patch, FRS Dome[1]	medium	high	[32]
suturing	DVRK	stereo camera	silicone, foam, FRS Dome	high	high	[11]
tissue manipulation, needle handling, knot-tying	DVRK	wrist cameras	silicone phantom, animal tissues	high	high	[33]
needle handling, vessel dilation, shunt insertion	DVRK	–	simulation, special phantom	high	high	[34]
peg transfer, needle handling, gauze retrieve	DVRK	stereo camera	simulation, special phantom	medium	high	[35][36]
ligation	EndoBot	–	special phantom	medium	high	[37]
palpation	DVRK	force sensor	special silicone phantom, FRS Dome	medium	medium	[12][38]
tumor palpation and resection	DVRK, RAVEN	force sensor	special silicone phantom, FRS Dome	high	medium	[39]
debridement	DVRK	stereo camera	tiny objects	medium	high	[40][10]
suction and debridement	DVRK	–	special phantom	medium	high	[28]
bowel anastomosis	STAR[2]	RGB-D camera	porcine bowel	high	high	[8]
peg transfer	DVRK	RGB-D camera	peg transfer board	medium	low	[41][42]
peg transfer	DVRK	RGB-D camera	peg transfer board	medium	low	[43][44] [45]
peg transfer	DVRK	stereo camera	peg transfer board, simulator	medium	low	[46]
peg transfer	DVRK	stereo camera	peg transfer board, simulator	medium	low	[47]

motion segment was parameterized and tested individually, the whole state machine was tested again, and the parameters updated, if necessary. The positions of the targets were estimated by computer vision, two pairs of stereo cameras were used to observe the field of operation. The debridement targets and the circle for cutting were detected in the images, and robot's trajectories were translated based on the 3D coordinates of the detected objects. To evaluate the autonomous execution, autonomous debridement was executed 10 times with 5 targets each, and shape-cutting were performed 20 times. The repeatability of the subtasks (ratio of successful trials) was respectively 96% and 70%.

Another work of the same research group [39], aimed at autonomous multilateral tumor resection based on palpation in phantom environment (Fig. 2.2d). To achieve the completion of these series of subtasks, custom-built instruments were designed, attachable to end effector of the da Vinci: a palpation probe, a scalpel, and a fluid injector. Also, a state machine was compiled to execute the series of subtasks: scan the phantom by palpation and localize the hard inclusion, making of the incision, removal of the inclusion (debridement), and fluid injection to close the wound. To evaluate the performance of the system, 10 end-to-end trials were performed, with overall 50% success rate. In 2 of the trials, the tumor location was estimated incorrectly, another 2 times retraction failed, and in 1 trial the tumor was not fully resected, which shows the challenges given the complexity of any surgical procedure. The authors asserted that the performance could be enhanced by visual feedback and are planned to include computer vision features in the future.

Suturing is probably the most intensively researched subtask of RAMIS, it occurs quite frequently in the workflow of surgical interventions, yet extremely time-consuming for the surgeon, and challenging for automation. Suturing has two, highly difficult aspects: needle guidance through the tissue on a given trajectory, and also thread manipulation, especially during knot-tying [53, 54]. In another work of the UC Berkeley Automation Lab [11], a solution for the former one is presented. To precisely grasp the needle, a positioning adapter (Suture Needle Angular Positioner—SNAP) was designed, that itself achieved a 3-fold error reduction in needle pose. The needle position was estimated using camera image, the needle size, trajectory, and control parameters using were optimized by sequential convex programming. During the trials, the system was able to complete 86% of attempted suture throws successfully. In a recent study by Kim et al. [33], the automation of surgical manipulation subtasks—tissue manipulation, needle handling, and knot-tying—,using imitation learning and action chunking with transformers, is presented. The inconsistency of kinematic data from the da Vinci is overcome by relative action formulation. Additionally, they tested and evaluated the use of wrist cameras in the proposed imitation learning technique.

Implementation of autonomous subtasks on cable-driven robots, like the da Vinci or the RAVEN can be challenging due to their inherent non-linearities. The inaccurate robot positioning causes no issue in teleoperation, as the human surgeon, who follows the tool position on the endoscopic camera stream, is part of the control loop. However, in the case of the automation of position-critical subtasks those inaccuracies can easily cause failures. In Seita et al. [10] a two-phase calibration method was presented, to decrease position errors of cable-driven surgical robots, using deep neural network and random

[1]Fundamentals of Robotic Surgery, Florida Hospital Nicholson Center, Celebration, FL

[2]Smart Tissue Autonomous Robot, Johns Hopkins University, Baltimore, MD

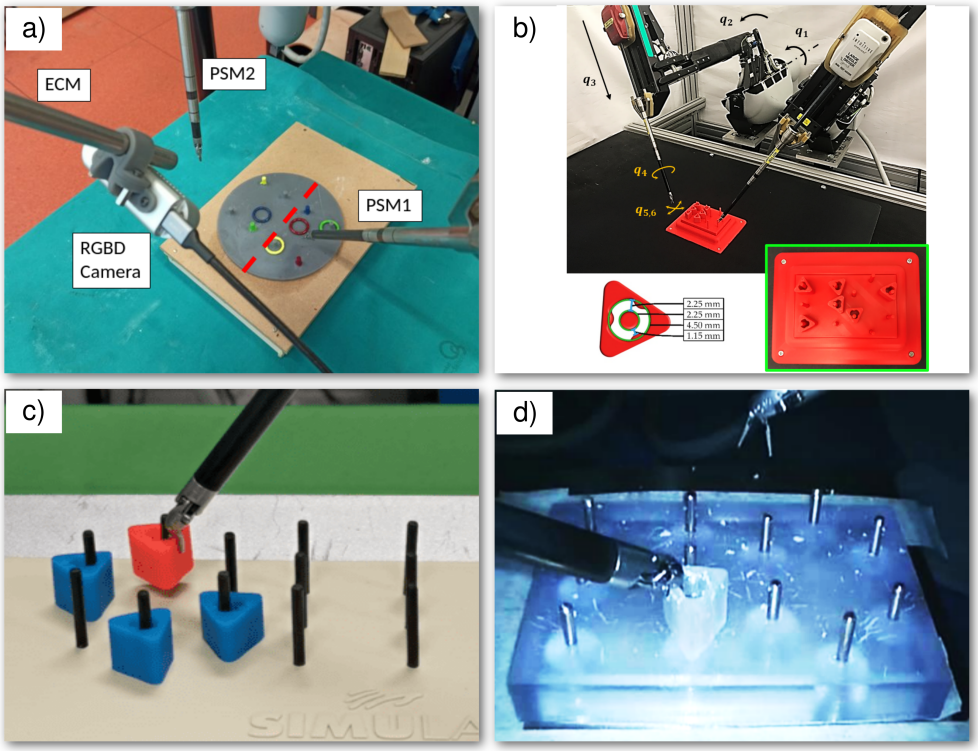

Fig. 2.3. Recent works and their experimental setups for autonomous peg transfer: a) The solution of Ginesi et al., utilizing answer set programming and dynamic movement primitives [41]; b) Automation using deep recurrent neural networks by Hwang et al. [43]; c) A reinforcement learning-based implementation by Xu et al., with the open-source SurRoL platform [46]; d) A shared control approach utilized learning by observation technique by Zhang et al. [47].

forest techniques. By precise calibration, the debridement subtask was automated with 94.5% success rate (Fig. 2.2e).

Since the peg transfer training exercise is probably one of the most intensively researched subtasks in the field of surgical automation, this exercise deserves to be mentioned separately. The first paper on surgical automation that mentions peg transfer dates 2014 [40], however, its automation received the most attention after the year of 2018.

Ginesi et al. presented a solution on autonomous peg transfer, with focus on the high-level decision making (Fig. 2.3a) [41]. In their study, the task was to transfer colored blocks (rings) to the appropriate colored pegs, that in instances required the removal of other, obstructing blocks first. This problem was solved by answer set programming; the tool trajectories were generated using dynamic movement primitives. Their later work [42] focuses on dynamic movement primitives and obstacle avoidance.

Hwang et al. proposed implementations with emphasis on decreasing the positioning error of the robotic instruments of the DVRK (Fig. 2.3b) [43, 44, 45]. This was achieved using deep recurrent neural networks and a 3D printed fiducial for hand–eye registration. The peg board utilized was also fully 3D printed; the source files for the board and also for the fiducial are available online,[3] making this setup highly usable as a benchmark environment.

[3] https://sites.google.com/view/surgicalpegtransfer

A reinforcement learning-based solution is presented in the work by Xu et al. (Fig. 2.3c) [46]. The teaching of the system was performed in the open-source SurRoL platform, offering a simulated gym environment for performing surgical tasks. Later, Huang et al. [35] used the same simulation environment for a demonstration-guided reinforcement learning-based method that narrows down exploration space by encouraging expert-like behaviors and enabling robust guidance when confronting states unobserved in demonstrations. Zhang et al. presented a solution to utilize a simulated environment to generate the data required for learning by observation, then translate the learned principles to the physical setup (Fig. 2.3d) [47].

The revolution of Large Language Models (LLMs) since the 2022 release of ChatGPT 3.5 [55] has also impacted this research field. Moghani et al. [34] introduced SuFIA, a framework for natural language-guided augmented dexterity that incorporates the reasoning capabilities of LLMs, enabling a learning-free approach without requiring in-context examples or motion primitives. SuFIA follows a human-in-the-loop paradigm, allowing the surgeon to regain control when information is insufficient, thereby mitigating unexpected errors in mission-critical tasks. Fu et al. [36] proposed a goal-conditioned decision transformer, which leverages large-scale transformers from LLMs to automate goal-conditioned surgical robot subtasks, such as peg transfer and needle picking.

Chapter 3

MATERIALS AND METHODS

3.1 Software Tools and Environment

To realize the data collection and processing for surgical subtask automation, complete research platforms have to be built and constructed, bringing computer technology to the operating room. On in academic domain, the Robot Operating System (ROS) [56] platform is widely used in the research of robotics, and also preferred by many in the medical robotics domain; most of the research centers, working on the two dominant platforms presented below, rely on ROS. ROS is undoubtedly a powerful, modular tool with already implemented solutions for most of the frequently occurring problems of the field, such as stereo-camera calibration or acquisition of sensory data. A new version of the platform—ROS 2—was released in 2017, offering, among other improvements, better support for real-time systems and enhanced communication, making it more suitable for large-scale and industrial applications. The transition to this new version is still ongoing; currently, both versions coexist. Some parts of the presented implementations have already been ported to ROS 2, while the discussion in this thesis remains independent from the specific ROS version.

3.2 Available Open Research Platforms for Automating Surgical Subtasks

The research projects on surgical subtask automation have utilized a number of robotic devices during the last decade, including medical, industrial, or custom-built robots. However, the two RAMIS research platforms presented in the followings appear most dominantly in the field, probably to ease the future translation of the developed methods to the clinical practice.

3.2.1 The da Vinci Research Kit

Roughly 8 years ago, when the 1st generation da Vinci robots (da Vinci classic) was sent to retirement due to the discontinued service and supply, the old systems found another purpose. Those systems were still functional and could be utilized in applications more

tolerant to malfunctions. At the Johns Hopkins University, the development of a research platform for those robots—the da Vinci Research Kit (DVRK)—was concluded, and only within a few years, an active community was built with more than 40 setups worldwide [57].

DVRK is a fully open-source platform, consisting of custom hardware and software elements, in order to open the possibility of programming the da Vinci arms. The controllers—developed to operate the arms—are built on custom boards: an IEEE-1394 FPGA board for computational power and low latency communication and a Quad Linear Amplifier (QLA) for high-frequency low-level robot control. The controllers are connected to PC using IEEE 1394a (FireWire). On the PC side, the open-source *cisst* libraries are responsible for the handling of FireWire communication and the mid-level control of the robot. The cisst libraries offer the functionality to program the arms themselves. Additionally, cisst is also interfaced with ROS, which interface is currently used to program the da Vinci arms at more than half of the DVRK locations [57, 58].

3.2.2 The RAVEN Platform

The RAVEN-I platform was originally developed at the University of Washington in the mid-2000s, aiming for space use, and other specific application areas. Hence most surgical robots were bulky, and meant to be used dominantly in the operating room. RAVEN-I was to be a new, lighter, portable, and still durable surgical robot, with possibility to be used on the field. After it has proved its versatility and durability in a number of experiments, e.g., in a trial on an underwater research station, as a part of the NASA's NEEMO program, its research potential was also soon discovered. In the beginning of 2010s, the University of Washington Biorobotics Lab and the University of California Santa Cruz Bionics Lab developed an updated design of the system, named RAVEN-II. Later, Applied Dexterity[4] was formed to support the RAVEN community, and also the development of the RAVEN-III platform was started. Today, there are 16 RAVEN sites worldwide within the cutting-edge research of surgical robotics.

The research platform is fully open-source, consists of two 3 DoF positioning arms, with 4 DoF attachable instruments—similar to the da Vinci. Like the DVRK, the system is Linux-based, and uses ROS interface for programming [59].

3.3 Description of Surgical Subtasks

This thesis focuses on two surgical subtasks as models to support the developed concepts and methodologies. These two subtasks presented below—blunt dissection and peg transfer—, are simple enough to enable work on the very fundamentals of surgical subtask automation, yet relevant from the clinical aspect.

3.3.1 The Blunt Dissection Surgical Subtask

Blunt dissection is a surgical subtask, where the surgeon carefully separates two tissue layers without using the instruments' cutting edges in an effort to avoid any damage to

[4] `http://applieddexterity.com/about/history`

sensitive tissue structures (e.g., vessels, nerves). During blunt dissection, the retractor holds the tissues, and the dissector is inserted between the two layers, then by opening of the dissector it forces the two layers apart. This surgical subtask is a recurring element in multiple procedures, where automation could ease the cognitive load on the surgeon by relieving him/her from the repetitive manipulation of instruments, so the surgeon may pay full attention to the patient-specific details of the surgery. Furthermore, robotically executed procedures can provide an increased accuracy compared to the human operator, therefore it can take effect the success of the operation. The frequency of blunt dissection depends on the type of surgery. It is commonly used in laparoscopic, robotic, and open surgeries, particularly in procedures involving delicate tissue separation. The duration of blunt dissection varies significantly based on the procedure and the surgeon's expertise but can range from a few minutes to a considerable portion of the operation, especially in cases requiring extensive tissue preparation [30, 60]. During Laparoscopic Cholecystectomy (LC) procedures, blunt dissection is a commonly employed subtask to expose the Calot triangle to avoid bile duct injuries [61].

3.3.2 The Peg Transfer Training Exercise

Peg transfer is probably the most frequently used exercise in MIS and RAMIS training to improve hand–eye coordination and motor skills. It is also one of the five tasks of the Fundamentals of Laparoscopic Surgery (FLS) exam [62]. The training task consists of a pegboard, with two sets of 6 pegs, an 6 blocks (Fig. 3.1). The exercise as defined by FLS:

1. grasp each block with the non-dominant hand;
2. transfer the block mid-air to the dominant hand;
3. place the block on a peg on the opposite side of the pegboard.

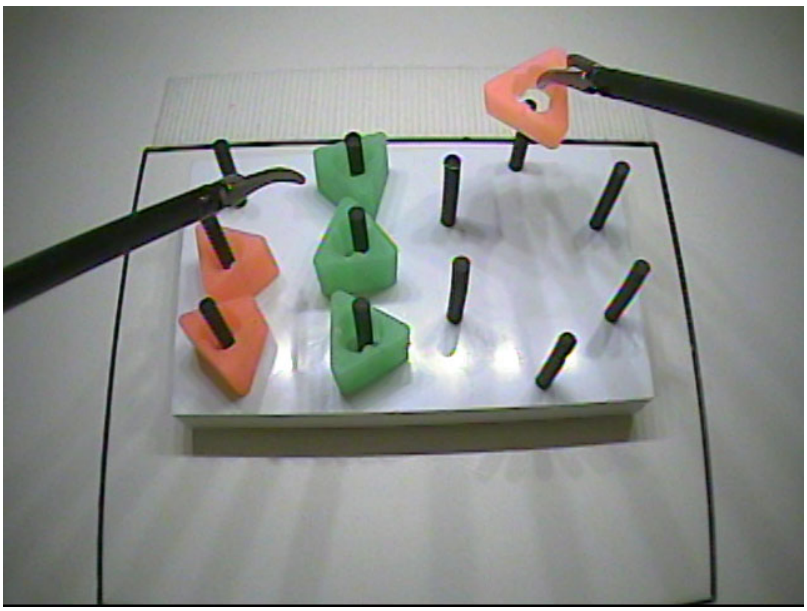

Fig. 3.1. The peg transfer exercise is a core of the five tasks in the Fundamentals of Laparoscopic Surgery (FLS) exam [62].

Once all six blocks have been transferred to the opposite side of the board, reverse the process and first grasp each block with the dominant hand, transferring mid-air to the non-dominant hand, and place it on the original side of the pegboard. FLS scores the task as follows: a penalty is applied if a block is dropped outside of the field of view; there is no penalty for dropping the block within the field of view, if the block can be retrieved; also, the task needs to be performed within a time limit (300 seconds for FLS exam).

Different variations can be defined for peg transfer. For instance, Hwang et al. [43] defined three variations (Fig. 3.2):

- *Unilateral Peg Transfer* – Transfers executed by a single arm;
- *Parallel Bilateral Peg Transfer* – Transfers executed by two arms in parallel;
- *Bilateral Handover Peg Transfer* – Transfers executed by both arms, with a mid-air transfer between the two arms.

Important to note that these variations have been defined for the purpose of automation research, instead of surgical training. The *Bilateral Handover* method differs from the *Parallel Bilateral* method in that it involves a transfer between the two arms mid-air, as opposed to simultaneous parallel actions. This can have advantages in terms of precision and flexibility, but introduces additional complexity and time due to the need for precise coordination between the two arms. On the other hand, the *Parallel Bilateral* method is more efficient, reducing the time required for the task.

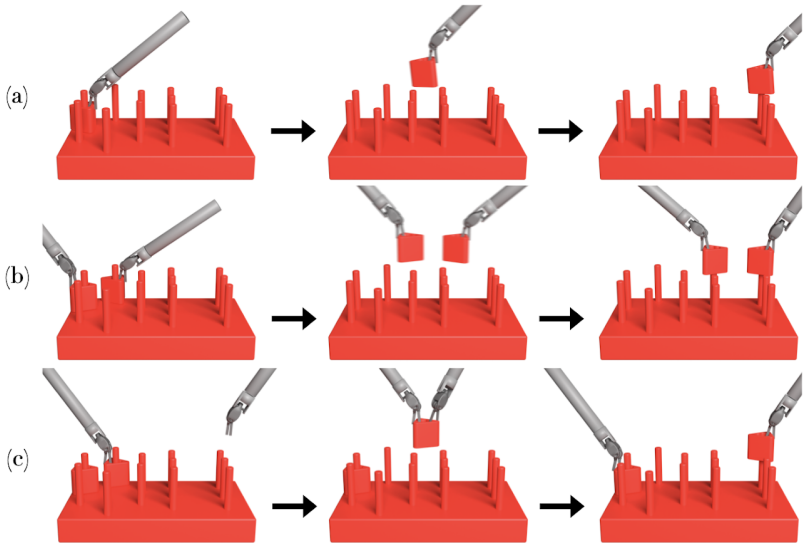

Fig. 3.2. Variations of the peg transfer exercise by Hwang et al. (a) *Unilateral:* transfers executed by a single arm; (b) *Parallel Bilateral:* transfers executed by two arms in parallel; (c) *Bilateral Handover* transfers executed by both arms, with a mid-air transfer between the two arms [43].

Chapter 4

METHODOLOGY FOR THE AUTOMATION OF SURGICAL SUBTASKS

4.1 Subtask Automation in Surgery

The first papers on RAMIS subtask automation appeared in the middle of '00s, with focus on knot-tying and suturing [LTNR1]. Currently, partial (or conditional) automation is the most intensively researched domain of surgical automation. The workflow of RAMIS interventions often contains subtask elements, that are time-consuming and repetitive, such as blunt dissection or grasping and retracting tissues [30]. Partial automation—the automation of such subtasks—may reduce the cognitive load and fatigue on the human surgeon, making possible them to pay more attention on the critical subtasks of the operation, requiring complex decision-making and high SA, such as the clipping and cutting of vessels or handling adverse events [63, 64, 65][LTNR5].

As the development of the technological background in the last couple of years offers a rising potential, like deep learning or mechatronics, the automation of surgical subtasks became a prevailing topic in the research of surgical robotics. A number of autonomous surgical subtasks are already implemented, or being currently developed by various research groups. A list of relevant subtasks in the research of surgical automation was presented in Table 2.2.

All of the mentioned surgical subtasks are to be performed on soft tissue, in a highly deformable environment. In contrast to subtasks involving hard tissue, like bone cutting, where the target organ can be fixed and registered to the surgical device via a navigation system, soft tissue presents new challenges from the aspect of automation, as the robot has to operate in unpredictable environment. Probably the biggest challenge is the development of perception algorithms; it is not trivial how the information, needed for the execution of the current subtask can be extracted from the surrounding soft, reflective environment. Despite the fact that working implementations could be found e.g., on instrument segmentation/pose estimation [66, 67] or organ segmentation and 3D reconstruction [68, 69, 70], autonomous navigation inside the patient's body still presents a huge challenge and being under intensive research. As of today, shared control is a more

viable option for these clinical routines [71, 72]. Furthermore, the generation of required motion patterns and the design of control methods for the manipulation of unknown soft tissues are also problematic [73].

My aim was to develop a methodology and an open-source framework to support such development projects; to provide software packages that contain already implemented basic functionalities, eventually becoming universal building blocks in surgical subtask automation. The architecture of this software package—the iRob Surgical Automation Framework, or `irob-saf`—is presented here.

4.2 Granularity Levels of Surgical Motion

A key enabling approach to manage complexity is dividing the surgical workflow into atomic elements. Identifying surgical subtasks makes it possible to guide the robotic instruments during soft tissue operations, following the decomposed surgical motions. This can be performed on different levels of granularity, and can be used to compile a parameterizable motion library.

One of the fundamental tasks in the development of this surgical automation framework was the hierarchical decomposition of surgical motion patterns. The workflow of surgical interventions, as well as the motion of the surgeon, can be decomposed into elements on different levels of granularity [54], similar to behavior trees [74]. In the literature, several different definitions of granularity levels of the surgical workflow have been found, such as : *Procedure, Task, Maneuver, Gesture* by Vedula et al. [54]; *Operation, Step, Subtask, Motion* by Mackenzie et al. [75]; *Task, Surgeme* by Gao et al. [53]; *Procedure, Phase, Step, Action* by Mascagni et al. [76]; *Procedure, Task, Gesture* by Ma et al. [30]. Although these different definitions share similarities and contain analogies between the various vocabularies, no consistent definition has been established for the entire domain. Also, most of the definitions serve the need of surgical skill assessment and workflow analysis. To decompose surgical motion and implement partial automation, it is necessary to define these levels as precisely as possible. For that manner, the levels of granularity are defined as follows (Fig. 4.1). Additionally, to better align with automation, an additional level was introduced—*Motion primitive*—which serves as the foundation for implementing robotic movements.

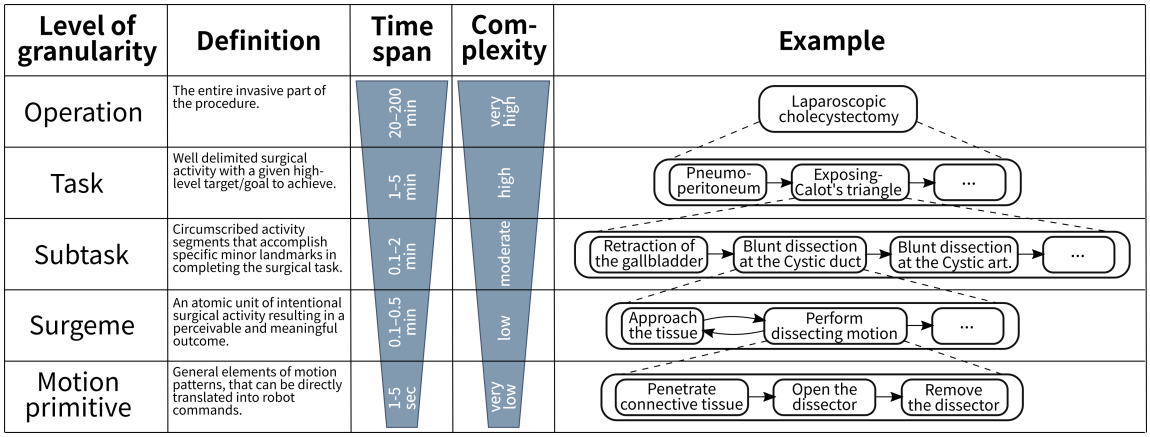

Fig. 4.1. Overview of surgical motion's granularity levels. Mapping of an example, Laparoscopic Cholecystectomy procedure onto different granularity levels.

1. **Operation:** The entire invasive part of the surgical procedure.
2. **Task:** Well delimited surgical activity with a given high-level target/goal to achieve.
3. **Subtask:** Circumscribed activity segments that accomplish specific minor landmarks in completing the surgical task.
4. **Surgeme:** An atomic unit of intentional surgical activity resulting in a perceivable and meaningful outcome.
5. **Motion primitive:** General elements of motion patterns, that can be directly translated into robot commands.

In most studies, the granularity level chosen for surgical automation is the level of subtasks (Table 2.2). The execution of those subtasks usually leads to the accomplishment of a specific milestone, which is in line with the term of partial automation. Subtasks can be further divided into *surgemes*, which are universal to different subtasks. Thus, from the viewpoint of automation, different subtasks can be built from a set of universal surgemes. Those thoughts lead to the assembly of a motion library (*irob-saf*), containing a set of universal surgeme implementations.

To develop this motion library, a number of surgical subtasks had to be decomposed into a set of universal surgemes. For that purpose, several features and events were defined that separates subsequent surgemes from one another. A prime one is the overall shape of motion; this distinguishes for example the cutting from free navigation. Another important feature is the presence of tissue interaction during the surgeme; the instrument can move freely in the abdomen, it can grasp a loose piece of tissue, or even manipulate a tissue layer anchored to the anatomy. If the type of tissue interaction changes during the subtask execution, it will surely mean the transition to another surgeme. The final aspect of decomposition was the instrument required to be used during the procedure, e.g., a grasping surgeme might not be performed using scissors, and a cutting might not be done using grasping tools.

4.3 The Architectural Design of a Framework for Surgical Subtask Automation

The ROS platform—used widely in robotics—offers solutions to build modular, reusable software on a large scale. A ROS-based architecture consists of so-called *nodes*, intercommunicating with each other over channels of three types:

- Topic: continuous data streaming;
- Service: request–response type communication with blocking behavior, has benefits for e.g., requesting calculations;
- Action: request–response type communication with non-blocking behavior, useful for environmental interactions.

Due to its advantages, the `irob-saf` framework was fully built on ROS, and tailored for use alongside the DVRK. However, thanks to the implemented ROS interface, the framework remains easily portable to other platforms. The control scheme of the developed framework operates as follows. Sensors and perception algorithms, managed by ROS nodes, are used for the purpose of the measurement and estimation of the properties of the environment. The information gained, including errors eventually, are all channeled into the subtask-level logic node. This node is responsible for the processing of the information regarding the environment, and the commands originating from the surgeon. Additionally, the surgical workflow is coded in this node; its elements are translated into surgemes and sent to the surgeme server in the form of ROS actions. Propagating down from the surgeme server, the robot motion is generated by a hierarchical network of nodes, then sent to the robot (the ROS nodes from DVRK). It is important to note that, due to the principle of partial automation, continuous monitoring by the surgeon remains essential during the execution of the subtask (Fig. 4.2).

In the following, the details of the implemented framework's packages are presented. As an example, Fig. 4.3 illustrates a typical system built using the nodes of `irob-saf`, including interfaces to robots and sensors.

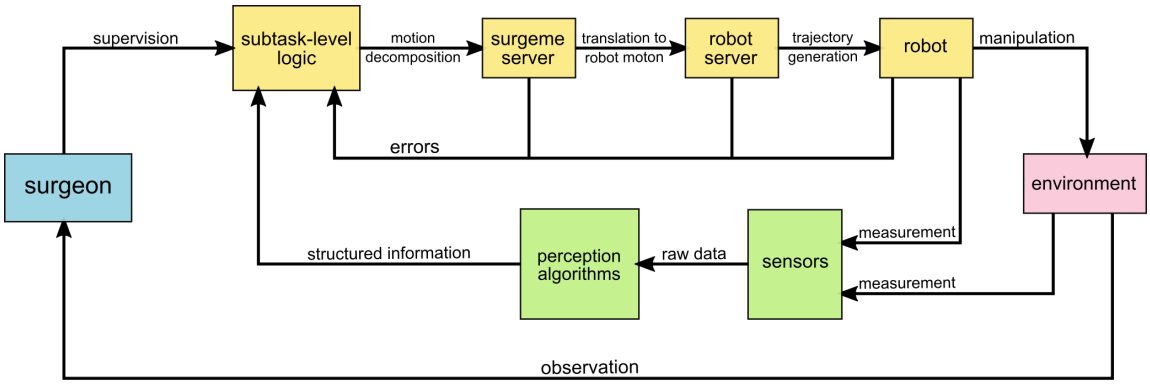

Fig. 4.2. The control scheme of partial automation offered by the framework. Perception nodes gather information from the environment. The subtasks-level logic contains the whole workflow of the subtask, processes the incoming information, and also communicates with the surgeon. This node also sends commands to the hierarchical set of nodes, appointing the surgemes to be executed. The generated motion is executed by the robot under permanent monitoring of the surgeon.

Camera image is one of the most important sources of information in the automation of RAMIS. The usage of the endoscopic camera image is undoubtedly the most obvious choice, since it does not require the placement of any additional instrument into the already crowded operating room. Nevertheless, in `irob-saf`, the video stream—preferably stereo—can be provided by a wide range of cameras as long as it is interfaced with a ROS topic. Example interfaces for USB webcameras and the stereo endoscope of the da Vinci are implemented in the framework. The calibration of the cameras—either mono or stereo—is performed by the built-in, easy-to-use camera calibration tool of the ROS environment, using a checkerboard pattern [77]. Furthermore, the basic stereo image processing algorithms, like disparity map calculation or the generation of the 3D point cloud are also performed with one of the built-in libraries of ROS [78].

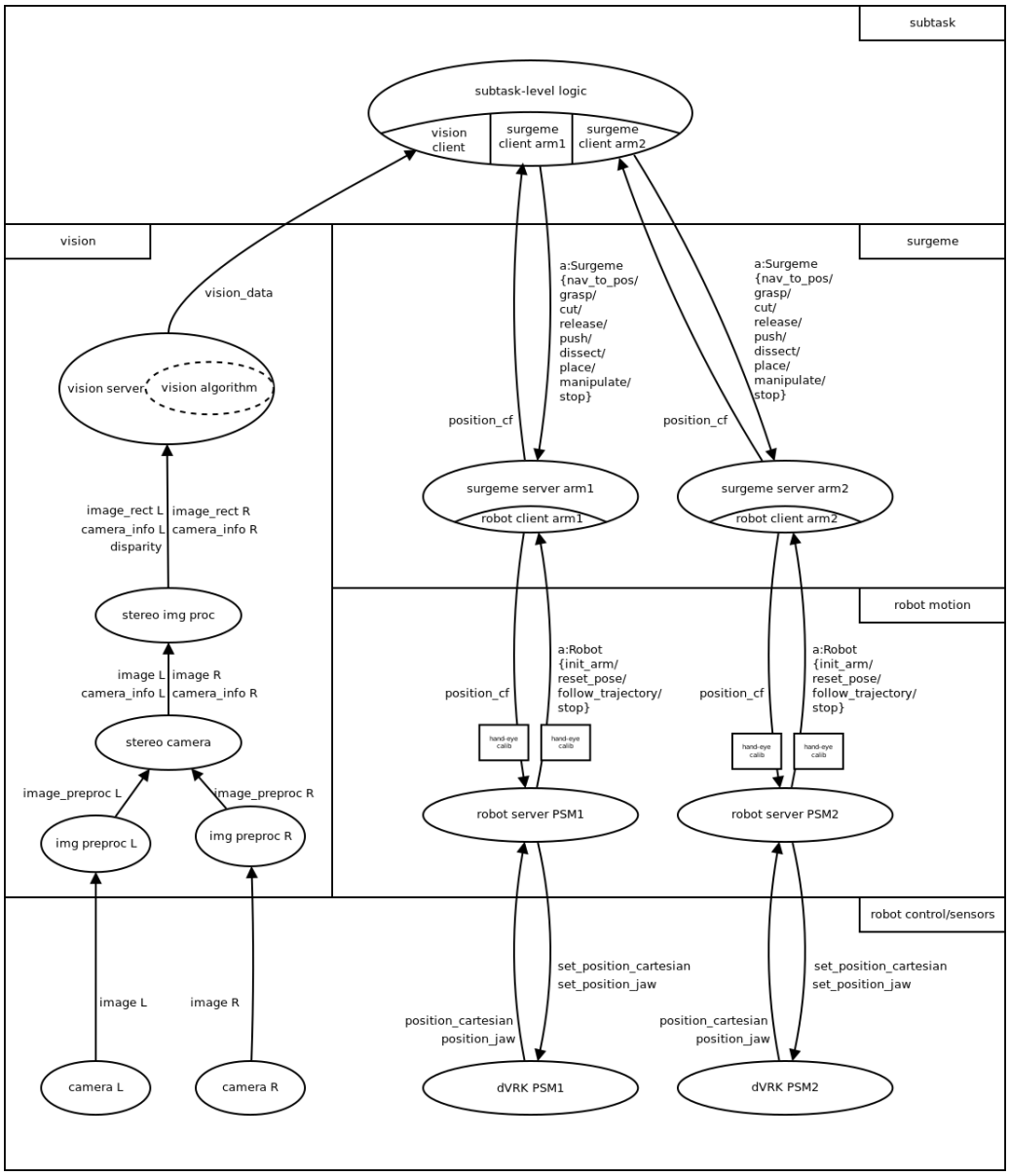

Fig. 4.3. An example of a system built of the ROS nodes offered by `irob-saf`.

The framework offers a pre-built infrastructure to run the algorithms usable for perception with the required input and output channels. These algorithms can be built using C++, Python, or even MATLAB. To ease development, the framework contains examples such as the detection of ChArUco markers [79]. As autonomous applications operate in increasingly realistic environments, approaching clinical translation, environmental perception becomes more critical. The implemented perception algorithms must handle dynamic soft tissues, enabling real-time anatomy recognition, tissue surface reconstruction, and even soft-tissue motion simulation [80]. Notably, additional sensor modalities, such as force sensors or RGB-D cameras, can be seamlessly integrated into the existing infrastructure.

The structured information from the perception nodes is finally being submitted to the subtask-level logic node, controlling the whole architecture. This node is subtask specific, an individual node needs to be implemented for each different surgical subtask. Here is where the information from the perception nodes is received and processed; all the errors and exceptions regarding the autonomous system, and user (surgeon) interactions are also channeled, and the surgeme level motion commands are generated. Subtask-level logic nodes are designed to contain and perform the specific workflow of the current subtask. The framework offers skeletons and also examples how to implement such nodes for the specific surgical subtask. At this level, behavior trees would offer a very structured representation of surgical knowledge and workflow [74], and it is planned to utilize this model in the future development of the framework.

Based on the processed information and the implemented workflow logic, the subtask-level logic node makes decisions on the execution of surgical actions—surgemes, and sends commands to the robot through hierarchic chains of nodes (one chain per arm). These chains communicate through ROS actions, making it possible for the higher-level nodes to do further work during action execution, e.g., monitoring the environment, or sending actions to other nodes. Moreover, actions provide the ability to send feedback and the result of the action, or preempt the action with another, if any environmental change makes it necessary, e.g., the location of the target changes or surrounding tissue moves during the execution.

The commands regarding the execution of surgical actions first reach the surgeme-level nodes. The surgical motion library, mentioned in Section 4.2, containing the implementation of universal surgemes, can be found in the package `irob_motion` of the framework. This surgeme library offers surgemes as parameterizable ROS actions, such as: *grasp*, *cut*, *place object*, *release object*, *navigate*, *dissect*, and *manipulate tissue*. The implemented surgemes are able to do the necessary safety checks, e.g., the proper instrument is used for the current surgeme. Further surgemes can be implemented based on the existing ones, and then added to the library.

Inside the surgeme-level nodes, the surgemes are translated into robot trajectories based on the pose of the Tool Center Point (TCP) and the angle between the jaws of the instrument. To ensure smooth robot movements, Linear Segments with Parabolic Blends (LSPB) trajectories are utilized [81]. The LSPB trajectory is characterized by constant velocity along a portion of the path, with acceleration and deceleration at the start and goal positions, resulting in a trapezoidal velocity profile (Fig. 4.4). In general, an LSPB trajectory is given by

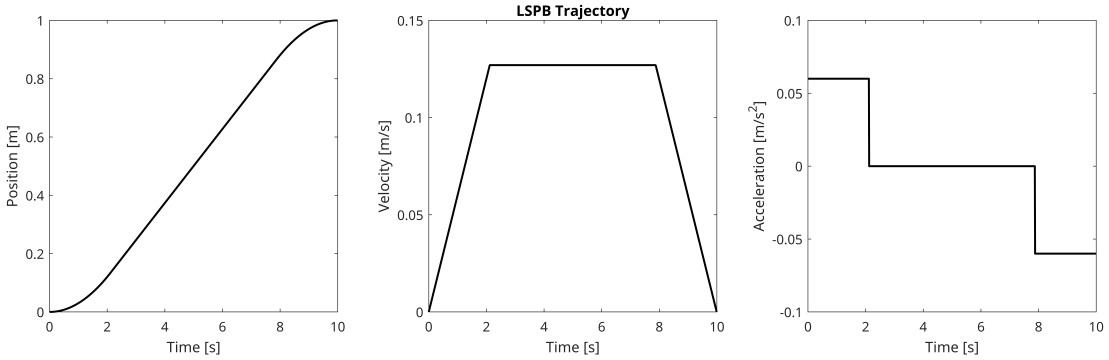

Fig. 4.4. Position, velocity, and acceleration trajectories during a Linear Segment with Parabolic Blends (LSPB) motion profile in one dimension, illustrating the smooth transition from acceleration to constant velocity and deceleration phases.

$$\mathbf{r}(t) = \begin{cases} \mathbf{r}_0 + \frac{\mathbf{a}}{2}t^2, & 0 \leq t < t_1 \\ \frac{\mathbf{r}_f + \mathbf{r}_0 - \mathbf{v_{max}} t_f}{2} + \mathbf{v_{max}} t, & t_1 \leq t < t_2 \\ \mathbf{r}_f - \frac{\mathbf{a} t_f^2}{2} + \mathbf{a} t_f t - \frac{\mathbf{a}}{2}t^2, & t_2 \leq t < t_f \end{cases} \tag{4.1}$$

where $\mathbf{r}$ represents the position, $\mathbf{r}_0$ is the initial position, $\mathbf{r}_f$ is the final position, $\mathbf{v}_{\text{max}}$ is the maximum velocity, $\mathbf{a}$ is the constant acceleration, t_1, t_2, and t_f represent the times corresponding to the transition points for acceleration, constant velocity, and deceleration, respectively. Also, to ensure smooth orientation interpolation between different poses, Spherical Linear Interpolation (SLERP) was employed [82]. SLERP is a technique used to interpolate between two unit quaternions (or rotation matrices), ensuring smooth and continuous rotational transitions. Given two quaternions $\mathbf{q}_0$ and $\mathbf{q}_1$, representing the initial and final orientations, the SLERP interpolation is defined as follows:

$$\mathbf{q}(t) = \frac{\sin((1-t)\theta)}{\sin(\theta)}\mathbf{q}_0 + \frac{\sin(t\theta)}{\sin(\theta)}\mathbf{q}_1 \tag{4.2}$$

where $\mathbf{q}(t)$ is the interpolated quaternion at time t, θ is the angle between $\mathbf{q}_0$ and $\mathbf{q}_1$, and $t \in [0, 1]$ is the interpolation parameter.

From the surgeme-level nodes, the commands—containing motion trajectories—propagate down to the high-level robot control nodes. The arms of the surgical robot are interfaced with the framework through high-level robot control nodes, one node per arm. These nodes are responsible for executing the trajectories generated by higher level nodes, while checking for errors originating from the robot.

The framework also offers solution for hand–eye calibration; namely the coordinate systems of the arms can be registered to the camera coordinate frame, that makes possible the generation of the robot motion relative to the camera. Visual markers attached to the instruments are used to estimate the tool positions based on the stereo or RGB-D camera stream. The hand–eye calibration can be performed using a Python script, that simultaneously records tool positions in the robot coordinate frame (received from DVRK through ROS) and in the camera coordinate frame (estimated using the visual markers), in

manually set positions. The hand-eye calibration process involves finding the optimal rigid transformation (rotation and translation) that aligns the robot coordinate system with the camera coordinate system [83, 84, 85]. This transformation is commonly represented as a rigid transformation matrix $\mathbf{T}_{\text{base,camera}}$ that consists of a rotation matrix and a translation vector. The basic equation for this transformation can be written as:

$$\mathbf{r}_{\text{base}} = \mathbf{T}_{\text{base,camera}} \cdot \mathbf{r}_{\text{camera}} \tag{4.3}$$

where $\mathbf{r}_{\text{base}}$ is the position vector in the robot coordinate frame and $\mathbf{r}_{\text{camera}}$ is the position vector in the camera coordinate frame; this equation holds for each pair of corresponding positions $\{\mathbf{r}_{\text{base}}, \mathbf{r}_{\text{camera}}\}$ recorded during the calibration process. The transformation matrix $\mathbf{T}_{\text{base,camera}}$ is computed by minimizing the sum of squared differences between the transformed robot positions and the measured camera positions using least-squares optimization. Once $\mathbf{T}_{\text{base,camera}}$ is determined, it is saved to a yaml file ("YAML Ain't Markup Language"), that is loaded by the corresponding high-level robot control node, which thus able to receive position commands in the camera coordinate frame from the higher-level nodes of the system.

These high-level robot control nodes are robot-specific, but their interface with the other framework nodes is universal. As a result, using a different robot arm only requires the implementation of the corresponding high-level control node, while the other framework nodes remain fully reusable.

4.4 Examples

The usage of the framework is explained through two examples on the automation of subtasks. It was decided to implement subtasks that require simpler perception methods; those algorithms are out of scope of the current work. The automation of a training exercise and an actual surgical exercise is presented in the followings.

4.4.1 Implementation of Autonomous Blunt Dissection

The first subtask example implemented using the framework was blunt dissection (Subsection 3.3.1). The development and testing of this algorithm was performed using a silicone phantom consisting of two harder layers of silicone connected with a softer, destructible silicone layer. This soft layer simulates the connective tissue, which can be penetrated and dissected with a blunt surgical tool. Naturally, the human abdominal space consists of much more complex tissue structures, with varying properties.

In our test environment (Fig. 4.5), two calibrated web cameras were utilized, with fixed focal length, attached onto a stable frame to provide the stereo image feed. The detection of the dissection profile relies on the depth map of the camera scene, calculated from the distance of each corresponding point pair on the rectified stereo pair.

The process presented in Fig. 4.6 is initiated by manually selecting a starting and an end point of the blunt dissection line. The precise dissection profile, where the dissection will be performed, is selected autonomously, by searching for the local minima of depth in the environment of the points of the manually selected dissection line (Fig. 4.6). The accuracy of the dissection line detection is further increased using Hampel filter to remove

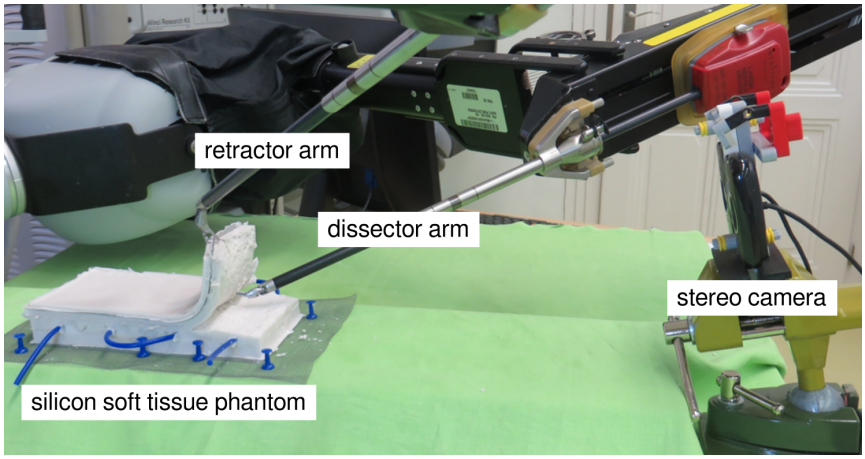

Fig. 4.5. The utilized automated blunt dissection test setup. The DVRK-enabled da Vinci Surgical System, a dissection phantom and a camera were involved.

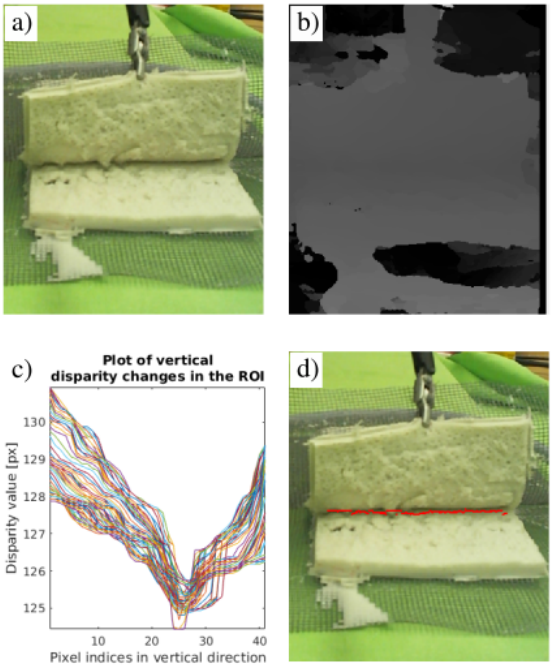

Fig. 4.6. Method for blunt dissection automation via computer vision. a) Image of blunt dissection phantom; b) disparity map of the field of view (greyscale represents the points' distances from the camera); c) plot of disparity changes in vertical direction; d) blunt dissection profile from the local minima of the disparity map.

outliers. To ensure to progress evenly inward between the tissues, the point with the lowest depth of the dissection profile is used for the location of the next dissection movement.

To estimate the depth in the field of view with a stereo system, it is crucial to calibrate the cameras. The stereo camera calibration was performed with 19 image pairs of a checkerboard pattern (with the checkerboard size being 25 x 25 mm). For every case, the pattern was fixed to a flat surface, as distortions in the pattern can greatly affect the calibration. To achieve better calibration accuracy, it is important for the checkerboard

pattern to be kept on an equal distance from the camera, within the expected field of interest. During the calibration, the pattern was placed at different orientations relative to the camera, and besides, the center points of the pattern were moved close to the frame edges as well to account for lens distortion. After the calibration the reprojection errors were calculated, which consisted of the error between the reprojected point in the camera and the detected point. MATLAB Stereo Camera Calibrator App calculated reprojection errors by projecting the checkerboard points from world coordinates (determined by the checkerboard) into image coordinates. The Camera Calibrator App then compared the reprojected points to the corresponding detected points. Reprojection errors are acceptable if they are closer than one pixel [86].

The success of this computer vision method depends on environmental factors such as light, noises, etc. To avoid complications caused by these factors, built-in functions are necessary. It may be important to know the earlier positions of the target objection and the dissection line. For this reason, a segmentation method was developed to detect the Region of Interest (ROI) on the image. This segmentation method is based on the depth of the start and end points of the dissection line; this way, if the surgeon chooses the right points, the ROI can be easily detected. The system keeps track of the last known position of the dissection line and searches for the corresponding local environment around it. Invalid disparity values are filtered to avoid inaccurate position coordinates.

As the subtask-level logic node receives the points of the dissection profile, so-called *dissect* surgemes are performed by the arm of the DVRK controlled da Vinci, consisting of the following primitives:

- **Dissect:**
  1. navigate to the point of dissection (Fig. 4.7a)
  2. slowly penetrate the tissue (Fig. 4.7b)
  3. open the jaws to separate layers (Fig. 4.7c)
  4. pull out the instrument in an open position (Fig. 4.7d)

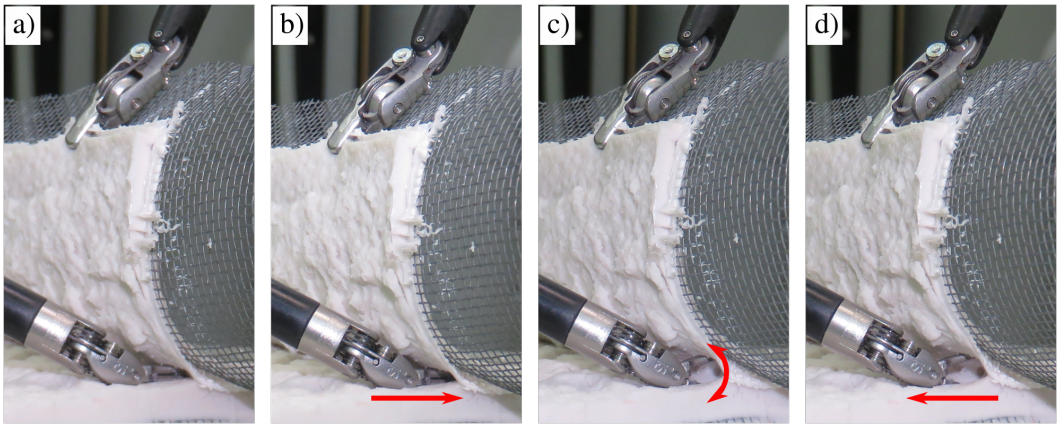

Fig. 4.7. Motion primitives of the surgical subtask automation. a) The surgical instrument (large needle driver) moves to the dissection target; b) the robot pushes the instrument into the phantom; c) the instrument is opened; d) the robot pulls out the instrument.

The system for autonomous blunt dissection is built using nodes from the `irob-saf` framework. It operates with a single arm, with computer vision implemented in MATLAB. The USB stereo camera pair is handled by the `irob_vision_support` package.

4.4.2 Implementation of Autonomous Peg Transfer

Another example for automation is a RAMIS training exercise, peg transfer (Subsection 3.3.2). This exercise is simple enough to present how an autonomous subtask execution can be built using my framework. In the followings, two versions of the implementation are shown, one with a custom peg transfer board and simple marker-based computer vision; while in the advanced version, the benchmarking environment by Hwang et al. [43] was utilized, and both the blocks and the board were detected markerless in RGB-D camera stream. Solutions on autonomous unilateral and bilateral handover variations of peg transfer were implemented of the peg transfer exercise both in the first end the second setup.

In the first version (Fig. 4.8), the position of the training board was estimated by the stereo camera stream of the built-in endoscope of the da Vinci. The video stream was captured by a DeckLink Blackmagic (Blackmagic Design Pty. Ltd., Port Melbourne, VIC) card, and forwarded to ROS using GStreamer [87]. The cameras were calibrated using the ROS built-in `camera_calibration` package. The board was marked by ArUco or ChArUco markers, that can be detected robustly by the camera, and can be used to estimate the board's position [79]. To start the nodes for computer vision, the launching of two launch files from the `irob_vision_support` package is necessary:

- `cam_blackmagic_raw.launch`: starts the node for streaming the camera image from one of the da Vinci's cameras
- `charuco_detector.launch`: for the pose estimation of the peg transfer board based on a ChArUco marker.

As it was mentioned in Section 2.3, since the 3D models are openly available, the pegboard designed by Hwang et al. is a perfect candidate for benchmarking autonomous peg transfer, as those can be printed by anyone. In the advanced version of the application,

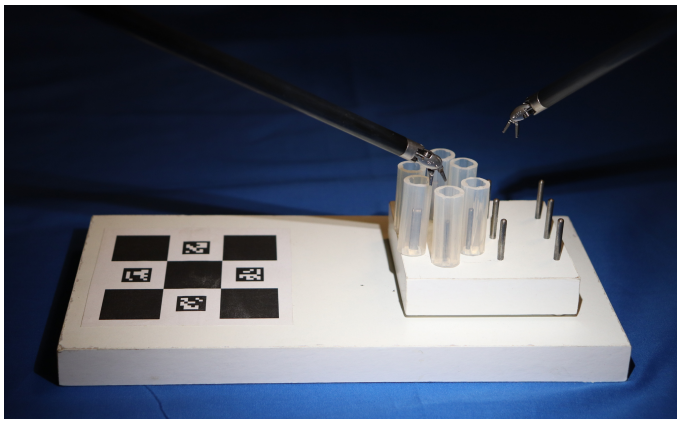

Fig. 4.8. The setup for the first version of the application performing the peg transfer exercise autonomously. The board is marked using a ChArUco marker for image-based pose estimation.

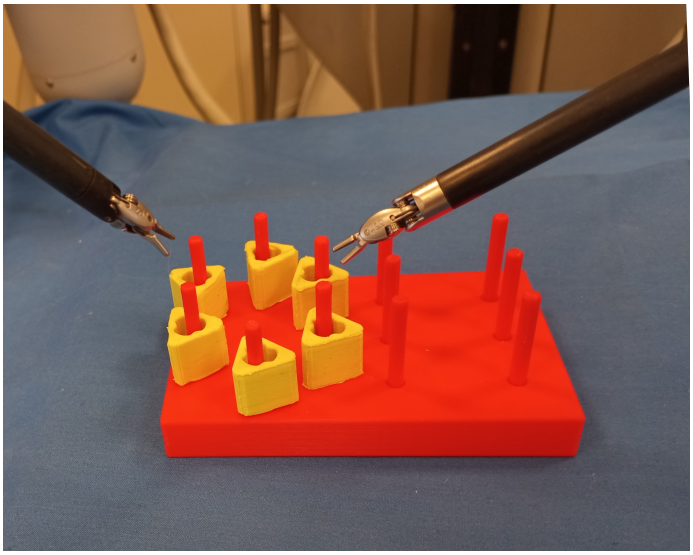

Fig. 4.9. The pegboard and blocks used in the advanced example. The board was fabricated using 3D printing, the blocks were moulded from silicone, to avoid snapping out of the jaws of the tools.

two modifications were made to this environment: the color and the material of the blocks (Fig. 4.9). Instead of printing those on a 3D printer, the blocks were moulded from Rubosil SR-20 silicone. For the moulding, inverse mould was designed using the original 3D model of the blocks, printable by 3D printer. The soft silicone blocks better represent a biological target, and unlike the rigid printed ones, are not prone to snapping out of the grippers. The silicone was also colored different than the board, to make detection and pose estimation easier.

The endoscopic stereo cameras used with the da Vinci Classic are typically more than 15 years old, those are now outdated; the image resolution and quality makes image processing quite challenging. Thus, a more up-to-date Intel RealSense D435i (Santa Clara, CA) RGB-D camera was utilized in the advanced version of the example.

The hand–eye registration for the surgical instruments was also performed by the 3D printed fiducial marker—graspable by the tool—designed by Hwang et al. [43] (Fig. 4.10). The colored spheres were segmented on the RGB-D camera stream by color, and their 3D position was estimated using the built-in triangulation function of the RealSense SDK. From the positions of the four spheres, the position of the instrument's TCP was calculated in the camera frame. At the same time, the TCP position was also received from the DVRK controller through ROS, in the coordinate frame of the robot. Moving the TCP to different positions within the field of the camera, the TCP position was collected in both frames in 15 instances. Those points then could be registered to get the transformation between the camera and the robot frame.

The first approach in developing the advanced version of the application was to estimate the positions of the blocks and the pegboard by fitting their known 3D models onto the 3D point cloud from the RealSense camera. Unfortunately, it was found that since the camera performs best with larger objects and at greater distances, the quality of the depth image—and consequently, the 3D point cloud—is insufficient for the scale of the blocks and the pegboard, rendering this method implausible. Thus, in the case of the blocks, I decided to use the 2D image as far as possible. The blocks were detected and their

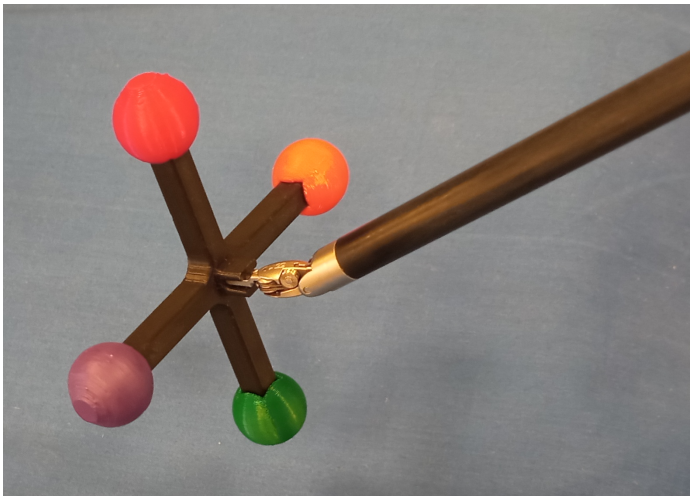

Fig. 4.10. The fiducial, graspable by the instrument, from the design of Hwang et al [43]. It is used to estimate the arms Tool Center Point (TCP) in the camera frame during hand-eye registration.

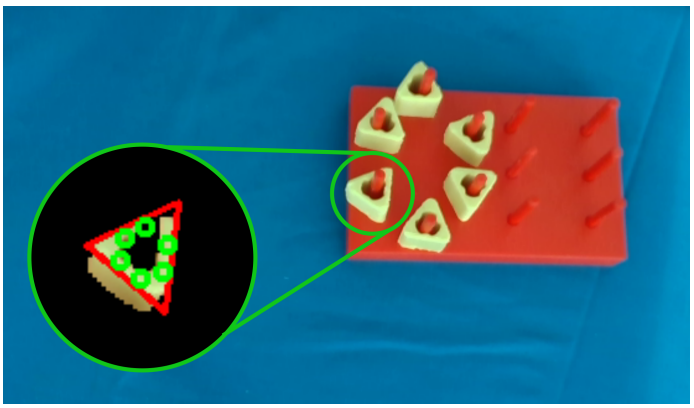

Fig. 4.11. The detection of the grasp locations of the blocks on the 2D image using color-based segmentation and edge detection.

positions were estimated using a traditional computer vision method from the *OpenCV* library [88]. The blocks were segmented by color, then the three outer edges were detected using Canny edge detector and lines were fitted using Hough transformation. Next, an affine transformation was calculated between these triangles and the known model, that also contained the grasp locations (Fig. 4.11). The 3D coordinates were only calculated using the mentioned triangulation algorithm of the RealSense SDK[5] after the grasp locations' image coordinates were calculated. Afterwards, these 3D coordinates were sent to the subtask-level logic node through ROS. The position of the pegboard was estimated by a similar method to the blocks, extended by RANSAC plane fitting applied on the 3D point cloud using the Open3D library [89], it was also forwarded to the top-level node of the system. The robustness and adaptiveness of object recognition and pose estimation—e.g., to variations in color and lighting conditions—could be enhanced by utilizing neural network models [90].

The nodes responsible for the generation and execution of surgical motion are oper-

[5] `https://www.intelrealsense.com/sdk-2/`

ating at 4 different levels of hierarchy. The uppermost level is the level of subtasks, with nodes of the `irob_subtask_logic` package. This level is built on a single node, that contains the workflow of the subtasks, receives the pose estimation of the peg transfer board, and chooses the surgemes for execution. The execution of surgemes is requested using ROS actions, that is sent to the proper node in the lower level. The second level of hierarchy contains the implementation of the universal surgemes. At this level, one surgeme server node is launched for each arm operating, receiving ROS actions from the subtask level, and sending ones to the lower, third level. This third level is responsible for the high level control of the arms, and consists of robot server nodes; one such node is responsible for the handling of one arm. These nodes accept ROS action commands for robot movements, and are also connected to the appropriate DVRK node at the fourth, lowermost level to execute the requested movements.

While the nodes of the three lower levels are universal for different subtasks, the uppermost, subtask-level logic node is unique. This node contains the workflow, basically a sequence of surgemes to execute, however, in the case of more complex subtasks, a state machine implementation can be useful as well. The motion—both in the case of one and two armed solutions—is composed of four surgemes: *grasp*, *navigate*, *place*, and *release* (Fig. 4.12). All surgemes of the framework including these four, are built of two motion primitives: spatial navigation of the instrument's endpoint, and the movement of the instrument's jaws. These motion primitives can be described well by only a few parameters, and based on the given parameters, the robot trajectories can be easily generated. These three surgemes are built up as follows:

- Grasp:
  1. navigate to approach position (waypoints can be added)
  2. navigate to grasp position
  3. close jaws
- Navigate:
  1. navigate to target position (waypoints can be added)
- Place:
  1. navigate to approach position (waypoints can be added)
  2. navigate to place position
- Release:
  1. open jaws
  2. navigate to leave position

The execution of these surgemes is requested by sending parameterized actions for to the surgeme server representing the chosen arm. The parameters of these surgeme action requests are calculated by the measured or estimated properties of the environment, received from the computer vision module. Such parameters can be the size of the object to grasp, the compression rate during grasping, or the approach and grasp positions of the instrument endpoint.

This hierarchy can be assembled by launching the following instances, in the case of bilateral handover execution:

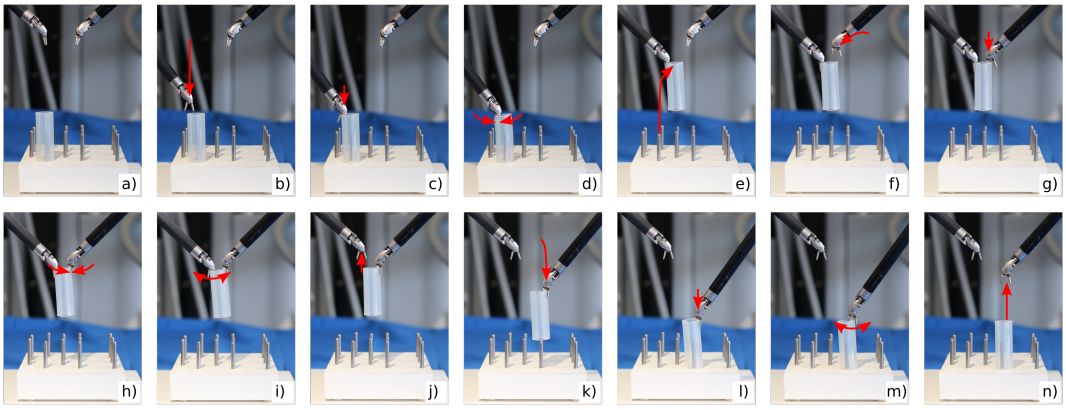

Fig. 4.12. The workflow used in the automation of bilateral handover variation of peg transfer. a) Setup before starting peg transfer. b–d) Left arm grasps the object. e) The object is lifted to the passing location. f–h) The object is grasped by the right arm. i–j) The object is released by the left arm. k–l) The object is placed on the target peg. m–n) The object is released by the right arm.

- `peg_transfer_dual.launch` from package `irob_subtask_logic`
- `surgeme_server.launch` from package `irob_motion`, in two instances, parameterized for each arms
- `dvrk_server` from package `irob_robot`, also in two instances, parameterized for each arms
- DVRK console, with the arms to be operated.

4.5 Summary of the Thesis

A methodology for the automation of surgical subtasks, based on the hierarchical decomposition of human surgical motions was proposed. Furthermore, an open-source, ROS-based software package was presented, which is based on the proposed methodology and able to ease surgical subtask automation research. This framework interfaces sensory inputs, perception algorithms and robots, and contains a surgeme-level motion library. The whole system can be controlled by a subtask-level logic ROS node, tailored to the needs of the current subtask to be automated. The iRob Surgical Automation Framework is available at `https://github.com/ABC-iRobotics/irob-saf`, and is being continuously developed and updated.

The framework can help in the implementation of further, more complex subtasks. In such development, it is straightforward to add new, necessary surgemes—like clipping or suturing. The implementation of new subtasks can be added to the motion library easily. Based on my experience, the most challenging aspect in automating more complex subtasks is the perception estimation of the environment, as computer vision usually struggles with light reflections or moving, deformable, and hardly recognizable tissue, even in phantom environment, or ex vivo.

The applicability of the framework was shown by the implementation of autonomous blunt dissection and an autonomous training exercise, the peg transfer, including the unilateral and the bilateral variations.

Chapter 5

ESTABLISHING STANDARD METHODS FOR THE EVALUATION OF AUTONOMOUS SURGICAL SUBTASKS

5.1 Introduction

As it was mentioned in Chapter 2.3, many research groups are working on partial automation in surgery currently. Despite how intensively researched surgical subtask automation is, there is no consensus on the choice of evaluation metrics on certain implementations yet, thus it is hard to compare those methods to each other or even the technique of human surgeons. Fontana et al. phrase the following on autonomous robotics: *"Within computer science, autonomous robotics takes the uneasy role of a discipline where the features of both systems (i.e., robots) and their operating environment (i.e., the physical world) conspire to make the application of the experimental scientific method most difficult."* [91]. According to their study, the difficulties caused by the large factor of uncertainties often lead to a methodological problem. Namely, it is practically challenging to perform accurate experimentation, thus methodological soundness often takes secondary role in robotic studies, detaining repeatability and reproducibility. The absence of those aspects forces even the best research works into the category of "proof of concept".

In this chapter, a contextual characterization model for surgical automation and evaluation metrics and techniques are presented, usable to compare autonomous surgical subtask execution to the performance of human surgeons, and to compare those autonomous methods to each other.

5.2 Characterization of Autonomy

Before the discussion of the evaluation metrics takes place, it is important to define the contextual classification for the automation of surgical tasks. The Autonomy Levels for Unmanned Systems (ALFUS) Ad Hoc Workgroup of National Institute of Standards and Technology paid a significant effort to define a framework for the characterization of Un-

manned Systems (UMS) from the aspect of autonomy. The resulted ALFUS Framework concerns a broad spectrum of UMS, including unmanned aerial, ground, maritime vehicles, and unattended ground sensors which are applicable in areas like military, manufacturing, search and rescue, or medical domains. Within the ALFUS Framework, a three-axis representation—the Contextual Autonomous Capability (CAC) model—was defined to characterize UMS from the perspectives of requirements, capabilities, and levels of difficulty, complexity, or sophistication. The individually established scores (1–10) along those three axes, namely *Human Independence*, *Environmental Complexity*, and *Mission Complexity* are used to give a standard and straightforward characterization of certain autonomous applications [19].

In this chapter's context, a specialized version of the CAC model, Surgical Contextual Autonomous Capability (SCAC), is introduced and customized to the domain of surgical robotics. SCAC extends the Level of Autonomy (LoA, Section 2.2) concept of RAMIS, presented in [7], offering a more detailed classification of autonomous surgical applications. The other foundation of the SCAC model is the Level of Clinical Realism (LoCR) scale for surgical automation, defined in [92] as:

- **LoCR 1** – *Training tasks with rigid phantoms;*
- **LoCR 2** – *Surgical tasks with simple phantoms;*
- **LoCR 3** – *Surgical tasks with realistic phantoms, but little or no soft-tissue interaction;*
- **LoCR 4** – *Surgical tasks with soft-tissue interaction;*
- **LoCR 5** – *Surgical tasks with soft-tissue topology changes.*

The LoCR concept (Fig. 5.1) can be interpreted as a composite scale, as it includes the complexity of both the environment and the surgical task. Thus, in this work it is decomposed to two individual scales: Level of Environmental Complexity (LoEC, Subsection 5.2.2) and Level of Task Complexity (LoTC, Subsection 5.2.3). LoA, LoEC, and LoTC are chosen to be the three aspects in the specialized, SCAC model (Fig. 5.2), matching the original concept of the ALFUS Framework. The SCAC model is formulated as a

Fig. 5.1. The Level of Clinical Realism (LoCR) scale for Robot-Assisted Minimally Invasive Surgery (RAMIS) with examples. *LoCR 1:* 3D printed board for the peg transfer training task, designed by Hwang et al. [43]; *LoCR 2:* Fundamentals of Robotic Surgery (FRS) training dome; *LoCR 3:* 3D printed bone phantom for drilling tasks; *LoCR 4:* anatomically relevant silicone pelvis phantom [93]; *LoCR 5:* in vivo human surgical environment [94].

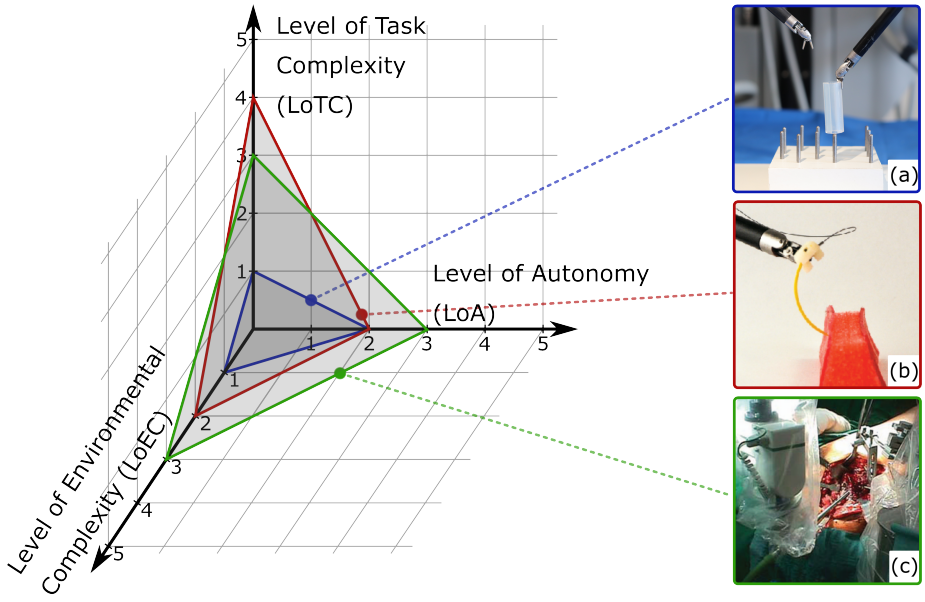

Fig. 5.2. The three-axis model of Surgical Contextual Autonomous Capability (SCAC). The x, y, and z axes represent the key characterizing aspects of autonomous surgical applications: Level of Autonomy (LoA), Level of Environmental Complexity (LoEC), and Level of Task Complexity (LoTC), respectively. The characterization of three example applications are illustrated: (**a**) *autonomous peg transfer surgical training exercise* [LT4] ($SCAC = F\{LoA = 2|LoEC = 1|LoTC = 1\}$); (**b**) *autonomous multi-throw multilateral surgical suturing* [11] ($SCAC = F\{LoA = 2|LoEC = 2|LoTC = 3\}$); (**c**) *autonomous bone drilling for total hip replacement surgery*, performed by the TSolution One system (THINK Surgical Inc., Fremont, California) [95] ($SCAC = F\{LoA = 3|LoEC = 3|LoTC = 3\}$).

function of the aforementioned key factors, thereby providing a structured approach to evaluate the capabilities of autonomous surgical systems. It is formally defined as follows:

$$SCAC = F\{LoA|LoEC|LoTC\} \tag{5.1}$$

The proposed SCAC model is able to characterize autonomous surgical applications from the perspectives of human independence and difficulty levels regarding the task and the environment. Although this thesis focuses on surgical subtask automation, my model concerns the whole domain of automation in surgery.

5.2.1 Level of Autonomy

Establishing objective conditions for autonomy has been a historical challenge for the robotics community [96]. First, the Degree of Autonomy (DoA) was introduced in ISO 8373:1994 *Robots and robotic devices — Vocabulary*, but was defined properly only decades later in *IEC/TR 60601-4-1: Medical electrical equipment – Part 4-1: Guidance and interpretation – Medical electrical equipment and medical electrical systems employing a degree of autonomy* as *"taxonomy based on the properties and capabilities of the medical electrical equipment or medical electrical system related to autonomy". IEC/TR 60601-4-1:2017* recommends the parameterization of DoA along four cognition related functions of a system, affecting options of a medical electrical system: *Generate* an option; *Execute* an option; *Monitor* an option; and *Select* an option [97]. The LoA concept

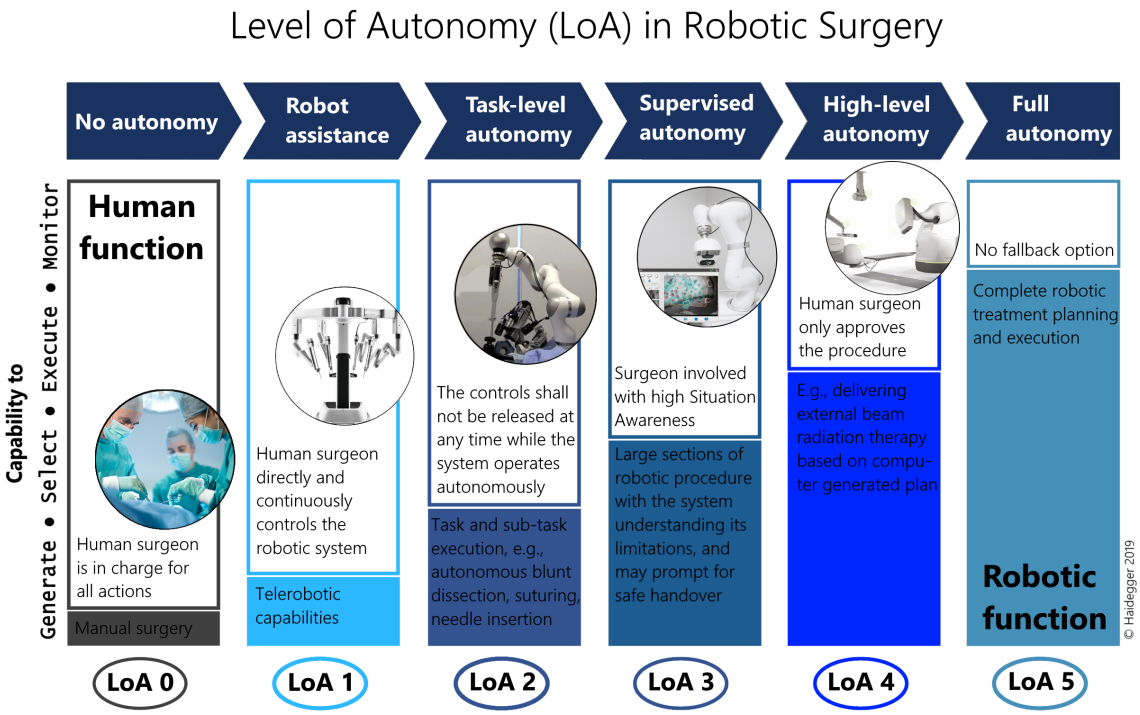

Fig. 5.3. The Level of Autonomy (LoA) concept for surgical robotics proposed by Haidegger [7] with a scale 0–5; from no autonomy to full autonomy.

of RAMIS—originating from the field of autonomous vehicles [LTNR6]—was proposed in [7], modified from [24], simplifying DoA to offer a taxonomy to generally assess the development phases of surgical robotics (Fig. 5.3). The proposed 6-grade scale is coherent to the mainstream standardization efforts, and defined as the following:

- **LoA 0** – *No autonomy*;
- **LoA 1** – *Robot assistance*;
- **LoA 2** – *Task-level autonomy*;
- **LoA 3** – *Supervised autonomy*;
- **LoA 4** – *High-level autonomy*;
- **LoA 5** – *Full autonomy*.

5.2.2 Level of Environmental Complexity

The common surgical environment can be described more accurately than the broad range of areas included in the ALFUS Framework. The proposed scale reading as follows:

- **LoEC 1** – *Training phantoms:* made for the training of surgical skills (e.g., hand–eye coordination), no or limited, highly abstract representation of the surgical environment, e.g., peg transfer;
- **LoEC 2** – *Simple surgical phantoms:* made for certain surgical subtasks, modeling one or few related key features of the real environment, e.g., silicone phantom for pattern cutting;

- **LoEC 3** – *Rigid, realistic surgical environment:* realistic surgical phantoms or ex/in vivo tissues/organs, little or no soft-tissue interaction, e.g., ex vivo bone for orthopedic procedures;
- **LoEC 4** – *Soft, realistic surgical environment:* realistic surgical phantoms or ex/in vivo tissues/organs, soft-tissue interaction, e.g., anatomically accurate phantoms for certain procedures or ex vivo environment;
- **LoEC 5** – *Dynamic, realistic surgical environment:* realistic surgical phantoms or ex/in vivo tissues/organs, soft-tissue topology changes, e.g., in vivo environment with all relevant physiological motions.

5.2.3 Level of Task Complexity

The LoTC represents the Mission Complexity from the ALFUS Framework in the surgical domain. Two components of complexity were compiled into the proposed scale: is it a training or an actual surgical task, and what are the Situation Awareness (SA) requirements of the execution. SA is defined on 3 levels based on the cognitive understanding of the (past–present–future) environment, and can be categorized into the following classes: spatial (locations), identity (salient objects), temporal, goal, and system awareness [98, 99]:

- **Level 1 SA** – *perception of the environment;*
- **Level 2 SA** – *comprehension of the current situation;*
- **Level 3 SA** – *projection of future status.*

Based on the mentioned considerations, the following LoTC scale is proposed:

- **LoTC 1** – *Simple training tasks:* no or limited, distant representation of surgical task, no or Level 1 SA required, e.g., peg transfer;
- **LoTC 2** – *Advanced training tasks:* no or distant representation of surgical task, basic reasoning and Level 2 or 3 SA required, e.g., peg transfer with swapping rings;
- **LoTC 3** – *Simple surgical tasks:* no or Level 1 SA required, e.g., debridement;
- **LoTC 4** – *Advanced surgical tasks:* Level 2 SA, spatial knowledge and understanding of the scene required, e.g., suturing;
- **LoTC 5** – *Complex surgical tasks:* Level 3 SA, clinical and anatomical knowledge required, e.g., stop acute bleeding.

5.3 Performance Metrics

For the purpose of validation of an autonomous surgical subtask, the most obvious idea would be to compare it with the task execution of human surgeons—even with varying skill levels. It is also viable to compare the performance of the autonomous system to another, already validated one. For both methods, i.e., human–machine and machine–machine comparison, the proper choice of performance metrics that describes how well the subtask is executed is crucial [100].

Since the research domain of surgical subtask automation is still in its infancy, no standard set of metrics has been established yet. In the case of surgical skill assessment,

the standard practice is to appoint a ground truth for that skill level, e.g., the *number of surgeries performed*, *years of practice* or manual scoring by an expert surgeon. Then, the metrics measured in the experimental study are correlated to the ground truth to find out which metrics are best related the surgical skill [100, 101]. To extend this methodology of finding the metrics best represent the quality of autonomous surgical subtask execution, there are three possibilities to correlate the metrics of the autonomous execution:

a) to the ground truth utilized in the case of human surgeons;
b) to the metrics from human execution that are found to be correlated with surgical skill;
c) to new ground truth for autonomous execution.

All of these options are found to be quite problematic: in the case of a), the ground truth metrics (e.g., *years of practice*) could not be interpreted for autonomous surgery; in the case of b) relies on the statement, that the same metrics represent the quality of both human and autonomous execution—which is not only not proved, but it is easy to see the opposite e.g., for metrics like *distance traveled* or *jerk*. The only viable solution appears to be option c) correlate to the new ground truth for autonomous execution. Unfortunately, there is nothing analogous to the population of human surgeons for autonomous agents to conduct such a study.

Although, it is not possible to experimentally prove which metrics represent the quality of an autonomous system or the clinical outcome best, a set of metrics can still be recommended based on their various properties. In the following subsections, the area of MIS skill assessment is reviewed briefly in terms of performance metrics, then the candidates for standard performance metrics will be presented, organized by modality. Those metrics are scored and evaluated along different aspects, and finally a recommendation on standard validation metrics is proposed. Also, it is important to note that choosing the fitting metrics to evaluate the performance of the given application greatly depends on the system's SCAC, the relationship of the metrics to choose and SCAC is also discussed.

5.3.1 Performance Metrics in Surgical Skill Assessment

The principle of RAMIS subtask automation is to take inspiration from human surgical actions as a reference for skillful execution. Thus, the search for metrics to evaluate the performance of an autonomous surgical application should start in the area of MIS skill assessment. Finding the metrics that best correlate with the surgical skill is far from trivial, and already has an extensive literature [100, 101]. For example, one could think that *mortality rate* after the surgery would somehow relate to the surgical skill; the better the surgeon the lower the *mortality rate* would be. However, there are a number of different factors contributing to this metric. For instance, a beginner surgeon may not undertake the operation of patients with poor health condition, and does low-risk surgeries instead, resulting in low *mortality rate*. In contrast, an expert surgeon may be more willing to undertake high-risk interventions, but that could result in higher *mortality rate* despite how well the interventions are performed [102].

One of the most widely used standard surgical skill assessment techniques is the Global Evaluative Assessment of Robotic Skills in Endoscopy (GEARS-E) [103], in which the

following aspects of execution are scored: *depth perception*, *bimanual dexterity*, *efficiency*, *tissue handling*, *autonomy*, and *endoscope control*. A quite similar, and also prevalent technique is the Robotic Objective Structured Assessments of Technical Skills (R-OSATS) [104], scoring *depth perception/accuracy of movements*, *force/tissue handling*, *dexterity*, and *efficiency of movements*. Both methods utilize manual scoring 1 to 5 using the Likert scale [105], thus being subjective, making it hard to be used in the validation of autonomous applications. Raison et al. [106] compiled their surgical simulation training study using a custom score set divided to general scores: *time to complete*, *economy of motion*, *master working space*, *instruments out of view*, *excessive force*, *instrument collision*, and task specific scores: *blood loss*, *broken vessels*, *misapplied energy time*, and *missed target*, *dropped instrument*.

The above mentioned metrics are all representing the technical skills of surgery. However, the outcome of the surgical procedure is also dependent on the non-technical skills of the surgeon, and those cannot be interpreted in the case of an autonomous application [100, 107]. Those skills are typically rated using a questionnaire filled by the subject, like the NASA Task Load Index (NASA-TLX) [108], evaluating *mental demand*; *physical demand*; *temporal demand*; *performance*; *effort*; and *frustration*. Non-technical skills are not discussed in more details due to the lack of their usefulness in automation, unless an adverse event occurs.

5.3.2 Metrics by Modality

In the followings, various performance metrics are overviewed, and their usability is discussed. Those metrics are collected from the literature of surgical subtask automation, and also from the fields of autonomous robotics, autonomous vehicles, and surgical skill assessment, and are organized into subsections by the modality of the measured values [100, 101, 109].

Temporal Metrics

Completion time is one of the most commonly used metric both in surgical skill assessment [100, 101] and in surgical subtask automation [10, 32, 40, 43, 49, 110, 111]. It characterizes the whole execution of the subtask, and could be defined in a number of ways, e.g., in the case of peg transfer, *completion time* can be taken as the average of each individual transfer, or as the average for whole subtask executions. For humans, temporal metrics tend to correlate with skill level, or give a measure of hesitation. In the case of automation, those connect more loosely to the quality of the execution. However, as the lower time requirement of surgical interventions is beneficial—for the patient, surgeon, or the hospital—, near-human, or even superhuman *completion time* is still an important factor in automation. For instance, Sen et al. [11] measured *completion time* for autonomous suturing and compared it to manual executions from the JHU—ISI Gesture and Skill Assessment Working Set (JIGSAWS) database [53]. Also, Ginesi et al. [41] validated their autonomous task planning algorithm for peg transfer by measuring the *task planning time* in different scenarios.

Temporal metrics could also be used to evaluate the elements of the whole system. *Time to compute* is one of the most important metrics in the field of computer science,

and could describe e.g., perception algorithms, trajectory generation, or planning in autonomous systems [41]. *Completion time* and *time to compute* become extremely useful, when working on benchmarks, then it can be quite a strong factor of comparison between different solutions, offering inter-comparability for different research groups.

Time is also critical in the reaction to adverse events. Current research efforts in the area of RAMIS typically targets LoA 2, where the surgeons' supervision is essential, as they need to recognize and solve adverse events. The *reaction time* of autonomous surgical systems becomes important—if not crucial—at LoA 3+, where the autonomous system have to recognize and react to unexpected events either by solving the emergency autonomously, or by sending a handover request to the human surgeon.

Outcome Metrics

Outcome metrics assess the end result of the whole procedure—or in this case subtask—or its elements individually, ignoring the way it is performed completely. Such metrics are e.g., *number of errors*, *quality of the outcome*, and *success rate*. *Success rate* is probably the most universal and easy to measure, thus utilized frequently in surgical subtask automation [10, 32, 39, 40, 43, 46, 110, 111]. McKinley et al. [39] evaluated autonomous tumor resection by end-to-end *success rate*, while Hwang et al. [43] defined it for the peg transfer training exercise as the percentile value of the ratio of $Success/Attempts$ for each individual transfer. Attanasio et al. [13] utilized *visible area* as outcome metric in autonomous retraction. Nguyen et al. [52] measured the accuracy of pattern cutting next to autonomous tensioning. In the study of Shademan et al. [8] autonomous end-to-end anastomosis is presented and validated in vivo on porcine, where *number of sutures*, *number of suturing mistakes*, *leak pressure*, *luminal diameter reduction*, *weight at surgery*, and *weight at sacrifice* were measured and compared to manual execution.

It is important to note that outcome metrics are highly task-specific, thus, making any comparison between different subtasks is very difficult. On the other hand, the implementation of the mentioned metrics requires less effort than most of the others.

Motion-based Metrics

Motion-based metrics utilize the position (and sometimes orientation) of the surgical instrument, the surgeon's hands, or other tracked object, as a trajectory or motion profile [100, 106, 112, 113, 114]. Some of the simplest ones are *distance traveled*, *economy of motion*, and *number of movements*, but it is also common to use hidden Markov models or other machine learning algorithms to compare the movement patterns to expert surgeons. Those metrics offer an objective way to assess the skill of human surgeons, hence, those give a measure of hesitation, dexterity, and motor skills. However, their usefulness is limited in the field of surgical subtask automation. For example, *distance traveled* could be decreased or increased programmatically without major effects on the quality of execution. Additionally, for the more advanced metrics, comparing the motion pattern to experts could also be misleading in the case of autonomy; the motion patterns of human experts contain the restraints and characteristics of human anatomy, but an autonomous robot could possibly execute the same task through different—even more beneficial or optimized—trajectories.

Velocity and Acceleration Metrics

Velocity and acceleration metrics are calculated as the first and second derivatives of the motion profiles mentioned regarding motion-based metrics [113, 115]. Those are widespread in surgical skill assessment, to name some without being exhaustive: *peak speed*, *normalized speed*, *number of changes in velocity over time*, *number of accelerations and decelerations*, *mean acceleration*, or the *integral of the acceleration vector*. Those are related to similar traits as the motion-based metrics, and thus their usefulness in automation is questionable.

Jerk Metrics

Jerk is a metric that derives from the third derivative of the motion profile. As it has a quite broad literature from surgical skill assessment [113, 116, 109] to the diagnostics of neurodegenerative diseases [117, 118], it deserves to be mentioned separately. Jerk metrics also tell us about the motor skills; motion patterns are usually become smoother by practice. However, its usage in automation is quite insignificant, since it is a highly human-specific metric.

Force-based Metrics

The amount of force applied to the tissues is a significant characteristics of surgery; it is important not to cause damage by excessive force, but it is also crucial to provide enough tension e.g., during tightening a knot [119][LTNR9]. Also, Trejos et al. [109] have shown in their experimental study that force-based metrics correlate better with experience levels than temporal metrics in manual MIS. Unfortunately, sensorizing MIS or RAMIS instruments is still very challenging, due to the small dimensions or sterilization requirements, and thus, the usage of force-based metrics is not prevalent in skill assessment or in the evaluation of automation. However, currently, there are a number of solutions to measure or estimate forces and torques on the shaft or the tip of the instruments [120, 121], the phantoms, or even to measure the applied grasping force [122, 123]. Using such devices, the metrics of the force utilized, e.g., *grasp maximum*, *grasp integral*, *Cartesian maximum*, or *Cartesian integral* could be included to the validation of surgical subtask automation. Such an example can be seen in the work of Osa et al. [124], where the applied force was measured during autonomous thread tightening.

Accuracy Metrics

Accuracy is a minor problem in surgical skill assessment, but it is extremely important in many areas of automation. Accuracy metrics usually characterize one or few subcomponents or aspects of the autonomous system, such as the *positioning accuracy* of the utilized robots and low-level controllers, *accuracy of hand–eye registration*, *pose estimation*, or *object detection*. It is the current best practice to measure the accuracy of the system's components in surgical subtask automation in order to validate the application. For instance, Lu et al. [125] validated their knot-tying application by measuring the *tracking error* on the instruments. Also, Lu et al. [111] measured the *tracking error* of the grasping point in their study on autonomous suture grasping. Sen et al. [11] measured the *accuracy*

of needle detection in their study on autonomous suturing. Elek et al. [LT1] measured *depth error of the camera*, *positioning error*, and *accuracy of dissection profile extraction* among automation metrics in their autonomous blunt dissection study. Besides validation, accuracy metrics could also help localizing problems during the implementation phase.

Although, measuring metrics such as the *robot positioning accuracy*, the *accuracy of hand–eye registration*, the *instrument tracking error*, or the *accuracy of pose estimation* can be highly beneficial, especially during the implementation phase, these properties and errors all contribute to the *end-to-end positioning accuracy*. Thus, for the purpose of validation, these measurements could be substituted by testing and measuring *end-to-end positioning accuracy*, or *application accuracy*, that can even tell more about the accuracy of the whole system, than the accuracy of its components [126].

The *application accuracy* could be measured most precisely by incorporating a high precision external system, such as industrial robots or tracking systems (visual or electromagnetic). While this method is highly recommended, in some cases this is not a viable option due to the high cost of such devices, and low fidelity methods need to be utilized. Pedram et al. [127] measured the *error of pose estimation* of the needle, then used this pose estimation to measure the application accuracy of the needle in their study on autonomous suturing. Seita et al. [10] used the following method to fine-tune the end-to-end positioning of their system for autonomous debridement: the tool was sent to points on some kind of grid, using positions from their computer vision solution, then the tool position was adjusted manually to the desired position on the grid, and from the kinematic data of the robot, the *application accuracy* was calculated. As this method, especially by utilizing a grid with physical constraints to help accurate manual positioning (e.g., holes or dips for the tool endpoint) offers a simple yet accurate method for the measurement of the *application accuracy*. The utilization of such methods is also viable for the validation of autonomous systems in RAMIS, but only recommended if the errors of the ground truth system are known and error propagation is taken into account.

5.3.3 Conclusions on Performance Metrics

The main scope of this chapter is to introduce a criteria set for the evaluation and validation of subtask-level automation in RAMIS. The most characteristic and meaningful metrics are compiled into Table 5.1, and scored for usability in the targeted area of the assessment of systems at LoA 2 in three aspects: *Task Independency*, *Relevance with Quality*, and *Clinical Importance*. The metrics with the best overall scores are highlighted with dark gray, those are highly recommended for the validation of autonomous subtask execution in RAMIS. The metrics highlighted with light gray were found to be moderately useful, the utilization of those could be considered in a number of subtasks as well.

Based on the overall scores, outcome and accuracy metrics (*accuracy of object detection*, *accuracy of pose estimation*, and *application accuracy*) were found to be the best amongst all for the validation of autonomous surgical systems, in general. Important to note that accuracy metrics perform slightly better in task independency, since, unlike outcome metrics, those offer inter-comparability between different tasks. Temporal and force-based metrics also received good overall scores, with *reaction time* and *Cartesian force* in the "highly recommended" category. Although the measurement of force-based metrics requires additional sensors, unlike other metrics, those metrics tell a lot about the

TABLE 5.1

CHARACTERISTIC PERFORMANCE EVALUATION AND COMPARISON METRICS FOR RAMIS SUBTASK AUTOMATION, GROUPED BY MODALITY. THE METRICS ARE SCORED IN THREE ASPECTS ON A SCALE 1–3: *Task Independency* (1–USABLE ONLY FOR SPECIFIC TASKS; 2–CAN BE USED FOR ANY TASK, BUT NOT INTER-COMPARABLE; 3–CAN BE USED FOR ANY TASK, DIFFERENT TASKS ARE COMPARABLE); *Relevance with Quality* OF TASK EXECUTION (1–IRRELEVANT; 2–RELEVANT, BUT MAY NOT CORRELATES WITH QUALITY; 3–RELEVANT AND CORRELATES WITH THE QUALITY OF TASK COMPLETION); AND *Clinical Importance* (1–NOT IMPORTANT; 2–MODERATELY IMPORTANT; 3–VERY IMPORTANT). THE SCORES ARE SUMMED IN THE LAST COLUMN FOR EACH METRIC AND THE ONES WITH THE BEST SCORES ARE HIGHLIGHTED WITH LIGHT GRAY (6–7) AND DARK GRAY (8–9).

Modality	**Metric**	**Task Independency**	**Relevance with Quality**	**Clinical Importance**	**Overall Score**
Temporal	Completion Time	2	2	2	6
	Time to Compute	2	2	2	6
	Reaction Time	3	2	3	8
Outcome	Rate of Errors	2	3	3	8
	Quality of the Outcome	2	3	3	8
	Success Rate	2	3	3	8
Motion-based	Distance Traveled	2	2	1	5
	Economy of Motion	2	2	1	5
	Number of Movements	2	2	1	5
Vel. and Acc.	Peak Speed	2	1	1	4
	Number of Accelerations	2	1	1	4
	Mean Acceleration	2	1	1	4
Jerk	Jerk	3	1	1	5
Force-based	Grasp Force	1	3	3	7
	Cartesian Force	2	3	3	8
Accuracy	Acc. of Pose Estimation	3	3	3	9
	Acc. of Object Detection	3	3	3	9
	Application Accuracy	2	3	3	8

quality of execution and also the utilized force is very important clinically. Motion-based, velocity, acceleration, and jerk metrics received relatively low scores, the usability for the performance assessment of autonomous surgical subtasks were found questionable.

The choice of metrics greatly depends on the certain subtask and the experimental setup. Thus, a flowchart is supplied that can be used to compile a list of metrics for the validation of the given autonomous surgical application (Fig. 5.4). The list of metrics

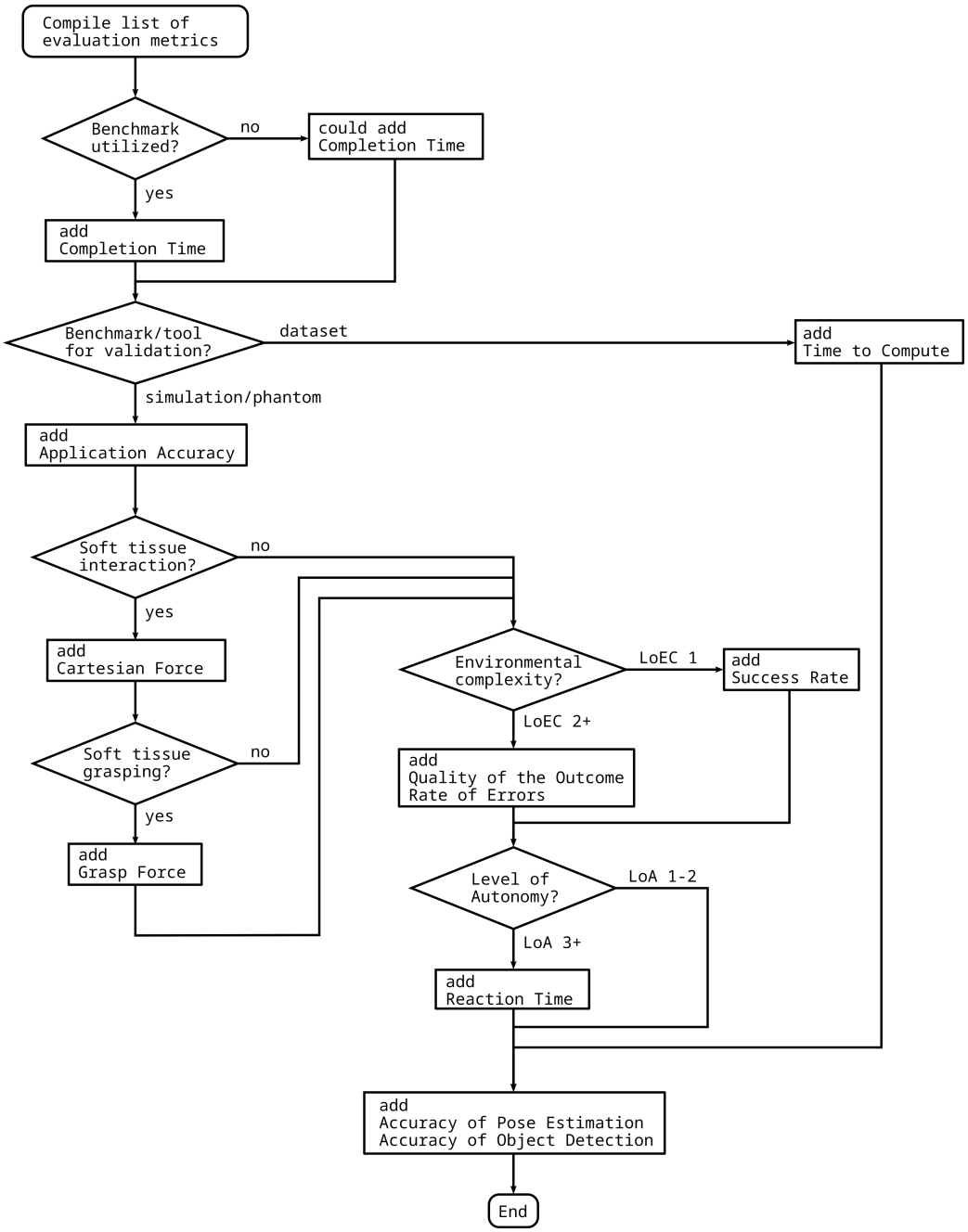

Fig. 5.4. Flowchart to compile a list of performance metrics for the validation of different autonomous applications. The proposed method requires different properties of the autonomous surgical application, the task, and the validation environment as input, and outputs the list of recommended validation metrics.

depends on, without being exhaustive, the type of benchmarking environment, the handing of soft tissues or the LoEC and LoA of the system. For example, the validation of an instrument pose estimation algorithm for surgical videos using a benchmark dataset requires *completion time*, *accuracy of object detection*, and *accuracy of pose estimation*, while the validation of a system performing autonomous peg transfer should be validated

using *completion time*, *application accuracy*, *success rate*, *accuracy of pose estimation*, and *accuracy of object detection*. Important to note that *accuracy of pose estimation* and *accuracy of object detection* are suggested to be added to the list in any case, since these can be used for the validation of almost any application in the field of surgical subtask automation, and also received the highest overall scores.

5.4 Benchmarking Techniques

The recommended metrics could make possible the comparison of different autonomous applications. Additionally, testing and evaluating autonomous applications in a benchmark environment—a standardized problem or test that serves as a basis for evaluation or comparison—offers an even more solid basis for both human–machine and machine–machine comparison [128]. Currently, the usage of benchmarking techniques is not prevalent in the field of surgical subtask automation, still it is used intensively in the area of autonomous driving. The development of autonomous vehicles is considered to be analogous to surgical subtask automation due to the high complexity of the environment and the presence of potentially life-threatening risks [LT9].

As the development of autonomous subtask execution in surgery progresses, and gets closer to clinical use, the need for evidence on the autonomous system's capabilities will rise. In [129], Fiorini states the followings: *"Benchmarks will be the key to build a solid confidence on the robots' capabilities to understand and react to clinical situations as a human surgeon would do, and the process of medical certification for surgical robots will need to be developed."* Next to low LoEC it is relatively convenient to create benchmarks, but unfortunately, as LoEC increases, meaning realistic or even dynamic environments, the development of benchmarks will be considerably more difficult.

In the research of self-driving, benchmarking techniques can be compiled into three categories: benchmark datasets to test system components like object detection accuracy [130, 131, 132, 133]; standard simulated environments like the scenarios developed for the CARLA Car Simulator [134, 135]; and physical test towns like mCity [136].

Autonomous surgical subtasks could also be evaluated using those types of benchmarks. Benchmark datasets for instrument segmentation are already available within the yearly MICCAI Challenges, including ground truth for training and evaluation of accuracy [137, 138]. Among the surgical simulators, the Asynchronous Multi-Body Framework (AMBF) [139] is worth to emphasize, being open source, offers da Vinci and Raven robot arms, and supports the simulation of deformable objects. Moreover, AMBF contains a built-in peg transfer board that could already serve as benchmark. Important to note the *2021 AccelNet Surgical Robotics Challenge*, also based on the AMBF simulator, offering a simulated setup to develop autonomous suturing applications [140].

Benchmarking physical setups presents a bigger challenge to implement [141]. Thanks to 3D printing technology it is possible to create and distribute some standard physical benchmarks, such as training phantoms, or rigid surgical phantoms. The peg transfer board (Fig. 2.3b) presented in [43], whose design files are freely available, is a perfect candidate for a standard physical benchmark. By defining a standard execution of the exercise and standard evaluation metrics a simple benchmark could be compiled to measure and compare the performance of different autonomous algorithms.

The utilization and spread of benchmarks alongside the standard performance metrics in RAMIS subtask automation would potentially make performance of different autonomous systems much more comparable. Since a number of challenges offer benchmark datasets, even standard simulated environments, the authors highly recommend those utilizations. In terms of physical benchmarks, 3D printing technology makes the sharing and reproducing of phantom environments and other objects easier and cheaper than ever [93], and hopefully the practice of using 3D printable surgical phantoms will soon spread in the research community.

5.5 Human–Machine Interface Quality

The quality of the HMI has an important role in clinical usability of autonomous surgical systems, especially between LoA 1 and LoA 4, where the human surgeon and the autonomous system perform the surgical intervention together. In the field of RAMIS subtask automation, it is quite uncommon to test and validate the HMI of the autonomous system. However, in related research fields, such as self-driving vehicles or image guided surgery, the HMI is validated much more frequently [142, 143]. Usually, the quality of the HMI is assessed using the performance metrics seen in Section 5.3, like *time to complete*, or *success rate*; the aim of those tests is to assess the performance of the human surgeon and the autonomous system together.

Another important aspect of HMI quality is the system's handover capability, especially at LoA 2–4 systems. During a handover, as it is described in Subsection 5.3.2, time is the most crucial factor. First, the autonomous system should recognize the adverse event and have to initialize a handover request to the surgeon as soon as possible, then the autonomous system has to yield the control safely. Secondly, in the case of surgeon initiated handovers, the system has to yield the control to the surgeon again, safely and with the lowest possible delay [LTNR10].

The assessment of general HMI quality and hand-over capabilities is rarely found in surgical subtask automation-related current studies. However, such tests could enhance the clinical relevance, and could also improve the trust in those systems, of the public and the relevant authorities.

5.6 Robustness

Robustness is defined in a number of ways [144], the followings are the best fitting to surgical subtask automation:

- *The ability...to react appropriately to abnormal circumstances (i.e., circumstances "outside of specifications"). [A system] may be correct without being robust.* [145];
- *Insensitivity against small deviations in the assumptions [146];*
- *The degree to which a system is insensitive to effects that are not considered in the design* [147].

Reaching higher LoA, LoEC, and LoTC robustness of autonomous surgical systems becomes crucial. It is conclusive that LoA cannot be increased without higher robustness,

hence unexpected events may result in unwanted human handover request, potentially causing a dangerous scenario. The surgeon's SA might decrease during autonomous execution, effecting handover performance negatively. Furthermore, as LoEC and LoTC are increasing, the number of uncertainties, not considered circumstances also rises.

To measure the robustness of an autonomous system, additional methods should be utilized, the performance of the system should also be tested next to unexpected events. In the case of autonomous training task, it might be relatively simple to even manually generate such events (e.g., accidentally dropping grasped objects). Another example of robustness testing can be seen in the work of Elek et al. [LT1], where the performance of the perception algorithm was measured on different textures. However, in the case of higher complexity, the utilization of automatic robustness testing software may be necessary [148].

It is also common in deep learning to add noise to the input to increase robustness and avoid overfitting, like in the autonomous soft-tissue tensioning application by Nguyen et al. [52].

5.7 Legal Questions and Ethics

During the academic research phase, legal and ethical aspects of an autonomous surgical application or system are usually less emphasized issues. However, at the point when development approaches clinical trials, those aspects become critical. Since autonomous surgical systems could potentially endanger the life of the patient, the introduction of new standards and regulations in the field can be extremely difficult and must be elaborated. The availability of best practices for the validation of such systems could support those processes.

In the field of automation, liability is usually a prickly issue, but in general, as LoA increases, the liability in RAMIS is gradually shifting from the surgeon to the manufacturer of the system. The regulating authorities, such as the European Commission in the case of the European Union, are to protect citizens from harm caused by, in this case, autonomous surgical applications. Thus, in order to commercialize such solutions, the manufacturer need to demonstrate adequately that the autonomous system is capable of performing the intervention with equal or better performance as a human surgeon would do [129]. Proper characterization model and standard evaluation metrics would probably be quite useful during the procedure of legal approval.

The effect of automation on the surgeons' performance is also a significant concern. The utilization of autonomous functions in surgery may increase the reliance on them, and could lead to a decrease of skills of human surgeons. This lack of skills can be crucial, and may even risk the life of the patient in cases when the autonomous system fails, and the human surgeon need to take over the execution suddenly—especially if such failure is infrequent.

The definitions and safety requirements of surgical robotics are established in the standard *IEC 80601-2-77* [97]. The future standardization of the SCAC model, metrics, benchmarks or even autonomous surgical applications could be initiated through the following organizations:

- International Organization for Standardization (ISO);
- International Electrotechnical Commission (IEC);
- Institute of Electrical and Electronics Engineers (IEEE)
- Strategic Advisory Group of Experts (SAGE), advising World Health Organization (WHO);
- European Society Of Surgery (ESS) in Europe;
- Food and Drug Administration (FDA) in the USA.

In addition to the medical ethics, like the Hippocratic oath of *"do no harm"*, Artificial Intelligence ethics are also touching on autonomous surgery. Such as, the FAST Track Principles (Fairness, Accountability, Sustainability, and Transparency) should be considered during development. Ensuring fairness, like avoiding algorithmic or statistical biases is essential, since in the case of autonomous surgery those biases may lead to fatal consequences. Also, for example, in the case of deep learning methods, transparency could not be ensured; although the result of the network shows good accuracy, those it learned could not be interpreted by humans, and no one could predict its output for inputs yet unseen by the network [149].

Concepts, such as sustainability, from *roboethics*, a research field, concentrating on ethics in robotics [150], should also be applied to the development of autonomous surgical systems. Moreover, regulations on safety and privacy must be followed, like the Medical Devices Regulation (EU) 2017/745 (MDR), the General Data Protection Regulation (EU) 2016/679 (GDPR), or the EU AI Act within the European Union. Furthermore, the contents of the recently published standard IEEE P7007™ - *Ontological Standard for Ethically Driven Robotics and Automation Systems* is also going to be of critical significance in the development of surgical subtask automation.

5.8 Examples

The principles of the proposed standard evaluation are shown on the two example applications presented in Section 4.3: autonomous blunt dissection and autonomous peg transfer.

5.8.1 Validation of Autonomous Blunt Dissection

The implementation of the autonomous blunt dissection subtask presented in Subsection 4.4.1 is validated as follows. According to the SCAC model, the implemented application can be classified as $SCAC = F\{LoA = 2|LoEC = 2|LoTC = 3\}$:

- *Level of Autonomy (LoA) 2* – Task-level autonomy;
- *Level of Environmental Complexity (LoEC) 2* – Simple surgical phantoms, made for certain surgical subtasks, modeling one or few related key features of the real environment;
- *Level of Task Complexity (LoTC) 3* – Simple surgical tasks, no or Level 1 SA required.

The application was validated in silico, and the perception in ex vivo as well. According to the flowchart in Fig. 4.2, the following metrics should be measured:

- *Completion Time,* optional;
- *Application Accuracy;*
- *Cartesian Force;*
- *Grasp Force;*
- *Quality of the Outcome;*
- *Rate of Errors;*
- *Accuracy of Pose Estimation;*
- *Accuracy of Object Detection.*

The *camera calibration accuracy* was validated with the mean pixel error from 10 calibrations. The average of the mean pixel errors was 0.104 px, with standard deviation of 0.0165 px. In each of the 10 calibration sessions 19 image pairs were used of whom averagely 2.2 pairs were rejected (checkerboard detection or outlier).

One aspect of the *accuracy of pose estimation*, the *accuracy of the depth estimation* of the system was tested on a planar white and a checkerboard pattern paper. The depth of these objects was measured on different distances from the camera pair (Fig. 5.5). The mean error and the average of the standard deviation was 4.1 and 0.7 mm respectively.

Application Accuracy was derived from 10 test cases, an average of 2.2 mm accuracy was achieved with a standard deviation of 0.5 mm in the camera view's plane. In the depth axis the algorithm achieved 1 mm accuracy with standard deviation of 0.6 mm.

The *rate of errors*, characterizing the overall performance of the system was measured on the silicone-based custom designed phantom. Single dissections were made on 25 different locations on the dissection profile, of whom 21 succeeded; in 4 of the locations, the tool missed the dissection profile by a maximum of 3 mm, resulting in a 16% *rate of errors*.

The accuracy of the dissection line detection method (*accuracy of object detection*) was measured in three different setups: on the silicon phantom next to different rotations; on surfaces with different textures; and on ex vivo tissues. The algorithm's sensitivity to

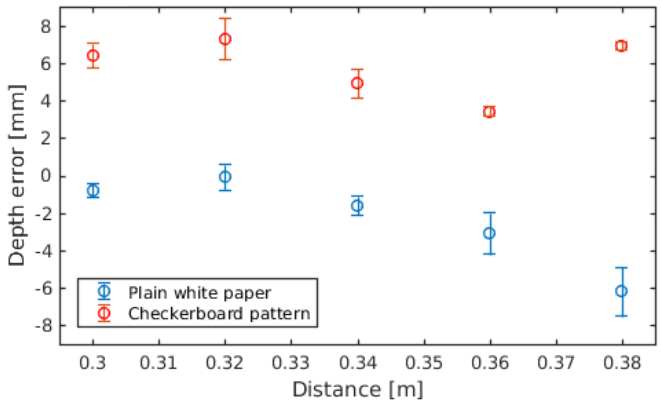

Fig. 5.5. Depth error of the objects with known surface on different distances from the stereo camera.

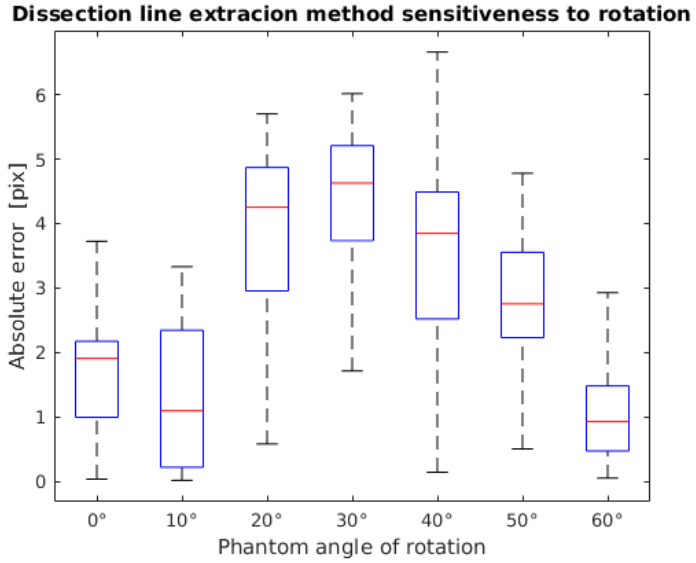

Fig. 5.6. Absolute error of the dissection line extraction method, demonstrating sensitivity to rotation. Boxplot showing the distribution of absolute errors, with the median indicated by the central line, the interquartile range (IQR) as the box, whiskers representing the range within 1.5 times the IQR. Rotation did not significantly influence the algorithm.

rotation was measured as follows. The silicon phantom was rotated $0 - 60$ °relative to the camera. It was found that my method is not significantly sensitive to rotation, as it worked acceptable in every cases (Fig. 5.6).

The dissection line detection method's sensitivity to texture was measured on four types of paper (plain white, checkerboard pattern, rough surfaced, and kraft paper) and the dissection phantom. The phantom and the papers were held in opened state to simulate retraction. In all of the cases, the algorithm had to find a linear dissection profile. The start and end points on the objects were chosen manually with 100 mm distance of each other; these points were the ground truth of the dissection line points. The objects placed from the stereo system approximately 500 mm distance. It was found that the tested method is highly sensitive to the texture and the pattern of the objects. The method worked well on feature-rich objects (with the checkerboard pattern, kraft paper, and the dissection phantom), but it failed on feature-poor objects (plain white paper and rough surface paper) (Fig. 5.7).

The accuracy of the algorithm was also measured in ex vivo environments—on chicken breast, pork shoulder, and duck liver. The sensitivity test was performed on the ex vivo objects: 6 points were selected as the basis of comparison between the ground truth points and the detected points. Based on the results, the method is sensitive to the texture of the object and to the lighting conditions. The method worked well on the pork shoulder, and it worked acceptable on the chicken breast and the duck liver. The reason is that pork is feature-rich, but the liver and the chicken breast are feature-poor and create reflections (Fig. 5.8).

The evaluation of the application would be full with measurements on *Cartesian force* and *grasp force*, since soft tissues were involved in the subtask. As it was mentioned, the measurement of the applied forces meets obstacles quite often due to the requirements

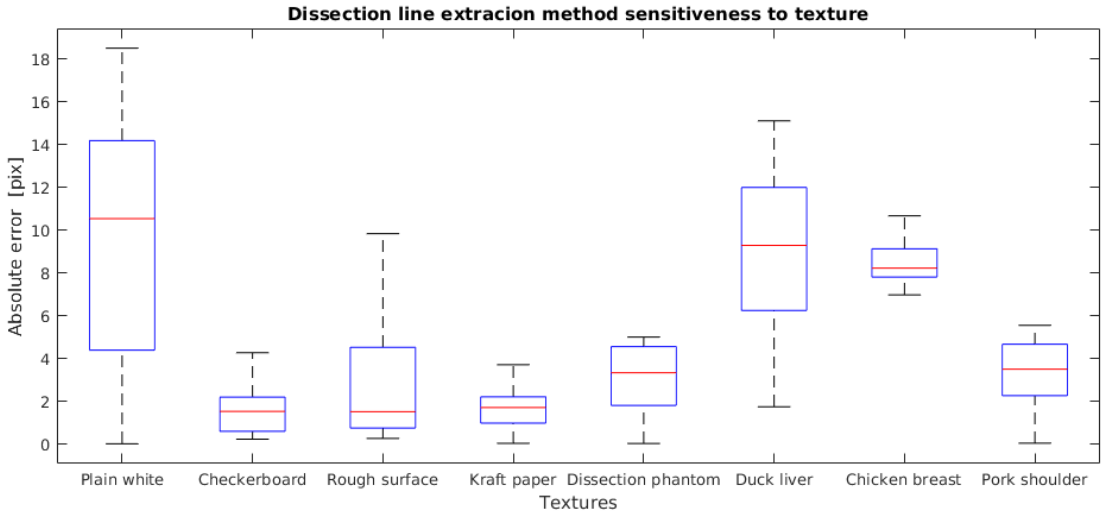

Fig. 5.7. Absolute error of the dissection line extraction method, demonstrating sensitivity to texture. Boxplot showing the distribution of absolute errors, with the median indicated by the central line, the interquartile range (IQR) as the box, whiskers representing the range within 1.5 times the IQR. The number of features and the shining of the objects are crucial in the detection of the dissection line.

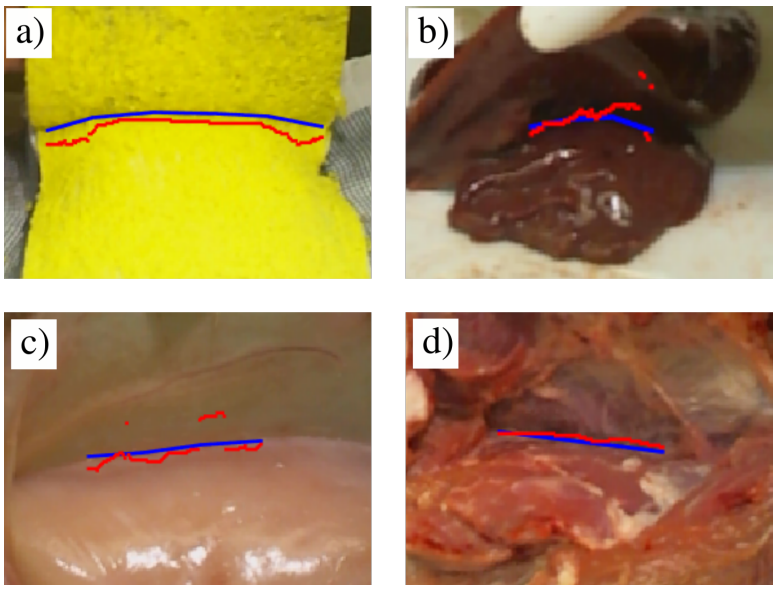

Fig. 5.8. Dissection line detection tests in vitro and ex vivo environment. a) Blunt dissection surgical phantom; b) duck liver; c) chicken breast; d) pork shoulder. The method is very sensitive to shining (see liver), and feature-richness (see chicken breast).

of additional, special devices. Sadly, the measurement of those metrics was not possible, since at the time of the measurements no sensors were available to measure these force values.

5.8.2 Validation of Autonomous Peg Transfer

The implementation of the autonomous peg transfer training exercise presented in Subsection 4.4.2 is validated and its performance is evaluated in the followings. The implemented application can be classified as $SCAC = F\{LoA = 2|LoEC = 1|LoTC = 1\}$:

- *Level of Autonomy (LoA) 2* – Task-level autonomy;
- *Level of Environmental Complexity (LoEC) 1* – Training phantoms, no or limited, highly abstract representation of the surgical environment;
- *Level of Task Complexity (LoTC) 1* – Simple training tasks, no or limited, distant representation of surgical task, no or Level 1 SA required.

The application was validated on the mentioned benchmarking environment. According to the flowchart in Fig. 4.2, the following metrics should be measured:

- *Completion Time;*
- *Application Accuracy;*
- *Cartesian Force;*
- *Grasp Force;*
- *Success Rate;*
- *Accuracy of Pose Estimation;*
- *Accuracy of Object Detection.*

The application was validated by completing 20–20 full unilateral and bilateral handover peg transfers. *Completion time* was expressed as *mean transfer time*—the mean of times for each individual transfers—to follow the convention by Hwang et al. [43]. *Success rate* is also expressed for individual transfers. The results, compared to the results of Hwang et al. are shown in Table 5.2.

The *application accuracy* of the shown implementation was measured as follows. The tip of the instruments were navigated to the key points of the scene. The navigation was

TABLE 5.2
THE PERFORMANCE OF THE IMPLEMENTED AUTONOMOUS PEG TRANSFER SOLUTION, INCLUDING COMPLETION TIME AND SUCCESS RATE, COMPARED TO THE WORK OF HWANG ET AL. [43].

Solution	Mean Transfer Time [s]	Success/Attempts	Success Rate [%]
Unilateral peg transfer by Hwang et al. [43]	5.2	120/120	100.0
Bilateral handover peg transfer by Hwang et al. [43]	8.1	111/118	94.1
Unilateral peg transfer using `irob-saf`	11.2	114/120	95.0
Bilateral handover peg transfer using `irob-saf`	16.3	109/120	90.8

TABLE 5.3
THE *application accuracy* OF THE IMPLEMENTED PEG TRANSFER SOLUTION, EXPRESSED IN POSITIONING ERROR.

	RMSE	SD	Max
Positioning error, Blocks [mm]	3.8	1.7	11.1
Positioning error, Board [mm]	3.1	1.2	7.3

performed based on the output of the implemented perception node, using the RGB-D camera stream. Then, the error of the positioning were corrected by manually moving the arms to the desired positions, while the difference of the two positions were recorded. By using the grasping points on the blocks and the top of the pegs, both the error regarding the blocks and the board were measured in 60–60 instances. The results of this measurement are shown in Table 5.3.

The evaluation of the application performing peg transfer autonomously would also be complete with measurements of *Cartesian force* and *grasp force* due to the involvement of soft tissues, and *accuracy of pose estimation* and *accuracy of object detection*. At the time of the measurements nor applicable force sensors nor additional tracking devices were available.

5.9 Summary of the Thesis

In this chapter, standard evaluation methods, metrics, and benchmarking techniques were proposed for performance evaluation and validation of systems executing surgical subtasks autonomously. A 3-axis SCAC model was proposed to represent more detailed characterization of the autonomous capabilities in RAMIS, and also in the wider field of surgery. This SCAC model uses the five and six grade scales of LoA, LoEC, and LoTC to represent autonomous surgical systems from a broad view.

Based on the review of literature, a set of performance metrics were presented and grouped by modality. After scoring the metrics by the aspect of usability of RAMIS subtask automation, the metrics were ranked, their properties were discussed. In the field of surgical subtask automation, I found the most widely used, and also the most meaningful metrics to be the outcome and accuracy metrics. Outcome metrics, like *success rate* are easy to implement, but—especially as SCAC increases—specific outcome metrics can also be defined to the given subtask. Accuracy metrics are also quite useful to validate components of the autonomous system, as *accuracy of pose estimation* and *accuracy of object detection* are among the ones that could and should be used for the validation of almost any autonomous surgical system, or their subsystems. The measurement of temporal metrics is very common, but as long as it is similar as at human execution, it is less informative, except in the case of benchmarks, where those become a good basis of comparison. Currently, the utilization of force-based metrics is quite uncommon, but as LoEC increases, the validation of delicate tissue handling will probably become critical issue.

A clear recommendation for the universal list of metrics cannot be given, since the

metrics best represent the quality of the autonomous execution greatly depend on the task and the validation environment. Instead, I proposed a methodology of choosing metrics for the validation of certain autonomous applications, and illustrated it in a flowchart (Figure 5.4). The principles of the proposed standard evaluation were presented on two examples: autonomous blunt dissection and autonomous peg transfer.

A review of current, and a proposal for further benchmarking techniquees are presented for surgical subtask automation. Benchmarks are going to have significant role in the future, during the introduction of subtask-level autonomy to the clinical practice by supporting the authorities' and the patients' trust in the autonomous systems. At present, a significant portion of research involves the automation of the peg transfer training task, hence, it serves as an adequate model to act as a foundation for surgical subtask automation. Additionally, the design files necessary for the 3D printing of the board and the pegs are available online [43]. In terms of benchmark datasets and simulations, the materials for a number of challenges are available, and even when the challenge is over, these could still serve as a good benchmark for future research projects.

The matter of HMI quality, robustness, legal and ethical issues were also discussed in brief. One of the most serious concerns regarding autonomous surgical systems—below LoA 5, *full autonomy*—is how a handover process could be performed in the case of emergency or malfunction. This concern is further strengthened by the trends in other areas, like self-driving vehicles, where automation led to increasing reliance on autonomous systems, bypassing the human operator.

The mentioned numerous concerns and fears related to autonomous surgical systems put an increasing need on the research community to perform through validation and testing on their developed applications. The proposed methodologies and recommendations could help the community to quantitatively and soundly measure the quality of the autonomous executions, and to provide a ground to compare the results of various research groups. At the point, when surgical subtask automation will break into the clinical practice, the proposed methodologies could also be used as the basis of emerging standards.

Chapter 6

ANALYSIS OF THE EFFECT OF AUTONOMOUS SUBTASK EXECUTION ON THE OPERATOR'S SITUATION AWARENESS

6.1 Introduction

Autonomous driving is often referred to as an analogous area to surgical automation, due to their similar complexity. This field is already one or two steps ahead of surgical automation. In this chapter, the effect of partial automation is analyzed in the field of self-driving, using a simulator interfaced to the DVRK-enhanced da Vinci Surgical System.

In the past few years, automotive technologies got a huge focus in research and development [26]. Fully automated vehicles are probable in the near future, however, there are still large milestones pending in development and safety [151]. Vehicle driving requires different skills from the driver, thus for autonomous solutions, new approaches are required, making it necessary to keep the human partially involved in control as well (when needed). The required skills on the HMI side depend on the level of automation. In automated driving technologies, there is a widely accepted classification of autonomy, which was introduced by the Society of Automotive Engineers (SAE) International (Fig. 6.1), and used in other research areas as well [7, 152]:

- **Level 0**: No automation. The vehicle is only permitted to send warning signals to the driver, it cannot interfere any of the controls. The human driver is responsible for controlling the vehicle in all aspects.
- **Level 1**: Driver assistance. The vehicle is allowed to control either steering or acceleration in cooperation with the human driver.
- **Level 2**: Partial automation. The vehicle performs complex actions by controlling both steering and acceleration in limited use-cases. The constant monitoring of the environment by the human driver is still required.
- **Level 3**: Conditional automation. The vehicle is prepared for the dynamic driving task by limited perception and decision-making abilities. The human driver is al-

lowed to divert its attention, in such a manner that he/she is able to take control back at any time if a fall-back event occurs.

- **Level 4**: High automation. The vehicle is equipped to perform the dynamical driving task in pre-defined driving modes. No real-time human–machine interaction is required, as the vehicle is able to move to a safe state from an emergency, at all possible conditions. In this safe state, the human driver could take over the control.
- **Level 5**: Full automation. The vehicle is able to accomplish the dynamic driving task in all the driving modes, regardless of the environment conditions.

The bold leap forward lies in between LoA 2 and LoA 3, not just in the technological terms, but in safety as well. LoA 2 still belongs to the well-known Advanced Driver Assistance Systems (ADAS), which means that the user is fully responsible for the driving and for the possible damages. In the case of LoA 3, the driver is partly responsible for the decision making: if the system gives a signal indicating it cannot handle the situation, the driver has to continue the process immediately and make decisions. Naturally, this solution requires much more advanced technologies to handle the different environments, such as AI, to manage diverse environments effectively [17, 153]. The main issue with LoA 3 is that the essential functions of driving are automated, and because of it the driver

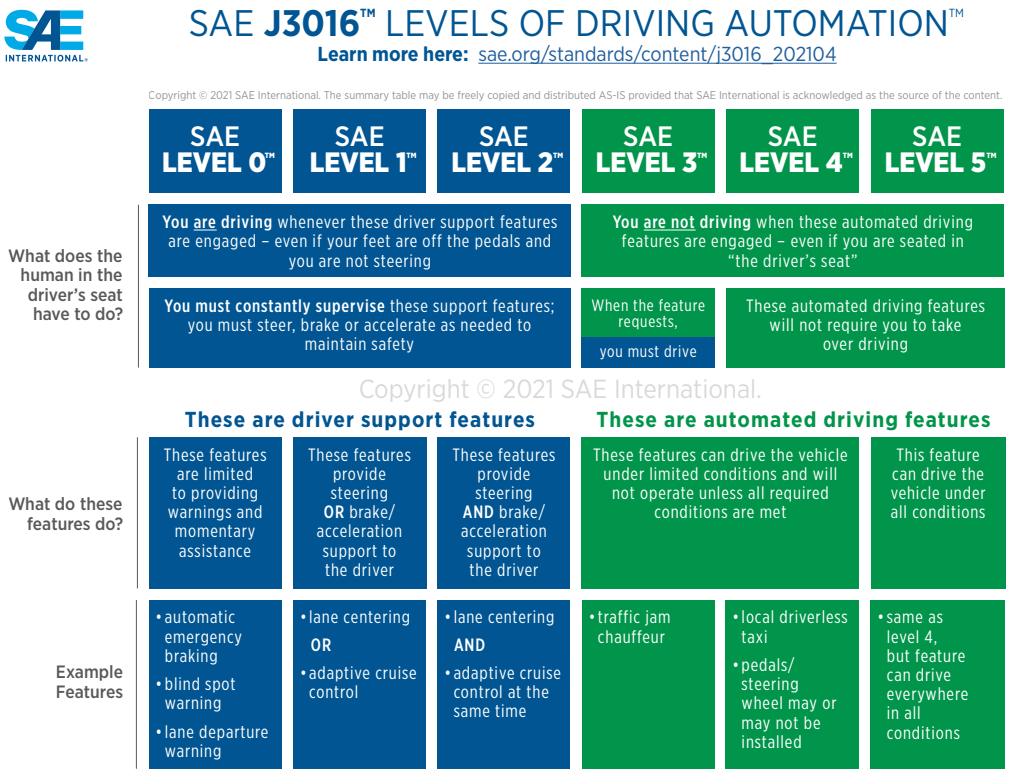

Fig. 6.1. Level of Autonomy (LoA) concept for automated vehicles introduced by the Society of Automotive Engineers International [152].

can easily be distracted—which can be crucial under critical conditions. Studies show that the human mind is not effective in long inactive monitoring tasks, and usually over-trust the automated system [154, 155, 156, 157, 158, 159, 160]. Because of these reasons, in LoA 3 safety considerations are indispensable.

SA is a key factor of driving safety, thus an important notion in LoA 3. SA is defined on 3 levels [98, 99, 161, 162][LTNR6]:

- **Level 1 SA**: Perception of the environment;
- **Level 2 SA**: Comprehension of the current situation;
- **Level 3 SA**: Projection of future status.

In the cases of LoA 0, 1 and 2 SA is obviously important: the driver has to constantly monitor and understand the environment, and estimate the future. The SA aspect of LoA 2 appears in the aviation automation: the pilot has to handle at least one function of the cruising, which can help not to lose SA [18, 99]. In LoA 3, SA is more special: while the autonomous functions are working well, the driver can easily lose SA, and when it is necessary, he/she cannot react well or acceptably fast [163, 164, 165, 166]. This situation, when the driver has to take back the control is called *handover*, and the necessary time for it is called *takeover* [164]. Under non-critical situations takeover is usually between 1.9 to 25.7 seconds, but it can be much more under critical conditions [167].

Situation awareness can be decomposed into 5 components (Fig. 6.2):

- *Spatial awareness*: knowledge of object locations;
- *Identity awareness*: knowledge of salient items;
- *Temporal awareness*: knowledge of the dynamic states;
- *Goal awareness*: knowledge of the maneuvering plan;
- *System awareness*: knowledge of the environment.

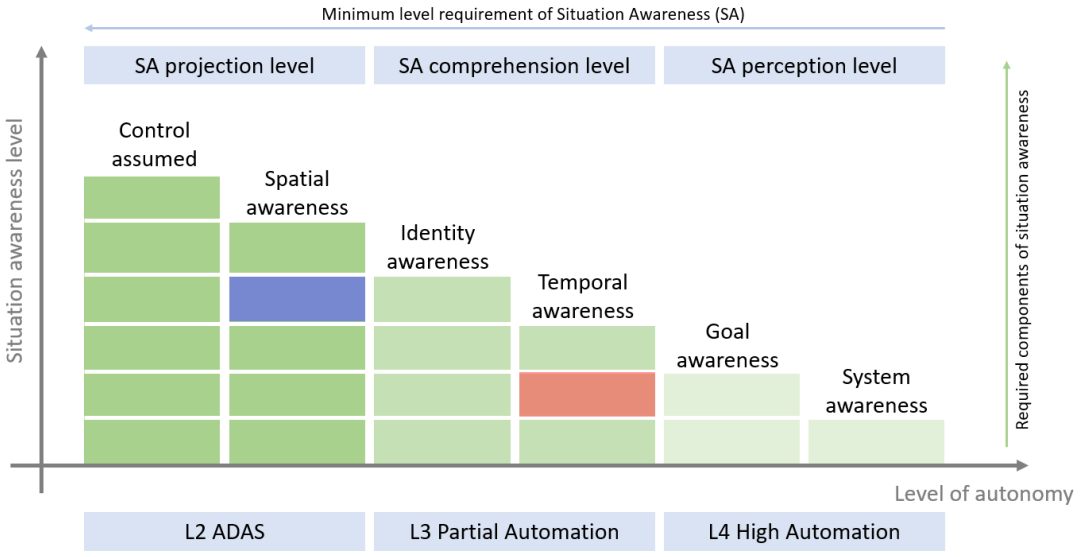

Fig. 6.2. Hierarchical representation of Situation Awareness blocks in automotive solutions. For each level of autonomy, the quantitative metrics must fulfill the requirements for each block.

The aim of this research was to assess the relationship between SA levels and performance during emergency situations, drawing on existing literature on the measurement of SA in similar contexts. Specifically, two research hypotheses were investigated:

- **Hypothesis 1:** SA can be accurately measured at participants during simulated emergency scenarios using a combination of freeze probe technique, self-rating, and performance measures.
- **Hypothesis 2:** The level of SA at participants during critical handover situations affects their performances, higher SA leads to better task performance during emergencies.

In the following sections, a measurement framework is introduced, designed to examine SA at self-driving technologies under critical conditions. To objectively measure handover, a novel system architecture was created, using the Master Console of the da Vinci Surgical System as HMI alongside the CARLA Simulator [57, 134]. The da Vinci system is capable of providing the limited view of the environment, which is critical in SA. Using the developed platform, the handover process can be modeled under emergency with different situations. Furthermore, an experimental study is presented, where the SA is measured and evaluated in the designed environment [LT11][LTNR6].

6.2 Measurement Framework to Assess Situation Awareness During Handover

In this section, the methodological principles for assessing SA during handover situations are introduced, forming the foundation of the proposed measurement framework. This framework integrates both a structured methodological approach and a purpose-built experimental platform to systematically evaluate the transition from autonomous to manual control under critical conditions. The framework builds on established techniques from the literature, incorporating both qualitative and quantitative methods to ensure a comprehensive assessment of SA in dynamic, high-stakes scenarios.

6.2.1 Measurement Methodology

The transition from automated to manual control represents a critical phase in which the driver must rapidly regain situational awareness and respond appropriately to dynamic environmental conditions. Given the complexity of SA as a cognitive construct, its assessment requires a combination of complementary techniques that capture both subjective and objective indicators of awareness and performance. The methodology adopted in this framework draws from established approaches in SA measurement, integrating elements from the literature to ensure objective evaluation.

One of the primary methods employed is the freeze probe technique, a widely recognized approach for assessing real-time cognitive processing. Originally developed as part of the Situation Awareness Global Assessment Technique (SAGAT) [168], the freeze probe method involves interrupting a task at predetermined points to assess the participant's awareness of critical situational elements. By temporarily halting the simulation

and requiring participants to recall specific environmental details, this technique provides a direct measure of SA at the moment of handover. Research has demonstrated that the freeze probe approach mitigates the limitations of post-hoc recall biases, offering a more reliable indicator of real-time cognitive processing [98].

In addition to the freeze probe method, self-rating scales were employed to capture participants' subjective perceptions of their SA during the handover scenarios. Self-assessment techniques such as the Situational Awareness Rating Technique (SART) [169] have been widely used in the literature to evaluate perceived cognitive workload and situational comprehension. While subjective measures are inherently influenced by individual biases, these provide valuable insight into participants' confidence in their understanding of the situation, as well as their perceived ability to respond effectively. Research has shown that self-rating scales can serve as a useful complement to objective measures, particularly when combined with performance-based assessments [170].

To further substantiate the assessment of SA, objective performance metrics were established, focusing on key behavioral indicators of effective response during the handover process. These metrics included reaction time, the accuracy of responses, and the effectiveness of decision-making in resolving the simulated emergency [154]. Studies have found that task performance can serve as an indirect indicator of SA, as individuals with higher levels of awareness tend to exhibit faster and more effective responses to dynamic scenarios [161, 171]. However, it is important to recognize that performance alone does not fully encapsulate SA, as a participant may achieve successful outcomes through compensatory strategies rather than an accurate mental model of the situation. To address this limitation, performance data were analyzed in conjunction with the freeze probe and self-rating measures to provide a more comprehensive evaluation of SA.

By integrating these three complementary methods—freeze probe assessments, self-rating scales, and performance metrics—this measurement framework adopts a multi-faceted approach to measuring SA during critical handover situations. The combination of real-time cognitive assessment, subjective self-evaluation, and objective performance analysis allows for a more nuanced understanding of how SA is maintained and influences decision-making in high-stakes scenarios.

6.2.2 Measurement Platform

Building upon the methodological considerations introduced earlier, this section outlines the design of a measurement platform, aimed at objectively and quantitatively assessing SA during the handover process at autonomous driving. The platform integrates the da Vinci Surgical System, traditionally used in teleoperated robot-assisted surgery, with modifications tailored for driving simulation. The da Vinci system is enhanced by the DVRK platform, allowing the system's HMI to be repurposed for handover experiments in autonomous vehicles (Fig. 6.3) [2].

In the Master Console, the da Vinci provides a fixed head position, where the operator can only see the display, but cannot see the environment around him, and vice versa, when his head is not inside the required area, he cannot see the display (Fig. 6.4a). The stereo monitors of the console makes possible to display the simulated environment in 3D enhancing immersion and realism in the driving experience. Furthermore, the da Vinci has a head presence sensor (IR-beam) to detect if the user is looking into the display or

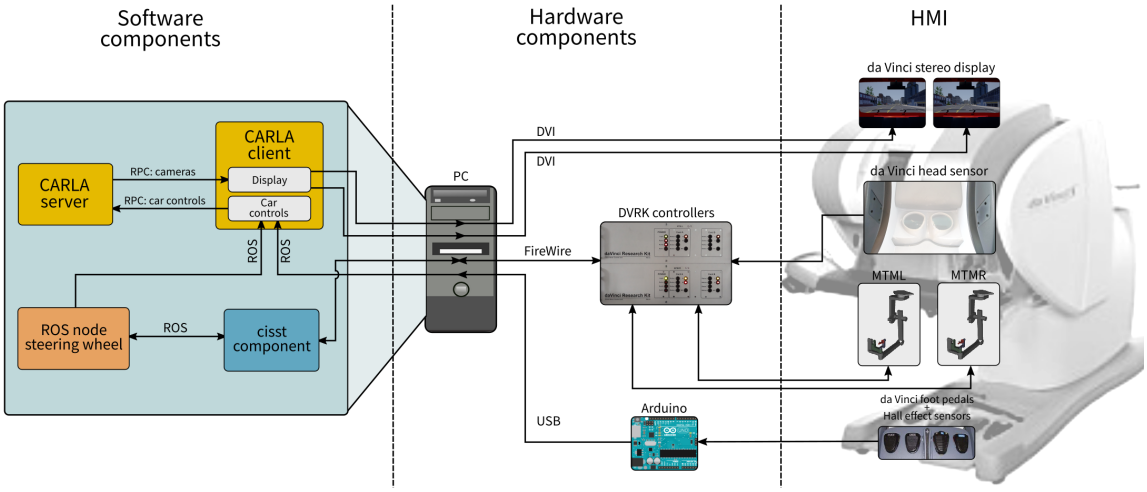

Fig. 6.3. Measurement platform with the DVRK-enhanced da Vinci Surgical System to examine situational awareness under critical conditions. The da Vinci master provides the display, the head sensor, the wheel, and the pedals to imitate a driving environment, and the setup is linked to the CARLA driving simulator with ROS components.

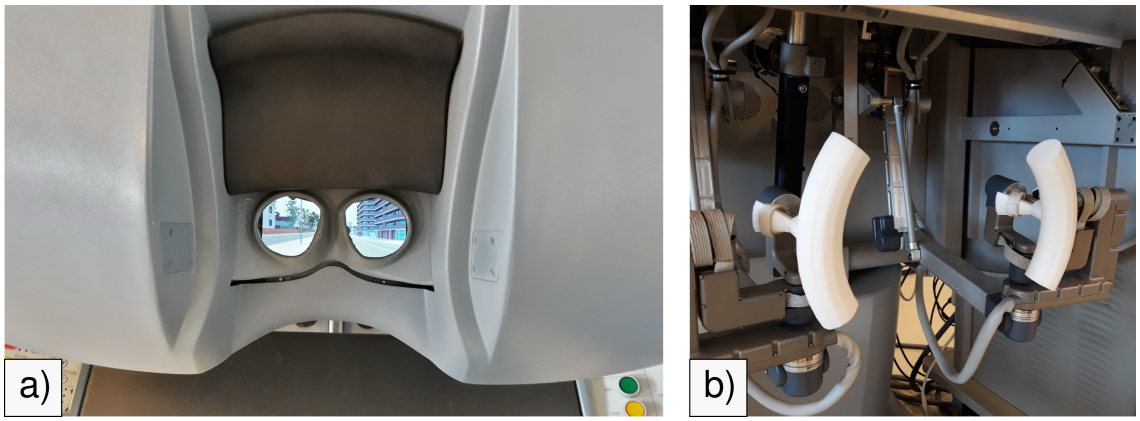

Fig. 6.4. The da Vinci Master Console modified for SA measurements in self driving handover situations. a) The da Vinci Master Console's stereo display with an integrated CARLA car simulator setup; b) the da Vinci Surgical System master arms amended with 3D printed wheel segments to imitate a steering wheel.

not (in the case of RAMIS it is a critical factor). With the fixed head position and built-in head sensor, attention can be monitored (at large). This setup mirrors a key aspect of the autonomous vehicle handover scenario, where the driver must be engaged and ready to take control at receiving a handover alert. Furthermore, the Master Tool Manipulators (MTMs) are well-suited for simulating a steering wheel and foot pedals in this context. By modifying the MTMs with 3D-printed components, a flexible and reproducible system can be created for use in other research laboratories (Fig. 6.4b). This approach facilitates rapid prototyping and offers consistent results across studies. The software for the da Vinci system supports impedance control, which limits the movement of the manipulator arms to a circular path, mimicking the motion of a steering wheel [172]. This approach enables accurate tracking of steering inputs and replicates the physical interaction that a driver would have during a handover in an autonomous vehicle.

The da Vinci pedals, originally designed for controlling medical equipment in surgery, were also modified for this experiment to provide continuous state readings for accelerator and brake inputs. Hall-effect sensors and magnets were incorporated into the pedals to detect precise input levels, which were then connected to an Arduino board (Arduino Co., Somerville, MA) [173] to track pedal movements.

The core software components of the experimental setup were the DVRK [57] and the CARLA Simulator[6]. CARLA is an open-source driving simulator, in this study it was interfaced to the da Vinci master. To link the components of the system, ROS was used, which is well-known library for robotic research [56]. DVRK, CARLA, and Arduino support ROS communication [134].

The core software architecture is designed to manage the interaction between these components. The CARLA server runs the simulation environment, while a CARLA client implemented in Python communicates with the server using Remote Procedure Calls (RPCs). The client forwards steering angles and pedal values received through ROS to control the driving simulation in real-time. Additionally, a second ROS node adjusts the impedance control gains to ensure smooth, realistic steering responses. The MTMs, through the DVRK software, are also programmable via ROS, providing fine-grained control over the manipulator arms (Fig. 6.3).

The developed measurement platform provides an immersive simulation environment, and also offers a framework to measure driver engagement and situational awareness during handover situations.

6.3 Experimental Study

This section presents an experimental study investigating the effects of partial automation on drivers' SA and handover performance in critical scenarios. Specifically, the study examines how drivers regain control and assess their surroundings after a period of automated driving, focusing on the cognitive and behavioral aspects of SA during the transition from automation to manual control. The study is supported by the previously introduced measurement framework.

6.3.1 Experimental Protocol

To analyze driver responses in emergency situations requiring a handover from automation, four distinct driving scenarios were developed. Each scenario required participants to react to an emergency auditory alarm signaling an imminent hazard. The study involved 15 participants, all of whom held a valid driver's license. Each participant completed only one experiment.

Before the experiment, a minute was given to the participants to get acquainted with the driving simulator. After each scenario, they completed a questionnaire assessing their awareness of the event, confidence in their response, and overall perception of the situation.

Each scenario began with a period of autonomous driving at LoA 3 automation. During this phase, participants were instructed to disengage from monitoring the road by per-

[6] http://carla.org/

forming a secondary task—typing a text message on a mobile phone. This ensured that their attention was diverted, reflecting a realistic challenge of driver re-engagement in automation-reliant conditions. The duration of the autonomous phase was randomized between 40 and 60 seconds, maintaining consistency across participants while introducing unpredictability.

At the end of the autonomous phase, an emergency audio alarm was triggered, signaling a critical event and initiating the handover process. Participants were required to assume manual control and respond to the situation, as the vehicle was no longer capable for handling the event autonomously. To discourage overly cautious responses, participants were informed that unnecessary braking would incur a penalty.

All scenarios were conducted at the same simulated location under clear weather conditions to ensure environmental consistency. The scenarios varied based on two key factors: the presence of a pedestrian emergency and the presence of oncoming traffic. Additionally, the validity of the pedestrian-related alarm varied, some alarms correctly indicated danger and others were false alarms. The four scenarios were constructed as follows:

- **True alarm:** The pedestrian stepped in front of the car from behind a vending machine (Fig. 6.5), close enough to hit him/her (alarm raised 3 seconds before the vehicle would reach the pedestrian).
- **False alarm:** The pedestrian was moving on the sidewalk, with safe distance from the car (alarm raised 3 seconds before the vehicle would reach the pedestrian).
- **Car arriving:** There was oncoming traffic.
- **No car arriving:** There was no oncoming traffic.

Fig. 6.5. Simulation screenshot of one scenario (true alarm, no oncoming traffic).

TABLE 6.1

QUESTIONS OF THE QUESTIONNAIRE THE SUBJECTS WERE ASKED TO FILL AT CERTAIN POINTS OF THE EXPERIMENT.

Before the experiment	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**
• Age; • Gender; • When did he/she gained driving license; • How frequently he/she drives; • Did she/he ever drive a car with ADAS;	• How lifelike were the simulator and the scenario; • Evaluate own reaction on a scale 1–5; • Who/what caused the emergency; • What was the color of the pedestrian's pants; • After the alarm what was the direction of the road;	• Evaluate own reaction on a scale 1–5; • Who/what caused the emergency; • Where was the bus station during the emergency (what side of the road); • Why did she/he not stop at the pedestrian crossing;	• Evaluate own reaction on a scale 1–5; • Who/what caused the emergency; • What was on the right side of the road during the emergency; • After the emergency what could be your maximum speed;	• How lifelike were the simulator and the scenario; • Evaluate own reaction on a scale 1–5; • Who/what caused the emergency; • Was there oncoming traffic; • What was on the left side of the road during the emergency;

From the combinations, 4 different scenarios were compiled:

1. True alarm, No car arriving;
2. False alarm, Car arriving;
3. True alarm, Car arriving;
4. False alarm, No car arriving.

Before the experiment and after each scenarios, subjects were asked to fill a Google Forms (Google LLC., Mountain View, CA) questionnaire about their experiences (Table 6.1.). Before the filling, they agreed the terms of the experiment and the data was anonymous.

6.3.2 Results

There were 15 subjects, 13 males and 2 females, mostly young adults (ages 21–34). The subjects have never driven a car with ADAS before, except for one, who was unsure.

The number of scenarios with collisions for each participants is shown in Fig. 6.6. There were relatively high number of collisions, most of the time the participants have collided with the curb. The number of collisions per scenarios are shown in Fig. 6.7. The number of collisions increased during the second scenario, probably due to the car coming from the front lane, regardless of the fact that there was a false alarm. The number of collisions was smaller at the last two scenarios, which were repeated scenarios in a sense that the participants had experienced front traffic and true and false alarms as well (i.e., all the components of the scenarios), and the participants had larger SA.

The SA of the subjects was calculated based on their answers concerning the environment in the questionnaire. Good answers gained 1 point, wrong answers resulted in -1 point, neutral answers 0 points. In the question about the direction of the road, the right answer was left, but straight was accepted as a correct answer with 0.5 point as well, since the left turn was not directly after the place of the potential accident. The mean SA for non-collision cases was 3.873, whereas for collision cases, it was, lower at 2.785

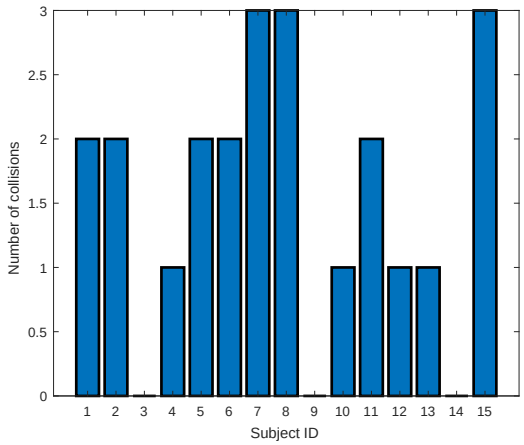

Fig. 6.6. The number of scenarios with collisions for each participants.

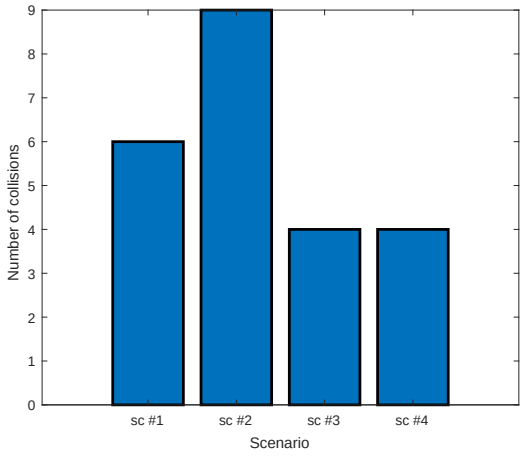

Fig. 6.7. The number of collisions for each scenario.

(Fig. 6.8), however, the difference was not statistically significant (t-test [174], $p = 0.078$, SA values follows the normal distribution). The correlation between SA and collisions was analyzed using Pearson's correlation coefficient [175]. For the whole dataset, SA and collisions shown weak negative correlation, but the result was not statistically significant (correlation coefficient $r = -0.229$, $p = 0.078$). The four scenarios were also analyzed separately; significant strong negative correlation was found between SA and collisions in *scenario 3* ($r = -0.596$, $p = 0.019$) and *scenario 4* ($r = -0.551$, $p = 0.033$).

The post-scenario questionnaire results further highlight gaps in awareness:

- **Scenario 1** (True Alarm, No Oncoming Traffic):
  - 80% correctly identified the pedestrian as the cause of the emergency;
  - Only 40% remembered the pedestrian's pant color;
  - 53.3% correctly identified the road's leftward turn.
- **Scenario 2** (False Alarm, Oncoming Traffic):
  - 73.3% recognized that the emergency was due to the automation system;
  - 86.7% correctly recalled that a forest was on the right;
  - Only 26.7% identified the correct speed limit.
- **Scenario 3** (True Alarm, Oncoming Traffic):
  - 80% correctly identified the pedestrian as the cause of the emergency;
  - Only 33.3% placed correctly the bus stop location;
  - 73.3% realized there was no pedestrian crosswalk.
- **Scenario 4** (False Alarm, No Oncoming Traffic):
  - 60% identified the automation system as the cause of the emergency;
  - 93.3% correctly noted the absence of oncoming traffic;
  - 93.3% correctly identified houses on the left.

These findings indicate progressive improvement in SA but also highlight areas where SA remained inconsistent, such as recognizing road infrastructure details. Fig. 6.9 illustrates the SA scores across the four scenarios, demonstrating the mentioned increasing trend. This learning effect was evaluated using linear regression analysis [176] and Cohen's d [177]. The SA score increased by 0.27 per scenario ($p = 0.044$, significant), the effect size was small (Cohen's $d = 0.467$). This suggests that the subjects' ability to assess and respond to the environment improved with exposure to the task. Such improvements could be attributed to increased familiarity with the task, reduced cognitive load, or greater confidence in decision-making as subjects progress through the experiment.

The observed increase in SA across scenarios, along with the correlation between SA and collisions emerging only in the final two scenarios, suggests that a minimum threshold of SA might be necessary for effectively managing emergency situations. Given the nature of the scenario, this threshold is likely at SA Level 2. However, quantifying SA levels based on questionnaire responses is not straightforward. A potential approach to determining this threshold is to identify the SA level at which the number of collisions decreases significantly. This could be achieved through a repeated study with a larger sample size to enhance statistical reliability.

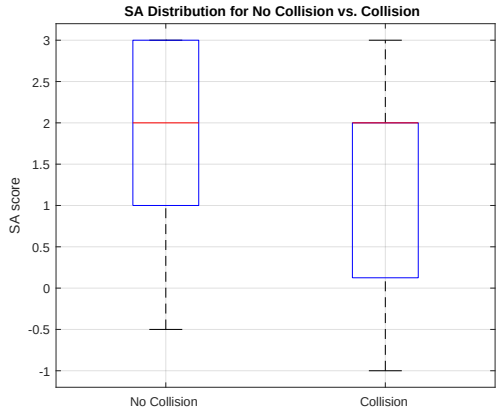

Fig. 6.8. The SA score distribution of the participants in the four scenarios, with and without collision. Boxplot showing the distribution of absolute errors, with the median indicated by the central line, the interquartile range (IQR) as the box, whiskers representing the range within 1.5 times the IQR.

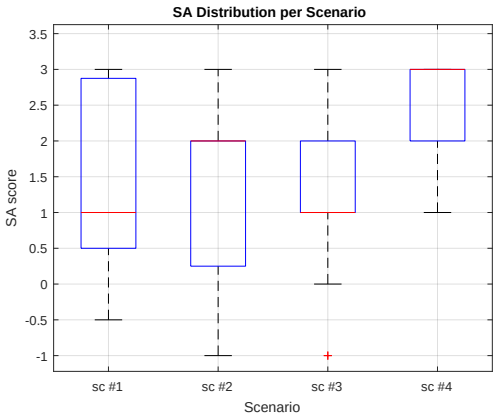

Fig. 6.9. The SA score of the participants in the four scenarios. Boxplot showing the distribution of absolute errors, with the median indicated by the central line, the interquartile range (IQR) as the box, whiskers representing the range within 1.5 times the IQR, and outliers beyond this range.

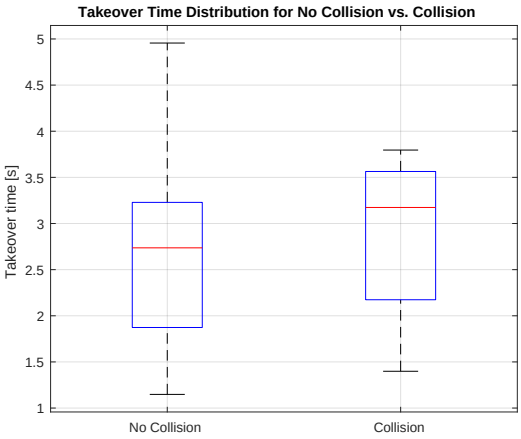

Fig. 6.10. The takeover times of the 15 subjects during the four scenarios, with and without collision. Boxplot showing the distribution of absolute errors, with the median indicated by the central line, the interquartile range (IQR) as the box, whiskers representing the range within 1.5 times the IQR.

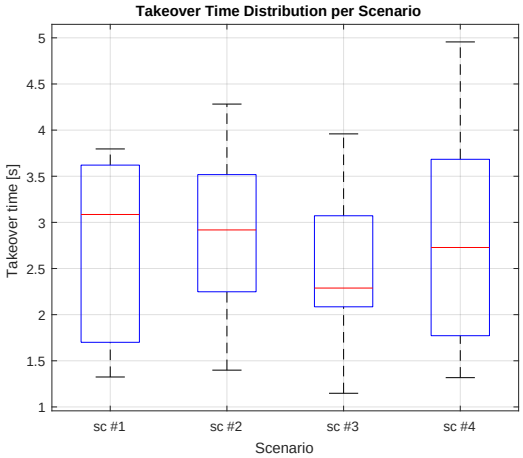

Fig. 6.11. The takeover times of the 15 subjects during the four scenarios. Boxplot showing the distribution of absolute errors, with the median indicated by the central line, the interquartile range (IQR) as the box, whiskers representing the range within 1.5 times the IQR.

Fig. 6.10 and Fig. 6.11 show the takeover times for scenarios with and without collision and for each scenario across all participants, respectively. The mean takeover time for collision cases was 2.82 seconds, while for non-collision cases, it was slightly lower at 2.7 seconds. However, no significant correlation was revealed between takeover time and collision occurrence ($r = 0.093$, $p = 0.489$) or between takeover time and SA ($r = 0.139$, $p = 0.297$) by statistical analysis. Moreover, as takeover time did not correlate with the number of collisions, while the SA score did, that might suggest that Level 2 SA of the driver (comprehension of the current situation) has a decisive role in handover

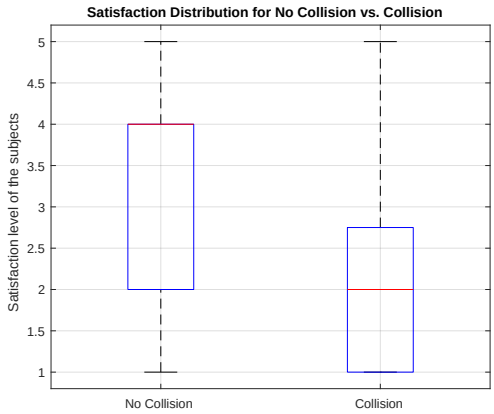

Fig. 6.12. The satisfaction of the 15 subjects during the four scenarios, with and without collision. Boxplot showing the distribution of absolute errors, with the median indicated by the central line, the interquartile range (IQR) as the box, whiskers representing the range within 1.5 times the IQR.

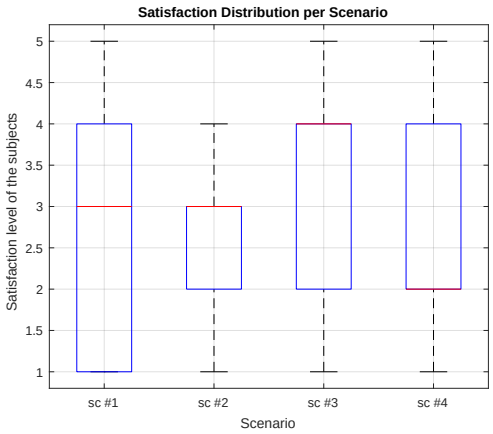

Fig. 6.13. The satisfaction distribution of the participants in the four scenarios. Boxplot showing the distribution of absolute errors, with the median indicated by the central line, the interquartile range (IQR) as the box, whiskers representing the range within 1.5 times the IQR.

performance, while Level 1 SA (perception of the environment) itself may not be enough for a proper reaction.

Additionally, the analysis of the effect of practice showed no significant impact on takeover time. These findings suggest that while SA increases and the number of collisions decreases over repeated scenarios, takeover time remains relatively stable. This implies that repetition does not necessarily decrease takeover time, but it improves SA which decreases the chance of collisions [178].

The satisfaction level of the subjects, based on the questionnaire, is shown for scenarios with and without collision and for each scenario across all participants in Fig. 6.12 and Fig. 6.13. The mean satisfaction level for the cases with no collision is 3.35, while the mean satisfaction level with the cases with collision is 1.93. The results imply strong, statistically significant negative correlation between satisfaction and collisions ($r = -0.512$, $p = 0.000029$), thus collision decreased the satisfaction of the participants. Also, the satisfaction did not improve with the scenarios, so gaining SA does not improve satisfaction.

The potential correlation between driving experience and key performance metrics, including SA, takeover time, and the number of collisions, was analyzed. The results indicated no statistically significant relationships, with correlation values of ($r = 0.0445$, $p = 0.875$) for SA, ($r = 0.185$, $p = 0.51$) for takeover time, and ($r = 0.0488$, $p = 0.863$) for collisions. One possible explanation is that automation levels the playing field, as all participants were disengaged from the driving task during autonomous operation, reducing any potential advantage of prior driving experience. Additionally, handover performance may depend more on automation-specific learning and cognitive adaptation rather than traditional driving skills [179].

6.4 Summary of the Thesis

In this chapter, a measurement framework was presented, which included methodological approaches and a measurement platform based on the DVRK-enhanced da Vinci Master Console and the CARLA driving simulator for the quantitative analysis of SA during LoA 3 handover scenarios. The current state of autonomous driving was reviewed, focusing on safety and the impact of SA on the driver's handover performance. An experimental study was conducted within the proposed framework, involving 15 test subjects who performed emergency handover tasks during autonomous driving.

The chapter proposed two primary hypotheses: first, *SA could be accurately measured at participants during simulated emergency scenarios using a combination of the freeze probe technique, self-rating, and performance measures*; and second, *the level of SA at participants during critical handover situations would affect their performance, with higher levels of SA leading to better task performance during emergencies*.

The results support the first hypothesis, demonstrating that SA can be measured effectively using the proposed combination of methods during the handover process. The freeze probe technique provided valuable insights into the participants' cognitive state during specific moments of the emergency scenarios, while the self-rating measures captured their subjective awareness of the situation. Additionally, performance measures offered objective data regarding task execution, enabling a holistic assessment of SA. By combining simulated driving environments, performance metrics, subjective self-assessments, and realistic, stress-inducing scenarios, the presented approach aligns with proven methodologies for SA measurement. The integration of controlled experimental design with advanced technology (e.g., CARLA simulator, DVRK system) ensures that the data collected is reliable, valid, and replicable—critical for understanding how SA influences human performance in the context of autonomous driving. Despite these strengths, discrepancies were noted between participants' subjective self-ratings and their objective performance, suggesting that while these methods provided a reliable measure of SA, there might be a need to integrate additional tools, such as physiological sensors or real-time situational monitoring, to capture a more accurate picture of SA in future research. The open-source implementation of the measurement platform is available at `https://github.com/ABC-iRobotics/dvrk_carla`.

Regarding the second hypothesis, the results partially support the notion that higher SA improves task performance in handover situations. The takeover time of the participants was shown to not decrease significantly in the four successive scenarios. In contrast, the SA scores, derived from the questionnaire responses, showed an increasing trend, interpretable as a learning curve, while the number of collisions decreased. These results suggest that the success of the handover maneuver is strongly influenced by the driver's SA. Additionally, the analysis revealed no correlation between takeover time and the number of collisions, while the SA score showed a partial correlation (notably in the last two scenarios). This implies that Level 2 SA (comprehension of the current situation) plays a crucial role in handover performance, while Level 1 SA alone may not be sufficient for an appropriate response.

These findings underline the crucial role that SA—particularly the higher-level cognitive processes associated with understanding and predicting the unfolding situation—plays in handover scenarios next to partial or conditional automation. The results suggest

that training drivers to develop and maintain this level of awareness could enhance their ability to safe transition control in emergency situations. Future research should further explore the measurement of SA, particularly by refining the existing methods and introducing more nuanced performance metrics, while also investigating the effects of real-time feedback and adaptive training systems on SA and handover performance.

Chapter 7

CONCLUSION

7.1 Summary of Contributions

This thesis investigates the automation of surgical subtasks in RAMIS, with a focus on recent advancements in surgical robot motion planning, perception, and human-machine interaction, while addressing the limitations of task-level autonomy. A framework to support the automation of RAMIS subtasks is proposed, alongside a characterization model for surgical automation, as well as a method for the performance evaluation and comparison of automated surgical subtasks. Additionally, the effects of automation on the performance of both surgeons and vehicle drivers are framed and explored.

A standardized methodology for automating surgical subtasks is presented, based on the hierarchical decomposition of human surgical motions. This methodology serves as the foundation for a framework designed to facilitate surgical subtask automation research. The framework integrates sensory inputs, perception algorithms, and robotic systems, and includes a surgeme-level motion library. The presented framework allows for the easy integration of new surgical subtasks into the motion library, such as clipping or suturing. The development of more complex subtasks remains a challenge, particularly regarding the perception and estimation of the environment, as computer vision techniques struggle with issues such as light reflections and the recognition of deformable or moving tissues, even in controlled phantom or ex vivo environments.

Standard evaluation methods, metrics, and benchmarking techniques were proposed for performance evaluation and validation of systems executing surgical subtasks autonomously. A 3-axis model of surgical autonomy is proposed to enhance the characterization of surgical automation. The thesis also reviews the existing metrics and proposes additional promising techniques for evaluating, comparing, and benchmarking automated surgical systems. The discussion includes the critical need for robustness in these systems, alongside considerations of legal and ethical implications in the field.

Finally, a measurement framework for the quantitative analysis of SA during LoA 3 handover scenarios is introduced. By combining the da Vinci Surgical System with the CARLA driving simulator, the research investigates how SA influences performance during emergency handover tasks at autonomous driving. An experimental study demonstrates that the SA has a key role in the drivers performance during emergency handover tasks.

7.2 New Scientific Results

Thesis 1

I developed a method to support the automation of surgical subtasks in RAMIS. The presented methodology is based on the hierarchical decomposition of human surgical motions, enabling high modularity. Additionally, based on the developed method, I have designed a system architecture and implemented a software framework, capable of realizing autonomous surgical applications effectively.

Related publications: [LT1, LT2, LT3, LT4, LT5, LT6]

Thesis 2

I developed a method for the validation of autonomous applications in the field of surgical subtask automation, originating from the human surgical skill assessment techniques. I created a model to represent the capabilities of autonomous surgical systems. I have reviewed, organized, and graded the metrics from the field of capability and performance evaluation in surgical subtask automation and related research areas, like autonomous robotics or surgical skill assessment. Additionally, I compiled a method to choose the metrics best describing the performance of applications performing surgical subtasks autonomously.

Related publications: [LT7, LT8]

Thesis 3

I addressed the objective monitoring and quantification of Situation Awareness (SA) and evaluated its impact within the context of partial automation.

Thesis 3/I: I conceived a measurement framework for the quantitative analysis of vehicle driver's SA during LoA 3 handover scenarios using the DVRK-enhanced da Vinci Surgical System. I showed that the proposed framework effectively enables the objective measurement of SA through the combination of methods during the handover process.

Thesis 3/II: I analyzed the effect of SA on the handover performance during LoA 3 emergency situations in the proposed measurement framework. I showed that the success of such handover maneuver highly depends on the driver's SA. Furthermore, I showed that SA exhibits an increasing trend across successive scenarios, suggesting that the subjects' ability to assess and respond to the environment improved with exposure to the task.

Related publications: [LT9, LT10, LT11]

Other publications related to the Ph.D. thesis and the accompanying research work: [LTNR1, LTNR2, LTNR3, LTNR4, LTNR5, LTNR6, LTNR7, LTNR8, LTNR9, LTNR10, LTNR11]

REFERENCES

[1] L. Márton, Z. Szántó, T. Haidegger, P. Galambos, and J. Kövecses, "Internet-based Bilateral Teleoperation Using a Revised Time-Domain Passivity Controller," *Acta Polytechnica Hungarica*, vol. 14, no. 8, pp. 27–45, 2017.

[2] Á. Takács, D. Á. Nagy, I. J. Rudas, and T. Haidegger, "Origins of Surgical Robotics: From Space to the Operating Room," *Acta Polytechnica Hungarica*, vol. 13, no. 1, pp. 13–30, 2016.

[3] A. Tewari, J. O. Peabody, M. Fischer, R. Sarle, G. Vallancien, V. Delmas, M. Hassan, A. Bansal, A. K. Hemal, B. Guillonneau, and M. Menon, "An Operative and Anatomic Study to Help in Nerve Sparing during Laparoscopic and Robotic Radical Prostatectomy," *European Urology*, vol. 43, no. 5, pp. 444–454, May 2003.

[4] R. Fagin, "Da Vinci prostatectomy: Athermal nerve sparing and effect of the technique on erectile recovery and negative margins," *Journal of Robotic Surgery*, vol. 1, no. 2, pp. 139–143, 2007.

[5] C. D'Ettorre, A. Mariani, A. Stilli, F. Rodriguez y Baena, P. Valdastri, A. Deguet, P. Kazanzides, R. H. Taylor, G. S. Fischer, S. P. DiMaio, A. Menciassi, and D. Stoyanov, "Accelerating Surgical Robotics Research: A Review of 10 Years With the da Vinci Research Kit," *IEEE Robotics & Automation Magazine*, vol. 28, no. 4, pp. 56–78, Dec. 2021.

[6] J. Klodmann, C. Schlenk, A. Hellings-Kuß, T. Bahls, R. Unterhinninghofen, A. Albu-Schäffer, and G. Hirzinger, "An Introduction to Robotically Assisted Surgical Systems: Current Developments and Focus Areas of Research," *Current Robotics Reports*, vol. 2, no. 3, pp. 321–332, Sep. 2021.

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

[8] A. Shademan, R. S. Decker, J. Opfermann, S. Leonard, A. Krieger, and P. C. W. Kim, "Supervised autonomous robotic soft tissue surgery," *Science Translational Medicine*, vol. 8, no. 337, 2016.

[9] M. Wartenberg, J. Schornak, P. Carvalho, N. Patel, I. Iordachita, C. Tempany, N. Hata, J. Tokuda, and G. Fischer, "Closed-loop autonomous needle steering during cooperatively controlled needle insertions for MRI-guided pelvic interventions," in *Proc. of the The Hamlyn Symposium on Medical Robotics*, 2017, pp. 33–34.

[10] D. Seita, S. Krishnan, R. Fox, S. McKinley, J. Canny, and K. Goldberg, "Fast and Reliable Autonomous Surgical Debridement with Cable-Driven Robots Using a Two-Phase Calibration Procedure," in *Proc. of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, 2017, pp. 6651–6658.

[11] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, 2016, pp. 4178–4185.

[12] A. Garg, S. Sen, R. Kapadia, Y. Jen, S. McKinley, L. Miller, and K. Goldberg, "Tumor localization using automated palpation with Gaussian Process Adaptive Sampling," in *Proc. of the 2016 IEEE International Conference on Automation Science and Engineering (CASE)*, Fort Worth, 2016, pp. 194–200.

[13] A. Attanasio, B. Scaglioni, M. Leonetti, A. Frangi, W. Cross, C. S. Biyani, and P. Valdastri, "Autonomous Tissue Retraction in Robotic Assisted Minimally Invasive Surgery - A Feasibility Study," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6528–6535, 2020.

[14] G. Q. Wei, K. Arbter, and G. Hirzinger, "Real-time visual servoing for laparoscopic surgery. Controlling robot motion with color image segmentation," *IEEE Engineering in Medicine and Biology Magazine*, vol. 16, no. 1, pp. 40–45, Jan. 1997.

[15] T. Yamamoto, B. Vagvolgyi, K. Balaji, L. L. Whitcomb, and A. M. Okamura, "Tissue property estimation and graphical display for teleoperated robot-assisted surgery," in *Proc. of the 2009 IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, May 2009, pp. 4239–4245.

[16] S.-Y. Chen, "Study on situation awareness for medical robots," National Taipei University of Technology, Technical Report, 2016.

[17] Á. Takács, D. A. Drexler, P. Galambos, I. J. Rudas, and T. Haidegger, "Assessment and Standardization of Autonomous Vehicles," in *Proc. of the IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, Las Palmas de Gran Canaria, Spain, Jun. 2018, pp. 185–192.

[18] D. B. Kaber and M. R. Endsley, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task," *Theoretical Issues in Ergonomics Science*, vol. 5, no. 2, pp. 113–153, Mar. 2004.

[19] H. M. Huang, "Autonomy levels for unmanned systems (ALFUS) Framework Volume II: Framework Models."

[20] ——, "Autonomy Levels for Unmanned Systems (ALFUS) Framework Volume I: Terminology."

[21] M. Delvaux, "Draft Report with rec-ommendations to the Commission on Civil Law Rules on Robotics," 2015/2103(INL), Tech. Rep., 2016.

[22] H. Alemzadeh, J. Raman, N. Leveson, Z. Kalbarczyk, and R. K. Iyer, "Adverse Events in Robotic Surgery: A Retrospective Study of 14 Years of FDA Data," *PLOS ONE*, vol. 11, no. 4, p. e0151470, Apr. 2016.

[23] B. Fei, W. S. Ng, S. Chauhan, and C. K. Kwoh, "The safety issues of medical robotics," p. 10, 2001.

[24] G.-Z. Yang, J. Cambias, K. Cleary, E. Daimler, J. Drake, P. E. Dupont, N. Hata, P. Kazanzides, S. Martel, R. V. Patel, V. J. Santos, and R. H. Taylor, "Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy," *Science Robotics*, vol. 2, no. 4, p. 8638, Mar. 2017.

[25] J. Rosen, "Surgical Robotics," in *Medical Devices: Surgical and Imaging-Guided Technologies*. Springer, 2013, pp. 63–97.

[26] A. Takacs, I. Rudas, D. Bosl, and T. Haidegger, "Highly Automated Vehicles and Self-Driving Cars [Industry Tutorial]," *IEEE Robotics & Automation Magazine*, vol. 25, no. 4, pp. 106–112, 2018.

[27] A. Mohamed, M. Hossny, S. Nahavandi, M. Dalvand, and H. Asadi, "Towards Trusted Autonomous Surgical Robots," in *Proc. of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, 2018, pp. 4083–4088.

[28] F. Richter, R. K. Orosco, and M. C. Yip, "Open-Sourced Reinforcement Learning Environments for Surgical Robotics," *arXiv preprint arXiv:1903.02090*, Mar. 2019.

[29] 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.

[30] R. Ma, "Surgical gestures—An emerging field for surgical assessment and training," *UroPrecision*, Feb. 2025.

[31] A. I. Károly, R. Fullér, and P. Galambos, "Unsupervised Clustering for Deep Learning: A tutorial survey," *Acta Polytechnica Hungarica*, vol. 15, no. 8, pp. 29–53, 2018.

[32] A. Murali, S. Sen, B. Kehoe, A. Garg, S. McFarland, S. Patil, W. D. Boyd, S. Lim, P. Abbeel, and K. Goldberg, "Learning by observation for surgical subtasks: Multilateral cutting of 3D viscoelastic and 2D Orthotropic Tissue Phantoms," in *Proc. of the IEEE International Conference on Robotics and Automation*, Seattle, 2015, pp. 1202–1209.

[33] J. W. Kim, T. Z. Zhao, S. Schmidgall, A. Deguet, M. Kobilarov, C. Finn, and A. Krieger, "Surgical robot transformer (srt): Imitation learning for surgical tasks," 2024.

[34] M. Moghani, L. Doorenbos, W. C.-H. Panitch, S. Huver, M. Azizian, K. Goldberg, and A. Garg, "Sufia: language-guided augmented dexterity for robotic surgical assistants," in *Proc. of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 6969–6976.

[35] T. Huang, K. Chen, B. Li, Y.-H. Liu, and Q. Dou, "Demonstration-Guided Reinforcement Learning with Efficient Exploration for Task Automation of Surgical Robot," in *Proc. of the 2023 IEEE International Conference on Robotics and Automation (ICRA)*, May 2023, pp. 4640–4647.

[36] J. Fu, Y. Long, K. Chen, W. Wei, and Q. Dou, "Multi-objective Cross-task Learning via Goal-conditioned GPT-based Decision Transformers for Surgical Robot Task Automation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, May 2024, pp. 13 362–13 368.

[37] Hyosig Kang and J. T. Wen, "EndoBot: A robotic assistant in minimally invasive surgeries," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2001)*, vol. 2, Seoul, South Korea, May 2001, pp. 2031–2036.

[38] K. A. Nichols and A. M. Okamura, "Autonomous robotic palpation: Machine learning techniques to identify hard inclusions in soft tissues," in *Proc. of the 2013 IEEE International Conference on Robotics and Automation*, Karlsruhe, 2013, pp. 4384–4389.

[39] S. McKinley, A. Garg, S. Sen, D. V. Gealy, J. McKinley, Y. Jen, and K. Goldberg, "Autonomous Multilateral Surgical Tumor Resection with Interchangeable Instrument Mounts and Fluid Injection Device," in *Proc. of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[40] B. Kehoe, G. Kahn, J. Mahler, J. Kim, A. Lee, A. Lee, K. Nakagawa, S. Patil, W. D. Boyd, P. Abbeel, and K. Goldberg, "Autonomous multilateral debridement with the Raven surgical robot," in *Proc. of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, 2014, pp. 1432–1439.

[41] M. Ginesi, D. Meli, A. Roberti, N. Sansonetto, and P. Fiorini, "Autonomous task planning and situation awareness in robotic surgery," in *Proc. of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 3144–3150.

[42] ——, "Dynamic Movement Primitives: Volumetric Obstacle Avoidance Using Dynamic Potential Functions," *Journal of Intelligent & Robotic Systems*, vol. 101, no. 4, pp. 1–20, Apr. 2021.

[43] M. Hwang, B. Thananjeyan, D. Seita, J. Ichnowski, S. Paradis, D. Fer, T. Low, and K. Goldberg, "Superhuman Surgical Peg Transfer Using Depth-Sensing and Deep Recurrent Neural Networks," *arXiv preprint arXiv:2012.12844*, 2020.

[44] M. Hwang, J. Ichnowski, B. Thananjeyan, D. Seita, S. Paradis, D. Fer, T. Low, and K. Goldberg, "Automating Surgical Peg Transfer: Calibration With Deep Learning Can Exceed Speed, Accuracy, and Consistency of Humans," *IEEE Transactions on Automation Science and Engineering*, pp. 1–14, 2022.

[45] S. Paradis, M. Hwang, B. Thananjeyan, J. Ichnowski, D. Seita, D. Fer, T. Low, J. E. Gonzalez, and K. Goldberg, "Intermittent Visual Servoing: Efficiently Learning Policies Robust to Instrument Changes for High-precision Surgical Manipulation," in *Proc. of the 2021 IEEE International Conference on Robotics and Automation (ICRA)*, May 2021, pp. 7166–7173.

[46] J. Xu, B. Li, B. Lu, Y.-H. Liu, Q. Dou, and P.-A. Heng, "SurRoL: An Open-source Reinforcement Learning Centered and dVRK Compatible Platform for Surgical Robot Learning," pp. 1821–1828, Aug. 2021.

[47] D. Zhang, Z. Wu, J. Chen, R. Zhu, A. Munawar, B. Xiao, Y. Guan, H. Su, W. Hong, Y. Guo *et al.*, "Human-robot shared control for surgical robot based on context-aware sim-to-real adaptation," in *Proc. of the 2022 International conference on robotics and automation (ICRA)*. IEEE, 2022, pp. 7694–7700.

[48] P. Fiorini, "Autonomy in Robotic Surgery: The First Baby Steps," in *Proc. of the International Conference on Robotics and Automation (ICRA)*, Montreal, Canada, 2019.

[49] M. M. Rahman, N. Sanchez-Tamayo, G. Gonzalez, M. Agarwal, V. Aggarwal, R. M. Voyles, Y. Xue, and J. Wachs, "Transferring Dexterous Surgical Skill Knowledge between Robots for Semi-autonomous Teleoperation," in *Proc. of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, New Delhi, India, Oct. 2019.

[50] M. Ginesi, N. Sansonetto, and P. Fiorini, "Overcoming Some Drawbacks of Dynamic Movement Primitives," *Robotics and Autonomous Systems*, no. 144, 2021.

[51] G. T. Gonzalez, U. Kaur, M. Rahman, V. Venkatesh, N. Sanchez, G. Hager, Y. Xue, R. Voyles, and J. Wachs, "From the Dexterous Surgical Skill to the Battlefield—A Robotics Exploratory Study," *Military Medicine*, vol. 186, pp. 288–294, Jan. 2021.

[52] T. Nguyen, N. D. Nguyen, F. Bello, and S. Nahavandi, "A New Tensioning Method using Deep Reinforcement Learning for Surgical Pattern Cutting," in *Proc. of the 2019 IEEE International Conference on Industrial Technology (ICIT)*, Melbourne, Australia, Feb. 2019, pp. 1339–1344.

[53] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, and D. D. Yuh, "JHU-ISI gesture and skill assessment working set (JIGSAWS): A surgical activity dataset for human motion modeling," in *Proc. of the MICCAI Workshop: M2CAI*, vol. 3, Boston, 2014.

[54] 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. Chen, "Analysis of the Structure of Surgical Activity for a Suturing and Knot-Tying Task," *PLOS ONE*, vol. 11, no. 3, p. e0149174, 2016.

[55] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[56] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "ROS: An open-source Robot Operating System," in *Proc. of the ICRA Workshop on Open Source Software*, vol. 3, no. 3.2, Kobe, Japan, 2009, pp. 5–11.

[57] 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 *Proc. of the IEEE International Conference on Robotics and Automation*, Hong Kong, 2014, pp. 6434–6439.

[58] Z. Chen, A. Deguet, R. H. Taylor, and P. Kazanzides, "Software Architecture of the Da Vinci Research Kit," in *Proc. of the IEEE International Conference on Robotic Computing (IRC)*, Taichung City, Taiwan, 2017, pp. 180–187.

[59] B. Hannaford, J. Rosen, D. W. Friedman, H. King, P. Roan, L. Cheng, D. Glozman, J. Ma, S. N. Kosari, and L. White, "Raven-II: An Open Platform for Surgical Robotics Research," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 954–959, 2013.

[60] R. M. Rogers and R. H. Taylor, "The Core of a Competent Surgeon: A Working Knowledge of Surgical Anatomy and Safe Dissection Techniques," *Obstetrics and Gynecology Clinics*, vol. 38, no. 4, pp. 777–788, Dec. 2011, publisher: Elsevier.

[61] X.-J. Cai, H.-N. Ying, H. Yu, X. Liang, Y.-F. Wang, W.-B. Jiang, J.-B. Li, and L. Ji, "Blunt Dissection: A Solution to Prevent Bile Duct Injury in Laparoscopic Cholecystectomy," *Chinese Medical Journal*, vol. 128, no. 23, pp. 3153–3157, Dec. 2015. [Online]. Available:

[62] "FLS Manual Skills Written Instructions and Performance Guidelines," https://www.flsprogram.org/wp-content/uploads/2014/03/Revised-Manual-Skills-Guidelines-February-2014.pdf, accessed: Apr. 9, 2025.

[63] O. Tolvanen, A.-P. Elomaa, M. Itkonen, H. Vrzakova, R. Bednarik, and A. Huotarinen, "Eye-Tracking Indicators of Workload in Surgery: A Systematic Review," *Journal of Investigative Surgery*, vol. 35, no. 6, pp. 1340–1349, Jun. 2022, publisher: Taylor & Francis.

[64] L. I. Sgaramella, A. Gurrado, A. Pasculli, N. d. Angelis, R. Memeo, F. P. Prete, S. Berti, G. Ceccarelli, M. Rigamonti, F. G. A. Badessi, N. Solari, M. Milone, F. Catena, S. Scabini, F. Vittore, G. Perrone, C. d. Werra, F. Cafiero, M. Testini, and S. I. C. Group, "The critical view of safety during laparoscopic cholecystectomy: Strasberg Yes or No? An Italian Multicentre study," *Surgical Endoscopy*, vol. 35, no. 7, p. 3698, Aug. 2020.

[65] M. E. Villarreal, C. Rothwell, and E. Huang, "Uncovering patient safety considerations in laparoscopic cholecystectomy using cognitive task analysis," *Surgical Endoscopy*, vol. 37, no. 5, pp. 3921–3925, May 2023.

[66] R. Sznitman, C. Becker, and P. Fua, "Fast Part-Based Classification for Instrument Detection in Minimally Invasive Surgery," in *Proc. of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, vol. 8674. Cham: Springer International Publishing, 2014, pp. 692–699.

[67] M. Allan, P.-L. Chang, S. Ourselin, D. J. Hawkes, A. Sridhar, J. Kelly, and D. Stoyanov, "Image Based Surgical Instrument Pose Estimation with Multi-class Labelling and Optical Flow," in *Proc. of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9349. Cham: Springer International Publishing, 2015, pp. 331–338.

[68] Yuan-Fang Wang, D. R. Uecker, and Wang Yulun, "Choreographed scope manoeuvring in robotically-assisted laparoscopy with active vision guidance," in *Proc. of the Third IEEE Workshop on Applications of Computer Vision. WACV'96*, Sarasota, US, Dec. 1996, pp. 187–192.

[69] A. Krupa, J. Gangloff, C. Doignon, M. F. de Mathelin, G. Morel, J. Leroy, L. Soler, and J. Marescaux, "Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 5, pp. 842–853, Oct. 2003.

[70] D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, D. Stoyanov, M. V. Scarzanella, P. Pratt, and G.-Z. Yang, "Real-Time Stereo Reconstruction in Robotically Assisted Minimally Invasive Surgery," in *Proc. of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, vol. 6361. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 275–282.

[71] J. J. Abbott, P. Marayong, and A. M. Okamura, "Haptic Virtual Fixtures for Robot-Assisted Manipulation," in *Robotics Research*, S. Thrun, R. Brooks, and H. Durrant-Whyte, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, vol. 28, pp. 49–64.

[72] M. Selvaggio, G. A. Fontanelli, F. Ficuciello, L. Villani, and B. Siciliano, "Passive Virtual Fixtures Adaptation in Minimally Invasive Robotic Surgery," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3129–3136, Oct. 2018.

[73] N. Abolhassani, R. Patel, and M. Moallem, "Needle insertion into soft tissue: A survey," *Medical Engineering & Physics*, vol. 29, no. 4, pp. 413–431, 2007.

[74] B. Hannaford, R. Bly, I. Humphreys, and M. Whipple, "Behavior Trees as a Representation for Medical Procedures," *arXiv preprint arXiv:1808.08954*, Aug. 2018.

[75] L. MacKenzie, J. A. Ibbotson, C. G. L. Cao, and A. J. Lomax, "Hierarchical decomposition of laparoscopic surgery: A human factors approach to investigating the operating room environment," *Minimally Invasive Therapy & Allied Technologies*, vol. 10, no. 3, pp. 121–127, 2001.

[76] P. Mascagni, D. Alapatt, L. Sestini, M. S. Altieri, A. Madani, Y. Watanabe, A. Alseidi, J. A. Redan, S. Alfieri, G. Costamagna, I. Boškoski, N. Padoy, and D. A. Hashimoto, "Computer vision in surgery: from potential to clinical value," *npj Digital Medicine*, vol. 5, no. 1, pp. 1–9, Oct. 2022, publisher: Nature Publishing Group.

[77] "ROS package for camera calibration," http://wiki.ros.org/camera_calibration, accessed: Apr. 9, 2025.

[78] "ROS package for stereo image processing," http://wiki.ros.org/stereo_image_proc, accessed: Apr. 9, 2025.

[79] 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, 2014.

[80] F. Liu, Z. Li, Y. Han, J. Lu, F. Richter, and M. C. Yip, "Real-to-sim registration of deformable soft tissue with position-based dynamics for surgical robot autonomy," in *Proc. of the 2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12 328–12 334.

[81] M. W. Spong, S. Hutchinson, M. Vidyasagar *et al.*, *Robot modeling and control*. Wiley New York, 2006, vol. 3.

[82] K. Shoemake, "Animating rotation with quaternion curves," in *Proc. of the 12th annual conference on Computer graphics and interactive techniques*, 1985, pp. 245–254.

[83] K. Strobl and G. Hirzinger, "Optimal Hand-Eye Calibration," in *Proc. of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, Oct. 2006, pp. 4647–4653.

[84] B. Bellekens, V. Spruyt, R. Berkvens, and M. Weyn, "A Survey of Rigid 3D Pointcloud Registration Algorithms," pp. 8–13, 2014.

[85] H. Nghia, "Optimal Rigid/Euclidean transform in 3D space," https://github.com/nghiaho12/rigid_transform_3D, accessed: Apr. 9, 2025.

[86] "Stereo Camera Calibrator App - MATLAB & Simulink," https://www.mathworks.com/help/vision/ref/stereocameracalibrator-app.html, accessed: Apr. 9, 2025.

[87] "GStreamer: Open source multimedia framework," https://gstreamer.freedesktop.org/, accessed: Apr. 9, 2025.

[88] G. Bradski, "The OpenCV library," *Dr. Dobb's Journal of Software Tools*, 2000.

[89] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv preprint arXiv:1801.09847*, 2018.

[90] Y. Bamba, S. Ogawa, M. Itabashi, H. Shindo, S. Kameoka, T. Okamoto, and M. Yamamoto, "Object and anatomical feature recognition in surgical video images based on a convolutional neural network," *International journal of computer assisted radiology and surgery*, vol. 16, no. 11, pp. 2045–2054, 2021.

[91] G. Fontana, M. Matteucci, and D. G. Sorrenti, "Rawseeds: Building a Benchmarking Toolkit for Autonomous Robotics," in *Methods and Experimental Techniques in Computer Engineering*, ser. Springer Briefs in Applied Sciences and Technology, F. Amigoni and V. Schiaffonati, Eds., Cham, 2014, pp. 55–68.

[92] P. Kazanzides and G. Fischer, "AccelNet: International Collaboration on Data Collection and Machine Learning," Jun. 2020.

[93] I. Nigicser, B. Szabo, L. Jaksa, D. A. Nagy, T. Garamvolgyi, S. Barcza, P. Galambos, and T. Haidegger, "Anatomically relevant pelvic phantom for surgical simulation," in *Proc. of the 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Wroclaw, Poland, Oct. 2016, pp. 427–432.

[94] A. A. Mohammed and S. H. Arif, "Midline gallbladder makes a challenge for surgeons during laparoscopic cholecystectomy; case series of 6 patients," *Annals of Medicine and Surgery*, vol. 40, pp. 14–17, Apr. 2019.

[95] P. Kazanzides, G. Fichtinger, G. D. Hager, A. M. Okamura, L. L. Whitcomb, and R. H. Taylor, "Surgical and Interventional Robotics - Core Concepts, Technology, and Design [Tutorial]," *IEEE Robotics & Automation Magazine*, vol. 15, no. 2, pp. 122–130, Jun. 2008.

[96] T. Haidegger, "Taxonomy and Standards in Robotics," in *Encyclopedia of Robotics*, M. H. Ang., O. Khatib, and B. Siciliano, Eds. Berlin, Heidelberg: Springer Nature, 2022, pp. 1–10.

[97] K. Chinzei, "Safety of Surgical Robots and IEC 80601-2-77: The First International Standard for Surgical Robots," *Acta Polytechnica Hungarica*, vol. 16, no. 8, pp. 174–184, Sep. 2019.

[98] M. Endsley, "Situation awareness global assessment technique (SAGAT)," in *Proc. of the IEEE 1988 National Aerospace and Electronics Conference*, Dayton, OH, USA, 1988, pp. 789–795.

[99] ——, "Situation Awareness in Aviation Systems," in *Handbook of Aviation Human Factors*. CRC Press, 1999, pp. 257–276.

[100] R. Nagyné Elek and T. Haidegger, "Robot-Assisted Minimally Invasive Surgical Skill Assessment—Manual and Automated Platforms," *Acta Polytechnica Hungarica, Special Issue on Platforms for Medical Robotics Research*, vol. 16, no. 8, pp. 141–169, Sep. 2019.

[101] J. H. Nguyen, J. Chen, S. P. Marshall, S. Ghodoussipour, A. Chen, I. S. Gill, and A. J. 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, Jul. 2020.

[102] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surgical Endoscopy*, vol. 25, no. 2, pp. 356–366, Feb. 2011.

[103] 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," *Endoscopy International Open*, vol. 6, no. 8, pp. E1065–E1069, Aug. 2018.

[104] 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," *American Journal of Obstetrics and Gynecology*, vol. 215, no. 5, Nov. 2016.

[105] A. Joshi, S. Kale, S. Chandel, and D. Pal, "Likert Scale: Explored and Explained," *British Journal of Applied Science & Technology*, vol. 7, no. 4, pp. 396–403, Jan. 2015.

[106] N. Raison, K. Ahmed, N. Fossati, N. Buffi, A. Mottrie, P. Dasgupta, and H. V. D. Poel, "Competency based training in robotic surgery: Benchmark scores for virtual reality robotic simulation," *BJU International*, vol. 119, no. 5, pp. 804–811, 2017.

[107] 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, Jan. 2021.

[108] J. C. Kwong, J. Y. Lee, and M. G. Goldenberg, "Understanding and Assessing Non-technical Skills in Robotic Urological Surgery: A Systematic Review and Synthesis of the Validity Evidence," *Journal of Surgical Education*, vol. 76, no. 1, pp. 193–200, Jan. 2019.

[109] A. L. Trejos, R. V. Patel, R. A. Malthaner, and C. M. Schlachta, "Development of force-based metrics for skills assessment in minimally invasive surgery," *Surgical Endoscopy*, vol. 28, no. 7, pp. 2106–2119, Jul. 2014.

[110] M. Hwang, D. Seita, B. Thananjeyan, J. Ichnowski, S. Paradis, D. Fer, T. Low, and K. Goldberg, "Applying Depth-Sensing to Automated Surgical Manipulation with a da Vinci Robot," in *Proc. of the 2020 International Symposium on Medical Robotics (ISMR)*, Nov. 2020, pp. 22–29.

[111] B. Lu, B. Li, W. Chen, Y. Jin, Z. Zhao, Q. Dou, P.-A. Heng, and Y. Liu, "Toward Image-Guided Automated Suture Grasping Under Complex Environments: A Learning-Enabled and Optimization-Based Holistic Framework," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, pp. 1–15, 2021.

[112] V. Datta, M. Mandalia, S. Mackay, A. Chang, N. Cheshire, and A. Darzi, "Relationship between skill and outcome in the laboratory-based model," *Surgery*, vol. 131, no. 3, pp. 318–323, Mar. 2002.

[113] F. Cavallo, G. Megali, S. Sinigaglia, O. Tonet, and P. Dario, "A biomechanical analysis of surgeon's gesture in a laparoscopic virtual scenario." *Medicine meets virtual reality. Studies in health technology and informatics*, vol. 119, pp. 79–84, 2006.

[114] I. Rivas-Blanco, C. J. P. Del-Pulgar, A. Mariani, G. Tortora, and A. J. Reina, "A surgical dataset from the da vinci research kit for task automation and recognition," in *Proc. of the 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. IEEE, 2023, pp. 1–6.

[115] E. Boyle, M. Al-Akash, A. G. Gallagher, O. Traynor, A. D. K. Hill, and P. C. Neary, "Optimising surgical training: Use of feedback to reduce errors during a simulated surgical procedure," *Postgraduate Medical Journal*, vol. 87, no. 1030, pp. 524–528, Aug. 2011.

[116] S. Cotin, N. Stylopoulos, M. Ottensmeyer, P. Neumann, R. Bardsley, and S. Dawson, "Surgical training system for laparoscopic procedures," US Patent US20 050 142 525A1, Jun., 2005.

[117] B. Rohrer, S. Fasoli, H. I. Krebs, R. Hughes, B. Volpe, W. R. Frontera, J. Stein, and N. Hogan, "Movement Smoothness Changes during Stroke Recovery," *Journal of Neuroscience*, vol. 22, no. 18, pp. 8297–8304, Sep. 2002.

[118] K. Takada, K. Yashiro, and M. Takagi, "Reliability and sensitivity of jerk-cost measurement for evaluating irregularity of chewing jaw movements," *Physiological Measurement*, vol. 27, no. 7, pp. 609–622, Apr. 2006.

[119] L. Moody, C. Baber, T. N. Arvanitis, and M. Elliott, "Objective Metrics for the Evaluation of Simple Surgical Skills in Real and Virtual Domains," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 2, pp. 207–221, Apr. 2003.

[120] T. Li, C. Shi, and H. Ren, "A High-Sensitivity Tactile Sensor Array Based on Fiber Bragg Grating Sensing for Tissue Palpation in Minimally Invasive Surgery," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 5, pp. 2306–2315, Oct. 2018.

[121] W.-J. Jung, K.-S. Kwak, and S.-C. Lim, "Vision-Based Suture Tensile Force Estimation in Robotic Surgery," *Sensors*, vol. 21, no. 1, p. 110, Dec. 2020.

[122] A. L. Trejos, R. V. Patel, M. D. Naish, A. C. Lyle, and C. M. Schlachta, "A Sensorized Instrument for Skills Assessment and Training in Minimally Invasive Surgery," *Journal of Medical Devices*, vol. 3, no. 4, Nov. 2009.

[123] D. Jones, H. Wang, A. Alazmani, and P. R. Culmer, "A soft multi-axial force sensor to assess tissue properties in RealTime," Sep. 2017, pp. 5738–5743.

[124] T. Osa, N. Sugita, and M. Mitsuishi, "Online Trajectory Planning in Dynamic Environments for Surgical Task Automation." Robotics: Science and Systems Foundation, Jul. 2014, pp. 1–9.

[125] B. Lu, H. K. Chu, and L. Cheng, "Robotic knot tying through a spatial trajectory with a visual servoing system," in *Proc. of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 5710–5716.

[126] T. Haidegger, *Enhancing Computer-Integrated SurgicalS Ystems—A Control Engineering Approach*. Saarbrucken: LAP LAMBERT Academic Publishing, 2011.

[127] S. A. Pedram, C. Shin, P. W. Ferguson, J. Ma, E. P. Dutson, and J. Rosen, "Autonomous Suturing Framework and Quantification Using a Cable-Driven Surgical Robot," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 404–417, Apr. 2021.

[128] T. Haidegger, P. Kazanzides, I. Rudas, B. Benyó, and Z. Benyó, "The importance of accuracy measurement standards for computer-integrated interventional systems," in *Proc. of the EURON GEM Sig Workshop on the Role of Experiments in Robotics Research at IEEE ICRA*, 2010.

[129] P. Fiorini, "Automation and Autonomy in Robotic Surgery," in *Robotic Surgery*. Cham: Springer International Publishing, 2021, pp. 237–255.

[130] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, Jun. 2012, pp. 3354–3361.

[131] D. Kondermann, R. Nair, K. Honauer, K. Krispin, J. Andrulis, A. Brock, B. Gussefeld, M. Rahimimoghaddam, S. Hofmann, C. Brenner, and B. Jahne, "The HCI Benchmark Suite: Stereo and Flow Ground Truth with Uncertainties for Urban Autonomous Driving," in *Proc. of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA, Jun. 2016, pp. 19–28.

[132] X. Song, P. Wang, D. Zhou, R. Zhu, C. Guan, Y. Dai, H. Su, H. Li, and R. Yang, "ApolloCar3D: A Large 3D Car Instance Understanding Benchmark for Autonomous Driving," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5452–5462.

[133] R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca, and M. A. Sotelo, "The PREVENTION dataset: A novel benchmark for PREdiction of VEhicles iNTentIONs," in *Proc. of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct. 2019, pp. 3114–3121.

[134] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. of the Conference on Robot Learning*, 2017, pp. 1–16.

[135] B. Osinski, P. Milos, A. Jakubowski, P. Ziecina, M. Martyniak, C. Galias, A. Breuer, S. Homoceanu, and H. Michalewski, "CARLA Real Traffic Scenarios – novel training ground and benchmark for autonomous driving," *arXiv preprint arXiv:2012.11329*, 2020.

[136] R. Chen, M. Arief, W. Zhang, and D. Zhao, "How to Evaluate Proving Grounds for Self-Driving? A Quantitative Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5737–5748, 2020.

[137] S. K. Hasan and C. A. Linte, "U-netplus: A modified encoder-decoder u-net architecture for semantic and instance segmentation of surgical instruments from laparoscopic images," in *Proc. of the 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2019, pp. 7205–7211.

[138] A. Zia, K. Bhattacharyya, X. Liu, Z. Wang, S. Kondo, E. Colleoni, B. van Amsterdam, R. Hussain, R. Hussain, L. Maier-Hein, D. Stoyanov, S. Speidel, and A. Jarc, "Surgical Visual Domain Adaptation: Results from the MICCAI 2020 SurgVisDom Challenge," *arXiv preprint arXiv:2102.13644*, 2021.

[139] A. Munawar, Y. Wang, R. Gondokaryono, and G. S. Fischer, "A Real-Time Dynamic Simulator and an Associated Front-End Representation Format for Simulating Complex Robots and Environments," in *Proc. of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, Nov. 2019, pp. 1875–1882.

[140] A. Munawar, J. Y. Wu, G. S. Fischer, R. H. Taylor, and P. Kazanzides, "Open Simulation Environment for Learning and Practice of Robot-Assisted Surgical Suturing," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3843–3850, Apr. 2022.

[141] K. Takacs, K. Moga, and T. Haidegger, "Sensorized Psychomotor Skill Assessment Platform Built on a Robotic Surgery Phantom," in *Proc. of the IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Herlany, Slovakia, Jan. 2020, pp. 95–100.

[142] J. Fernandez-Lozano, J. de Gabriel, V. Munoz, I. Garcia-Morales, D. Melgar, C. Vara, and A. Garcia-Cerezo, "Human-machine interface evaluation in a computer assisted surgical system," in *Proc. of the IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, vol. 1, Apr. 2004, pp. 231–236.

[143] Z. Li, A. Gordon, T. Looi, J. Drake, C. Forrest, and R. H. Taylor, "Anatomical Mesh-Based Virtual Fixtures for Surgical Robots," in *Proc. of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Jul. 2020, pp. 3267–3273.

[144] J. W. Baker, M. Schubert, and M. H. Faber, "On the assessment of robustness," *Structural Safety*, vol. 30, no. 3, pp. 253–267, May 2008.

[145] B. Meyer, *Object-Oriented Software Construction*. Englewood Cliffs: Prentice hall, 1997.

[146] P. J. Huber, *Robust Statistics*. John Wiley & Sons, 2004, vol. 523.

[147] J.-J. E. Slotine, W. Li *et al.*, *Applied Nonlinear Control*. Prentice hall Englewood Cliffs, NJ, 1991, vol. 199.

[148] C. Hutchison, M. Zizyte, P. E. Lanigan, D. Guttendorf, M. Wagner, C. L. Goues, and P. Koopman, "Robustness testing of autonomy software," in *Proc. of the 40th International Conference on Software Engineering: Software Engineering in Practice*, Gothenburg Sweden, May 2018, pp. 276–285.

[149] F. Rudzicz and R. Saqur, "Ethics of Artificial Intelligence in Surgery," *arXiv preprint arXiv:2007.14302*, 2020.

[150] D. B. O. Boesl and M. Bode, "Signaling Sustainable Robotics – a Concept to Implement the Idea of Robotic Governance," in *Proc. of the IEEE 23rd International Conference on Intelligent Engineering Systems (INES)*, Apr. 2019, pp. 143–146.

[151] T. Litman, *Autonomous Vehicle Implementation Predictions: Implications for Transport Planning*. Victoria Transport Policy Institute, 2019.

[152] *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, Std. J3016_202 104, April 2021.

[153] R. Parasuraman, T. Sheridan, and C. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, May 2000.

[154] D. Maggi, R. Romano, and O. Carsten, "Transitions Between Highly Automated and Longitudinally Assisted Driving: The Role of the Initiator in the Fight for Authority," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, p. 001872082094618, Aug. 2020.

[155] N. Du, J. Kim, F. Zhou, E. Pulver, D. M. Tilbury, L. P. Robert, A. K. Pradhan, and X. J. Yang, "Evaluating Effects of Cognitive Load, Takeover Request Lead Time, and Traffic Density on Drivers' Takeover Performance in Conditionally Automated Driving," in *Proc. of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2020, pp. 66–73.

[156] V. A. Banks, K. L. Plant, and N. A. Stanton, "Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016," *Safety Science*, vol. 108, pp. 278–285, Oct. 2018.

[157] S. S. Banerjee, S. Jha, J. Cyriac, Z. T. Kalbarczyk, and R. K. Iyer, "Hands Off the Wheel in Autonomous Vehicles?: A Systems Perspective on over a Million Miles of Field Data," in *Proc. of the 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Luxembourg City, Jun. 2018, pp. 586–597.

[158] T. Tamás and K. Szabó, "Combined Mathematical Modeling of Different Transport Networks, Considerations and Complex Analysis," *Acta Polytechnica Hungarica*, vol. 14, no. 2, pp. 7–26, May 2017.

[159] G. Weller and B. Schlag, "Behavioral Aspects of Driver Assistance Systems," in *Handbook of Driver Assistance Systems: Basic Information, Components and Systems for Active Safety and Comfort*, H. Winner, S. Hakuli, F. Lotz, and C. Singer, Eds. Cham: Springer International Publishing, 2016, pp. 91–107.

[160] R. Palin, D. Ward, I. Habli, and R. Rivett, "ISO 26262 safety cases: Compliance and assurance," in *Proc. of the 6th IET International Conference on System Safety 2011*, Birmingham, UK, 2011, pp. B12–B12.

[161] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human factors*, vol. 37, no. 1, pp. 32–64, 1995.

[162] P. M. Salmon, N. A. Stanton, G. H. Walker, D. Jenkins, D. Ladva, L. Rafferty, and M. Young, "Measuring Situation Awareness in complex systems: Comparison of measures study," *International Journal of Industrial Ergonomics*, vol. 39, no. 3, pp. 490–500, May 2009.

[163] M. Walch, K. Lange, M. Baumann, and M. Weber, "Autonomous driving: Investigating the feasibility of car-driver handover assistance," in *Proc. of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*. Nottingham, United Kingdom: ACM Press, 2015, pp. 11–18.

[164] P. Morgan, C. Alford, and G. Parkhurst, "Handover issues in autonomous driving: A literature review," University of the West of England, Bristol, Project Report, 2016.

[165] K. Saleh, M. Hossny, and S. Nahavandi, "Towards trusted autonomous vehicles from vulnerable road users perspective," in *Proc. of the 2017 Annual IEEE International Systems Conference (SysCon)*, Montreal, QC, Canada, Apr. 2017, pp. 1–7.

[166] P. M. Salmon, N. A. Stanton, and G. H. Walker, "Distributed situation awareness and vehicle automation: Case study analysis and design implications," in *Handbook of human factors for automated, connected, and intelligent vehicles*. CRC Press, 2020, pp. 293–317.

[167] A. Eriksson and N. A. Stanton, "Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and From Manual Control," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 59, no. 4, pp. 689–705, Jun. 2017.

[168] M. R. Endsley, "Measurement of situation awareness in dynamic systems," *Human factors*, vol. 37, no. 1, pp. 65–84, 1995.

[169] R. M. Taylor, "Situational awareness rating technique (sart): The development of a tool for aircrew systems design," in *Situational awareness*. Routledge, 2017, pp. 111–128.

[170] M. R. Endsley and D. J. Garland, *Situation awareness analysis and measurement*. CRC press, 2000.

[171] P. Salmon, N. Stanton, G. Walker, and D. Green, "Situation awareness measurement: A review of applicability for c4i environments," *Applied ergonomics*, vol. 37, no. 2, pp. 225–238, 2006.

[172] N. Hogan, "Impedance Control: An Approach to Manipulation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 17, Mar. 1985.

[173] A. D'Ausilio, "Arduino: A low-cost multipurpose lab equipment," *Behavior Research Methods*, vol. 44, no. 2, pp. 305–313, Jun. 2012.

[174] T. K. Kim, "T test as a parametric statistic," *Korean journal of anesthesiology*, vol. 68, no. 6, pp. 540–546, 2015.

[175] K. Pearson, "Vii. mathematical contributions to the theory of evolution.—iii. regression, heredity, and panmixia," *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, no. 187, pp. 253–318, 1896.

[176] C. D. Wickens, "Situation awareness: Review of mica endsley's 1995 articles on situation awareness theory and measurement," *Human factors*, vol. 50, no. 3, pp. 397–403, 2008.

[177] J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[178] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics-theory and Methods*, vol. 6, no. 9, pp. 813–827, 1977.

[179] H. E. Russell, L. K. Harbott, I. Nisky, S. Pan, A. M. Okamura, and J. C. Gerdes, "Motor learning affects car-to-driver handover in automated vehicles," *Science Robotics*, vol. 1, no. 1, 2016.

PUBLICATIONS RELATED TO THE THESIS

[LT1] 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 *Proc. of the IEEE 21st International Conference on Intelligent Engineering Systems (INES)*, Larnaca, 2017, pp. 253–258.

[LT2] T. D. Nagy, M. Takács, I. J. Rudas, and T. Haidegger, "Surgical Subtask Automation—Soft Tissue Retraction," in *Proc. of the 16th IEEE World Symposium on Applied Machine Intelligence and Informatics*, Kosice, 2018, pp. 55–60.

[LT3] T. D. Nagy and T. Haidegger, "An Open-Source Framework for Surgical Subtask Automation," in *Proc. of the 2018 IEEE International Conference on Robotics and Automation ICRA Workshop on Supervised Autonomy in Surgical Robotics*, Brisbane, 2018.

[LT4] ——, "A DVRK-based Framework for Surgical Subtask Automation," *Acta Polytechnica Hungarica, Special Issue on Platforms for Medical Robotics Research*, vol. 16, no. 8, pp. 61–78, 2019.

[LT5] ——, "Autonomous Surgical Robotics at Task and Subtask Levels," in *Advanced Robotics and Intelligent Automation in Manufacturing*, ser. Advances in Computational Intelligence and Robotics (ACIR) Book Series. IGI Global, 2020, pp. 296–319.

[LT6] ——, "Autonomous Peg Transfer—a Gateway to Surgery 4.0," in *Proc. of the IEEE 10th Jubilee International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC 2022)*, Reykjavík, Iceland, Jul. 2022, pp. 69–76.

[LT7] ——, "Towards Standard Approaches for the Evaluation of Autonomous Surgical Subtask Execution," in *Proc. of the 25th IEEE International Conference on Intelligent Engineering Systems (INES 2021)*, Budapest, Hungary, Jul. 2021, pp. 67–74.

[LT8] ——, "Performance and Capability Assessment in Surgical Subtask Automation," *Sensors*, vol. 22, no. 7, p. 2501, Jan. 2022.

[LT9] T. D. Nagy, N. Ukhrenkov, D. A. Drexler, Á. Takács, and T. Haidegger, "Enabling quantitative analysis of situation awareness: System architecture for autonomous vehicle handover studies," in *Proc. of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy, 2019, pp. 914–918.

[LT10] T. D. Nagy, N. Uhrenkov, D. A. Drexler, Á. Takács, and T. Haidegger, “System Architecture for Situation Awareness Quantification Employing the Da Vinci Console,” in *Proc. of the 2019 International Conference on Robotics and Automation (ICRA) Workshop on ”Open Challenges and State-of-the-Art in Control System Design and Technology Development for Surgical Robotic Systems”*, Montreal, Canada, 2019.

[LT11] T. D. Levendovics, D. A. Drexler, N. Ukhrenkov, Á. Takács, and T. Haidegger, “Quantitative Analysis of Situation Awareness at Autonomous Vehicle Handover on the Da Vinci Research Kit,” *Sensors*, 2025, submitted.

OTHER PUBLICATIONS

[LTNR1] R. Elek, T. D. Nagy, D. Á. Nagy, G. Kronreif, I. J. Rudas, and T. Haidegger, "Recent Trends in Automating Robotic Surgery," in *Proc. of the 20th IEEE Jubilee International Conference on Intelligent Engineering Systems*, Budapest, 2016, pp. 27–32.

[LTNR2] R. Elek, T. D. Nagy, D. Á. Nagy, B. Takács, P. Galambos, I. J. Rudas, and T. Haidegger, "Robotic Platforms for Ultrasound Diagnostics and Treatment," in *Proc. of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Banff, 2017, pp. 1752–1757.

[LTNR3] D. Nagy, T. D. Nagy, R. Elek, I. J. Rudas, and T. Haidegger, "Ontology Based Surgical Subtask Automation," in *Proc. of the 2017 IEEE International Conference on Robotics and Automation ICRA Workshop C4 Surgical Robots: Compliant, Continuum, Cognitive, and Collaborative*, Singapore, 2017.

[LTNR4] T. D. Nagy, G. Vadai, and Z. Gingl, "Digital phonocardiographic experiments and signal processing in multidisciplinary fields of university education," *European Journal of Physics*, vol. 38, no. 5, 2017.

[LTNR5] 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. 3, no. 3, 2018.

[LTNR6] D. A. Drexler, Á. Takács, T. D. Nagy, and T. Haidegger, "Handover Process of Autonomous Vehicles –Technology and Application Challenges," *Acta Polytechnica Hungarica*, vol. 16, no. 9, pp. 235–255, 2019.

[LTNR7] ——, "HandoverProcess Modeling in AutonomousVehicles—a Control Engineering Approach," in *Proc. of the IEEE Joint 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics : CINTI-MACRo 2019*, Szeged, Hungary, 2019, pp. 155–160.

[LTNR8] D. A. Drexler, Á. Takács, T. D. Nagy, P. Galambos, I. Rudas, and T. Haidegger, "Situation Awareness and System Trust Affecting Handover Processes in Self-Driving Cars up to Level 3 Autonomy," in *Proc. of the IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, Budapest, Hungary, 2019, pp. 179–184.

[LTNR9] T. D. Nagy and T. Haidegger, “Recent Advances in Robot-Assisted Surgery: Soft Tissue Contact Identification,” in *Proc. of the 13th International Symposium on Applied Computational Intelligence and Informatics (SACI 2019)*, Timisoara, Romania, 2019, pp. 99–105.

[LTNR10] Á. Takács, D. A. Drexler, T. D. Nagy, and T. Haidegger, “Handover Process of Autonomous Driver Assist Systems—a call for critical performance assessment,” in *Proc. of the 2019 IEEE/RSJ International Conference on Intelligent Robots And Systems (IROS)*, 2019, p. 3390.

[LTNR11] C. Molnár, T. D. Nagy, R. N. Elek, and T. Haidegger, “Visual servoing-based camera control for the da Vinci Surgical System,” in *Proc. of the IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY)*, Sep. 2020, pp. 107–112.