UNIVERSITY OF CALIFORNIA SAN DIEGO

Machine Intelligence for Performant Surgical Robotics

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science (Computer Engineering)

by

Christopher Dambrosia

Committee in charge:

Professor Henrik I. Christensen, Chair
Professor Eliah Aronoff-Spencer
Professor Virginia De Sa
Professor Linda Hill
Professor Nadir Weibel

2023

The Dissertation of Christopher Dambrosia is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

I dedicate this work to my family, my friends, and all of my teachers who gifted me with the curiosity, love, and inspiration to keep me fascinated during this long and challenging but fulfilling journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I thank my advisor, Professor Henrik Christensen, for his continuous support and mentorship. He gave me the creative space, intellectual freedom, and guidance to succeed as a PhD student.

I thank Professor Eliah Aronoff-Spencer for his unwavering belief at difficult times during my studies and research, his support that allowed me to remain at student at UCSD, his mentorship and guidance throughout my doctoral program, and perhaps most importantly, the moonlit paddle-boarding sessions during the pandemic that meant more to me than he knows.

I thank Professor Linda Hill for being a beacon of idealism and compassion, her enthusiasm for scholarly collaborations, her support of my medical-engineering journey, and her boundless energy for intellectual discovery in the service of public health and the alleviation of suffering in the lives of others.

I thank my committee members, Professor Nadir Weibel and Professor Virginia De Sa for their encouragement, advice, and insights.

I express my deep gratitude to Professor Laurel Riek for giving me the opportunity to become a student in the doctoral program at UCSD.

I am immensely thankful for the efforts of my co-authors at the Center for the Future of Surgery: Dr. Estella Huang, Dr. Nicole Goldhaber, Dr. Ryan Broderick, and Dr. Lawrence Appelbaum. I am similarly grateful for the work done by my co-authors in the ARCLab: Dr. Florian Richter, Nikhil U. Shinde, Zih-Yun (Sarah) Chiu, Dr. Fei Liu, and Professor Michael Yip.

Lastly, I thank all the members of the Cognitive Robotics Lab, the ARCLab, LegoLand and the Contextual Robotics Institute with whom I had the pleasure of sharing time at UCSD, including:

Dr. Jacob Johnson, Flora Liu, Srinidhi Bharadwaj Kalgundi Srinivas, Sander Tonkens, Dr. Dimitri Schreiber, Elizabeth Peiros, Dr. Priyam Parashar, Dr. Vikas

Dhiman, Pol Mestres, Nicklas Hansen, Seth Farrell, Dr. David Paz-Ruiz, Hengyuan Zhang, Jiaming Hu, Anwesan Pal, Andi Frank, Jing-Yan Liao, Rohan Patil, Luobin Yang, Shrutheesh Raman Iyer.

Chapter 1, in part, is a reprint of material from:

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R., Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C., 2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Richter, F., Aronoff-Spencer, E., Broderick, R.C. and Christensen, H., 2023. SCALPEL Please: Personalized, Performance-Oriented Operator Models Using Biometrics and Intra-task Error Detection During Robot-Assisted Surgery. IEEE Robotics and Automation Letters, (in submission).

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 2, in part, is a reprint of material from:

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen,

H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R., Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C., 2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Richter, F., Aronoff-Spencer, E., Broderick, R.C. and Christensen, H., 2023. SCALPEL Please: Personalized, Performance-Oriented Operator Models Using Biometrics and Intra-task Error Detection During Robot-Assisted Surgery. IEEE Robotics and Automation Letters, (in submission).

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 3, in part, is a reprint of material from:

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R., Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C.,

2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Richter, F., Aronoff-Spencer, E., Broderick, R.C. and Christensen, H., 2023. SCALPEL Please: Personalized, Performance-Oriented Operator Models Using Biometrics and Intra-task Error Detection During Robot-Assisted Surgery. IEEE Robotics and Automation Letters, (in submission).

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 4, in part, is a reprint of material from:

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 5, in part, is a reprint of material from:

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative

error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R., Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C., 2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Richter, F., Aronoff-Spencer, E., Broderick, R.C. and Christensen, H., 2023. SCALPEL Please: Personalized, Performance-Oriented Operator Models Using Biometrics and Intra-task Error Detection During Robot-Assisted Surgery. IEEE Robotics and Automation Letters, (in submission).

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

VITA

2008	Bachelor of Arts in English Literature, Princeton University
2019	Masters of Science in Computer Science, Columbia University
2023	Doctor of Philosophy in Computer Science (Computer Engineering), University of California San Diego

PUBLICATIONS

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R., Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C., 2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Christensen, H., Aronoff-Spencer, E. (2020). Computing SARS-CoV-2 Infection Risk From Symptoms, Imaging, and Test Data: Diagnostic Model Development. Journal of Medical Internet Research, 22(12), e24478.

Nicoli F, D'Ambrosia C, Lazzeri D, et al. Microsurgical dissection of facial nerve in parotidectomy: a discussion of techniques and long-term results. Gland Surg. 2017;6(4):308-314.

Li K, Xia L, Liu NF, Constantinides J, D'Ambrosia C, et al. Far infrared ray (FIR) therapy: An effective and oncological safe treatment modality for breast cancer related lymphedema. J Photochem Photobiol B, Biol. 2017;172:95-101.

Fu MC, D'Ambrosia C, Mclawhorn AS, Schairer WW, Padgett DE, Cross MB. Malnutrition Increases With Obesity and Is a Stronger Independent Risk Factor for Postoperative Complications: A Propensity-Adjusted Analysis of Total Hip Arthroplasty Patients. J Arthroplasty. 2016;31(11):2415-2421.

Nicoli F, Ciudad P...D'Ambrosia C, et al. Potential Use of Transferred Lymph Nodes as Metastasis Detectors after Tumor Excision. Arch Plast Surg. 2015;42(4):478-83.

McKillop M, Gorman S, Goli S, D'Ambrosia C, Knoll C, Ryan P, Elhadad N. Endometriosis Phenotype Development, Validation, and Characterization from

Observational Health Databases. Observational Health Data Sciences and Informatics (OHDSI) Symposium 2018; 2018 Oct 11- 13; Bethesda, MD.

Saliaj M, Sedaliu T, D'Ambrosia C, Khan-Manji S. A Rare Case of Drug-Induced Liver Injury Following Prophylactic Doses of Subcutaneous Unfractionated Heparin. New York American College of Physicians Resident and Medical Student Forum; 2017 Oct 28; Staten Island, NY.

ABSTRACT OF THE DISSERTATION

Machine Intelligence for Performant Surgical Robotics

by

Christopher Dambrosia

Doctor of Philosophy in Computer Science (Computer Engineering)

University of California San Diego, 2023

Professor Henrik I. Christensen, Chair

Intelligent algorithms for measuring and augmenting performance during robotic-assisted surgery (RAS) in both human-robot collaborative and autonomous system settings have the potential to benefit both surgeons and patients. Successful RAS depends on both human operator and robotic system performance. Measuring performance requires integrating, synchronizing and analyzing contemporaneous data from humans, robots, and task environments. Safety-critical tasks in dynamic, unstructured environments, such as RAS, require both high performing operators and robotic systems. Surgeons operate in mentally and physically demanding workspaces where the impact of error is highly consequential, and uncertainties in operating room

(OR) robotic systems, particularly in kinematics and perception for autonomous applications, have meaningful implications for clinical outcomes.

The purpose of this dissertation is to develop novel machine intelligence algorithms to quantitatively model and augment performance during robot-assisted surgery for both human operators and autonomous systems. For human operators, we analyze operator biometric data to detect and predict intraoperative error. For autonomous systems, we use perception algorithms to measure tool localization accuracy and visual scene uncertainty in surgical environments. Our results show that we can detect and predict intraoperative human error using a set of biometric features and that our perception algorithms can more accurately localize surgical tools and measure visual scene uncertainty in surgical environments.

Chapter 1

Introduction and Background

The purpose of this dissertation is to develop novel machine intelligence algorithms to quantitatively model and augment performance during robot-assisted surgery for both human operators and autonomous systems. For human operators, we analyze operator biometric data to detect and predict intraoperative error. For autonomous systems, we use perception algorithms to measure tool localization accuracy and visual scene uncertainty in surgical environments. Our results show that we can detect and predict intraoperative human error using a set of biometric features and that our perception algorithms can more accurately localize surgical tools and measure visual scene uncertainty in surgical environments.

Effective human-robot collaboration depends on both human operator and robotic system performance. Measuring performance requires integrating, synchronizing and analyzing contemporaneous data from humans, robots, and task environments. Safety-critical tasks in dynamic, unstructured environments, such as robotic-assisted surgery (RAS), require both high performing operators and robotic systems. Surgeons operate in mentally and physically demanding workspaces where the impact of error is highly consequential [102] [30], and uncertainties in operating room (OR) robotic systems, particularly in kinematics and perception for autonomous applications, have meaningful implications for clinical outcomes. Intelligent algorithms for measuring and augmenting

performance during RAS have the potential to benefit both surgeons and patients.

Both human operator and robotic platform and performance are critical to task success [70] during RAS but understanding operator-specific and robotic platform-specific contributions to task performance remains a challenge [35]. Automated task performance analysis requires selecting accurate markers of intra-task progress and detecting these markers in rapidly changing environments. Identifying operator-specific contributors to task performance requires examining correspondences between task performance and biometric signals. Identifying platform-specific contributors to task performance requires associating task performance with platform attributes such as dexeme (surgical motion primitives) difficulty [157] or measuring uncertainty in surgical tool localization and surgical image representations.

1.1 Automated Surgical Performance Analysis

Automated objective error detection at a high temporal resolution is required to accurately associate objective surgical errors with operator-specific data or robotic platform-specific data. These associations are necessary for building performance-oriented operator models and modelling robotic system uncertainties associated with surgical errors.

Prior work in automated surgical performance analysis has primarily focused on kinematic, system event, and haptic data obtained from RAS platforms. This data is then used to find associations between features such as instrument traveling distance (kinematics), camera clutch engagement (system event), and grip force (haptics) and self-reported level of surgical experience [27]. Other research has used surgical video analysis for competency assessment [161] [125] [150]. An alternative approach is use of automated surgeme (surgical gesture) recognition and classification for proficiency labeling or error detection [156].

However, due to limitations in objective task-environment analysis at high temporal resolution, the unit of analysis for most prior research in RAS performance is often an entire procedure or task rather than milliseconds to seconds within a procedure. Additionally, rather than using actual surgical error, prior work often utilizes features such as instrument travel distance or camera clutch engagement known to be proxies for error rather than actual error. Where actual intraoperative error detection is studied, research often relies on retrospective, manual annotations by expert surgeons [27]. These manual annotations provide sub-task level resolution for error detection but sub-tasks within surgical procedures can take seconds to minutes. A sub-task that is labeled with an error then creates an imprecise error marker of seconds to minutes in length and results in a low temporal resolution error signal.

1.2 Operator-specific Models

Building accurate, performance-oriented models of human operators requires selecting operator characteristics of interest, accurately measuring these characteristics, and correlating these characteristics with objective markers of task performance. Across a wide range of studies, the examined operator characteristics are often physiological signals due to their correlation with operator cognitive and affective states as well as various aspects of human performance [106, 55]. Electroencephalography (EEG), electrokardiogram (EKG), and eye tracking metrics are often the operator characteristics of interest during various tasks including RAS [22].

EEG spectral band powers, for example, have been used as indicators of cognitive and affective operator states as well as task performance in mentally and physically demanding workspaces [16, 117, 61, 69, 121, 148, 78]. EKG metrics such as interbeat interval (IBI), standard deviation of N-N intervals (SDNN), root-mean-square of successive differences in N-N intervals (RMSSD), low frequency power, and high frequency power have also been

validated as indicators of operator cognitive and affective states as well as correlates of task performance during RAS [63, 15, 72, 39, 164]. Eye tracking metrics such as pupil location, gaze trajectory, and fixation length have been correlated with performance in various surgical tasks [155]. These biometric features are a natural choice for building RAS operator models given their extensive validation as indicators of operator cognitive and affective states and their correlation with task performance during RAS.

More specifically, understanding the relationship between operator physiology and performance during robot-assisted surgery (RAS) has the potential to improve operating outcomes, surgical education, training, and skills assessment [37, 104, 142]. The cognitive and affective states of surgeons in the operating room (OR) are primary determinants of surgical performance [169] [138] [67]. Measuring changes in these states contributes to insightful performance analysis, training development and skills improvement [49] [21]. However, these cognitive and affective state measurements, which can be derived from neurophysiological metrics, must be coupled with measures of surgical task success or failure to enable any conclusions about the correspondences between psychological indicators and surgical performance [103]. Unfortunately, the majority of existing tools to measure surgical task performance are retrospective, reliant on subjective, manual expert annotation and often focused on entire procedures rather than on intraoperative performance [101].

Time and frequency-domain electrokardiogram (EKG) metrics provide high-temporal resolution measures of surgeon physiology that are robust to sensing challenges in dynamic operating room (OR) environments [63]. These metrics have been associated with cognitive and affective features such as mental workload, acute stress, and cognitive fatigue that are known to impact surgical performance [133, 124, 15]. To explore the relationships between EKG measurements and performance, subjective, manual ratings of surgical performance or error are then post-processed and correlated with physiological data and their implied cognitive or affective features [163, 39]. Developing a more robust

relationship between intraoperative EKG metrics and surgical performance, however, requires rater-independent, objective error annotation, high-temporal resolution error identification, and precise alignment of error markers with physiological signals.

Prior works examining EKG statistics in the context of surgical performance typically consist of retrospective analyses that often aggregate physiological metrics over the entire surgical procedure and use manual rating schemes to characterize per-procedure surgical performance [88, 15, 76, 164, 5]. All of these studies use aggregated physiological statistics and subjective performance outcomes. Higher temporal resolution experiments examining intraoperative, rather than per-procedure performance and physiology, have been published but are largely reliant on retrospective, rather than on-line or real-time, analysis due to the need for manual annotation of surgical performance by expert surgeons [8].

EEG indicators of psychological features have also been shown to impact task performance in mentally and physically demanding workspaces. Global θ power [23], and global β power [123] as quantitative EEG indicators of error recognition. "Focused attention" or "vigilance" is the ability of an individual to maintain sustained attention on a task [110]. Decreases in focused attention increase error rates in human-machine teamwork [57]. Fluctuations in levels of attention can be measured through changes in EEG spectral power analyses [16]. Global β: $(\alpha + \theta)$ power ratio [117], occipital β power [61], occipital δ power [69], occipital θ power [121], frontal θ power [148], and global γ power [78] have been used as quantitative indicators of attention.

Cognitive load and fatigue are similarly relevant to human-robot teaming. High levels of cognitive load lead to task errors [149]. Cognitive fatigue, a decrease in cognitive performance after sustained cognitive load [1], is of particular importance in safety-critical aviation, transportation, aerospace, military, medicine, and industrial settings where fatigued individuals routinely operate complex, automated systems [171]. EEG spectral power ratios [134] can also be used as measures of cognitive load or fatigue. Frontal θ power [143] [122] [145], frontal α power [111] and parietal α power [53] have been used as

indicators of cognitive load. Global α power [135], occipital α power [153] [178], frontal θ power [148], and θ:α power ratio at Fz [13] have been used as indicators of cognitive fatigue.

Affective characteristics such as valence, arousal, and dominance (VAD) may also influence human-robot system performance. Valence, the positivity or negativity of emotion, modulates attention and has shown an inverted u-shaped relationship with human performance in various tasks [19]. Arousal, often referred to as "stress", has a similar inverted u-shaped relationship with cognitive load [151] [146]. Dominance, feelings of being "in control" of a situation, has a positive relationship with performance during sustained tasks [32]. EEG correlates of valence include α power asymmetry [62] [109], global β power [96], global γ power [108] and global α power [3]. Indicators of arousal include global θ power, global β power and global γ power [96]. EEG correlates of dominance include global α:β power ratio and parietal β power [158].

Prior work examining EEG changes during RAS has primarily focused on correlating EEG measurements with retrospective, subjective surveys of perceived mental workload [168] [65] [174] [176] [170]. While mental workload is a significant contributor to surgical performance [136] [71] [9], it is not a proxy for actual performance. Understanding the relationship between operator EEG measurements and actual surgical performance requires objective, high temporal-resolution measurements of intraoperative performance that can be synchronized with changes in EEG indicators.

While research that connects human physiology with robotic systems has appeared in assistive or rehabilitation robotics, surgical robotics, drone operation, and rescue robotics. Human-in-the-loop systems for assistive and rehabilitation robotics primarily use raw physiological signals to optimize control parameters [4]. These systems do not consider higher-level cognitive or affective features and their relationships with objective task performance. Similar research in drone operation has examined the relationship between EKG statistics and perceived task difficulty [38]. These analyses are focused on intra-task

physiology and perceived task difficulty but are not related to objective markers of task performance. While work in rescue robotics has measured cognitive workload via EEG during teleoperation, this research is focused on adapting operator control interfaces to reduce cognitive load rather than understanding how intra-task levels of cognitive load affect intra-task errors or success [40].

Prior research on RAS-specific operator models uses post-surgery, manual ratings of task performance or operator workload and retrospective analysis of physiological signals to develop correspondences among task performance, and operator-specific or platform-specific components of performance [107] [60] [41] [167]. Yu et al. [168] have pioneered significant work in RAS human factors research, however, their work does not include task environment perception for intra-task performance analysis. They instead focus on aggregate levels of subjective workload and retrospectively calculate task performance measures over entire procedures. While Zhou et al. [175, 174] have published work using multimodal physiological signals combined with task performance and operator workload annotations, these studies still rely on retrospective task performance metrics, post-procedure subjective workload surveys, and post-processed physiological data to analyze aggregate correspondences over entire procedures. These studies rely on retrospective task performance metrics, post-procedure subjective workload surveys, and post-processed physiological data to analyze aggregate correspondences over entire procedures.

Importantly, this work includes neither task environment perception nor intra-task performance measurement, making it impossible to analyze correspondences among task performance, operator-specific and platform-specific components of performance at intra-task resolution. Without these correspondences, for example, a surgical procedure rated as an overall "success" could have multiple intraoperative errors, all of indeterminate origin (platform or operator). A "failed" procedure would likely have many intraoperative errors, also of indeterminate origin. Inability to track intraoperative error

and analyze operator-specific or platform-dependent determinants of error makes operator and platform performance improvement more difficult. Combining intra-task EKG, EEG, or eye tracking metrics and intra-task performance enables understanding of intra-task responses to objective error during RAS. To construct performance-oriented, accurate models of human operators during RAS that have the capacity to guide real-time, adaptive surgical robotic systems, integrating physiological indicators of operator states with task error detection at intra-task resolution is necessary.

1.3 Robotic System Uncertainty

Robotic-system specific uncertainty, particularly for potentially autonomous platforms, may also contribute to decreased RAS performance. Controlling robotic manipulators via visual feedback requires a known coordinate frame transformation between the robot and the camera. Uncertainties in mechanical systems as well as camera calibration create errors in this coordinate frame transformation. These errors result in poor localization of robotic manipulators and create a significant challenge for applications, such as RAS, that rely on precise interactions between manipulators and the environment.

Prior work in estimating the base-to-camera transform includes solving homogeneous linear systems based on multiple images that capture markers affixed to robotic links [50, 112], Solve-PnP approaches using detected markers on robotic manipulators [93] or marker-less keypoint detection [87, 86, 91, 98]. These methods assume that there is no error or uncertainty in joint angle measurements or the forward kinematic transforms.

To compensate for error and uncertainty in the base-to-camera transform, prior work has considered interactive methods for calibrating remote center of motion (RCM) robots [173] and kinematic remote center coordinate systems (KCS) that accounts for error in the RCM camera frame-to-base transform [172]. Markers, learned features, silhouette

matching, and online template matching from images have been used for KCS tracking [94, 126, 127, 97]. These methods, however, fail to explicitly account for error or uncertainty in joint angle measurements.

To estimate joint angle measurement errors in the context of cable-driven surgical robotics, prior work has applied data-driven approaches, real-time inverse kinematics, and physical modeling of cable transmission friction and hysteresis [113, 160].

Learning-based approaches for estimating cable stretch and Gaussian processes for measurement error compensation have also been used [74, 115, 100]. Image-based methods have used unscented Kalman filters and deep learning tools to accurately estimate cable-driven joint angles [66, 139]. While all of these approaches have been effective in the laboratory setting, they require significant investments in additional sensors and calibration time. Moreover, learned joint calibration parameters become less applicable over time due to mechanical effects such as cable stretch.

Estimating both camera-to-base transform and joint angle offsets simultaneously using joint calibration techniques have been proposed [120, 89]. These methods, however, can not account for dynamic uncertainties such as non-constant joint angle errors. Methods to achieve real-time estimation of dynamic joint angle errors have used iterative closest point matching based on 3D point clouds and Kalman filters [82]. Probabilistic approaches using observation models parameterized by physical parameters have also been explored [31]. While these works prioritize integrating additional sensors such as depth cameras into real-time estimation, current RAS platforms only have access to endoscopic images.

To address these challenges, prior work on lumped error formulations account for uncertainty in the camera-to-base transform, camera calibration, and joint transforms in a forward-kinematic model for partially visible kinematic chains [131]. Estimating this lumped error has enabled more accurate, probabilistic tool tracking for surgical systems [29]. However, this approach remains vulnerable to degradation in image quality, false

positive and negative feature detections, fragile associations due to hand-tuned parameters, and a lack of empirical, data-driven priors for Bayesian filter models.

In addition to end-effector localization uncertainty, visual uncertainty with respect to endoscopic images remains a significant obstacle to high performance RAS, both in human-robot collaborative settings and in autonomous applications. Estimating depth from binocular imagery, particularly in dynamic unstructured environments such as minimally invasive surgical scenes, is a necessity for human operators as well as autonomous agents. Prior work on stereo algorithms for local correspondences often result in inconsistent matching on poorly-textured and ambiguous surfaces [137, 162]. Global correspondence algorithms require large computational efforts and high memory capacity [18, 80, 51, 165, 28, 81]. Bayesian approaches for depth estimation have been demonstrated to be both accurate and fast enough for real-time approaches [54].

While this work enables measurement of visual uncertainty, its applicability to RAS remains unknown. "active vision" techniques that manipulate camera movement to reduce the visual uncertainty of surgical scenes presented to human operators or robotic system perception modules have not been explored. Prior research on active vision in the RAS domain have largely focused on centering objects or features of interest in the endoscopic field-of-view (FOV) [11, 45, 129, 75]. This higher-level approach depends on a priori identification of objects or features of interest, relies on 2D image metrics for optimized camera movement, and neglects the significant impact of visual depth information on operators or autonomous systems performing RAS.

1.4 Our Work

This dissertation seeks to develop intelligent algorithms for augmenting surgical performance during RAS in both human-robot collaborative and autonomous system settings. For adaptive, high-fidelity operator models, we construct on-line, automated

detection, analysis and alignment methods for intraoperative errors and biometric signals including in order to identify relationships between operator-specific data and objective surgical performance. We build and deploy SCALPEL, a system for on-line integration, synchronization and analysis of biometric, robot and task data to characterize human operator and robotic platform-specific measurements during intraoperative error. Changes in EEG and EKG metrics are compared relative to personalized baselines during periods of operator error to those captured during periods of no error. Pupil localization metrics are compared across operators.

For robotic system uncertainty parametrization, we first develop a novel, data-driven method for robust tool tracking that uses deep feature detection and association to exploit surgical tool geometric primitives for pixel-based probabilistic models using empirically derived priors. We use a laparoscopic tool insertion-shaft detection approach based on a Deep Neural Network (DNN) with multiple observation models for Bayesian filtering. We then create a novel formulation to capture image-based visual uncertainty and derive an active vision method to reduce visual uncertainty in the RAS environment.

Our results demonstrate that automated, accurate detection of intraoperative error synchronized with surgical motion primitive analysis and operator biometric patterns can distinguish operator-specific and platform-specific contributions to surgical error. Additionally, our results show that our novel approaches to surgical tool tracking and visual uncertainty reduction enable higher performance RAS.

Our motivation is improving the performance of both human-robot teams and autonomous systems during RAS. Understanding operator neurophysiology during intraoperative error can help guide more accurate human performance assessments and training for surgeons and other teleoperators. Understanding robotic platform uncertainty can assist development of more capable teleoperated and autonomous robotic systems. Both complementary avenues of investigation may ultimately lead to improved patient outcomes following RAS.

In the remainder of this dissertation, we describe our development of novel machine intelligence algorithms to quantitatively characterize performance during robot-assisted surgery for both human operators and autonomous systems.

1.5 Acknowledgements

Chapter 1, in part, is a reprint of material from:

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R., Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C., 2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Richter, F., Aronoff-Spencer, E., Broderick, R.C. and Christensen, H., 2023. SCALPEL Please: Personalized, Performance-Oriented Operator Models Using Biometrics and Intra-task Error Detection During Robot-Assisted Surgery. IEEE Robotics and Automation Letters, (in submission).

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 2

Measuring Surgical Performance

To measure surgical performance, we focus on objective intra-task error detection, operator biometric data capture, analysis and synchronization, robotic tool localization, and robotic perception uncertainty. Capturing these task, operator and robot-specific data streams required development and construction of a novel data capture and analysis platform for use in an RAS-equipped OR (Fig. 2.1) as well as access to a dVRK.

2.1 SCALPEL Platform

SCALPEL (**S**ynchronized **C**apture and **A**nalysis for Te**L**eoperation **P**erformance **E**va**L**uation) is a capture and analysis platform that examines task performance and operator physiology in tandem, provides high temporal resolution for intra-task performance and physiology metrics, and is suitable for on-line and retrospective data analysis. SCALPEL is intended for high-consequence, safety-critical teleoperation tasks such as robot-assisted surgery (RAS), drone operation, and disaster robotics to help identify intra-task moments of difficulty or failure, monitor operator responses to challenging situations, guide precision training and skill acquisition, and inform ratings of operator proficiency. We use task environment images to detect intraoperative errors. We analyze operator brain function, heart rate, and pupil movement to identify operator characteristics during intra-task errors. Unlike prior work, SCALPEL includes a

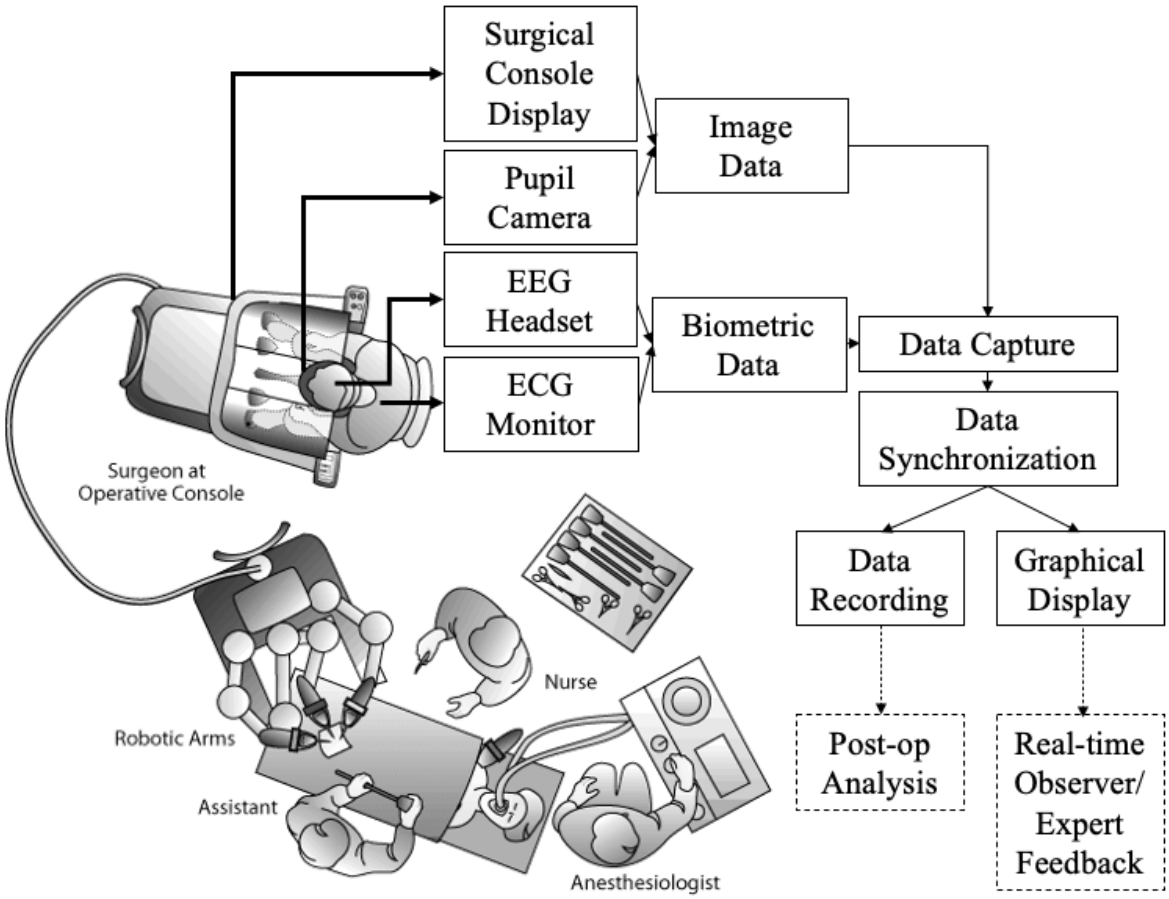


Figure 2.1. SCALPEL is an on-line, intra-task performance and biometric data platform for measuring human operator performance during teleoperation. OR schematic specifies the layout of a typical OR during a robot-assisted surgical procedure with SCALPEL concepts included. Dashed lines indicate potential applications.

perception module for automated intraoperative error detection at high temporal resolution that is synchronized with human operator biometric data on a moment-by-moment basis. (Fig. 2.2)

2.2 System Components

SCALPEL is composed of 5 hardware components and 6 software modules. 4 hardware components acquire separate surgical or physiological data streams, and 1 central compute node runs 6 software modules to collect, synchronize, and integrate

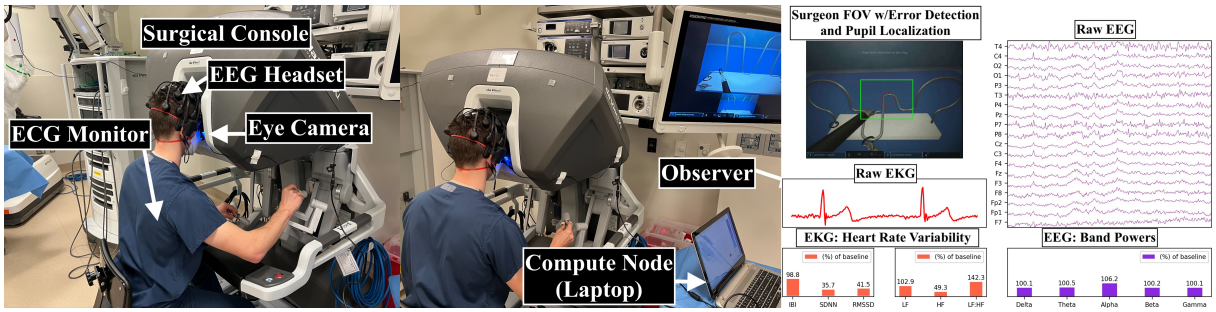


Figure 2.2. SCALPEL deployed in OR during surgical training. "Wearing EEG Headset" shows the positioning of the EEG headset on the surgeon. "At Console" shows the surgeon at the surgical console, the fit of the EEG headset within the confines of the console, and the eye camera mounted within the console. "Expert Observer at Display" shows the compute node (laptop) recording the EEG, ECG, surgical console, and eye camera data streams and displaying those streams for the observer. "Display Streams" shows the real-time display of all data streams as presented to the observer via the laptop screen.

physiological measurements with task-based information. (Fig. 2.3)

2.2.1 Hardware

The hardware components include a surgical console display, an infrared-sensitive, short focal length USB camera, a 20-channel, dry-electrode wireless EEG headset, and a chest-strap ECG monitor. (Table II)

Surgical Console Display

We use a daVinci Xi surgical console (Intuitive Surgical, Inc.) for image-based surgical data. The Xi console has two VGA-video output ports for 640 x 480 pixel (px) RGB images of the surgeon's FOV at 30 fps. Image analysis can provide surgical procedure information such as task and sub-task progress, error, duration, and manipulator or endoscope movement. At a frame rate of 30 fps, frame-by-frame analysis of the surgical procedure can be used to correlate task information with captured physiological signals over 33 ms intervals.

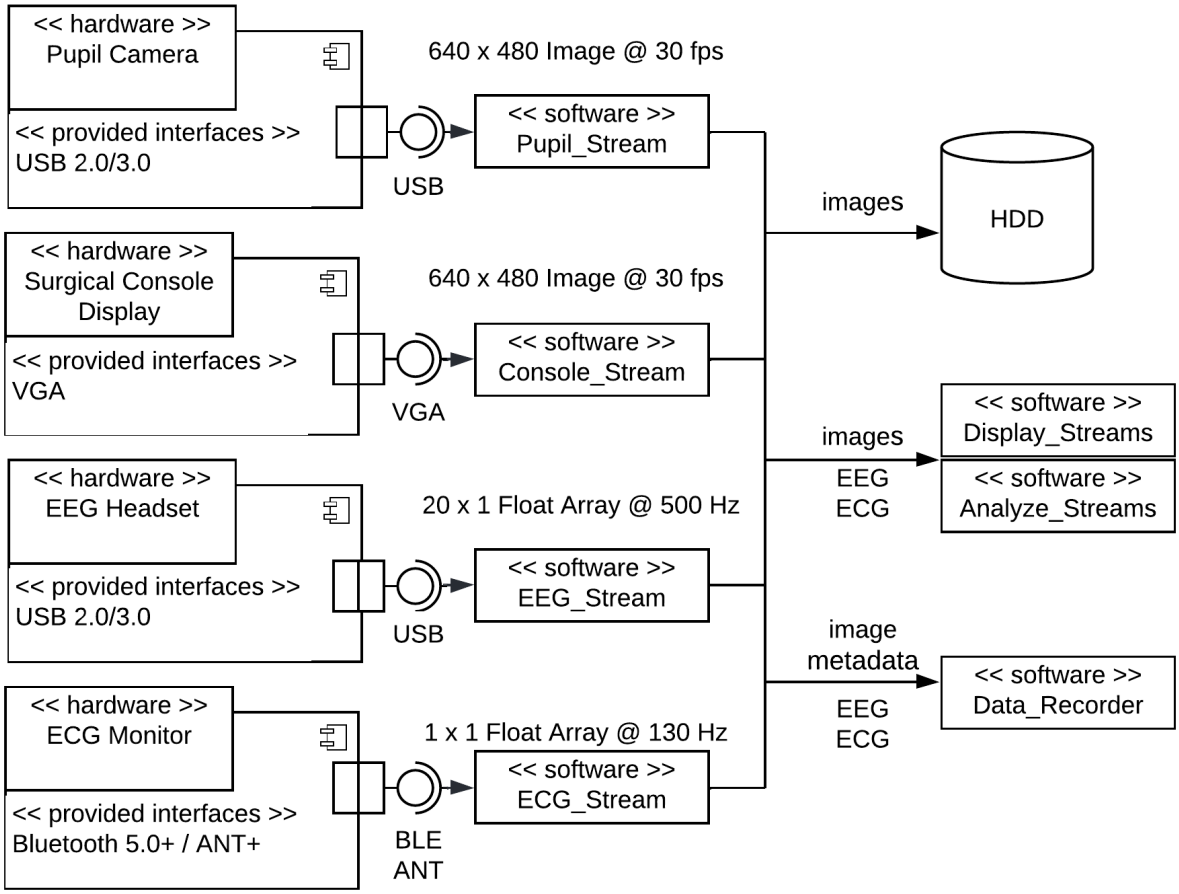


Figure 2.3. Software module input and output streams. Hardware component biometric output streams.

Eye Camera

We use a short focal length USB camera with a frame rate of 30 fps and a resolution of 640 x 480 px. The camera's short focal length facilitates quality eye image capture in the confined space of the surgical console. The surgical console prevents ambient light from reaching the operator's eyes so we remove the infrared (IR)-blocking filter on the camera lens to increase pupil detection accuracy in low-light settings [105]. We also modify the camera case to minimize possible occlusion of the surgeon's FOV when the camera is mounted on the surgical console. Rather than integrating the camera into a wearable frame, we choose to mount the camera on the console to maximize the

probability of accurate gaze estimation. By fixing the position of the camera relative to both the position of the operator's head in the surgical console and the position of the surgical console display, gaze analysis is more robust.

EEG Headset

For neurophysiological data capture, we use a Cognionics Quick-20 EEG headset (Cognionics, Inc.). We choose a dry-electrode rather than a gel-electrode system because dry-electrode systems have less set-up and tear-down time. Instrumenting surgeons rotating through simulated procedures in a fast-paced OR requires minimal set-up time.

ECG Monitor

We use a Polar H10 (Polar, Inc.) chest-strap monitor because the H10 has comparable accuracy to medically-prescribed ECG monitors [56]. The H10's signal accuracy and sampling rate are high enough to detect in-task, real-time fluctuations of time and frequency domain ECG statistics.

Compute Node

We use an Intel Core i7-5500U (Intel, Inc.)-based Windows (Microsoft, Inc.) laptop as the central collection and compute node for the data streams provided by the previous 4 hardware components.

2.2.2 Software

SCALPEL's 6 software modules are run on the central compute node. 4 modules transfer raw data from the hardware components to a recording module (`Data_Recorder`) that synchronizes the raw data streams and stores them on disk. These 4 modules also transfer raw data to a graphical display module (`Display_Streams`) that synchronizes and displays the input data. To transfer data between components and software modules, we use TCP-based interprocess communication coupled with interprocess shared memory

buffers to enable data synchronization and integration. Shared memory buffers minimize the differences in synchronization and integration times across large-bandwidth data streams (images from the surgical console and eye camera) and small-bandwidth data streams (EEG and ECG samples). Synchronization is achieved using a linear fit between the sample timestamp and the effective sampling rate for each separate data stream.

Image Stream Modules

`Eye_Stream` receives images from the eye camera and uses separate threads to write these images to video, send frame-by-frame image metadata to `Data_Recorder`, and send shared memory buffer image pointers to `Display_Streams`. `Console_Stream` receives images of the surgeon's FOV from the surgical console but is otherwise identical to `Eye_Stream`. We send image metadata rather than actual images to `Data_Recorder` to eliminate the need for writing identical videos to disk at both the image stream modules and the data recording module. We send shared memory buffer image pointers rather than actual images to `Display_Streams` to avoid lag between image capture and image display. Sending pointers reduces the time required for image data serialization at the image stream modules and subsequent image data deserialization at the display module.

Biometric Stream Modules

`EEG_Stream` sends EEG data from the Cognionics EEG headset to `Data_Recorder` and `Display_Streams`. `ECG_Stream` receives ECG voltages from the wearable heart rate monitor and sends this data to the recording and display modules. Each sample in these data streams is 20 x 32 bits (`EEG_Stream`) or 1 x 32 bits (`ECG_Stream`).

Display Module

`Display_Streams` uses separate threads to subscribe to all image and physiological data streams. These 4 streams are synchronized and displayed on the compute node's

Table 2.1. Target signals and sensors for an integrated task information and biometric data analysis platform. Required sensor sampling rates for capturing features of interest.

Target Signal	Biometric Sensor: Feature(s)	Minimum Required Sampling Rate (Hz)
Task Information	Console Display: Surgeon POV	20 - 77+
Brain Function	EEG: Spectral Powers	100 - 500+
Autonomic Nervous System Tone	ECG: R-R Intervals, Heart Rate Variability	100 - 500+ [84]
Eye Movement	Eye Camera: Intra-task Gaze Trajectory	$\geq$ Task Information Rate (20 - 77+)

Table 2.2. Hardware components for the SCALPEL system. Interfaces and reported sampling rates.

Component	Interface	Sampling Rate (Hz)
Surgical Console Display	VGA (Out)	30
Eye Camera	USB 2.0 / 3.0 (Out)	30
EEG Headset	USB 2.0 / 3.0 (Out)	500
ECG Monitor	Bluetooth 5.0 / ANT+ (Out)	130
Compute Node (Laptop)	3x USB 2.0 (In) 1x Bluetooth 5.0 (In)	-

monitor. `Display_Streams` receives shared memory buffer image pointers from the image stream modules to minimize the time required for interprocess data transport.

2.3 SCALPEL Validation Experiments

We conducted three separate trials to validate SCALPEL's sample loss, latency, jitter, and synchronization error during actual RAS simulation training sessions for surgical residents. For each trial, we instrumented a different surgical resident and had them complete three simulation lessons on the surgical console.

Table 2.3. Hardware components for the SCALPEL system. Interfaces, channels and resolutions.

Component	Interface	Channels	Resolution
Surgical Console Display	VGA (Out)	3: RGB	640 x 480 pixels
Eye Camera	USB 2.0 / 3.0 (Out)	3: RGB	640 x 480 pixels
EEG Headset	USB 2.0 / 3.0 (Out)	20	+/- 1 microvolt
ECG Monitor	Bluetooth 5.0 / ANT+ (Out)	1	+/- 1 microvolt
Compute Node (Laptop)	3x USB 2.0 (In) 1x Bluetooth 5.0 (In)	-	-

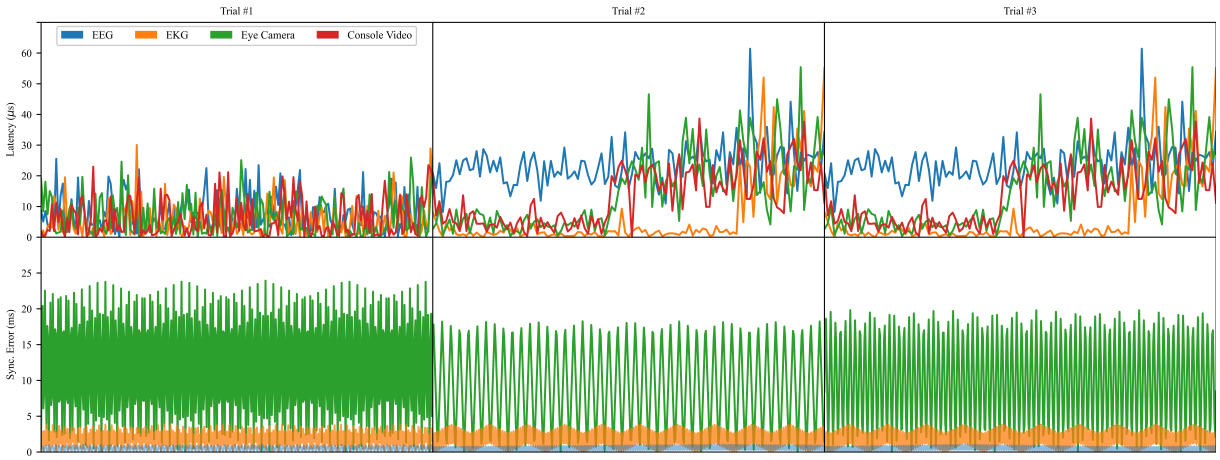


Figure 2.4. Latency and synchronization error for all validation trials.

2.3.1 Sample Loss, Latency, Jitter and Synchronization Error

Given a set of sensing modalities M:{console video (cv), eye camera (ec), eeg, ekg} with sampling rates SM:$\{sm_{cv}, sm_{ec}, sm_{eeg}, sm_{ekg}\}$, an experiment duration D, a sample acquisition process acq, and a sample recording process rec running on a central compute node for each sensing modality $m \in M$, we define the sample loss sl_m for sensing modality m as the minimum of the difference between the expected number of samples to be recorded for the experiment and the actual number of recorded samples and zero.:

$$sl_m = \min(D * sm_m - count(s_{i,m} \forall i \in \{1, ..., N\}), 0)$$

We define the latency $l_{i,m}$ associated with the ith sample from modality $m \in M$ as

the absolute error between the time t at which the sample was acquired $t^{\text{acq}}_{i,m}$ and the time at which the sample was recorded $t^{\text{rec}}_{i,m}$:

$$l_{i,m} = |t^{\text{acq}}_{i,m} - t^{\text{rec}}_{i,m}|$$

We generate these timestamps by having the acquisition and recording processes query the local cpu clock each time a sample is acquired or recorded.

The jitter j_m for each sensing modality $m \in M$ is the standard deviation of the latency for all N samples acquired and recorded by that sensor:

$$j_m = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(l_{i,m} - \frac{1}{N}\sum_{i=1}^{N} l_{i,m})^2}$$

We calculate the synchronization error for each sample from each sensing modality relative to the samples from the console video stream. We choose to synchronize to the console video because the physiological responses of the operator depend on committing and observing errors through the surgical console display.

We define the ith sample synchronization error $\text{syn}_{i,m}$) for $m \in M \setminus \text{cv}$ as the absolute error between the time the ith console video sample is recorded $t^{\text{rec}}_{i,\text{cv}}$ and the time of the nearest sample from modality $m \in M \setminus \text{cv}$:

$$\text{syn}_{i,m} = \min_{j \in 1 \ldots N} |t^{\text{rec}}_{i,\text{cv}} - t^{\text{rec}}_{j,m}|$$

Only the video streams had sample loss (Table I). The maximum sample loss from the console video stream was 0.6%. The eye camera had sample loss ranging from 9-30%. We attribute this sample loss to two factors. First, we use an inexpensive USB camera with a maximum sampling rate of 30 Hz. These cameras are known to throttle image capture and transport at the camera processor level [73]. Second, we use an inexpensive central compute node with USB 2.0 serial buses. This limits the throughput of frames

from both video streams able to be received by the central compute node.

The maximum system latency across all streams was at most 14 μs, and the maximum jitter across all streams was 12 μs. Both of these results were generated by the eye camera; again, an expected result given our pairing of a low-resource USB camera to an low-resource central compute node. Importantly, the latency and jitter of this sensor stream is orders of magnitude smaller than physiological responses to stimulus which take place at the hundreds of milliseconds scale [166].

Synchronization error was also largest for the eye camera. We expected this result given the number of dropped samples as well as the lower sampling rate from this sensor. Fewer eye camera samples recorded implies a larger synchronization error as each console video sample to which an eye camera sample must be synchronized is necessarily further in time from any of the sparsely recorded eye camera samples. Synchronization error for EEG and EKG streams was consistent at 1.9 ms and 0.5 ms. This consistency was expected given the low rates of sample loss from both of those streams. The waveform-patterns of synchronization error in Fig. 2.4 is a result of the consistent offsets between regularly sampled data streams. These synchronization errors were orders of magnitude smaller than physiological response time to stimulus.

2.3.2 Surgical Error Detection

Understanding the relationships between physiological indices and operator performance requires objective task-based data. This data can be acquired from several possible sources including manual annotation of task performance and image-based analysis of task videos. We choose image-based task analysis because we use only three simulation environments for all participants. After annotating a single simulation in each environment for one participant, we use those annotation parameters to analyze the simulations for all participants. These annotations are performed on videos captured at a frame rate of 30 frames per second (fps). Frame-by-frame annotation provides 30 Hertz

Table 2.4. Results from validation trials.

	Validation Trials		
	#1	#2	#3
Trial Duration			
Seconds (s)	781.5	501.3	795.2
Nominal Sampling Rate (Hz)			
Console Video	30	30	30
EEG	500	500	500
EKG	130	130	130
Eye Camera	30	30	30
No. of Samples Expected			
Console Video	23445	15039	23856
EEG	390750	250650	397600
EKG	101595	65169	103376
Eye Camera	23445	15039	23856
No. of Samples Captured			
Console Video	23306	15039	23858
EEG	390755	250672	397632
EKG	101908	65335	103514
Eye Camera	16358	13669	19720
Sample Loss (% of expected)			
Console Video	0.6%	0.0%	0.0%
EEG	0.0%	0.0%	0.0%
EKG	0.0%	0.0%	0.0%
Eye Camera	30.2%	9.1%	17.3%
Effective Sampling Rate (Hz)			
Console Video	29.8	30.0	30.0
EEG	500.0	500.0	500.0
EKG	130.4	130.3	130.2
Eye Camera	20.9	27.3	24.8
Sample Transmission Latency ± Jitter (s)			
Console Video	6.07E-06±5.90E-06	1.32E-05±9.31E-06	1.22E-05±5.93E-06
EEG	8.07E-06±5.83E-06	2.37E-05±6.92E-06	8.72E-06±6.20E-06
EKG	5.64E-06±5.43E-06	6.81E-06±1.14E-05	7.70E-06±6.23E-06
Eye Camera	6.31E-06±6.12E-06	1.43E-05±1.19E-05	9.46E-06±6.64E-06
Synchronization Error$\pm 1\sigma$ (s)			
Console Video	-	-	-
EEG	5.00E-04±2.89E-04	5.00E-04±2.89E-04	5.00E-04±2.89E-04
EKG	1.92E-03±1.11E-03	1.92E-03±1.11E-03	1.92E-03±1.11E-03
Eye Camera	1.20E-02±6.91E-03	9.15E-03±5.28E-03	9.93E-03±5.73E-03

(Hz) resolution for task-based data. This resolution aligns with estimates of human visual processing speed which range from 13 to 50 milliseconds (ms) per image i.e. 20 to 77 fps [119] and implies that we can map post-visual processing operator reactions to in-task stimuli using frame-by-frame analysis of task videos.

We use the daVinci Simulation Software (dvSS) for the surgical simulation task environment. When an operator makes an error during surgical simulation, the operator receives visual feedback from dvSS via the surgical console display (i.e. the surgeon's field-of-view or FOV). We exploit this visual feedback for intraoperative error detection by constructing hue, saturation, and value (hsv) thresholds that are tuned to the visual feedback received by the operator in the surgical console display during error. We use these thresholds to mask and segment each video frame. We sum the pixel intensities in the masked and segmented image and compare them to a tuned error threshold in order to classify that frame as an "error" or "no-error" frame. (Fig. 2.5)

Given a set of intraoperative errors E: {Excessive Force on Manipulated Object, Manipulated Object Collision with Obstacle, Instrument Collision with Obstacle, Instrument Out of View of Endoscope, Instrument-Instrument/Endoscope Collision}, we create an hsv threshold hsv_e for each $e \in E$. From the console video stream cv, we obtain the ith video image frame (i.e. the ith sample) $s_{i,\text{cv}}$ and apply each hsv threshold to that sample:

$$\text{hsv}_e(s_{i,\text{cv}}) \forall e \in E$$

This generates a set of masked and segmented images corresponding to each type of intraoperative error $k_e \forall e \in E$. We then sum the pixel intensities for all P pixels p in each of the masked and segmented images k_e to generate a value $v_{i,e}$ for the ith sample and error type e where $I(k_e)$ indicates the vector of pixel intensities for all pixels in masked and segmented image k_e:

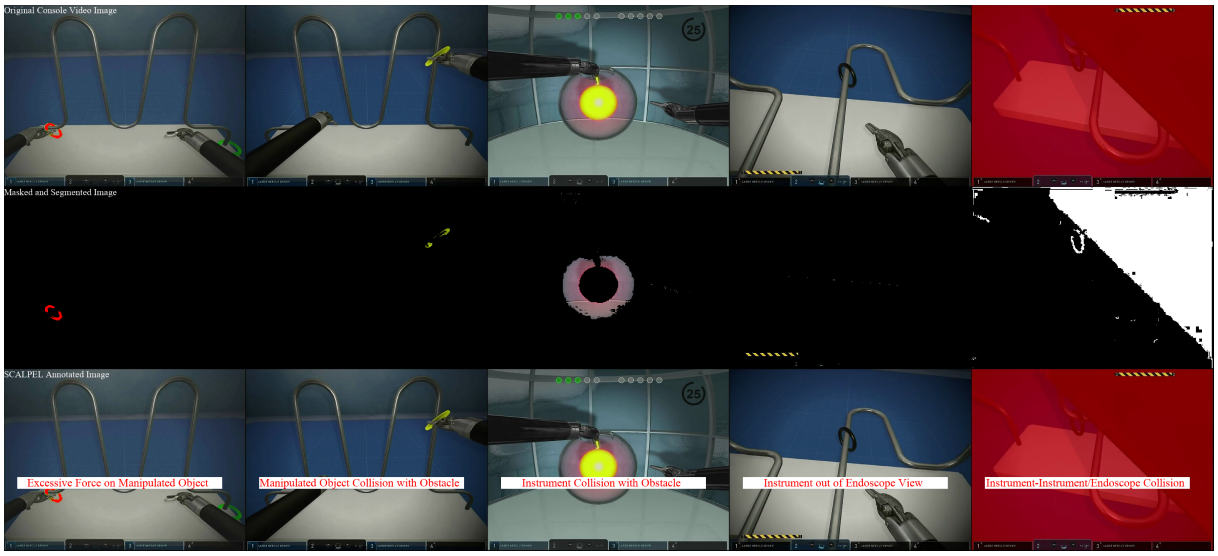


Figure 2.5. SCALPEL error annotation pipeline. The top row of images are video frames captured from the surgical console. The middle row of images are masked and segmented images resulting from applying our hsv thresholds to each original image. The bottom row of images are the annotations produced by SCALPEL.

$$v_{i,e} = \sum_{p=1}^{P} I(k_e)$$

We then compare each $v_{i,e}$ to the corresponding error threshold r_e in order to classify the console video sample $s_{i,\text{cv}}$ as an "error" or "no-error" frame:

$$s_{i,\text{cv}} := \begin{cases} \text{error} & \text{if } v_{i,e} \geq r_e \\ \text{no error} & \text{if } v_{i,e} \leq r_e \end{cases}$$

Comparing a single expert's manual annotation of a random selection of 18,887 frames from three validation trials to SCALPEL annotations demonstrated that our automated error detection algorithms had an overall accuracy of 98.0%, precision of 90.4%, and recall of 93.7%.

Table 2.5. Intraoperative Error Detection Accuracy. Numbers shown indicate the number of operating console video frames from the validation set labeled with "Error" and "No Error" by SCALPEL as compared to manual annotation.

	Manual Annotations		
SCALPEL	Error	No Error	Total
Error	2135	226	2361
No Error	143	16383	16526
Total	2278	16609	18887

2.4 Acknowledgements

Chapter 2, in part, is a reprint of material from:

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R., Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C., 2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Richter, F., Aronoff-Spencer, E., Broderick, R.C. and Christensen, H., 2023. SCALPEL Please: Personalized, Performance-Oriented Operator Models Using Biometrics and Intra-task Error Detection During Robot-Assisted Surgery. IEEE Robotics and Automation Letters, (in submission).

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 3

Modeling Human Operators

To build interpretable operator models that can detect and predict intra-task error based on operator biometric features, we captured, analyzed and synchronized operator biometric data with high temporal resolution intra-task error markers. Due to intersubject variability and our interest in comparing biometric features across participants, we used each participant as their own experimental control. Therefore, all EEG and EKG biometric data are reported as relative to that individual's average baseline measure.

3.1 Methods

3.1.1 Experimental Protocol

Fifty-seven participants signed informed consent and participated in a study that was approved by the local institutional IRB. Of these 57 participants, nine had no medical or surgical training, and the remainder were general surgery residents or attending physicians. All participants were naïve to the hypotheses of the study prior to participation.

During the experimental session, participants were first informed of the procedures, then fitted with a Polar H10 EKG monitor (Polar Inc. USA) and CGX Quick-20 (CGX Inc. San Diego CA USA), twenty-channel mobile EEG headset arranged according to the international 10-20 EEG system. These systems were checked for appropriate positioning, data transfer, and signal quality. If necessary, sensor positions were adjusted to improve

Table 3.1. Study demographics.

Cohort	Final
Number of participants	57
Gender: M/F	36/21
Error % (mean, [range])	17.2 [2.3 - 35.3]
Time to completion (s) (mean, [range])	818.3 [306.6 - 2028.8]
Years of experience (mean, [range])	2.3 [0 - 5]

signal quality. All data, including EKG, EEG, and video data from the operator's point of view, were recorded on the same external computer.

After confirming EKG and EEG signal quality, participants performed four minutes of eyes-open, baseline resting data collection with each participant sitting at the operating console. Due to intersubject variability and the interest in comparing biometric features across participants, each participant's baseline data was used as their own experimental control with subsequent EKG and EEG features reported relative to that individual's baseline measurements.

Following baseline recording, each participant was directed to start the simulation tasks on the Da Vinci surgical robot. Each participant completed the "Ring Rollercoaster 1," "Ring Rollercoaster 3," and "Wrist Articulation 1" simulation tasks without breaks, in this order, as EKG was collected at 130 Hz and EEG was collected at 500 Hz. During the simulations, each participant's head rested against a custom mount affixed to the console to minimize pressure on the EEG headset, and they were asked to be conscious of limiting pressure on the electrodes. After the experimental session, participants completed a brief demographic survey.

3.1.2 EEG Analysis

To pre-process the raw EEG data, we use the PREP bad-channel elimination pipeline [14], highpass filter the resulting signal at 0.5Hz. We then eliminate EEG noise using artifact subspace reconstruction [24]. This pre-processed data is then filtered in parallel into four spectral power bands: theta (θ: 4-8 Hz), alpha (α: 8-12 Hz), beta (β: 12-30 Hz), and gamma (γ: 30-50 Hz).

To accurately synchronize each operating console video frame with a 1-second window of EEG data, we align the ith console video sample with the jth EEG sample that minimizes the synchronization error $\text{syn}_{i,\text{eeg}}$. The timestamp $t^{\text{rec}}_{j,\text{eeg}}$ is the "start EEG window" timestamp that corresponds to the start of the ith console video frame. We then we align the $i+1$th console video sample with the kth EEG sample that minimizes the synchronization error $\text{syn}_{i+1,\text{eeg}}$. The timestamp $t^{\text{rec}}_{k,\text{eeg}}$ is the "start EEG window" timestamp that corresponds to the start of the $i+1$th console video frame.

The EEG window corresponding to the ith console video frame therefore includes all samples received in the interval

$$\left[t^{\text{rec}}_{j,\text{eeg}}, t^{\text{rec}}_{j,\text{eeg}} + 1\right)$$

We then compute the amplitude of each EEG spectral band for every EEG channel. Channel-specific spectral band amplitudes are aggregated into global spectral band amplitudes for all samples in the EEG window. We take the root-mean-square of the amplitudes for each band in the 1-second window to generate the global spectral band power features for the EEG window. These features are compared to the average feature value for the baseline period, and these relative feature values can be associated with the presence or absence of a detected intraoperative error in the synchronized console video frame.

Table 3.2. EEG correlates of selected cognitive and affective features from previously published research. All electrodes: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, P7, P3, Pz, P4, P8, O1, O2.

Cognitive / Affective Feature	EEG Correlate	Spectral Bands: EEG Electrodes
Attention	global β: $(\alpha + \theta)$ power ratio [117]	β, α, θ: All
	occipital β power [61]	β: O1, O2
	occipital δ power [69]	δ: O1, O2
	occipital θ power [121]	θ: O1, O2
	frontal θ power [148]	θ: F4, Fz, F3
	global γ power [78]	γ: all
Cognitive Load	frontal θ power [143] [122] [145]	θ: F4, Fz, F3
	frontal α power [111]	α: F4, Fz, F3
	parietal α power [53]	α: P3, P4, Pz, P7, P8
Cognitive Fatigue	global α power [135]	α: All
	occipital α power [153] [178]	α: O1, O2
	frontal θ power [148]	θ: F4, Fz, F3
	θ:α power ratio [13]	θ, α: Fz
Valence	α power asymmetry [62] [109]	α: (F7, Fp1, F3, C3, P7, T3, P3, O1) / (Fp2, F8, F4, P8, P4, O2, C4, T4)
	global β power [96]	β: all
	global γ power [108]	γ: all
	global α power [3]	α: all
Arousal	global θ power	θ: all
	global β power	β: all
	global γ power [96]	γ: all
Dominance	global α:β power ratio	α, β: all
	parietal β power [158]	β: P3, P4, Pz, P7, P8
Error Recognition	global θ power [23]	θ: all
	global β power [123]	β: all

3.1.3 EKG Analysis

Interbeat interval (IBI), standard deviation of N-N interval (SDNN), root-mean-square successive differences in N-N interval (RMSSD), low frequency (LF) signal (0.04 – 0.15 Hz), and high frequency (HF) signal (0.15 – 0.40 Hz) features are calculated from raw EKG data [140]. SDNN and RMSSD are calculated using a 30 second (s) sliding window [141] with a single sample stride resulting in 30 s SDNN and RMSSD measures at 130 Hz resolution. LF and HF signal components were isolated from raw data using parallel passband filters of (0.04 – 0.15 Hz) and (0.15 – 0.40 Hz), respectively. LF and HF signal powers were calculated using root-mean-square (RMS) amplitude over a 25-second window.

The IBI value at $t^{\text{rec}}_{j,\text{ekg}}$ is the most recently captured heartbeat interval. The SDNN and RMSSD values at $t^{\text{rec}}_{j,\text{ekg}}$ are computed over all EKG samples in the interval:

$$\left[t^{\text{rec}}_{j,\text{ekg}}, t^{\text{rec}}_{j,\text{ekg}} + 30\right].$$

As an additional example, the SDNN and RMSSD values at $t^{\text{rec}}_{j+1,\text{ekg}}$ are computed over all samples in the interval:

$$\left[t^{\text{rec}}_{j+1,\text{ekg}}, t^{\text{rec}}_{j+1,\text{ekg}} + 30\right]$$

The LF and HF signal powers at $t^{\text{rec}}_{j,\text{ekg}}$ are the RMS amplitude values of LF and HF signal components over the window:

$$\left[t^{\text{rec}}_{j,\text{ekg}}, t^{\text{rec}}_{j,\text{ekg}} + 25\right]$$

The value of each EKG feature (relative IBI, SDNN, RMSSD, LF power, HF power) is thus associated with a timestamp $t^{\text{rec}}_{j,\text{ekg}}$ corresponding to the jth sample $s_{j,\text{ekg}}$.

To accurately synchronize each operating console video frame with these heart rate variability (HRV) features, we align the ith console video sample with the jth EKG sample that minimizes the synchronization error $\text{syn}_{i,\text{ekg}}$. The timestamp $t^{\text{rec}}_{j,\text{ekg}}$ is the "start EKG window" timestamp that corresponds to the start of the ith console video frame. We then align the $i+1$th console video sample with the kth EKG sample that minimizes the synchronization error $\text{syn}_{i+1,\text{ekg}}$. The timestamp $t^{\text{rec}}_{k,\text{ekg}}$ is the "start EKG window" timestamp that corresponds to the start of the $i+1$th console video frame.

We compare the HRV feature values associated with each sample in this window with the average baseline feature value. We average these ratios over all samples in the window to generate relative IBI, SDNN, RMSSD, LF power, and HF power values for the ith console video frame that we can then associate with the presence or absence of an intraoperative error.

Time-domain EKG statistics: IBI reflects the amount of time between successive

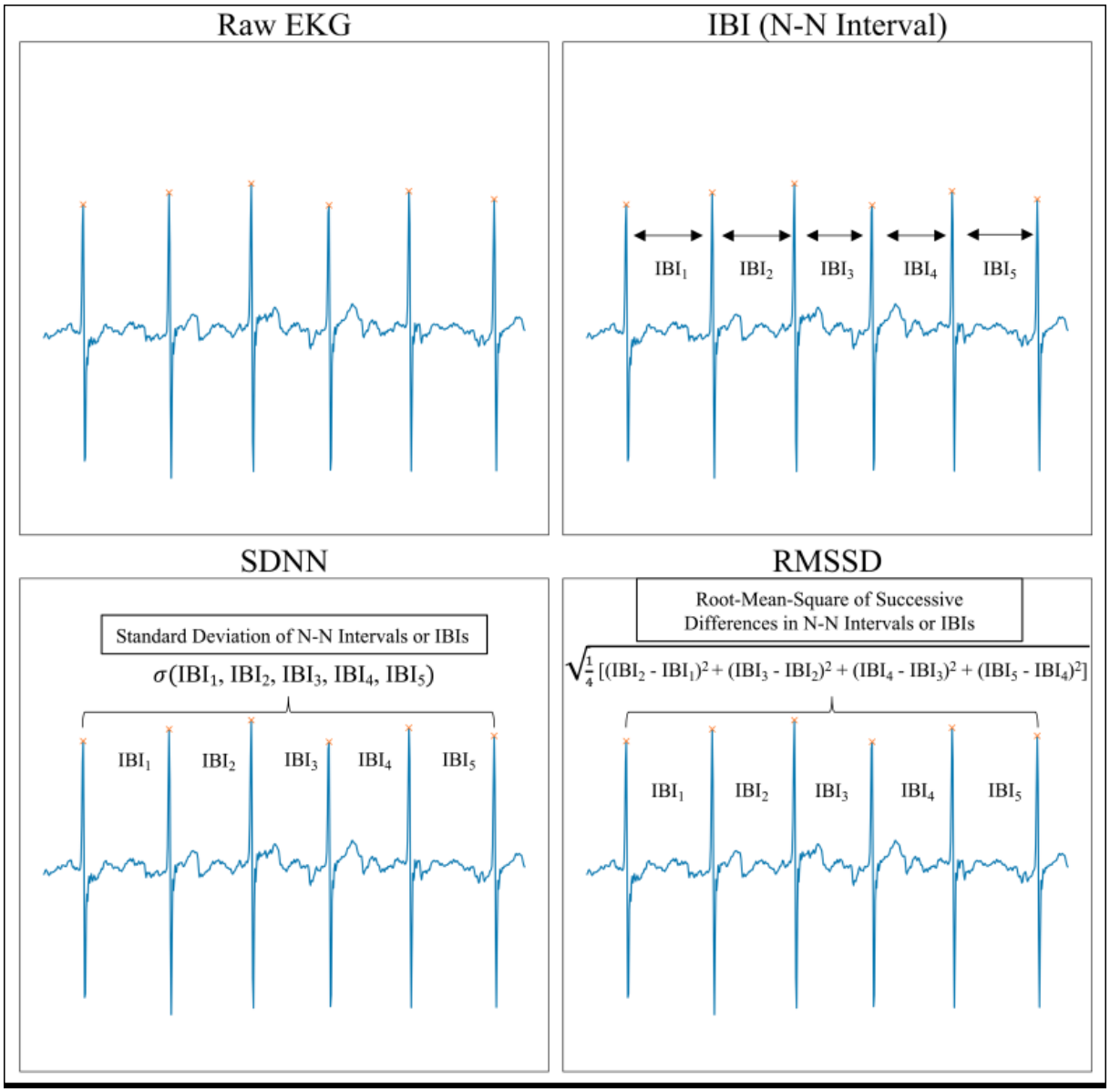


Figure 3.1. Time-Domain EKG Statistics. Interbeat interval (IBI) measures the time (ms) between two heart beats as indicated by the N–N interval. Standard deviation of N–N intervals (SDNN) measures the standard deviation of N–N intervals as measured by IBIs over a specified window of time. RMSSD measures the root mean square of successive differences in N–N intervals as measured by successive differences in IBIs over a window of time.

heart beats (Fig. 3.1). A decrease in IBI is equivalent to an increase in heart rate, which is a marker of increased sympathetic autonomic nervous system (SANS) activity or decreased parasympathetic autonomic nervous system (PANS) activity [118]. Increased

SANS activity has been associated with increased stress, mental and physical workload, and suppression of negative emotions [140].

Increases in SDNN are associated with increased vagal tone and therefore increased PANS activity [133]. Vagal tone and increased PANS activity have been correlated with resilience to mental, emotional, and physical stress as well as lower cognitive workload [116, 42, 118]. Increases in RMSSD are also associated with higher vagal tone and PANS activity [25].

Frequency-domain EKG statistics: LF power is an indicator of mixed SANS and PANS activity as well as baroreflex activity, whereas HF power reflects vagal tone [141]. Increases in LF are associated with higher mental workload [90], and decreases in HF are associated with stress, panic, anxiety, or worry [177].

3.1.4 Eye Tracking

To generate the pixel coordinates of the operator's pupil location x_i, y_i in $s_{i,\text{ec}}$, we first fine-tune a pre-trained eye tracking CNN [44] on manually annotated eye camera video frames. Using this fine-tuned network we generate the x_i, y_i operator pupil pixel coordinates for the ith image frame obtained from the eye camera ($s_{i,\text{ec}}$). We measure the frequency of pupil localizations at each pixel over all eye camera samples to generate a frequency-based heatmap of pupil localizations for each operator over all simulations.

3.1.5 Operator Models

For operator dexeme identification, we create a taxonomy of dexemes for the surgical simulations. (Fig. 3.2) We manually annotate operating console video for dexeme identification on a frame-by-frame basis for individual operators. (Fig. 3.3)

Surgeon performance was calculated based on the combination of simulation error percentage and time to completion. Error percentage was calculated as the percentage of video frames in which an error was detected relative to the total number of video frames

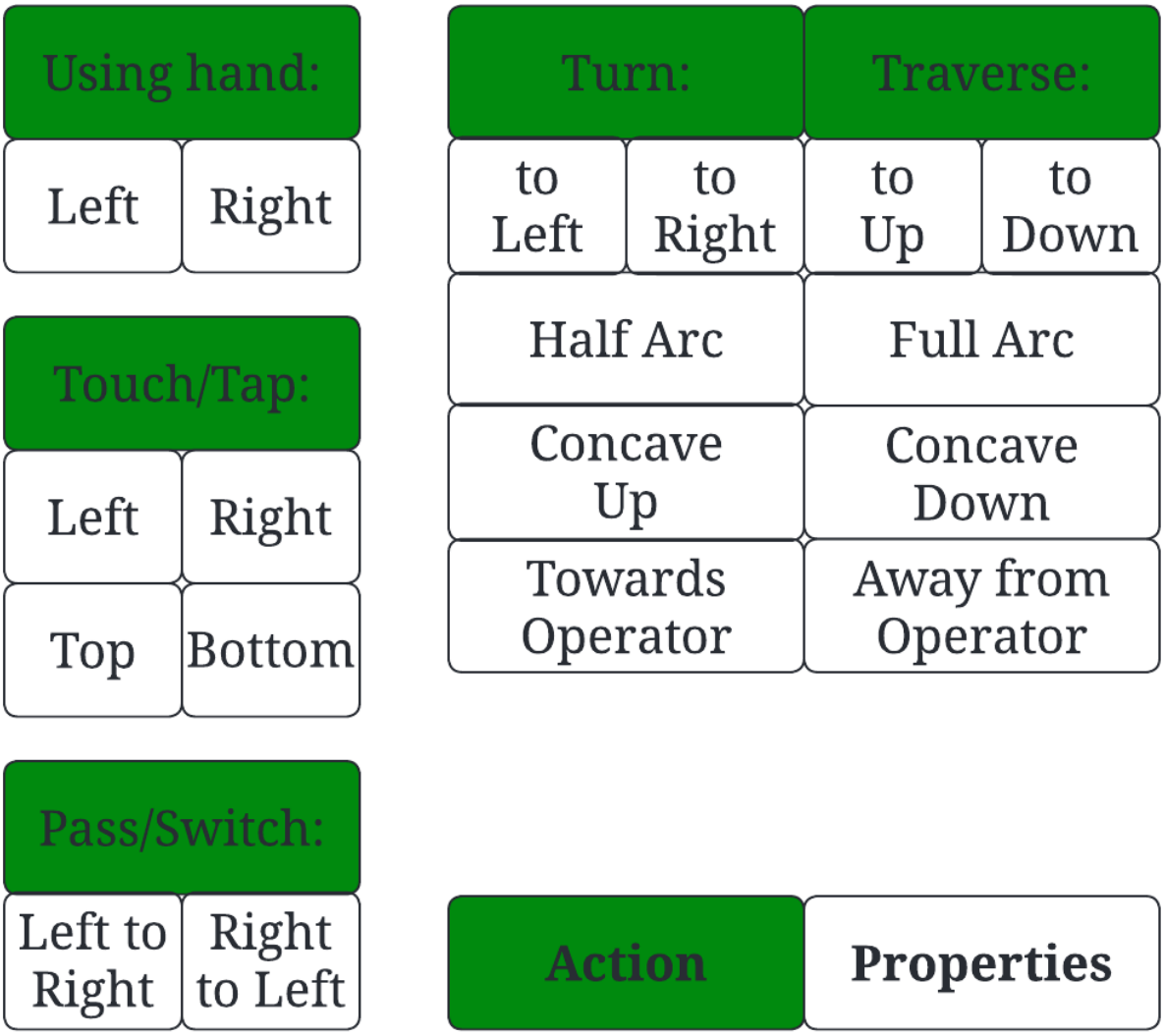


Figure 3.2. Dexeme Taxonomy. "Action" indicates the possible actions taken by the operator. "Properties" indicates the possible properties of the action taken by the operator.

captured during all three simulation tasks. Time to completion was calculated as the total time taken to complete all three tasks. To stratify the participants into performance groups, high performers were defined as the top 33% of participants in rank order by the lowest combined error percentage rank and time to completion rank. Low-performing participants were defined as the bottom 33% of participants in rank order by the highest combined error percentage rank and time to completion rank.

Separate linear mixed-effects models were used for the analysis of EEG and EKG

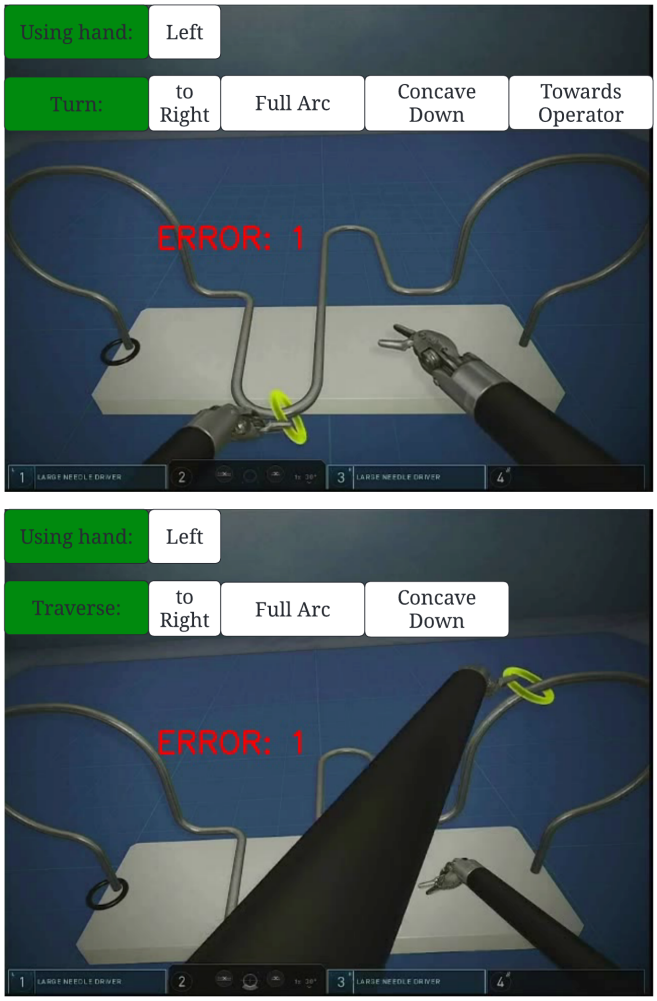


Figure 3.3. Frame-by-Frame Intraoperative Error and Dexeme Annotation. Illustrative example of an operating console video frame that has been labeled as "Error" and has been annotated with the relevant dexemes.

features. Based on previous literature (Table:3.2), we studied a total of 19 EEG features: global delta, theta, alpha, beta, and gamma powers, occipital delta, theta, alpha, beta, and gamma powers, frontal theta and alpha powers, parietal alpha and beta powers, the ratio of theta power to alpha power at the Fz electrode, alpha power left-to-right asymmetry, the ratio of global beta to the sum of global alpha and theta power, and the ratio of global alpha power to beta power. We also examinex 5 EKG features: IBI, SDNN, RMSSD, LF power, and HF power.

Each feature was evaluated in a separate linear mixed-effects model. In each model,

the dependent variable was a log-scaled feature, while the fixed effect was the presence or absence of error. Individual participant variance was modeled as a random effect in all models. Four study groups were considered: all participants, high-performing operators, low-performing operators, and operators that were neither high nor low-performing (labeled "None"). Therefore, a total of 96 models were evaluated, one for each standardized EEG or EKG feature as a function of error status for each of the three subgroups. Given the independent testing of 96 unique models, the baseline significance level of $P = 0.01$ is corrected by a factor of $\frac{1}{100}$ leading to a multiple comparisons corrected significance level of $P = 0.0001$. Effects that met this significance threshold were then further evaluated using visual inspection of residual plots such as the Q-Q plots for evaluation of normality. Only effects that were both significant and passed visual inspection of residual plots are reported as significant. An example of an effect (LF Power for all participants) that met our significance threshold but did not pass our residual plot inspection demonstrated a Q-Q plot as seen in Fig. 3.4 whereas the LF:HF Ratio effect for all participants that met our significance threshold and passed our residual plot inspection had a Q-Q plot as seen in Fig. 3.5.

3.2 Results

3.2.1 Performance Groups

Among the 57 total participants, there was a wide range of completion times (from 300 to 2000 seconds) and error percentages (from 2 to 35%) (Figure 1). The average error percentage and time to completion for high-performing surgeons was 8.7% (sd=2.8) and 780.9s (sd=284.7), respectively, as compared to 26.4% (sd=5.4) and 763.0s (sd=396.6) for low-performing surgeons. The median number of years of training was two years for both the high and low-performing groups.

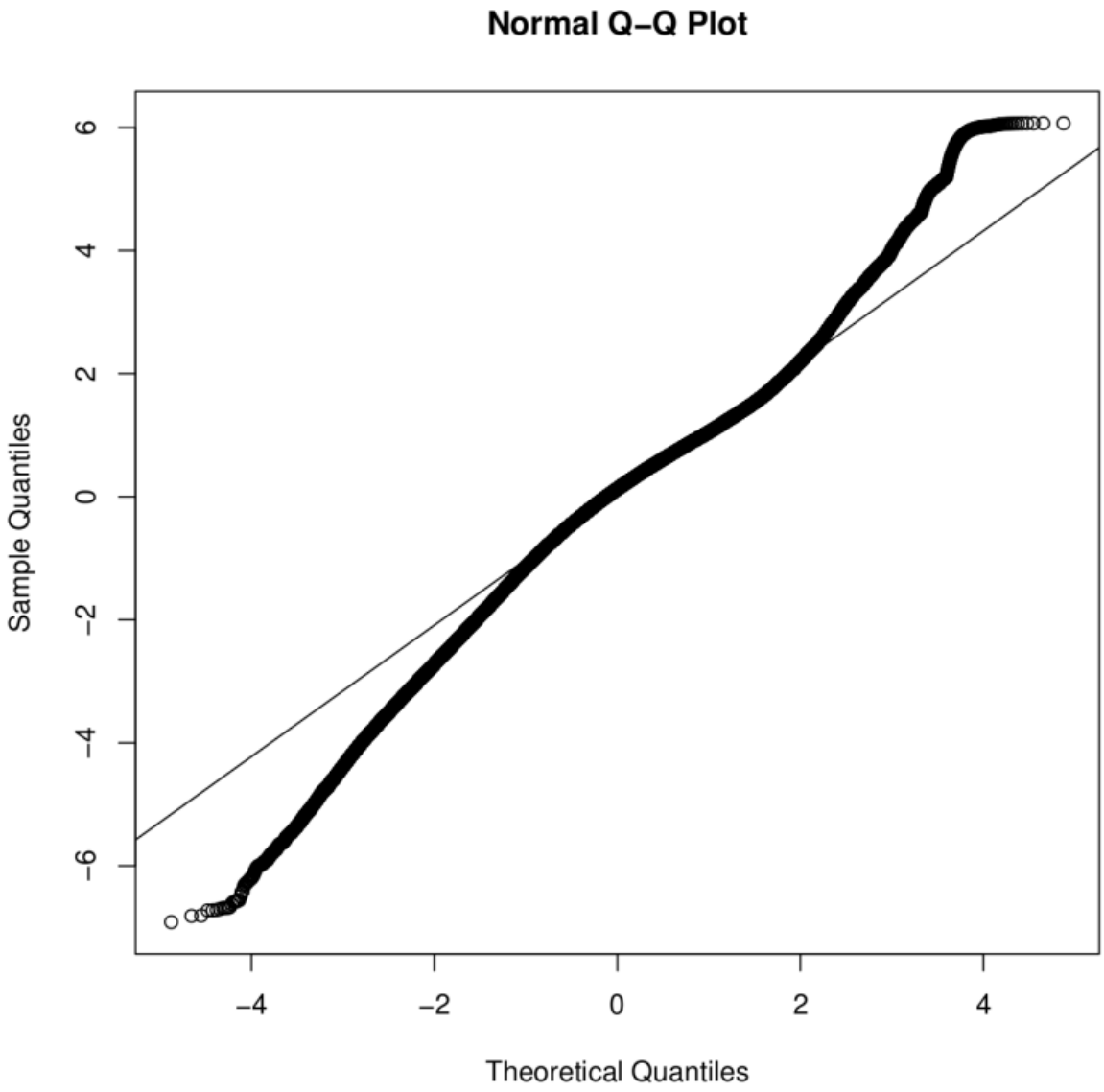


Figure 3.4. Q-Q plot to evaluate normality of residuals for LF Power model.

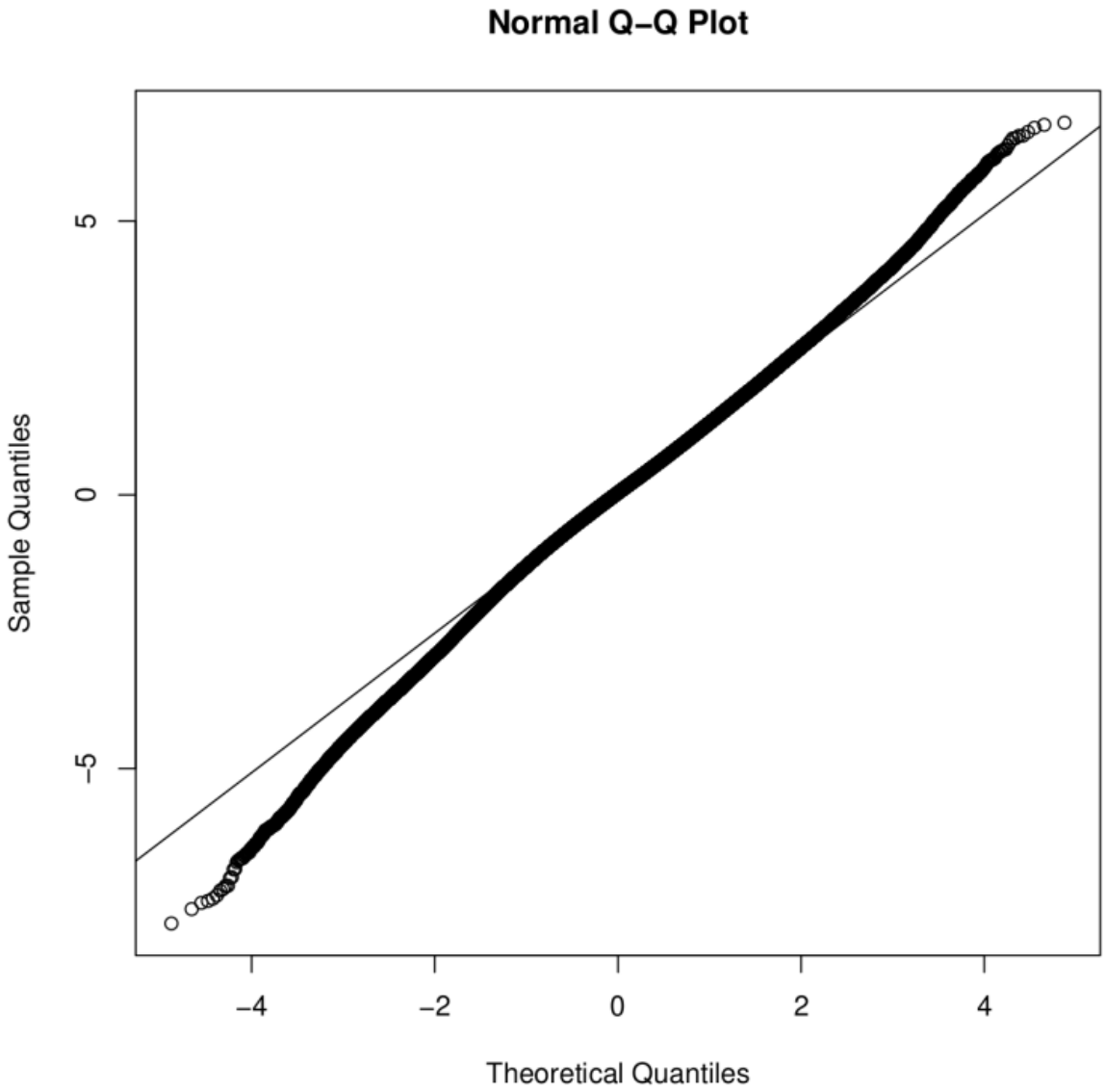


Figure 3.5. Q-Q plot to evaluate normality of residuals for LF:HF Power Ratio model.

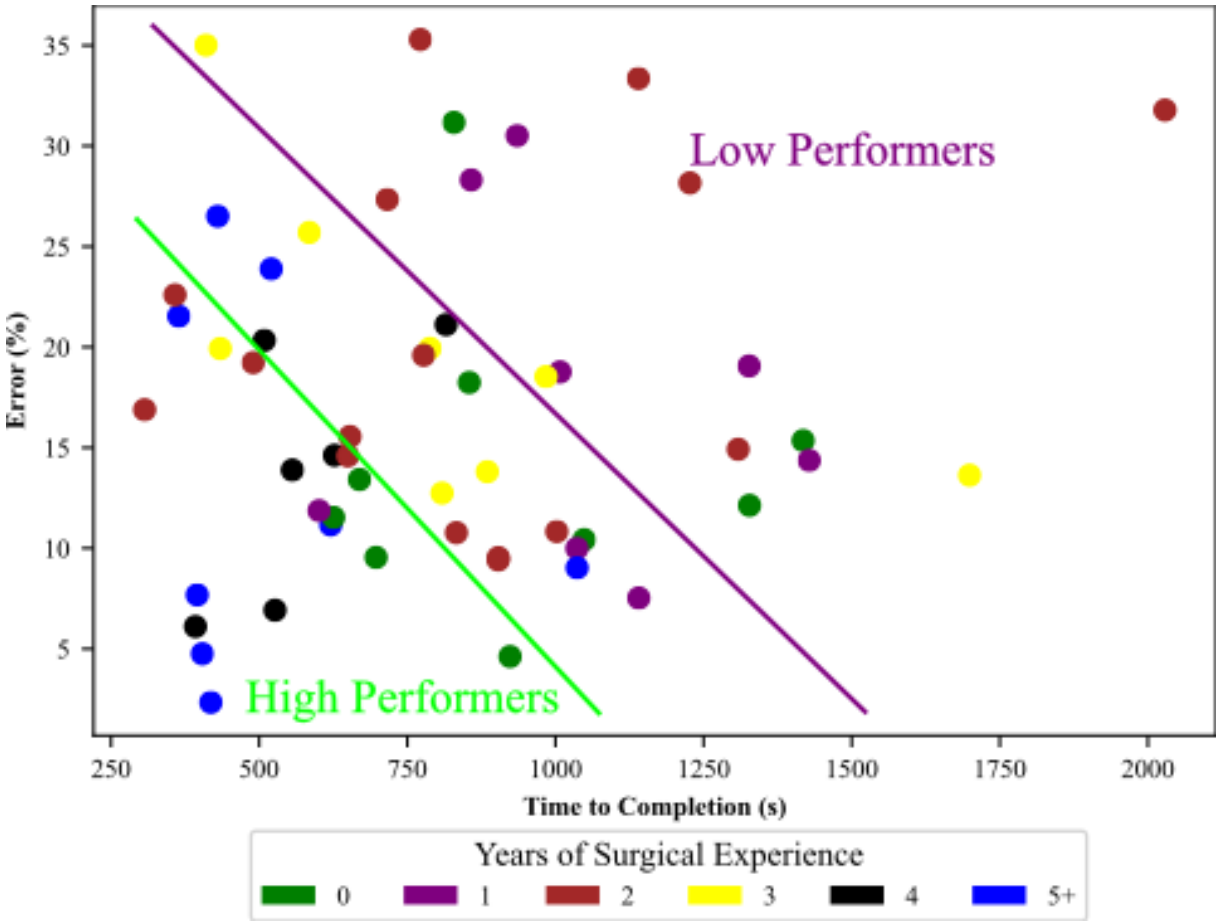


Figure 3.6. High and low-performance groups based on error percentage and time to completion. High-performing surgeons have a relatively lower combined error percentage and time to completion than low-performing surgeons. The color scale indicates the number of years of surgical training experience.

3.2.2 EEG and EKG Features During Error

EEG analyses tested over all participants revealed significant differences between error and non-error intervals for occipital delta power (-0.007318804; se=0.0009969239; p=2.115369e-13) and the ratio of theta power to alpha power at the Fz electrode (-0.01120121; se=0.0009093942; p=7.356278e-35). Among the five tested EKG features, differences in HF power (0.03082602; se=0.002861553; p=4.660412e-27), the ratio of LF to HF power (-0.02625609; se=0.003857423; p=9.996478e-12), and RMSSD (0.00811906; se=0.00145106; p=2.203534e-0) were significant for error, as compared to non-error

intervals, for all participants.

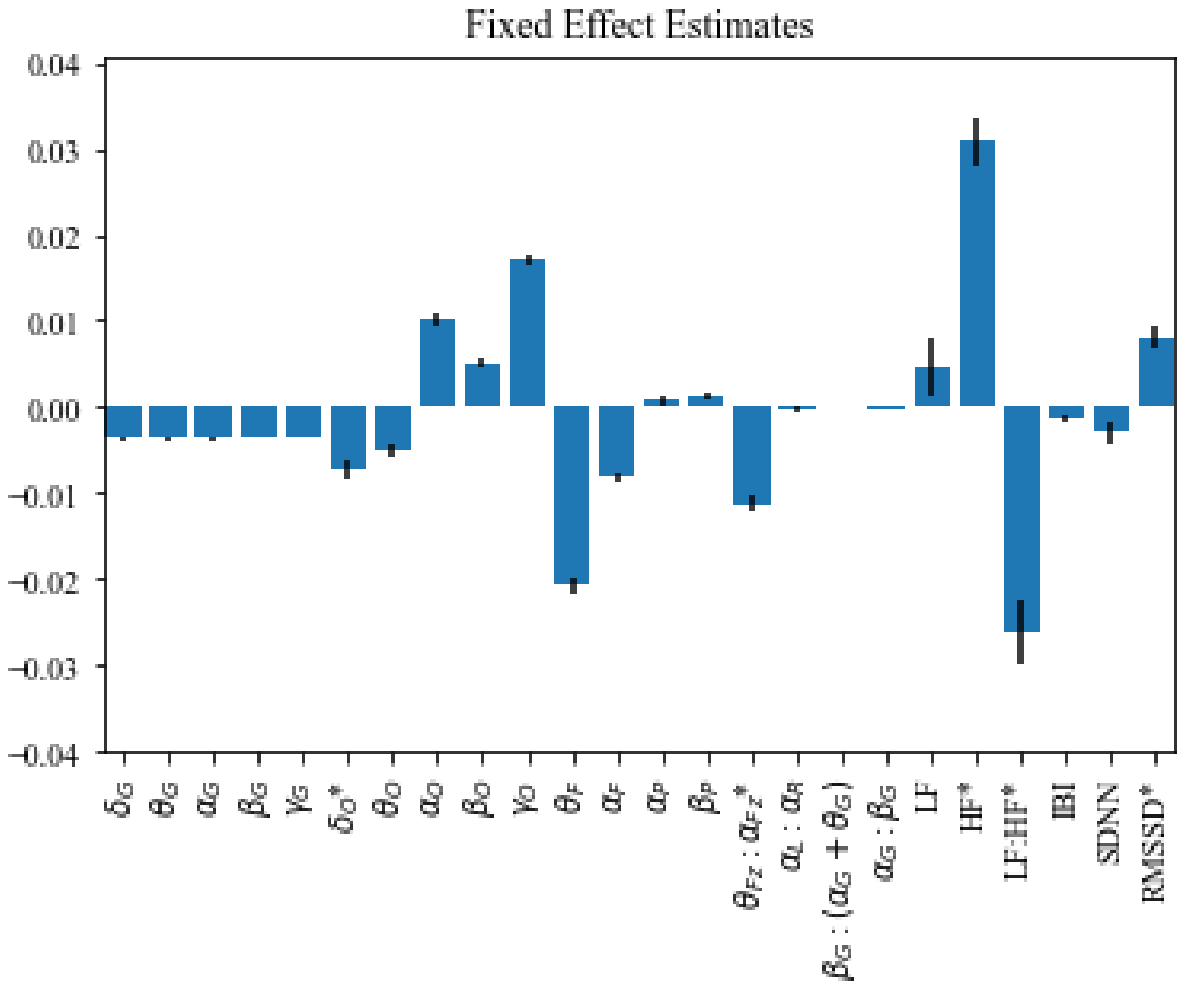


Figure 3.7. Fixed Effect Estimates for Changes in EKG and EEG features during error for All Participants. Corrected significance level for all participants is $P = 0.0001$. *: $P < 0.0001$.

3.2.3 EKG and EEG Features During Error for High and Low Performing Surgeons

In order to test the neurophysiological responses associated with the commission of errors in high and low performing operators, EEG and EKG features were compared for error versus non-error intervals in random, mixed-effects models for each performance level.

For high performers, differences in occipital delta power (0.009580559; se=0.002340957; p=4.266862e-05), occipital alpha power (0.009870165, se=0.001625928, p=1.276493e-09), and the ratio of LF to HF power (0.05823631; se=0.00817304; p=1.039713e-12) were significant during error versus non-error intervals.

For low performers, differences in occipital alpha power (0.01289403; se=0.0009594165; p=3.638445e-41), frontal theta power (-0.01468344; se=0.001018718; p=4.385957e-47), the ratio of theta power to alpha power at the Fz electrode (-0.01198618; se=0.001376407; p=3.097818e-18), alpha power left-to-right asymmetry (0.002659913; se=0.0006095249; p=1.277945e-05), and the ratio of LF to HF power (-0.07776039; se=0.005591473; p=5.930055e-44) were significant during errors versus non-error intervals.

For operators who were neither high nor low performers, differences in the ratio of theta power to alpha power at the Fz electrode (-0.01212955; se=0.001490252; p=3.988714e-16) were significant during errors versus non-error intervals.

3.2.4 Intraoperative Error Density by Dexeme

We choose two participants with different error timelines to illustrate the necessity of high temporal-resolution error detection for insights into platform and operator-specific contributors to intraoperative error. (Fig. 3.9) For Participant #1, the dexemes in which the operator made the most number of errors are "L T:RFU" (gripping the object with the left hand end effector, rotating the left wrist through a full concave up arc while traveling from to the right), "Pass/Switch: L to R" (gripping the object with the left hand end effector and passing it to the right hand end effector), and "R Touch/Tap: LB" (using the right hand end effector to touch the left and bottom hemispheres of the target).

For Participant #2, however, the dexemes with the highest error densities are "L T:RFD" (gripping the object with the left hand end effector, rotating the left wrist through a full concave up arc while traveling to the right), "L T:RHDD" (gripping the object with the left hand end effector, rotating the left wrist through a half concave down

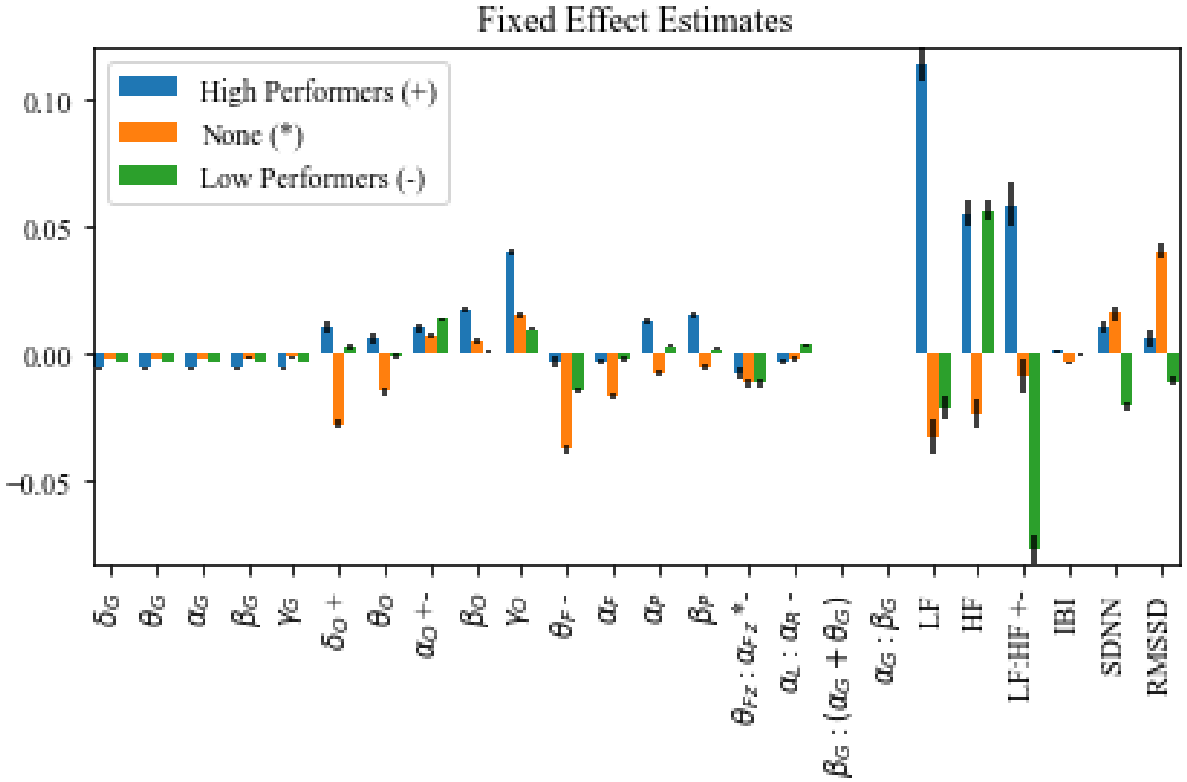


Figure 3.8. Fixed Effect Estimates for Changes in EKG and EEG features during error for High, Low, and neither High nor Low performing Participants. Corrected significance level is $P = 0.0001$. +: significant effect for High performers. *: significant effect for 'None' group i.e. neither high nor low performers. -: significant effect for Low performers.

arc while traveling down and to the right), and "L Touch/Tap: RT" (using the left hand end effector to touch the right and top hemispheres of the target).

We can use synchronized intraoperative error detection and dexeme identification for operator-specific performance improvement. Participant #1 could benefit from training in left-handed surgical instrument manipulation with an emphasis on left wrist, counter-clockwise rotational control. Additionally, practice in targeting objects on the opposite side of the surgical field using right-handed instruments may be beneficial. Participant #2, by contrast, may benefit from training left wrist, clockwise rotational control and targeting objects on the opposite side of the surgical field using left-handed instruments.

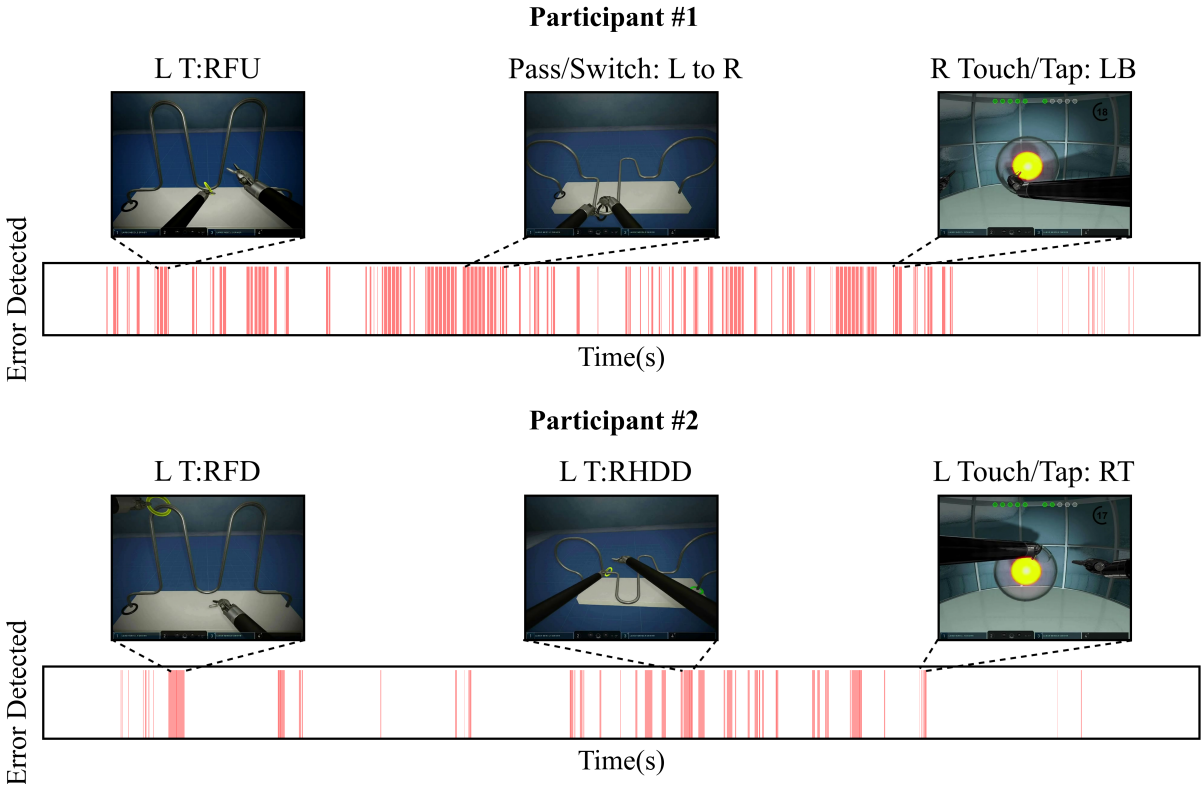


Figure 3.9. High temporal resolution error detection for comparing intraoperative error density between participants. Each vertical red line in the error timeline represents an operating console video frame in which an intraoperative error was detected. Clusters of vertical red lines indicate simulation segments with a high density of errors. The images above each error timeline represent operator actions associated with high error densities for each participant.

We can also use synchronized intraoperative error detection and dexeme identification to guide platform-specific development. For both participants, the common dexemes with the highest intraoperative error density were (a) "Using hand: Left; Traverse: Away from Operator" (using left hand, traverse linearly away from operator into the field), (b) "Using hand: Left; Turn: to Right, Full Arc, Concave Down" (using left hand, rotate left wrist in full concave down arc towards the right) and (c) "Using hand: Left; Turn: to Right, Full Arc, Concave Down, Away from Operator" (using left hand, rotate left wrist in a full concave down arc away from the operator and to the right). (Fig. 3.10)

The operator-specific interpretation is teleoperation performance improvement

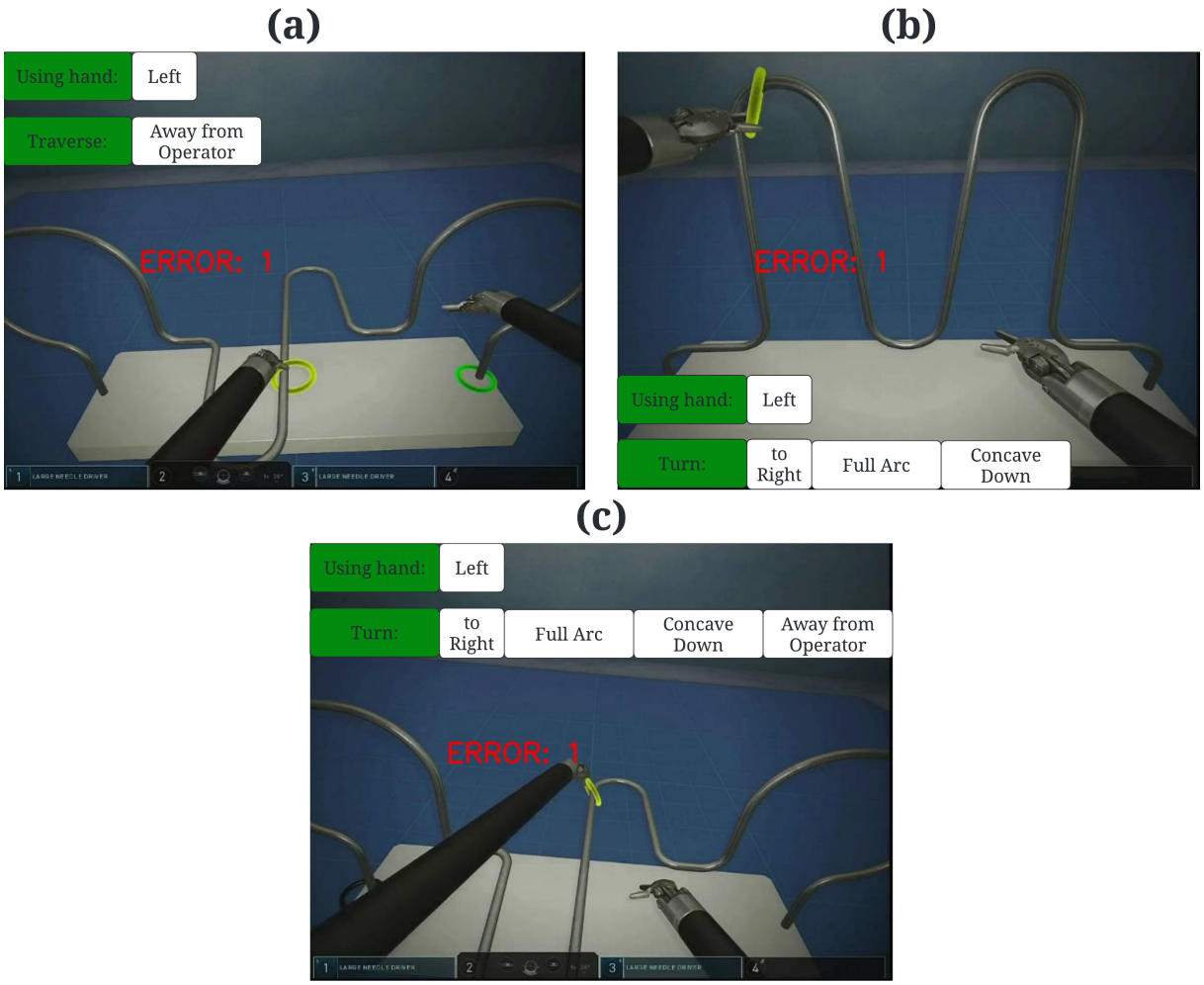


Figure 3.10. Common dexemes with highest intraoperative error density.

depends on operators improving their execution of these common dexemes. An alternative, platform-specific interpretation is a need for improvement in platform controls. Performance during the traverse dexeme, for example, could benefit from more robust, high-pass controller-level filtering of operator inputs during end-effector extension to the deepest region of the workspace in order to mitigate the difficulty of executing smooth, linear motion along the depth axis. The rotational dexemes might benefit from adaptive motion scaling to minimize the anatomical constraints of wrist rotation in both the vertical and horizontal axes [132].

3.2.5 Personalized Intraoperative Performance Analysis

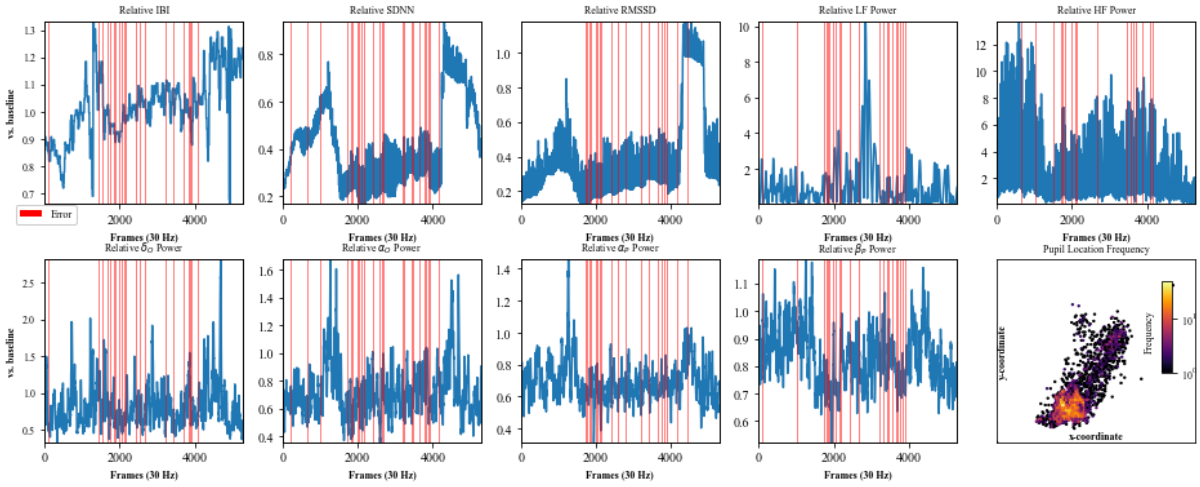


Figure 3.11. High-performing operator-specific error and physiology analysis. Vertical red lines represent detected intraoperative errors during the simulation tasks. The line graphs represent physiological metrics captured during the simulation tasks. Pupil dispersion shows the frequency of pupil localization coordinates in the pupil camera image frame.

While Fig. 3.2, 3.3, 3.10 demonstrate the benefits of high temporal resolution error detection, without synchronized biometric data, the associations between intraoperative error and operator physiology can not be examined. Fig. 3.11 demonstrates SCALPEL's capacity to integrate error detection with physiological signals and identify potential operator-specific contributions to intraoperative error.

For example, SDNN represents the parasympathetic nervous system contribution to heart rate variability. Decreases in SDNN have been associated with periods of stress and arousal [83], possible contributors to and results of intraoperative error. For this high-performing participant, SDNN ranges remains less than 90% of baseline, potentially indicating that they experienced more stress throughout the simulations than during the baseline period. Targeted training interventions, such as SDNN-based biofeedback to reduce stress during task execution, may improve OR performance [92].

The pupil dispersion heat map shows areas of concentration of pupil localization during the simulation tasks. For this participant, pupil localization across all three

simulations is concentrated in the lower-left quadrant of the pupil camera image frame. We do not distinguish between pupil localization during error as compared to no error.

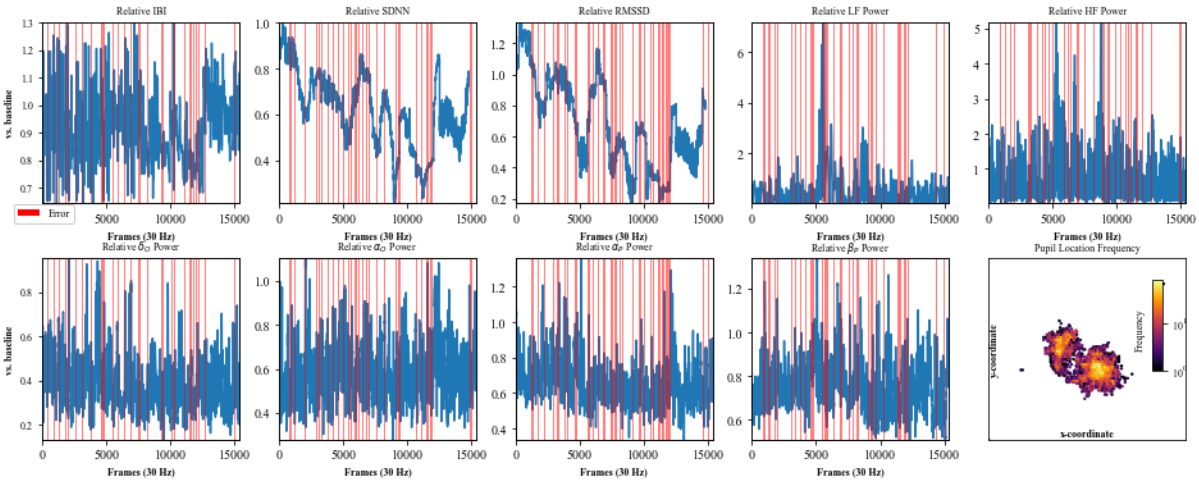


Figure 3.12. Low-performing operator-specific error and physiology analysis. Vertical red lines represent detected intraoperative errors during the simulation tasks. The line graphs represent physiological metrics captured during the simulation tasks. Pupil dispersion shows the frequency of pupil localization coordinates in the pupil camera image frame.

The low-performing operator time series are displayed in Fig. 3.12. For this operator, for example, SDNN trends also indicate that may have experienced more stress throughout the simulations than during the baseline period. Targeted training interventions, such as SDNN-based biofeedback to reduce stress during task execution, may also improve their OR performance [92].

The pupil dispersion heat map shows areas of concentration of pupil localization during the simulation tasks. For this participant, pupil localization across all three simulations has a bimodal concentration in both the center and center-left quadrants of the pupil camera image frame. We do not distinguish between pupil localization during error as compared to no error.

Differentiating individual surgeons using unidimensional metrics such as time to completion, instrument distance, or instrument velocity is ineffective [77, 159, 48, 47]. Fig. 3.13 demonstrates the utility of combining high temporal resolution error detection with

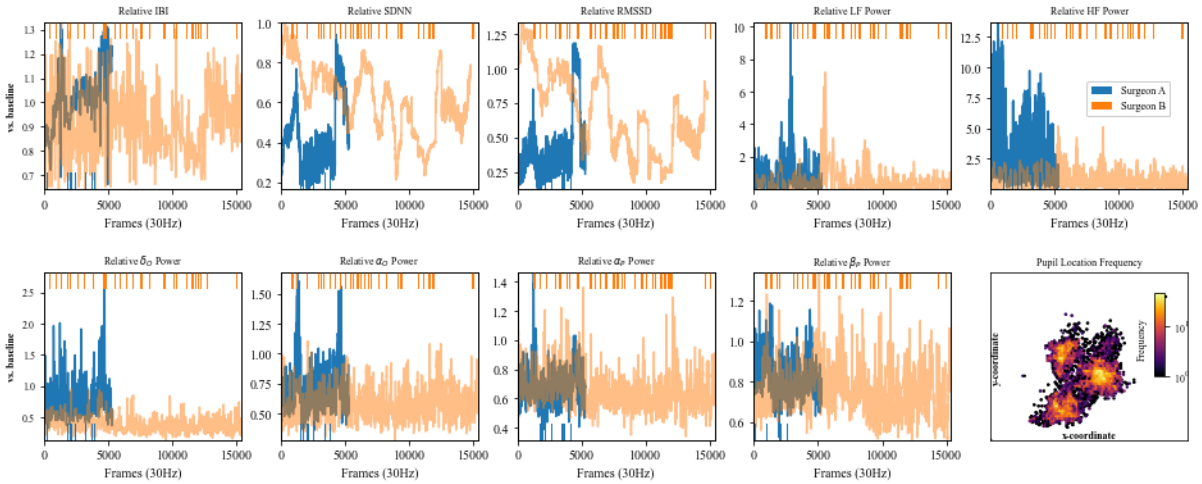


Figure 3.13. Comparison of detected individual error profiles and synchronized biometric trajectories over time for two different operators. Vertical bars represent a detected intra-task error at that task video frame. Line graphs indicate biometric time series synchronized to task video frames. The pupil localization map depicts the pupil localization frequencies in the eye camera image frame with higher intensity pixels (red) indicating more frequent localizations at that pixel coordinate for Surgeon A as compared to Surgeon B. Using SCAPEL's high temporal intra-task error detection synchronized with individual operator biometrics, we are able to derive personalized operator models for surgeons of vastly different performance levels such as Surgeon A and B shown here.

other measures of surgical efficiency to examine individual operator behavior profiles. Error rate represents the percentage of total operating console video frames from all three simulation tasks in which an intra-task error is detected. Time to completion is the total time taken to complete all simulations. Distance is the total distance traveled by both end effectors for all simulations. Average velocity is the average velocity of both end effectors for all simulations.

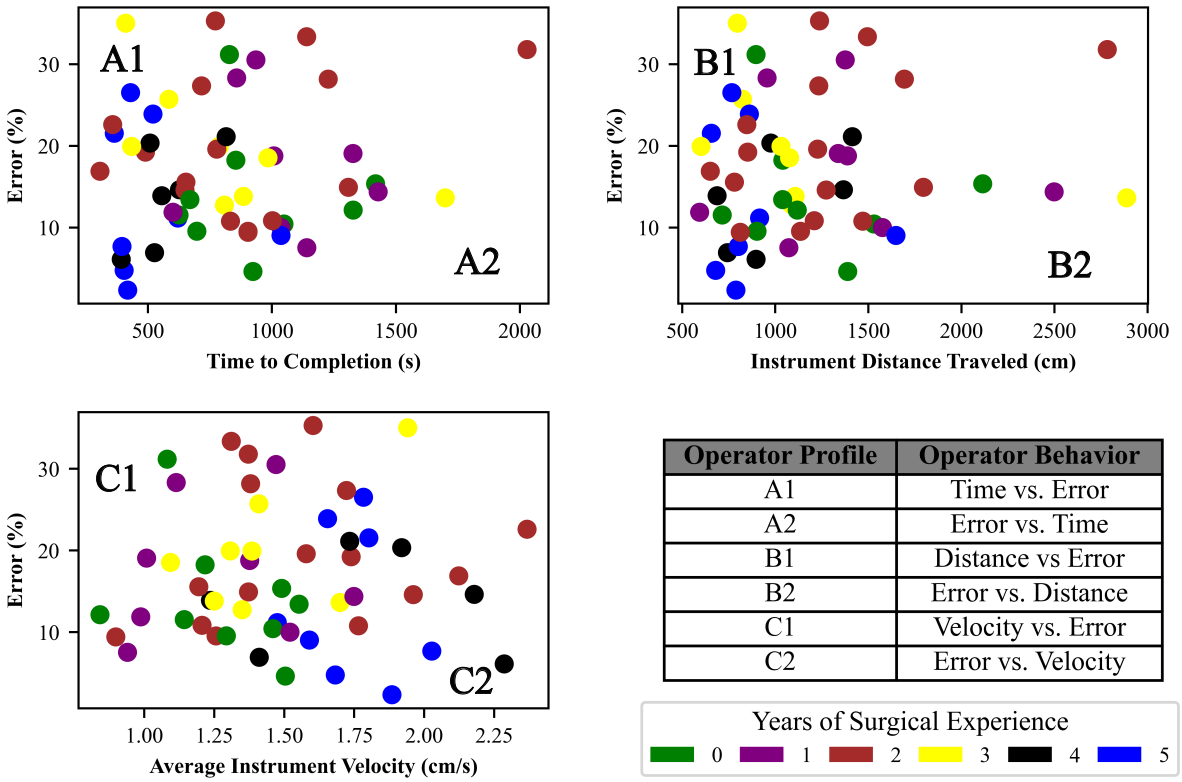
Combining intra-task error data with other surgical metrics enables recognition of a diverse set of individual operator behavior profiles. The scatter plots show a wide range of priorities with respect to error rate, time to completion, instrument distance traveled, and instrument velocity. Without this combination of high temporal resolution intra-task error detection and surgical metrics, differentiation of operator behavior profiles would not be possible. The spread of operator behaviors across different performance axes demonstrates

the need for individualized operator models.

To develop these individual operator models and to explore the error patterns and biometrics unique to each individual operator, we synchronized individual operator biometrics with high temporal resolution intra-task error detection. Fig. 3.13 displays the error patterns and biometric profiles of two different operators. High resolution intra-task error detection showed errors for Surgeon B were more evenly distributed throughout the simulations than those of Surgeon A. Surgeon A had fewer errors but those errors were more heavily concentrated in the middle segment of the surgical simulation. The biometric profiles synchronized to these error signals were also different. IBI was consistently lower and more variable for Surgeon B than Surgeon A. SDNN and RMSSD were higher and more variable throughout the simulations for Surgeon B as compared to Surgeon A. LF and HF power were greater and more variable for Surgeon A as compared to Surgeon B. Occipital delta, alpha, and parietal beta powers were also greater and more variable for Surgeon A as compared to Surgeon B. These synchronized error and biometric profiles that can be used to build interpretable, personalized operator models.

3.2.6 Operator-specific Surgical Performance Clustering

SCALPEL's capabilities provide insight into the underlying structure of operator-specific surgical performance data. Fig. 3.14 demonstrates the utility of high-resolution error detection for inferring surgical behavior. Participants in Region A1 have relatively higher error rates and lower times to completion. A personalized operator model for participants in this cluster would optimize surgical tasks for shortest time rather than least error. Participants in Region B1 have relatively higher error rates and lower total distance traveled by surgical instruments. An operator model for these participants would optimize for efficiency of movement rather than least error. Participants in Region C1 have relatively higher error rates and lower surgical instrument velocities. These participants appear to be optimizing their surgical tasks for the lowest



Operator Profile	Operator Behavior
A1	Time vs. Error
A2	Error vs. Time
B1	Distance vs Error
B2	Error vs. Distance
C1	Velocity vs. Error
C2	Error vs. Velocity

Figure 3.14. Operator surgical behavior clustering based on aggregate statistics across three simulation tasks. Error % represents the percentage of total operating console video frames from all three simulation tasks in which an intraoperative error was detected. Time to completion is the total time taken to complete all simulations. Distance is the total distance traveled by both end effectors for all simulations. Average velocity is the average velocity of both end effectors for all simulations. Regions represent inferred operator behavior.

possible speed of instrument movement rather than least error.

We can use the performance and behavior clusters from Fig. 3.14 and Fig. 3.6 to examine physiological models informed by reactions to error and segregated by performance. After clustering participants on error and time to completion, we identify High Performer (HP: low combined error and time) and Low Performer (LP: high combined error and time) groups based on their relative performance. The top 33% of performers are "High Performers". The bottom 33% of performers are "Low Performers". Fig. 3.6 displays the performance clusters.

3.3 Limitations

First, these results may be affected by other variables such as movement at the console and time at the console. Based on our observations, low performers tended to move and shift body positions more frequently than high performers, potentially indicative of discomfort at the console. Low performers also required more time to complete the simulation tasks. Second, while our linear mixed models showed significant effects, the effect sizes were small in magnitude. Third, while some metrics were reported as significant, the directionality of the effects for the High, Low, and None Performance groups warrants further investigation, particularly for effects that demonstrated a consistent direction for High and Low Performers but an opposite direction effect for the None group. Fouth, the error detection video pipeline we use is based on color and text-based error indicators that appear in the operating console during surgical simulations. These error indicators are part of the built-in simulation software and cannot be turned off or altered during simulation tasks. This presents the possibility that the neurophysiological changes that accompany error may be confounded by a response to the specific error indicator in the simulator. Fifth, we use an EKG window of 1-second, an SDNN and RMSSD window of 30-seconds, and an LF and HF power window of 25-seconds. The start of these EEG and EKG windows are sychronized to the start of the temporally closest console video frame. Because the EEG and EKG intervals are longer than the interval described by the synchronized console video frame, the EEG and EKG intervals include biometric data captured during video frames that occur after the synchronized video frame. These subsequent video frames may include both error and non-error intervals, altering the measured error or non-error effect magnitude. Lastly, the results of our study should be considered preliminary and should be verified on an independent group of participants in future studies.

3.4 Acknowledgements

Chapter 3, in part, is a reprint of material from:

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R., Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C., 2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Richter, F., Aronoff-Spencer, E., Broderick, R.C. and Christensen, H., 2023. SCALPEL Please: Personalized, Performance-Oriented Operator Models Using Biometrics and Intra-task Error Detection During Robot-Assisted Surgery. IEEE Robotics and Automation Letters, (in submission).

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 4

Modeling Robotic System Uncertainty

We focused on measuring uncertainty with respect to robotic tool localization during RAS as well as deriving novel visual uncertainty parameterizations in RAS scenes.

4.1 Methods

4.1.1 Tool Localization

Any joint-link on a surgical robot can be localized in the camera frame by combining the robots forward kinematics and the base-to-camera transform. Written explicitly, the homogeneous transform matrix describing the j-th joint-link in the camera frame is

$$^{c}\mathbf{T}_{j} = ^{c}\mathbf{T}_{b} \prod_{i=1}^{j} {}^{i-1}\mathbf{T}_{i}(q_i) \tag{4.1}$$

where $^{c}\mathbf{T}_{b} \in SE(3)$ is the base-to-camera transform and $^{i-1}\mathbf{T}_{i}(q_i) \in SE(3)$ is the i-th joint transform with joint angle q_i.

In an ideal case, the joint transforms are provided by the robot manufacturer and the joint angles are measured by encoders hence only the base-to-camera transform needs to be solved for. However, due to the unique workspace challenges faced in surgical environments (e.g. narrow corridors), surgical robotic systems often rely on cable-drive

Figure 4.1. The insertion-shaft on surgical robotic tools is an excellent feature for localization. In this work, we present a novel approach for detecting the insertion-shaft and incorporating it into Bayesian filtering to probabilistically localize and track surgical robotic tools. We stress test our method in a challenging deformable tissue dataset where a surgical robotic tool is in the back of the scene with low light and commanded to deform tissue. The tissue is registered to a simulated scene using a separate approach in the camera frame [95] and notice how well aligned the grasp between the localized surgical robotic tool and the tissue is in the 3D rendering.

designs without encoders at the joint-link location such as the dVRK [79] and RAVEN robotic systems [99, 68]. Instead joint angles are read at the motor which leads to inaccurate joint readings from mechanical phenomena such as cable stretch, hysteresis and cable stretch [66, 74]. Furthermore, the kinematic chains of surgical robots are only partially visible in the camera frame. For example, the endo-wrist and insertion-shaft from da Vinci ®surgical platform are often the only joint-links visible in the endoscopic frame.

Localizing the surgical robot in the camera frame from endoscopic images by estimating the hand-eye transform and joint angle errors is not feasible since the parameters are not identifiable from the camera frame due to the partially visible robotic chain [131].

In our previous work, we derive a smaller parameter set that is identifiable from the camera frame and enables complete localization of the surgical robotic links visible in the camera frame. The forward kinematics in (4.1) can be written with the new parameter set as

$$ {}^{c}\mathbf{T}_{j} = \mathbf{E}\prod_{i=1}^{j} {}^{i-1}\mathbf{T}_{i}(\tilde{q}_i) \tag{4.2} $$

where $\mathbf{E} \in SE(3)$ is the lumped error transform and $\tilde{q}_i$ are the noisy joint angle readings. The lumped error transform captures both the camera-to-base transform and errors in joint angle readings. Please see [131] for a complete derivation.

To estimate the lumped error, features are detected on the surgical robotic tool. The scope of this work focuses on using the insertion-shaft as a feature which can be modelled as a cylinder with the following parameters: a position on the center-line, $\mathbf{p}_0^j \in \mathbb{R}^3$, direction vector of the centerline, $\mathbf{d}^j \in \mathbb{R}^3$, and radius $r \in \mathbb{R}^+$. Through the kinematics equation in (4.2), the insert-shaft cylinder on the j-th joint can be transformed to the camera frame: $\mathbf{p}_0^c = {}^{c}\mathbf{T}_j\mathbf{p}_0^j$ and $\mathbf{d}^c = {}^{c}\mathbf{T}_j\mathbf{d}^j$. Finally, a point, $\mathbf{p} \in \mathbb{R}^3$, is defined to be on the insertion-shaft in the camera frame if it satisfies the following equations

$$\begin{cases} (\mathbf{p} - \mathbf{p}_a^c)^\top (\mathbf{p} - \mathbf{p}_a^c) - r^2 = 0 \\ (\mathbf{d}^c)^\top (\mathbf{p} - \mathbf{p}_a^c) = 0 \end{cases} \tag{4.3}$$

where $\mathbf{p}_a^c$ is a point on the center-line

$$\mathbf{p}_a^c = \mathbf{p}_0^c + \lambda \mathbf{d}^c \tag{4.4}$$

and $\lambda \in \mathbb{R}$.

Chaumette showed that projecting (4.3) with the camera pin-hole model results in two line segments that represent the two edges of the projected cylinder [26]:

$$a_1 X + b_1 Y + c_1 = 0 \qquad a_2 X + b_2 Y + c_2 = 0 \tag{4.5}$$

where (X, Y) are pixel coordinates on a unit camera (i.e. $(X, Y) = (\frac{u - c_u}{f_x}, \frac{v - c_v}{f_y})$ where $(c_u, c_v), (f_x, f_y)$ are the principle point and focal length of a non-unit camera),

$$\begin{aligned} A_{1,2} &= \frac{r\left(x_0^c - a^c (\mathbf{p}_0^c)^\top \mathbf{d}^c\right)}{\sqrt{(\mathbf{p}_0^c)^\top \mathbf{p}_0^c - (\mathbf{p}_0^c)^\top \mathbf{d}^c - r^2}} \pm (c^c y_0^c - b^c z_0^c) \\ B_{1,2} &= \frac{r\left(y_0^c - b^c (\mathbf{p}_0^c)^\top \mathbf{d}^c\right)}{\sqrt{(\mathbf{p}_0^c)^\top \mathbf{p}_0^c - (\mathbf{p}_0^c)^\top \mathbf{d}^c - r^2}} \pm (a^c z_0^c - c^c x_0^c) \\ C_{1,2} &= \frac{r\left(z_0^c - c^c (\mathbf{p}_0^c)^\top \mathbf{d}^c\right)}{\sqrt{(\mathbf{p}_0^c)^\top \mathbf{p}_0^c - (\mathbf{p}_0^c)^\top \mathbf{d}^c - r^2}} \pm (b^c x_0^c - a^c y_0^c) \end{aligned} \tag{4.6}$$

and $\mathbf{p}_0^c = \begin{bmatrix} x_0^c & y_0^c & z_0^c \end{bmatrix}^\top$, $\mathbf{d}^c = \begin{bmatrix} a^c & b^c & c^c \end{bmatrix}^\top$.

By combining the forward kinematic model in (4.2) with a the insertion-shaft projection (4.5), estimation methods can be used to find the lumped error, $\mathbf{E}$, from insertion-shaft feature detections hence localizing the surgical robotic tools in the camera frame.

To estimate the lumped error, a particle filter tracks its probability distribution

temporally using observations from the endoscopic camera. The lumped error is parameterized with a translation and axis-angle vector, $\mathbf{E}(\mathbf{b}_t, \mathbf{w}_t)$ where $\mathbf{b}_t, \mathbf{w}_t \in \mathbb{R}^3$. We describe the motion model, insertion-shaft line detection algorithm, and observation model to fully define the necessary components for a Bayesian Filter to track the lumped error and thereby localize the surgical robotic tool in the camera frame.

Motion model: The lumped error's motion model is defined with additive Gaussian noise due to its ability to generalize over a large number of random processes.

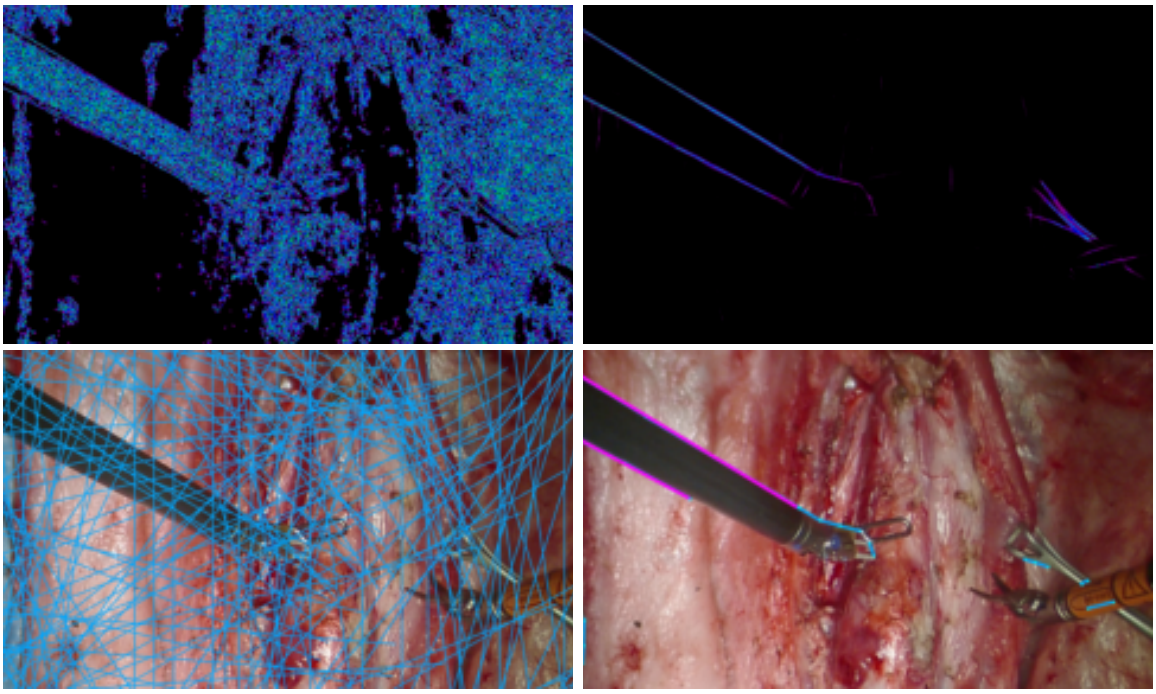
Written explicitly, the motion model is

$$\left[\mathbf{b}_{t+1}, \mathbf{w}_{t+1}\right]^\top = \mathcal{N}\left(\left[\mathbf{b}_t, \mathbf{w}_t\right]^\top, \mathbf{\Sigma}_{\mathbf{b},\mathbf{w},t+1}\right) \tag{4.7}$$

where $\mathbf{\Sigma}_{\mathbf{b},\mathbf{w},t+1} \in \mathbb{R}^{6\times6}$ is a covariance matrix.

Point Feature Detection: We detect point features in the camera image ${}^{im,d}p$ using hue, saturation, and value-based segmentation based on the current state-of-the art benchmark for surgical tool tracking [131].

Insertion-Shaft Line Detection: In our previous work, we manually tuned a canny edge detection algorithm [20] followed by a hough transform to detect lines [10]. From there, a greedy association strategy is employed to find which detected lines are associated with the edges of the insertion-shaft. In this work, we use SOLD2, a self-supervised occlusion-aware line description and detection DNN [114], for more reliable, robust insertion-shaft line detection in unstructured environments. The SOLD2 model at inference is provided a reference image of the surgical robotic tool and outputs the detected insertion-shaft lines by finding which of the detected lines are associated with the insertion-shaft from the reference image. SOLD2 also provides a heatmap of all the detected pixels associated with the detected lines.



(a) Canny (Baseline) **(b)** SOLD2 [114]

Figure 4.2. The columns show our previous [131] and proposed approach for detecting the insertion shaft when deployed in a live surgery [130]. The top row shows a heatmap of the pixels potentially associated with a line segment and the bottom row shows all the detected line segments in light blue from both approaches. Additionally, the purple line segments for the SOLD2 approach highlight the detected lines associated with the insertion shaft of the surgical tool which was not possible with the previous approach. As shown in the images, the previous approach is unable to detect and isolate shaft lines in an actual surgical scene with low light, low contrast, and significant amounts of image noise.

Written explicitly, SOLD2 outputs

$$\mathbf{e}_{1,2}^{a}, \mathbf{e}_{1,2}^{b}, \mathbf{H} = f(\mathcal{I}; \mathcal{I}_{ref}) \tag{4.8}$$

where $\mathbf{e}_{1,2}^{a}, \mathbf{e}_{1,2}^{b} \in \mathbb{R}^2$ are the endpoints of the detected insertion-shaft line segments, $\mathbf{H} \in \mathbb{I}^{H \times W}$ is the heatmap, $\mathcal{I} \in \mathbb{I}^{H \times W \times 3}$ is the input image, and $\mathcal{I}_{ref} \in \mathbb{I}^{H \times W \times 3}$ is the reference image.

An example insertion-shaft detection is shown in Fig. 4.2.

SOLD2 relies on similarity-based image features to evaluate candidate line endpoints. Surgical endoscope images, however, may have barrel distortion and lens vignetting [144] which significantly increases image feature variance between the center and periphery of the image. Lines with one endpoint located at the center of the image and one endpoint located at the periphery of the image (e.g. insertion-shaft edges) are often falsely eliminated due to the distortion and artifact-based dissimilarities in the neighborhoods of proposed endpoints. We augment our surgical endoscope images through cropping and downsampling to reduce the number of false negative endpoint detections.

Observation Model: From the insertion-shaft line detection, we present four different observation models and compare them experimentally. The first two are similar to our previous work where the uncertainty is modelled on line parameters, distance to the origin and slope angle. Meanwhile the second two observation models work with pixels directly by deriving a random variable that describes how well the pixel fits on to the projected insertion-shaft.

Observation Model #1: Endpoint Intensities to Polar . SOLD2 architecture directly outputs endpoint candidates and then completes the line segments though the heatmap [114]. Similarly, the first observation model uses the detected endpoints, $\mathbf{e}_{1,2}^{a}, \mathbf{e}_{1,2}^{b}$, to derive detected line parameters, $\rho_{1,2}^{e}, \theta_{1,2}^{e}$, which represents the distance to origin and slope angle. To increase the robustness, we derive the detected line parameters from a set of

points near the end-points that have a high heat-map value hence should be part of a line.

Written explicitly, the endpoint set is:

$$\mathcal{E}_e(\mathbf{e}) = \{\mathbf{p} \mid ||\mathbf{p} - \mathbf{e}|| \leq \alpha_e \cap \mathbf{H}(\mathbf{p}) \geq \beta\} \tag{4.9}$$

where α_e is the end-point search radius and β is the heatmap threshold.

From the end-point sets, $\mathcal{E}_e(\mathbf{e}^a_{1,2})$ and $\mathcal{E}_e(\mathbf{e}^b_{1,2})$, sequential RANSAC [52] is employed to robustly detect two line segments parameterized by $\rho^e_{1,2}, \theta^e_{1,2}$. The probabilistic observation model for time t is defined in the line parameter space and written as a summation of Gaussians, similar to our previous work:

$$\begin{aligned} P(\rho^e_{1,2,t}, &\theta^e_{1,2,t} | \mathbf{b}_t, \mathbf{w}_t) \propto \\ &\sum_{i=1,2} e^{-\gamma_\rho |\rho^e_{i,t} - \rho_i(\mathbf{b}_t, \mathbf{w}_t)| - \gamma_\theta |\theta^e_{i,t} - \theta_i(\mathbf{b}_t, \mathbf{w}_t)|} \end{aligned} \tag{4.10}$$

where γ_ρ and γ_θ describe the standard deviation of the $\rho^e_{1,2}$ and $\theta^e_{1,2}$ detections and $\rho_{1,2}(\mathbf{b}_t, \mathbf{w}_t), \theta_{1,2}(\mathbf{b}_t, \mathbf{w}_t)$ are the projected insertion-shaft lines from the estimated lumped error.

The projected insertion-shaft lines from the estimated lumped error come from (4.2) and (4.5) to generate line parameters $a_{1,2}(\mathbf{b}_t, \mathbf{w}_t), b_{1,2}(\mathbf{b}_t, \mathbf{w}_t), c_{1,2}(\mathbf{b}_t, \mathbf{w}_t)$ and subsequently converted as follows:

$$\begin{aligned} \theta_{1,2}(\mathbf{b}_t, \mathbf{w}_t) &= -\tan^{-1}\left(\frac{a_{1,2}(\mathbf{b}_t, \mathbf{w}_t)}{b_{1,2}(\mathbf{b}_t, \mathbf{w}_t)}\right) \\ \rho_{1,2}(\mathbf{b}_t, \mathbf{w}_t) &= -\frac{c_{1,2}(\mathbf{b}_t, \mathbf{w}_t)}{b_{1,2}(\mathbf{b}_t, \mathbf{w}_t)} \sin\left(\theta_{1,2}(\mathbf{b}_t, \mathbf{w}_t)\right) \quad . \end{aligned} \tag{4.11}$$

Observation Model #2: Line Intensities to Polar . The Endpoint Intensities to Polar observation model is extended in this observation model by expanding the set of points to

include points along the line segment rather than just near the detected endpoints to improve robustness by including more candidate points associated with the detected insertion-shaft line. Written explicitly, the new set of points is defined as:

$$\mathcal{E}_l(\mathbf{e}^a, \mathbf{e}^b) = \{\mathbf{p} \mid \min_{l \in L(\mathbf{e}^a, \mathbf{e}^b)} ||\mathbf{p} - l|| \leq \alpha_l \cap \mathbf{H}(\mathbf{p}) \geq \beta\} \tag{4.12}$$

where $L(\mathbf{e}^a, \mathbf{e}^b)$ is the set of points on the line segment between $\mathbf{e}^a, \mathbf{e}^b$ and α_l is the search radius about the line.

Similar to the previous observation model, line segments parameters $\rho^l_{1,2}, \theta^l_{1,2}$ are derived from the line point sets, $\mathcal{E}_l(\mathbf{e}^a_{1,2}, \mathbf{e}^b_{1,2})$, using sequential RANSAC. Observation Model #3: Detected Lines. The same probability model in (4.11) is used to define the observation model for the detected insertion-shaft lines, $\rho^l_{1,2,t}, \theta^l_{1,2,t}$, at timestep t.

Observation Model #4: Endpoint Intensities . The previous observation models use a noise distribution on derived line parameters. Instead, the pixel based observation works directly in the pixel space by deriving a random variable $R(\cdot)$ which describes how well the detected pixel fits to the projected insertion-shaft:

$$R(x, y|\theta, \rho) = \cos(\theta)x + \sin(\theta)y - \rho \tag{4.13}$$

where x, y is an input pixel coordinate.

The derived variable $R(\cdot)$ is inspired from [29] where a point is fitted to an ellipse projected from a suture needle. The fitting-based derived random variables allow for a pixel-based probability distribution without requiring pixel-to-pixel association rather a pixel-to-line association. We assume the noise of the pixels associated with the insertion-shaft's edges, $\mathbf{p}$, are independent and normally distributed from the projected line, $\mathbf{p} \sim \mathcal{N}([X, Y]^\top, \sigma^2)$ where X, Y come from the projected insertion-shaft, i.e (4.5). Therefore, the derived random variable is Gaussian whose distribution is:

$$R(\mathbf{p}|\theta,\rho) \sim \mathcal{N}\left(0, \cos(\theta)^2\sigma^2 + \sin(\theta)^2\sigma^2\right) \tag{4.14}$$

hence providing a probability distribution for a detected insertion-shaft pixel without requiring explicit pixel-to-pixel association.

Given a projected pixel $({}^{im,p}x, {}^{im,p}y)$ on a projected line ${}^{im,p}\rho, {}^{im,p}\theta$ in the camera image, we parameterize the line: ${}^{im,p}x \cos {}^{im,p}\theta + {}^{im,p}y \sin {}^{im,p}\theta = {}^{im,p}\rho$.

Given the detection uncertainty in pixel location, we create a random variable R to represent the parameters of a detected line:
$R = ({}^{im,p}x + \varepsilon_x)\cos {}^{im,p}\theta + ({}^{im,p}y + \varepsilon_y)\sin {}^{im,p}\theta - {}^{im,p}\rho$ where
$({}^{im,d}x, {}^{im,d}y) = ({}^{im,p}x + \varepsilon_x, {}^{im,p}y + \varepsilon_y)$. We assume that the pixel-wise detection errors (ε) in the x and y directions of the camera image projection are independent with zero mean such that: $\varepsilon_x \sim \mathbf{N}(0, \sigma_x^2)$, $\varepsilon_y \sim \mathbf{N}(0, \sigma_y^2)$, and $\varepsilon_x \perp\!\!\!\perp \varepsilon_y$.

The expectation of R is then $\mathrm{E}[R] = 0$. The variance of R can be derived:

$$
\begin{aligned}
\mathrm{Var}[R] &= \mathrm{Var}[({}^{im,p}x + \varepsilon_x)\cos{}^{im,p}\theta \\
&\quad + ({}^{im,p}y + \varepsilon_y)\sin{}^{im,p}\theta - {}^{im,p}\rho] \\
&= \mathrm{Var}[{}^{im,p}x\cos{}^{im,p}\theta + {}^{im,p}y\sin{}^{im,p}\theta - {}^{im,p}\rho \\
&\quad + \varepsilon_x\cos{}^{im,p}\theta + \varepsilon_y\sin{}^{im,p}\theta] \\
&= \mathrm{Var}[\varepsilon_x\cos{}^{im,p}\theta + \varepsilon_y\sin{}^{im,p}\theta] \\
&= \mathrm{E}[\varepsilon_x^2\cos^2{}^{im,p}\theta + 2\varepsilon_x\varepsilon_y\cos{}^{im,p}\theta\sin{}^{im,p}\theta \\
&\quad + \varepsilon_y^2\sin^2{}^{im,p}\theta] \\
&\quad - \left[\mathrm{E}[\varepsilon_x\cos{}^{im,p}\theta + \varepsilon_y\sin{}^{im,p}\theta]\right]^2 \\
&= \mathrm{E}[\varepsilon_x^2\cos^2{}^{im,p}\theta + 2\varepsilon_x\varepsilon_y\cos{}^{im,p}\theta\sin{}^{im,p}\theta \\
&\quad + \varepsilon_y^2\sin^2{}^{im,p}\theta] \\
&\quad - \left[\cos{}^{im,p}\theta\mathrm{E}[\varepsilon_x] + \sin{}^{im,p}\theta\mathrm{E}[\varepsilon_y]\right]^2 \\
&= \mathrm{E}[\varepsilon_x^2\cos^2{}^{im,p}\theta + 2\varepsilon_x\varepsilon_y\cos{}^{im,p}\theta\sin{}^{im,p}\theta \\
&\quad + \varepsilon_y^2\sin^2{}^{im,p}\theta] \\
&= \cos^2{}^{im,p}\theta\mathrm{E}[\varepsilon_x^2] \\
&\quad + 2\cos{}^{im,p}\theta\sin{}^{im,p}\theta\mathrm{E}[\varepsilon_x\varepsilon_y] \\
&\quad + \sin^2{}^{im,p}\theta\mathrm{E}[\varepsilon_y^2] \\
&= \cos^2{}^{im,p}\theta\left[\mathrm{Var}[\varepsilon_x] - [\mathrm{E}[\varepsilon_x]]^2\right] \\
&\quad + \sin^2{}^{im,p}\theta\left[\mathrm{Var}[\varepsilon_y] - [\mathrm{E}[\varepsilon_y]]^2\right] \\
&= \cos^2{}^{im,p}\theta\sigma_x^2 + \sin^2{}^{im,p}\theta\sigma_y^2
\end{aligned}
\tag{4.15}
$$

This demonstrates that detection uncertainty in line features can be directly estimated from raw pixel data. Further, this uncertainty incorporates information about the angle at which the surgical tool is visible in the image plane as well as camera

calibration error.

We use this distribution $R \sim \mathcal{N}(0, \cos^2{}^{im,p}\theta\sigma_x^2 + \sin^2{}^{im,p}\theta\sigma_y^2)$ to associate detected pixels in $\mathcal{E}$ with projected lines. Given a set of detected pixels $e_j \in \mathcal{E} | 0 \leq j \leq m$ where $e_j \overset{i.i.d}{\sim} R$ and a set of projected line features $\{{}^{im,p}\rho_i, {}^{im,p}\theta_i | 0 \leq i \leq n\}$ in a camera image, we define an association function $c_{j,i} = p\left(e_j | \sim N(0, \cos^2{}^{im,p}\theta_i\sigma_x^2 + \sin^2{}^{im,p}\theta_i\sigma_y^2)\right)$.

This generates a association matrix $\mathbf{C}_{m \times n}$ for all pixels in $\mathcal{E}$ and all projected line features. The j, ith entry in this association matrix is the probability that detected pixel j in $\mathcal{E}$ is associated with projected line feature i. We then use a greedy approach to match each detected pixel to the projected line feature with the highest probability. This generates a set of probability values $\mathcal{A}_x$ for all pixels in $\mathcal{E}$.

This observation model applies the derived random variable over the endpoint sets of the detected insertion-shaft, $R(\mathbf{p}_t | \theta_{1,2}(\mathbf{b}_t, \mathbf{w}_t), \rho_{1,2}(\mathbf{b}_t, \mathbf{w}_t)) \forall \mathbf{p}_t \in \mathcal{E}_e(\mathbf{e}^a_{1,2,t}), \mathcal{E}_e(\mathbf{e}^b_{1,2,t})$, for a timestep t.

Observation Model #5: Line Intensities . The Line Intensities bservation model also uses the derived random variable $R(\cdot)$ defined in (4.13) with the same uncertainty distribution from (4.14). The only difference with the previous observation model is the point sets which the derived random variable is applied to now covers the entire line, $R(\mathbf{p}_t | \theta_{1,2}(\mathbf{b}_t, \mathbf{w}_t), \rho_{1,2}(\mathbf{b}_t, \mathbf{w}_t)) \; \forall \; \mathbf{p}_t \in \mathcal{E}_l(\mathbf{e}^a_{1,2,t}, \mathbf{e}^b_{1,2,t})$, for a timestep t.

4.1.2 Visual Uncertainty

Let $\mathbf{p} = [x, y, z]^\top$ be a point of interest in 3D space. With an ideal pin-hole camera model, it will be projected to 2D space with:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{4.16}$$

where $[u, v]^\top$ is the 2D, projected pixel coordinate.

The uncertainty of recovering $\mathbf{p}$ when using stereo cameras comes from two challenges:

1. The pixel space, $[u, v]^\top$, is discretized (i.e. spatial resolution)
2. The depth, z is recovered from stereo-disparity which is also discretized (i.e. depth resolution)

In this problem, we will use stereo cameras to recover a distribution of depth and incorporate discretization effects into the spatial uncertainty.

The true pixel point, $[u, v]^\top$, corresponding to the point of interest, $\mathbf{p}$, is discretized to pixel resolution. The discretization process is modelled as rounding

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} \text{round}(u) \\ \text{round}(v) \end{bmatrix} \tag{4.17}$$

where $[\tilde{u}, \tilde{v}]^\top$ is the observed pixel of interest.

Therefore, we can model the uncertainty of the true pixel point with a uniform distribution given an observed pixel of interest:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \tilde{u} + \epsilon_u \\ \tilde{v} + \epsilon_v \end{bmatrix} \tag{4.18}$$

where $\epsilon_u, \epsilon_v \sim \mathcal{U}(-0.5, 0.5)$

From the true depth, z, of the point of interest, $\mathbf{p}$, a true disparity is derived using the stereo cameras focal lengths and baseline.

The expression for true disparity is:

$$d = \frac{tf}{z} \tag{4.19}$$

where t is the baseline of the stereo cameras and f is the focal length.

Since the observed disparity is in pixels, the discretization process is modelled as rounding:

$$\tilde{d} = \text{round}(d) \tag{4.20}$$

where $\tilde{d}$ is the observed disparity.

Therefore, the uncertainty of the true depth is modelled with a uniform distribution given an observed pixel of interest:

$$z = \frac{tf}{\tilde{d} + \epsilon_d} \tag{4.21}$$

where $\epsilon_d \sim \mathcal{U}(-0.5, 0.5)$.

Given an observed pixel of interest, $\tilde{u}, \tilde{v}$, and its observed disparity, $\tilde{d}$, a distribution of the true point of interest's 3D location can be recovered with the pin-hole camera model in (4.16).

The expression for true point, $\mathbf{p}$, when using models (4.18) and (4.21) is:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = z \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{4.22}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{tf}{\tilde{d} + \epsilon_d} \begin{bmatrix} 1/f & 0 & -c_x/f \\ 0 & 1/f & -c_y/f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{u} + \epsilon_u \\ \tilde{v} + \epsilon_v \\ 1 \end{bmatrix} \tag{4.23}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{t}{\tilde{d} + \epsilon_d} \begin{bmatrix} \tilde{u} - c_x + \epsilon_u \\ \tilde{v} - c_y + \epsilon_v \\ f \end{bmatrix} . \tag{4.24}$$

From (4.24), the probability of a point of interest, $\mathbf{p}$, given the observed pixel $\tilde{u}, \tilde{v}$

and disparity, $\tilde{d}$ can be derived.

Assume two random variables $(\mathcal{E}_u, \mathcal{E}_v)$ represent discretization errors in observed x,y-coordinate projections onto the camera image (u, v). We also define another random variable D_o to represent the true disparity at pixel coordinate (u, v) associated with object O. We define functions of those random variables Z, X, Y which are themselves also random variables:

$$Z = \frac{Bf}{D_o}$$

$$X = \frac{Z(u + \mathcal{E}_u + c_x)}{f}$$

$$Y = \frac{Z(u + \mathcal{E}_v + c_y)}{f}$$

$\underline{\boldsymbol{f_Z(z)}}$: To determine the distribution of $Z = \frac{Bf}{D_o}$ we must:

- define the distribution of D_o:
- compute $\mathcal{D}_1$: the inverse distribution of D_o
- compute $\mathcal{D}_2$: the distribution of $\mathcal{D}_1$ multiplied by the constant Bf

Define the distribution D_o:
Let $S_o = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, ..., \boldsymbol{s}_m\}$ represent the set of support points for the distribution of the disparity measure D_o. Each of these support points $\boldsymbol{s}_m$ represents a tuple of (u_m, v_m, d_m) consisting of the x-coordinate (u_m), y-coordinate (v_m), and calculated disparity (d_m) of a pixel associated with segmented object O in the camera image.

We choose D_o to be a truncated normal distribution $D_o \sim TN(\mu_o, \sigma_o, a, b)$ with a lower bound of $a = 1$ and an upper bound of $b = im_w - 1$ where im_w is equal to the image width

in pixels. The distribution is further specified by the mean calculated disparity μ_o and the standard deviation of the calculated disparities σ_o over all points in the support S_o.

We choose a truncated normal distribution rather than the normal distribution used in ELAS due to the fact that the true disparity value for any pixel must be greater than or equal to 1 and less than the image width in pixels. We parameterize the mean and variance of this distribution using a support of all pixels in the segmented object rather than using a support of any pixel occupying a similar image feature space (as in ELAS) to avoid (1) hand tuning a set of parameters to find candidate support points (as in ELAS which uses horizontal and vertical Sobel filter responses over 9x9 pixel windows, Sobel masks of size 3x3 with a stride of 5 pixels, matching correspondences based on both right-to-left and left-to-right l_1 distances with best to next-best threshold elimation of 0.9); (2) capture semantic information from the image: pixels that are part of the same object should have more similar disparity measurements.

$$D_o \sim TN(\mu_o, \sigma_o, a, b)$$

$$f_{D_o}(d_o; \mu_o, \sigma_o, a, b) = \left(\frac{1}{\sigma_o}\right) \frac{\varphi\left(\frac{d_o-\mu_o}{\sigma_o}\right)}{\Phi\left(\frac{b-\mu_o}{\sigma_o}\right) - \Phi\left(\frac{a-\mu_o}{\sigma_o}\right)} = \left(\frac{1}{\sigma_o}\right) \frac{\varphi\left(\frac{d_o-\mu_o}{\sigma_o}\right)}{\Phi\left(\frac{im_w-1-\mu_o}{\sigma_o}\right) - \Phi\left(\frac{1-\mu_o}{\sigma_o}\right)}$$

Support of $f_{D_o}(d_o; \mu_o, \sigma_o, a, b)$: $d_o \in [a, b] = d_o \in [1, im_w - 1]$

Where φ is the pdf of the standard normal distribution:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) \tag{4.25}$$

And Φ is the cdf of the standard normal distribution:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{1}{2}t^2)dt = \frac{1}{2}\left(1 + \text{erf}(\frac{x}{\sqrt{2}})\right) \tag{4.26}$$

We can numerically approximate Φ using the following:

(https://www.ijser.org/researchpaper/Approximations-to-Standard-Normal-Distribution-Function.pdf)

$$\Phi(x) \approx 1 - \frac{-\frac{x^2}{2}}{\frac{44}{79} + \frac{8}{5}x + \frac{5}{6}\sqrt{x^2+3}} \tag{4.27}$$

To find $\mathcal{D}_1$:

$W = g(D_o) = \dfrac{1}{D_o}$

$$\begin{aligned} F_W(w) &= P(W \leq w) \\ &= P(g(D_o) \leq w) \\ &= P(\frac{1}{D_o} \leq w) \\ &= P(D_o \geq \frac{1}{w}) \\ &= 1 - P(D_o \leq \frac{1}{w}) \\ &= 1 - F_{D_o}(\frac{1}{w}) \end{aligned} \tag{4.28}$$

$$
\begin{aligned}
f_W(w) &= \frac{\partial F_W(w)}{\partial w} \\
&= \frac{\partial \left(1 - F_{D_o}(\frac{1}{w}) \right)}{\partial w} \\
&= \frac{1}{w^2} f_{D_o}(\frac{1}{w})
\end{aligned}
\tag{4.29}
$$

$$
f_W(w) = \mathcal{D}_1 = \begin{cases} \left(\frac{1}{w^2 \sigma_o} \right) \frac{\varphi\left(\frac{\frac{1}{w} - \mu_o}{\sigma_o} \right)}{\Phi\left(\frac{im_w - 1 - \mu_o}{\sigma_o} \right) - \Phi\left(\frac{1 - \mu_o}{\sigma_o} \right)} & \frac{1}{im_w - 1} \leq w \leq 1 \\ 0 & \text{otherwise} \end{cases}
$$

To find $\mathcal{D}_2$:

$$
\begin{aligned}
Y &= \frac{1}{D_o} \\
Y &\sim \mathcal{D}_1
\end{aligned}
$$

$$
f_Y(y) = \mathcal{D}_1 = \begin{cases} \left(\frac{1}{y^2 \sigma_o} \right) \frac{\varphi\left(\frac{\frac{1}{y} - \mu_o}{\sigma_o} \right)}{\Phi\left(\frac{im_w - 1 - \mu_o}{\sigma_o} \right) - \Phi\left(\frac{1 - \mu_o}{\sigma_o} \right)} & \frac{1}{im_w - 1} \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}
$$

$W = g(Y) = BfY$

$$
\begin{aligned}
F_W(w) &= P(W \leq w) \\
&= P(g(Y) \leq w) \\
&= P(BfY \leq w) \\
&= P(Y \leq \frac{w}{Bf}) \\
&= F_Y(\frac{w}{Bf})
\end{aligned}
\tag{4.30}
$$

$$
\begin{aligned}
f_W(w) &= \frac{\partial F_W(w)}{\partial w} \\
&= \frac{\partial F_Y(\frac{w}{Bf})}{\partial w} \\
&= \frac{1}{Bf} f_Y(\frac{w}{Bf})
\end{aligned}
\tag{4.31}
$$

$$
f_W(w) = \mathcal{D}_3 = \begin{cases} \left(\frac{1}{Bfy^2\sigma_o}\right) \frac{\varphi\left(\frac{\frac{Bf}{w}-\mu_o}{\sigma_o}\right)}{\Phi\left(\frac{im_w-1-\mu_o}{\sigma_o}\right) - \Phi\left(\frac{1-\mu_o}{\sigma_o}\right)} & \frac{Bf}{im_w - 1} \leq w \leq Bf \\ 0 & \text{otherwise} \end{cases}
$$

Therefore the distribution $f_Z(z)$:

$$f_Z(z) = \begin{cases} \left(\frac{1}{Bfz^2\sigma_o}\right) \frac{\varphi\left(\frac{\frac{Bf}{z}-\mu_o}{\sigma_o}\right)}{\Phi\left(\frac{im_w-1-\mu_o}{\sigma_o}\right) - \Phi\left(\frac{1-\mu_o}{\sigma_o}\right)} & \frac{Bf}{im_w-1} \leq z \leq Bf \\ 0 & \text{otherwise} \end{cases}$$

As above, $f_{X|Z}(x;z)$ and $f_{Y|Z}(y;z)$ are the same:

$$f_{X|Z}(x;z) = \begin{cases} \frac{f}{z} & \frac{z}{f}(u-0.5-c_x) \leq x \leq \frac{z}{f}(u+0.5-c_x) \\ 0 & \text{otherwise} \end{cases}$$

$$f_{Y|Z}(y;z) = \begin{cases} \frac{f}{z} & \frac{z}{f}(v-0.5-c_y) \leq y \leq \frac{z}{f}(v+0.5-c_y) \\ 0 & \text{otherwise} \end{cases}$$

Therefore:

$$\begin{aligned} f_{X,Y,Z}(x,y,z) &= f_{X|Z}(x;z) f_{Y|Z}(y;z) f_Z(z) \\ &= \left(\frac{f}{z}\right)\left(\frac{f}{z}\right)\left(\frac{1}{Bfz^2\sigma_o}\right) \frac{\varphi\left(\frac{\frac{Bf}{z}-\mu_o}{\sigma_o}\right)}{\Phi\left(\frac{im_w-1-\mu_o}{\sigma_o}\right) - \Phi\left(\frac{1-\mu_o}{\sigma_o}\right)} \\ &= \left(\frac{f}{Bz^4\sigma_o}\right) \frac{\varphi\left(\frac{\frac{Bf}{z}-\mu_o}{\sigma_o}\right)}{\Phi\left(\frac{im_w-1-\mu_o}{\sigma_o}\right) - \Phi\left(\frac{1-\mu_o}{\sigma_o}\right)} \end{aligned} \tag{4.32}$$

$$f_{X,Y,Z}(x,y,z) = \begin{cases} \left(\frac{f}{Bz^4\sigma_o}\right) \frac{\varphi\left(\frac{\frac{Bf}{z}-\mu_o}{\sigma_o}\right)}{\Phi\left(\frac{im_w-1-\mu_o}{\sigma_o}\right) - \Phi\left(\frac{1-\mu_o}{\sigma_o}\right)} & \frac{Bf}{im_w-1} \leq z \leq Bf \\ & \frac{z}{f}(u-0.5-c_y) \leq x \\ & x \leq \frac{z}{f}(u+0.5-c_y) \\ & \frac{z}{f}(v-0.5-c_y) \leq y \\ & y \leq \frac{z}{f}(v+0.5-c_y) \\ 0 & \text{otherwise} \end{cases}$$

4.2 Acknowledgements

Chapter 4, in part, is a reprint of material from:

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 5

Conclusion

In this dissertation, we developed novel machine intelligence algorithms to quantitatively model and augment performance during robot-assisted surgery for both human operators and autonomous systems. For human operators, we analyzed intraoperative error and operator biometric data. For autonomous systems, we used perception algorithms to measure tool localization accuracy and visual scene uncertainty in surgical environments.

We tested the feasibility and utility of measuring physiological signals during performance on RAS simulations to determine if these signals could be assocated with intraoperative error. Our results demonstrate that it is possible to measure physiological indices associated with intraoperative errors and opens the possibility for application as a surgical training tool and as an early warning system during RAS.

Surgical education and surgical performance improvement are of paramount importance to patients, surgeons, and surgical trainees [37]. Research into the relationship between surgeon physiology and performance during RAS can help enhance surgical outcomes and improve training and education ([37, 104, 142, 49, 21]). Precision surgical skill evaluation and development may drive improved patient outcomes through higher operator technical proficiency [63].

Past studies have shown that physiological metrics can be utilized to provide insights

into the cognitive and emotional states of surgeons and their impact on surgical performance [138, 169, 67]). A recent comprehensive review investigated the impact of cognitive biases in surgery, revealing that biases like overconfidence, anchoring, and confirmation bias can negatively impact surgical performance and patient outcomes, highlighting a research gap in understanding the sources of these biases and strategies to mitigate their effects [6]. The current study utilized physiological metrics previously associated with cognition to create an empirical platform to link such cognitive states to objective biomarkers.

Several previous research studies have focused on cardiac activity during surgery as a way to monitor surgeon performance. EKG-based biometrics coupled with objective intraoperative error detection offer a potential method for personalized surgical performance evaluation and targeted surgical skill improvement [104]. For example, in one systematic review, heart rate variability was identified as a reliable method for measuring mental stress in surgical settings, pinpointing stressors, and determining the most stressful surgical techniques for the performer [152]. Other potential cognitive and affective determinants of surgical performance, including stress, mental workload, and valence, have also been explored in prior research
[37, 104, 142, 63, 133, 124, 15, 72, 163, 39, 88, 76, 164, 5, 8]. These determinants have been associated with both physiological and neurophysiological metrics, including EKG statistics [140, 141, 85, 128]. Time and frequency-domain EKG measures are particularly relevant given their proposed mappings to cognitive and affective features [42, 118, 116]. Additionally, these EKG measures have been associated with surgical performance independent of any cognitive or affective mediators [133].

Prior work, however, neglects the potential of on-line EKG analysis coupled with automated surgical error detection to capture high temporal-resolution indicators of intraoperative success or failure [133, 124, 15, 72, 163, 39, 88, 76, 164, 5, 8]. While promising, these previous studies have primarily performed retrospective evaluations of

EKG metrics and correlated them with manual ratings of surgical performance or error [15, 88, 164, 5]). However, given the retrospective nature of these methods and oftentimes subjective nature of these results, generalizability and applicability are limited. Our work attempts to link on-line intraoperative error detection during simulated surgery with biometric data captured during surgery for potential uses in both performance monitoring and education.

EEG also has the potential to provide indicators of attention, cognitive load, fatigue, affect, and error recognition, which have been shown to correlate with task performance in demanding workspaces [58, 17, 2, 147], [33]). Similar to EKG, previous studies linking EEG with surgical task performance are largely retrospective and subjective [65, 168, 170, 174]), necessitating the integration of high-resolution intraoperative performance data with real-time neuro-physiological data. In a compelling, recent proof-of-concept study, however, Armstrong and colleagues demonstrated that EEG could predict technical errors in surgery as far out as 17 seconds before they occur, suggesting potential for an EEG-based early warning system for enhancing surgical training and patient safety [7]. The current study attempts to associate errors with changes in neurophysiology.

We designed SCALPEL to integrate with a surgical robotics platform. However, the utility of our system extends beyond surgical robotics and encompasses human-robot systems. SCALPEL provides a novel platform for potential on-line integration, synchronization and analysis of robot, biometric and task data to characterize robotic platform-specific and human operator-specific contributions to error during human-robot teamwork. In this dissertation, we have defined a set of physiological requirements for capturing and integrating biometric data with task performance data, identified a set of cognitive and affective factors relevant to human operator performance during human-robot collaboration in RAS, developed a hardware and software platform to measure human operator performance, demonstrated the platform's capability in an OR

environment, and validated the integrity of our data collection process.

Our results demonstrate that SCALPEL is reliably able to capture, synchronize and analyze data from human operators and the surgical task environment that provides insight into potential features for human operator modeling during RAS. We showed that utilizing SCALPEL for contemporaneous operator and task environment analysis results in discovery of operator and performance characteristics that can drive informative, accurate operator models. High performing and low performing operators demonstrated differences in physiological time series trajectories over the course of all simulation tasks.

Additionally, our novel localization and perception algorithms suggest that we can measure tool location and RAS visual scene uncertainty with a high degree of precision. Our experimental results demonstrate that our tool tracking algorithms are not only more accurate in structured environments than the current state-of-the art approach but are also more generalizable to less structured environments. Our active vision experiments show that visual uncertainty could be used an perception parameter to augment surgical performance during RAS.

In summary, this disseration described and demonstrated the capabilities of our novel machine intelligence algorithms to quantitatively characterize and augment performance during robot-assisted surgery for both human operators and autonomous systems.

5.1 Acknowledgements

Chapter 5, in part, is a reprint of material from:

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. The neurophysiology of intraoperative error: An EEG study of trainee surgeons during robotic-assisted surgery simulations. Frontiers in Neuroergonomics, 3, p.39.

D'Ambrosia, C., Aronoff-Spencer, E., Huang, E.Y., Goldhaber, N.H., Jacobsen, G.R.,

Sandler, B., Horgan, S., Appelbaum, L.G., Christensen, H. and Broderick, R.C., 2023. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. Surgical Endoscopy, pp.1-10.

D'Ambrosia, C., Richter, F., Aronoff-Spencer, E., Broderick, R.C. and Christensen, H., 2023. SCALPEL Please: Personalized, Performance-Oriented Operator Models Using Biometrics and Intra-task Error Detection During Robot-Assisted Surgery. IEEE Robotics and Automation Letters, (in submission).

D'Ambrosia, C., Richter, F., Chiu, Z-Y., Shinde, N., Liu, F., Christensen, H.I. and Yip, M.C. 2023. Robust Surgical Tool Tracking with Pixel-based Probabilities for Projected Geometric Primitives. IEEE International Conference on Robotics and Automation (ICRA), (in submission).

D'Ambrosia, C., Huang, E.Y., Goldhaber, N.H., Christensen, H.I., Broderick, R.C. and Appelbaum, L.G., 2023. Surgical Training, Performance, and Intraoperative Error During Robot-Assisted Surgery. Annals of Surgery, (in submission).

The dissertation author is the primary author of these papers.

Chapter 6

Reflections and Recommendations

My experiences in both the OR and the engineering lab have given me ample opportunity to reflect on possibilities for accelerating the safety, quality, and delivery of surgical care during RAS. First, an improvement in vision systems is necessary to not only improve execution of surgical procedures but also to build useful training datasets for in-depth analysis. Current endoscopic camera systems used during RAS are often of lower-quality than the cameras found on state of the art mobile phones. By enabling high-fidelity image capture, and potentially highly accurate stereo depth mapping, advanced endoscopic cameras would facilitate human operator performance, more informative datasets, better materials for augmented reality simulations, higher quality data for perception algorithms, and improved training information for autonomous systems.

Second, to improve localization accuracy and perception of surgical tools in RAS images or videos, built-in optical markers or RFID chips could be incorporated into tool construction. Adding colored markers to end effectors, similar to the ones we used in our localization experiments, would enable more accurate localization as well as aid in pose estimation of the surgical tools. RFID chips embedded in the surgical tools would allow for highly precise localization in 3D space that would not require the 2D to 3D inference pipeline that is necessary for marker-based localization.

Third, development of lower profile, wearable biometric sensors would facilitate more comprehensive monitoring of surgical personnel and enable a leap forward in surgical performance assessment and precision training. Current EEG headsets are particularly cumbersome in an OR environment, are prone to motion artifact, and have minimal tolerance for movement in confined spaces such as a robotic operating console. The development of in-ear EEG monitoring has the potential to address these challenges, but the lack of spatial resolution and paucity of available EEG channels for in-ear systems will require that collected data undergo additional post processing for interpretable, comparable results. Adapting eye tracking systems for OR and RAS OR use will also require significant investment but will also offer novel insights into surgical performance and training. Current eye tracking systems require either a monitor-mounted sensor with a clear line-of-sight to the target's pupils or a wearable set of eyeglasses. Both of these constructions are difficult to use in an OR setting. Surgeons in non-RAS ORs gaze downwards into their operating field. Their pupils are not visible to screen mounted sensors. Surgeons in RAS ORs use operating consoles that do not accommodate screen mounted sensors and have space constraints that prohibit eye tracking glasses. Finally, motion tracking systems for both individuals and surgical teams in the OR could provide valuable information about individual motion, posture, and physical efficiency, team dynamics, efficiency and cooperation, and even OR utilization, physical layout efficiency, and resource demands.

Lastly, demonstrating the utility of high-fidelity biometric and robotic system data would allow for improved performance assessments and training. In many other domains from professional sports to competitive chess, performance coaches use multimodal physiological and neurophysiological data to understand individual responses to training, improve recovery, and elevate "game-day" performance. The surgical domain could adopt a similar approach. Every patient wants their surgeon at their cognitive, physical, and psychological peak on their OR day because every surgical outcome depends on the

surgeon's performance. Using tools such as heart rate variability analysis, sleep monitoring, cognitive load assessment, and stress levels to guide OR staffing and even in-OR interventions by autonomous systems could improve surgeon health and skill, and more importantly, patient outcomes.

Bibliography

[1] Phillip L Ackerman. *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications.* APA, 2011.

[2] Phillip L Ackerman. *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications.* American Psychological Association, 2011.

[3] Geoffrey L Ahern and Gary E Schwartz. Differential lateralization for positive and negative emotion in the human brain: Eeg spectral analysis. *Neuropsychologia*, 23(6):745–755, 1985.

[4] Maryam Alimardani and Kazuo Hiraki. Passive brain-computer interfaces for enhanced human-robot interaction. *Front. Robot. AI*, 7, 2020.

[5] I Amirian, L Toftegård Andersen, J Rosenberg, and I Gögenur. Decreased heart rate variability in surgeons during night shifts. *Can J Surg*, 57(5):300–304, 2014. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

[6] Bonnie A Armstrong, Ilinca A Dutescu, Arthur Tung, Diana N Carter, Patricia L Trbovich, Sherman Wong, Gustavo Saposnik, and Teodor Grantcharov. Cognitive biases in surgery: systematic review. *British Journal of Surgery*, 110(6):645–654, 2023.

[7] Bonnie A Armstrong, Dan Nemrodov, Arthur Tung, Simon J Graham, and Teodor Grantcharov. Electroencephalography can provide advance warning of technical errors during laparoscopic surgery. *Surgical Endoscopy*, 37(4):2817–2825, 2023.

[8] S Arora, N Sevdalis, D Nestel, et al. The impact of stress on surgical performance: a systematic review of the literature. *Surgery*, 147(3):318–330.e6, 2010. [PubMed] [CrossRef] [Google Scholar].

[9] Charles M Balch, Julie A Freischlag, and Tait D Shanafelt. Stress and burnout among surgeons: understanding and managing the syndrome and avoiding the adverse consequences. *Archives of surgery*, 144(4):371–376, 2009.

[10] Dana H Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981.

[11] Mehrdad J Bani. Autonomous camera movement for robotic-assisted surgery: A survey. *International Journal of Advanced Engineering, Management and Science*, 3(8):239897, 2017.

[12] Marni YV Bekkedal, John Rossi III, and Jaak Panksepp. Human brain eeg indices of emotions: delineating responses to affective vocalizations by measuring frontal theta event-related synchronization. *Neuroscience & Biobehavioral Reviews*, 35(9):1959–1970, 2011.

[13] Chris Berka, Daniel J Levendowski, Michelle N Lumicao, Alan Yau, Gene Davis, Vladimir T Zivkovic, Richard E Olmstead, Patrice D Tremoulet, and Patrick L Craven. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(5):B231–B244, 2007.

[14] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A Robbins. The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Frontiers in neuroinformatics*, 9:16, 2015.

[15] Bartholomäus Böhm, Nikolaus Rötting, Wolfgang Schwenk, Susanne Grebe, and Ulrich Mansmann. A prospective randomized trial on heart rate variability of the surgical team during laparoscopic and conventional sigmoid resection. *Archives of Surgery*, 136(3):305–310, 2001.

[16] Gianluca Borghini, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia, and Fabio Babiloni. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.*, 44:58–75, 2014.

[17] Gianluca Borghini, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia, and Fabio Babiloni. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44:58–75, 2014.

[18] Yuri Boykov, Olga Veksler, and Ramin Zabih. Markov random fields with efficient approximations. In *CVPR*, pages 648–655, 1998.

[19] Hua Cai and Yingzi Lin. Modeling of operators' emotion and task performance in a virtual driving environment. *Int. J. Hum. Comput. Stud.*, 69(9):571–586, 2011.

[20] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[21] C Melody Carswell, Duncan Clarke, and W Brent Seales. Assessing mental workload during laparoscopic surgery. *Surgical innovation*, 12(1):80–90, 2005.

[22] Ken Catchpole, Ann Bisantz, M Susan Hallbeck, Matthias Weigl, Rebecca Randell, Merrick Kossack, and Jennifer T Anger. Human factors in robotic assisted surgery: Lessons from studies 'in the wild'. *Applied ergonomics*, 78:270–276, 2019.

[23] James F Cavanagh and Michael J Frank. Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences*, 18(8):414–421, 2014.

[24] Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini, and Tzyy-Ping Jung. Evaluation of artifact subspace reconstruction for automatic eeg artifact removal. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1242–1245. IEEE, 2018.

[25] RL Charles and J Nixon. Measuring mental workload using physiological measures: a systematic review. *Appl Ergon*, 74:221–232, 2019. [PubMed] [CrossRef] [Google Scholar].

[26] François Chaumette. *La relation vision-commande: théorie et application à des tâches robotiques*. PhD thesis, L'Université de Rennes I, 1990.

[27] Jian Chen, Nathan Cheng, Giovanni Cacciamani, Paul Oh, Michael Lin-Brande, Daphne Remulla, Inderbir S Gill, and Andrew J Hung. Objective assessment of robotic surgical technical skill: a systematic review. *The Journal of urology*, 201(3):461–469, 2019.

[28] Li Cheng and Terry Caelli. Bayesian stereo matching. *Computer Vision and Image Understanding*, 106:85–96, 2007.

[29] Zih-Yun Chiu, Florian Richter, and Michael C Yip. Real-time constrained 6d object-pose tracking of an in-hand suture needle for minimally invasive robotic surgery. *arXiv preprint arXiv:2210.11973*, 2022.

[30] Caprice K Christian, Michael L Gustafson, Emilie M Roth, Thomas B Sheridan, Tejal K Gandhi, Kathleen Dwyer, Michael J Zinner, and Meghan M Dierks. A prospective study of patient safety in the operating room. *Surgery*, 139(2):159–173, 2006.

[31] Cristina Garcia Cifuentes, Jan Issac, Manuel Wüthrich, Stefan Schaal, and Jeannette Bohg. Probabilistic articulated real-time tracking for robot manipulation. *IEEE Robotics and Automation Letters*, 2(2):577–584, 2016.

[32] Iris Cohen, Nadia den Braber, Nanja JJM Smets, Jurriaan van Diggelen, Willem-Paul Brinkman, and Mark A Neerincx. Work content influences on cognitive task load, emotional state and performance during a simulated 520-days' mars mission. *Comput. Hum. Behav.*, 55:642–652, 2016.

[33] Iris Cohen, Nadia den Braber, Nanja JJM Smets, Jurriaan van Diggelen, Willem-Paul Brinkman, and Mark A Neerincx. Work content influences on cognitive

task load, emotional state and performance during a simulated 520-days' mars mission. *Computers in Human Behavior*, 55:642–652, 2016.

[34] Marco Cuturi and Mathieu Blondel. Soft-dtw: a differentiable loss function for time-series. In *International conference on machine learning*, pages 894–903. PMLR, 2017.

[35] Praveen Damacharla, Ahmad Y Javaid, Jennie J Gallimore, and Vijay K Devabhaktuni. Common metrics to benchmark human-machine teams (hmt): A review. *IEEE Access*, 6:38637–38655, 2018.

[36] Christopher D'Ambrosia, Eliah Aronoff-Spencer, Estella Y Huang, Nicole H Goldhaber, Henrik I Christensen, Ryan C Broderick, and Lawrence G Appelbaum. The neurophysiology of intraoperative error: An eeg study of trainee surgeons during robotic-assisted surgery simulations. *Frontiers in Neuroergonomics*, 3:39, 2023.

[37] A Darzi, S Smith, and N Taffinder. Assessing operative skill needs to become more objective. *BMJ*, 318(7188):887–888, 1999. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

[38] Fabio Dell'Agnola, Leila Cammoun, and David Atienza. Physiological characterization of need for assistance in rescue missions with drones. In *IEEE ICCE*, pages 1–6. IEEE, 2018.

[39] Roger D Dias, Minhtran C Ngo-Howard, Marko T Boskovski, Marco A Zenati, and Steven J Yule. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *Journal of British Surgery*, 105(5):491–501, 2018.

[40] Joseph P Distefano, Hemanth Manjunatha, Souma Chowdhury, Karthik Dantu, David Doermann, and Ehsan T Esfahani. Using physiological information to classify task difficulty in human-swarm interaction. In *IEEE SMC*, pages 1198–1203. IEEE, 2021.

[41] Jeffrey D Doyle, Eric M Webber, and Ravi S Sidhu. A universal global rating scale for the evaluation of technical skills in the operating room. *The American journal of surgery*, 193(5):551–555, 2007.

[42] S Duschek, M Muckenthaler, N Werner, and del Paso GAR. Relationships between features of autonomic cardiovascular control and cognitive performance. *Biol Psychol*, 81(2):110–117, 2009. [PubMed] [CrossRef] [Google Scholar].

[43] Christopher D'Ambrosia, Eliah Aronoff-Spencer, Estella Y Huang, Nicole H Goldhaber, Garth R Jacobsen, Bryan Sandler, Santiago Horgan, Lawrence G Appelbaum, Henrik Christensen, and Ryan C Broderick. The physiology of intraoperative error: using electrokardiograms to understand operator performance during robot-assisted surgery simulations. *Surgical Endoscopy*, pages 1–10, 2023.

[44] Shaharam Eivazi, Thiago Santini, Alireza Keshavarzi, Thomas Kübler, and Andrea Mazzei. Improving real-time cnn-based pupil detection through domain-specific data augmentation. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pages 1–6, 2019.

[45] Shahab Eslamian, Luke A Reisner, and Abhilash K Pandya. Development and evaluation of an autonomous camera control algorithm on the da vinci surgical system. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 16(2):e2036, 2020.

[46] L. et al. Feature selection: A data perspective. *ACM computing surveys (CSUR)*, 50(6):1–45, 2017.

[47] N. et al. Robotic surgery training and performance: identifying objective variables for quantifying the extent of proficiency. *Surgical Endoscopy And Other Interventional Techniques*, 20:96–103, 2006.

[48] S. et al. Surgical robotics and laparoscopic training drills. *Journal of Endourology*, 18(1):63–67, 2004.

[49] A Eversbusch and TP Grantcharov. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. *Surgical Endoscopy And Other Interventional Techniques*, 18(10):1514–1518, 2004.

[50] Irene Fassi and Giovanni Legnani. Hand to sensor calibration: A geometrical interpretation of the matrix equation ax= xb. *Journal of Robotic Systems*, 22(9):497–506, 2005.

[51] Pedro Felzenszwalb and Daniel Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70:41–54, 2006.

[52] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[53] Julia Natascha Frey, Philipp Ruhnau, and Nathan Weisz. Not so different after all: The same oscillatory processes support different types of attention. *Brain research*, 1626:183–197, 2015.

[54] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient large-scale stereo matching. In *Asian conference on computer vision*, pages 25–38. Springer, 2010.

[55] Miriam Gil, Manoli Albert, Joan Fons, and Vicente Pelechano. Designing human-in-the-loop autonomous cyber-physical systems. *International journal of human-computer studies*, 130:21–39, 2019.

[56] Rahel Gilgen-Ammann, Theresa Schweizer, and Thomas Wyss. Rr interval signal quality of a heart rate monitor and an ecg holter at rest and during exercise. *Eur. J. Appl. Physiol.*, 119(7):1525–1532, 2019.

[57] Karamjit S Gill. *Human machine symbiosis: The foundations of human-centred systems design*. Springer Science & Business Media, 2012.

[58] Karamjit S Gill. *Human machine symbiosis: The foundations of human-centred systems design*. Springer Science & Business Media, 2012.

[59] Vera C Goessl, Joshua E Curtiss, and Stefan G Hofmann. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. *Psychological medicine*, 47(15):2578–2586, 2017.

[60] Alvin C Goh, David W Goldfarb, James C Sander, Brian J Miles, and Brian J Dunkin. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *The Journal of urology*, 187(1):247–252, 2012.

[61] Mateusz Gola, Mikołaj Magnuski, Izabela Szumska, and Andrzej Wróbel. Eeg beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects. *International Journal of Psychophysiology*, 89(3):334–341, 2013.

[62] Ross Gordon, Joseph Ciorciari, and Tom van Laer. Using eeg to examine the role of attention, working memory, emotion, and imagination in narrative transportation. *European Journal of Marketing*, 2018.

[63] PD Grantcharov, T Boillat, S Elkabany, Katarzyna Wac, and H Rivas. Acute mental stress and surgical performance. *BJS open*, 3(1):119–125, 2019.

[64] John H Gruzelier. Eeg-neurofeedback for optimising performance. i: A review of cognitive and affective outcome in healthy participants. *Neuroscience & Biobehavioral Reviews*, 44:124–141, 2014.

[65] Khurshid A Guru, Somayeh B Shafiei, Atif Khan, Ahmed A Hussein, Mohamed Sharif, and Ehsan T Esfahani. Understanding cognitive performance during robot-assisted surgery. *Urology*, 86(4):751–757, 2015.

[66] Mohammad Haghighipanah, Muneaki Miyasaka, Yangming Li, and Blake Hannaford. Unscented kalman filter and 3d vision to improve cable driven surgical robot joint angle estimation. In *2016 IEEE international conference on robotics and automation (ICRA)*, pages 4135–4142. IEEE, 2016.

[67] Faizal A Haji, David Rojas, Ruth Childs, Sandrine de Ribaupierre, and Adam Dubrowski. Measuring cognitive load: performance, mental effort and simulation task complexity. *Medical education*, 49(8):815–827, 2015.

[68] Blake Hannaford, Jacob Rosen, Diana W Friedman, Hawkeye King, Phillip Roan, Lei Cheng, Daniel Glozman, Ji Ma, Sina Nia Kosari, and Lee White. Raven-ii: an open platform for surgical robotics research. *IEEE Transactions on Biomedical Engineering*, 60(4):954–959, 2012.

[69] Thalía Harmony. The functional significance of delta oscillations in cognitive processing. *Frontiers in integrative neuroscience*, 7:83, 2013.

[70] Caroline E Harriott and Julie A Adams. Modeling human performance for human–robot systems. *Rev. Hum. Factors Ergon.*, 9(1):94–130, 2013.

[71] AN Healey, N Sevdalis, and CA Vincent. Measuring intra-operative interference from distraction and interruption observedin the operating theatre. *Ergonomics*, 49(5-6):589–604, 2006.

[72] Jeroen Heemskerk, H Reinier Zandbergen, Sander WM Keet, Ingrid Martijnse, Gust Van Montfort, Rob JA Peters, Vesna Svircevic, R Arthur Bouwman, Cor GMI Baeten, and Nicole D Bouvy. Relax, it's just laparoscopy! a prospective randomized trial on heart rate variability of the surgeon in robot-assisted versus conventional laparoscopic cholecystectomy. *Digestive surgery*, 31(3):225–232, 2014.

[73] Shikun Hou and Elizabeth J Glover. Pi usb cam: A simple and affordable diy solution that enables high-quality, high-throughput video capture for behavioral neuroscience research. *Eneuro*, 2022.

[74] Minho Hwang, Brijen Thananjeyan, Samuel Paradis, Daniel Seita, Jeffrey Ichnowski, Danyal Fer, Thomas Low, and Ken Goldberg. Efficiently calibrating cable-driven surgical robots with rgbd fiducial sensing and recurrent neural networks. *IEEE Robotics and Automation Letters*, 5(4):5937–5944, 2020.

[75] Jessica J Ji, Sanjay Krishnan, Vatsal Patel, Danyal Fer, and Ken Goldberg. Learning 2d surgical camera motion from demonstrations. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pages 35–42. IEEE, 2018.

[76] B Joseph, S Parvaneh, T Swartz, et al. Stress among surgical attending physicians and trainees: a quantitative assessment during trauma activation and emergency surgeries. *J Trauma Acute Care Surg*, 81(4):723–728, 2016. [PubMed] [CrossRef] [Google Scholar].

[77] Timothy N Judkins, Dmitry Oleynikov, and Nick Stergiou. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surgical endoscopy*, 23:590–597, 2009.

[78] Jochen Kaiser and Werner Lutzenberger. Human gamma-band activity: a window to cognitive processing. *Neuroreport*, 16(3):207–211, 2005.

[79] Peter Kazanzides, Zihan Chen, Anton Deguet, Gregory S Fischer, Russell H Taylor, and Simon P DiMaio. An open-source research kit for the da vinci® surgical system. In *2014 IEEE international conference on robotics and automation (ICRA)*, pages 6434–6439. IEEE, 2014.

[80] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision*, pages 508–515, 2001.

[81] Desen Kong and Haidong Tao. Stereo matching via learning multiple experts behaviors. In *BMVC*, pages 90–97, 2006.

[82] Michael Krainin, Peter Henry, Xiaofeng Ren, and Dieter Fox. Manipulator and object tracking for in hand model acquisition. In *Proceedings, IEEE International Conference on Robots and Automation*, pages 1817–1824, 2010.

[83] Sylvia D Kreibig. Autonomic nervous system activity in emotion: A review. *Biological psychology*, 84(3):394–421, 2010.

[84] Ohhwan Kwon, Jinwoo Jeong, Hyung Bin Kim, In Ho Kwon, Song Yi Park, Ji Eun Kim, and Yuri Choi. Electrocardiogram sampling frequency range acceptable for heart rate variability analysis. *J. Healthc. Inform. Res.*, 24(3):198–206, 2018.

[85] S Laborde, E Mosley, and JF Thayer. Heart rate variability and cardiac vagal tone in psychophysiological research—recommendations for experiment planning, data analysis, and data reporting. *Front Psychol*, 8:213, 2017. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

[86] Jens Lambrecht. Robust few-shot pose estimation of articulated robots using monocular cameras and deep-learning-based keypoint detection. In *2019 7th International Conference on Robot Intelligence Technology and Applications (RiTA)*, pages 136–141. IEEE, 2019.

[87] Jens Lambrecht and Linh Kästner. Towards the usage of synthetic data for marker-less pose estimation of articulated robots in rgb images. In *2019 19th International Conference on Advanced Robotics (ICAR)*, pages 240–247. IEEE, 2019.

[88] C Langelotz, M Scharfenberg, O Haase, and W Schwenk. Stress and heart rate variability in surgeons during a 24-hour shift. *Arch Surg*, 143(8):751–755, 2008. [PubMed] [CrossRef] [Google Scholar].

[89] Quoc V Le and Andrew Y Ng. Joint calibration of multiple sensors. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3651–3658. IEEE, 2009.

[90] Y Lean and F Shan. Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors Ergon Manuf Serv Indust*, 22(3):177–187, 2012. [CrossRef] [Google Scholar].

[91] Timothy E Lee, Jonathan Tremblay, Thang To, Jia Cheng, Terry Mosier, Oliver Kroemer, Dieter Fox, and Stan Birchfield. Camera-to-robot pose estimation from a single image. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9426–9432. IEEE, 2020.

[92] Paul M Lehrer and Richard Gevirtz. Heart rate variability biofeedback: how and why does it work? *Frontiers in psychology*, page 756, 2014.

[93] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Ep n p: An accurate o (n) solution to the p n p problem. *International journal of computer vision*, 81:155–166, 2009.

[94] Yang Li, Florian Richter, Jingpei Lu, Emily K Funk, Ryan K Orosco, Jianke Zhu, and Michael C Yip. Super: A surgical perception framework for endoscopic tissue manipulation with surgical robotics. *IEEE Robotics and Automation Letters*, 5(2):2294–2301, 2020.

[95] Fei Liu, Zihan Li, Yunhai Han, Jingpei Lu, Florian Richter, and Michael C Yip. Real-to-sim registration of deformable soft tissue with position-based dynamics for surgical robot autonomy. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 12328–12334. IEEE, 2021.

[96] Yisi Liu and Olga Sourina. Eeg databases for emotion recognition. In *2013 international conference on cyberworlds*, pages 302–309. IEEE, 2013.

[97] Jingpei Lu, Ambareesh Jayakumari, Florian Richter, Yang Li, and Michael C Yip. Super deep: A surgical perception framework for robotic tissue manipulation using deep learning for feature extraction. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4783–4789. IEEE, 2021.

[98] Jingpei Lu, Florian Richter, and M Yip. Robust keypoint detection and pose estimation of robot manipulators with self-occlusions via sim-to-real transfer. *arXiv preprint arXiv:2010.08054*, 2020.

[99] Mitchell JH Lum, Diana CW Friedman, Ganesh Sankaranarayanan, Hawkeye King, Kenneth Fodero, Rainer Leuschke, Blake Hannaford, Jacob Rosen, and Mika N Sinanan. The raven: Design and validation of a telesurgery system. *The International Journal of Robotics Research*, 28(9):1183–1197, 2009.

[100] Jeffrey Mahler, Sanjay Krishnan, Michael Laskey, Siddarth Sen, Adithyavairavan Murali, Ben Kehoe, Sachin Patil, Jiannan Wang, Mike Franklin, Pieter Abbeel, et al. Learning accurate kinematic control of cable-driven surgical robots using data cleaning and gaussian process regression. In *2014 IEEE international conference on automation science and engineering (CASE)*, pages 532–539. IEEE, 2014.

[101] M Maruthappu, MA El-Harasis, M Nagendran, DP Orgill, P McCulloch, A Duclos, and MJ Carty. Systematic review of methodological quality of individual

performance measurement in surgery. *Journal of British Surgery*, 101(12):1491–1498, 2014.

[102] Hemel N Modi, Harsimrat Singh, Felipe Orihuela-Espina, Thanos Athanasiou, Francesca Fiorentino, Guang-Zhong Yang, Ara Darzi, and Daniel R Leff. Temporal stress in the operating room: brain engagement promotes "coping" and disengagement prompts "choking". *Annals of surgery*, 267(4):683–691, 2018.

[103] Hemel Narendra Modi, Harsimrat Singh, Guang-Zhong Yang, Ara Darzi, and Daniel Richard Leff. A decade of imaging surgeons' brain function (part i): Terminology, techniques, and clinical translation. *Surgery*, 162(5):1121–1130, 2017.

[104] K Moorthy, Y Munz, SK Sarker, and A Darzi. Objective assessment of technical skills in surgery. *BMJ*, 327(7422):1032–1037, 2003. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

[105] Carlos Hitoshi Morimoto, Dave Koons, Arnon Amir, and Myron Flickner. Pupil detection and tracking using multiple light sources. *Image Vis. Comput.*, 18(4):331–335, 2000.

[106] Sirajum Munir, John A Stankovic, Chieh-Jan Mike Liang, and Shan Lin. Cyber physical system challenges for human-in-the-loop control. In *8th International Workshop on Feedback Computing (Feedback Computing 13)*, 2013.

[107] Hiroaki Niitsu, Naoki Hirabayashi, Masanori Yoshimitsu, Takeshi Mimura, Junya Taomoto, Yoich Sugiyama, Shigeru Murakami, Shuji Saeki, Hidenori Mukaida, and Wataru Takiyama. Using the objective structured assessment of technical skills (osats) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surgery today*, 43(3):271–275, 2013.

[108] Desmond J Oathes, William J Ray, Alissa S Yamasaki, Thomas D Borkovec, Louis G Castonguay, Michelle G Newman, and Jack Nitschke. Worry, generalized anxiety disorder, and emotion: Evidence from the eeg gamma band. *Biological psychology*, 79(2):165–170, 2008.

[109] Rafal Ohme, Dorota Reykowska, Dawid Wiener, and Anna Choromanska. Application of frontal eeg asymmetry to advertising research. *Journal of economic psychology*, 31(5):785–793, 2010.

[110] Barry S Oken, Martin C Salinsky, and SM2865224 Elsas. Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin. Neurophys.*, 117(9):1885–1901, 2006.

[111] Satu Palva and J Matias Palva. New vistas for α-frequency band oscillations. *Trends in neurosciences*, 30(4):150–158, 2007.

[112] Frank C Park and Bryan J Martin. Robot sensor calibration: solving ax= xb on the euclidean group. *IEEE Transactions on Robotics and Automation*, 10(5):717–721, 1994.

[113] Peter Pastor, Mrinal Kalakrishnan, Jonathan Binney, Jonathan Kelly, Ludovic Righetti, Gaurav Sukhatme, and Stefan Schaal. Learning task error models for manipulation. In *2013 IEEE International Conference on Robotics and Automation*, pages 2612–2618. IEEE, 2013.

[114] Rémi Pautrat, Juan-Ting Lin, Viktor Larsson, Martin R Oswald, and Marc Pollefeys. Sold2: Self-supervised occlusion-aware line description and detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11368–11378, 2021.

[115] Haonan Peng, Xingjian Yang, Yun-Hsuan Su, and Blake Hannaford. Real-time data driven precision estimator for raven-ii surgical robot end effector position. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 350–356. IEEE, 2020.

[116] T Pinna and DJ Edwards. A systematic review of associations between interoception, vagal tone, and emotional regulation: potential applications for mental health, wellbeing, psychological flexibility, and chronic conditions. *Front Psychol*, 11:1792, 2020. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

[117] Alan T Pope, Edward H Bogart, and Debbie S Bartolome. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology*, 40(1-2):187–195, 1995.

[118] SW Porges. Cardiac vagal tone: a physiological index of stress. *Neurosci Biobehav Rev*, 19(2):225–233, 1995. [PubMed] [CrossRef] [Google Scholar].

[119] Mary C Potter, Brad Wyble, Carl Erick Hagmann, and Emily S McCourt. Detecting meaning in rsvp at 13 ms per picture. *Atten. Percept. Psychophys.*, 76(2):270–279, 2014.

[120] Vijay Pradeep, Kurt Konolige, and Eric Berger. Calibrating a multi-arm multi-sensor robot: A bundle adjustment approach. In *Experimental Robotics: The 12th International Symposium on Experimental Robotics*, pages 211–225. Springer, 2014.

[121] Peter Putman, Bart Verkuil, Elsa Arias-Garcia, Ioanna Pantazi, and Charlotte van Schie. Eeg theta/beta ratio as a potential biomarker for attentional control and resilience against deleterious effects of stress on attention. *Cognitive, Affective, & Behavioral Neuroscience*, 14(2):782–791, 2014.

[122] Bujar Raufi and Luca Longo. An evaluation of the eeg alpha-to-theta and theta-to-alpha band ratios as indexes of mental workload. *Frontiers in Neuroinformatics*, 16, 2022.

[123] William J Ray and Harry W Cole. Eeg alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science*, 228(4700):750–752, 1985.

[124] I Reijmerink, M van der Laan, and F Cnossen. Heart rate variability as a measure of mental stress in surgery: a systematic review. *Int Arch Occup Environ Health*, 93(7):805–821, 2020. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

[125] Carol E Reiley and Gregory D Hager. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. In *International conference on medical image computing and computer-assisted intervention*, pages 435–442. Springer, 2009.

[126] Austin Reiter, Peter K Allen, and Tao Zhao. Feature classification for tracking articulated surgical tools. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part II 15*, pages 592–600. Springer, 2012.

[127] Austin Reiter, Peter K Allen, and Tao Zhao. Appearance learning for 3d tracking of robotic surgical tools. *The International Journal of Robotics Research*, 33(2):342–356, 2014.

[128] GA Reyes del Paso, W Langewitz, LJM Mulder, A van Roon, and S Duschek. The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: a review with emphasis on a reanalysis of previous studies: Lf hrv and sympathetic cardiac tone. *Psychophysiology*, 50(5):477–487, 2013. [PubMed] [CrossRef] [Google Scholar].

[129] Rogério Richa, Antônio PL Bó, and Philippe Poignet. Robust 3d visual tracking for robotic-assisted cardiac interventions. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part I 13*, pages 267–274. Springer, 2010.

[130] Florian Richter, Emily K Funk, Won Seo Park, Ryan K Orosco, and Michael C Yip. From bench to bedside: The first live robotic surgery on the dvrk to enable remote telesurgery with motion scaling. In *2021 International Symposium on Medical Robotics (ISMR)*, pages 1–7. IEEE, 2021.

[131] Florian Richter, Jingpei Lu, Ryan K Orosco, and Michael C Yip. Robotic tool tracking under partially visible kinematic chain: A unified approach. *IEEE Transactions on Robotics*, 38(3):1653–1670, 2021.

[132] Florian Richter, Ryan K Orosco, and Michael C Yip. Motion scaling solutions for improved performance in high delay surgical teleoperation. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 1590–1595. IEEE, 2019.

[133] A Rieger, R Stoll, S Kreuzfeld, K Behrens, and M Weippert. Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress. *Int Arch Occup Environ Health*, 87(2):165–174, 2014. [PubMed] [CrossRef] [Google Scholar].

[134] Kilseop Ryu and Rohae Myung. Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *Int. J. Ind. Ergon.*, 35(11):991–1009, 2005.

[135] Joan Santamaria and Keith H Chiappa. The eeg of drowsiness in normal adults. *Journal of clinical Neurophysiology*, 4(4):327–382, 1987.

[136] Sudip K Sarker and Charles Vincent. Errors in surgery. *International Journal of Surgery*, 3(1):75–81, 2005.

[137] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47:7–42, 2002.

[138] M Schuetz, I Gockel, J Beardi, P Hakman, F Dunschede, S Moenk, W Heinrichs, and TH Junginger. Three different types of surgeon-specific stress reactions identified by laparoscopic simulation in a virtual scenario. *Surgical endoscopy*, 22(5):1263–1267, 2008.

[139] Daniel Seita, Sanjay Krishnan, Roy Fox, Stephen McKinley, John Canny, and Ken Goldberg. Fast and reliable autonomous surgical debridement with cable-driven robots using a two-phase calibration procedure. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6651–6658, 2018.

[140] Fred Shaffer and Jay P Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, page 258, 2017.

[141] Fred Shaffer, Zachary M Meehan, and Christopher L Zerr. A critical review of ultra-short-term heart rate variability norms research. *Frontiers in neuroscience*, 14:594880, 2020.

[142] J Shah and A Darzi. Surgical skills assessment: an ongoing debate: assessing surgical skills. *BJU Int*, 88(7):655–660, 2001. [PubMed] [CrossRef] [Google Scholar].

[143] Guofa Shou and Lei Ding. Frontal theta eeg dynamics in a real-world air traffic control task. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5594–5597. IEEE, 2013.

[144] Warren E Smith, Nimish Vakil, and Seth A Maislin. Correction of distortion in endoscope images. *IEEE Transactions on Medical Imaging*, 11(1):117–122, 1992.

[145] Winnie KY So, Savio WH Wong, Joseph N Mak, and Rosa HM Chan. An evaluation of mental workload with frontal eeg. *PloS one*, 12(4):e0174949, 2017.

[146] Mark A Staal. *Stress, cognition, and human performance: A literature review and conceptual framework*. Citeseer, 2004.

[147] Mark A Staal. Stress, cognition, and human performance: A literature review and conceptual framework. 2004.

[148] Arjen M Strijkstra, Domien GM Beersma, Berdine Drayer, Nynke Halbesma, and Serge Daan. Subjective sleepiness correlates negatively with global alpha (8–12 hz) and positively with central frontal theta (4–8 hz) frequencies in the human resting awake electroencephalogram. *Neuroscience letters*, 340(1):17–20, 2003.

[149] John Sweller. Cognitive load theory. In *Psychology of learning and motivation*, volume 55, pages 37–76. Elsevier, 2011.

[150] Lingling Tao, Ehsan Elhamifar, Sanjeev Khudanpur, Gregory D Hager, and René Vidal. Sparse hidden markov models for surgical gesture classification and skill evaluation. In *International conference on information processing in computer-assisted interventions*, pages 167–177. Springer, 2012.

[151] Karl Halvor Teigen. Yerkes-dodson: A law for all seasons. *Theory Psychol.*, 4(4):525–547, 1994.

[152] Anne-Fleur The, Iris Reijmerink, Maarten van der Laan, and Fokie Cnossen. Heart rate variability as a measure of mental stress in surgery: a systematic review. *International archives of occupational and environmental health*, 93:805–821, 2020.

[153] Gregor Thut, Annika Nietzel, Stephan A Brandt, and Alvaro Pascual-Leone. α-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience*, 26(37):9494–9502, 2006.

[154] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[155] Tony Tien, Philip H Pucher, Mikael H Sodergren, Kumuthan Sriskandarajah, Guang-Zhong Yang, and Ara Darzi. Eye tracking for skills assessment and training: a systematic review. *journal of surgical research*, 191(1):169–178, 2014.

[156] Beatrice van Amsterdam, Matthew J Clarkson, and Danail Stoyanov. Gesture recognition in robotic surgery: a review. *IEEE Transactions on Biomedical Engineering*, 68(6), 2021.

[157] Balakrishnan Varadarajan, Carol Reiley, Henry Lin, Sanjeev Khudanpur, and Gregory Hager. Data-derived models for segmentation with application to surgical assessment and training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 426–434. Springer, 2009.

[158] Gyanendra K Verma and Uma Shanker Tiwary. Affect representation and recognition in 3d continuous valence–arousal–dominance space. *Multimedia Tools and Applications*, 76(2):2159–2183, 2017.

[159] von B. et al. Does practice make perfect? laparoscopic training mainly improves motion efficiency: a prospective trial. *Updates in Surgery*, pages 1–13, 2023.

[160] Feng Wang, Kai Chen, and Xiaoping Chen. An online calibration method for manipulator with joint clearance. *Robot*, 35(5):521–526, 2013.

[161] Ziheng Wang and Ann Majewicz Fey. Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. *International journal of computer assisted radiology and surgery*, 13(12):1959–1970, 2018.

[162] Michael Weber, Martin Humenberger, and Wilfried Kubinger. A very fast census-based stereo matching implementation on a graphics processing unit. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 786–793. IEEE, 2009.

[163] M Weenk, APB Alken, LJLPG Engelen, SJH Bredie, TH van de Belt, and H van Goor. Stress measurement in surgeons and residents using a smart patch. *Am J Surg*, 216(2):361–368, 2018. [PubMed] [CrossRef] [Google Scholar].

[164] Cordula M Wetzel, Stephen A Black, George B Hanna, Thanos Athanasiou, Roger L Kneebone, Debra Nestel, John HN Wolfe, and Maria Woloshynowych. The effects of stress and coping on surgical performance during simulations. *Annals of surgery*, 251(1):171–176, 2010.

[165] Oliver Woodford, Philip Torr, Ian Reid, and Andrew Fitzgibbon. Global stereo reconstruction under second-order smoothness priors. *PAMI*, 31:2115–2128, 2009.

[166] David L Woods, John M Wyma, E William Yund, Timothy J Herron, and Bruce Reed. Factors influencing the latency of simple reaction time. *Frontiers in human neuroscience*, 9:131, 2015.

[167] Chuhao Wu, Jackie Cha, Jay Sulek, Tian Zhou, Chandru P Sundaram, Juan Wachs, and Denny Yu. Eye-tracking metrics predict perceived workload in robotic surgical skills training. *Human factors*, 62(8):1365–1386, 2020.

[168] Denny Yu, Cem Dural, Melissa Morrow, Liyun Yang, Justin W Collins, Susan Hallbeck, Magnus Kjellman, and Mikael Forsman. Intraoperative workload in robotic surgery assessed by wearable motion tracking sensors and questionnaires. *Surgical endoscopy*, 31(2):877–886, 2017.

[169] Yuliya Y Yurko, Mark W Scerbo, Ajita S Prabhu, Christina E Acker, and Dimitrios Stefanidis. Higher mental workload is associated with poorer laparoscopic performance as measured by the nasa-tlx tool. *Simulation in healthcare*, 5(5):267–271, 2010.

[170] Thorsten O Zander, Kunal Shetty, Romy Lorenz, Daniel R Leff, Laurens R Krol, Ara W Darzi, Klaus Gramann, and Guang-Zhong Yang. Automated task load detection with electroencephalography: towards passive brain–computer interfacing in robotic surgery. *Journal of Medical Robotics Research*, 2(01):1750003, 2017.

[171] Chong Zhang, Chong-Xun Zheng, and Xiao-Lin Yu. Automatic recognition of cognitive fatigue from physiological indices by using wavelet packet transform and kernel learning algorithms. *Exp. Sys. App.*, 36(3):4664–4671, 2009.

[172] Tao Zhao, Wenyi Zhao, Brian D Hoffman, William C Nowlin, and Hua Hui. Efficient vision and kinematic data fusion for robotic surgical instruments and other applications, March 3 2015. US Patent 8,971,597.

[173] Fangxun Zhong, Zerui Wang, Wei Chen, Kejing He, Yaqing Wang, and Yun-Hui Liu. Hand-eye calibration of surgical instrument for robotic surgery using interactive manipulation. *IEEE Robotics and Automation Letters*, 5(2):1540–1547, 2020.

[174] Tian Zhou, Jackie S Cha, Glebys Gonzalez, Juan P Wachs, Chandru P Sundaram, and Denny Yu. Multimodal physiological signals for workload prediction in robot-assisted surgery. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(2):1–26, 2020.

[175] Tian Zhou, Jackie S Cha, Glebys T Gonzalez, Chandru P Sundaram, Juan P Wachs, and Denny Yu. Jisap: Joint inference for surgeon attributes prediction during robot-assisted surgery. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2246–2251. IEEE, 2019.

[176] Tian Zhou, Jackie S Cha, Glebys T Gonzalez, Juan P Wachs, Chandru Sundaram, and Denny Yu. Joint surgeon attributes estimation in robot-assisted surgery. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 285–286, 2018.

[177] J Zhu, L Ji, and C Liu. Heart rate variability monitoring for emotion and disorders of emotion. *Physiol Meas*, 40(6):064004, 2019. [PubMed] [CrossRef] [Google Scholar].

[178] Johanna M Zumer, René Scheeringa, Jan-Mathijs Schoffelen, David G Norris, and Ole Jensen. Occipital alpha activity during stimulus processing gates the information flow to object-selective cortex. *PLoS biology*, 12(10):e1001965, 2014.