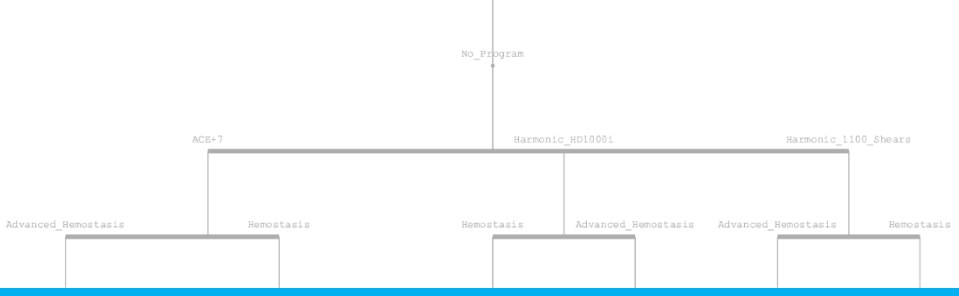


Department of Precision and Microsystems Engineering

Real Time State Estimation of Surgical Energy Devices for Use in an Integrated Operating Room

Pepijn van Esch

Report no	:	2022.016
Coach	:	AP. Andres Hunt; Prof. I. Sakuma; Prof. E. Kobayashi
Professor	:	Prof. Jenny Dankelman
Specialisation	:	BME & HTE
Type of report	:	Thesis
Date	:	24 May 2022

TU Delft Delft University of Technology

Challenge the future

Real Time State Estimation of Surgical Energy Devices for Use in an Integrated Operating Room

by

P. van Esch

To obtain the degree of Master of Science at the Delft University of Technology, to be defended publicly on Tuesday May 24, 2022 at 09:00 AM

Student number	4443136	
Project duration:	September 1, 2021 - May 24, 2022	
Thesis committee:	Prof. I. Sakuma	The University of Tokyo, Daily Supervisor
	Prof. E. Kobayashi	The University of Tokyo, Daily Supervisor
	AP. A. Hunt	TU Delft, Chair-HTE
	Prof. J. Dankelman	TU Delft, Chair-BME
	AP. S. Iskander-Rizk	TU Delft

This thesis is confidential and cannot be made public until May 2024.
An electronic version of this thesis is available at http://repository.tudelft.nl/

Contents

Abstract

Integrated operating rooms (OR) have shown to be promising in meeting the challenges created by the increased complexity in the OR, by improving the quality of care, simplifying the clinical workflows, and reducing equipment-related incidents and surgical errors. Although several systems that integrate medical devices have been developed, surgical energy devices have yet to be integrated into these systems. Current measurement systems of surgical energy devices are not yet suitable for this integration, due to a lack of being manufacturer-independent and being limited in the amount of measured information. Therefore, the main objective of this thesis is the development of a manufacturer-independent measuring system that can estimate and communicate the state of any surgical energy device in real time. To realise this goal, a measurement system has been developed that measures the state of any surgical energy device through image recognition. The system is integrated into the OPeLiNK system of the Smart Cyber Operating Theatre, SCOT. It can however, be integrated in any of the other integrated ORs. The system has been tested through several experiments, showing that the accuracy and measuring speed meet the requirements of 90% and 0.2s. The system outperforms the state of the art in the amount of data that is being measured in real time at the cost of a small delay of 0.07s. Currently the bottleneck is the communication through the OPeLiNK system, where the communication frequency has a set limit of 1Hz. It can be concluded that the developed measurement system can be used for the integration of surgical energy devices in the integrated operating room. The next step is to further improve the system in terms of robustness, easier use, and increased measurement capabilities.

1 Introduction

1.1 Medical device interoperability and OR integration

The amount of complex technical systems that are used in the operating room (OR) has been increasing for many years. Currently the work of interdisciplinary teams consists of information ranging from many different sources, including highly specialized medical devices from different brands [1–4]. Examples of these devices are endoscopes, intraoperative neuromonitoring (IONM) systems, anesthesia systems, surgical microscopes, MRI scanners, and complex surgical robotic systems such as the da Vinci System [5]. This increased complexity, comes with surgical errors and equipment-related incidents having become a large percentage of the errors occurring in the OR [6–8]. The communication between these systems is limited, resulting in the increase of the overall complexity of the OR [9]. There is a growing need for integrated medical systems in a holistic clinical infrastructure. Integrated ORs have shown to be promising in meeting the challenges created by this increased complexity, by improving the quality of care, simplifying the clinical workflows, and reducing equipment-related incidents and surgical errors [1, 10]. However, the absence of manufacturer-independent interoperability frequently prevents the development and use of these integrated assistive systems. Most medical devices that are used within the OR are not yet able to communicate with each other. This results in the isolation of information, and therefore prevents access to time-synchronized qualitative data for the development and implementation of not only the integrated OR, but also the further development of future semi-automated surgery. Currently, the state-of-the-art in integrated ORs consist of the research projects OR.NET, OpenICE, and SCOT that have been working on the interoperability of medical devices and the integration into the OR [11].

1.1.1 OR.NET

The OR.NET project was first introduced in 2012 in Germany, with the focus to extend and evaluate the safe dynamic component-interactions in the OR [12]. The most important goal of the interaction between devices is to realize this without a central system as middleware through standardization of the communication and data model. During this project a new standard was introduced, named the IEEE 11073 Service-oriented Device Connectivity (SDC) [13]. This standard is based on the concept of Service-Oriented Medical Device Architecture (SOMDA), and aims at the automatic recognition of device services and applications that are connected to the network. For a medical device to be integrated, it must meet the standardized specification requirements that are set in the IEEE 11073 medical device communication standard. This means that it can communicate through Ethernet according to the interface description. When they comply to the IEEE 11073 SDC standard, the data of the medical devices are presented according to a domain information model, that distinguishes the medical device description and the state of the system in a medical information database. Each medical device has its own unique device identification, where information can be sent or requested from the device. Through this, the interaction between medical devices, and the interaction between the medical devices and the clinical IT-infrastructure, can be realized.

1.1.2 OpenICE

OPenICE is an open-source software project introduced by the Medical Device Plug-and-Play Interoperability Program (MD PnP) at Massachusetts General Hospital [14]. This project is an implementation of the ASTM F2761 Integrated Clinical Environment (ICE) standard [15]. This standard defines an architecture for building a safe patient-centric Integrated Clinical Environment. The platform contains software device adapters for medical devices and standard middleware for the OMG Data-Distribution Service for Real-Time Systems (DDS). These equipment interfaces mainly consist of small single-board computers that can be connected and attached to the back of a medical device. When the connected medical device announces updates to the data, it acts as a so-called 'publisher'. When requesting data, it acts as a 'subscriber'. The DDS middleware matches the publisher to the subscriber, resulting in the integration of their data. This allows the data from apps to become indistinguishable from data from physical medical devices. Furthermore, it enables the development and use of processing apps that could generate data for use by other system components. For a participant to be integrated into the system, it must comply to a subset of the ISO/IEEE 11073-10101 nomenclature, meaning that the components are semantically interoperable. For this, the medical device manufacturer must produce a device with both a suitable electronic interface and the required data elements, that will work with any ICE application.

1.1.3 SCOT

SCOT is a Smart Cyber Operating theatre that was first introduced by Tokyo Women's Medical University Hospital in Japan. This integrated OR gathers information of any surgical device that is connected to the OR. Various information can be stored and displayed such as diagnostic images, and the status of several surgical devices in the OPeLiNK integrated communication interface. An example of this can be seen in Figure 1.

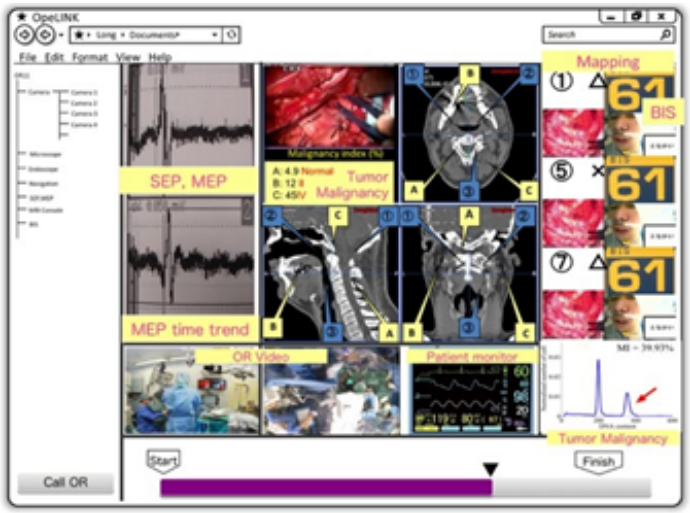


Figure 1: Picture of the decision-making navigation screen of the OPeLiNK system [16].

SCOT has three main functions:

1. By logging of time-synchronized data of any surgical equipment in the OR, postsurgical error evaluation can be performed.
2. It allows for telementoring by communicating real time data through the decision-making navigation screen [17]. Furthermore, by only communicating the data relevant for the phase of the surgery, the information overload is reduced [18].
3. With the generated data, future semi-automated surgery could be a possibility [19].

The OPeLiNK interface is based on Orin, which allows for the integration of applications and devices. Examples of systems that are supported by Orin are systems that use the HL7 and DICOM standards. Other examples of supported systems are surgical navigation and intraoperative monitoring systems. The system uses the Simple Object Access Protocol (SOAP) for standard Ethernet connection between the server and the devices. A Controller Access Object (CAO) engine is used as an interface for the connection of the devices. This interface executes all device independent tasks and accesses devices for dependent operations.

While all three system are quite similar, their focus is slightly different. The focus of the OR.NET project is on integrating medical devices in the operating room. OpenICE focuses on a more clinic-wide interoperability. SCOT not only focuses on the integration of medical devices, but also on the integration of imaging modalities, like MRI, and surgical robotic modalities. All three systems have the ability to integrate several medical devices. Surgical energy devices, however, have yet to be integrated. This is because most energy devices do not follow the required protocol for integration. Furthermore, their software is often not accessible and protected by the manufacturer.

1.2 Surgical Energy device integration

1.2.1 Energy devices

Surgical energy devices are devices that are used to seal blood vessels, cut tissue and stop bleeding. Within the OR, several types of energy devices are used. At Japan's National Cancer center, for example, 7 types of energy devices are used. These consist of the Conmed System 2450, the Valleylab ForceTriad (Ligasure), the Valleyab FT10, the ERBE VIO 300D, the ERBE VIO3, the OLYMPUS Thunderbeat, and the ETHICON Harmonic & EnSeal. These devices support electrocautery, electrosurgery, and ultrasonic scalpels. Electrocautery is the process where heat is generated by passing direct or alternating current through an electrode. This electrode is then used to destruct tissue or to achieve hemostasis. In electrocautery, the current does not pass through the patient. Electrosurgery is the process were electricity is used for the thermal destruction of tissue through dehydration, coagulation, or vaporization. Within electrosurgery, the current does pass through the patient. In electrocautery, the tissue is always cauterized, while within electrosurgery the tissue can also be cut without searing it. The ultrasonic scalpel uses vibration to create frictional heat in contact with tissue. This allows for simultaneous cutting and cauterization of the tissue.

1.2.2 State-of-the-art in the measuring of Energy devices

To determine the state-of-the-art in the recognition of the states of energy devices, a literature survey was performed. In this literature survey the relevant articles were found through the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model [20]. SCOPUS was the search engine used. Various keywords were used for the search on the measurement of the energy devices. A few of these keywords were the terms Energy devices, electrosurgery, electrocautery, and harmonic scalpel, together with keywords indicating the measurement of different types of information on these systems. Synonyms, different spellings, and different suffixes were accounted for using operators. The exclusion criteria consist of that the article must be published within the last 20 years ranging from 2000 till 2022.

The PRISMA Diagram in Figure 2 shows the process of selecting the literature to be included. In total 281 articles were identified through he SCOPUS data base. From these, zero duplicates have been removed. 61 articles were excluded based on their title or abstract. Finally, 216 full-text articles were excluded based on their content, resulting in four articles to be included in the review. From these four articles two are from the same author, using the same measuring device [21, 22]. Therefore, we conclude that the state-of-the-art consists of three different approaches in measuring the state of the energy device.

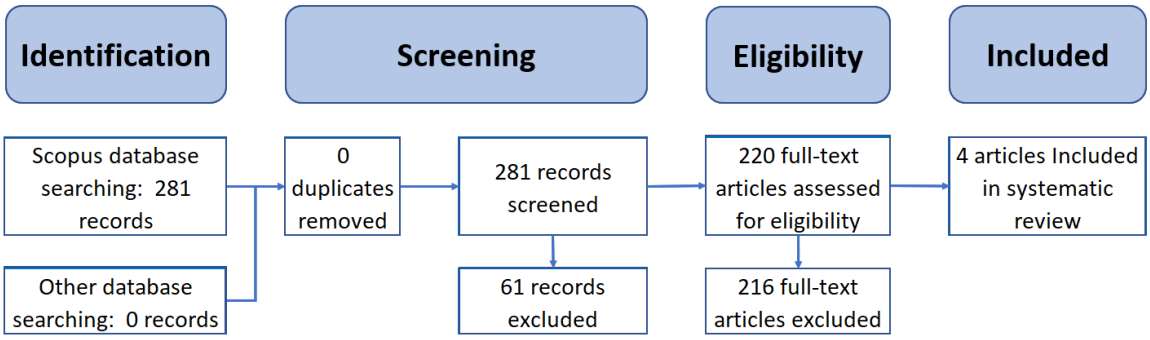


Figure 2: PRISMA flow diagram describing the article selection process.

The first approach consists of measuring the activations of a electrocautery system by Yuki Ushimaru et al [23]. In this research a counter electrode plate cable detector coil was attached to the electrocautery generator. Based on the high-frequency output of the electrocautery, the activation rates and activation time could be measured. This data is then captured and used for postoperative evaluation by visualising the instrument usage in specific procedures.

The second approach consists of a measurement system with a current sensor connected in between the power plug of a electrosurgical device and socket. This was proposed by Annetje C. P. Guédon et al [21]. In this approach the electric current delivered to the device is measured, and when this current exceeds a certain threshold, the activation of the electrosurgical tool is measured. The measuring frequency of the system was 8Hz. This data was evaluated postoperatively for the quantification of the difference between a skilled surgeon and a resident. Furthermore this approach was also applied in real time in a different research, by applying the collected data for the real time estimation of the surgical procedure duration [22].

The final state-of-the-art approach was proposed by Jose Dums, Bertoldo Schneider, and Alceu Badin [24]. In this approach an electronic structure was developed to estimate the output electrical signals of a electrosurgical device. This structure consisted of common resistors and high frequency distortion compensation circuits that are used to measure both the output current and the output voltage. From this, the output power is calculated. The average measurement error of this system is a maximum of 3.34%. The future goal of this research is applying these measurements in a control system for the real-time power adjustment for burn-prevention.

1.2.3 The gap in the state-of-the-art

To summarize, we see three somewhat similar approaches that measure the current (and voltage) of either an electrocautery or an electrosurgical device. In the first two researches the activation rate and duration are estimated. In the third approach the output power of the electrosurgical unit is estimated. The gap in this research is that although the activations and power are measured, the other settings of the energy device are not measured. There is for example no information on the selected power level, instruments, or type of modes being used. This information is mainly communicated visually through the display of the energy device. Furthermore, these systems have been designed for the measurement of either electrocautery or electrosurgical devices. In the OR however, multiple types of energy devices are used. These systems are for example not designed for the measurement of a ultrasonic scalpel.

1.3 Aim of this thesis

Integrated ORs have shown to be promising in meeting the challenges created by the increased complexity in the OR, by improving the quality of care, simplifying the clinical workflows, and reducing equipment-related incidents and surgical errors. Although several systems that integrate medical devices have been developed, surgical energy devices have yet to be integrated into these systems. Current measurement systems of surgical energy devices are not suitable for this integration, due to a lack of being manufacturer-independent and being limited in the amount of measured information. Therefore, the aim of this thesis is the development of a manufacturer-independent measuring system that can estimate and communicate the state of any surgical energy device in real time.

1.4 Approach and outline thesis

To realise this goal, a measurement system has been developed that measures the state of any surgical energy device through image recognition. The reason for this, is that most information is communicated visually by the energy devices. The system is integrated into the OPeLiNK system of the Smart Cyber Operating Theatre, SCOT. It can however, be integrated in any of the other integrated ORs.

This thesis discusses the development of this measuring system in three parts. First a literature review was performed where an overview of the state-of-the-art in different image recognition methods for different types of indicators is given. Furthermore, an overview of noise-handling methods is given. This part can be found in Appendix A. In the other two parts the developed measurement system is discussed. These can be found in section 2 and section 3. Below, a short description of the three parts is given.

An overview of Image Recognition Methods and Noise Handling Methods for Different Types of Indicators To be able to design a measurement system that can measure the state of a surgical energy device through image recognition, the system must be able to recognize the different type of visual indicators that the surgical energy device uses to communicate its states. These states can be communicated through light indicators, characters, and symbols. To develop this measuring system, it is vital to know the state-of-the-arts in the real time recognition of these indicators. Therefore, this literature review gives an overview on these methods and evaluates them based on their speed and accuracy. Furthermore, since the operating room is an environment with lots of noise, this review presents an overview of noise handling methods that deal with reflections, shadows, the movement of the camera, and occlusions. The results can be applied in the development of the measuring system.

Measuring System for the Data Collection of an Energy Device In the next section, the development of a measurement system for the data collection of one energy device is discussed. This system uses image recognition in the form of template matching to determine the state of the energy device. The system consists of a camera connected to the OPeLiNK system, that gathers visual data from the display of the energy device. The OPeLiNK system creates multiple video files, that are given as input to the image recognition, for postsurgical evaluation. The energy device communicates its states through light indicators and character indicators. The developed measuring system locates these indicators and then uses a newly developed light indicator recognition method and a standard character recognition method to determine the states of the energy device. The system is tested on its limits in terms of camera placement, its delay between the acquired data and endoscopic images, and its recognition accuracy.

Manufacturer Independent Measuring System for the Real Time State Communication of any Energy Device In this final section, the previously developed measurement system is expanded to a system that can measure any energy device in real time. For this, a communication protocol is developed that describes the state of any energy device. Based on this communication protocol a reading strategy is developed that increases the accuracy of the system, and reduces the computation time. A setup program is developed to register any energy device in the preoperative stage. A main program is developed to measure the state of any energy device in real time during the intraoperative stage.The measuring system is integrated into the OPeLiNK system. Several cameras are used to gather visual data from the displays of several energy devices. This visual data is then given as input to the measuring system. The measurement system determines the states of the energy devices and communicates this to the surgeon and the remote experts through the decision-making navigation screen of the OPeLiNK system. Several experiments are performed to evaluate the system on its accuracy, measurement speed and communication speed. Furthermore the system is validated through testing in comparison to the state-of-the-arts.

Finally, the work that is presented in this thesis is discussed in section 4. The methods used for the recognition of the states of energy devices are considered, The robustness and speed of the system is discussed, and the future development and use of the system is highlighted.

References

[1] Heinz U. Lemke and Michael W. Vannier. "The operating room and the need for an IT infrastructure and standards". In: 1.3 (Nov. 2006), pp. 117–121. DOI: 10.1007/s11548-006-0051-7.

[2] Nicolas Padoy et al. "Statistical modeling and recognition of surgical workflow". In: *Magazine* 16.3 (Apr. 2012), pp. 632–641. DOI: https://doi.org/10.1016/j.media.2010.10.001.

[3] Armin Janss et al. "Development of Medical Device UI-Profiles for Reliable and Safe Human-Machine-Interaction in the Integrated Operating Room of the Future". In: *Advances in Human Aspects of Healthcare* 3 (Jan. 2014), p. 274.

[4] Martin Kasparick et al. "New IEEE 11073 standards for interoperable, networked point-of-care Medical Devices". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2015, pp. 1721–1724. DOI: 10.1109/EMBC.2015.7318709.

[5] Simon DiMaio, Mike Hanuschik, and Usha Kreaden. "The da Vinci Surgical System". In: ed. by Jacob Rosen, Blake Hannaford, and Richard M. Satava. Boston, MA: Springer US, 2011, pp. 199–217. DOI: 10.1007/978-1-4419-1126-1_9.

[6] Annetje C. P. Guédon et al. "Safety status system for operating room devices". In: 22 (2014), pp. 795–803. DOI: 10.3233/THC-140854.

[7] I. Wubben et al. "Equipment-related incidents in the operating room: an analysis of occurrence, underlying causes and consequences for the clinical process". In: *Magazine* 19.6 (2010), e64.

[8] Ruwan A. Weerakkody et al. "Surgical technology and operating-room safety failures: a systematic review of quantitative studies". In: *Magazine* 22.9 (2013), p. 710. DOI: 10.1136/bmjqs-2012-001778.

[9] Kathy Lesh et al. "Medical Device Interoperability-Assessing the Environment". In: *2007 Joint Workshop on High Confidence Medical Devices, Software, and Systems and Medical Device Plug-and-Play Interoperability (HCMDSS-MDPnP 2007)*. 2007, pp. 3–12. DOI: 10.1109/HCMDSS-MDPnP.2007.22.

[10] Kevin Cleary, Audrey Kinsella, and Seong K. Mun. "OR 2020 workshop report: Operating room of the future". In: *Magazine* 1281 (May 2005), pp. 832–838. ISSN: 0531-5131. DOI: https://doi.org/10.1016/j.ics.2005.03.279.

[11] Xiao Sun et al. "Robotic Technology in Operating Rooms: a Review". In: 2.3 (Sept. 2021), pp. 333–341. DOI: 10.1007/s43154-021-00055-4.

[12] M. Rockstroh et al. "OR.NET: multi-perspective qualitative evaluation of an integrated operating room based on IEEE 11073 SDC". In: 12.8 (Aug. 2017), pp. 1461–1469. DOI: 10.1007/s11548-017-1589-2.

[13] Okan Yilmaz et al. *Development and Evaluation of a Platform-Independent Surgical Workstation for an Open Networked Operating Theatre Using the IEEE 11073 SDC Communication Standard*. Springer International Publishing, 2020, pp. 79–92. ISBN: 978-3-030-49904-4.

[14] David Arney, Jeffrey Plourde, and Julian M. Goldman. "OpenICE medical device interoperability platform overview and requirement analysis". In: *Magazine* 63.1 (2018), pp. 39–47. DOI: doi:10.1515/bmt-2017-0040.

[15] ASTM F2761-09. *Medical Devices and Medical Systems – Essential safety requirements for equipment comprising the patient-centric integrated clinical environment (ICE) – Part 1: General requirements and conceptual model*. 2013. URL: http://www.astm.org/Standards/F2761.htm..

[16] Jun Okamoto et al. "Development concepts of a Smart Cyber Operating Theater (SCOT) using ORiN technology". In: *Magazine* 63.1 Number (2018), pp. 31–37. DOI: doi:10.1515/bmt-2017-0006. URL: https://doi.org/10.1515/bmt-2017-0006.

[17] Toshihiro Ogiwara et al. "Endoscopic Endonasal Approach in the Smart Cyber Operating Theater (SCOT): Preliminary Clinical Application". In: *Magazine* 147 (Mar. 2021), e533–e537. DOI: 10.1016/j.wneu.2020.12.114.

[18] N. Bitterman. "Technologies and solutions for data display in the operating room". In: 20.3 (June 2006), pp. 165–73. DOI: 10.1007/s10877-006-9017-0.

[19] Yoshihiro Muragaki et al. "Smart Cyber Operating Theater (SCOT): Strategy for Future OR". In: ed. by Makoto Hashizume. Singapore: Springer Singapore, 2022, pp. 389–393. ISBN: 978-981-16-4325-5. DOI: 10.1007/978-981-16-4325-5_53.

[20] Padhraig S. Fleming, Despina Koletsi, and Nikolaos Pandis. "Blinded by PRISMA: Are Systematic Reviewers Focusing on PRISMA and Ignoring Other Guidelines?" In: 9.5 (2014), e96407. DOI: 10.1371/journal.pone.0096407.

[21] F. C. Meeuwsen et al. "The Art of Electrosurgery: Trainees and Experts". In: 24.4 (Aug. 2017), pp. 373–378. DOI: 10.1177/1553350617705207.

[22] Annetje C. P. Guédon et al. "'It is Time to Prepare the Next patient' Real-Time Prediction of Procedure Duration in Laparoscopic Cholecystectomies". In: *Magazine* 40.12 (2016), pp. 271–271. DOI: 10.1007/s10916-016-0631-1.

[23] Yuki Ushimaru et al. "Innovation in surgery/operating room driven by Internet of Things on medical devices". In: 33.10 (Oct. 2019), pp. 3469–3477. ISSN: 1432-2218. DOI: 10.1007/s00464-018-06651-4.

[24] Jose Dums, Bertoldo Schneider, and Alceu Badin. "Low cost system to measure active power in electrosurgical units". In: *Magazine* 33 (Nov. 2017). DOI: 10.1590/2446-4740.03217.

2 Measuring System for the Data Collection of an Energy Device

In this paper a novel measuring system is developed for the postoperative state estimation of one surgical energy device. This can be used for postsurgery evaluation by integrating the resulting information with endoscopic images. The further contributions discussed in this paper are a new image recognition approach in combining several existing and new image recognition methods for the recognition of different types of indicators used by the energy device to communicate its states. This approach consists of a novel indicator localization approach, a new light indicator recognition method, and an existing character recognition method.

Development of a measuring system for the data collection of a surgical energy device

Pepijn van Esch[1,2,3*]

[1*]Biomedical engineering, Delft University of Technology, Mekelweg 2, Delft, 2628 CD, Netherlands.
[2*]Mechanical engineering, Delft University of Technology, Mekelweg 2, Delft, 2628 CD, Netherlands.
[3*]The Graduate School of Engineering, The University of Tokyo, Bunkyo-ku 7-3-1, Tokyo, 113-8656, Japan.

Corresponding author(s). E-mail(s): p.vanesch@student.tudelft.nl;

Abstract

Purpose: Research about intraoperative images for surgical analysis is currently popular. It is expected that the analysis will become more accurate by adding the usage status of surgical energy devices. This paper proposes a new method for the state recognition of an energy device. The measurement system will be integrated into the OPeLiNK system of the Smart Cyber Operating Theatre, SCOT, but could also be implemented in other integrated operating rooms.
Methods: The developed system uses image recognition in the form of template matching to estimate the state of the device. A new approach based on predetermined locations with respect to a reference image is proposed for indicator localization. To align the video image with the reference image, this approach uses the existing methods of SURF feature detection, RANSAC, homography transformation, and an Enhanced Correlation Coefficient optimization method. Furthermore, a novel approach to evaluate the accuracy of the transformation is introduced. For light indicator recognition, a new method based on the luminance level with respect to template images with threshold estimation is proposed. For character recognition, an existing Normalized Cross Correlation method is applied.
Results: The system has been tested on the Conmed System 2450. The limit of the indicator localization method is a maximum camera angle of **35°** with respect to the normal of the display. The delay between the video data of the endoscope and the measurement system, has

a median of 0.077s. 2963 videos of nine colon surgeries performed at Japan's National Cancer Center were evaluated. The results show a 98.2% accuracy.
Conclusion: The results show a high accuracy and a low delay that fits within our requirements. This indicates that our system can be integrated into the OPeLiNK system for postsurgery error evaluation. The system outperforms the state-of-the-art in measuring more significant data on the use of energy devices.

Keywords: OPeLiNK, Conmed System 2450, image recognition, template matching

1 Introduction

Research on surgical analysis using intraoperative images is popular nowadays. It is expected that more accurate analysis will be possible by adding the usage status of surgical energy devices. There are several systems that integrate data of multiple surgical devices within the operating room. These consist of OR.NET, OpenICE, and SCOT. The OR.NET project is an interoperability concept that allows for a dynamic manufacturer-independent integration of point-of-care medical devices in the operating room [1]. OpenICE is an open source implementation of the ASTM F2761 Integrated Clinical Environment (ICE) standard [2]. This platform included software device adapters for medical devices and standard middle ware for the OMG Data-Distribution Service for Real-Time Systems (DDS). Our developed method has been applied within SCOT, but could be used in any of these other systems.

SCOT is a Smart Cyber Operating Theatre that integrates data of multiple surgical devices in the surgical room [3]. This operating theatre measures as much information as possible of any surgical device in the operating room. It stores and displays time-synchronized patient condition data, diagnostic images, and the status of surgical devices in the OPeLiNK integrated communication interface as can be seen in Figure 1.

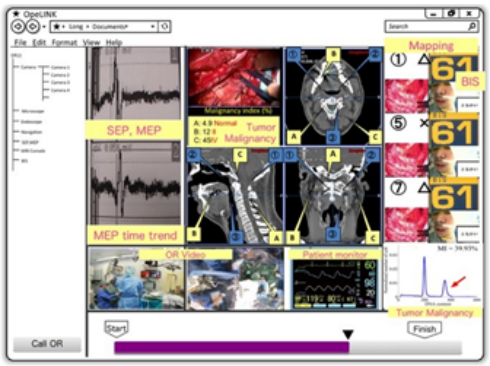


Fig. 1: Picture of the decision making navigation screen of the OPeLiNK system [3].

SCOT has three main functions:

1. First, it allows for the logging of time synchronized data of any surgical equipment in the operating room for error evaluation.
2. Second, only the data relevant for the phase of the surgery can be communicated through the decision-making navigation screen. This reduces the information overload [4] and allows for telementoring, where the surgeon can be assisted by a remote expert during surgery [5].
3. Finally, the generated data could allow for future semi-automated surgery [6].

Data that still must be integrated is that of the surgical energy devices. To acquire the status of the energy device, one would like to retrieve this directly through the software. The software is however most of the time protected by the manufacturer. Access to this software can either be bought for significant sums of money or is not available. To circumvent this problem a different method needs to be used to extract the status of these devices. When looking at previous research on measuring the state of an energy device, there are three main researches that have been performed [7–9]. All three measure the current and voltage, from which then the activation durations and number of activations are estimated. On top of measuring the current and voltage, one research also estimates the output power. However, the problem is that it only measures the activations and not the other settings of the machine.

The other settings of the machine are communicated with the surgeon through visual and auditory indicators. These settings give important information on for example the modes, and types of instruments that are used during the surgery that largely affect the interaction between tissue and tool. This information is vital for more accurate and elaborate postsurgical evaluation. Since most information is communicated through visual indicators, this study proposes a new method for the acquisition of the states of the energy device through image recognition.

The aim of this research is to develop an image recognition system that can acquire the usage status of a surgical energy device and integrate it with endoscopic images. By time synchronizing and logging the data of this system in combination with other surgical equipment in the operating room, complications that might have arisen during surgery can be evaluated better by making it more reliable and objective.

The focus of this study is the time synchronization and logging of the status of one specific energy device, the Conmed System 2450, from here on referred to as the Conmed [10]. An image of the Conmed can be seen in Figure 2. This is an electrosurgical tool with the capability to cut, coagulate and use bipolar. The developed system will be integrated into the OPeLiNK system of SCOT, but could be used in any other integrated operating theatre. This research will focus on using the measurement system for postsurgical analysis. In the future however, its real time communication capabilities for telementoring will be investigated.

Fig. 2: Picture of the Conmed System 2450.

This paper proposes a new method for the recognition of the states of a surgical energy device through image recognition. The system uses a camera that is connected to the OPeLiNK system and gathers visual data of the display of the energy device during surgery. The OPeLiNK system creates several video files which then postoperatively are evaluated by using image recognition. The recognized states can then be integrated with the endoscopic images.
The contributions can be summarized as follows:

- A novel measuring system has been developed for the measurement of the states of a surgical energy device.
- A new approach in combining several types of existing and new image recognition methods has been proposed consisting of:
  - A novel light indicator recognition method based on the intensity with respect to template images, with threshold estimation based on the environment.
  - An existing method for character recognition consisting of template matching based on Normalized Cross Correlation (NCC).
  - A novel indicator localization approach based on a reference image with predetermined locations with respect to a reference frame is proposed. A perspective transformation function is applied to transform the video image such that it aligns with the reference frame. This approach combines the existing methods of, SURF feature detection, RANSAC, homography transformation, and an Enhanced Correlation Coefficient (ECC) optimization method. Then a new approach to the evaluation of the accuracy of the transformation is introduced. If the transformation is deemed inaccurate, the process of transformation estimation is restarted with different initial conditions.

2 Method

To develop the measuring system first a system analysis is performed to identify what needs to be detected of the Conmed. Based on this analysis, the types of indicators that need to be recognized are determined. Afterwards, an environmental analysis is performed to determine the requirements for the

measuring system. This is then used to develop the measurement system and decide what image recognition methods and noise handling methods need to be applied.

System requirements

It was determined that the desired accuracy of the system is at least 90% to be considered for integration into the OPeLiNK system. This requirement is based on the desire for future real time applications. The system is required to outperform a human, which has a 90% recognition accuracy with a 0.2s reading time [11]. In this paper, the accuracy is the only requirement for the image recognition method. The recognition speed will be considered in future research. To make sure that the real time recognition is a possibility in the future, the delay between the measured data and the acquired endoscopic data may not exceed the future requirement of 0.2s reading time. If this requirement would not be met, it would not matter how fast the future real time recognition is, since the input data for the image recognition system would always have a minimum delay that exceeds the requirement.

The measuring system needs to be designed such that it does not hinder the surgeons during the surgery. This means that no parts should stick out too far from the Conmed. Furthermore a minimal clearance of 15cm from the display is desired, such that interaction with the display is not hindered. Finally, no permanent changes must be made to the Conmed. This means that any marker or other object attached to the Conmed must be attached and removed for every surgery.

2.1 Reading strategy: possible states of The Conmed system

To be able to read the states of the Conmed, first an analysis was done on the states it can communicate and how they are communicated [10]. The states can be divided into the state names 'Machine on/off', 'Instrument', 'Activation on/off', 'Major mode', 'Minor mode', 'Power level', and 'Error'. All the states combined will influence the resulting interaction between the tissue and the electrosurgical tool. Not only the result is influenced, but also the availability of each state is dependent on other states. This dependence can be described by the state tree shown in Figure 3.

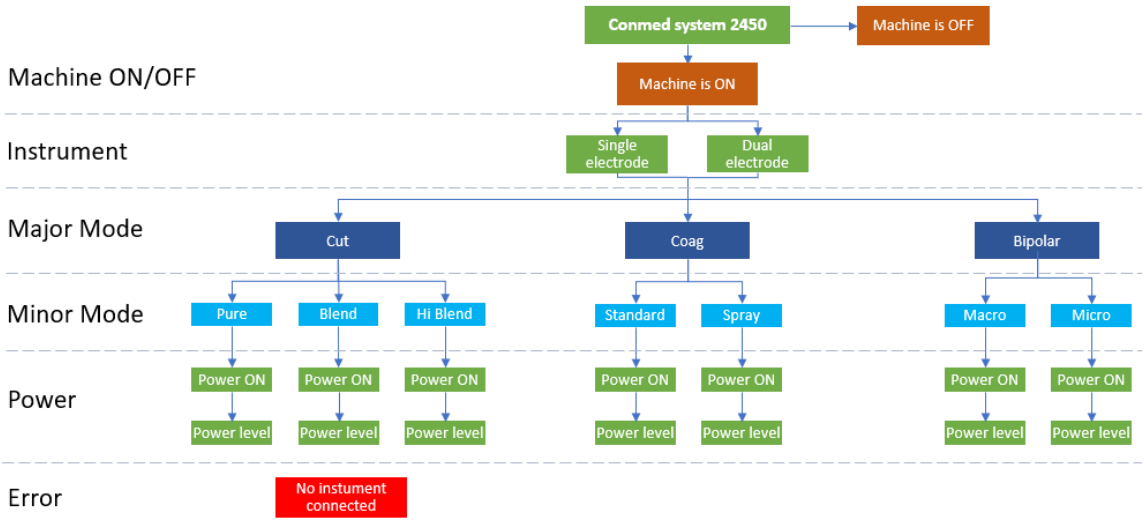


Fig. 3: State tree describing all possible states of the Conmed system 2450.

Machine on/off

The Conmed does not show if it is turned on through a specific indicator. This state can however be determined by the indicator lights of the Instruments, Minor modes, and Power levels that will be turned on when the machine is turned on.

Instrument

The Conmed has two types of dispersive electrodes that can be connected. These consist of a single dispersive electrode and a dual dispersive electrode. Each dispersive electrode has its own light indicator. When the machine is turned on, always one of these indicators is shown.

Major mode

The Major mode is the main function of the energy device. The Conmed has three types of Major modes, consisting of Cut, Coagulation, and Bipolar. Each Major mode has its own light indicator and sound indicator. The Cut mode is described by a yellow indicator, while the Coagulation and Bipolar mode are described by a blue indicator. In terms of sound, the Cut mode has a higher tone, and Coagulation and Bipolar mode have a lower tone.

Minor mode

The Minor mode describes the settings of the Major mode. Each Major mode has certain Minor modes that can be chosen. For the Conmed, the Minor modes that are available for each Major mode can be seen in the Minor mode layer in Figure 3.

Power: Activation on/off & Power Level

The Conmed system communicates in two ways if the power is activated or not, and thus if cutting, coagulating, etc. is being performed. The activation is communicated with the same indicator as the Major mode. The first communication method is through three light indicators, where each indicator represents a different Major mode. The second method is through sound indicators, where for each Major mode a different tune is used.

The Power level, describes the power that is going to be used to perform a cut, coagulate or any other Major mode that is being performed. For the recognition of the Power level, this is divided into three character indicators. These indicators consist of digits, tens, and hundreds. Each of these indicators can have a value of 0 up to 9. The states of all 3 indicators combined determine the state of the Power level. If no candidate state is found for either the parts of the hundreds or the tens, the indicator value must be below that. The digit indicator is always visible. Depending on the Major mode the availability of each indicator can vary from digits to hundreds. For the Conmed, the Cut and Coagulation mode can go up to hundreds, while for the Bipolar mode it is only possible to go up to tens.

Error

When the machine has no equipment attached in the form of dispersive electrodes, or electrosurgical tools, the machine will indicate that there is an error. This is done by coloring the indicator lights of the dispersive electrodes in red and alternately turning them on and off.

The types of indicators that the Conmed uses to communicate the state can be summarized into the visual indicators consisting of light indicators and character indicators, and auditory indicators. For this research, the focus is only on recognizing the visual indicators.

2.2 Camera setup

The measuring system needs to operate in a surgical environment. This means that the system needs to deal with the following different types of noise:

- *Reflections*: In the operating room there are several operating lights that are used. Due to this, reflections on the display of the machine can appear when these lights are moved in a certain position.
- *Shadows*: Sometimes the surgeons or assistants will walk past the machine which can cast shadows on the display.
- *Camera movement*: In general, the camera is assumed stationary. Sometimes however, the machine needs to be moved or the surgeon or assistant accidentally touches the camera. This can result in that the camera position and orientation is changed with respect to the display of the machine.

- *Obstructions*: When an assistant is changing the settings of the machine or attaching a different instrument, it can be that the camera view is temporarily obstructed.

The measuring system is not allowed to be permanently attached to the Conmed but must be attached and removed during each surgery. During the preparation there is a small time-window where the measurement system needs to be setup. The system must be made ready just after the anesthesia has been applied and before the surgery starts.

2.3 Measuring system

Figure 4 shows the layout of the measuring setup. The proposed measuring system consists of a camera mounted to the Conmed. During a surgery, the camera collects visual data from the display of the Conmed with 30 frames per second. This camera is connected to the OPeLiNK system. The OPeLiNK system collects video data from the camera, and time synchronizes it with the data from the other surgical equipment in the operating room. During a surgery the OPeLiNK system stores all data in a database, which can be analyzed after the surgery. In terms of the video data, these are divided into multiple video files, each consisting of 930 frames, or approximately of 31 seconds duration. A folder consisting of video files of one surgery is given as input to the measuring system, which then evaluates the states of the Conmed system, for each frame. These states would consist of the time in the video, the Instrument, the Major mode, the Minor mode, the Power level, and Errors if they occurred. The measuring system consists of a laptop with a 9th generation Intel Core i7 9750H /2.6GHz/6 core CPU, 16GB RAM, and a GeForce GTX 1660Ti graphics card.

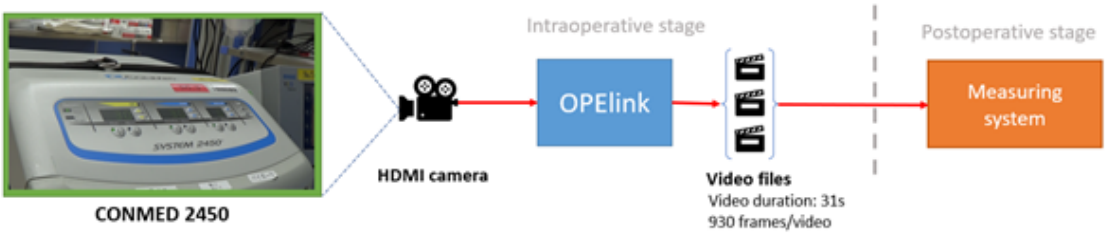


Fig. 4: Layout of the measuring setup consisting of the Conmed system 2450, a HDMI camera, the OPeLiNK system, and the measuring system. In the intraoperative stage, the OPeLiNK system collects visual data through the HDMI camera. This data is then stored in the data base in the form of video files, each with 930 frames. In the postoperative stage, these video files are evaluated to determine the state of the Conmed system 240, for each frame.

2.3.1 Image recognition program design

The image recognition program consists of an indicator localization method and an indicator state estimation method. For this system, both of these methods consist of non-machine learning methods. The reason is that machine learning methods require large datasets to reach sufficient accuracy. For this application there are no such datasets available. Because of the varying surgical environment, creating such a dataset would be difficult to realize. Therefore, for the first version of the system, non-machine learning methods were used.

In Figure 5 a flow diagram shows the recognition process of the measurement system. The program is written in MATLAB R2021a. The process starts with selecting a video file that is to be evaluated. Then the first frame of the video file is taken to locate the indicators of the Conmed. To be able to locate these indicators, first a transformation matrix is estimated to compensate for the perspective of the camera such that the image will be correctly aligned with the reference image. After this, the estimated transformation is evaluated on its correctness. If it is deemed incorrect, the system will try the estimation again with different parameters, up to seven tries based on trial and error. If after seven tries this does not result in a correct estimation, then an error code is generated and the estimation of the video file is terminated. If the transform is deemed correct, the indicators are located within the image. After the location has been found, the correct thresholds for light indicator recognition are determined based on the environment. Then the state of each indicator can be determined through different types of image recognition, depending on the type of indicator. When there is a change in the state of the Conmed system, this is stored in a table containing the timestamp within the video, together with all states that have been measured in the frame. Finally, when the system has evaluated the final frame of the video file, the table is saved in CSV-format and the measurement is finished.

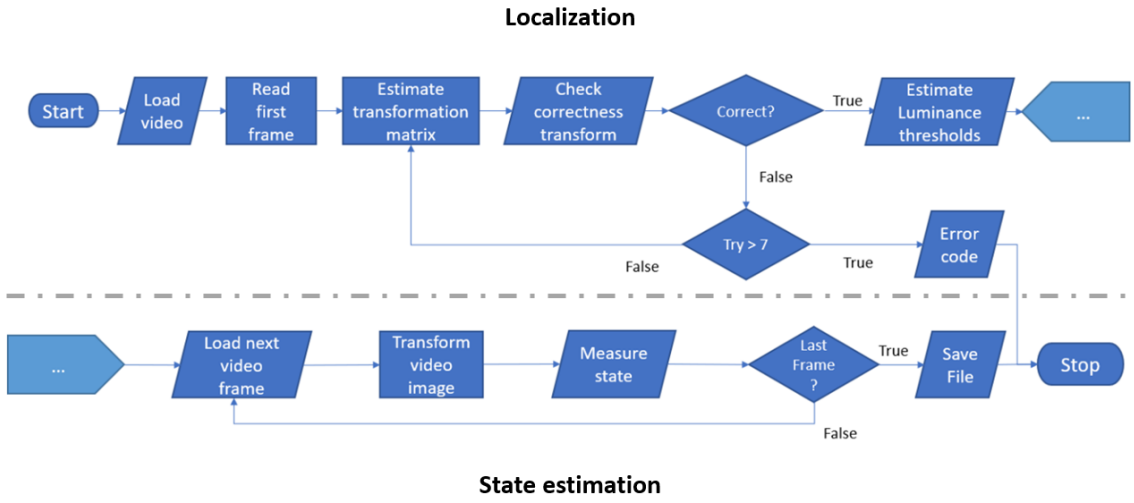


Fig. 5: Flow diagram showing the image recognition process of evaluating one video file with visual data of the Conmed System 2450.

2.3.2 Indicator localization

Since the system has a layout of indicators that does not vary, the location of the indicator can partially be known beforehand. Each indicator has its location and size stored with respect to a reference frame. Figure 6 shows an example of all known locations for each indicator. Through this, the region of interest (ROI) can be determined by only having to find the correct location of the reference frame.

When starting the surgery, the measuring setup is prepared, and the camera is attached to the Conmed. Due to this, the position and perspective of the camera will change per surgery. Therefore, a method needs to be used to compensate for the changed perspective and locate the reference frame. This reference frame is located by transforming the video image of the camera such that it aligns with a reference image, corresponding to the reference frame. In Figure 7 you can see an example of the varying perspective, and the reference image with reference frame and ROI for each indicator.

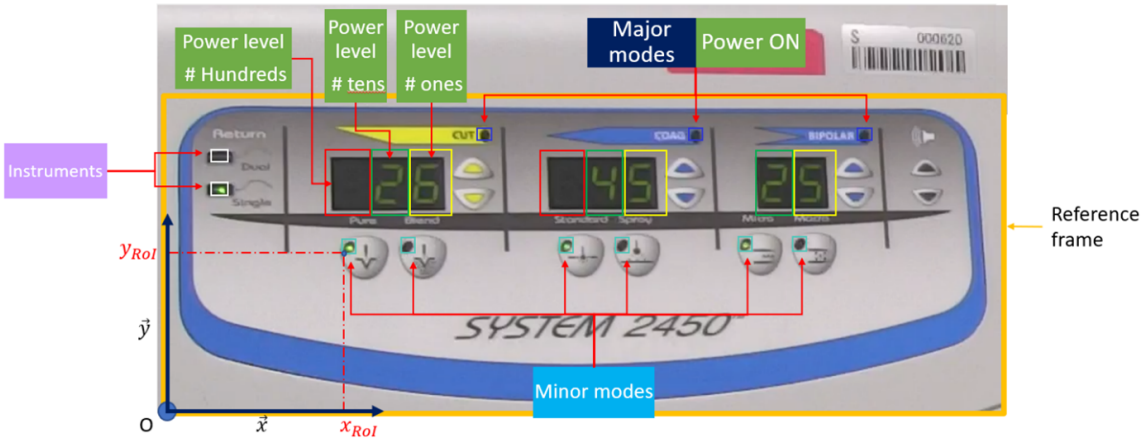


Fig. 6: All regions of interest for the Conmed system 2450. The indicator size together with the location of each indicator with respect to the reference frame are stored in a database.

(a) (b)

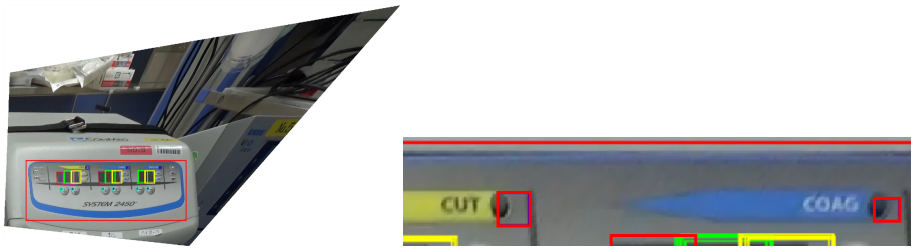
Fig. 7: Example of (a) the camera recording the display of the Conmed System 2450 with a different perspective, and (b) the reference image with corresponding reference frame and the ROIs of all indicators.

(a) (b)

Fig. 8: Example of (a) the detected features based on SURF feature detection, and (b) the RANSAC filtered result.

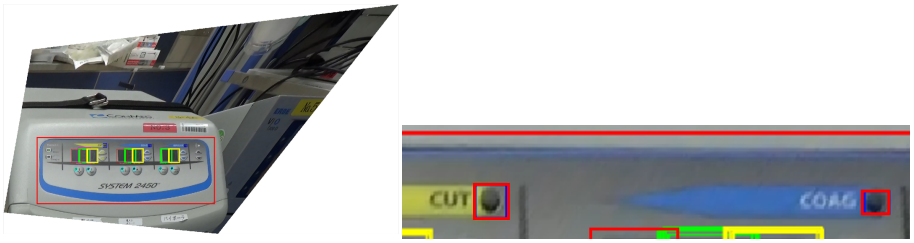
The alignment method consists of Speeded Up Robust Features (SURF) - feature detection, RANdom SAmple Consensus (RANSAC) with Homography transformation, and an Enhanced Correlation Coefficient (ECC) optimization model. This creates a transformation matrix that can be used to transform and align the video image with the reference image. These three algorithms were used in MATLAB through altered functions from the Image Alignment Toolbox (IAT) [12]. First, the features are detected through a SURF-feature detector [13]. After which RANSAC is applied to filter out outliers and estimate the initial transformation matrix [14]. In Figure 8 an example is shown of the features being detected and filtered after applying RANSAC. The parameters of the RANSAC algorithm consist of a tolerance of 0.005, a maximum invalid count of 100, and a maximum iteration of 300. These parameters have been chosen based on trial and error, with the default values in the MATLAB function as a starting position. The transformation estimated by the RANSAC algorithm is set to a perspective transformation, also called homography transformation.

The combination of SURF and RANSAC works great for estimating the transformation matrix for larger transformations, but still needs to be enhanced to align the regions of interest correctly. As can be seen in Figure 9, the alignment is for most indicators already correct, but there is still a slight skew in the image, which make the ROI misaligned for the Cut mode and the Coagulation mode.



(a) (b)

Fig. 9: Example of (a) the first estimation of the transformation matrix, and (b) zoomed in version on the display. It can be seen that for the Cut mode and the Coagulation mode, the ROIs do not align properly.



(a) (b)

Fig. 10: Example of (a) the final estimation of the transformation matrix, and (b) zoomed in version on the display. It can be seen that now the ROI for each indicator correctly aligns.

After the first version of the transformation matrix has been estimated, the result is refined by applying a forward additive Enhanced Correlation Coefficient (ECC) optimization method [15]. This method varies the parameters of the transformation matrix, trying to maximize the Enhanced Correlation Coefficient. The number of iterations is set to 100 based trial and error. The result is a refined transformation matrix, that allows for the correct alignment of the reference frame and the corresponding ROI. The enhanced alignment is shown in Figure 10.

Since RANSAC randomizes the initial conditions and the environment changes over time, the results are not deterministic. This means that in some rare occasions the results can be completely off due to the combination of 'bad' initial conditions together with the complexity of the scene and the preset number of iterations. To deal with this, the measurement system needs to evaluate if the estimated transformation is correct or not. If the transformation matrix is deemed wrong, this results in an error flag and the transformation estimation process needs to restart with different initial conditions. This is done by applying several checks to evaluate if the transformation is correct. First the transformation is evaluated based on the normalized cross correlation between the newly transformed image and the reference image. If the correlation coefficient is below a certain predetermined threshold, then the transformation is incorrect. Next, the Instruments are evaluated. If the transformation is correct, the measurement system should be able to estimate either one of the connected instruments based on the light indicator recognition method. If this results in an error due to initial thresholds not being met, then the transformation matrix is deemed wrong. The same is applied for the Minor modes and the Power levels since these also are always visible. Since the Major mode is only visible when the power is activated, this cannot be evaluated. If all these checks do not result in an error flag, the transformation is deemed correct.

2.3.3 State estimation

After all indicators have been located, the states of the indicators need to be recognized. The states are recognized according to the state tree in the order of Instrument, Major mode, Minor mode, and Power level. Depending on the type of indicator, either a light indicator recognition method or a character

recognition method is used. Based on the first recognized state, the state tree is followed, where the possible states that are being evaluated are the child-states of the current state.

Light indicator recognition

For the light indicator recognition, a new method based on the known position and the luminance ratio is proposed to evaluate the state of the light indicator. After the light indicator locations are known due to the indicator localization method, each light indicator is compared to a corresponding template image of the active state from the template database. First, for each indicator, the video image is cropped to the ROI with the same size as the template image and then converted to grayscale. After this, the sum of the luminance ratio of each pixel of the cropped image is calculated. The same is applied to the corresponding template image. Finally, the ratio of the luminance sum of the cropped image to the luminance sum of the template image is calculated. The corresponding formula is

$$Luminance\ ratio = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} I_{cropped\,image}(i,j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} I_{template\,image}(i,j)} \tag{1}$$

where m denotes the height, and n denotes the width of the template image. If this ratio is higher than a certain threshold, the state of the light indicator is deemed active. Only one indicator can be active for each state. For example, either only the indicator of the dual electrode or the indicator of the single electrode can be active. In case multiple indicators are deemed active, then the maximum value for the ratio is chosen to be the correct active state.

Light indicator threshold estimation

Since the light in the operating theatre can vary for each surgery, the thresholds for the luminance ratio of each state need to be adjusted. For this, a new method for threshold estimation is proposed. This method is based on the different luminance levels from the video image, and stored template images. First the difference between an active indicator from a template image and an inactive indicator of another template image with the same environmental lighting is calculated. This is denoted by $\Delta L2$. By adding this value to the measured inactive luminance level of the video image, the expected maximum luminance level for an active indicator in the current environmental lighting can be estimated. Since the indicator can only have an active or an inactive state, the threshold needs to be somewhere between the measured inactive luminance level, and the maximum expected luminance level. For the threshold estimation, the threshold is placed at the measured inactive luminance level with an additional 40% of $\Delta L2$. Normally you would put this threshold close to 50% to make an equal division between the two states. However, since the actual luminance level of the activated state can be lower than the estimated luminance level, the correct threshold would lower as well. During testing, the actual luminance level of the active state could be as low as 80% of the

estimated maximum. To account for this, the threshold was set to 40%. When the threshold would be lowered further, the variation in environmental lighting results in too many falsely measured activations. The threshold estimation is represented by the following formula:

$$\begin{aligned} \Delta L2 &= LH2 - LL2 \\ Threshold &= \frac{LL1 + \Delta L2 \cdot 0.4}{LH2} \end{aligned} \tag{2}$$

where $LL1$ is the measured minimum luminance level of all inactive light indicators from the video image, $LL2$ is the minimum luminance level of all template images of the inactive light indicators, $LH2$ is the average luminance level value of all template images of the active light indicators, and $\Delta L2$ is the difference in luminance level between the template images of the active and inactive state of the light indicator. This varying threshold allows for the recognition under different lighting conditions.

Character recognition

For the character recognition, template matching in the form of normalized cross correlation (NCC) is applied. The ROI of the located indicator are compared to several template images that describe all possible states of the indicator. First these images are converted to grayscale and then given as input to a Normalized 2-D cross-correlation (normxcorr2) MATLAB function [16, 17]. With this, the template image is moved over the other image and the normalized cross-correlation at each position is computed. The output is a correlation matrix containing all correlation coefficients. After this, the top ten peaks from the correlation matrix are selected for evaluation. To filter out any sharp peaks that are the result of noise, each peak is averaged with their neighboring points. Next, the maximum value of the ten averaged peaks is chosen to determine the state of the indicator. If this value is above a predetermined threshold, the template image is seen as a candidate for the state. Finally, the correlation values of each candidate state are compared, and the state with the maximum correlation value is selected as describing the state of the indicator.

Output data

The output data consists of a table containing a timestamp of the frame of the video file together with all other recognized states. In Table 1 you can see an example of the output data. Only when there is a change in the state of the Conmed System, this is recorded. The Instruments can be read independent of the Activation. Then when there is an activation, the Major mode is recognized, and corresponding Minor mode and Power level is stored. If an error occurred during the indicator localization, this is also stored.

Table 1: Example of the output data of the measurement system

Time [s]	Activation	Instrument	Major mode	Minor mode	Power level
8.162467	Off	Dual electrode	—	—	—
10.40308	On	Dual electrode	Coagulation	Standard	20
12.87417	Off	Dual electrode	—	—	—

Absence of information is displayed with '—'

2.3.4 Camera setup design

The camera setup consists of a camera attached to a flexible arm. This arm is connected to a clamp that can be mounted on the back of the Conmed. Figure 11 shows the camera setup connected to the Conmed. The camera consists of a HD camera with a lens attached to it. The HD camera is a MISUMI-VONA HDTV color camera with a resolution of 1280 x 720p and a build in Infrared (IR) filter [18]. The image sensor consists of an 1/3" Inter line CCD. The output of the camera is HDMI, which can be directly connected to the OPeLiNK system. The lens consists of a TECHSPEC® 4mm UC Series Fixed Focal Length Lens with a 61.9° field of view (FoV) when combined with the sensor format of the selected camera. It has a maximum distortion of -20.43% [19].

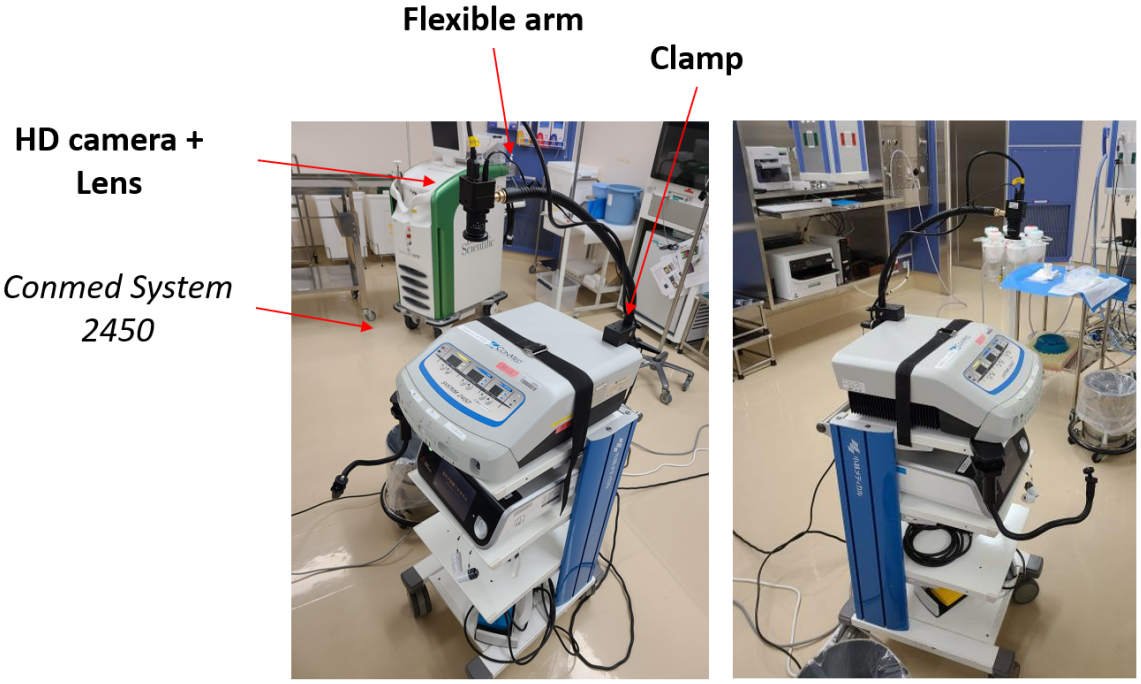


Fig. 11: Example of the camera setup consisting of the HD camera with lens, a flexible arm, and a clamp attached to the Conmed System 2450.

3 Results

For the evaluation of the system three experiments were performed. First, the limit of the perspective transformation method is evaluated to determine the range in which the camera needs to be attached to the Conmed. Then an experiment is performed to evaluate the time synchronization by determining the delay between the video data of the endoscope and the video data from the measurement system. Finally, the measurement system is evaluated based on it's accuracy to determine if it can be integrated into the OPeLiNK system.

3.1 Experiment 1: The limit of the perspective transformation

Purpose

To estimate the maximum angle at which the camera is allowed to be placed, an experiment has been performed to test the limit of the perspective transform estimation.

Methods

A MATLAB program was written that varies the perspective angle of the image in vertical direction, horizontal direction, and a combination of these two. In Figure 12 an example is shown of the range of transformations. The middle image is the untransformed input image. This image is taken normal to the display with an approximate distance of 30cm. The input image is the larger image from which the reference image was cropped. The image can be rotated around the horizontal axis with α, around the vertical axis with β, and the combination of the two. The angles are varied within a range of $-60°$ to $60°$ in steps of $5°$, assuming a field of view of $61.9°$.

Fig. 12: Rotation of the input image with $-60°$ to $60°$ around the horizontal axis with α, around the vertical axis with β, and the combination of the two.

Results

The results show that the limit of the transformation estimation for correct alignment is a maximum rotation of $\alpha = \pm 35°$, $\beta = \pm 45°$, and a combined rotation of $\pm 35°$. If the orientation of the image is beyond this limit, the transformation cannot be estimated and the result is a random transformation as can be seen in Figure 13.

Fig. 13: Example of the incorrect perspective transformation after exceeding the limit of 35° for the camera angle with respect to the normal of the screen.

3.2 Experiment 2: Delay between endoscope data and video data

Purpose

For the integration with the endoscopic data for error evaluation, and future real time performance, it is necessary to know what the delay is after the OPeLiNK system tries to time synchronize it. An experiment has been performed to evaluate this delay.

Methods

Both the endoscope and the camera setup of the measurement system are connected to the OPeLiNK system and used to film the display of the Conmed. In Figure 14 the setup of this experiment is shown. The Conmed is activated multiple times, varying between the Cut mode and Coagulation mode. This video data is shown in real-time by the decision-making navigation screen and recorded by the OPeLiNK system. The camera has a frame rate of 30 fps, and the endoscope has a frame rate of 60fps. A MATLAB program was designed to evaluate each frame of the combined video data. The measurements were performed over a period of 120 seconds. Within this time, 38 changes in the state were performed, consisting of 19 activations and 19 deactivations.

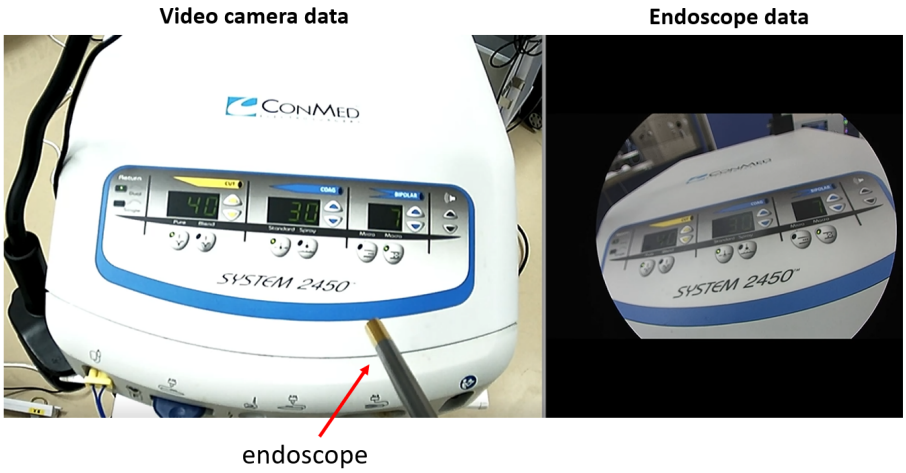
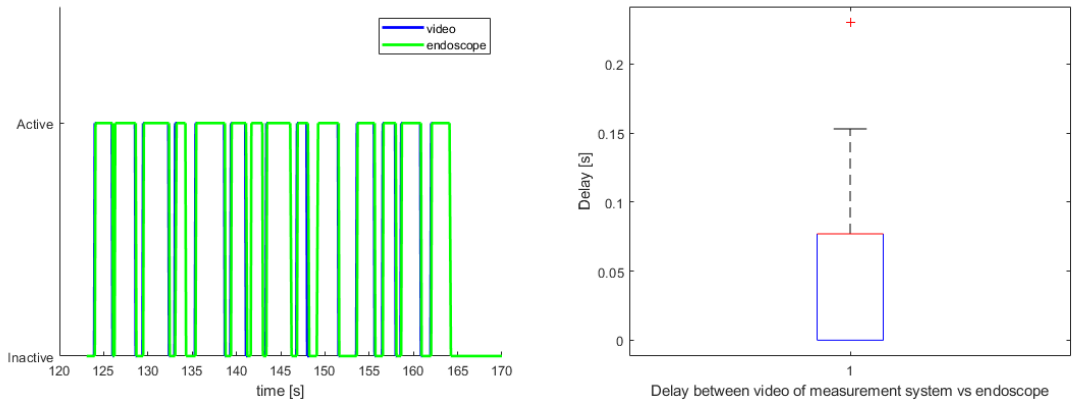


Fig. 14: Setup for the second experiment on measuring the delay between the endoscope data and video data. In this setup, the display of the Conmed is recorded through the video camera of the measurement system and an endoscope. Both are connected to and displayed by the OPeLiNK system. On the left the video data is shown, and on the right the camera data is shown.

Results

In Figure 15 (a) a part of the measurement is shown. It can be perceived that the endoscope has a slight delay compared to the HD camera. In Figure 15 (b) the box plot of this delay is shown. The median of the delay is a value of 0.077s.



(a) (b)

Fig. 15: Example of (a) the measured activations and deactivations over time, and (b) the box plot indicating the delay between the video of the measurement system and the endoscope. The median has a value 0.077s.

3.3 Experiment 3: Applicability of the measurement system

Purpose

To determine if the system can be integrated into the operating room, experiments have performed to measure its accuracy and to evaluate the overall use during real surgeries. The measurement system needs to be accurate enough, so that the measured data can be used for postsurgery evaluation, To evaluate the accuracy of the system, an experiment is performed to measure the accuracy of the total system and the individual indicator localization and recognition methods separately.

Methods

The setup consists of the camera filming the display of the Conmed, which is connected to the OPeLiNK system. This then creates multiple video files, each consisting of 930 frames with 31s duration. The recording is done during the intraoperative stage, while the state evaluation is done postoperatively through a MATLAB program. The scheme of the setup can be seen in Figure 4 on page 8. The measuring system has been tested on 2963 videos consisting of nine colon surgeries performed at Japan's National Cancer Center. With an average operating time of 3h. Data acquisition is conducted with the approval of the ethic committee of Japan's National Cancer Center and the University of Tokyo.

The accuracy of the system Is measured by postoperatively evaluating each video file and manually comparing it to the results that are given by the measurement system. If there is at least one error in the video file, the video is counted as an error. The formula for the accuracy is described by

$$accuracy = \frac{N_{video\,files} - N_{errors}}{N_{video\,files}} \cdot 100\% \tag{3}$$

where $N_{video\,files}$ indicates the total number of evaluated video files and N_{errors} indicates the total number of video files with a detected error.

Results

The measuring system was setup during each surgery by an assistant. The setup was deemed easy, quick, and did not hinder the surgeon in any way. From the 2963 videos in total 130 errors were detected. From these errors, 77 are due to circumstances where the display is either out of view, or the camera is moved. This is the results of either the camera losing connection, the camera view being obstructed, or the camera accidentally being moved. In Figure 16 an example is shown of these errors. Since these 77 errors do not reflect the accuracy of the system, these errors have been excluded. This means that the number of included video files is $N_{video\,files} = 2963 - 77 = 2886$ and the number of detected errors is $N_{errors} = 53$. The resulting accuracy is 98.2% for the overall measuring system. When we look at the accuracy of each

(a) (b) (c)

Fig. 16: Example of the excluded errors due to (a) loss of connection, (b) movement of the camera, and (c) obstruction of the camera.

component, we see that the perspective transform estimation is wrong in 28 videos, resulting in an accuracy of 99.03%. For the light indicator recognition, four mistakes were made in the Major mode detection, and one mistake was made in the Minor mode detection. This results in an accuracy of 99.97% for the light indicator recognition method. In 24 videos, the wrong power level was estimated. This results in a 99.17% accuracy for the character recognition method. Figure 17 and Figure 18 show an example of the measured activations during a surgery performed by a resident and a skilled surgeon, respectively. In these datasets, the identified errors have been removed. The other examples of the datasets of the Conmed can be found in Appendix B.

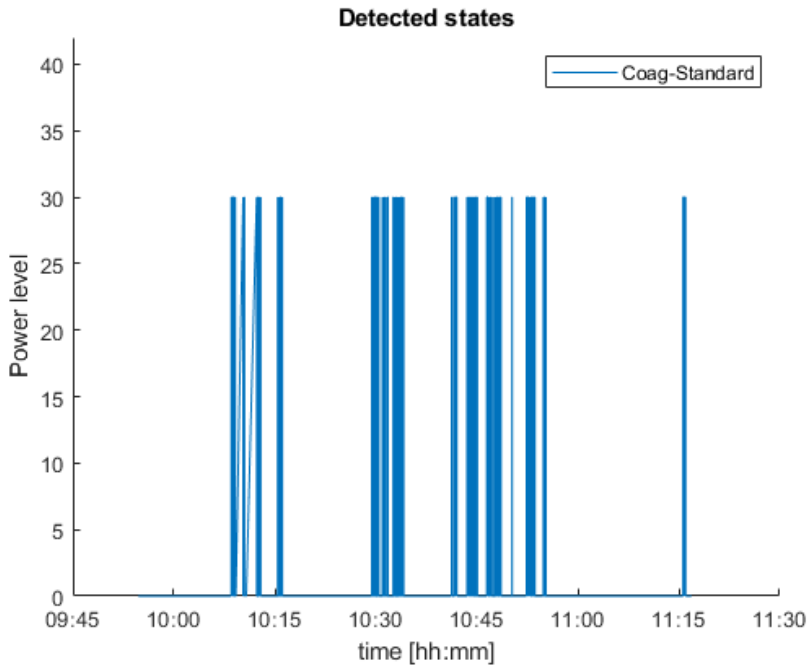


Fig. 17: Example of the measured activation data during a surgery performed by a resident. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. In this surgery, no Cut was performed with the Conmed. The instrument used was a dual electrode.

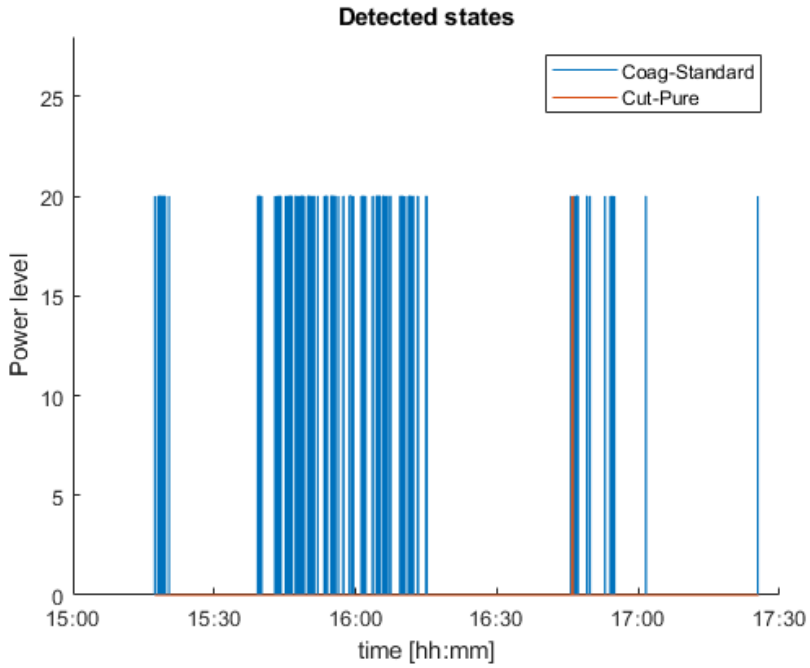


Fig. 18: Example of the measured activation data during a surgery performed by a skilled surgeon.The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

4 Discussion

A new image recognition method is proposed for the localization of the indicators. The novelty of this type of image recognition with respect to, for example license plate detection, is that only the locations of the indicators are predefined with respect to a reference image [20–28]. This allows for faster localization through only having to align the video image with the reference image. Furthermore the correctness of the alignment is evaluated. In terms of indicator recognition, the novelty is in reading multiple types of indicators, through a novel light indicator recognition method, and an existing character recognition method. The novel light indicator recognition method differs from conventional methods used in for example traffic light detection, by not using the color distribution in combination with thresholding, but rather uses the luminance levels with respect to a template image [29–33]. The developed measuring system could be integrated in systems like OR.NET, OpenICE, and SCOT. Furthermore, this novel type of image recognition could be generalized and applied for any type of energy device, with any layout and any combination of indicators.

The results of the first experiment showed that the limit of the alignment method is a maximum angle of 35° with respect to the normal of the display of the Conmed. This is however based on the ideal situation where the initial image that is being warped exactly matches the reference template image. Both have the same lighting conditions and distance with respect to the display.

When the environment changes, the estimation of the transformation becomes more difficult. The number of correctly detected feature points will decrease, and the number of incorrect detected feature points will increase. This means that the noisier the environment is, the lower the maximum angle for correct alignment becomes. Therefore, it is important to try to place the camera as close to the display as possible wile maintaining the minimum required distance of 15cm, while minimizing the angle. This results in the highest chance for the correct localization of the indicators.

The results of the second experiment show that there is a slight delay between the video data of the HD camera compared to the video data of the endoscope. This delay can be seen as the effect of asynchronization between the frames of the endoscope and the camera of the measuring system, together with the communication and processing time of the OPeLiNK system. The delay has a median of 0.077s. For error evaluation, this value would not seem that significant. It could however become a problem, if more complex analysis is performed regarding quantification of a surgeons skill for AI based decision making or (semi-)automated surgery. To improve this result, either a camera with a higher frame rate could be used, or the hardware and software of the OPeLiNK system need to be further optimized.

The result of the final experiment showed that the overall system has an accuracy of 98.2%. Furthermore, the perspective transformation method, the light indicator recognition method, and the character recognition method, have an accuracy of 99.03%, 99.97%, and 99.17%, respectively. These accuracies are above the required accuracy of 90%, and therefore are deemed suitable for the implementation in the measurement system. The errors that were included in the estimation of the accuracy were mainly the result of the changing lighting conditions. This would be either due to reflections of the surgery lights, or due to strong shadows being cast on parts of the display. This creates the situation where for example for the Power level recognition, a different template image would have a higher correlation coefficient than the correct template image, resulting in a wrong state estimation. Regarding the perspective transformation estimation method, the lighting conditions would result in a lower number of detected features for the feature recognition, causing incorrect estimations. In most cases, the incorrect transformations are detected by the transformation accuracy detection. However, in a few cases this detection was bypassed, by having an almost correct transformation, but where light indicators are falsely located due to reflections.

To increase the accuracy, the hardware could be improved by adding a light coming from the camera, to limit the influence of the lighting of the environment. Another method would be by creating a model that would detect these reflections and shadows, and then compensate for this by preprocessing the video image before the recognition stage. This would however increase the recognition time, which can be detrimental for future applications like real time state communication. Another option to increase the accuracy would be to replace certain non-machine learning recognition methods with machine

learning methods in the form of for example Convolutional Neural Networks (CNN) or Support Vector Machine (SVM) [34, 35]. This however would require a significant dataset to train these algorithms.

The estimated accuracy excluded the errors that occurred due to the camera losing visibility of the display. These errors do not reflect the accuracy of the image recognition program but do show the limitations of using image recognition in a surgical environment. If the visual data is somehow limited due to various reasons, the output of the system is unreliable and important information gets lost. There are various ways the system can be improved. One method is to limit the movement of the camera by having a rigid connection with a standard connecting point to the Conmed, instead of a flexible arm. Another option would be to introduce some active or passive stabilizing mechanism that directs the view of the camera always towards the display. Furthermore, a future real time application, the system could recognize errors due to loss of visibility and warn the surgeon or assistant to fix the problem.

According to the surgeon assistant, the system was easy to use and could be quickly setup. The system outperforms the state of the art in the amount of data that can be measured. It not only measures all activations, but also measures new valuable data on the Instrument, the Major modes, Minor modes and Power level used during surgery. This kind of information can be significant for postsurgery evaluation.

5 Conclusion

In this paper a new image recognition based method for the recognition of the states of a surgical energy device is introduced. This method allows for the postsurgical evaluation of the intraoperative data of surgical energy devices in the operating room. This can be applied in systems like OR.NET, OpenICE, and SCOT. In this paper the developed system has been integrated in the Smart Cyber Operating Theatre, SCOT. The developed method has been applied on one energy device, the Conmed System 2450, referred to as the Conmed. First a system analysis and environment analysis were performed to determine the requirements of the measuring system. The visual indicators communicating the state of the Conmed consist of light indicators and character indicators. The states that can be communicated consist of the state names ‘Machine on/off’, ‘Instrument’, ‘Activation on/off’, ‘Major mode’, ‘Minor mode’, ‘Power level’, and ‘Error’. The developed system consists of a camera attached to the Conmed, gathering visual data of the display. This information is communicated to the OPeLiNK system, which creates video files for postsurgery analysis. The video files are given as input to the measuring system, which determines the state of the Conmed for each frame.

The developed system uses image recognition in the form of template matching to determine the state of the Conmed. The indicators are located through a novel method, combining the stored location information with respect to a reference frame and a transformation estimation method. This

transformation estimation method consists of the combination of the existing methods SURF-feature detection, RANSAC with homography transformation, and a forward additive ECC optimization model, together with a novel method of evaluating the correctness of the estimated transformation. For the state estimation of light indicators, a new method is proposed based on the luminance values with respect to a template image. Furthermore, a novel method is proposed for the threshold estimation for light indicator recognition, based on the detected environment. For the character recognition a common method is used based on Normalized Cross Correlation. The output of the measurement system consists of a CSV file containing all detected states of the Conmed and the corresponding timestamp within the video.

The developed system has been evaluated through three experiments. The first experiment showed that the limit of indicator localization method is a maximum camera angle of 35° with respect to the normal of the display of the Conmed. This result is under ideal conditions, and therefore it is advised to minimize the camera angle and put the camera as close to the display as possible with a minimum clearance of 15cm. The second experiment showed that the median of the delay between the video data of the measurement system and the endoscopic data is equal to 0.077s. This small delay is seen as insignificant for postsurgery error evaluation. Finally, the accuracy of the measurement system was evaluated. The accuracy was determined to be 98.2% under the condition that the display of the Conmed is within the view of the camera. Furthermore, the measured accuracies are 99.03%, 99.97%, and 99.17%, for the perspective transformation method, the light indicator recognition method, and the character recognition method, respectively. The developed system meets the requirement of a minimum accuracy of 90%, and therefore the system could be integrated within the OPeLiNK system.

Besides improvements on the accuracy of the system, the next step would be to make the system modular to not only measure the state of the Conmed, but also any other energy device. Furthermore, the system could be designed such that the measured states can directly be communicated to the surgeon and remote experts through telementoring. This would be realized, by integrating the system into the decision-making navigation screen of the OPeLiNK system. For this, the system needs to be optimized in terms of speed such that it can recognize and communicate the states in real time. The current system uses the display for estimating the transformation matrix. In the future, it might be possible to introduce a standardized QR code or other marker on the energy device to be able to recognize the perspective of the energy device more efficiently.

Finally, only detecting visual information does have limitations in the robustness of the measurement system. This could be overcome by either adding auditory data or other data on the current and voltage of the system, or a combination of these.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Ethics approval Data acquisition is conducted with the approval of the ethic committee of Japan's National Cancer Center and the University of Tokyo.

Funding This study is supported by AMED (grant No. 21he2102001h1003).

References

[1] Rockstroh, M., Franke, S., Hofer, M., Will, A., Kasparick, M., Andersen, B., Neumuth, T.: Or.net: multi-perspective qualitative evaluation of an integrated operating room based on ieee 11073 sdc **12**(8), 1461–1469 (2017). https://doi.org/10.1007/s11548-017-1589-2

[2] Arney, D., Plourde, J., Goldman, J.M.: Openice medical device interoperability platform overview and requirement analysis **63**(1), 39–47 (2018). https://doi.org/10.1515/bmt-2017-0040

[3] Okamoto, J., Masamune, K., Iseki, H., Muragaki, Y.: Development concepts of a smart cyber operating theater (scot) using orin technology. Magazine **63**(1), 31–37 (2018). https://doi.org/10.1515/bmt-2017-0006

[4] Bitterman, N.: Technologies and solutions for data display in the operating room. Magazine **20**(3), 165–73 (2006). https://doi.org/10.1007/s10877-006-9017-0

[5] Ogiwara, T., Goto, T., Fujii, Y., Nakamura, T., Suzuki, Y., Hanaoka, Y., Ito, K., Horiuchi, T., Hongo, K.: Endoscopic endonasal approach in the smart cyber operating theater (scot): Preliminary clinical application. Magazine **147**, 533–537 (2021). https://doi.org/10.1016/j.wneu.2020.12.114

[6] Muragaki, Y., Okamoto, J., Masamune, K., Iseki, H.: Smart cyber operating theater (scot): Strategy for future or. In: Hashizume, M. (ed.) Conference Name, pp. 389–393. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-4325-5_53

[7] Meeuwsen, F.C., Guédon, A.C.P., Arkenbout, E.A., van der Elst, M., Dankelman, J., van den Dobbelsteen, J.J.: The art of electrosurgery: Trainees and experts. Magazine **24**(4), 373–378 (2017). https://doi.org/10.1177/1553350617705207

[8] Dums, J., Schneider, B., Badin, A.: Low cost system to measure active power in electrosurgical units. Magazine **33** (2017). https://doi.org/10.1590/2446-4740.03217

[9] Ushimaru, Y., Takahashi, T., Souma, Y., Yanagimoto, Y., Nagase, H., Tanaka, K., Miyazaki, Y., Makino, T., Kurokawa, Y., Yamasaki, M., Mori, M., Doki, Y., Nakajima, K.: Innovation in surgery/operating room driven by internet of things on medical devices **33**(10), 3469–3477 (2019). https://doi.org/10.1007/s00464-018-06651-4

[10] CONMED Corporation: Conmed System 2450 Service Manual. https://www.conmed.com/en/products/orthopedics/total-joint-replacement/

electrosurgical-generators/system-2450-electrosurgical-generator-esu (2013)

[11] Geirhos, R., Janssen, D.H.J., Schütt, H.H., Rauber, J., Bethge, M., Wichmann, F.: Comparing deep neural networks against humans: object recognition when the signal gets weaker. ArXiv **abs/1706.06969** (2017)

[12] Evangelidis, G.: IAT: A Matlab toolbox for image alignment. https://sites.google.com/site/imagealignment/ (2013)

[13] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Magazine **110**(3), 346–359 (2008). https://doi.org/10.1016/j.cviu.2007.09.014

[14] Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**, 381–395 (1981)

[15] Evangelidis, G.D., Psarakis, E.Z.: Parametric image alignment using enhanced correlation coefficient maximization. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(10), 1858–1865 (2008). https://doi.org/10.1109/TPAMI.2008.113

[16] MathWorks: normxcorr2. https://nl.mathworks.com/help/images/ref/normxcorr2.html (2006)

[17] Lewis, J.P.: Fast Normalized Cross-Correlation (1995). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.6062

[18] Misumi-Vona: MISUMI-VONA HDTV color camera. https://jp.misumi-ec.com/vona2/detail/110400400970/ (2012)

[19] Edmund Optics: 4mm UC Series Fixed Focal Length Lens. https://www.edmundoptics.com/p/4mm-uc-series-fixed-focal-length-lens/2966/ (2014)

[20] Pervej, M., Das, S., Hossain, M.P., Atikuzzaman, M., Mahin, M., Rahaman, M.A.: Real-time computer vision-based bangla vehicle license plate recognition using contour analysis and prediction algorithm **21**(4) (2021). https://doi.org/10.1142/S021946782150042X

[21] Wahyono, Jo, K.H.: Information retrieval of LED text on electronic road sign for driver-assistance system using spatial-based feature and Nearest Cluster Neighbor classifier. Institute of Electrical and Electronics Engineers Inc. (2014). https://doi.org/10.1109/ITSC.2014.6957744

[22] Jain, A., Sharma, J.: Classification and interpretation of characters in

multi-application OCR system. Institute of Electrical and Electronics Engineers Inc. (2014). https://doi.org/10.1109/ICDMIC.2014.6954231

[23] Ahmed, A.A., Ahmed, S.: A real-time car towing management system using ml-powered automatic number plate recognition **14**(11) (2021). https://doi.org/10.3390/a14110317

[24] Ajanthan, T., Kamalaruban, P., Rodrigo, R.: Automatic number plate recognition in low quality videos (2013). https://doi.org/10.1109/ICIInfS.2013.6732046

[25] Zheng, L., He, X., Samali, B., Yang, L.T.: An algorithm for accuracy enhancement of license plate recognition **79**(2), 245–255 (2013). https://doi.org/10.1016/j.jcss.2012.05.006

[26] Fu, M., Chen, N., Hou, X., Sun, H., Abdussalam, A., Sun, S.: Real-time vehicle license plate recognition using deep learning **494**, 35–41 (2019). https://doi.org/10.1007/978-981-13-1733-0_5

[27] Biyabani, A.A., Al-Salman, S.A., Alkhalaf, K.S.: Embedded real-time bilingual ALPR. Institute of Electrical and Electronics Engineers Inc. (2015). https://doi.org/10.1109/ICCSPA.2015.7081311

[28] Massoud, M.A., Sabee, M., Gergais, M., Bakhit, R.: Automated new license plate recognition in egypt. Magazine **52**(3), 319–326 (2013). https://doi.org/10.1016/j.aej.2013.02.005

[29] Diaz-Cabrera, M., Cerri, P., Medici, P.: Robust real-time traffic light detection and distance estimation using a single camera **42**(8), 3911–3923 (2015). https://doi.org/10.1016/j.eswa.2014.12.037

[30] Chen, Q., Shi, Z., Zou, Z.: Robust and real-time traffic light recognition based on hierarchical vision architecture. Institute of Electrical and Electronics Engineers Inc. (2014). https://doi.org/10.1109/CISP.2014.7003760

[31] Ying, J., Tian, J., Lei, L.: Traffic light detection based on similar shapes searching for visually impaired person. Institute of Electrical and Electronics Engineers Inc. (2015). https://doi.org/10.1109/ICICIP.2015.7388200

[32] Wang, W., Sun, S., Jiang, M., Yan, Y., Chen, X.: Traffic lights detection and recognition based on multi-feature fusion **76**(13), 14829–14846 (2017). https://doi.org/10.1007/s11042-016-4051-5

[33] Tran, T.H.-P., Pham, C.C., Nguyen, T.P., Duong, T.T., Jeon, J.W.: Real-Time traffic light detection using color density. Institute of Electrical and

Electronics Engineers Inc. (2016). https://doi.org/10.1109/ICCE-Asia.2016.7804791

[34] Liu, Q., Chen, S.L., Li, Z.J., Yang, C., Chen, F., Yin, X.C.: Fast recognition for multidirectional and multi-type license plates with 2d spatial attention **12824 LNCS**, 125–139 (2021). https://doi.org/10.1007/978-3-030-86337-1-9

[35] Gonçalves, G.R., Menotti, D., Schwartz, W.R.: License plate recognition based on temporal redundancy (2016). https://doi.org/10.1109/ITSC.2016.7795970

3 Manufacturer Independent Measuring System for the Real Time State Communication of any Energy Device

In the following paper the previous measuring system has been expanded, and a novel manufacturer independent measuring system is developed for the real-time state estimation and communication of any surgical energy device. This system allows not only for postoperative evaluation, but now allows for telementoring in the form of real time communication to the surgeon and remote experts. This system uses the same image recognition techniques as discussed in the previous paper. The further contributions are a novel communication protocol describing the states of any energy device and a new dynamic reading strategy that can deal with any display layout and that improves the accuracy, and recognition speed of the system. The system is integrated into the OPeLiNK system of the Smart Cyber Operating Theatre, SCOT.

Development of a manufacturer independent method for the real-time state communication of surgical energy devices

Pepijn van Esch[1,2,3*]

[1*]Biomedical engineering, Delft University of Technology, Mekelweg 2, Delft, 2628 CD, Netherlands.
[2*]Mechanical engineering, Delft University of Technology, Mekelweg 2, Delft, 2628 CD, Netherlands.
[3*]The Graduate School of Engineering, The University of Tokyo, Bunkyo-ku 7-3-1, Tokyo, 113-8656, Japan.

Corresponding author(s). E-mail(s): p.vanesch@student.tudelft.nl;

Abstract

Purpose: At the moment, surgical analysis through intraoperative images is a widely researched field in medicine and engineering. The addition of various surgical device data is expected to improve the accuracy of this surgical analysis. The aim of this study is to develop an image recognition system that can acquire the usage status of various types of energy devices in real time and integrate them with endoscopic images.

Methods: To be able to read the status of any energy device, a common communication protocol was developed. A reading strategy was developed, where each state-option is determined by its parent-state. The developed system uses image recognition in the form of template matching. A setup program has been developed to create a template database. For Light indicator recognition, the luminance values are evaluated. For numbers, text, and symbols, Normalized Cross Correlation is used. The states of the energy device are integrated with endoscopic images using the OPeLiNK system of the Smart Cyber Operating Theatre, SCOT.

Results: Experiments showed a 98.2%, and a 99.6% accuracy for the recognition of the Conmed and Harmonic, respectively. Furthermore, results showed a mean recognition time of 0.037s and 0.166s, for one and five active energy devices, respectively. Compared to the state-of-the-art, both have a 100% accuracy for the recognition of the power

activation. The delay between them is 0.07s. Finally, our system was integrated in the OPeLiNK system and evaluated on the delay in communication time between video, endoscopic, and system-communicated data. This showed that there is a delay with a median of 0.308s and a mean measuring time of 0.373s. Our system has a mean measuring time of 0.11s, showing that the delay is in the communication. The fifth experiment showed that the OPeLiNK system cannot handle messages send with a higher frequency than 1Hz.
Conclusion: The measurement system meets the set requirements of a minimum accuracy of 90% and a maximum recognition speed of 0.2s. Furthermore, it showed that the developed system can compete with the state-of-the-art in terms of measuring more information of multiple types of energy devices at the cost of a slight delay, in real time. Currently, the bottle neck is the communication with the OPeLiNK system. To improve this, either the TCP model needs to be changed, or the OPeLiNK system needs to be updated, such that it can handle more messages in a shorter time.

Keywords: OPeLiNK, energy devices, image recognition, template matching

1 Introduction

Recently, using intraoperative imaging for surgical analysis has been frequently researched. Improvement of the accuracy of the surgical analysis could be achieved by reading and storing surgical device data. An example of a system that integrates data of multiple surgical devices is the Smart Cyber Operating Theatre (SCOT). Time-synchronized patient condition data, diagnostic images, and the status of surgical devices are stored. Furthermore, this data is displayed in real time in the OPeLiNK integrated communication interface as can be seen in Figure 1 [1]. One of the main functions of SCOT is that it allows for the communication of only the data relevant for the phase of the surgery through the decision-making navigation screen. This screen allows for telementoring, and reduces the information overload [2]. Through telementoring the surgeon can be assisted by a remote expert during surgery [3]. This decision-making navigation screen can show various data in real time, that ranges from the patient monitor to, for example, the operating room video. This information is chosen by the remote expert and can be displayed in any layout. The remote expert can then communicate with the surgeon that is performing the surgery and gives the surgeon directions based on the displayed information.

The purpose of this research is to develop a measuring system that allows for the integration of the status of any energy device into the integrated operating room. Although several types of surgical equipment have already been integrated into the system, surgical energy devices have yet to be integrated. Surgical energy devices can range from electrosurgical tools to harmonic

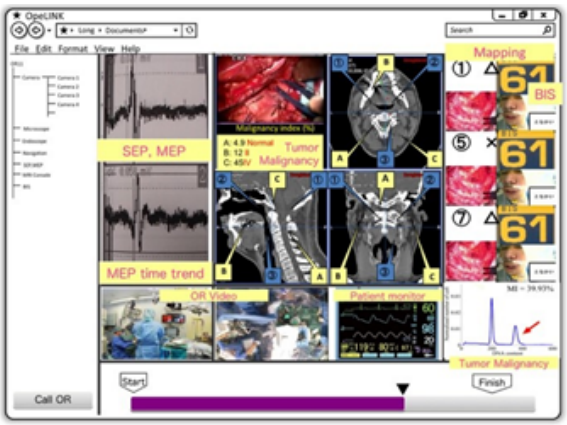


Fig. 1: Picture of the decision-making navigation screen of the OPeLiNK system [1].

scalpels, or other vessel sealing devices. Within Japan's National Cancer Center, there are for example seven types of energy devices used. Depending on the type of surgery and the surgeon, one or multiple of these devices are used. Currently, the problem for integration into the OPeLiNK system is that these devices all communicate their state in a different way. Furthermore, you cannot just access their software since this is protected by the manufacturer. Therefore, a different method needs to be used to measure and communicate the status of these devices.

There are three main researches that have been performed on the state acquisition for an energy device [4–6]. All three measure the activation durations and number of activations through the current and voltage. One of these also estimates the output power. The measured information is however limited since the other settings of the devices are not measured. Also, the states of the energy device are not communicated with the surgeon in real time since that was not the purpose of these studies. Finally, these measurement systems only work for one type of device, the electrosurgical tool, while many other energy devices are used during surgery.

The energy device communicates most of its states visually, while only some states are communicated through auditory data. Therefore, this study proposes a new method for the acquisition of the states of any energy device through image recognition in real time. The developed system uses image recognition in the form of template matching. The basic design of the image recognition system for postsurgery error evaluation has been discussed in the previous paper. In this paper there is less focus on the image recognition techniques for the different types of indicators, and more focus on the design of the full system with its real time application and integration into the OPeLiNK system. The system is integrated into the OPeLiNK system for real time communication through the decision-making navigation screen for telementoring. Our system is designed for any energy device, with as reference the seven different types of energy devices used in Japan's National Cancer Center. To be able to read any energy device a common state protocol was developed. Based on this protocol a reading strategy was created to improve

the accuracy and reduce the reading time. A setup program has been developed for the registration of any surgical energy device in the preoperative stage. In the intraoperative stage the system gathers visual data from multiple energy devices. This data is given as input to the main program, which estimates and communicates the states of the energy devices in real time to the surgeon and remote experts through the decision-making navigation screen.

The contributions of this paper can be summarized as follows:

- A novel measuring system has been developed for the real time communication of the states of any energy device.
- A novel manufacturer independent communication protocol has been developed for any energy device.
- A new reading strategy has been designed to increase the accuracy and decrease the computation time of the measuring system.
- A modular setup program for the registration of new energy devices has been developed.
- The system can read and communicate the states of multiple energy devices in parallel with real time performance.
- The system can communicate to the surgeon and remote expert through the decision-making navigation screen of the OPeLiNK system. This allows for telementoring.

2 Method

To develop the system, first a common state analysis was performed to create the communication protocol of any energy device. From this, a reading strategy was developed to improve the accuracy and the reading time of the system. With the communication protocol and reading strategy, the measurement system has been developed. The following section discusses these topics in the presented order.

Requirements

The initial requirements of the system are set to outperform a human. This means that the system must have a minimum accuracy of 90% with a maximum recognition time of 0.2s [7]. It is estimated that a maximum of five different energy devices are used during one surgery. Of these five energy devices it is estimated that a maximum of two energy devices can be used at the same time by two surgeons. The other 3 machines would be expected to be inactive, meaning that they are turned on and display some information, but are not activated when the other two machines are actively used. The system has been designed for these requirements and can read the states of up to five machines in parallel.

2.1 Common State Analysis

To make sure the measuring system can measure any energy device, a common state analysis was performed, based on which a state tree was developed describing all possible states of any energy device. This communication protocol is based on seven energy devices that are used within Japan's National Cancer Center. These devices can be seen in Figure 2, and consist of the following:

1. Conmed System 2450
2. Valleylab ForceTriad (Ligasure)
3. Valleyab FT10
4. ERBE VIO 300D
5. ERBE VIO3
6. OLYMPUS Thunderbeat
7. ETHICON Harmonic & EnSeal

Fig. 2: Images of all energy devices used within Japan's National Cancer Center consisting of (1)Conmed System 2450, (2)Valleylab ForceTriad (Ligasure), (3)Valleyab FT10, (4)ERBE VIO 300D, (5)ERBE VIO3, (6)OLYMPUS Thunderbeat, (7)ETHICON Harmonic & EnSeal.

The manuals of the machines were analyzed to determine the states they can communicate, and the methods they use to communicate this information [8–14]. This is then generalized to a common state communication protocol, where the resulting state representation scheme is shown in Figure 3. The scheme consists of several layers, each describing a certain type of state. These layers consist of the names 'Machine ON/OFF', 'Program', 'Instruments', 'Major mode', 'Minor mode', 'Power level', 'Extra options', and 'Errors'. The states are communicated visually through four types of indicators, consisting of light indicators, number indicators, text indicators, and symbol indicators.

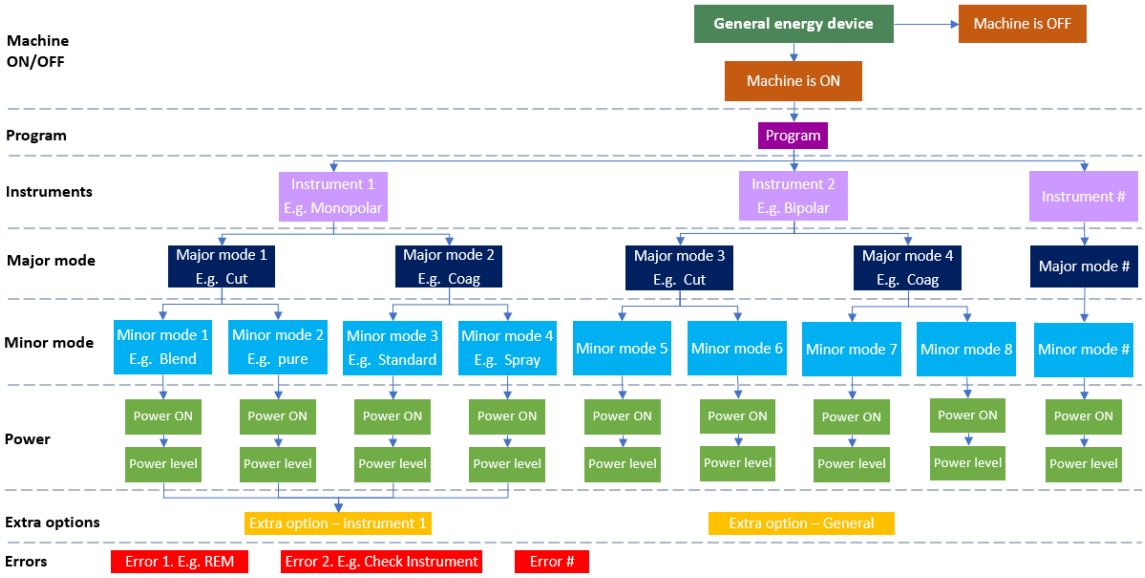


Fig. 3: Generalized classification tree describing all possible states of an energy device. The colors are used to emphasize different layers for better identification in the reading strategy.

Machine on/off

Depending on the machine there might be a light indicator to show if the machine is turned on or off. But in general, the state can be recognized by either the screen showing information, or the indicators of the other states showing the state of the device.

Program

Some machines like the ERBE VIO3 and ERBE VIO300D have a certain program that can be created, edited, and selected by the user. This program can determine the layout of the display, the symbols used for certain indicators, but also program settings, such as mode and power level (called 'effect' for the ERBE – series). This state is mainly communicated through text indicators, indicating the name of the program.

Instrument

Practically every energy device shows what type of instrument is connected to it. Depending on the type of energy device the instrument can vary from monopolar to bipolar, and single electrode to dual electrode for electrosurgical units. For the harmonic scalpel like the ETHICON Harmonic & Seal, it can for example determine the availability of certain modes. The instrument can be communicated through, light indicators, text, symbols, or any combination of these types of indicators. The type of indicator is dependent on the machine that is used or the program that is selected.

Major mode

The Major mode is the main mode of the device. This can vary from cutting, coagulating, bipolar to hemostasis, or any other cutting/vessel sealing method.

The major mode is always communicated with some type of indicator light, and sometimes with additional indicators like text or symbols.

Minor mode

The minor mode describes the specific type of mode, like blend cut or pure cut. Depending on the major mode, the number of minor modes can vary from non-existing to several minor modes that can be selected. The minor modes can be communicated through any type of indicator depending on the machine.

Power: Activation on/off & Power Level

The Power layer is split up in two sublayers: the activation and the Power level. This is because not every Minor mode has a Power level. Sometimes the power can only be 'On' or 'Off'.

Every energy device shows on the display if the power is activated or not and thus if cutting, coagulating, etc. is being performed. This is in general shown by a light indicator and sometimes accompanied by a text indicator or symbols. This indicator is always shared with the Major mode, meaning that when the power is activated, the Major mode becomes visible.

The Power level, or effect, describes the power that is going to be used to cut, coagulate, or the power of any other Major mode that can be performed. Depending on the machine, this level can be adjusted from single digits to tens and sometimes even hundreds. Depending on the type of machine, or Major mode, the power level might not always be available. The power level is always communicated through one or several number indicators.

Extra options

Some machines have extra options that can be selected. These options can be either independent or dependent on the selected equipment.

Errors

Some machines give information if an error occurs. This is also sometimes dependent on the type of instrument attached.

Almost all these states are dependent on their parent state in the state tree. The parent state can influence the states in the next layer in different ways. It can influence the availability of certain states, or it can influence the way certain states are represented. Furthermore, the combination of the measured states in each layer determines the interaction between the tool and the tissue. This standardized communication protocol not only allows us to develop a measurement system that can read the state of any energy device, but also creates an opportunity to develop a reading strategy where each state-option is determined by its parent. This would reduce the number of possible state options, thereby increasing the accuracy while reducing the recognition time. The classification tree of each energy device used in Japan's National Cancer Center can be found in Appendix C.

2.2 Reading strategy

Based on the developed communication protocol, a reading strategy was created for the measurement system as shown in Figure 4. The reading strategy of the program is as follows. Before the surgery starts, the machines that are going to be used are turned on, and if available, the program is selected. This only must be read once since this is not going to change during the surgery. This reduces the real-time workload for the system.

During the surgery, all other states are estimated continuously. The instrument is rarely changed during surgery and therefore does not have to be read at a high frequency. The reading period for the instrument is chosen to be every 10s to reduce the real time workload. Every 0.2s the other states are read. These are in the order of the Major mode and Power ON, which are given by the same indicator. Then from there the Minor mode, Power level and Extra options are read, where each state-possibility is determined by its parent. If an error occurs during this process this is also measured. Through this reading strategy, the amount of state options to be evaluated are minimized, resulting in a reduced estimation time and increased accuracy.

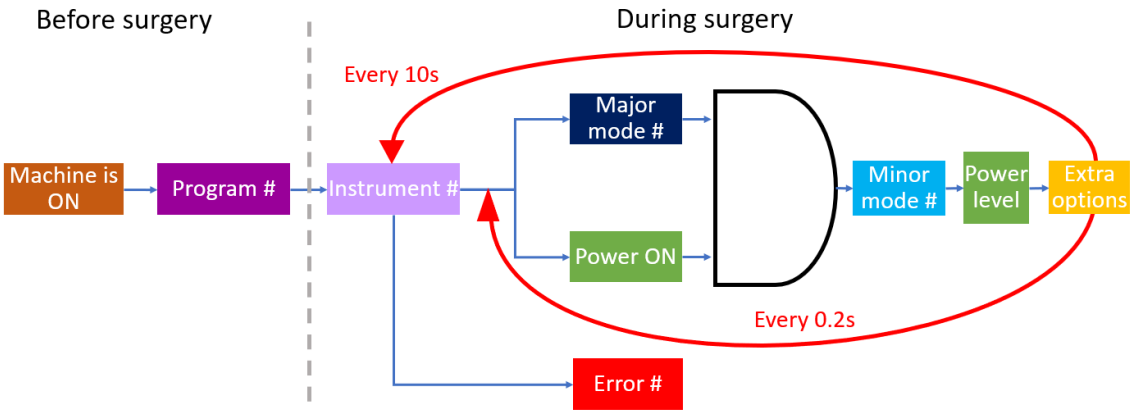


Fig. 4: Reading strategy for measuring the state of any energy device. Each state-possibility is determined by its parent in the previous layer.

2.3 Measuring system

Figure 5 shows an overview of the developed system. The measuring system uses image recognition in the form of template matching to determine the state of the device. It consists of two programs, a setup program for the preoperative stage, and a main program for the intraoperative stage.These programs are run on a laptop with a 9th generation Intel Core i7 9750H /2.6GHz/6 core CPU, 16GB RAM, and a GeForce GTX 1660Ti graphics card. During the preoperative stage, we can register any surgical energy device and make a classification tree for it. Based on this we then classify, preprocess and store template images in the template database. The data from this database is then used during the intraoperative stage. During this stage, we use the main

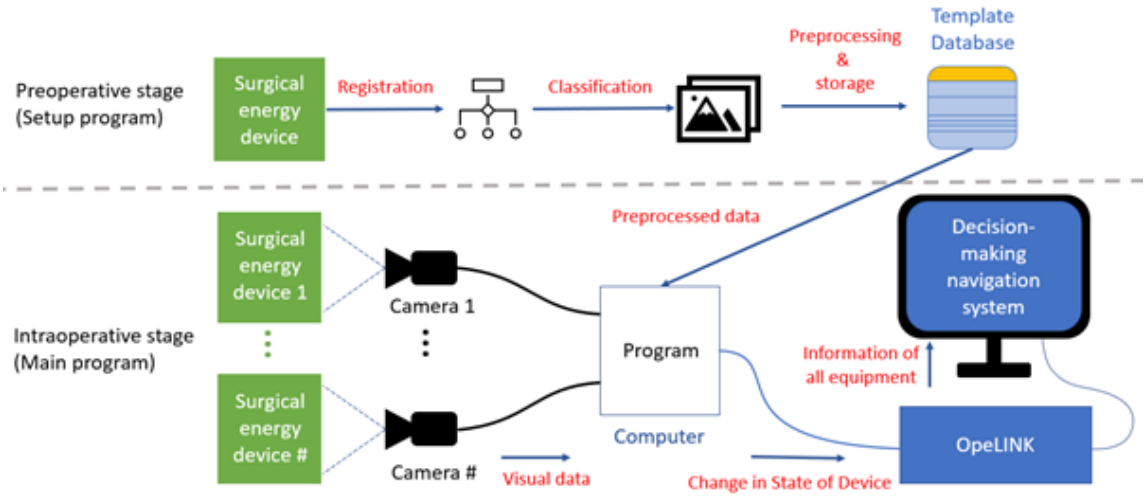


Fig. 5: Overview of the developed measuring system. In the preoperative stage, a setup program is used to register new energy devices, and create the template database. In the intraoperative stage, the main program uses image recognition to measure the state of one or several energy devices. This is communicated through the OPeLiNK system to the surgeon and remote experts.

program to gather visual data from one or multiple cameras filming the display of each energy device. This data is then compared to the preprocessed data from the database, and the state of the energy device is estimated. When there is a change in the state of the device, this is communicated to the OPeLiNK system. This is communicated to the surgeon and remote expert through the decision-making navigation screen.

2.3.1 Setup program

In the preoperative stage, a new energy device can be registered through the setup program. The setup program allows for the registration of any energy device through a simple GUI. The setup program is written in MATLAB R2021a. Figure 6 describes the process of registering a new energy device. To register a new device, first the name of the device is written as input by the user. Then a classification tree describing the relation between all the states of the energy device needs to be developed. The first layer of the classification tree consists of the name of the energy device. The consecutive layers can be added by the user through first selecting the layer the state belongs to, and then selecting a parent state in the upper layer from a choice list. Finally, the name of the new state is given as input and the classification tree is updated.

Depending on the machine, not every layer is applicable. For example, the Conmed does not have a program that can be selected. In this case, instead of giving the name of the state as input, the user can create a placeholder state, which does not have to be evaluated during the state recognition in the intraoperative state. The name of such a state would be 'No_ {layer name}'. Sometimes the user might make a mistake in adding the wrong state. To deal with this, the setup program allows the user to delete any state they want to remove. If a state is selected to be removed, all child states of this state are

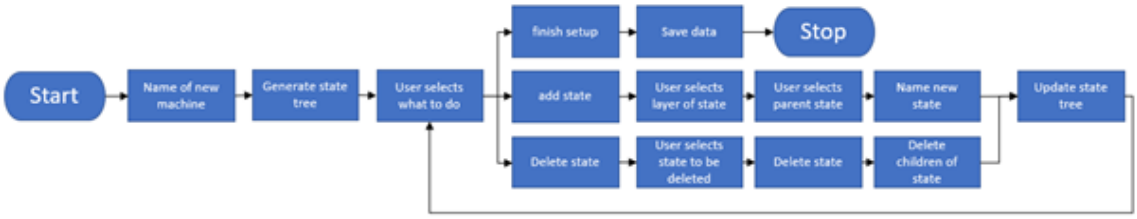


Fig. 6: Flow diagram showing the process of registering a new energy device, and creating a classification tree for it. The classification tree will determine the relationship between the states of each layer.

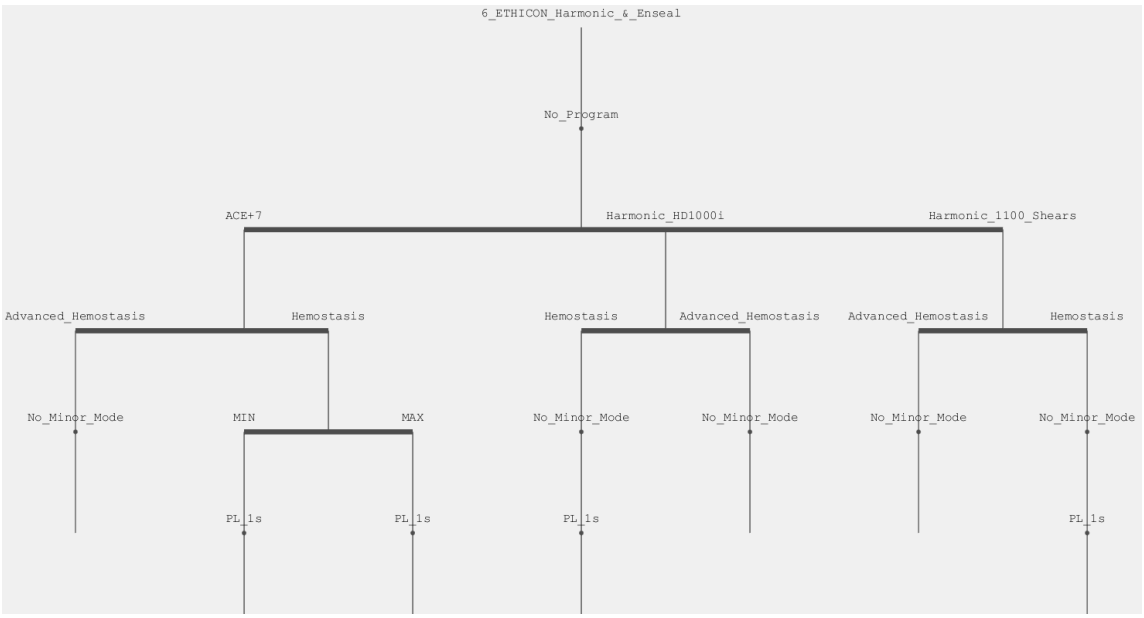


Fig. 7: Example of the generated classification tree for the Harmonic. This classification tree determines the relationship between the states of each layer. The layers consist of the name of the energy device on top, followed by the Program, the Instruments, the Major modes, the Minor modes, and the Power levels.

also automatically removed. This process continues until the full classification tree has been created and then the classification tree is stored in the database. In Figure 7 an example of the generated classification tree for the Harmonic is shown. The classification tree is visualized through a MATLAB function developed by Tinevez, J.-Y [15].

When the full classification tree has been created, the user can start creating the template image database. Figure 8 shows the process of creating this database. To be able to create this template database, the system requires a video file where the display of the to be registered energy device is shown. This video file needs to show all states, the user wants to add. First the user is asked to create a reference image for the future localization of the indicators. Then, the user is asked to give the time stamp of the video frame that they want to use, and the image is shown to the user. After this, the user is asked to crop the reference image out of the video frame, by dragging a rectangle

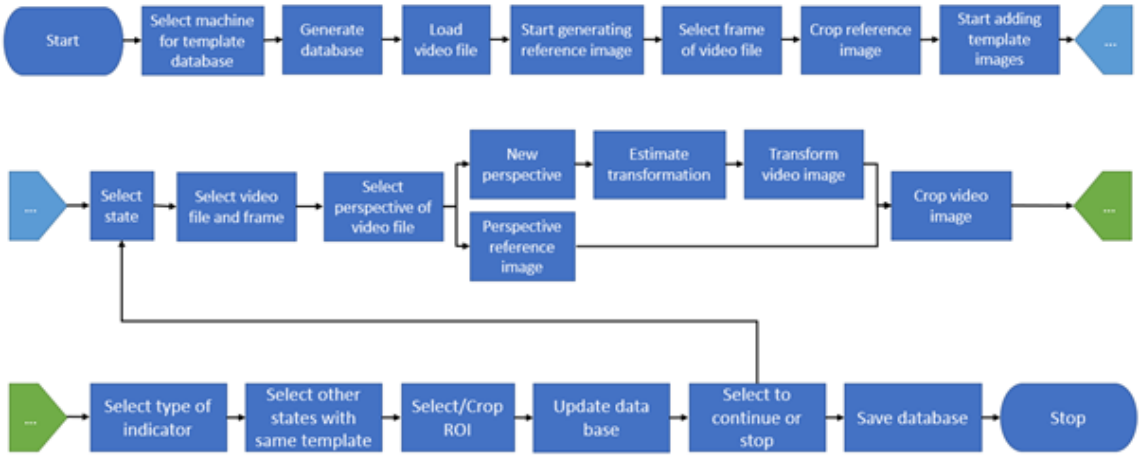


Fig. 8: Flow diagram showing the process of creating a reference image and template images for the template database.

over the display of the energy device. The image is then cropped, and the corresponding Region of Interest (ROI) is stored in the database. Any template image that will be added after this, will have the information of the location of the indicator with respect to the reference frame, stored in the database. After having created the reference image and corresponding reference frame, the user can start adding template images. The user selects a state, for which they want to create a template image, and is asked to select the desired video frame and crop the image to the desired template image. When the template image is created, the user is asked to select what type of indicator the image represents. This can be either a light indicator, a text indicator, a number indicator, or a symbol indicator. The choice is stored in the database and will determine during the intraoperative stage, what type of image recognition will be used to estimate the state of the indicator.

Sometimes the created template image can also be used for other states if they use the same type of indicator. The user can then select the other states from the state tree that also can use this template image. The user then needs to indicate the locations of the indicators by either selecting an existing location of a different state or creating a new location by cropping the ROI in the video image.

In some situations, multiple video files with varying perspective need to be used to create additional template images. To handle this, the user can indicate that the perspective of the new video file is different than the original video file which was used to create the reference image. The frames of the video file are then transformed and aligned with the reference image through a transformation estimation method that was proposed in the previous paper in this thesis. From there, the user can perform the same process of adding template images, where the locations of the template images are stored with respect to the reference frame.

After all template images have been created, the images are preprocessed and stored in the database. This database is then used in the intraoperative stage for the state recognition of the energy devices.

2.3.2 Main program

Figure 9 shows a flow diagram describing the main program The main program is written in MATLAB R2021a. During the intraoperative stage, visual data of the displays of the energy devices is collected and given as input to the main program. This in turn tries to estimate the state of the energy devices and communicates this information to the OPeLiNK system. The program has two phases, one phase just before the surgery starts and the other phase during the surgery. In the first phase the cameras are calibrated and the data from the database is collected and preprocessed. After this, the indicators are located within the video images, and the light indicator thresholds are estimated based on the environment. Then the program follows the reading strategy, where the Program state of the energy device is recognized. In the second phase, during the surgery, the system continuously estimates the other states of the energy devices and communicates this to the OPeLiNK system.

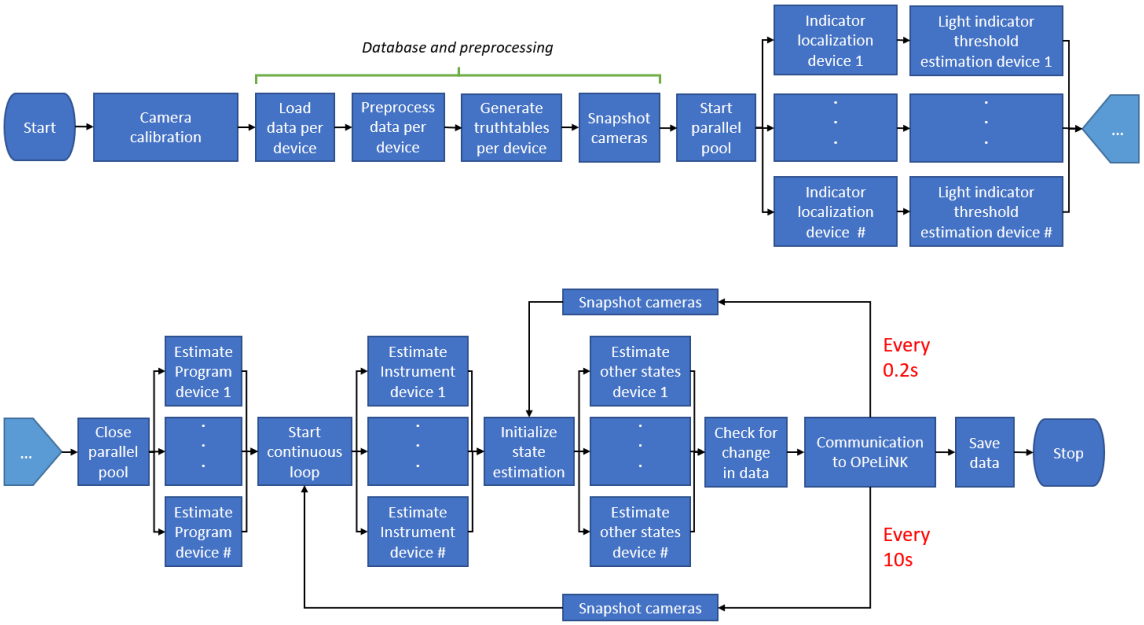


Fig. 9: Flow diagram showing the process of the main program in the intraoperative stage. Just before the surgery starts, several steps from calibrating the cameras and selecting the energy devices, to perspective, threshold and the estimation of the Program state, are performed. During the surgery, every 10s the Instruments are recognized, and every 0.2s the other states are recognized and communicated to the OPeLiNK system.

Camera calibration

Before the surgery starts, the cameras are connected to the computer that runs the main program and the program is started. Then the program shows the view of each camera. Based on this, the user can adjust the position of each camera, such that a proper view of the display of the energy device can be realized. After this, the program asks which machine is within the view of the camera. Based on these choices, the program will change the settings based on the stored data related to the machine of each camera in terms of resolution, backlight compensation, exposure, focus, and white balance. The parameters are chosen for each machine based on trial and error. In Figure 10 an example of the view of two cameras is shown.

(a) (b)

Fig. 10: The view of two connected cameras with one viewing the display of (a) the ETHICON Harmonic & EnSeal, and the other viewing the display of (b) the Conmed System 2450.

Database and preprocessing

After that the cameras have been initialized and the energy devices that are going to be used during the surgery have been selected, the program starts loading the specific data of the energy devices from the database. This data consists of the reference images for indicator localization, the template images of each indicator, the data of each indicator on the locations with respect to the reference frame, the type of indicator for each state, and data on the interrelationship between all the states, as described by the corresponding state tree. Then for each state the images are preprocessed according to the indicator recognition method of the specific type of indicator of the state. After the data has been preprocessed, a snapshot is taken with each camera for indicator localization and initial state estimation.

Indicator localization

Next, the indicators are located of each energy device. Since the cameras are attached and detached every surgery, the perspectives of the cameras change. For this, an indicator localization method has been developed which is discussed in the previous paper in this thesis. This method transforms the camera images to align them with the corresponding reference image. After this, the prestored locations of the indicators can be applied to locate the indicators.

Light indicator threshold estimation
After the indicators have been located, the thresholds for the light indicator recognition need to be estimated. Since the positions of the energy devices and the lighting vary for each surgery, the thresholds of the light indicators need to be adjusted for this. This threshold estimation method is discussed in detail in the previous paper in this thesis.

State estimation
After the thresholds have been estimated and everything has been setup, the first states are estimated of the energy device before the surgery starts. These consist of the Program and the Instrument. As mentioned in the reading strategy, the Program does not change and must be read only ones. The Instrument can change during the surgery and is estimated every 10s in the second phase, after the initial estimation in the first phase. In the second phase, when the surgery starts, all other states are estimated every 0.2s and communicated to the OPeLiNK system. Every 0.2s a snapshot is taken of the display with each camera. Then the image of each camera is transformed according to the corresponding perspective transformation matrix. After this, the indicators are cropped out of the larger transformed image, and their states are evaluated. The order of the recognition is according to the reading strategy. The state recognized in each layer is stored in a truth table connected to the layer, specific to each energy device. The resulting state of the layer will determine which options for the state in the next layer becomes available for evaluation. Through this, the number of states that need to be evaluated is minimized, reducing the reading time and increasing the accuracy of the system. The types of indicators that need to be recognized consist of light indicators, text indicators, number indicators, and symbol indicators. All indicators are recognized through some form of template matching. For light indicators a novel light indicator recognition method based on the relative position and luminance values is used. For text indicators, number indicators and symbol indicators, an existing method based on Normalized Cross Correlation (NCC) is used. Both methods are discussed in more detail in the previous paper in this thesis.

Communication to OPeLiNK
When the measuring system sends information to the OPeLiNK system, this is stored in the database of the OPeLiNK system. To prevent the generation of large unclear datasets, the information is sent only when there is a change in state of an energy device.

Speed optimization
Next to applying the reading strategy, several techniques have been applied to optimize the speed of the system. Almost all data is converted to single precision at the phase when the variables are initialized. Standard MATLAB functions have been replaced by more efficient altered versions. To reduce

the time, to estimate the perspective transformation of each machine, parallel computing in the form of a parallel for loop is applied. For the other functions in real time, regular for loops were found to be more efficient when dealing with multiple machines and cameras. To reduce the computation time of the real time perspective transformation, the corresponding function has been converted from MATLAB code to a MEX-function containing CUDA code, through the MATLAB GPU coder [16]. This allows for fast parallel calculations on the GPU instead of the CPU. For the indicator recognition functions, both functions were converted to a MEX-function containing C++ code to optimize the calculation speed through the CPU. This was done with help of the MATLAB coder [17].

2.3.3 Camera setup

The camera setup consists of multiple cameras that can be attached to each energy device through a clamp with a flexible arm. In Figure 11 shows the camera setup used for both the Conmed and the Harmonic. ELP-USB4K02AF-V100 cameras were used for the system. The housing is printed from AR-M2 resin with the Keyence Agilista-3200 inkjet 3d printer [18]. The cameras have a size of 42mm x 42mm x 36mm with a resolution ranging from 1280 x 720p to 3840 x 2160p, and a frame rate of 30fps. The lens has a 94.5° field of view (FoV), minimal distortion, and built-in Infrared (IR) filter [19].

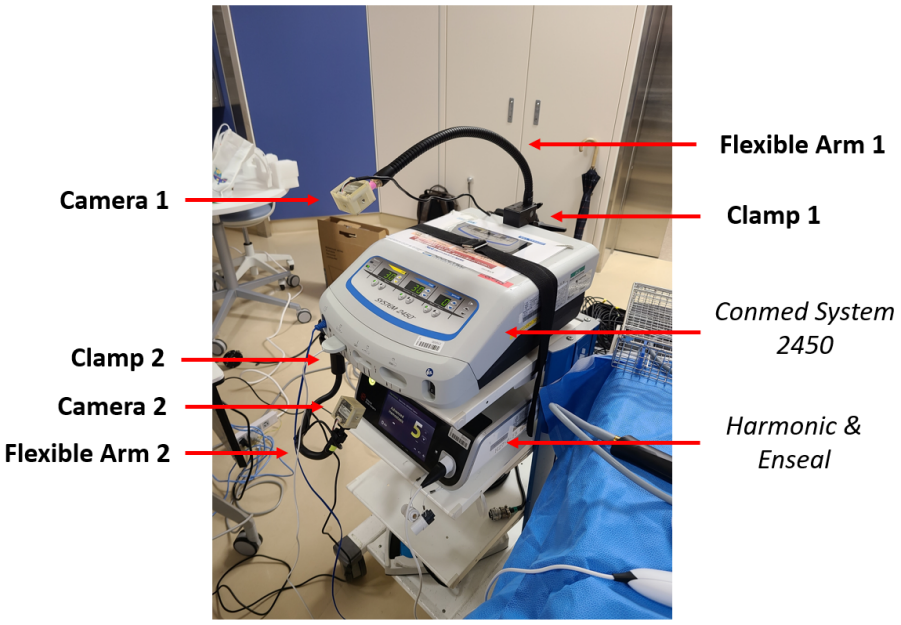


Fig. 11: The camera setup of the measuring system. In this situation 2 cameras are used to record the Conmed System 2450 and the Ethicon Harmonic & EnSeal. Each camera is attached to a flexible arm, which is connected to the corresponding energy device through a clamp.

2.3.4 The OPeLiNK system: the decision-making navigation screen

The OPeLiNK system uses the decision-making navigation screen to communicate the information of all surgical equipment to the surgeon. The OPeLiNK system runs on a SQL server that continuously gathers information of the different surgical devices connected to the OPeLiNK system. Regarding the energy devices, there are several ports reserved for them. The measurement system is connected to the OPeLiNK system through ethernet cables. The communication is performed through the Transmission Control Protocol (TCP) [20]. When there is a change in the state of any energy device, the measurement system sends a string to the appropriate port consisting of the Machine name, Power status, Program name, Instrument name, Major mode, Minor mode, Power level, and Error message. This information is then shown to the surgeon and remote expert through the decision-making navigation screen. In Figure 12 you can see an example of the output of the measurement system together with additional video data of a camera and an endoscope. In the square for the results of the measuring system, the name of one energy device is shown at the top. When the energy device is not activated, the screen will show this by the text OFF in the top left corner, together with no information on the rest of the states. Then when the system is activated, the screen will show all other states of the energy device. The program is displayed in the middle. The Instrument, Major mode, and Minor mode are displayed in the top right corner. The Power level is displayed at the bottom under the OUTPUT section. To view the state of any other energy device, the user can right-click on the screen and select the energy device from the list.

Fig. 12: Example of the output of the measuring system displayed on the decision-making navigation screen. On the left the output video data of a connected camera is shown. In the middle the output video data of a connected endoscope is shown. On the right, in the middle window, the output of the measurement system is shown on the states of an energy device.

3 Results

The system is tested on two types of energy devices. The first energy device is the electrosurgical tool, the Conmed System 2450, from here on referred to as the Conmed [8]. The second energy device is the harmonic scalpel, the ETHICON Harmonic & EnSeal, from here on referred to as the Harmonic [14]. Both devices are shown in Figure 13. Several experiments are performed that measure the performance of the system in terms of accuracy and speed in a real operating environment. Furthermore, the system is validated through an experiment that compares the system to the state-of-the-art. Finally, the system is integrated into the OPeLiNK system and evaluated on its ability to communicate the states of any energy device in real time.

(a) (b)

Fig. 13: Images of (a) the Conmed System 2450, and (b) the ETHICON Harmonic & EnSeal

3.1 Experiment 1: The accuracy of the system

Purpose

The measurement system needs to have a minimum accuracy of 90% to be eligible for integration in the OPeLiNK system and to safely communicate the information to the surgeon and remote experts. To evaluate the system on its accuracy for multiple energy devices an experiment was performed to measure the accuracy of the total system and the individual indicator localization and recognition methods separately.

Method

In the previous research the accuracy of the system was evaluated to be 98.2% for the Conmed, under the condition that the display of the energy device is within view of the camera. To further evaluate the system for other energy devices, a similar approach was done. The setup is slightly different then the final use. For the evaluation of the accuracy, the same setup was used as in the previous article. Figure 14 shows the layout of the setup. This consists

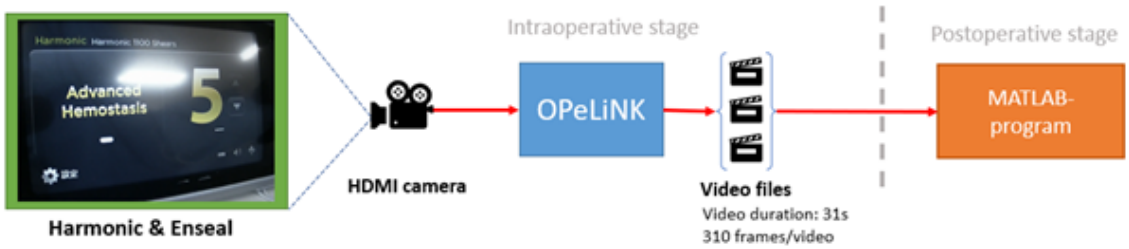


Fig. 14: Example of the output of the measuring system displayed on the decision-making navigation screen.

of a camera filming the display of the Harmonic, which is connected to the OPeLiNK system. This then creates multiple video files, each consisting of 930 frames with 31s duration. The recording is done in the intraoperative stage, while the state evaluation is done postoperatively through the Main program. The Main program was tested on 520 videos consisting of two colon surgeries performed at Japan's National Cancer Center. The average operating time was three hours. Data acquisition is conducted with the approval of the ethic committee of Japan's National Cancer Center and the University of Tokyo.

The accuracy of the system Is measured by postoperatively evaluating each video file and manually comparing it to the results that are given by the measurement system. If there is at least one error in the video file, the video is counted as an error. The formula for the accuracy is described by

$$accuracy = \frac{N_{video\,files} - N_{errors}}{N_{video\,files}} \cdot 100\% \tag{1}$$

where $N_{video\,files}$ indicates the total number of evaluated video files and N_{errors} indicates the total number of video files with a detected error.

Results

From the 520 videos in total two errors were detected. This means that $N_{video\,files} = 520$ and the number of detected errors is $N_{errors} = 2$. The resulting accuracy is 99.6%. The accuracy of both the indicator localization method and the light indicator recognition method was 100%. The other recognition method for text, numbers and symbols does have two errors in the estimation one specific Major mode, the Advanced Hemostasis. Specifically with these two errors, a large specular reflection can be seen on the text indicator, as shown in Figure 15. This results in the text indicator recognition method not meeting it's threshold, and thus not being able to read the status of the energy device. This error does not reflect the performance of the recognition method, since the input data is not even legible for a human. Therefore, if we would exclude these two errors, the other recognition methods would also have a performance of 100% accuracy. Figure 16 and Figure 17 show an example of the measured activations during a surgery performed by a resident and a skilled surgeon, respectively. In these datasets, the two identified errors have been removed.

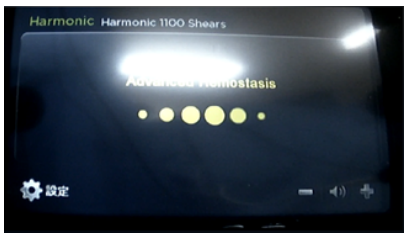


Fig. 15: Example of the error in recognizing the Advanced Hemostasis, due to a large specular reflection.

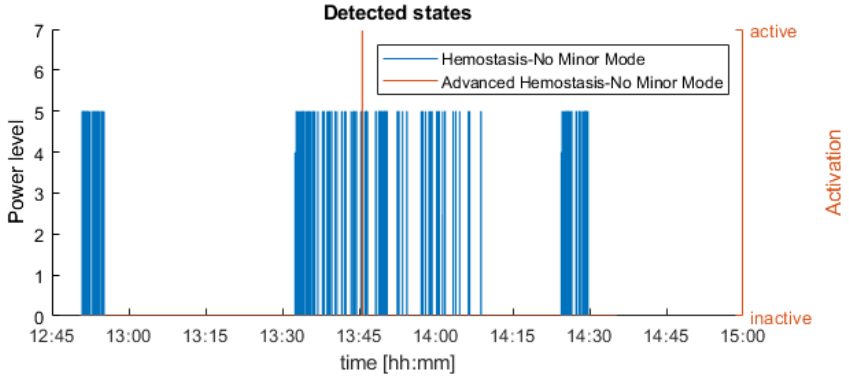


Fig. 16: Example of the measured activation data during a surgery performed by a resident. The left y-axis indicates the power level of the major mode when activated. The right y-axis indicates the activation of the Advanced Hemostasis. This activation has no power level. The x-axis indicates the time during the surgery. The instrument used was the Harmonic 1100 Shears.

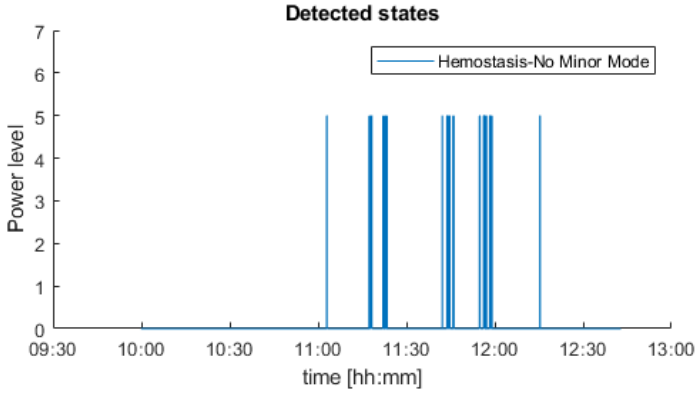


Fig. 17: Example of the measured activation data during a surgery performed by a skilled surgeon. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was the Harmonic 1100 Shears. There was no Advanced Hemostasis measured due to the specular reflection error.

3.2 Experiment 2: The recognition speed of the system

Purpose

In the second experiment the recognition speed of the system is evaluated to determine its real time performance in terms of state recognition.

Method

In this experiment, the Main program has been slightly altered by giving video files of the Conmed as input instead of the cameras. The cameras are still turned on and take pictures, but the images of the video files are used to measure the states. These video files consist each of 930 frames, with a duration of 31s. To evaluate the system on its recognition speed, the system is tested on two types of video files of the Conmed. The first video consists of the Conmed being turned on and displaying its instrument, but the power is never activated. This will be referred to as an inactive machine. The second video consists of the Conmed being turned on and having the power being activated in the form of cutting and coagulating. This will be referred to as an active machine. These two video files are used to simulate the real operating environment where up to five energy devices are used during a surgery. The recognition time is measured through a timer function and stored together with the measured states of the energy device.

Results

The system is tested for the scenarios of one to five active energy devices having their power activated at the same time, combined with zero inactive machines. Furthermore, the system is tested on the heaviest realistic scenario of two active machines used at the same time, in combination with three inactive machines. The time it takes to read the states consists of the full process of taking the snapshot with the camera, to the reading and storing the states of the energy devices. This does however exclude the communication time between the Main program and the OPeLiNK system. The results are shown in Table 1.

Table 1: Average recognition speed for different number of activate and inactive machines

Number of active machines	Number of inactive machines	Average inactive measuring time [s]	Average active measuring time [s]
1	0	0.025	0.037
2	0	0.048	0.075
3	0	0.065	0.105
4	0	0.082	0.133
5	0	0.102	0.166
2	3	0.103	0.130

For the performance of the Main program also the individual average time for the main functions is evaluated. These consist of the snapshot time of the camera, the transformation time of the perspective transformation, and the active state estimation time. The results can be seen in Table 2. The results are based on the evaluation of one active machine. With this, the performance of each function for multiple machines can be estimated by multiplying the average function time with the number of machines.

Table 2: Processing time of the main functions of the Main program

Function machines	Recognition time [s] for 1 active machine	Recognition time [s] for 5 active machines
Snapshot camera	0.004	0.02
Perspective transformation	0.02	0.10
State estimation	0.01	0.05
Total program	0.034	0.17

3.3 Experiment 3: Comparison to the state-of-the-art

Purpose

In this experiment the system is evaluated based on its speed and accuracy compared to the state-of-the-art.

Method

An illustration of the experimental setup is shown in Figure 18, and an image of the setup can be seen in Figure 19. For this, a system has been developed that is similar to the state of art proposed by Annetje C. P. Guédon et al [4]. Instead of a measuring box between the socket and the energy device, an oscilloscope with a current probe around the cable that is connected to the socket was used. The oscilloscope is connected to a computer with a program that evaluates the collected current data. The current probe is a URD Co. High frequency CT current sensor [21]. The oscilloscope is the Tektronix TBS 2102 digital oscilloscope [22]. When the measured current is above a certain threshold, this is indicated as the power being activated. When there is a change in the state of the energy device, this information is sent to a database together with a time stamp of the computer running the program. Our developed system consists of the camera gathering visual data of the display and communicated to the main program. The program uses image recognition to estimate the states. When there is a change in the state of the energy device, this information is stored in the database, together with a time stamp of the computer running the program.

To evaluate our developed system, both measurement setups are set to measure the Conmed. Both recognition programs are run on the same computer in parallel, to ensure some form of time synchronization with regards to the time stamp that is communicated to the database. The measuring frequency of the current probe program is set to 10Hz, according to the state-of-the-art. While the measuring frequency of our system is set to 5Hz, according to the designed limit of the system.

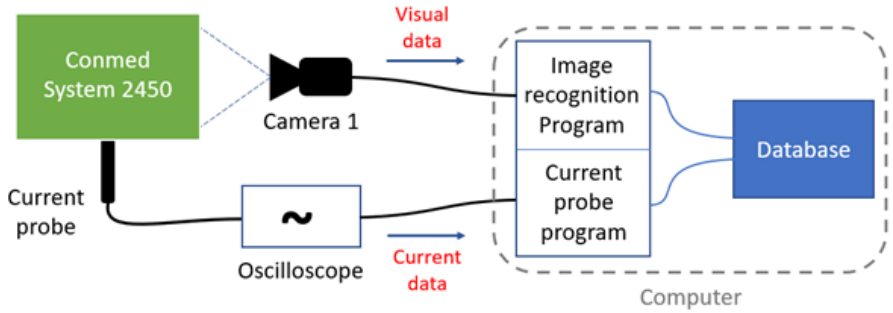


Fig. 18: Illustration of the setup for the comparison between the developed image recognition system and the state-of-the-art in the form of a current probe system. Both systems are used to measure states of the Conmed System 2450.

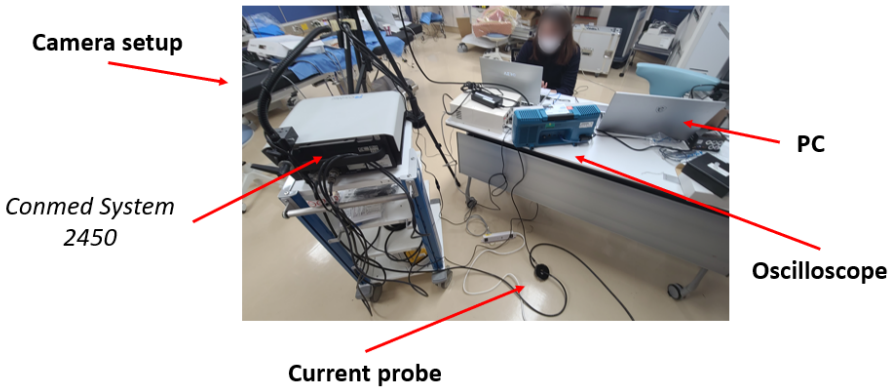


Fig. 19: Image of the setup for the third experiment. Both the image recognition system and the current probe system measure the states of the Conmed System 2450.

Results

The experiment duration was 16 minutes and 41 seconds. This means that in total 5005 frames have been evaluated by the image recognition system. The results show a 100% accuracy for the recognition of the activation of the Conmed for both measuring systems. Regarding the measurement of the other states, in 11 frames an error was detected. This error consisted of the image recognition system temporarily not being able to read the Minor Mode of the Conmed. This results in a 99.8% accuracy in recognizing the other states, that are not measured by the current probe setup. The delay between the two systems showed a mean delay of 0.07s. In Figure 20 an example of the measured data is shown. Both systems were also tested on the Harmonic, but it was found that the current probe cannot read information on the Harmonic, since there are no significant peaks that can be measured during activation. In Figure 21 an example is shown of measuring the activation of both the Conmed and the Harmonic. It can be seen that for the Conmed a clear peak during activation is visible, while during multiple activations of the Harmonic, no clear peak can be found.

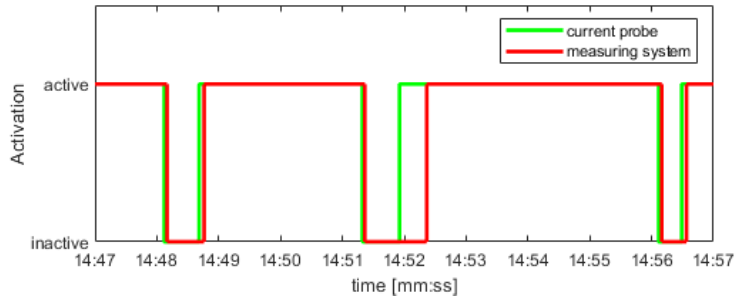
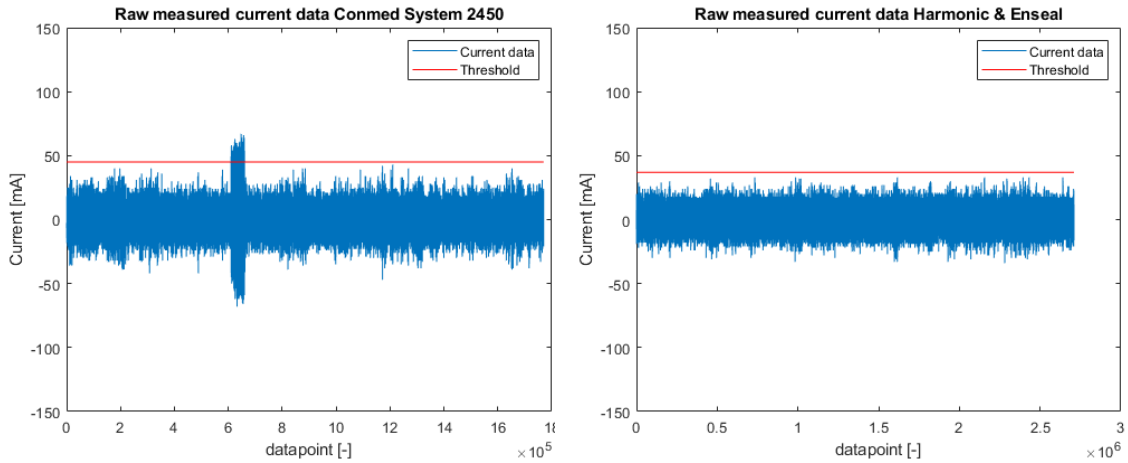


Fig. 20: Example of the measured activation data of the Conmed System 2450 for both the current probe and the measurement system. The y-axis indicates the activation of the power, and the x-axis indicates the time in minutes and seconds.



(a) (b)

Fig. 21: Example of the measured current data of (a) the Conmed System 2450, and (b) the Ethicon Harmonic & EnSeal. When the power of the energy device is activated, the current exceeds should exceed the set threshold. For the Conmed one activation was performed. For the Harmonic, the power is activated several times, but no clear peaks are perceived that exceed the threshold. Due to this, the power activation could not be measured for the Harmonic.

3.4 Experiment 4: Implementation in the OPeLiNK system

Purpose

In the fourth experiment the system is evaluated on real time communication while it is integrated in the OPeLiNK system.

Method

Figure 22 shows an image of the experimental setup setup. The setup consists of two cameras of the measurement system gathering visual data of the Conmed. These are connected to a computer with the Main program running on it, which in turn is connected to another computer running the OPeLiNK system. To compare the communication time of our system to other equipment

integrated into the OPeLiNK system, an endoscope and a HDMI camera are directly connected to the OPeLiNK system. All cameras give visual data from the display of the Conmed. The information of each system is communicated through the decision making navigation screen. The displayed information is recorded by the computer using screen recording. The video data created from this screen recording is given as input to an evaluation program in MATLAB. With this program, the recognized activations and deactivations of the Conmed are detected for the camera, the endoscope and the output of the measuring system.

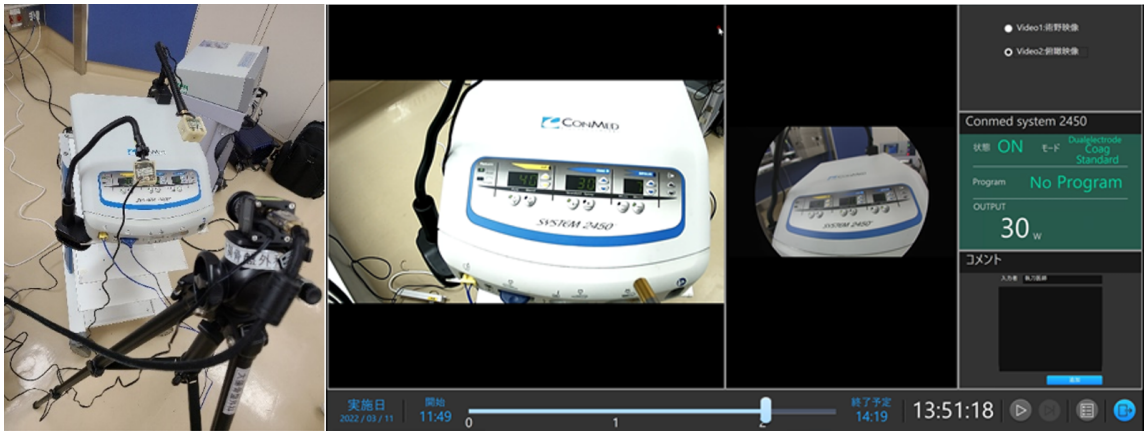


Fig. 22: Image of the setup for the fourth experiment. The image recognition system consisting of two cameras estimates the state of the Conmed System 2450. An additional camera and an endoscope provide extra visual data on the display of the Conmed. All results are shown on the decision-making navigation screen.

Results

In Figure 23 an example of the resulting data is shown. The data consists of a recording of 120s duration. Within this time, 38 changes in the state were performed, consisting of 19 activations and 19 deactivations. In Figure 24 the box plot of the delay between the video camera and the endoscope, the delay between the video camera and the measuring system, and the delay between endoscope and the measuring system is shown. The median for the delay between the video and the endoscope data is 0.077s, and the mean is 0.059s. The median for the delay between the video and measuring system data is 0.308s, and the mean is 0.373s. The median for the delay between the endoscope and the measuring system data is 0.231s, and the mean is 0.314s.

To evaluate what the delay regarding the measurement system consists of, the average measuring time during the activations were evaluated. This consists of the total time to make a snapshot with the two cameras, the perspective transformation, state estimation, and the sending of the data to the OPeLiNK system. The mean of the measuring time during activations is 0.11s and the median is 0.10s. Based on this the communication time can be estimated to have a mean of $0.375s - 0.11s = 0.265s$.

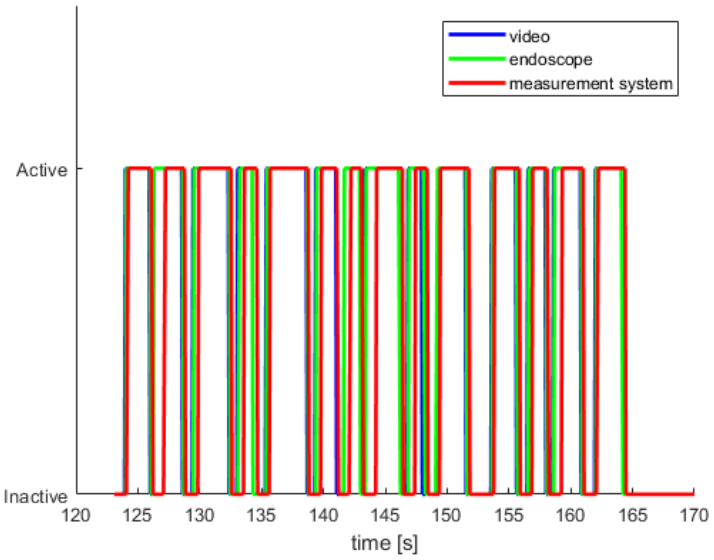


Fig. 23: Example of the measured activations and deactivations over time.

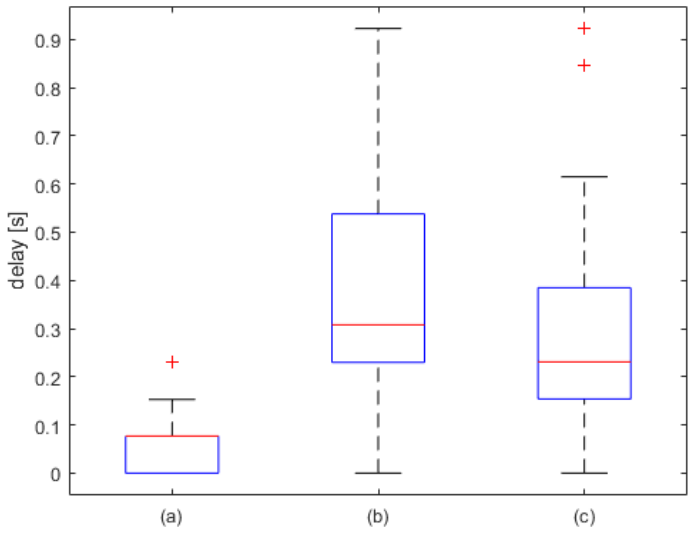


Fig. 24: Delay between (a) video vs endoscope, (b) video vs measurement system, and (c) endoscope vs measurement system.

3.5 Experiment 5: Communication speed limit of the OPeLiNK system

Purpose

In this final experiment the OPeLiNK system is evaluated on its limit in terms of communication speed. This experiment does not involve the measurement system, but rather evaluates the communication limitations of solely the OPeLiNK system.

Method

For this experiment, a MATLAB program has been created that simulates the output of the measurement system. The program is run on the computer of the measurement system and sends two types of messages alternately at a certain frequency. Both messages consist of simulated output data on the states of an energy device with one of the states being replaced with a timestamp. One message indicates that the energy device is activated, the other indicates that the energy device is deactivated. The OPeLiNK system is evaluated on its ability to show the information correctly at different frequencies ranging from 0.1Hz to 10Hz. If the OPeLiNK system can handle the frequency, the output on the decision making navigation screen should display the messages at approximately these respective frequencies.

Results

The results show that the OPeLiNK system can display information up to a frequency of 1Hz. Beyond this frequency, messages are delayed or missing on the decision making navigation screen.

4 Discussion

The results of the previous paper together with the results of the first experiment in this paper showed an accuracy of 98.2% for the postsurgery state estimation of the Conmed and a 99.6% accuracy for the postsurgery state estimation of the Harmonic. Both accuracies exceed the minimum required accuracy of 90%. The errors in the state estimation of the Harmonic were of a specific type where strong reflections were visible. These reflections resulted in the loss of information on the Major mode, Advanced Hemostasis. To improve the accuracy of the system, these reflections need to be filtered or avoided. To solve this, a polarizing filter was applied to try to filter out the reflections on the Harmonic. This however did not work due to either the reflections being not polarized because of the type of material of the screen, or the reflection angle not being within the range that can be filtered. Another option would be to apply some type of noise handling method that filters and compensates for these reflections. This however is only possible if the reflections do not completely saturate the pixels. Furthermore, this would increase the recognition time of the system. The final option to solve this, would be to change the camera angle slightly downwards. This would ensure that no strong reflection of the operating lights would appear.

For the second experiment the system showed a total recognition speed that varied for the number of machines that are actively used. All combinations showed a recognition time below the requirement of 0.2s. This however excluded the communication time between the OPeLiNK system and the measuring system. In that case, using five machines actively at the same time might take more time to communicate. This was tested for the Conmed video since this system has more states to be evaluated than the Harmonic. Other systems

with more states could have a bit larger recognition time. This effect is however limited by the reading strategy with the state options being dependent on the recognized state in the previous layer. In terms of how to further optimize the speed of the program, the most efficient way would be by improving the speed of the perspective transformation since this takes the most time. Currently this function consists of a MATLAB function that was converted to a MEX-function containing CUDA code. Although the speed is higher this way, it still takes more time for MATLAB to call this function due to it being in a different language. By changing the programming language from MATLAB to some other function, this communication could be optimized. In general, MATLAB is quite a slow environment. The whole system could for example benefit from rewriting and compiling the code in C++. This in general would result in a higher performance in terms of speed [23].

For the third experiment, the measuring system was compared to the performance of the state-of-the-art, in the form of a current probe. The results showed that the state-of-the-art has a 100% accuracy in terms of measuring the activation. If we compare this to the accuracy of our measurement system, we must isolate the accuracy of our system with regards to purely the power activation state. In terms of measuring the activation, our system also reaches a 100% accuracy. However, if we look at the accuracy with regards to measuring the other states, the accuracy becomes 99.8% when measuring all other states of the energy device. The delay between the two systems is an average of 0.07s. This is mainly due to the difference in measuring frequency. In theory the measuring frequency can significantly be increased for the state-of-the-art, which could increase this delay. With regards to trying to measure the Harmonic, it was perceived that the current probe setup could not measure the activation of this energy device. This probably has to do with the type of cable that is connected to the socket in combination with the power level. The maximum power level of the Harmonic is only 5W, compared to the Conmed that regularly uses around 30W. A lower power level decreases the height of the peaks when the energy device is activated. The inability to read the Harmonic shows that the reading capability of the state-of-the-art is limited to electrosurgical and electrocautery tools, while our developed system can be used for any energy device.

For the fourth experiment, the measuring system was integrated into the OPeLiNK system, and evaluated on it's communication speed. The results showed a significant delay between the communicated data of the measurement system compared to the video data and the endoscope data. When looking more detailed at what the delay consists of, it was perceived that the measuring time was an average of 0.11s. This indicates that the rest of the delay, purely consists of the communication time from when the measurement system has sent the data, up to when the OPeLiNK system shows the data on its decision making navigation screen. It can be seen in Figure 23 of the measured activations, that the delay becomes larger when there is a relative short time between two activations. This indicates that if the time between two messages

is too short, it can cause significant delay when displaying the information on the decision making navigation screen. This can be due to multiple factors such as the reading frequency of the SQL server with regards to the messages send by the measurement system, the programmed frequency of displaying the information, and the communication speed of the TCP model. Based on the measured data of the surgeries in the first experiment, it was perceived that the activation duration is no shorter than 0.3s. This means that the expected minimum time between two messages send by the measuring system would be 0.3s.

In the final experiment, the limit of the OPeLiNK system with regards to the communication speed was tested. This showed that the limit is a communication frequency of 1Hz, or a communication period shorter than 1s. This confirms the suspicion in the fourth experiment, that the communication method with the OPeLiNK system causes a significant delay on the measured data. When there would be a quick change in the state of the energy device, the measurement system would still be able to read it with a minimum time between two messages of 0.3s, so a maximum communication frequency of around 3Hz. This however can currently not be handled by the OPeLiNK system. To further reduce the delay of the total system, it is necessary to change the settings of the OPeLiNK system such that the reading frequency matches the maximum communication speed of the measuring system of 0.2s, or 5Hz. Through this, it can be guaranteed that the overall system can handle the minimum time between two messages of 0.3s.

The reading strategy in combination with the communication protocol creates a dynamic measuring strategy that can be applied on any energy device with any display. Due to the registered relationship between the states, even if a state like a Program, or an Instrument completely changes the layout of the display, the system still can locate and identify the indicators of the children states in the next layers. Furthermore, because of this dependence, the reading speed is less influenced by the number of states that the energy device has, and is significantly decreased for all energy devices. The number of states that need to be evaluated are decreased from for example 30 states for one layer to three states for one layer. This significantly reduces the reading time, and increases the accuracy of the system.

The setup program is used to register any energy device to create the database with the template images and to register the relationship between all the states. The program has several options to deal with different situations where either a mistake is made by the user in registration, or the video data has a different perspective. Before the program is used to register a new device, it is advised to already have drawn out the state tree. This way, it is easier to recreate this tree in the program. This does, however require a thorough understanding of how the machine works, how it displays its states, and what the relationship between the states is. Depending on the machine, the adding of states and images can be quite time consuming. The registration can take from 30 minutes to 1.5 hours if there are many states. Currently there is no

indication through the GUI of which template images you already have added to the database. This can only be seen by accessing the database folder.

The measurement system uses a camera setup that is connected through USB-cables to a computer which is connected to the OPeLiNK system through Ethernet. Because of this, setting up the system is more time consuming than desired. First the cameras are attached to the energy device with a clamp and a flexible arm. Then all cables need to be guided such that it does not hinder the surgeons and assistants. Then the system is started up and the cameras need to be adjusted such that they have a clear view of the displays of the energy devices. All these processes take up quite some time, and should be made easier in the future such that there is no added complexity in the operating room and that the system does not add any stress in a stressful environment.

When using image recognition to recognize the state of the energy device through its display, the output is limited. Although all these energy devices do give some information on the type of instrument, the information lacks important details. The instrument tip for example, is not registered by these devices. This is however important since the size of the electrode influences the current density in electrosurgery. A smaller electrode would have a higher current density and thus a more concentrated heating effect on the tissue.

5 Conclusion

In this paper a new manufacturer independent method is introduced for the real-time state communication of surgical energy devices. For this, a common state communication protocol has been developed that describes the states of any energy device. Based on this communication protocol a reading strategy was developed, where each state option is dependent on the result of the previous layer. This reduces the number of state options and thereby improves the accuracy and the recognition time. With the communication protocol and reading strategy the measuring system was developed. The measuring system uses image recognition in the form of template matching to determine the state of any energy device. For the registration of a new energy device in the preoperative stage, a setup program was developed. This program is used to create a template database with a state tree describing the relationship between the states of an energy device. In the intraoperative stage the main program is used to read out the states of one up to several energy devices in real time.

The developed system has been evaluated based on five experiments. The first experiment showed a 98.2% accuracy for the recognition of the Conmed, and a 99.6% accuracy for the recognition of the Harmonic. The system therefore meets the requirements of minimum accuracy of 90%.

The second experiment showed a mean recognition time of 0.037s for measuring one active energy device, up to a mean recognition time of 0.166s for five active energy devices. The recognition time therefore fits within the requirement of a maximum recognition time of 0.2s.

In the third experiment the system is compared to the state-of-the-art, that consists of a measuring system based on a current probe. The results show that both the image recognition system and the state-of-the-art have a 100% accuracy in terms of recognizing the activation of the Conmed. The image recognition system, however, also gave information on the other states communicated on the display of the energy device with an accuracy of 99.8%. The measuring frequency of the image recognition system has a maximum frequency of 5Hz. The state-of-the-art has a measuring frequency set to 10Hz but can be set even higher. The delay between our developed measuring system and the state-of-the-art is 0.07s. Therefore, in terms of recognition time, the state-of-the-art outperforms the image recognition system. Although the state-of-the-art can recognize the activations faster, it gives no information on the other states of the energy device. Furthermore, the state-of-the-art is limited to recognizing the activations of electrosurgical tools. Other energy devices like the Harmonic could not be measured.

In the fourth and fifth experiment the developed measuring system was integrated in the OPeLiNK system and evaluated based on the delay in communication time between, general video data, endoscopic data and the data communicated by the measuring system through the decision making navigation screen. This showed that there is a significant delay with a median of 0.308s and an average mean measuring time of 0.373s. This delay was found to be largely due to the communication speed between the measuring system and the decision making navigation screen of the OPeLiNK system. The measuring system has a mean measuring time of 0.11s and sends information every 0.2s if there is a change in the state of the energy device. The fifth experiment showed that the OPeLiNK system cannot handle messages send with a higher frequency than 1Hz, or one message per second. This confirmed that there is a significant delay in sending and displaying messages due to the limit of the communication with the OPeLiNK system.

All these experiments show that the measurement system itself meets the set requirements in terms of accuracy and recognition speed. Furthermore, it showed that the developed system can compete with the state-of-the-art in terms of communicating more information of multiple types of energy devices at the cost of a slight delay, in real time. Currently the bottle neck is the communication with the OPeLiNK system. To improve this, either the TCP model needs to be changed, or the OPeLiNK system needs to be updated such that it can handle more messages in a shorter time. Other steps would be improving the system in it's accuracy and recognition speed.

Finally, only detecting visual information does have limitations in the robustness of the measurement system. This could be overcome by either adding auditory data or other data on the current and voltage of the system, or a combination of these.

The next step would be to apply this measuring system to try to quantify the skill of expert surgeons, for future semi-automated surgery.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Ethics approval Data acquisition is conducted with the approval of the ethic committee of Japan's National Cancer Center and the University of Tokyo.

Funding This study is supported by AMED (grant No. 21he2102001h1003).

References

[1] Okamoto, J., Masamune, K., Iseki, H., Muragaki, Y.: Development concepts of a smart cyber operating theater (scot) using orin technology. Magazine **63**(1), 31–37 (2018). https://doi.org/10.1515/bmt-2017-0006

[2] Bitterman, N.: Technologies and solutions for data display in the operating room. Magazine **20**(3), 165–73 (2006). https://doi.org/10.1007/s10877-006-9017-0

[3] Ogiwara, T., Goto, T., Fujii, Y., Nakamura, T., Suzuki, Y., Hanaoka, Y., Ito, K., Horiuchi, T., Hongo, K.: Endoscopic endonasal approach in the smart cyber operating theater (scot): Preliminary clinical application. Magazine **147**, 533–537 (2021). https://doi.org/10.1016/j.wneu.2020.12.114

[4] Meeuwsen, F.C., Guédon, A.C.P., Arkenbout, E.A., van der Elst, M., Dankelman, J., van den Dobbelsteen, J.J.: The art of electrosurgery: Trainees and experts. Magazine **24**(4), 373–378 (2017). https://doi.org/10.1177/1553350617705207

[5] Dums, J., Schneider, B., Badin, A.: Low cost system to measure active power in electrosurgical units. Magazine **33** (2017). https://doi.org/10.1590/2446-4740.03217

[6] Ushimaru, Y., Takahashi, T., Souma, Y., Yanagimoto, Y., Nagase, H., Tanaka, K., Miyazaki, Y., Makino, T., Kurokawa, Y., Yamasaki, M., Mori, M., Doki, Y., Nakajima, K.: Innovation in surgery/operating room driven by internet of things on medical devices **33**(10), 3469–3477 (2019). https://doi.org/10.1007/s00464-018-06651-4

[7] Geirhos, R., Janssen, D.H.J., Schütt, H.H., Rauber, J., Bethge, M., Wichmann, F.: Comparing deep neural networks against humans: object recognition when the signal gets weaker. ArXiv **abs/1706.06969** (2017)

[8] Corporation, C.: Conmed System 2450 Service Manual. https://www.conmed.com/en/products/orthopedics/total-joint-replacement/electrosurgical-generators/system-2450-electrosurgical-generator-esu (2013)

[9] Covidien: Covidien Valleylab ForceTriad service manual. https://www.medtronic.com/animal-health/en-us/products/electrosurgical-hardware/forcetriad-energy-platform.html (2013)

[10] Covidien: Covidien Valleylab FT10 service manual. https://www.medtronic.com/covidien/en-us/products/electrosurgical-hardware/valleylab-ft10-energy-platform.html (2016)

[11] ERBE: ERBE VIO300D Service manual. https://us.erbe-med.com/us-en/products/electrosurgery/vior-300-d/ (2019)

[12] ERBE: ERBE VIO3 Service manual. https://vio.erbe-med.com/ (2019)

[13] OLYMPUS: OLYMPUS ESG-400 System Service Manual. https://medical.olympusamerica.com/products/thunderbeat (2013)

[14] ETHICON: ETHICON Harmonic & EnSeal. https://www.ethicon.com/emea/epc/search/platform/energy%20sealing%20and%20dissecting?lang=en-default (2016)

[15] Tinevez, J.-Y.: Tree data structure as a MATLAB class. https://github.com/tinevez/matlab-tree) (2020)

[16] MathWorks: Generating CUDA Code from MATLAB: Accelerating Embedded Vision and Deep Learning Algorithms on GPUs. https://nl.mathworks.com/products/gpu-coder.html (2019)

[17] MathWorks: MATLAB Coder. https://nl.mathworks.com/help/coder/ (2021)

[18] Keyence: Keyence Agalista 3D printer. https://www.keyence.co.jp/ss/products/3d-printers/agilista/ (2012)

[19] ELP: ELP-USB4K02AF-V100. http://www.elpcctv.com (2021)

[20] MATLAB: TCP/IP Communication. https://nl.mathworks.com/help/matlab/tcpip-communication.html (2022)

[21] Co., U.: URD CT Standard CT only for high frequency -2kHz - 100MH. https://www.u-rd.com/english/products/CTL-28-S90-05Z-1R1.html (2022)

[22] Tektronix: Tektronix TBS2000 Series. https://www.tek.com/en/oscilloscope/tbs2000-digital-storage-oscilloscope-manual/tbs2000-series (2016)

[23] Andrews, T.: Computation time comparison between matlab and c++ using launch windows (2012)

4 Discussion and Conclusion

The main aim of this thesis is the development of a manufacturer-independent measuring system that can estimate and communicate the state of any surgical energy device in real time, such that it can be implemented into the integrated operating rooms. This thesis proposed a new method based on image recognition and was integrated into the OPeLiNK system of the Smart Cyber Operating Theatre, SCOT.

The literature review in Appendix A was performed to determine the state-of-the-art in image recognition for different types of indicators. The results showed that indicator recognition is largely researched and many methods for each indicator already exist. The gap is however that each indicator recognition method is designed for the recognition of one type of indicator while for automation multiple types of indicators are used. Regarding light indicators, there is less research done within this field. Currently light indicators are mainly evaluated based on their color distribution, but not on their relative position to each other or their luminance value. Currently machine learning methods are the most heavily researched and have the highest potential in terms of robustness. They do however require sufficient datasets, that might not be available when entering a new application. Since we do not have sufficient datasets within our application, we will have to combine multiple nonmachine learning methods to determine the state of each energy device.

The first paper in section 2 showed the development of a measuring system that measures the state of a surgical energy device for postoperative evaluation. The system used a camera to gather visual information of the energy device and communicated this to the OPeLiNK system. This in turn created video files that are evaluated by the image recognition system postoperatively. The system was tested on its limit in terms of camera angle, the delay between the evaluated data and the endoscopic video data, and the measuring accuracy of the system. The results showed that the camera angle needs to be within a maximum of 35° with respect to the normal of the display of the energy device. The average delay between the measured data and the endoscope data is 0.07s. The system has an overall accuracy of 98.2%. The system therefore fits within the requirements of a minimum accuracy of 90%.

The second paper in section 3 showed the further development of the measuring system. Now the system is completely integrated into the OPeLiNK system. It can recognize the state of any energy device in real time through image recognition in the form of template matching, and it can communicate this information to the surgeon and remote experts through the decision-making navigation screen. To be able to recognize any energy device a communication protocol was developed. Then, a reading strategy was developed to increase the accuracy and reduce the measuring time. A setup program was developed to register any surgical energy device in the preoperative stage. In the intraoperative stage, several cameras gather the visual data of the displays of several energy devices. This data is given as input to the developed main program, which recognizes the states of any energy device in real time, and communicates this information to the OPeLiNK system. This data is communicated in real time for decision making and is recorded for postoperative evaluation. Several experiments were performed to evaluate the system. The accuracy of the system is 98.2% for the state estimation of the Conmed System 2450, and 99.7% for the state estimation of the Ethicon Harmonic & EnSeal. The system has an average recog-

nition time of 0.037s for measuring one active energy device, up to an average recognition time of 0.166s for five active energy devices. Both the state-of-the-art and our developed system have a 100% accuracy in the recognition of the activation of an energy devices. Our developed system however, also estimated the other states with an accuracy of 99.8%. The delay between our developed measuring system and the state-of-the-art is 0.07s. Finally, the communication speed of the system was evaluated. The results showed that there is a significant delay in communication of the measured states due to to the limit of the communication with the OPeLiNK system, where the OPeLiNK system cannot handle more than one message per second. The results show that the measuring system meets the set requirements in terms of accuracy and recognition speed. Furthermore, it showed that the developed system can compete with the state-of-the-art in terms of communicating more information of multiple types of energy devices at the cost of a slight delay, in real time. Currently the bottle neck is the communication with the OPeLiNK system. To improve this, either the TCP model needs to be changed, or the OPeLiNK system needs to be updated such that it can handle more messages in a shorter time. Other steps would be improving the system in it's accuracy and recognition speed.

Overall, the system adds significant value to the measurement of energy devices by not only measuring the activation, but also measuring new information that has not been measured before, such as the used Program, Instruments, Major modes, Minor modes, and Power Levels. This information can be very valuable to communicate to the remote experts in the form of telesurgery for better decision making. Furthermore, it largely increases the value of the data for postsurgery evaluation.

The first developed system for the recognition of one energy device was deemed by the assistant to be very easy to apply in the operating room, and did not hinder the surgeon. The limit of this system however was that it could measure the states of only one energy device postoperatively. The second system tackled this limitation, and increased the value of the measurements by measuring multiple energy devices in real time and communicating this to the surgeon and remote experts through the decision-making navigation screen. But to do measurement in real time of multiple devices, the complexity of the measuring system increased. This made the system more difficult to use during surgery. Now, multiple cameras had to be set up, calibrated and connected to the computer that runs the measurement program. The alignment of multiple cameras takes time and is more complex. Furthermore, all the cables and the computer running the program need to be placed somewhere without hindering the surgeon. Since the time window for setting up the system is relatively small, the system needs to be improved in terms of easier use.

Finally, only detecting visual information does have limitations in the robustness of the measurement system. The system is sensitive to obstructions, movement of the camera, and strong reflections. Besides applying noise handling techniques to increase the robustness, other types of data could be measured to increase this robustness and increase the value of the measured data. For example auditory data or other data on the current and voltage of the system, or a combination of these could increase the value of the measured data even further.

Conclusion The newly developed measurement system can be used for the integration of surgical energy devices in the integrated operating room. The next step is to further improve the system in terms of robustness, easier use, and increased measurement capabilities.

5 Reflections and Recommendations

When designing the system, the programming was done with MATLAB. However the integration of the measurement system became complicated due this, since the OPeLiNK system could not communicate yet with the measurement system. For this, The company OPExPARK had to develop an interface that allowed for communication through TCP. In the future it would be better to have the system programmed in either C++ or C#. This could not only make the integration smoother, but also would increase the overall speed of the system.

Regarding setting up the measurement system in the operating room, this is still too complicated. The assistants in the operating room had to setup the system for measurements. They have a very small time window for this between the applied anaesthesia and the start of the surgery. There was difficulty in placing the camera in such way that there was a clear view of the displays of the energy devices. Furthermore it was perceived that it was not always clear what the measurement system was doing during the setup. This resulted in for example switching between settings while the system was trying to locate the indicators. To solve this, a more clear Graphical User Interface (GUI) needs to be developed to clearly communicate with the people in the operating room what it is doing and what they need to do. Furthermore a manual and troubleshooting guide could help them in better understanding the measurement system. In the future, however, the measuring system must be almost completely automated to reduce the setup time and prevent any added complexity in the operating room.

Currently, the OPeLiNK system communicates the measured information in real time to the surgeon and remote experts through the decision-making navigation screen. When an energy device is activated, all states are communicated. But when the energy device is not activated, only the information of the name of the device and the activation is communicated with the term 'Off'. The problem is that this creates the situation where if there are very short activation, which can be of 0.3s duration, the information of all other states are only displayed for this time period. For easier telementoring, some form of past data needs to be shown. This gives time to the surgeon and remote expert to review recent events.

For the measurement setup based on the state of the art, an interesting discovery was made. In general the system measures the current through the current probe, and when the value is above a certain threshold the activation of the electrosurgical tool was measured. But when the electrosurgical tool makes contact with tissue, a drop in the current was perceived. Based on this, the setup was later slightly enhanced with a second threshold to measure the contact and activations separately. It could be interesting to research this further, because currently all energy device state estimation systems (including our own developed image recognition system) only measure the activation and not the actual contact. The data could be enhanced with also tissue-contact information, which might give significant information for the quantification of a surgeon's skill and is essential for the training of machine learning systems for future semi-automated surgery. Examples of the measured current data can be found in Appendix D.

A An overview of Image Recognition Methods and Noise Handling Methods for Different Types of Indicators

For the development of the image recognition system it is vital to know the state of the arts in the real time recognition of the different types of indicators that the energy devices use to communicate their state. This literature review gives an overview of these recognition methods and evaluates them based on their speed and accuracy. Furthermore, since the operating room is a very noisy environment, this review presents an overview of different noise handling methods that deal with reflections, shadows, the movement of the camera, and occlusions. Finally, gaps are identified and a recommendation is done on what type of methods would be suitable for a new application within image recognition. This paper already has been graded.

A literature review on image recognition and classification techniques for real-time state estimation of light indicators, characters, and symbols

P. van Esch
4443136

March 12, 2022

Abstract

Purpose: Image recognition is a vital method for the automation of many processes. The types of indicators describing the state of an object or system can be divided into light indicators, characters, and symbols. This paper presents an overview of the state of arts in the real time recognition of these indicators and noise handling methods.

Methods: The literature is reviewed through the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method. 945 articles have been identified through Scopus. 7 duplicates were removed. 340 articles have been excluded based on their abstract and title. 452 articles have been excluded based on their content. This resulted in 146 articles to be included in this review.

Results: Popular localization and state estimation methods are discussed. These methods are compared based on their performance. For the recognition of light indicators, morphological operations and filtering rules in combination with thresholding of the color distribution are used. For the recognition of characters, common machine learning methods give the highest accuracy. A combination of edge detection and template matching is used when datasets are insufficient. For the recognition of Symbols, machine learning methods have the highest potential. Morphological operations and filtering rules in combination with several feature extraction methods and template matching can be used when datasets are insufficient. To increase the accuracy of these methods, different noise handling methods can be used. The suitable method depends on the type of noise and application.

Conclusion: Each method evaluates one type of indicator, while for automation multiple types of indicators are used. The next step would be to combine multiple methods to determine the total state of the measured object or system. For this, machine learning or a combination of different non-machine learning methods can be used depending on the availability of sufficient datasets.

I Introduction

Image recognition is an essential method for the automation of many processes. It is used within automated driving through traffic light, traffic sign and object recognition. [1] Within medical imagery it is used to improve the diagnosis for tumor detection. [2] Or it can be used for example within license plate recognition [3] or automated reading of handwriting. [4] Basically, any of these methods are designed to recognize some type of state of an object or system. When evaluating the state of a system through image recognition, different types of indicators need to be considered. There are three different types of indicators that can communicate information visually: light indicators such as traffic lights, characters consisting of text and number recognition, and other indicators that are characterized by some shapes which are summarized as symbol indicators. In this paper a comprehensive overview of different image recognition methods is presented for these different types of indicators. The literature is reviewed through the application of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method. First the different types of indicator recognition methods are investigated in the order of light indicators, characters, and symbols. The light indicator and character recognition methods are divided into indicator localization methods and indicator state estimation methods. For symbol recognition this is divided into indicator localization methods, feature extraction methods and indicator state estimation methods. After this, the performance of each indicator recognition method is analyzed. The environment and application can largely influence the performance of the indicator recognition method. Noise in the form of light fluctuations due to reflections and shadows can impact the accuracy of these methods. Noise in the form of unintentional movement of the image can also decrease the accuracy of these methods. Finally, obstruction of the tracked object or system will influence the performance of the recognition method. Therefore, an overview of different noise handling methods is presented. These methods provide insight on how to improve the accuracy of the indicator recognition methods in noisy environments. Finally, we identify which indicator recognition methods are most suitable per type of indicator, what the challenges are, what the gaps in the areas of research are, and what the most suitable methods for a new research field would be.

II Method

This section gives an overview of the research method used to find relevant articles by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model. [5][6] The search engine used was SCOPUS. For this search engine, multiple search queries were created for each topic. The topics consisted of the following:

- Realtime light indicator recognition methods
- Realtime character recognition methods
- Realtime symbol recognition methods
- Reflection handling methods
- Shadow handling methods
- Vibration handling methods
- Occlusion handling methods

The keywords were selected based on these topics. Synonyms, different spellings, and different suffixes were accounted for using operators. Articles written in any language other than English or published outside of the year 2012 till 2022 were excluded. Duplicate and inaccessible articles were also excluded. Furthermore, only research articles containing clear methods and results are included. For symbol recognition, a regular search with the applied exclusion criteria resulted in more than 4000 articles. Therefore, for this topic, only review

articles were included in this review.

The PRISMA Diagram in Figure 1 shows the process of selecting the literature to be included. A total of 945 articles were identified through the Scopus database. From this, 7 duplicates were removed. 938 records were screened based on their title and their abstract, resulting in the exclusion of 340 articles. Then 598 full-text articles were assessed for eligibility based on their content. From there, 452 full-text articles were excluded, resulting in 146 articles to be included in this review. Table 1 gives a more detailed view of the selection process per topic.

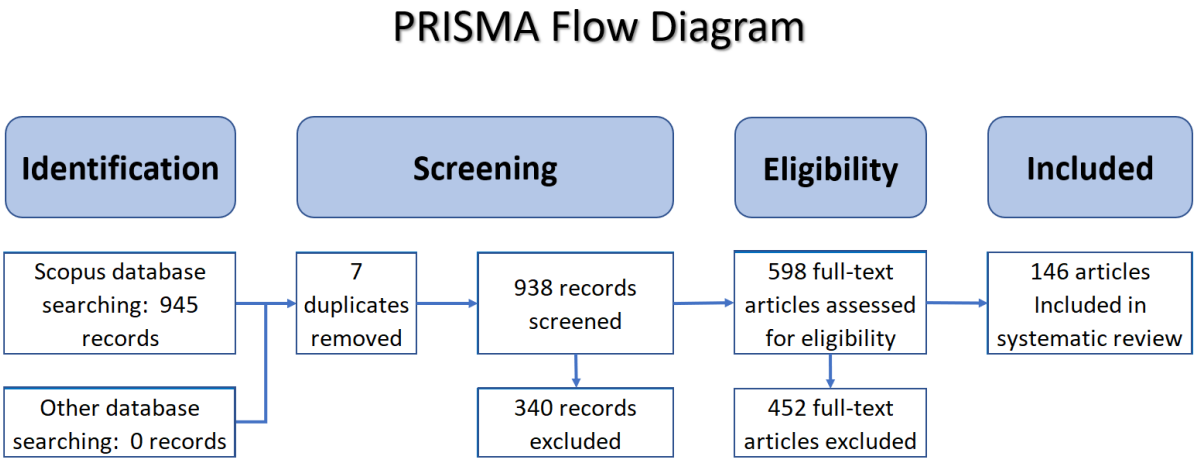


FIGURE 1: PRISMA flow diagram describing the article selection process.

Research topic	Identification Scopus	Identification other	Total identification	Duplicates removed	Records screened	Records excluded	Full-text articles assessed for eligibility	Full-text articles excluded	Articles included
Light indicator recognition	59	0	59	0	59	20	39	33	6
Character recognition	469	0	469	2	467	183	284	208	76
Symbol recognition	115	0	115	0	115	15	100	92	8
Reflection filtering	21	0	21	0	21	5	16	12	4
Shadow filtering	65	0	65	0	65	36	29	16	13
Vibration filtering	138	0	138	5	133	56	77	53	24
Occlusion filtering	78	0	78	0	78	25	53	38	15
TOTAL	**945**	**0**	**945**	**7**	**938**	**340**	**598**	**452**	**146**

TABLE 1: Detailed overview of the article selection process per topic.

III Results

In the following section, different indicator recognition methods are discussed. The indicator recognition methods are discussed in the order of real-time light indicator, recognition, real-time Character recognition, and symbol recognition.

After this, an overview of different noise handling methods for each category of noise is given. These methods are discussed in the order of reflection handling, shadow handling, vibration handling, and Occlusion handling.

1) Real-time Light indicator recognition methods

When it comes to light indicator recognition, automatic traffic light detection is the main field where it is being researched. Traffic light recognition works as follows: An image is acquired through a camera which is then preprocessed for the designated recognition method. The recognition method consists of two parts. First the light indicator needs to be located within the image and separated from the background through a localization method. After this, the features of the light indicator need to be detected, extracted, and then used by a state estimation method to recognize the state of the traffic light.

Light indicator localization methods

For the light indicator localization methods, there are two main methods to locate the traffic lights. The first method is through morphological operations and filtering rules and the second method is through machine learning.

The most popular method for traffic light localization is through morphological operations and filtering rules based on the features of the traffic lights. It consists of applying multiple filters and other morphological operations to remove the background and noise until the traffic light is the only visible part in the whole image.

Diaz-Cabrera, Cerri, and Medici [7] proposed to use color segmentation based on fuzzy logic clustering. Five clusters were made based on images corresponding to red, amber, green, black, and white colors. Uncategorized colors would be stored into a false positive cluster. If a pixel would fall under the false positive cluster, it would be colored white in daytime images and black in nighttime images. After this, to filter out noise and enhance the image, morphological processing in the form of erosion and dilation were performed on red, amber, and green images. This process can be seen in Figure 2. Finally, the found bounding boxes were evaluated based on their diameter and if they were within a certain threshold, the bounding box was deemed valid.

FIGURE 2: Morphological filtering operations for the localization of traffic lights. (a) Fuzzy-logic filtered image, (b) eroded image and, (c) dilated image. [7]

Another method was proposed by Ying, Tian, and Lei [8], where for the localization of the traffic light the image is first preprocessed through a normalization of the brightness. This is performed to reduce the effect of the sunlight. Then since the traffic light was found to be only in the top half of the image, the image was cropped. After this, a median filter was applied to enhance the image and remove noise. To reduce the noise of other object like buildings, signs and the sky, top-hat transformation was used to remove the large uniform areas in the image. Based on the traffic light having high intensities, the regions with high brightness were extracted through threshold segmentation and connected regions

were selected. Then to remove noise in the form of small holes, morphological processing in the form of erosion and dilation were performed. After this, the circular shapes are extracted from the image based on their circularity. Then, the rectangularity of the found regions were evaluated based on the ratio between their length and width to determine if the regions are candidate traffic lights. Finally based on if the extracted circular shapes are within the candidate traffic light region, it can be determined if it is a traffic light.

Wang, Sun, Jiang, *et al.* [9] proposed a method where first color segmentation is performed based on Red-Green-Blue (RGB) and Hue-Saturation-Value (HSV) color space. By combining the segmented images on HSV and RGB, the number of false targets can be reduced. After the segmentation, false targets are removed based on thresholds of the geometric features in the form of the area size and the ratio of the lengths in the horizontal direction compared to the vertical direction.

The final filtering rule-based method is proposed by Tran, Pham, Nguyen, *et al.* [10]. Just like the previous method, first the top halve of the image is cropped. Then the image is converted from RGB to HSV color space after which a new image is created that consists of the S and V channels. After this, the image is binarized through thresholding and dilated to amplify the traffic light signals. Then, an eight-connected tuned contours extractor algorithm is used to extract the circular and rectangular boundaries of each blob. Finally, the blobs are filtered based on the radius of the blobs and the width to height ratio.

Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) are both very popular machine learning methods for image recognition in general. So, it is not surprising that also these methods have been used within traffic light localization. Vishal, Arvind, Mishra, *et al.* [11] proposed to use a CNN based detector for the localization of the traffic light within the image. In this case the CNN based detector was a You Only Look Once (YOLO) detector. YOLO divides the input image into smaller grids, where for each grid cell bounding boxes are created. The CNN then predicts the position and probability of the class for those bounding boxes. This results in the detection of the class ”traffic light box”. Chen, Shi, and Zou [12] proposed the use of SVM for the localization and recognition of the traffic lights. To realize this, first some preprocessing is done. The RGB image is converted to normalized RGB and Hue-Saturation-Intensity (HSI) color space. The color space values are used to create a vector which is then given as input to the SVM classifier to create candidate regions for the traffic lights. Finally noise is removed by applying filtering rules based on the dimensions of the candidate regions.

Light indicator state estimation methods

After that the traffic lights have been located within the image, it is time to detect the state of the traffic light. To do this first the features need to be extracted which are then used to detect the state. For this, three main detection methods are used. The first method is through color density identification based on fuzzy logic clustering. The second method is through color density identification based on color distribution. The third method is a machine learning method based on Support Vector Machine (SVM)

After that the traffic light has been located, Diaz-Cabrera, Cerri, and Medici [7] proposed to use the five Gaussian fuzzy membership functions to evaluate the pixels within the found bounding boxes in terms of color. To validate the pixel, the component values must satisfy all the mean values for the five functions. If the ratio of the number of validated pixels over the area is higher than a certain threshold, the state of the traffic light is determined

Ying, Tian, and Lei [8] proposed to recognize the state of the traffic light based on the color distribution. To do this, first the RGB image was converted from RGB to HSI color space. Then both the RGB image and HSI image are tested through thresholding, where the state can be determined by the combination of the Hue value of the HSI image and the value of each channel of the RGB image.

Tran, Pham, Nguyen, *et al.* [10] took a similar approach by evaluating the color densities, but only in the RGB color space. The found contour blobs were evaluated based on the ratios of the RGB channels where if two of the four ratios are above a certain threshold, the state of the traffic light could be determined.

In terms of machine learning methods, only SVM is used to determine the state of the traffic lights. The most popular method is by combining Histogram of Oriented Gradients (HOG)-features with SVM to determine the state. Both Chen,

Shi, and Zou [12] and Wang, Sun, Jiang, *et al.* [9] used this method to determine the state of the traffic lights. The HOG features are extracted from an image that has been binarized through Otsu's method. The HOG-features give information on the direction of the traffic light. This is then combined with the RGB components and used to train a SVM classifier to determine the state of the traffic light. In Figure 3 an example is given of a dataset that is used for the training of the SVM classifier in terms of the direction of the traffic light. Vishal, Arvind, Mishra, *et al.* [11] took a slightly different approach in the use of the SVM. In this case there was no need to determine the direction of the traffic lights but purely the colors red, yellow, and green. To accomplish this, histograms were created based on each channel of the RGB color space and given as input to a SVM classifier that was trained to estimate the state of the traffic light.

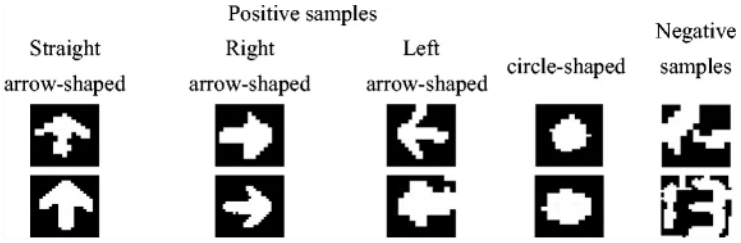


FIGURE 3: Partial training samples for the recognition of traffic lights and their direction. [9]

Performance comparison of real-time light indicator recognition methods

Table 2 summarizes all found studies regarding real-time light indicator recognition methods. In this table we can see that in terms of accuracy, the localization method based on morphological operations and filtering rules in combination with the detection method based on Fuzzy logic cluster has the highest performance. If we look in terms of recognition speed, the localization method based on morphological operations and filtering rules in combination with the detection method based on the RGB color distribution has the highest performance. This however can be biased due to the use of an i7 processor having higher calculation power and thus being faster than the 2 Quad CPU. This could indicate that it is not necessarily the method that improves the speed, but purely the used hardware. If not only the color of the light needs to be determined, but also the direction of the light, it is more appropriate to use SVM in combination with HOG-features.

Study name	Location detection	State estimation	Accuracy [%]	Speed [ms]	processor and RAM
Diaz-Cabrera, Cerri, and Medici [7]	Morphological operations and filtering rules	Fuzzy logic clustering	99.40	0.055	CPU: Intel(R) Core(TM)2 Quad CPU Q9550 @ 2.83 GHz RAM: 8GB
Chen, Shi, and Zou [12]	SVM	RGB + HOG + SVM	99.18	0.084	Core i5–2450M RAM: 2GB
Ying, Tian, and Lei [8]	Morphological operations and filtering rules	RGB +HIS - Color distribution	98	0.09	-
Wang, Sun, Jiang, *et al.* [9]	Morphological operations and filtering rules	RGB + HOG + SVM	96.50	0.115	-
Tran, Pham, Nguyen, *et al.* [10]	Morphological operations and filtering rules	RGB - Color distribution	85	0.02	CPU: Intel i7
Vishal, Arvind, Mishra, *et al.* [11]	CNN: YOLO	RGB + SVM	96	0.143	CPU: Intel Xeon 3.2GH; GPU: Nvidia Quadro K2200

TABLE 2: Overview of light indicator recognition methods and their performance.

2) Real-time character recognition methods

Character recognition is one of the most researched subjects within image recognition. Within text and number recognition two research fields are dominant: license plate recognition and handwriting detection. An image is acquired through a camera which is then preprocessed for the designated recognition method. The recognition method consists of two parts. First the text needs to be located within the image and separated from the background through a localization method. Then the features of the characters need to be extracted and recognized based on a recognition method. In handwriting recognition, the localization step is sometimes skipped and only the recognition step is researched.

Character localization methods

The most popular method in license plate detection, is the use of edge detection methods in combination with some type of filtering rules. Pervej, Das, Hossain, *et al.* [13] proposed to use the Sobel edge detection in the extraction of the license plate. For this, first the acquired image is converted from RGB to gray scale after which Histogram Equalization is applied to enhance the contrast. Then edges are detected by using a Sobel Edge filter. After this, the image can be cropped to the license plate by applying adaptive thresholding. Finally, the image is aligned horizontally through rotation and some noise is removed through Gaussian filtering. Figure 4 shows the process of the localization of the the license plate. Wahyono and Jo [14] proposed to use Canny edge detection for the localization of electronic road signs. The RGB image is directly converted to an edge image. In this case since the text consists of multiple small dot regions, called blobs, supporting points defined as center points of a segment are extracted. Then based on their properties, the supporting points are merged to generate the character region. Jain and Sharma [15] proposed to use a different type of edge detection in the form of the Prewitt Operator. First the image is acquired, converted to grayscale and then binarized. To remove noise, a median filter is applied. Then the Prewitt Operator is used for edge detection such that features can easily be extracted. The lines, words and characters are then segmented based on horizontal and vertical projection techniques, where the horizontal and vertical histograms determine how they can be segmented.

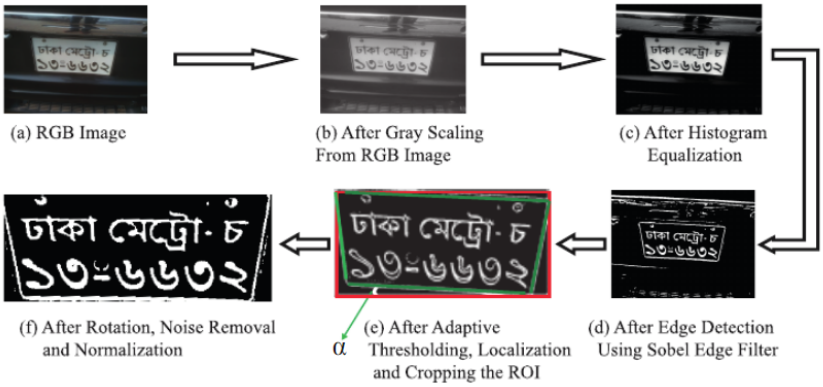


FIGURE 4: The process of finding the location of the license plate through edge detection. [13]

Ahmed and Ahmed [16] proposed to use Connected Component Analysis (CCA) for the localization and segmentation of the characters for Automatic Number Plate Recognition (ANPR). To realize this, first the image is converted to grayscale. Then the image is binarized through the adaptive threshold method. After this, morphological transformations are applied to reduce the irrelevant contours from the threshold image. These transformations consist of erosion, opening, closing and gradient operations. The CCA method is used to remove remaining irrelevant contours and objects in the image. CCA groups connected pixels belonging to the same object and then labels each found area with a unique number. After this, a bounding box Is created for the segmentation of the overall text and each character. In Figure 5 the process of character segmentation is shown after having found the candidate region for the license plate.

With regards to machine learning methods, the Viola and Jones method and the Convolutional Neural Network (CNN) are the popular methods used in license plate detection. Where CNN is by far the most popular.

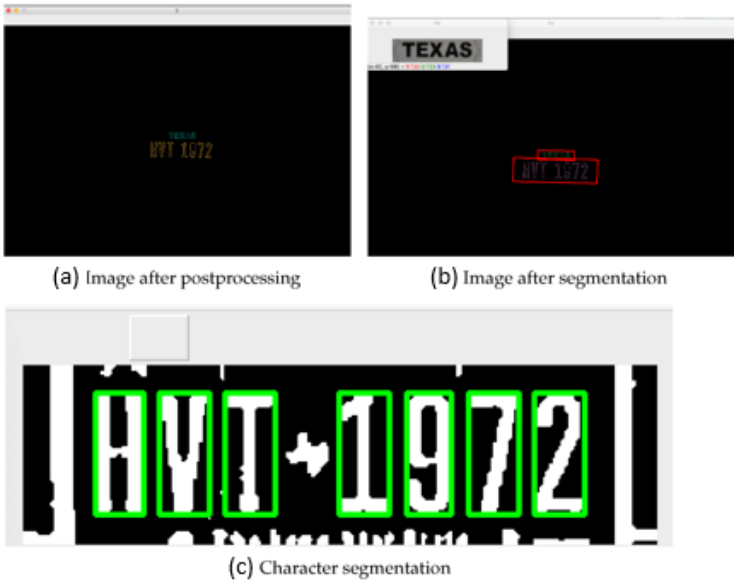


FIGURE 5: The process of character segmentation through CCA after the candidate region of the license plate has been detected. [16]

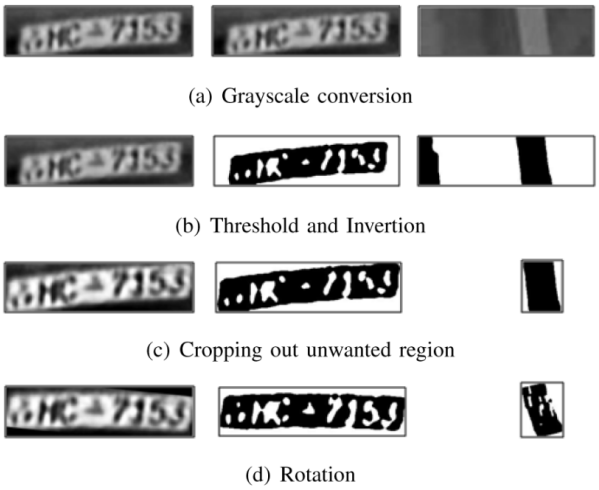
Ajanthan, Kamalaruban, and Rodrigo [17] proposed to use the Viola and Jones method to locate license plates. This method consists of a trained network, called the AdaBoost algorithm, which receives Haar-like features as input. This detector does give false positives due to its method of looking for dark regions in a bright background. To cope with this, first the plate region is enhanced through thresholding and morphological opening. The unwanted regions are then cropped through horizontal and vertical scanning. Finally, to identify the plate, the detected region is evaluated on its width-to-height ratio. In Figure 6 the process of finding the location of the license plate is shown. Zheng, He, Samali, *et al.* [18] proposed an extension of the Viola and Jones method by not only using Haar-like features, but also global features based on edge density and edge density variance. First Sobel edge detection is applied for vertical edge detection. After converting the image to a vertical edge map, Haar-like local features and Global features based on edge density and edge density variance are extracted. Both types of features are then used by the AdaBoost algorithm to detect the license plate.

Fu, Chen, Hou, *et al.* [19] proposed to use a CNN, named YOLO9000, for the detection of license plates. First the created image is resized to 416x416 pixels and given as input to the YOLO9000 algorithm. The special thing about YOLO9000 compared to other YOLO CNNs, is that it uses batch normalization to increase the model convergence speed before every convolutional layer. The output of the CNN is in the form of bounding boxes within the image, indicating the location of the license plate.

Character state estimation methods

A common method for character recognition is template matching. Ahmed Biyabani, Al-Salman, and Alkhalaf [20] proposed to use template matching in the form of a cross correlation function. To realize this, first Sobel edge detection in combination with horizontal and vertical projection is used to locate the license plate. Then the same method of horizontal and vertical projection is used to segment each character. After segmentation, the segmented characters are correlated with a template image dataset. The cross correlation function is described by the following:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \tag{1}$$



(a) Grayscale conversion

(b) Threshold and Invertion

(c) Cropping out unwanted region

(d) Rotation

FIGURE 6: The process of license plate detection through Viola and Jones method together with the filtering of false detections.The left column indicates the license plate image to be used for state estimation. The middle and right column show the process of removing the false detection, where the right column is a false detection. [17]

In this formula, A represents the pixels of one image and B the pixels of the other. The result is a correlation coefficient that describes how much the two images correlate. The highest correlation value r indicates the recognized character. Massoud, Sabee, Gergais, *et al.* [21] proposed to use template matching in the form of a normalized cross correlation (NCC) function. For this, first the license plate is detected through Sobel edge detection with a few morphological filtering methods in the form of erosion, dilation, and a median filter. After the license plate has been segmented, it is first divided into two halves, where the left halve consists of letters and the right halve consists of numbers. Then each individual character is segmented through horizontal projection. Finally, each segmented character is correlated with the template data base images through the NCC function. The formula for NCC is given below. F_1 indicates the pixels of the input image and F_2 indicates the pixels of the template image. M and N indicate the parent image size, while m and n indicate the pixel location in the parent image. j and k indicate the pixel location in the template image. The highest correlation value R indicates the recognized character.

$$R(m,n) = \frac{\sum_j \sum_k F_1(j,k) F_2(j-m+(M+1)/2, k-n+(N+1)/2)}{[\sum_j \sum_k |F_1(j,k)|^2]^{1/2} [\sum_j \sum_k |F_2(j-m+(M+1)/2, k-n+(N+1)/2)|^2]^{1/2}} \tag{2}$$

Salahshoor, Broumandnia, and Rastgarpour [22] proposed to use template matching combined with the method of Euclidian Distance for vehicle plate recognition. First the plate is located through edge detection and morphological operations. After the license plate has been located, the image gets prepared for character segmentation. The license plate angle is corrected, and the contrast is enhanced. After this the image is binarized through Otsu thresholding. Finally, the characters get segmented based on their newly created method called Averaging of White Pixels in Objects (AWPO). After the segmentation the Euclidian distance is calculated for character recognition, where template matching is used to classify the character.

Bague, Jorda, Fortaleza, *et al.* [23] proposed to use a binary CNN, named B-CEDNet, in combination with a bidirectional recurrent neural network (bi-RNN) for scene text recognition. In this research the text does not have to be located within a larger image, rather only the characters have to be recognized. To realize this, first an input image is given to the B-CEDNet. This CNN normalizes and binarizes the image through an Adapter. After this the Binary convolutional encoder performs binary convolution where the output is the activation. Finally, the strongest activation is filtered out through 2 x 2 max pooling and binarized by a Binrz layer. The binarized activations are then given to the Binary convolutional decoder which gives as output a posterior probability distribution for each pixel in the image. After this,

morphological filtering is applied to find the candidates of the character regions through thresholding to remove false detections at the edges of the image. Finally, the Bi-RNN is trained to remove false detections with high confidence value.

Ramteke, Gurjar, and Deshmukh [24] proposed to use an SVM algorithm for the recognition of Handwritten Marathi text. To do this, first the image is preprocessed by converting it to grayscale. Then it is binarized through thresholding. To reduce noise, morphological operations are applied. After this, the preprocessed image is segmented through line, word and character detection and segmentation. Lines are segmented through the horizontal projection method and thresholding. Then words are segmented through a vertical projection method and thresholding. After this Characters are segmented through a combination of vertical and horizontal projection with thresholding. For the character recognition, features are extracted through the curvelet transform. This consists of four steps: Subband decomposition, Smooth partitioning, Renormalization and Ridgelet analysis. After this, dimensionality reduction is applied on the large dimensional feature space through Principal Component Analysis (PCA). After this the resulting feature vector is optimized through an Adaptive Cuckoo Search (ACS) algorithm. Finally, to recognize the characters, a SVM is trained on the optimized feature vector.

Kour and Saabne [25] proposed to use a k-nearest neighbors (KNN) algorithm for the recognition of handwritten Arabic characters. To realize this, features are extracted through Shape Context (SC) and the Multi Angular Descriptor (MAD). After this, the features vectors are embedded by an approximation of the Earth Mover's Distance (EMD) using the Manhattan distance into normed wavelet coefficients domain, in the form of Haar-wavelet coefficients. Then the dimensionality is reduced through Principal Component Analysis (PCA) followed by linear discrimination analysis (LDA). The feature vectors are then given as input to a KNN classifier to classify the characters.

Performance comparison of real-time character recognition methods

Table 3 gives an overview of different real-time character recognition methods and their performance. If we investigate which methods perform the best, we can see that for localization and character recognition the CNNs result in the highest accuracy, closely followed by edge detection methods for localization in combination with CNNs for recognition. Other machine learning based methods like SVM also have extremely high accuracies within recognition. Also in terms of speed, we see that CNNs in combination with edge detection have the highest performance closely followed by edge detection in combination with template matching based on a correlation function. This difference in speed however is not necessarily reflected by the type of method used, but is also influenced by the complexity of the input image. Although the CNN based method performs faster, the application is based on reading banknotes. The template matching based method is used for automatic license plate recognition, which although performs slower, has much more noise to deal with for both the localization and detection of the characters.

Study name	**Location detection**	**State estimation**	**Accuracy [%]**	**Speed [ms]**	**processor and RAM**
Ahmed Biyabani, Al-Salman, and Alkhalaf [20]	Horziontal and vertical projection method + Sobel edge detection	Template matching: correlation function	84	1.3	Altera Cyclone III EP3C25F324 FPGA device
Ahmed and Ahmed [16]	Connected component analysis (CCA);	KNN	95	10	4.50 GHz Intel Core™ i7-16MB CPU processor, 16 GB of RAM
Ahn and Cho [26]	CNN: YOLO	OCR-net	95.70	-	-
Ajanthan, Kamalaruban, and Rodrigo [17]	Viola and Jones method: AdaBoost + Haar-like features and cascade of classifiers	SVM	90	38	-
Alghyaline [27]	CNN	CNN	87	33.3	-

Ali and Suresha [28]	CNN	SVM	96,9	-	-
Al-Jubouri and Abusaimeh [29]	-	SVM	92.20	-	-
Alzubaidi and Latif [30]	-	KNN	90.60	-	-
Arafat, Khairuddin, and Paramesran [31]	Morphological operations and filtering rules	Connected component analysis (CCA) + template matching for recognition	95.40	520	Intel® Core™ i3 CPU Ram: 4GB
Awalgaonkar, Bartakke, and Chaugule [32]	Mobilenet V1	Easy OCR	90.	-	Jetson Nano
Bachchan, Gorai, and Gupta [33]	Color feature extraction using fuzzy logic and HSI model. Texture feature extraction using scale-invariant feature transform (SIFT) and advanced local binary pattern (ALBP) are used for plate area detection. Extraction Region of interest with Local Binary Pattern (LBP) descriptor	Segmentation with Minimum bounding rectangle (MBR) + KNN	89.35	-	-
Bague, Jorda, Fortaleza, *et al.* [23]	-	Keras + CNN: VGG16	98.84	1950	-
Bhutta, Mahmood, and Malik [34]	HSV filter with aspect ratio	Multi layer perceptron (MLP) neural network	98	2740	Core i3 Ram: 3GB
Chen, Yang, and Lk [35]	CNN: YOLOv3	CNN: YOLOv3	84.30	25	GeForce RTX2080ti GPU
Duan, Cui, Liu, *et al.* [36]	CNN	CNN	99	33.3	-
Dun, Zhang, Ye, *et al.* [37]	sliding window technique + vertical edge density + Machine learning method	Connected Component analysis + vertical projection + stacked denoising autoencoder + Hog features	83.29	-	-
Edlin, Kiantono, Martin, *et al.* [38]	CNN	CNN	87.53	530	-
Elsaid, Alharthi, Alrubaia, *et al.* [39]	-	Connected component analysis (CCA) + feature matching: Regionprops matlab function;	94.70	2000	-
Farhad, Nafiul Hossain, Hossain, *et al.* [40]	Blob detection	neural network (NN)	91.40	1320	-
Feng and Xia [41]	Sobel edge detection + Morphological operations and filtering rules	Vertical projection for character segmentation. Feature matching: concavity, curvature and points of intersection	94.67	-	-
Fu, Chen, Hou, *et al.* [19]	CNN: YOLO9000	CNN	99.98	17.25	-
Gao, Ge, Lu, *et al.* [42]	DCNN	deep convolutional neural network (DCNN)	94.50	407.51	-

Ge, Liu, Sun, *et al.* [43]	Gravity-center method + Sobel edge projection method	Multidirectional line scanning (MLS) for segmentation. Point Cloud registration for DOt-matrix Character (PC4DOC) method for recognition.	93.84	830	-
Godage and Wimalaratne [44]	Morphological operations and filtering rules	Recognition based on template matching based on pieces of characters	92.50	-	-
Kour and Saabne [25]Gonçalves, G. R. Menotti, D.	SVM + HOG	SVM classifier + HOG features	89.60	29.4	Intel(R) Xeon(R) X5670 CPU 32GB RAM
Grębowiec and Protasiewicz [45]	-	CNN + histogram	89	-	-
Guo, Shi, Bao, *et al.* [46]	-	Preliminary character recognition based on distance histogram. Character recognition based on angle histogram. Comprehensive recognition based on weighted combination of both distance and angle histrogram.	95	180	Intel Celeron III processor 1.0 GHz CPU
Jain and Sharma [15]	Prewitt Operator for edge detection	Connected Component Labelling + artificial neural network (ANN)	96	-	-
Jain, Sasindran, Rajagopal, *et al.* [47]	CNN	CNN	98.64	200	-
Jang, Suh, and Lee [48]	Edge detection + morphological operations and filtering rules	CNN	99.85	0.56	-
Jeon, Nguyen, and Jeon [49]	DNN	DNN	98.23	-	-
Jia, Gonnot, and Saniie [50]	Pixel statistics methodology + the Hough Line transformation	Otsu method + redpass filter for red characters + Template matching	95.10	-	-
Joseph and Hameed [51]	-	SVM	90	520	-
Kamble and Hegadi [52]	-	Feedforward NN + SVM + Rectangular-HOG (R-HOG)	97.15	-	-
Kathigi and Kariputtaiah [53]	-	AlexNet + SVM	99.97	-	-
Keerthi Prasad, Khan, Chanukotimath, *et al.* [54]	-	Principle component analysis (PCA) + euclidian distance	88	800	-
Khazaee, Tourani, Soroori, *et al.* [55]	CNN: YOLOv3	CNN: YOLOv3	98	38	-
Khosravi [56]	Sobel edge detection	NN: AdaBoost M2	97.80	240	-

Kour and Saabne [25]	Feature extraction with Shape Context (SC) and the Multi Angular Descriptor (MAD). Embedding through the Earth Mover's Distance (EMD). Reduction of dimensionality through Principle component analysis (PCA).	KNN	96	4.4	-
Laroca, Severo, Zanlorensi, *et al.* [57]	CNN: Fast-YOLO + YOLOv2	CNN: CR-NET	78.33	28.6	-
Lee, Hung, and Wang [58]	NN	NN	95.50	-	-
Liu, Huang, Cao, *et al.* [59]	Sobel edge detection	CNN + connected component analysis (CCA) + projection analysis (PA)	93.74	318	-
Liu, Zhang, Zhang, *et al.* [60]	difference of Gaussian (DoG) filters	Difference of Gaussian (DoG) filters + thresholdding	70.1	262.1	3.4GHz CPU
Liu, Li, Ren, *et al.* [61]	-	Binary convolutional encoderdecoder network (B-CEDNet) + bidirectional recurrent neural network (Bi-RNN).	88	less than 1ms	XNOR kernel on TITAN X GPU
Lukic, Makarov, and Spanovic [62]	-	Tesseract OCR	95.51	-	-
Massoud, Sabee, Gergais, *et al.* [21]	Sobel edge detection	Template matching: normalized cross correlation (NCC)	91	-	-
Muresan, Szabo, and Nedevschi [63]	-	CNN	97.50	-	Intel i5-5200 CPU with 2.20 GHz; NVIDIA GeForce 920M GPU
Odate and Goto [64]	-	KNN	94.37	-	-
Onim, Akash, Haque, *et al.* [65]	-	CNN	90.50	70	-
Pavaskar and Budihal [66]	-	KNN	85	-	-
Pervej, Das, Hossain, *et al.* [13].	Sobel edge detection	Feature recognition: contour vectors	95.41	35.8	-
Pirgazi, Sorkhi, and Kallehbasti [67]	Connected component analysis (CCA) + Morphological operations and filtering rules	Random Forest Model + F-score criterion	97.50	125	-
Prasad, Khan, and Chanukotimath [68]	-	Principle component analysis (PCA) + euclidian distance	86	800	-
Qin and Liu [69]	CNN	CNN	92.7 - 99.7	38	-
Rahaman, Mahin, Ali, *et al.* [70]	-	CNN	97.43	44.95	-

Ramteke, Gurjar, and Deshmukh [24]	Morphological operations and filtering rules	SVM	97.15	6.55	-
Rosalina, Hutagalung, and Sahuri [71] P.	CNN	CNN	99.56 for digits, 90.78 for Hiragana	-	-
Salahshoor, Broumandnia, and Rastgarpour [22]	Detector for the Blue Area (DBA)	Template matching + euclidic distance	95.09	450	-
Sethy and Patra [72]	-	NN	94.6-96.3	-	-
Shahed, Udoy, Saha, *et al.* [73]	-	Template matching: correlation coefficient + edge detection	95	750	-
Shelke and Apte [74]	-	Template matching: correlation coefficient	92.50	-	-
Siddique, Iqbal, Mahmud, *et al.* [75]	Sobel edge detection	Segmentation based on pixel connectivity; template matching based on dc offset, coefficent of inclination and coefficient of curvature	85	800	-
Silva and Jung [76]	CNN	CNN	92.41	114.5	-
Srivastava, Rajitha, and Agarwal [77]	-	Gradient features + SVM	91.3 - 93.5	-	-
Wahyono, Filonenko, and Jo [78]	Color probability density functions (PDF) of normal distributions for RGB channels	Local spatial pattern + random forest classifier	21 - 0.77	200	-
Wahyono and Jo [14]	Canny edge detection + Morphological operations and filtering rules	Density-based spatial region generation (DB-SREN) + k-nearest cluster neighbor (k-NCN)	61	87.2	-
Wibowo, Sigit, and Barakbah [79]	-	Shape Energy: + Backpropagation neural network (BPNN)	90.30	-	-
Xu, Wang, Niu, *et al.* [80]	CNN: YOLOv3 + Deeplabv3plus	SVM	92.33	-	-
Yang, Zhuang, Zhou, *et al.* [81]	Horizontal and vertical projection + edge detection	Template matching: simple binary template matching + stroke direction density characteristics method.	85 under normal lighting and max 15 degree angle	-	-
Yuan, Zhang, Wu, *et al.* [82]	-	BPNN	99.90	-	-
Zhai, Bensaali, and Sotudeh [83]	-	Feed forward ANN	97.30	8.4	3GB RAM
Zhang, Luan, Xu, *et al.* [84]	CNN: YOLOv5 + NMS module	CRNN model	72.80	-	-
Zhao, Cao, and Meng [85]	Morphological operations and filtering rules	Vertical projection method for segmentation. No character recognition	97	-	CPU 3.0 RAM: 2GB
Zheng, He, Samali, *et al.* [18]	Haar-like features + Sobel edge density and magnitude of the verical edge + cascaded classifier	AdaBoost + Blob detection + Connected Component Analysis (CCA)	94.30	204	Pentium 2.8 GHz

Zhuang, Liu, Qiu, *et al.* [86]	-	CNN	90.91	-	-
Zohra and Rajeswara Rao [87]	-	NN	CNN: 98.4; KNN 96.7; SVM: 97.6	-	-

TABLE 3: Overview of character recognition methods and their performance.

3) Real-time symbol recognition methods

In terms of symbol recognition, we look at standard image recognition methods. There are three main fields where this is researched: Traffic sign detection, medical image recognition and fingerprint recognition. For these types of recognition, the methods are divided in three steps: localization, feature extraction and state estimation.

Symbol localization methods

Zhu, Yuen, Mihaylova, *et al.* [88] investigated the different methods for traffic sign detection. In terms of the localization, they concluded that most researches use some form of morphological operations and filtering rules, while other methods use some type of machine learning in terms of a cascade classifier with AdaBoost. In terms of traffic sign detection, often different types of morphological operations and filtering rules are applied to extract the sign out of the larger image. Often this is based on thresholding in a certain color space. This can be either HSI, HSV, RGB, or Luminance-Chroma-Hue (LCH) color space. After this the image is further enhanced through for example contrast normalization, RGB normalization or Ohta normalization. Khan, Sajjad, Hussain, *et al.* [89] did research on the classification for White Blood Cells (WBC) in blood smear images. For segmentation, often morphological operations and filtering rules are used. In this case a combination of filters is used consisting of a median filter, a low-pass filter, a high-pass filter and a Gabor filter. In terms of machine learning methods, the most often used methods are cascade classifiers and CNNs for finding the locations. In his research on Deep Learning Applications in MRI Images, Datta and Rohilla [90] showed that MRI images get segmented based trained machine learning algorithms. In this research segmentation was done either though different types of CNN or a voxel-wise residual network (VoxResNet).

Symbol Feature extraction methods

To be able to recognize the symbols, the features defining the symbol need to be extracted. There are many different methods to extract features depending on what type of image needs to be recognized. Features are either extracted based on visual feature extraction methods or deep feature extraction through machine learning.

Zhu, Yuen, Mihaylova, *et al.* [88] and Fulco, Devkar, Krishnan, *et al.* [91] showed that withing traffic sign detection either color-based approaches are used in the form of color-based feature selection or shape-based approaches. Shape based approaches consist of Edge detection, HOG-features, Haar wavelet-like features, Fast Fourier Transform (FFT), Shape signatures, Tangent function, Simple image patches. Wei, Tran, Xu, *et al.* [92] showed other feature extraction techniques for retail product recognition. In this case either SIFT, Speeded Up Robust Features (SURF) or machine learning methods in the form of CNN were used. Wang, Zhang, and Xia [93] used various feature extraction methods for the recognition of medical wubfigures. These consisted of Bag of features (BoF), Color histogram (CH), Gray-level co-occurrence matrices (GLCM), local binary patterns (LBP), Gabor features, CENsus TRansform hISTogram (CENTRIST), pyramid histogram of oriented gradients (PHOG), edge histogram descriptor (EHD), auto color correlogram (ACC), and color and edge directivity descriptor (CEDD). Then to reduce dimensionality, dimensionality reduction was applied in the form of Principle Component Analysis (PCA) or Locality Preserving Projections (LPP). In terms of fingerprint recognition morphological operations and filtering rules are commonly applied to extract the minutiae. This would be through converting the image from RGB to gray and apply binarization. Other methods are based on machine learning.

In terms of machine learning there are several types of deep feature extraction. CNN is by far the most common approach applied in all fields of research. The most common types are YOLO and SSD. Other machine learning methods consist more of more specific types like RBF neural networks and reinforcement learning in fingerprint recognition.

Symbol state estimation methods

In terms of recognition methods, these highly depend on the types of features that have been extracted. Fulco, Devkar, Krishnan, *et al.* [91] showed the use of the Hough transform for edge features. For the detection of different shapes of traffic signs a radial symmetry detector or fast radial symmetry transform can be used. In terms of fingerprint recognition, Hasan and Abdul-Kareem [94] and Ali, Khan, and Aslam [95] showed that often a basic form of template matching is done by overlaying the found image and a template and measuring the alignment. Some techniques use Fuzzy Logic techniques to match the fingerprints. Other techniques are based on machine learning.

In terms of machine learning methods, Fulco, Devkar, Krishnan, *et al.* [91] showed how HOG-features can be used to classify a traffic sign through either SVM or a cascade classifier. Other methods like Multilayer perceptron neural networks (MPNN) or genetic algorithm can also be used in combination with different types of feature extraction methods. Next to traffic sign detection, SVMs are also widely used for medical image detection for WBC detection by Khan, Sajjad, Hussain, *et al.* [89] and for medical subfigure detection by Wang, Zhang, and Xia [93]. Other types of machine learning methods for WBC detection are KNN, Naïve Bayes (NB), or regular Artificial Neural Networks (ANN). Datta and Rohilla [90] showed that Recurrent Neural Networks (RNN) and CNNs are common recognition methods for the detection of MRI images. CNNs remain the most used machine learning method, which are applied in every research field.

Performance comparison of real-time symbol recognition methods

In terms of symbol recognition methods, it is hard to say which method outperforms the other. In Table 4 an overview is given of the identified symbol localization, feature detection and symbol state estimation methods. What we see is a large increase in popularity in the use of machine learning based methods for symbol recognition. What we can say based on the previous character recognition section is that in general machine learning based methods are more robust and have higher accuracy within image recognition. CNNs and SVMs maintain to be the most popular used method in combination with HOG or Haar-like features within symbol recognition. The large downside is that to increase the accuracy large datasets are necessary. Depending on the field of research the availability of such datasets can be quite limited. When this is the case, it can be more efficient by using simple template matching.

Study name	Location detection	Feature extraction	State estimation
Zhu, Yuen, Mihaylova, *et al.* [88]	Morphological operations and filtering rules: cascade classifier + AdaBoost	Color based approaches: color based feature selection; Shape based appraoches: edge detection methods (canny, Sobel, etc), HOG-features, Haar wavelet-like features, and other methods like FFT, shape signatures, tangent function, simple image patches.	Hough transform with edge features; radial symmetry detector, fast radial symmetry transform; SVM with HOG features; cascade classifier with HOG features; Multilayer perceptron Neural Networks (MPNN); Genetic algorithm
Wei, Tran, Xu, *et al.* [92]	Segmentation + transformation + enhancement	SIFT, SURF, CNN	CNN: YOLO, SSD
Wang, Zhang, and Xia [93]	-	Deep feature extraction: CNN; Visual Feature extraction: BoG, CH, GLCM, LBP, CENTRIS, PHOG, EHD, ACC, CEDD; Dimensionality reduction: PCA; LPP	SVM
Khan, Sajjad, Hussain, *et al.* [89]	Morphological operations and filtering rules: median filter, low-pass filter, high-pass filter, and Gabor filter	CNN or other machine learning methods	SVM, Naïve Bayes (NB), KNN, Artificial neural network (ANN)
Hasan and Abdul-Kareem [94]	-	Morphological operations and filtering rules: Minutiae extraction with RGB to gray conversion with binarization. Parallel iterative thinning algorithm called MB2, RBF neural network, reinforcement learning	Template matching: allignment, fuzzy logic, neural network

Fulco, Devkar, Krishnan, *et al.* [91]	-	maximally stable extremal regions (MSER) , HOG, CNN, color based feature selection	CNN, SVM
Datta and Rohilla [90]	Segmentation based on CNN + voxel-wise residual network (VoxResNet)	-	Various deep learning methods: CNN, Recurrent Neural Networks (RNN)
Ali, Khan, and Aslam [96]	-	-	Template matching, Multi-Layer Perceptron neural network, CNN

TABLE 4: Overview of symbol recognition methods and their performance.

4) Reflection handling methods

Reflection handling is mainly researched within iris detection or the specular removal of objects or faces. It consists of two parts: detection and filtering. Sometimes reflections are not even detected but rather limited through certain filtering methods.

Reflection detection methods

There are two main methods for the detection of reflections in an image. This is either through morphological operations and filtering rules, or through albedo estimation and intensity estimation.

Alice Nithya and Lakshmi [97] proposed to use simple thresholding for the removal of specular reflection in Iris recognition. First the image is converted to gray scale. Then the intensity is measured, where if the intensity is above a certain threshold, then this is detected as specular reflection. In this case the threshold was determined to be the top 20% of intensities. Ge, Han, and Shen [98] proposed to use intensity and chromaticity analysis to detect specular reflections on objects. First the intensity of every pixel is measured. Then based on thresholding high level and medium level intensity pixels are detected. High level intensities are labeled as specular reflection while for medium level intensity pixels the diffuse chromaticity's are estimated. Then if the chromaticity distance is larger than a certain threshold, the pixel is highlighted as specular reflection.

Muhammad, Dailey, Farooq, *et al.* [99] took a different approach for the detection of specular reflection in faces. First the image is converted from RGB to HSV color space. Then the albedo is estimated. After this a Levenberg-Marquardt nonlinear least squares optimization is used to estimate intensities and position of the lights, which is required for creating a dataset of images with and without specular reflection. This in turn is used in the filtering step for the training of a deep learning model.

Reflection filtering methods

For the filtering of the reflection, Rajeev Kumar and Arthi [100] proposed a method for an iris recognition system where to deal with specular reflection, no detection was used. Rather the effect of specular reflection was limited by first converting the image to black and gray, normalizing the image and applying a Hybrid Median filter. Alice Nithya and Lakshmi [97] proposed to use simple thresholding and morphological binarization operators on the detected specular reflections. First dilation is applied after which the detected pixels are set to zero. In Figure 7 the process is shown from reflection detection to filtering the reflection. After having detected the specular reflection pixels, Ge, Han, and Shen [98] removed the specular reflection component through the estimation of diffuse chromaticity based on chromaticity similarity. Muhammad, Dailey, Farooq, *et al.* [99] used machine learning in the form of a trained deep learning model named Spec-CGAN to filter out the reflections. The image is given directly as input to the model, which in turn gives an image filtered without specular reflection as output.

Table 5 gives an overview of the different detection and filtering methods for reflection handling. Based on the current findings it is impossible to compare these methods based on their performance. It rather shows that there are several methods to deal with reflections, where the method is determined by the type of image recognition used the field of application.

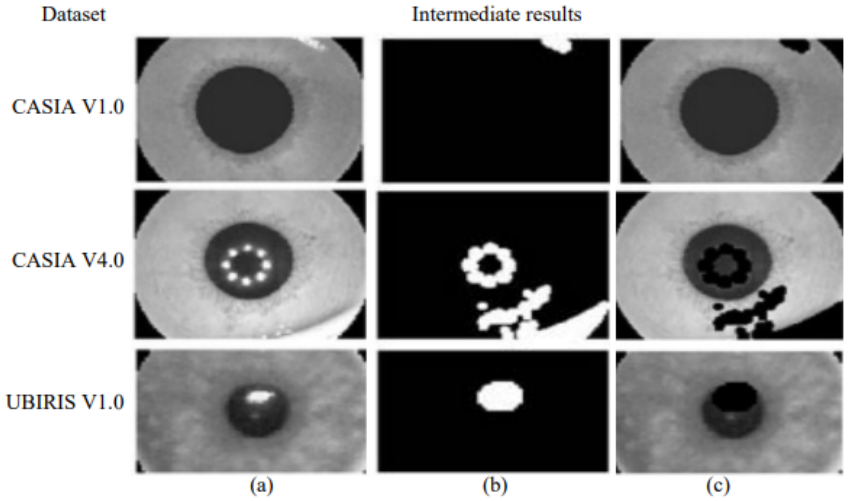


FIGURE 7: Reflection filtering based on thresholding and reflection filtering for different datasets, with (a) the input image, (b) the detected reflection, and (c) the filtered image. [97]

Study name	Detection	Filtering
Muhammad, Dailey, Farooq, *et al.* [99]	Skin pixel detection through converting to HSV and then estimating the albedo. Levenberg-Marquardt nonlinear least squares optimization to estimate intensities and position of the lights.	Reconstruction through Deep learning model
Ge, Han, and Shen [98]	High level intensity pixels are detected. Diffuse chromaticities are estimated on medium level intensity pixels. Highlighted pixel is determined if chromaticity distance is large.	Specular reflection component is removed based on the robust estimation of diffuse chromaticity under the guidance of chromaticity similarity.
Rajeev Kumar and Arthi [100]	-	Use hybrid median filter
Alice Nithya and Lakshmi [97]	High intensity threshold	Thresholding and morphological binarization operators. Set detected pixels to zero.

TABLE 5: Overview of reflection handling methods and their performance.

5) Shadow handling methods

In terms of shadow handling methods, this is mainly done within the field of object, people, or vehicle detection. Shadow handling methods can be divided into shadow detection and shadow filtering. Depending on the research, either only the shadow is detected, or it is both detected and filtered afterwards.

Shadow detection methods

The most applied method for shadow detection consists of morphological operations and filtering. Other methods consist of fuzzy logic, Gaussian Mixture Model (GMM) and machine learning in the form of CNN.

Macedo, Nascimento, and Souza [101] proposed a real-time shadow detection method based on morphological operations and filtering for object tracking. First the acquired image is converted from RGB to multiple different color spaces consisting of grayscale, YCrCb and CIE L*a*b*. After this, the relevant channels are split into separate images. This allows to filter out non-shadow regions with high intensities in only one of the channels. Then binarization is applied

through Otsu thresholding. Finally, segmentation of the shadows is done through thresholding.

Niranjil Kumar and Sureshkumar [102] used fuzzy logic for the detection of shadows. First the background is subtracted based on morphological operations and filtering rules. Then edge detection is used to create an edge image. After this image is denoised with some morphological operations. After the image has been processed the shadow is detected through fuzzy logic. The fuzzy rule consists of two membership functions consisting of background pixel distributions and shadow pixel distributions. Each function has a membership value for every detected region indicating how much the region belongs to the class of background or shadow. The maximum membership value determines the class of the region. In Figure 8 the process is shown for the detection of a shadow in an image.

FIGURE 8: Detection process of an shadow withing an image with (a) the input image, (b) the subtracted background, (c) the detected shadow, and (d) the detected foreground. [102]

Li, Li, Tian, *et al.* [103] used a Gaussian Mixture model (GMM) for the detection of shadows in vehicle detection. First both the moving vehicle and its shadow are detected as a foreground object through an edged Gaussian Mixture Modeling method (GMM). After this the background can be used as a reference to estimate a brightness threshold for shadow detection. Since the brightness of the shadow is usually darker than the background, this distortion in brightness is used to detect the shadow.

Fan, Han, and Li [104] used machine learning in the form of CNN for the detection of shadows. The CNN was named RSnet, which consist of an encoding-decoding stage and a refinement stage. The encoder network is based on the pre-trained convolution layer of a VGG16 network. It takes the image, and creates feature representations to transfer them to shadow masks for the detected shadow. The following decoding and refinement stage are part of filtering the shadow and will be discussed in the next section.

Regarding shadow filtering, either the pixels are replaced with a predetermined value based on thresholding, or a model is applied to estimate the restored values of the pixels. After using morphological operations and filtering rules to detect shadows in an image, Gad, Yaghi, Alkhedher, *et al.* [105] proposed to replace the identified pixels with background pixels through horizontal and vertical scanning. In Figure 9 an example is shown of the shadow removal process. Abdusalomov and Whangbo [106] also used morphological operations and filtering rules to detect shadows in an image. In this case the detected pixels are simply removed by setting the classified shadow pixels in the binarized image to zero. Then to remove the last undetected pixels, simple morphological restoration is applied to fill in small holes. This results in a quite effective removal of the shadows. Salim, Cheng, and Xiao [107] used the Buffer area method to filter the shadows for unstructured road detection. After having detected the shadows through morphological operations and filtering, the shadow is removed through using an area around the shadow, called the buffer area, which compensates the shadow using the mean and variances of the shadow region.

Fan, Han, and Li [104] both used machine learning in the form of CNN for the detection of shadows and filtering. After having used the encoder network to detect the shadow, a decoder needs to be used for the filtering of the shadow. The decoder network consist of a deconvolution network which is mirror-symmetric with the convolutional network. For the activation function the parameter rectification linear unit (PReLU) was used instead of rectified linear unit (ReLu). The output of the deconvolution network is a predicted shadow mask factor α which if multiplied with a shadow-free image would create an image with a shadow. This factor can thus be used to restore a shadow-free image. Finally the resulting shadow mask factor is refined through using a small CNN for local detail correction.

Table 6 gives an overview of the different detection and filtering methods for shadow handling and their reported performance.

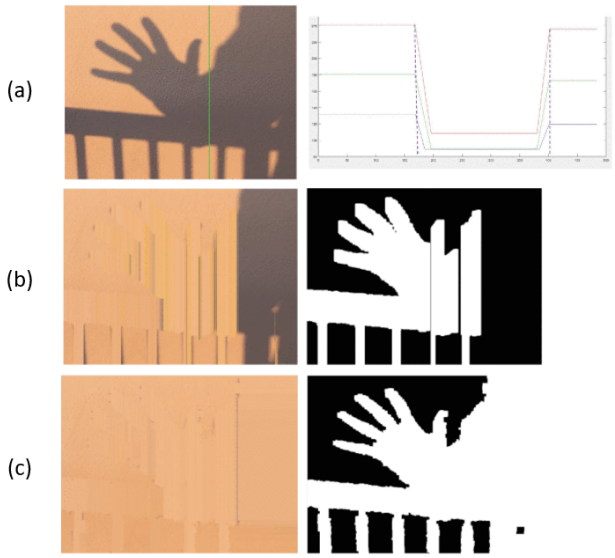


FIGURE 9: Shadow filtering process based on replacing shadow pixels with the background consisting of: (a) shadow detection, (b) horizontal scanning, and (c) vertical scanning. [105]

Study name	**Detection**	**Filtering**	**Accuracy [%]**	**Speed [ms]**
Silva, Carneiro, Doth, *et al.* [108]	Morphological operations and filtering rules: RGB conversion to CIELCh color space (polar form of CIE L*a*b* color space). Smoothing of L and h channels. Calculation modified Specthem ratio. Shadow segmentation thresholding based on Specthem ratio and Otsu's method. Erosion and dilation for the removal of shadow regions.	Shadow removal model using the estimation of illumination ratio. Relighting of shadow regions by multiplying pixels with the illumination ratio.	93	35 - 36
Xiao, She, Xiao, *et al.* [109]	GMM: probability density function composed of a set of Gaussian models. Global brightness detection is used as reference	-	-	-
Macedo, Nascimento, and Souza [101]	Morphological operations and filtering rules: RGB conversion to: - grayscale, - YCrCb, and - CIEL*a*b*, Discard non-shadow regions with high intensity in each channel. Otsu binarization. Segmentation based on thresholds.	-	-	13
Gad, Yaghi, Alkhedher, *et al.* [105]	Morphological operations and filtering rules: noise reduction using a Gaussian filter + color variation and structure through vertical & horizontal scanning; The mean values of the group of pixels are used as knee points where drop points exist.	Replace pixels with background based on horizontal and vertical scanning	96	35.7
Fan, Han, and Li [104]	CNN	CNN	-	-

Yuan, Wu, and Cheng [110]	Morphological operations and filtering rules: RGB conversion to HSV. According to the shadow's characters such as the color variation and the structure, two shadow detection algorithms which are respectively based on the RGB color model and the HSV color model.	-	RGB: 75 - 92; HSV: 78 - 84	RGB: 15 - 25; HSV: 50 - 65 (Including RGB to HSV conversion)
Abdusalomov and Whangbo [106]	Morphological operations and filtering rules: Improve contrast: substracting two consequent grayscale values of the image. Upgrade shadowy scenery and low-level pixel intensities caused by non-natural illumination sources. Median filter for median noise removing + Color variation and structure for geometry feature information.	Morphological restoration through filling small gaps and holes based on thresholds	92	270
Niranjil Kumar and Sureshkumar [102]	Fuzzy logic: background subtraction, edge detection, image denoising. Fuzzy logic through two membership functions indicating backround, shadow or edge pixel.	-	-	-
Salim, Cheng, and Xiao [107]	Morphological operations and filtering rules: RGB conversion to HSV. The ratio-image (H +1)/(V +1) is obtained through a spectral ratio technique. Otsu segmentation method is used over the histogram of the ratio image to determine the segmentation threshold for the ratio image. The image is segmented and the shadow region image is obtained. The segmented image is filtered by median filter for noise removal, then processed by morphological operations consisting of erosion and dilation techniques to extract the shadow region.	Buffer area method	-	-
Shedlovska and Hnatushenko [111]	Morphological operations and filtering rules: RGB conversion to HSV. Normalized saturation-value difference index is used for obtaining shadow mask. Otsu thresholding used for extration shadow.	Shadow formation model consisting of two types of illuminance: direct light and reflected illuminance. Non-shadowed regions are lit by both types of illuminance. Shadowed regions are lit only by reflected light.	-	-
Oh, Min, and Heo [112]	Morphological operations and filtering rules: gradient vector points are are used with edge detection. the edge detection and quantity is the magnitude of this vector, named the gradient.	The shadow removal by replacing the corresponding pixel values with those of the background.	-	-
Li, Li, Tian, *et al.* [103]	GMM: Firstly, moving vehicles and their following shadows have the same motion. They are detected as foreground objects through edged mixture Gaussian modelling method. The shadow is detected by the brightness distortion, with knowing that the shadow is usually darker than background.	-	-	-
Yan, Hu, Su, *et al.* [113]	Morphological operations and filtering rules: Fore ground removal + calculate brightness distortion and evaluate with thresholds.	Literal removal of all nonshadow pixels	97	-

TABLE 6: Overview of shadow handling methods and their performance.

6) Vibration handling methods

Vibration handling is a research field that has been researched a lot. One field that has been researched mainly consists of general video stabilization for systems that involves a lot of movement like handheld devices, moving robots, or drones for example. Other fields apply vibration handling methods for Surveillance or monitoring systems. All systems work by first detecting the movement or vibration camera and then apply filters to compensate for this.

Vibration detection methods

In terms of vibration detection, the most used method is through feature matching. Other popular methods consist of using a Kanade–Lucas–Tomasi (KLT) feature tracker and block matching. There are many ways to apply feature matching for the detection of movement. The following feature matching methods are the most common.

Rodriguez-Padilla, Castelle, Marieu, *et al.* [114] used an edge detection in the form of a canny edge detector to extract features. By converting the image into an edge image, the computation costs can be greatly reduced without losing accuracy. After creating the edge image, keypoints can be extracted and matched against themselves by estimating the two-dimensional displacement with respect to a reference frame. Figure 10 shows the process of vibration estimation and filtering.

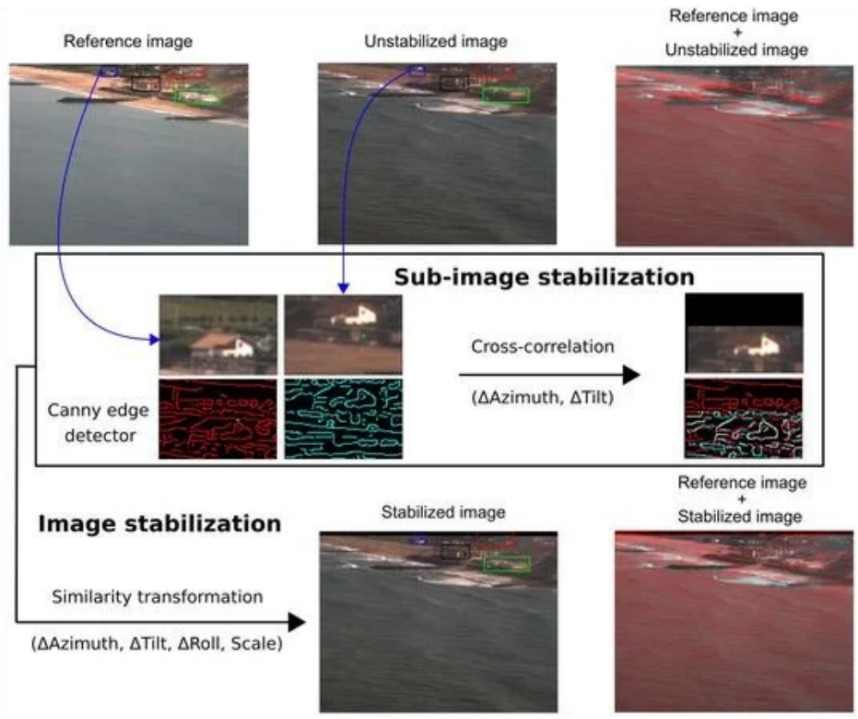


FIGURE 10: Process of vibration estimation and filtering through motion compensation.[114]

Rasti and Sadeghi [115] simply used the SIFT algorithm for feature point extraction. After this, Random sample consensus (RANSAC) was applied to remove outliers and estimate an affine transformation matrix.

Alam, Singh, and Abeyratne [116] used the Features from Accelerated Segment Test (FAST) algorithm for feature point extraction. After this the Fast Retina Keypoint (FREAK) algorithm is used for matching based on the hamming distance and the lowest cost path. Finally, RANSAC is used to filter out outliers and estimate an affine transformation matrix.

Aguilar and Angulo [117] did something similar but instead used the SURF algorithm for feature detection. After this, RANSAC is applied to filter out outliers and estimate a homography transform matrix. This projective transformation is

then used to estimate an affine transformation model to compute a mean rotation angle of the image.

Okade and Biswas [118] proposed to use MSER feature matching for video stabilization. To do this, an ellipse is chosen as a measurement region. Then it is assumed that the center of gravity the ellipse is invariant. Based on this the covariance matrix of the measurement region Q is calculated, where the eigenvectors are the direction of the ellipse axis. Finally, MSER feature matching is used to estimate the motion of the video.

Li, Xu, and Zhang [119] proposed to use the Oriented FAST and rotated BRIEF (ORB) algorithm for corner extraction. ORB consists of a FAST corner detector in combination with a Binary Robust Independent Elementary Features (BRIEF) feature description. The extracted features are matched based on the Hamming distance. The Hamming distance consist of the ratio of the minimum distance and the second minimum distance below a certain threshold. After the feature matching, the motion is estimated based on a 2D affine model. This is done through triangle matching in combination with RANSAC to update the model.

Umrikar and Tade [120] proposed to use block matching for the estimation for video stabilization. Block matching uses a series of frames and tries to locate and match macro blocks. The current frame is divided into macro blocks which are compare with a candidate block. The similarity is calculated through the Mean Square Error (MSE) method. Where based on the lowest MSE, the block is matched and the movement is detected.

Another popular method is to use the off-the-shelf KLT feature tracker to extract features. This is often combined with other feature matching methods. Hamza, Hafiz, Khan, *et al.* [121] for example used RANSAC to filter out outliers and estimate a transformation matrix describing the movement.

Vibration filtering methods

In terms vibration filtering there are 3 ways to filter vibrations, dependent on the field it is used. either a transformation matrix is created and used to transform the image, or a filter is applied to smooth the measured motion path, or the smear created by the motion is tackled through a smear removal method.

In terms of the transformation matrix, this is often already created through the motion detection method. To summarize, most of the time RANSAC is used for feature matching to remove outliers and estimate the transformation matrix. The resulting matrix is most of the time an affine transformation matrix, and sometimes a homography transformation matrix. Souza and Pedrini [122] however took a different approach and used a similarity matrix for the transformation of the image. This similarity matrix was created by combining multiple feature detection algorithms consisting of MSER, SIFT, and STAR, and the applying RANSAC to filter out outliers and estimate the transformation matrix. After having estimated the transformation matrix, the image is transformed through geometric transformation. Koh, Sim, and Kim [123] proposed to use mesh grid warping with a least squares optimization method to transform the image. First the features were extracted through a KLT feature tracker. After this, the rolling-free feature trajectories were smoothed through minimizing a cost function based on the least squares method.

With regards to smoothing the motion path, this is achieved by first detecting and measuring the motion path, consisting of the displacement of the features. This motion path consists of intended motion and unintended motion caused by vibrations. These unintended motions are then filtered out by applying some filter. A common approach is applying a Kalman filter to filter out the vibrations. Tanakian, Rezaei, and Mohanna [124] proposed to use adaptive fuzzy Kalman filtering for video stabilization. For the motion path detection and estimation, first local motion vectors (LMV) are generated through block matching. Then the global motion vector (GMV) is estimated based on fuzzy clustering. This is done by creating a histogram of valid LMV. The GMV is the position in the histogram of the cluster with the maximum votes. After having found the motion path, a Kalman filter is applied to smooth out the motion path and used to remove the unintentional motions. Other methods use Gaussian discrete filter (Okade and Biswas [118]), Lowpass filter (Favorskaya, Jain, and Buryachenko [125]), Alpha trimmed filter (Minh and Hong [126]), or any combination of these or other filters. Figure 11 gives an example of a filtered motion path. After the motion path has been smoothed, sometimes the empty areas created with the smoothing need to be resolved. Sunny, Srikanth, and Eswar [127] proposed to use mosaicing for this. With mosaicing information from previous frames is used to match and fill in empty spaces.

In some researches not only the image is stabilized, but also the blur created by the movement is restored. Saxena,

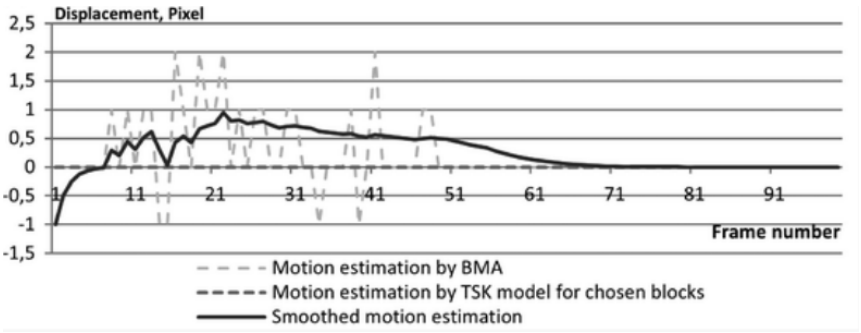


FIGURE 11: Filtering of the motion path of an object in a video.[125]

Verma, Ghosh, *et al.* [128] proposed to use a Point Spread Function (PSF) for smear removal. First the global motion is estimated through a SURF algorithm in combination with affine transformation. This affine transformation is then used to recover the geometric distortion and used to estimate the unintentional motion vectors area. After this a point spread function is used to remove the motion blur. This method requires the global translation noise and rotational noise as inputs to deblur the image.

Table 7 gives an overview of the different detection and filtering methods for vibration handling.

Study name	Detection	Filtering
Rodriguez-Padilla, Castelle, Marieu, *et al.* [114]	Canny edge detector + Keypoints are matched against themselves after computing their two-dimensional displacement with respect to a reference frame.	Geometric transformation based on estimated transformation matrix: Pairs of keypoints are subsequently used as control points to fit a geometric transformation in order to align the whole frame with the reference image.
Alam, Singh, and Abeyratne [116]	FAST algorithm for feature points detection. FREAK for matching based on hamming distance and lowest cost paths; RANSAC for outlier filtering and Affine transform estimation.	Affine transformation
Aguilar and Angulo [117]	SURF feature detection; RANSAC + Homography transform; The projective transformation is used to estimate the affine model. With this a mean rotation angle of the image is computed.	Motion control
Hamza, Hafiz, Khan, *et al.* [121]	KLT feature tracker + RANSAC	-
Javadi, Kadim, Woon, *et al.* [129]	SURF + RANSAC	Homography transfomation
Chereau and Breckon [130]	RGB conversion to Gray. Laplacian of Gaussian (LoG) convolution filter + block matching. Outlier removal through threshold of similarity measure indicative of an aperture effected block and threshold of minimum correlation distance. Global motion vector (GMV) is the mean of all local motion vectors (LMV).	Smoothing of motion path: temporally weighted filter for smoothed path. Correlation vector for zooming and uniform cropping.
Junwei, Jianjun, Jiaqi, *et al.* [131]	GFTT feature point detection + sparse optical flow feature point matching	Smoothing of motion path: Kalman filter + Gaussian discrete filter
Sunny, Srikanth, and Eswar [127]	Global and local motion estimation based on two builders: MotionEstimatorRansacL2Builder estimates global motion between two 2D point clouds and also minimizes the L2 error. The class Motion Estimator L1 Builder is also use to estimate 2D global motion by minimizing the L1 error which provides the least absolute deviations.	Smoothing of motion path: motion smoothing through TwoPassStabilizer. Image deblurring through transfering/interpolating blurry pixels from sharper frames. Mosaicking is used to fill the empty areas of the video caused by stabilization of the video.

Verma, Andri, Singh, *et al.* [132]	RGB to gray; Features from Accelerated Segment Test (FAST) algorithm are used to detect the features of the frame for the estimation of motion vectors. Iterative Dichotomiser 3 (ID3) algorithm is used to train the decision tree classifier. RANSAC is used for estimating transformation.	Transformation matrix is calculated and used. Synthesizer module combines all stabilized frames to produce online stabilized video.
Umrikar and Tade [120]	Feature extraction: Block Matching.	Affine transformation
Tanakian, Rezaei, and Mohanna [124]	Block matching: Block-based motion estimation to generate Local Motion Vectors (LMV). LMV validation through smoothness test and complexity test. Global Motion Vector (GMV) estimation through clustering process: Construct histogram of valid LMV. position of cluster with maximum votes is considered GMV.	Smoothing of motion path: Kalman filter used to estimate intentional motion and derive unintentional motion.
Saxena, Verma, Ghosh, *et al.* [128]	Global motion estimation through SURF algorithm for feature detection. Affine transformation is estimated using inliers feature points. Affine transformation is used to recover geometric distortions in an image. After this the unintentional motion vectors area is estimated	Smear removal through Point Spread Function.
Rasti and Sadeghi [115]	SIFT features for motion estimation. 2D affine model for motion model; RANSAC for outlier removal.	Smoothing of motion path: Motion smoothing through spatio temporal Gausian lowpass filter. Frame mosaicing for the reconstruction of undefined regions.
Qu and Song [133]	L1 used for feature tracking and 2D linear motion model fitting to compute the original camera path. the optimization is related to L1 optimization, which minimizes the first, second and third derivatives of the resulting camera path with some linear constraints. Both L1 norm of smooth path and L2 norm of the difference between smoothed and original camera paths are optimized.	-
Okade and Biswas [118]	MSER feature extraction. Choose elipse as measurement region. Choose center of gravity of ellipse as an invariant. Calculate the covariance matrix of the measurement region. The center of gravity is the center of the elipse. The eigenvectors are directions of the ellipse axis. MSER feature matching is used for motion esitmation.	Smoothing of motion path: motion smoothing through Gaussian filter
Minh and Hong [126]	Extract feature points with KLT tracking. Cluster directional features. Cluster magnitude features.	Smoothing of motion path: Smooth motion trajectory through alpha-trimmed filter
Li, Xu, and Zhang [119]	Corner extration with ORB algorithm: FAST corner detector + BRIEF feature discription. Feature matching using Hamming distance: ratio of minimum distance and second-minium minimum distance below a certain threshold. Motion estimation base on 2D affine model. Triangle matching in combination with RANSAC.	Stabilized filter through mosaicing and Kalman filter.
Koh, Sim, and Kim [123]	KLT feature tracker. Smoothing of the rolling-free feature trajectories through minimizing the cost function;	Mesh grid warping for stabilization with minimizing cost-function based on least squares.
Favorskaya, Jain, and Buryachenko [125]	Fast block matching algorithm (BMA) for local motion estimation. Improving accuracy using Takagi-Sugeno-Kang model. Estimate global motion with clustering model.	Smoothing of motion path: low-pass filter and Kalman filter
Fang, Ma, and Cao [134]	KLT tracking + feature point classification on homography consistency to local subspace video stabilization.	Homography transfomation
Souza and Pedrini [122]	MSER + SIFT + STAR + RANSAC	Similarity matrix
Avramelos, Wallendael, and Lambert [135]	Motion vectors ectraction through High Efficiency Video Coding (HEVC).The motion vectors are divided into grid cells and the median is calculated for every block. Hampel filter for remving outliers. If a vector differs from the median by more than a given number of MADs (Median Absolute Deviation), it is replaced with the median. Finally, the median is calculated again resulting in the global motion vector (GMV);	Smoothing of motion path: fast fourier transform with low-pass filter; After this motion compensation is applied according to the smoothened curver;

Auysakul, Xu, and Pooneeth [136]	Inertial Measurement Unit (IMU) for large motions + KLT tracking for small motions	Homography transformation + Kalman filtering
Gunawan, Han, and Lee [137]	SURF feature detection. Brute force matching with brute force KNN. Homography through RANSAC	Homography transformation

TABLE 7: Overview of vibration handling methods and their performance.

7) Occlusion handling methods

Regarding occlusion handling, the main field where it is applied is in vehicle tracking, multi-person tracking and object tracking. In almost all cases the object is being tracked by some type of tracking model. To deal with occlusions, either occlusions are detected through a detection method and an occlusion filtering method is applied, or the effect of occlusion is directly limited through a filtering method.

Occlusion detection methods

A common method of occlusion detection is the sudden absence of tracking information or the merging of two targets. Zhou, Zhong, Zhang, *et al.* [138] proposed an occlusion handling method for person tracking. In this case, the motion model is based on a Kalman filter. When the target is occluded for a long time, the appearance and motion information are unavailable, and the occlusion is detected. When this happens the information just before the occlusion is kept as a reference to reidentify the target. Wen, Xu, and Zhan [139] detect occlusions when two targets overlap. In this situation, the bounding the boxes of the two objects are merged by the object tracker, indicating an occlusion.

Other methods are based on some type of thresholding to detect occlusion. Sarwar, Rao, and Khan [140] proposed to use Normalized Cross Correlation (NCC) to detect an occlusion. In this case, a candidate region is compared to a template image through NCC. When the correlation value is below a certain threshold, the occlusion is detected. Another thresholding-based method is by detecting the size of a tracked object. Velazquez-Pupo, Sierra-Romero, Torres-Roman, *et al.* [141] used occlusion detection for vehicle detection. In this the case the width of a detected vehicle would be compared to a threshold, defined by the width of two driving lanes. If the width of the vehicle is larger than the width of two driving lanes, occlusion is detected. Something similar was done by Xiong and Li [142], where instead of looking at the width of the vehicle, the number of back windows were detected. The image would be converted to a brightness curve, where the bumps would indicate back car windows. The amount of back car windows would indicate the number of cars, and thus can be used to detect occlusion. Dong, Shen, Yu, *et al.* [143] proposed to use the occlusion distance to detect occlusions. When the occlusion distance is above a certain threshold, the occlusion is detected. Huang, Ju, Hu, *et al.* [144] showed that with a sharp drop of the Peak-to-Correlation Energy (APCE) of a tracked person can indicate that the target is occluded. This APCE is calculated with the correlation filter response peak.

Another method for occlusion detection is through machine learning by training a CNN. Sarkar, Venugopalan, Reddy, *et al.* [145] proposed to first convert an image to an edge map and then use a trained CNN to directly detect occlusions.

Occlusion filtering methods

In terms of occlusion filtering, either the location of the tracked object is predicted through some type of filtering, or the effect of occlusions are limited by training a machine learning algorithm to detect an object even with partial occlusion. After Sarwar, Rao, and Khan [140] detect the occlusion through template matching, the object tracker starts predicting the location of the object through a Kalman filter. Heimbach, Ebadi, and Wood [146] also used a Kalman filter to predict the location of the object. But rather than detecting occlusions and then activating the occlusion handling method, the Kalman filter is applied continuously during the tracking of an object. Figure 12 gives an example of the application of a Kalman filter to predict the position of an occluded target.

FIGURE 12: Example of the use of a Kalman filter for the filtering of occlusions during person tracking. The detection window is shown in purple. The ground truth is indicated in red. The HOG estimation is indicated in green. The Kalman filter estimation is shown in yellow. [146]

Lomaliza and Park [147] proposed to use machine learning in the form of a CNN to deal with occlusions in object tracking. In this case, the CNN was trained with sub-parts of an object. Through this, even if the object is partially occluded, the CNN-based tracker can still recognize and track the object. Something similar was done by Cong, Tian, Feng, *et al.* [148], where an occluded dataset was created such that the object still can be detected during self-occlusion through filtering and Hough voting. The Hough voting is used to create candidates, after which nearest neighbor search is used to filter out false hypothesis. After this, the pose estimation of the object is refined through an iterative closest point strategy.

Table 8 gives an overview of the different detection and filtering methods for occlusion handling.

Study name	**Detection**	**Filtering**
Zhou, Zhong, Zhang, *et al.* [138]	Kalman Filter based motion model. When a target is occluded for a long time, both its appearance and motion cues are unavailable. Appearance cues are stored just before the occlusion, and used to re-identify the targets.	-
Xiong and Li [142]	Track backwindow ofvbehicles with brightnesscurve. Amount of backwindows show amount of cars.	-
Wen, Xu, and Zhan [139]	RGB conversion to HSV. When two tracked objects collide, they merge in one bounding box	-
Velazquez-Pupo, Sierra-Romero, Torres-Roman, *et al.* [141]	Occlusion is detected when the width of a vehicle is greater than one lane inside the region of interest and occlusion is detected when the width of a vehicle is greater than two lanes for the detection line	-
Van Pham and Lee [149]	-	Partial occluded dataset + SVM
Sarwar, Rao, and Khan [140]	Occlusion is detected by dynamically thresholding the Normalized Cross Correlation (NCC) between templates and selected candidate regions.	Once the object is declared occluded, the tracker starts predicting the location of the object using a Kalman filter.
Sarkar, Venugopalan, Reddy, *et al.* [145]	CNN	-

Ramisetty, Qu, Aktar, *et al.* [150]	Occlusion detection based on spatial impairment rate and temporal impairment rate. This is based on intensity difference of pixels	-
Lomaliza and Park [147]	-	CNN sub-part detection to reduce effect of occlusion.
Lin and Chang [151]	Occlusion is detected when two tracked objects start to overlap and cannot be detected any more	Once the object is declared occluded, the tracker starts predicting the location of the object using Kalman filter.
Huang, Ju, Hu, *et al.* [144]	A sharp drop of the Peak-to-Correlation Energy (APCE) of a tracked person indicates that the target is occluded. APCE is calculated with the correlation filter response peak (Fmax).	-
Heimbach, Ebadi, and Wood [146]	-	Kalman filter for noise surpression.
Fang, Zhao, Yuan, *et al.* [152]	Normalized area of intersection is used for matching of a template and an object by constructing a bipartite association graph between the template and the object. When two tracked objects merge, occlusion is detected.	-
Dong, Shen, Yu, *et al.* [143]	Occlusion is detected when the occlusion distance larger than a threshold.	-
Cong, Tian, Feng, *et al.* [148]	-	Create dataset with occlusion for nearest neighbor search. Create candidates based on Hough voting. Filter the false hypotheses and refine the pose estimation via the iterative closest point strategy.

TABLE 8: Overview of occlusion handling methods and their performance.

IV Discussion

In subsection 1 to subsection 3 an overview is given of different methods to recognize different types of indicators. In general, the standard method consists of an indicator localization method and indicator state estimation method. Depending on the application, environment, and type of indicator, different methods are more suitable. For the recognition of light indicators, the most common method is to apply morphological operations and filtering rules for the detection of the traffic light. Then depending on the type of light indicator different methods need to be used. If only the light color needs to be determined, either simple thresholding of the color distribution or fuzzy logic clustering can be used. When the direction of the light also needs to be determined, HOG-features in combination with SVM can be used. In all cases, the color of the light indicator is determined, while this is not entirely necessary. In case of traffic lights, the position of the light with regards to the other lights in combination with the intensity can also determine the state of the traffic light. For the recognition of characters, machine learning methods like CNNs or SVMs give the highest accuracy. This however does require a large training dataset to train the models. Many datasets for character recognition are readily available. If in a special case this is not available, then edge detection methods for the localization and template matching for the detection are a good alternative. In terms of speed, these methods can approach the same range as machine learning based methods. The downside is that they are less robust. It is possible to increase robustness by applying noise handling methods, but this will reduce the recognition speed. For the recognition of symbols, many different methods are available. Depending on the application, environment, and type of indicator, different methods are more suitable. CNNs and SVMs have the highest potential just as within character recognition. But again, this can only be used if a large enough training dataset is available. If this is not the case, morphological operations and filtering rules can be applied to locate the symbol. Furthermore, many different feature extraction methods can be applied depending on what suits the application. To increase the robustness a combination of different types of features can be beneficial. This increased complexity will however affect the computation time. For the actual state estimation, template matching can be a solid method. Again, the robustness of the non-machine learning methods is often lower than those of the machine learning methods. To counter this, different noise handling methods can be applied, which will reduce the recognition speed.

In subsection 4 to subsection 7an overview of different noise handling methods was given. Depending on the application

different noise handling methods can be used. For the detection of both shadows and reflections, morphological operations and filtering rules are popular. Then to filter these noises, either a more complex restoration model can be used, or a simpler replacement of the noisy pixels can be applied. The complexity often will increase the accuracy, but will be at the cost of more computation time. For the vibration handling methods, the type of application determines the suitable method. If dealing with a stationary camera that has unintentional movements, the estimation of a transformation matrix can be useful. If dealing with a moving camera where the unintentional movement needs to be filtered, but the intentional movement needs to be kept, a motion path filtering method can be applied. Finally, if one deals with blurry images due to movement, a deblurring method can be used. Not every method can be applied in real time, some methods that for example measure the motion path, need to have data from the past to compensate for the present. This means that it becomes more difficult to realize real time vibration filtering, since more data needs to be gathered in a shorter time frame to estimate the current motion path. An interesting approach to reduce the computation time while keeping the same accuracy is to simplify the image being tracked by converting it to for example a Sobel edge image. This method however is dependent on the type of object being tracked and might not always be applicable. For the occlusion handling methods, it depends on the application if this can be used or not. In all cases, the occlusion can be detected through various methods. The filtering of the occlusion and estimation of the ground truth is only possible for certain types of applications. If the goal of the image recognition method is the tracking of a target, the position of the target can be estimated. If there is a partial occlusion, a state of an indicator can still be detected if enough visual cues are available, and the indicator detection model is designed to handle them. In the case of full occlusion or the low availability of visual cues, the state cannot be accurately predicted. In that case the only sensible thing to do is to show an error and warn the user of the occlusion.

The methods of machine learning in the form of CNNs and SVMs have a huge potential with regards to reading all types of indicators. These however require big training data sets that might not always be available. In the case of a new application, either new data sets need to be created first, or other non-machine learning methods need to be used. To create a new dataset, it could be interesting to create template images and apply various types of digital noise to create a larger dataset for training the model. By expanding the small template dataset with artificial noisy images, the robustness and thus the accuracy of a machine learning model could increase.

V Conclusion

The development of real time image recognition methods is important for the monitoring, automation, and smartification of systems. This paper presents an overview of the state of arts in the real time recognition of light indicators, characters, and symbols. First, the most common localization methods are discussed. Second, the most common state estimation methods are given. Then the methods are compared based on their performance in accuracy and speed. Furthermore, this paper provides an overview of most common noise handling methods that can be applied to improve the accuracy in noisy environments. Ongoing challenges consist of increasing the robustness in challenging noisy environments while maintaining real time performance. As there are many methods to recognize each type of indicator, there needs to be more performance evaluations to compare each real time method. In this paper different methods are compared to each other with their own application, environment and hardware that is used. This can highly affect the performance, especially in terms of speed. Therefore, a standardized test to evaluate the performance would be advised. Currently, each method only evaluates one type of indicator, while for automation multiple types of indicators are used. Therefore, the next step would be to combine multiple methods to read each type of indicator to determine the total state of the object or system being measured. This can be either achieved through some type of machine learning in the form of for example CNNs or SVMs, or a combination of different non-machine learning methods. When dealing with a new application, machine learning becomes a bit more difficult due to the lack of datasets. That is why for the first version of a new image recognition system, non-machine learning methods could be useful.

Bibliography

[1] H. Fujiyoshi, T. Hirakawa, and T. Yamashita, "Deep learning-based image recognition for autonomous driving," *Magazine*, vol. 43, no. 4 Number, pp. 244–252, Dec. 2019, ISSN: 0386-1112. DOI: https://doi.org/10.1016/j.iatssr.2019.11.008.

[2] M. Efimenko, A. Ignatev, and K. Koshechkin, "Review of medical image recognition technologies to detect melanomas using neural networks," *Magazine*, vol. 21, no. 11 Number, p. 270, Sep. 2020, ISSN: 1471-2105. DOI: 10.1186/s12859-020-03615-1.

[3] M. Bakhtan, M. Abdullah, and R. A, "A review on license plate recognition system algorithms," in *Book A review on License Plate Recognition system algorithms*. 2016, ch. Chapter. DOI: 10.1109/ICICTM.2016.7890782.

[4] S. Rosyda and T. Purboyo, "A review of various handwriting recognition methods," *Magazine*, no. Number, Jan. 2018.

[5] I. Balki, A. Amirabadi, J. Levman, *et al.*, "Sample-size determination methodologies for machine learning in medical imaging research: A systematic review," *Magazine*, vol. 70, no. 4 Number, pp. 344–353, Nov. 2019, ISSN: 0846-5371. DOI: 10.1016/j.carj.2019.06.002.

[6] P. S. Fleming, D. Koletsi, and N. Pandis, "Blinded by prisma: Are systematic reviewers focusing on prisma and ignoring other guidelines?" *Magazine*, vol. 9, no. 5 Number, e96407, 2014. DOI: 10.1371/journal.pone.0096407.

[7] M. Diaz-Cabrera, P. Cerri, and P. Medici, "Robust real-time traffic light detection and distance estimation using a single camera," *Magazine*, vol. 42, no. 8 Number, pp. 3911–3923, 2015, ISSN: 09574174. DOI: 10.1016/j.eswa.2014.12.037.

[8] J. Ying, J. Tian, and L. Lei, *Traffic light detection based on similar shapes searching for visually impaired person*. Institute of Electrical and Electronics Engineers Inc., 2015, ISBN: 9781479917174. DOI: 10.1109/ICICIP.2015.7388200.

[9] W. Wang, S. Sun, M. Jiang, *et al.*, "Traffic lights detection and recognition based on multi-feature fusion," *Magazine*, vol. 76, no. 13 Number, pp. 14 829–14 846, 2017, ISSN: 13807501. DOI: 10.1007/s11042-016-4051-5.

[10] T.-P. Tran, C. Pham, T. Nguyen, *et al.*, *Real-Time traffic light detection using color density*. Institute of Electrical and Electronics Engineers Inc., 2016, ISBN: 9781509027439. DOI: 10.1109/ICCE-Asia.2016.7804791.

[11] K. Vishal, C. Arvind, R. Mishra, *et al.*, *Traffic light recognition for autonomous vehicles by admixing the traditional ML and DL*. SPIE, 2019, vol. 11041, ISBN: 9781510627482. DOI: 10.1117/12.2523105.

[12] Q. Chen, Z. Shi, and Z. Zou, *Robust and real-time traffic light recognition based on hierarchical vision architecture*. Institute of Electrical and Electronics Engineers Inc., 2014, ISBN: 9781479958351. DOI: 10.1109/CISP.2014.7003760.

[13] M. Pervej, S. Das, M. P. Hossain, *et al.*, "Real-time computer vision-based bangla vehicle license plate recognition using contour analysis and prediction algorithm," *Magazine*, vol. 21, no. 4 Number, 2021, ISSN: 02194678. DOI: 10.1142/S021946782150042X.

[14] Wahyono and K. H. Jo, *Information retrieval of LED text on electronic road sign for driver-assistance system using spatial-based feature and Nearest Cluster Neighbor classifier*. Institute of Electrical and Electronics Engineers Inc., 2014, ISBN: 9781479960781. DOI: 10.1109/ITSC.2014.6957744.

[15] A. Jain and J. Sharma, *Classification and interpretation of characters in multi-application OCR system*. Institute of Electrical and Electronics Engineers Inc., 2014, ISBN: 9781479946754. DOI: 10.1109/ICDMIC.2014.6954231.

[16] A. A. Ahmed and S. Ahmed, "A real-time car towing management system using ml-powered automatic number plate recognition," *Magazine*, vol. 14, no. 11 Number, 2021, ISSN: 19994893. DOI: 10.3390/a14110317.

[17] T. Ajanthan, P. Kamalaruban, and R. Rodrigo, *Automatic number plate recognition in low quality videos*. 2013, ISBN: 9781479909100. DOI: 10.1109/ICIInfS.2013.6732046.

[18] L. Zheng, X. He, B. Samali, *et al.*, "An algorithm for accuracy enhancement of license plate recognition," *Magazine*, vol. 79, no. 2 Number, pp. 245–255, 2013, ISSN: 00220000. DOI: 10.1016/j.jcss.2012.05.006.

[19] M. Fu, N. Chen, X. Hou, *et al.*, "Real-time vehicle license plate recognition using deep learning," in *4th International Conference on Signal and Information Processing, Networking and Computers, ICSINC 2018*, S. Sun, Ed., vol. 494, Springer Verlag, 2019, pp. 35–41, ISBN: 9789811317323. DOI: 10.1007/978-981-13-1733-0_5.

[20] A. Ahmed Biyabani, S. A. Al-Salman, and K. S. Alkhalaf, *Embedded real-time bilingual ALPR*. Institute of Electrical and Electronics Engineers Inc., 2015, ISBN: 9781479965328. DOI: 10.1109/ICCSPA.2015.7081311.

[21] M. A. Massoud, M. Sabee, M. Gergais, *et al.*, "Automated new license plate recognition in egypt," *Magazine*, vol. 52, no. 3 Number, pp. 319–326, 2013, ISSN: 11100168. DOI: 10.1016/j.aej.2013.02.005.

[22] M. Salahshoor, A. Broumandnia, and M. Rastgarpour, *An intelligent and real-time system for plate recognition under complicated conditions*. IEEE Computer Society, 2013, ISBN: 9781467361842. DOI: 10.1109/IranianMVIP.2013.6779988.

[23] L. R. Bague, R. J. L. Jorda, B. N. Fortaleza, *et al.*, "Recognition of baybayin (ancient philippine character) handwritten letters using vgg16 deep convolutional neural network model," *Magazine*, vol. 8, no. 9 Number, pp. 5233–5237, 2020, ISSN: 23473983. DOI: 10.30534/ijeter/2020/55892020.

[24] S. P. Ramteke, A. A. Gurjar, and D. S. Deshmukh, "A streamlined ocr system for handwritten marathi text document classification and recognition using svm-acs algorithm," *Magazine*, vol. 11, no. 3 Number, pp. 186–195, 2018, ISSN: 2185310X. DOI: 10.22266/IJIES2018.0630.20.

[25] G. Kour and R. Saabne, *Fast classification of handwritten on-line Arabic characters*. Institute of Electrical and Electronics Engineers Inc., 2014, ISBN: 9781479959341. DOI: 10.1109/SOCPAR.2014.7008025.

[26] H. Ahn and H. J. Cho, "Research of automatic recognition of car license plates based on deep learning for convergence traffic control system," *Magazine*, no. Number, 2021, ISSN: 16174909. DOI: 10.1007/s00779-020-01514-z.

[27] S. Alghyaline, "Real-time jordanian license plate recognition using deep learning," *Magazine*, no. Number, 2020, ISSN: 13191578. DOI: 10.1016/j.jksuci.2020.09.018.

[28] A. A. A. Ali and M. Suresha, *Arabic Handwritten Character Recognition Using Machine Learning Approaches*. Institute of Electrical and Electronics Engineers Inc., 2019, vol. 2019-November, ISBN: 9781728108988. DOI: 10.1109/ICIIP47207.2019.8985839.

[29] M. Al-Jubouri and H. Abusaimeh, "Offline arabic handwritten isolated character recognition system using support vector machine and neural network," *Magazine*, vol. 95, no. 10 Number, pp. 2315–2322, 2017, ISSN: 19928645.

[30] L. Alzubaidi and G. Latif, "Real time license saudi plate recognition using raspberry pi," *Magazine*, vol. 8, no. 1 Number, pp. 42–47, 2019, ISSN: 22783091. DOI: 10.30534/ijatcse/2019/0981.12019.

[31] M. Y. Arafat, A. S. M. Khairuddin, and R. Paramesran, "A vehicular license plate recognition framework for skewed images," *Magazine*, vol. 12, no. 11 Number, pp. 5522–5540, 2018, ISSN: 19767277. DOI: 10.3837/tiis.2018.11.019.

[32] N. Awalgaonkar, P. Bartakke, and R. Chaugule, *Automatic License Plate Recognition System Using SSD*. Institute of Electrical and Electronics Engineers Inc., 2021, ISBN: 9781665433235. DOI: 10.1109/IRIA53009.2021.9588707.

[33] A. K. Bachchan, A. Gorai, and P. Gupta, "Automatic license plate recognition using local binary pattern and histogram matching," in *13th International Conference on Intelligent Computing, ICIC 2017*, D. S. Huang, K. H. Jo, and J. C. Figueroa-Garcia, Eds., vol. 10362 LNCS, Springer Verlag, 2017, pp. 22–34, ISBN: 9783319633114. DOI: 10.1007/978-3-319-63312-1_3.

[34] M. U. M. Bhutta, H. Mahmood, and H. Malik, *An intelligent approach for robust detection and recognition of multiple color and font styles automobiles license plates: A feature-based algorithm*. Institute of Electrical and Electronics Engineers Inc., 2014, ISBN: 9781479939022. DOI: 10.1109/ICALIP.2014.7009936.

[35] G. W. Chen, C. M. Yang, and T. U. Lk, *Real-Time License Plate Recognition and Vehicle Tracking System Based on Deep Learning*. Institute of Electrical and Electronics Engineers Inc., 2021, ISBN: 9784885523328. DOI: 10.23919/APNOMS52696.2021.9562691.

[36] N. Duan, J. Cui, L. Liu, *et al.*, "An end to end recognition for license plates using convolutional neural networks," *Magazine*, vol. 13, no. 2 Number, pp. 177–188, 2021, ISSN: 19391390. DOI: 10.1109/MITS.2019.2898967.

[37] J. Dun, S. Zhang, X. Ye, *et al.*, "System recognizing bahamian license plate with touching characters," *Magazine*, vol. 25, no. 6 Number, 2016, ISSN: 10179909. DOI: 10.1117/1.JEI.25.6.063009.

[38] C. Edlin, D. Kiantono, S. Martin, *et al.*, "Real-time license plate detection and recognition for odd-even plate rationing system," *Magazine*, vol. 14, no. 8 Number, pp. 751–759, 2020, ISSN: 1881803X. DOI: 10.24507/icicel.14.08.751.

[39] S. A. Elsaid, H. Alharthi, R. Alrubaia, *et al.*, "Arabic real-time license plate recognition system," in *Communications in Computer and Information Science*, A. Alfaries, H. Mengash, A. Yasar, *et al.*, Eds., vol. 1098 CCIS, Springer, 2019, pp. 126–143, ISBN: 9783030363673. DOI: 10.1007/978-3-030-36368-0_12.

[40] M. M. Farhad, S. M. Nafiul Hossain, M. I. Hossain, *et al.*, *A real time numeric character recognition system using artificial neural network*. IEEE Computer Society, 2013, ISBN: 9781479924653. DOI: 10.1109/ICAEE.2013.6750364.

[41] Y. q. Feng and L. Xia, "Image recognition and wireless transmission-based intelligent vehicle access control system," in *2012 2nd International Conference on Materials Science and Information Technology, MSIT 2012*, vol. 532-533, Xi'an, Shaan, 2012, pp. 934–938, ISBN: 9783037854389. DOI: 10.4028/www.scientific.net/AMR.532-533.934.

[42] F. Gao, Y. Ge, S. Lu, *et al.*, "Vehicle tire text reader: Text spotting and rectifying for small, curved and rotated characters," *Magazine*, no. Number, 2021, ISSN: 00189456. DOI: 10.1109/TIM.2021.3111973.

[43] J. Ge, L. Liu, J. Sun, *et al.*, "Automatic recognition of hot spray marking dot-matrix characters for steel-slab industry," *Magazine*, no. Number, 2021, ISSN: 09565515. DOI: 10.1007/s10845-021-01830-y.

[44] L. H. Godage and G. D. S. P. Wimalaratne, *Real-Time Mobile Vehicle License Plates Recognition in Sri Lankan Conditions*. Institute of Electrical and Electronics Engineers Inc., 2019, ISBN: 9781728137063. DOI: 10.1109/ICIIS47346.2019.9063258.

[45] M. Grębowiec and J. Protasiewicz, *A neural framework for online recognition of handwritten Kanji characters*. Institute of Electrical and Electronics Engineers Inc., 2018, ISBN: 9788394941970. DOI: 10.15439/2018F140.

[46] S. Guo, X. Shi, L. Bao, *et al.*, *A rotated character recognition method based on geometry correction*. Institute of Electrical and Electronics Engineers Inc., 2015, ISBN: 9781467386111. DOI: 10.1109/ICISSEC.2015.7370970.

[47] V. Jain, Z. Sasindran, A. Rajagopal, *et al.*, *Deep automatic licence plate recognition system*. Association for Computing Machinery, 2016, ISBN: 9781450347532. DOI: 10.1145/3009977.3010052.

[48] U. Jang, K. H. Suh, and E. C. Lee, "Low-quality banknote serial number recognition based on deep neural network," *Magazine*, vol. 16, no. 1 Number, pp. 224–237, 2020, ISSN: 1976913X. DOI: 10.3745/JIPS.04.0160.

[49] H. M. Jeon, V. D. Nguyen, and J. W. Jeon, *Real-time multi-digit recognition system using deep learning on an embedded system*. Association for Computing Machinery, 2018, ISBN: 9781450363853. DOI: 10.1145/3164541.3164641.

[50] Y. Jia, T. Gonnot, and J. Saniie, *Design flow of vehicle License Plate reader based on RGB color extractor*. IEEE Computer Society, 2016, vol. 2016-August, ISBN: 9781467399852. DOI: 10.1109/EIT.2016.7535290.

[51] S. Joseph and A. Hameed, *Online handwritten malayalam character recognition using LIBSVM in matlab*. Institute of Electrical and Electronics Engineers Inc., 2014, ISBN: 9781479966462. DOI: 10.1109/NCCSN.2014.7001151.

[52] P. M. Kamble and R. S. Hegadi, *Handwritten Marathi character recognition using R-HOG feature*, C. Elsevier B.V., 2015, vol. 45. DOI: 10.1016/j.procs.2015.03.137.

[53] A. Kathigi and K. H. Kariputtaiah, "Handwritten character recognition using unsupervised feature selection and multi support vector machine classifier," *Magazine*, vol. 14, no. 6 Number, pp. 301–310, 2021, ISSN: 2185310X. DOI: 10.22266/ijies2021.1231.27.

[54] G. Keerthi Prasad, I. Khan, N. R. Chanukotimath, *et al.*, *On-line handwritten character recognition system for Kannada using Principal Component Analysis Approach: For handheld devices*. 2012, ISBN: 9781467348041. DOI: 10.1109/WICT.2012.6409161.

[55] S. Khazaee, A. Tourani, S. Soroori, *et al.*, "An accurate real-time license plate detection method based on deep learning approaches," *Magazine*, vol. 35, no. 12 Number, 2021, ISSN: 02180014. DOI: 10.1142/S0218001421600089.

[56] H. Khosravi, "A sliding and classifying approach towards real time persian license plate recognition," *Magazine*, vol. 28, no. 1 Number, pp. 76–82, 2015, ISSN: 17281431. DOI: 10.5829/idosi.ije.2015.28.01a.10.

[57] R. Laroca, E. Severo, L. A. Zanlorensi, *et al.*, *A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector*. Institute of Electrical and Electronics Engineers Inc., 2018, vol. 2018-July, ISBN: 9781509060146. DOI: 10.1109/IJCNN.2018.8489629.

[58] R. C. Lee, K. C. Hung, and H. S. Wang, "Real-time vehicle license plate recognition based on scanning and 2d haar discrete wavelet transform," in *2nd International Conference on Engineering and Technology Innovation 2012, ICETI 2012*, vol. 284-287, Kaohsiung, 2013, pp. 2402–2406, ISBN: 9783037856123. DOI: 10.4028/www.scientific.net/AMM.284-287.2402.

[59] Y. Liu, H. Huang, J. Cao, *et al.*, "Convolutional neural networks-based intelligent recognition of chinese license plates," *Magazine*, vol. 22, no. 7 Number, pp. 2403–2419, 2018, ISSN: 14327643. DOI: 10.1007/s00500-017-2503-0.

[60] Y. Liu, D. Zhang, Y. Zhang, *et al.*, *Real-time scene text detection based on stroke model*. Institute of Electrical and Electronics Engineers Inc., 2014, ISBN: 9781479952083. DOI: 10.1109/ICPR.2014.537.

[61] Z. Liu, Y. Li, F. Ren, *et al.*, *SqueezedText: A real-time scene text recognition by binary convolutional encoder-decoder network*. AAAI press, 2018, ISBN: 9781577358008.

[62] V. Lukic, A. Makarov, and M. Spanovic, *An application of image point operators for OCR improvement in ALPR algorithm*. 2012, ISBN: 9781467329842. DOI: 10.1109/TELFOR.2012.6419305.

[63] M. P. Muresan, P. A. Szabo, and S. Nedevschi, *Dot Matrix OCR for Bottle Validity Inspection*. Institute of Electrical and Electronics Engineers Inc., 2019, ISBN: 9781728149141. DOI: 10.1109/ICCP48234.2019.8959762.

[64] R. Odate and H. Goto, *Highly-accurate fast candidate reduction method for Japanese/Chinese character recognition*. IEEE Computer Society, 2016, vol. 2016-August, ISBN: 9781467399616. DOI: 10.1109/ICIP.2016.7532887.

[65] M. S. H. Onim, M. I. Akash, M. Haque, *et al.*, *Traffic surveillance using vehicle license plate detection and recognition in Bangladesh*. Institute of Electrical and Electronics Engineers Inc., 2020, ISBN: 9781665422543. DOI: 10.1109/ICECE51571.2020.9393109.

[66] S. Pavaskar and S. Budihal, "Real-time vehicle-type categorization and character extraction from the license plates," in *International Conference on Cognitive Informatics and Soft Computing, CISC 2017*, A. K. Bhoi, V. E. Balas, A. F. Zobaa, *et al.*, Eds., vol. 768, Springer Verlag, 2019, pp. 557–565, ISBN: 9789811306167. DOI: 10.1007/978-981-13-0617-4_54.

[67] J. Pirgazi, A. G. Sorkhi, and M. M. P. Kallehbasti, "An efficient robust method for accurate and real-time vehicle plate recognition," *Magazine*, vol. 18, no. 5 Number, pp. 1759–1772, 2021, ISSN: 18618200. DOI: 10.1007/s11554-021-01118-7.

[68] G. K. Prasad, I. Khan, and N. Chanukotimath, *On-line Hindi handwritten character recognition for mobile devices*. 2012, ISBN: 9781450311960. DOI: 10.1145/2345396.2345568.

[69] S. Qin and S. Liu, "Towards end-to-end car license plate location and recognition in unconstrained scenarios," *Magazine*, no. Number, 2021, ISSN: 09410643. DOI: 10.1007/s00521-021-06147-8.

[70] M. A. Rahaman, M. Mahin, M. H. Ali, *et al.*, *BHCDR: Real-Time Bangla Handwritten Characters and Digits Recognition using Adopted Convolutional Neural Network*. Institute of Electrical and Electronics Engineers Inc., 2019, ISBN: 9781728134451. DOI: 10.1109/ICASERT.2019.8934476.

[71] Rosalina, J. P. Hutagalung, and G. Sahuri, "Hiragana handwriting recognition using deep neural network search," *Magazine*, vol. 14, no. 1 Number, pp. 161–168, 2020, ISSN: 18657923. DOI: 10.3991/ijim.v14i01.11593.

[72] A. Sethy and P. K. Patra, *Discrete cosine transformation based approach for offline handwritten character and numeral recognition*, 1st ed. IOP Publishing Ltd, 2021, vol. 1770. DOI: 10.1088/1742-6596/1770/1/012004.

[73] M. T. Shahed, M. R. I. Udoy, B. Saha, *et al.*, *Automatic Bengali number plate reader*. Institute of Electrical and Electronics Engineers Inc., 2017, vol. 2017-December, ISBN: 9781509011339. DOI: 10.1109/TENCON.2017.8228070.

[74] S. Shelke and S. Apte, *Real-time character reading system for marathi script using raspberry PI*, CP739. Institution of Engineering and Technology, 2016, vol. 2016, ISBN: 9781785618260. DOI: 10.1049/cp.2016.1552.

[75] N. A. Siddique, A. Iqbal, F. Mahmud, *et al.*, *Development of an automatic vehicle license plate detection and recognition system for Bangladesh*. 2012, ISBN: 9781467311519. DOI: 10.1109/ICIEV.2012.6317529.

[76] S. M. Silva and C. R. Jung, "Real-time license plate detection and recognition using deep convolutional neural networks," *Magazine*, vol. 71, no. Number, 2020, ISSN: 10473203. DOI: 10.1016/j.jvcir.2020.102773.

[77] D. Srivastava, B. Rajitha, and S. Agarwal, "Gradient feature-based classification of patterned images," in *2nd International Conference on Computing, Communications, and Cyber-Security, IC4S 2020*, P. K. Singh, S. T. Wierzchon, S. Tanwar, *et al.*, Eds., vol. 203 LNNS, Springer Science and Business Media Deutschland GmbH, 2021, pp. 1007–1016, ISBN: 9789811607325. DOI: 10.1007/978-981-16-0733-2_71.

[78] Wahyono, A. Filonenko, and K. H. Jo, *Automatic LED text recognition method on electronic road sign using local spatial pattern and random forest classifier*. Institute of Electrical and Electronics Engineers Inc., 2015, vol. 2015-August, ISBN: 9781467372664. DOI: 10.1109/IVS.2015.7225686.

[79] G. H. Wibowo, R. Sigit, and A. Barakbah, *Feature extraction of character image using shape energy*. Institute of Electrical and Electronics Engineers Inc., 2016, ISBN: 9781509016402. DOI: 10.1109/ELECSYM.2016.7861052.

[80] Z. Xu, L. Wang, S. Niu, *et al.*, *A method of positioning and recognition of electronic scale characters based on deep learning*, 1st ed. IOP Publishing Ltd, 2020, vol. 1693. DOI: 10.1088/1742-6596/1693/1/012122.

[81] H. Yang, W. Zhuang, Q. Zhou, *et al.*, "A portable intelligent license plate recognition system based on off-the-shelf mini camera," in *6th International Conference on Artificial Intelligence and Security, ICAIS 2020*, X. Sun, J. Wang, and E. Bertino, Eds., vol. 12239 LNCS, Springer Science and Business Media Deutschland GmbH, 2020, pp. 635–645, ISBN: 9783030578831. DOI: 10.1007/978-3-030-57884-8_56.

[82] S. Yuan, G. Y. Zhang, J. H. Wu, *et al.*, "Study of digital character recognition based on bp neural networks," in *2013 2nd International Conference on Measurement, Instrumentation and Automation, ICMIA 2013*, vol. 333-335, Guilin, 2013, pp. 856–859, ISBN: 9783037857502. DOI: 10.4028/www.scientific.net/AMM.333-335.856.

[83] X. Zhai, F. Bensaali, and R. Sotudeh, *OCR-based neural network for ANPR*. 2012, ISBN: 9781457717741. DOI: 10.1109/IST.2012.6295581.

[84] F. Zhang, J. Luan, Z. Xu, *et al.*, "Detreco: Object-text detection and recognition based on deep neural network," *Magazine*, vol. 2020, no. Number, p. 2 365 076, Jul. 2020, ISSN: 1024-123X. DOI: 10.1155/2020/2365076.

[85] Q. J. Zhao, P. Cao, and Q. X. Meng, "Image capturing and segmentation method for characters marked on hot billets," in *Advanced Materials Research*, vol. 945-949, Trans Tech Publications Ltd, 2014, pp. 1830–1836, ISBN: 9783038351221. DOI: 10.4028/www.scientific.net/AMR.945-949.1830.

[86] Y. Zhuang, Q. Liu, C. Qiu, *et al.*, *A Handwritten Chinese Character Recognition based on Convolutional Neural Network and Median Filtering*, 1st ed. IOP Publishing Ltd, 2021, vol. 1820. DOI: 10.1088/1742-6596/1820/1/012162.

[87] M. Zohra and D. Rajeswara Rao, "A comprehensive data analysis on handwritten digit recognition using machine learning approach," *Magazine*, vol. 8, no. 6 Number, pp. 1449–1453, 2019, ISSN: 22783075.

[88] H. Zhu, K. V. Yuen, L. Mihaylova, *et al.*, "Overview of environment perception for intelligent vehicles," *Magazine*, vol. 18, no. 10 Number, pp. 2584–2601, 2017, ISSN: 15249050. DOI: 10.1109/TITS.2017.2658662.

[89] S. Khan, M. Sajjad, T. Hussain, *et al.*, "A review on traditional machine learning and deep learning models for wbcs classification in blood smear images," *Magazine*, vol. 9, no. Number, pp. 10 657–10 673, 2021, ISSN: 21693536. DOI: 10.1109/ACCESS.2020.3048172.

[90] P. Datta and R. Rohilla, *An Introduction to Deep Learning Applications in MRI Images*. Institute of Electrical and Electronics Engineers Inc., 2019, ISBN: 9781728117935. DOI: 10.1109/PEEIC47157.2019.8976727.

[91] J. Fulco, A. Devkar, A. Krishnan, *et al.*, *Empirical evaluation of convolutional neural networks prediction time in classifying German traffic signs*. SciTePress, 2017, ISBN: 9789897582424. DOI: 10.5220/0006307402600267.

[92] Y. Wei, S. Tran, S. Xu, *et al.*, "Deep learning for retail product recognition: Challenges and techniques," *Magazine*, vol. 2020, no. Number, 2020, ISSN: 16875265. DOI: 10.1155/2020/8875910.

[93] H. Wang, J. Zhang, and Y. Xia, "Jointly using deep model learned features and traditional visual features in a stacked svm for medical subfigure classification," in *7th International Conference on Intelligence Science and Big Data Engineering, IScIDE 2017*, Y. Sun, H. Lu, L. Zhang, *et al.*, Eds., vol. 10559 LNCS, Springer Verlag, 2017, pp. 191–199, ISBN: 9783319677767. DOI: 10.1007/978-3-319-67777-4_17.

[94] H. Hasan and S. Abdul-Kareem, "Fingerprint image enhancement and recognition algorithms: A survey," *Magazine*, vol. 23, no. 6 Number, pp. 1605–1610, 2013, ISSN: 09410643. DOI: 10.1007/s00521-012-1113-0.

[95] S. F. Ali, M. A. Khan, and A. S. Aslam, "Fingerprint matching, spoof and liveness detection: Classification and literature review," *Magazine*, vol. 15, no. 1 Number, 2021, ISSN: 20952228. DOI: 10.1007/s11704-020-9236-4.

[96] ——, "Fingerprint matching, spoof and liveness detection: Classification and literature review," *Magazine*, vol. 15, no. 1 Number, 2021, ISSN: 20952228. DOI: 10.1007/s11704-020-9236-4.

[97] A. Alice Nithya and C. Lakshmi, "Towards enhancing non-cooperative iris recognition using improved segmentation methodology for noisy images," *Magazine*, vol. 10, no. 3 Number, pp. 76–84, 2017, ISSN: 19945450. DOI: 10.3923/jai.2017.76.84.

[98] Q. G. Ge, T. Q. Han, and H. L. Shen, *Non-Lambertian photometric stereo by color and noise analysis*. Institute of Electrical and Electronics Engineers Inc., 2015, ISBN: 9781467381253. DOI: 10.1109/CompComm.2015.7387612.

[99] S. Muhammad, M. N. Dailey, M. Farooq, *et al.*, "Spec-net and spec-cgan: Deep learning models for specularity removal from faces," *Magazine*, vol. 93, no. Number, 2020, ISSN: 02628856. DOI: 10.1016/j.imavis.2019.11.001.

[100] M. Rajeev Kumar and K. Arthi, "An effective non-cooperative iris recognition system using hierarchical collaborative representation-based classification," *Magazine*, vol. 76, no. 8 Number, pp. 5835–5848, 2020, ISSN: 09208542. DOI: 10.1007/s11227-019-03007-0.

[101] M. C. F. Macedo, V. P. Nascimento, and A. C. S. Souza, "Real-time shadow detection using multi-channel binarization and noise removal," *Magazine*, vol. 17, no. 3 Number, pp. 479–492, 2020, ISSN: 18618200. DOI: 10.1007/s11554-018-0799-3.

[102] A. Niranjil Kumar and C. Sureshkumar, "Fuzzy based shadow removal and integrated boundary detection for video surveillance," *Magazine*, vol. 9, no. 6 Number, pp. 2126–2133, 2014, ISSN: 19750102. DOI: 10.5370/JEET.2014.9.6.2126.

[103] Y. Li, Z. Li, H. Tian, *et al.*, *Vehicle detecting and shadow removing based on edged mixture Gaussian model*, 1 PART 1. IFAC Secretariat, 2011, vol. 44, ISBN: 9783902661937. DOI: 10.3182/20110828-6-IT-1002.01717.

[104] H. Fan, M. Han, and J. Li, "Image shadow removal using end-to-end deep convolutional neural networks," *Magazine*, vol. 9, no. 5 Number, 2019, ISSN: 20763417. DOI: 10.3390/app9051009.

[105] A. Gad, M. Yaghi, M. Alkhedher, *et al.*, *Real-time Shadow Detection and Removal by Illumination Drop Point Analysis*. Institute of Electrical and Electronics Engineers Inc., 2020, ISBN: 9781728196732. DOI: 10.1109/3ICT51146.2020.9311979.

[106] A. Abdusalomov and T. K. Whangbo, "Detection and removal of moving object shadows using geometry and color information for indoor video streams," *Magazine*, vol. 9, no. 23 Number, 2019, ISSN: 20763417. DOI: 10.3390/app9235165.

[107] N. N. A. Salim, X. Cheng, and D. Xiao, *Improved shadow removal for unstructured road detection*. CSREA Press, 2013, vol. 1, ISBN: 9781601322524.

[108] G. F. Silva, G. B. Carneiro, R. Doth, *et al.*, "Near real-time shadow detection and removal in aerial motion imagery application," *Magazine*, vol. 140, no. Number, pp. 104–121, 2018, ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2017.11.005.

[109] C. Xiao, R. She, D. Xiao, *et al.*, "Fast shadow removal using adaptive multi-scale illumination transfer," *Magazine*, vol. 32, no. 8 Number, pp. 207–218, 2013, ISSN: 01677055. DOI: 10.1111/cgf.12198.

[110] J. Yuan, J. Wu, and Y. Cheng, *Shadow detecting algorithms research for moving objects base on self-adaptive background*. 2012, ISBN: 9780956715715.

[111] Y. I. Shedlovska and V. V. Hnatushenko, *Shadow removal algorithm with shadow area border processing*. Institute of Electrical and Electronics Engineers Inc., 2016, ISBN: 9781467388412. DOI: 10.1109/YSF.2016.7753827.

[112] J. Oh, J. Min, and B. Heo, "Development of an integrated system based vehicle tracking algorithm with shadow removal and occlusion handling methods," *Magazine*, vol. 46, no. 2 Number, pp. 139–150, 2012, ISSN: 01976729. DOI: 10.1002/atr.148.

[113] T. Yan, S. Hu, X. Su, *et al.*, *Moving object detection and shadow removal in video surveillance*. Institute of Electrical and Electronics Engineers Inc., 2016, ISBN: 9781509032976. DOI: 10.1109/SKIMA.2016.7916189.

[114] I. Rodriguez-Padilla, B. Castelle, V. Marieu, *et al.*, "A simple and efficient image stabilization method for coastal monitoring video systems," *Magazine*, vol. 12, no. 1 Number, 2020, ISSN: 20724292. DOI: 10.3390/RS12010070.

[115] M. Rasti and M. T. Sadeghi, "Video stabilization and completion using the scale-invariant features and ransac robust estimator," in *1st International Conference on Innovative Computing Technology, INCT 2011*, vol. 241 CCIS, Tehran, 2011, pp. 274–281, ISBN: 9783642273360. DOI: 10.1007/978-3-642-27337-7_26.

[116] S. Alam, S. P. N. Singh, and U. Abeyratne, *Considerations of handheld respiratory rate estimation via a stabilized Video Magnification approach*. Institute of Electrical and Electronics Engineers Inc., 2017, ISBN: 9781509028092. DOI: 10.1109/EMBC.2017.8037805.

[117] W. G. Aguilar and C. Angulo, "Real-time model-based video stabilization for microaerial vehicles," *Magazine*, vol. 43, no. 2 Number, pp. 459–477, 2016, ISSN: 13704621. DOI: 10.1007/s11063-015-9439-0.

[118] M. Okade and P. K. Biswas, "Video stabilization using maximally stable extremal region features," *Magazine*, vol. 68, no. 3 Number, pp. 947–968, 2014, ISSN: 13807501. DOI: 10.1007/s11042-012-1095-z.

[119] J. Li, T. Xu, and K. Zhang, "Real-time feature-based video stabilization on fpga," *Magazine*, vol. 27, no. 4 Number, pp. 907–919, 2017, ISSN: 10518215. DOI: 10.1109/TCSVT.2016.2515238.

[120] D. P. Umrikar and S. L. Tade, *Implementation of Video Stabilization Algorithm Using Block Based Method*. Institute of Electrical and Electronics Engineers Inc., 2017, ISBN: 9781538640081. DOI: 10.1109/ICCUBEA.2017.8463721.

[121] A. Hamza, R. Hafiz, M. M. Khan, *et al.*, "Stabilization of panoramic videos from mobile multi-camera platforms," *Magazine*, vol. 37, no. Number, pp. 20–30, 2015, ISSN: 02628856. DOI: 10.1016/j.imavis.2015.02.002.

[122] M. R. e Souza and H. Pedrini, "Combination of local feature detection methods for digital video stabilization," *Magazine*, vol. 12, no. 8 Number, pp. 1513–1521, 2018, ISSN: 18631703. DOI: 10.1007/s11760-018-1307-8.

[123] Y. J. Koh, J. Y. Sim, and C. S. Kim, *Robust video stabilization based on mesh grid warping of rolling-free features*. Institute of Electrical and Electronics Engineers Inc., 2014, ISBN: 9781479957514. DOI: 10.1109/ICIP.2014.7026169.

[124] M. J. Tanakian, M. Rezaei, and F. Mohanna, "Digital video stabilization system by adaptive fuzzy kalman filtering," *Magazine*, vol. 1, no. 4 Number, pp. 223–232, 2013, ISSN: 23221437. DOI: 10.7508/jist.2013.04.003.

[125] M. N. Favorskaya, L. C. Jain, and V. Buryachenko, "Digital video stabilization in static and dynamic scenes," in *Intelligent Systems Reference Library*, vol. 73, Springer Science and Business Media Deutschland GmbH, 2015, pp. 261–309. DOI: 10.1007/978-3-319-10653-3_9.

[126] T. P. Minh and M. C. Hong, *Video Stabilization Using Feature-Based Classification*. Institute of Electrical and Electronics Engineers Inc., 2018, ISBN: 9781538658079. DOI: 10.1109/ICCE-ASIA.2018.8552107.

[127] N. Sunny, M. Srikanth, and K. Eswar, "Full frame video motion detection and stabilization using mosaicing and deblurring," *Magazine*, vol. 8, no. 7 Number, pp. 597–601, 2019, ISSN: 22783075.

[128] H. Saxena, K. Verma, D. Ghosh, *et al.*, *Digital Video Stabilization with Preserved Intentional Camera Motion and Smear Removal*. Institute of Electrical and Electronics Engineers Inc., 2019, ISBN: 9781538659069. DOI: 10.1109/ICCCNT45670.2019.8944856.

[129] M. S. Javadi, Z. Kadim, H. H. Woon, *et al.*, *Video stabilization and tampering detection for surveillance systems using homography*. Institute of Electrical and Electronics Engineers Inc., 2015, ISBN: 9781479979523. DOI: 10.1109/I4CT.2015.7219580.

[130] R. Chereau and T. P. Breckon, *Robust motion filtering as an enabler to video stabilization for a tele-operated mobile robot*. 2013, vol. 8897, ISBN: 9780819497666. DOI: 10.1117/12.2028360.

[131] W. Junwei, Q. Jianjun, L. Jiaqi, *et al.*, *Wide Range and Dynamic Video Image Stabilization Applied in UAV Aerial Photographing*. Institute of Electrical and Electronics Engineers Inc., 2020, ISBN: 9781728177380. DOI: 10.1109/IICSPI51290.2020.9332430.

[132] K. Verma, Andri, R. K. Singh, *et al.*, *FAST Based Video Stabilization with Preserved Intentional Motion and Smear Removal*. Institute of Electrical and Electronics Engineers Inc., 2020, ISBN: 9781728158464. DOI: 10.1109/ICE348803.2020.9122965.

[133] H. Qu and L. Song, *Video stabilization with L1-L2 optimization*. 2013, ISBN: 9781479923410. DOI: 10.1109/ICIP.2013.6738007.

[134] S. Fang, X. Ma, and Z. Cao, *Video stabilization based on adaptive local subspace of feature point classification*. Institute of Electrical and Electronics Engineers Inc., 2016, vol. 0, ISBN: 9781509041657. DOI: 10.1109/ICDSP.2016.7868623.

[135] V. Avramelos, G. V. Wallendael, and P. Lambert, *Real-time low-complexity digital video stabilization in the compressed domain*. IEEE Computer Society, 2018, vol. 2018-September, ISBN: 9781538660959. DOI: 10.1109/ICCE-Berlin.2018.8576211.

[136] J. Auysakul, H. Xu, and V. Pooneeth, "A hybrid motion estimation for video stabilization based on an imu sensor," *Magazine*, vol. 18, no. 8 Number, 2018, ISSN: 14248220. DOI: 10.3390/s18082708.

[137] H. Gunawan, C. C. Han, and C. H. Lee, *Video Stabilization and Text-Based Sentence Identification for Video-Based Answering Systems*. Institute of Electrical and Electronics Engineers Inc., 2018, ISBN: 9781538663509. DOI: 10.1109/CCOMS.2018.8463237.

[138] Q. Zhou, B. Zhong, Y. Zhang, *et al.*, "Deep alignment network based multi-person tracking with occlusion and motion reasoning," *Magazine*, vol. 21, no. 5 Number, pp. 1183–1194, 2019, ISSN: 15209210. DOI: 10.1109/TMM.2018.2875360.

[139] J. Wen, Y. Xu, and Y. Zhan, *Ghost, occlusion and distractors handling in object tracking system*. 2011, ISBN: 9781612841984. DOI: 10.1109/ICCIS.2011.6070311.

[140] S. Sarwar, N. I. Rao, and M. F. Khan, *Real-time object tracking using Powell's direct set method for object localization and kalman filter for occlusion handling*. 2012, ISBN: 9781467321815. DOI: 10.1109/DICTA.2012.6411705.

[141] R. Velazquez-Pupo, A. Sierra-Romero, D. Torres-Roman, *et al.*, "Vehicle detection with occlusion handling, tracking, and oc-svm classification: A high performance vision-based system," *Magazine*, vol. 18, no. 2 Number, 2018, ISSN: 14248220. DOI: 10.3390/s18020374.

[142] C. Xiong and L. Li, "Vehicle occlusion detection and segment based on windows," in *2011 International Conference on Material Science and Information Technology, MSIT2011*, vol. 433-440, Singapore, 2012, pp. 3186–3191, ISBN: 9783037853191. DOI: 10.4028/www.scientific.net/AMR.433-440.3186.

[143] X. Dong, J. Shen, D. Yu, *et al.*, "Occlusion-aware real-time object tracking," *Magazine*, vol. 19, no. 4 Number, pp. 763–771, 2017, ISSN: 15209210. DOI: 10.1109/TMM.2016.2631884.

[144] Y. Huang, C. Ju, X. Hu, *et al.*, "An anti-occlusion and scale adaptive kernel correlation filter for visual object tracking," *Magazine*, vol. 13, no. 4 Number, pp. 2094–2112, 2019, ISSN: 19767277. DOI: 10.3837/tiis.2019.04.020.

[145] S. Sarkar, V. Venugopalan, K. Reddy, *et al.*, "Deep learning for automated occlusion edge detection in rgb-d frames," *Magazine*, vol. 88, no. 2 Number, pp. 205–217, 2017, ISSN: 19398018. DOI: 10.1007/s11265-016-1209-3.

[146] M. Heimbach, K. Ebadi, and S. Wood, *Improving Object Tracking Accuracy in Video Sequences Subject to Noise and Occlusion Impediments by Combining Feature Tracking with Kalman Filtering*. IEEE Computer Society, 2018, vol. 2018-October, ISBN: 9781538692189. DOI: 10.1109/ACSSC.2018.8645175.

[147] J. P. Lomaliza and H. Park, "Initial pose estimation of 3d object with severe occlusion using deep learning," in *20th International Conference on Advanced Concepts for Intelligent Vision Systems, ACIVS 2020*, J. Blanc-Talon, P. Delmas, W. Philips, *et al.*, Eds., vol. 12002 LNCS, Springer, 2020, pp. 325–336, ISBN: 9783030406042. DOI: 10.1007/978-3-030-40605-9_28.

[148] Y. Cong, D. Tian, Y. Feng, *et al.*, "Speedup 3-d texture-less object recognition against self-occlusion for intelligent manufacturing," *Magazine*, vol. 49, no. 11 Number, pp. 3887–3897, 2019, ISSN: 21682267. DOI: 10.1109/TCYB.2018.2851666.

[149] H. Van Pham and B. R. Lee, "Front-view car detection and counting with occlusion in dense traffic flow," *Magazine*, vol. 13, no. 5 Number, pp. 1150–1160, 2015, ISSN: 15986446. DOI: 10.1007/s12555-014-0229-7.

[150] R. R. Ramisetty, C. Qu, R. Aktar, *et al.*, *Dynamic Computation Off-loading and Control based on Occlusion Detection in Drone Video Analytics*. ICST, 2020, vol. Part F165625, ISBN: 9781450377515. DOI: 10.1145/3369740.3369793.

[151] D. T. Lin and Y. H. Chang, *Occlusion handling for pedestrian tracking Using partial object template-based component Particle Filter*. 2013, ISBN: 9789728939892.

[152] W. Fang, Y. Zhao, Y. Yuan, *et al.*, *Real-time multiple vehicles tracking with occlusion handling*. 2011, ISBN: 9780769545417. DOI: 10.1109/ICIG.2011.140.

B Examples measured activations of a surgical energy device

The following section shows all the measured data of the Conmed System 2450. Video data was acquired during surgery and evaluated postoperatively as described in the first article in section 2. Data acquisition is conducted with the approval of the ethic committee of Japan's National Cancer Center and the University of Tokyo. In these datasets, the errors have been removed to create a clearer view of how the energy device was used during each surgery. In total six surgeries were performed my expert surgeons, described by Figures 3 to 8, and three surgeries were performed by residents described by Figures 9 to 11.

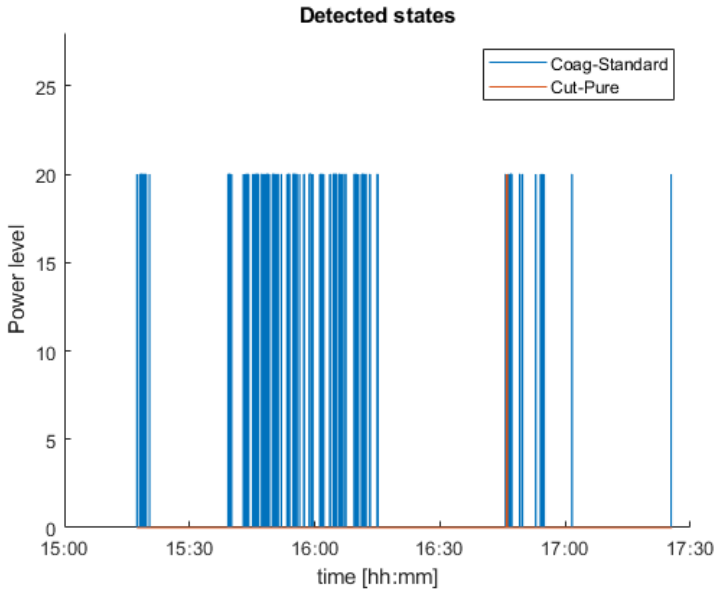


Figure 3: Measured activations during a surgery performed by a surgeon with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

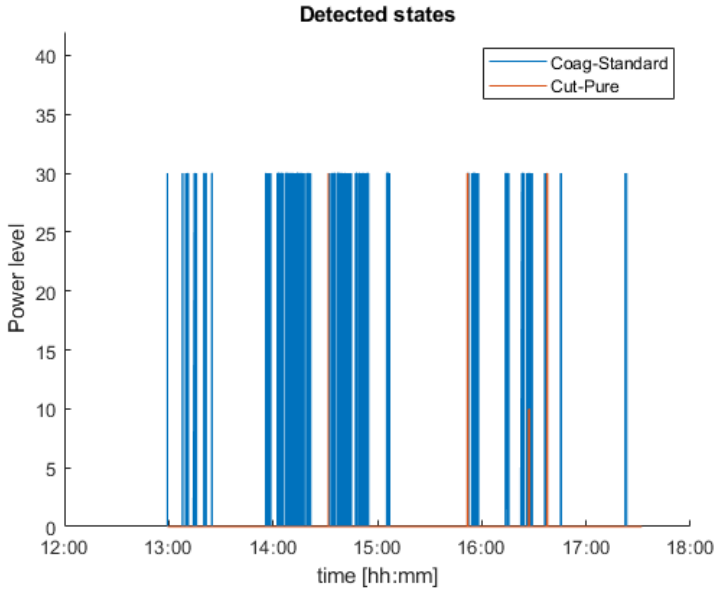


Figure 4: Measured activations during a surgery performed by a surgeon with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

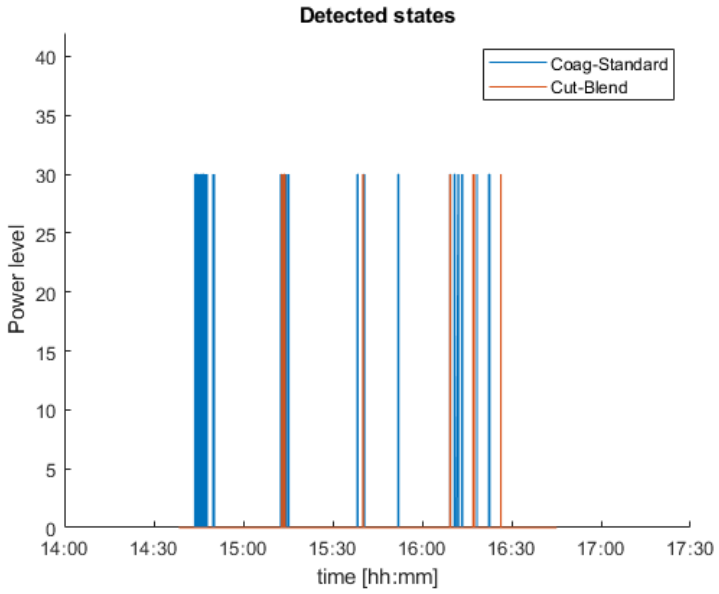


Figure 5: Measured activations during a surgery performed by a surgeon with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

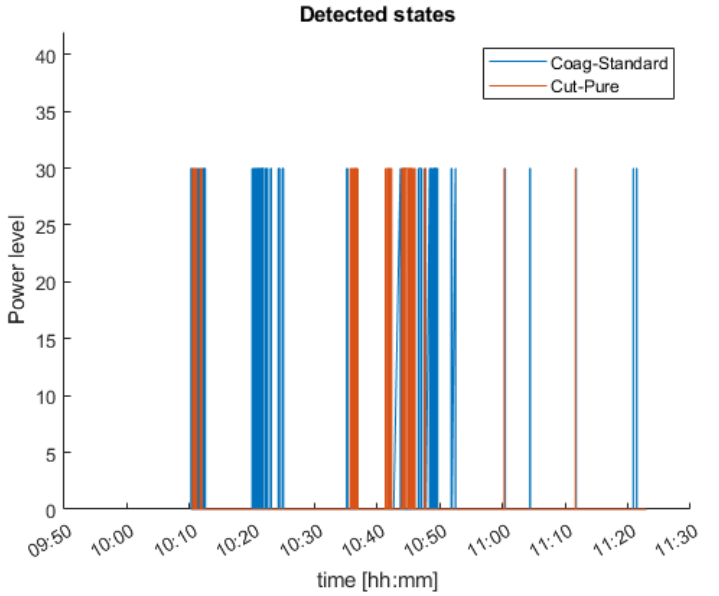


Figure 6: Measured activations during a surgery performed by a surgeon with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

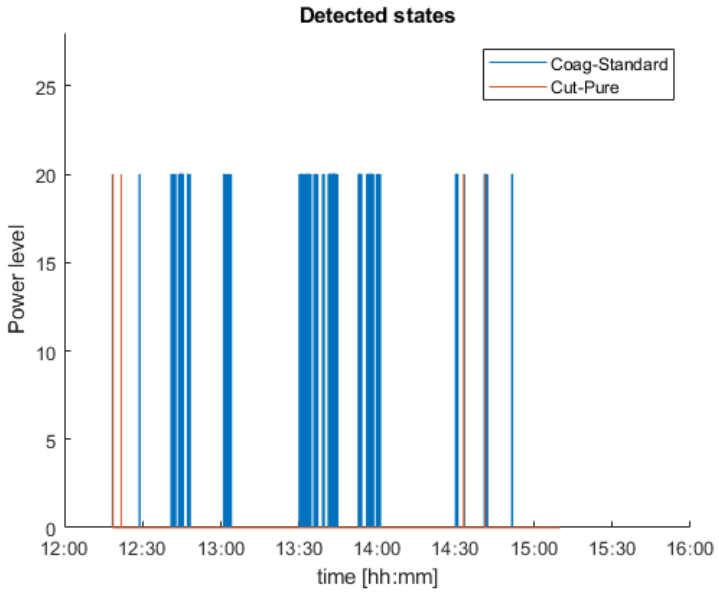


Figure 7: Measured activations during a surgery performed by a surgeon with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

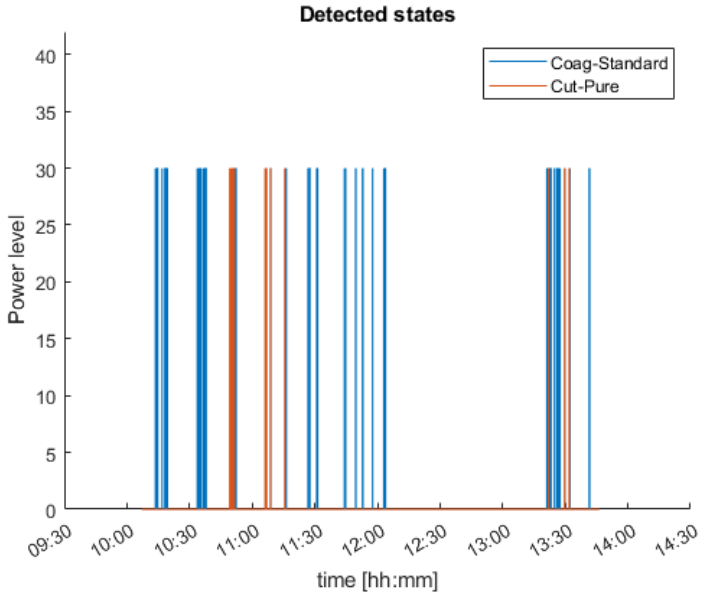


Figure 8: Measured activations during a surgery performed by a surgeon with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

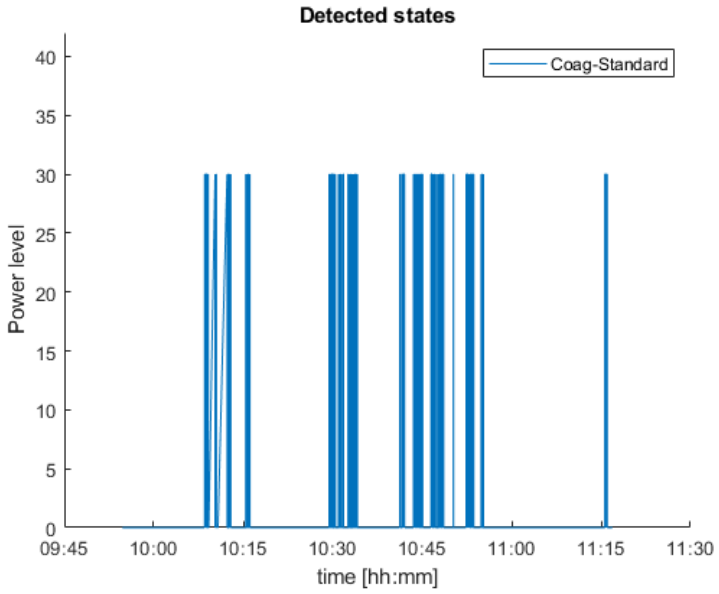


Figure 9: Measured activations during a surgery performed by a resident with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

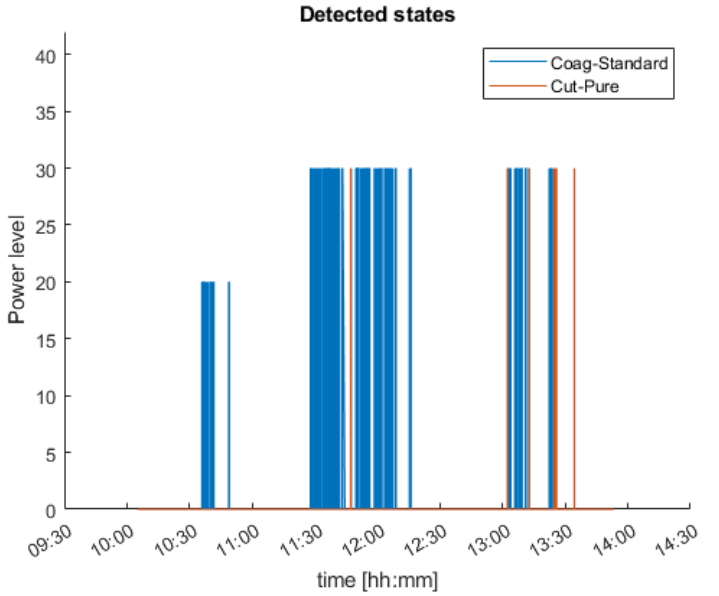


Figure 10: Measured activations during a surgery performed by a resident with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

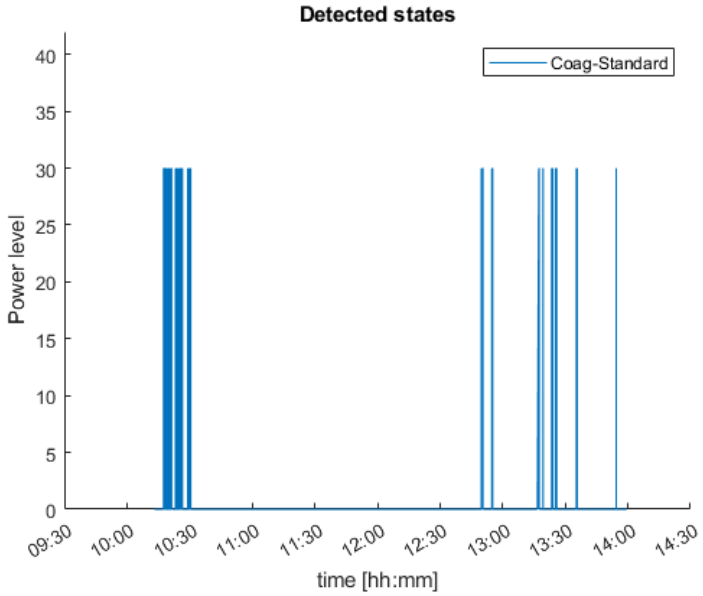


Figure 11: Measured activations during a surgery performed by a resident with a Conmed System 2450. The y-axis indicates the power level of the major mode when activated. The x-axis indicates the time during the surgery. The instrument used was a dual electrode.

C State representation schemes of all energy devices, used at Japan's National Cancer Center

In the following section an overview is given of all the state trees describing the seven energy devices used within Japan's National Cancer Center. These consist of the following energy devices:

1. the Conmed System 2450, described by Figure 12
2. the Valleylab ForceTriad (Ligasure), described by Figure 13
3. the Valleyab FT10, described by Figure 14
4. the ERBE VIO 300D, described by Figure 15
5. the ERBE VIO3, described by Figure 16
6. the OLYMPUS Thunderbeat, described by Figure 17
7. the ETHICON Harmonic & EnSeal, described by Figure 18

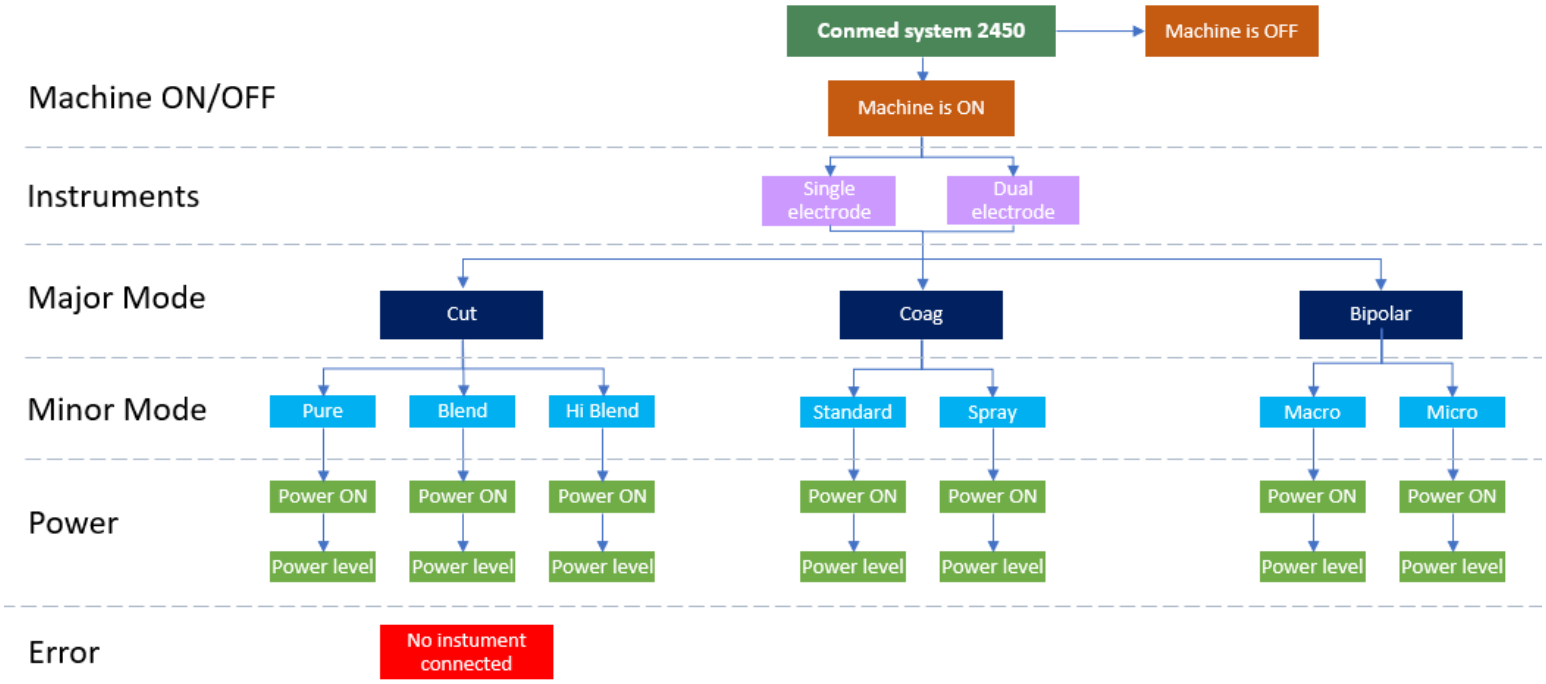


Figure 12: State tree representing the states of the Conmed System 2450.

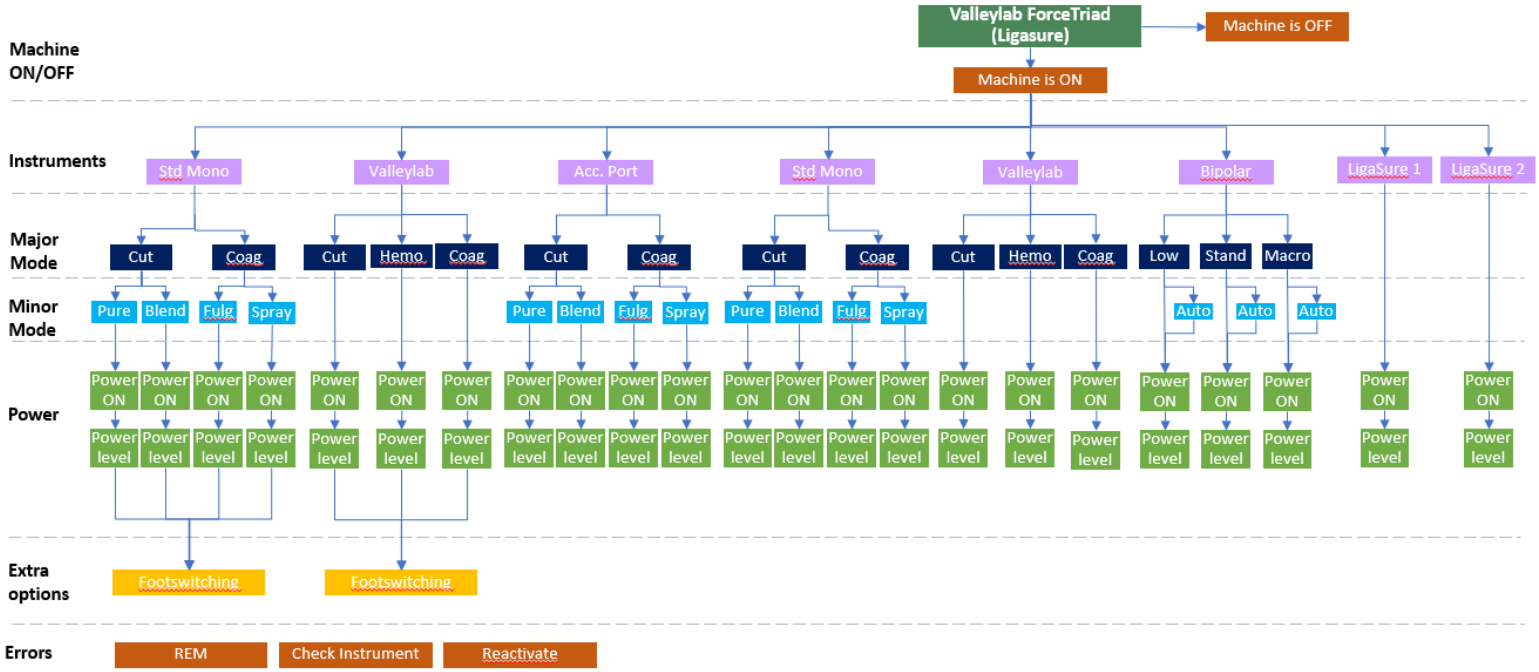


Figure 13: State tree representing the Valleylab ForceTriad (Ligasure).

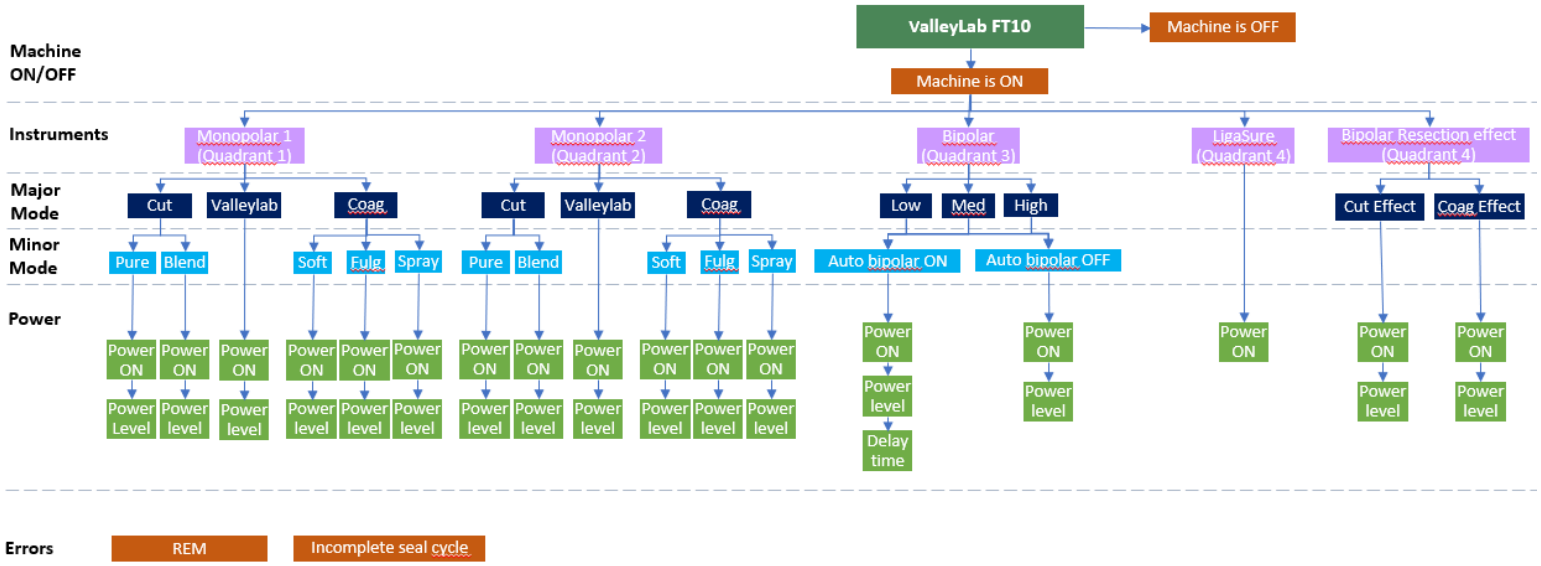


Figure 14: State tree representing the Valleyab FT10.

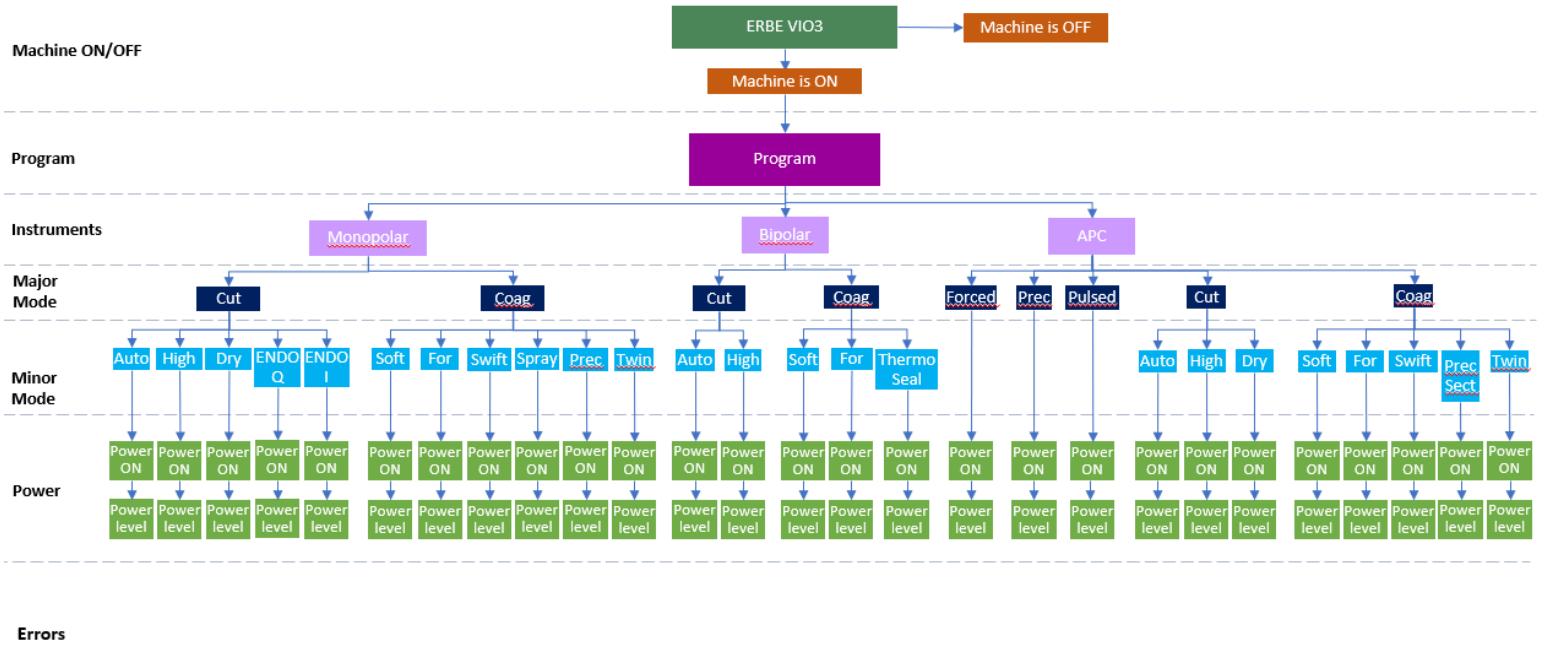


Figure 15: State tree representing the ERBE VIO 300D.

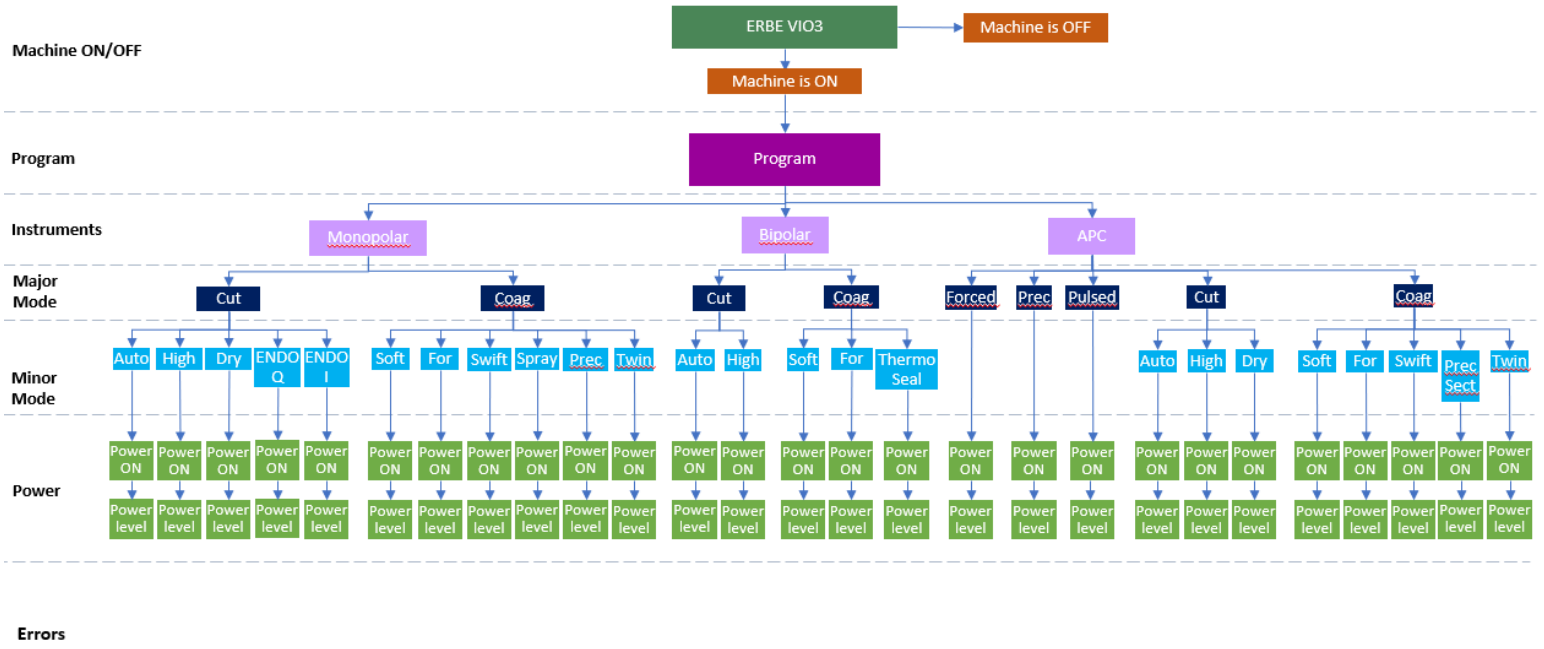


Figure 16: State tree representing the ERBE VIO3.

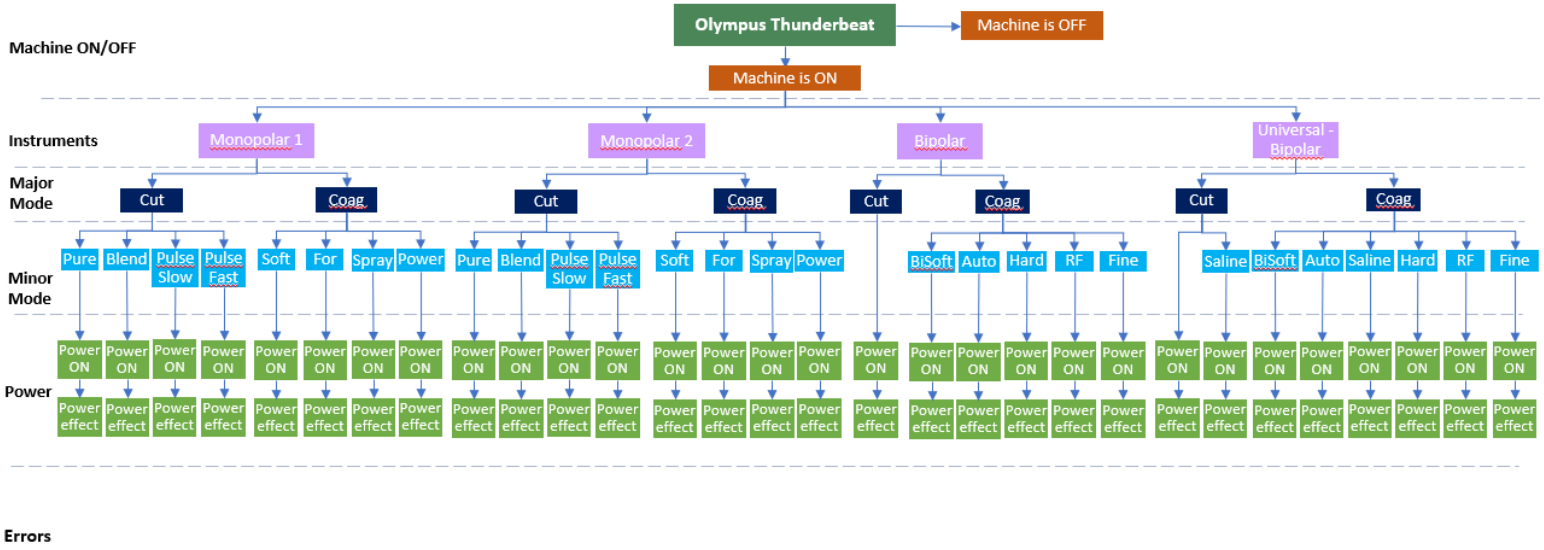


Figure 17: State tree representing the OLYMPUS Thunderbeat.

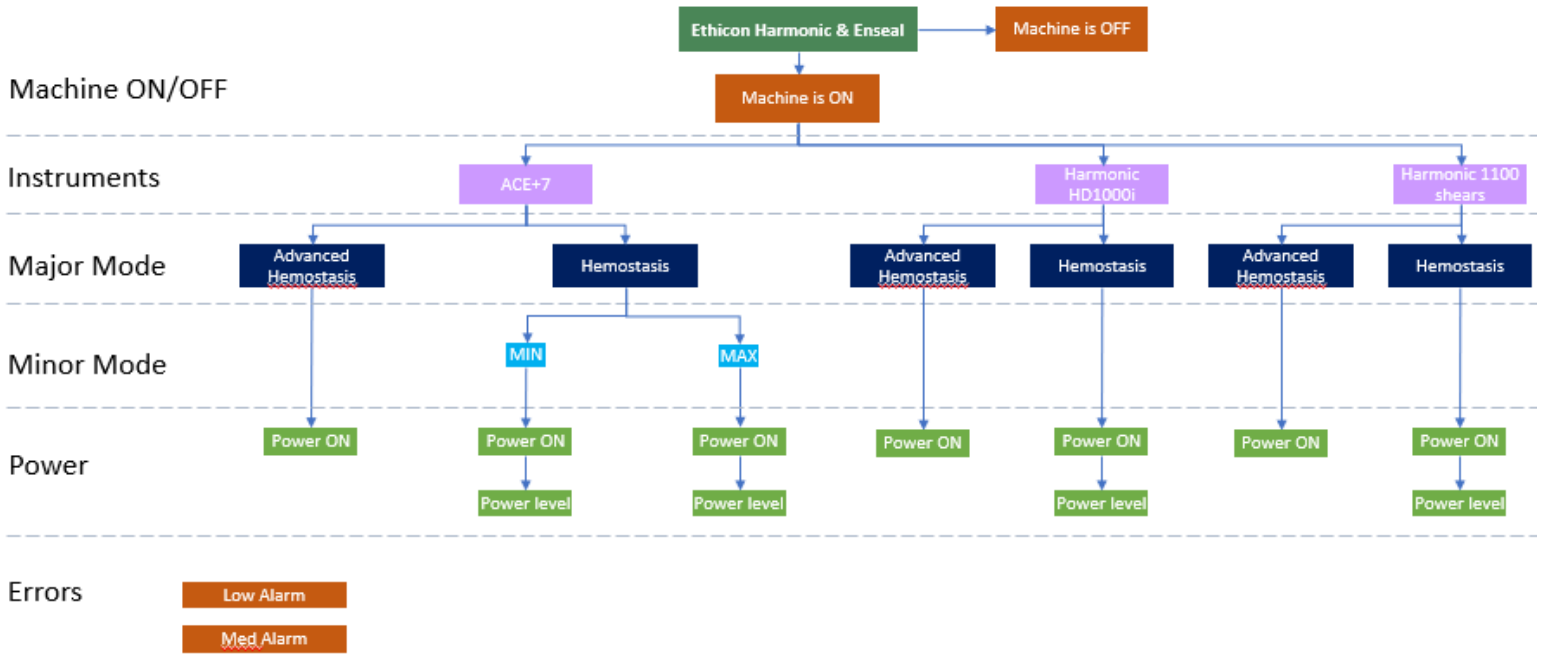


Figure 18: State tree representing the ETHICON Harmonic & EnSeal.

D Examples measured currentprobe data

In the following section raw data of the measurement of the Conmed system 2450 through the current probe is described. In Figure 19 an example is shown of measuring the activation of the Conmed. When the current exceeds the threshold, the major mode of the Conmed is activated. In Figure 20 data is shown of measuring the Ethicon Harmonic & EnSeal. In this case no clear peaks are visible and therefore the activation cannot be measured. In Figure 21 an example is shown of measuring the Conmed on both the activation of the major mode and the actual contact between the instrument and the tissue. In this situation it can be perceived that when the measured current exceeds the higher threshold, the major mode is activated, but the instrument is not in contact with the tissue. When the instrument comes into contact with the tissue while activating the power of the major mode, the measured current lowers to a value below the higher threshold and above the lower threshold. Through this, a distinction can be made between merely activating the energy device, and making contact with the tissue when activating the energy device.

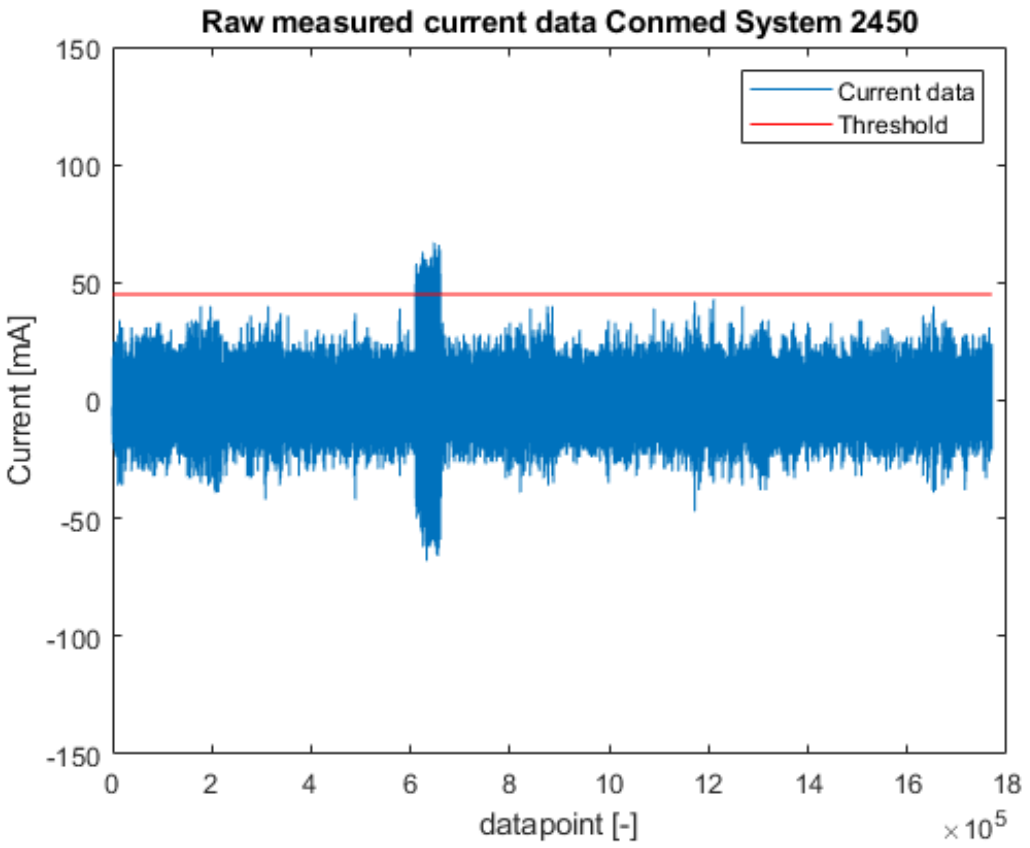


Figure 19: Measured current data of the Conmed System 2450. When the power of the energy device is activated, the current exceeds the set threshold. In this case, one activation is performed.

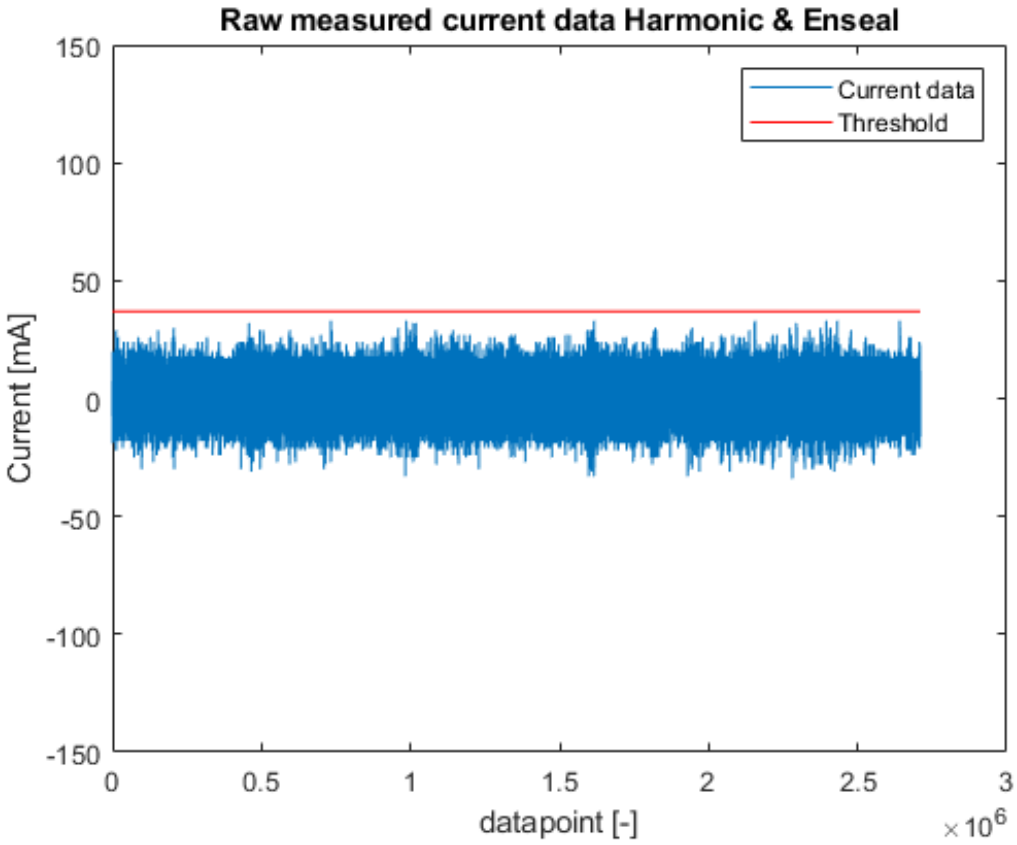


Figure 20: Measured current data of the Ethicon Harmonic & EnSeal. The power is activated several times, but no clear peaks are perceived that exceed the threshold. Due to this, the power activation could not be measured.

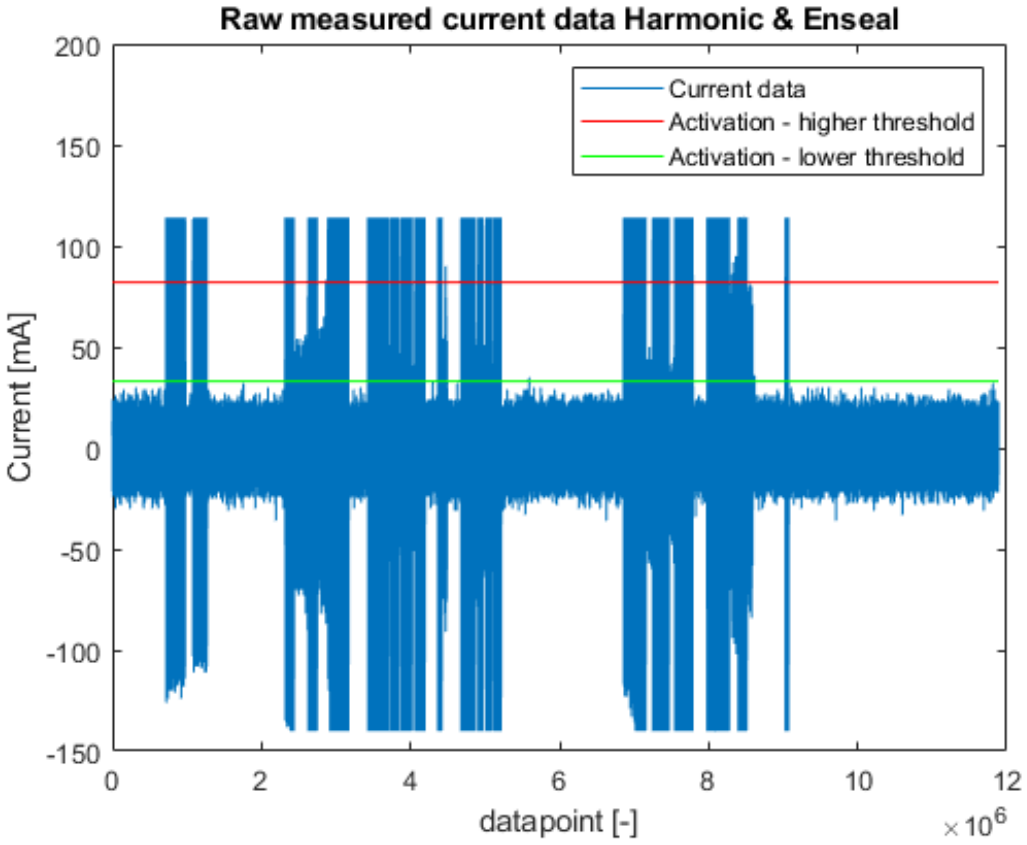


Figure 21: Measured current data of the Conmed System 2450. In this situation the power of the energy device is activated multiple times and used to cut some tissue. When the energy device is activated, but not in contact with the tissue, the measured current exceeds the higher threshold. When the energy device is activated, and the instrument is in actual contact with the tissue, the measured current is below the higher threshold and above the lower threshold. When the energy device is not activated, the measured current is below the lower threshold.