HHS Public Access

Author manuscript

IEEE Trans Med Robot Bionics. Author manuscript; available in PMC 2022 February 01.

Published in final edited form as:

IEEE Trans Med Robot Bionics. 2021 February ; 3(1): 2–10. doi:10.1109/tmrb.2020.3040002.

Computer Vision in the Operating Room: Opportunities and Caveats

Lauren R. Kennedy-Metz,
Medical Robotics and Computer-Assisted Surgery (MRCAS) Laboratory, affiliated with Harvard Medical School in Boston, MA 02115 and the VA Boston Healthcare System in West Roxbury, MA 02132.

Pietro Mascagni,
ICube at the University of Strasbourg, CNRS, IHU Strasbourg, France and Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy

Antonio Torralba,
Computer Science and Artificial Intelligence Laboratory (CSAIL) at Massachusetts Institute of Technology in Cambridge, MA 02139.

Roger D. Dias,
Harvard Medical School in Boston, MA 02115 and STRATUS Center for Medical Simulation in the Department of Emergency Medicine at Brigham and Women's Hospital in Boston, MA 02115.

Pietro Perona,
Computer Vision Laboratory at CalTech and Amazon Inc. in Pasadena, CA 91125.

Julie A. Shah,
Computer Science and Artificial Intelligence Laboratory (CSAIL) at Massachusetts Institute of Technology in Cambridge, MA 02139.

Nicolas Padoy,
ICube at the University of Strasbourg, CNRS, IHU Strasbourg, France.

Marco A. Zenati
Medical Robotics and Computer-Assisted Surgery (MRCAS) Laboratory, affiliated with Harvard Medical School in Boston, MA 02115 and the VA Boston Healthcare System in West Roxbury, MA 02132.

Abstract

Effectiveness of computer vision techniques has been demonstrated through a number of applications, both within and outside healthcare. The operating room environment specifically is a setting with rich data sources compatible with computational approaches and high potential for direct patient benefit. The aim of this review is to summarize major topics in computer vision for surgical domains. The major capabilities of computer vision are described as an aid to surgical teams to improve performance and contribute to enhanced patient safety. Literature was identified through leading experts in the fields of surgery, computational analysis and modeling in medicine,

Lauren.Kennedy-Metz@va.gov.

and computer vision in healthcare. The literature supports the application of computer vision principles to surgery. Potential applications within surgery include operating room vigilance, endoscopic vigilance, and individual and team-wide behavioral analysis. To advance the field, we recommend collecting and publishing carefully annotated datasets. Doing so will enable the surgery community to collectively define well-specified common objectives for automated systems, spur academic research, mobilize industry, and provide benchmarks with which we can track progress. Leveraging computer vision approaches through interdisciplinary collaboration and advanced approaches to data acquisition, modeling, interpretation, and integration promises a powerful impact on patient safety, public health, and financial costs.

Index Terms

Artificial intelligence; computer vision; patient safety; surgery; robotics; vigilance

I. INTRODUCTION

Advancements in operative techniques, multidisciplinary management, and the introduction of multiple technologies make modern operating rooms (ORs) more effective but also more complex and error prone than they used to be [1]. Indeed, modern ORs are complex sociotechnical systems with many people (surgeons, anesthesiologists, perfusionists, nurses), tools and technologies (scalpel, patient monitors), tasks (cutting, communicating, monitoring), environmental features (patient table, lighting), and organizational aspects (unspoken hierarchy, rules, policies) interacting with each other. All of these system complexities affect healthcare professionals' performance, which then influences patient safety and patient outcomes [2].

Efforts to improve OR surgical safety have recently shifted from the individual to the team, and the awareness of surgical team quality and its impact on OR safety has considerably increased [3]. Teamwork is critical for safe patient care and interventions like team training or the introduction of cognitive aids (e.g. checklists) to improve teamwork and safety; however, their impact is limited by the lack of systematic sensor-based (e.g. video recording) measurement approaches of team performance [4]. Cognitive engineering approaches in surgery focus on integrating multiple data sources to understand cognitive demands on providers to minimize their impact [5]. Because of the recent emphasis on quality measurements, data acquisition needs to be systematic, mandatory, and automated to allow for subsequent detailed prospective analysis in a learning healthcare system [2].

In the current practice of surgery however, the only data that are routinely collected for the record of an operation come from the dictated, subjective, post-hoc report. For the most part the operative report lacks objective data such as images, video sequences, time sequences, or real-time annotations. Unlike conventional approaches to team performance measurement, sensor-based measurement is automated and objective activity in which data are collected in real-time. In addition, understanding task performance and team behavior is crucial to the success of the procedure and may influence the optimal design, development, and operations of next-generation OR. Current surgical safety research relies on trained human observers

collecting potentially biased data manually during the surgical procedure; this model is poorly scalable for delivering just-in-time information and improving real-time vigilance and situational awareness.

In addition to the need for more advanced and comprehensive data acquisition, there is tremendous utility in algorithmic manipulation and transformation of these data into a tangible source of cognitive support for clinical individuals and teams. To date, various cognitive aids have been developed and deployed with the goal of supporting clinician cognition (most notably the surgical safety checklist [6]), but these tools are largely static and agnostic to the data-rich context of the operating room. Computer vision techniques provide a transformative opportunity to leverage computational and technological advances to not only support human cognition and performance in healthcare, but to augment it through the acquisition, manipulation, and delivery of data to incite behavioral change in a meaningful way.

There is a clear need for novel approaches to help bridge this safety and quality gap. The purpose of this review is to highlight opportunities and limitations of computer vision technology, with implications for surgical and patient safety.

II. COMPUTER VISION AND OR VIGILANCE

Computer vision is the engineering discipline aiming to give sight to machines. Recent advances in algorithms and computer speed have catalyzed much progress in this area and the number of practical applications is exploding – examples are self-driving cars and trucks [7], precision agriculture [8], the analysis of satellite images [9], robotic assembly of mechanical parts [10], surveillance and security systems [11], and food inspection and quality control [12]. Computer vision has transformative potential in medicine – methods have been developed for application in pathology [13], [14], radiology [15], [16], cell counting [17], skin lesion classification [18], [19], and medical image enhancement [20]. In the endoscopic suite, computer vision applications for early stage cancer diagnosis have already shown expert-like performance in both upper and lower gastrointestinal endoscopy [21], [22].

Computer vision is an excellent sensory modality to detect, measure and study human behavior [23], [24], and to interface flexibly and naturally with people and machines. This is because cameras can be small non-contact passive sensors that work well in the human environment without impeding human motion, do not require complex ad hoc set-up and calibration, and provide a rich signal which may be used to capture a great number of informative variables, such as facial expression [25], body pose [26], skin tone and color, motion, gestures, and identity.

The study of behavior analysis from video or tracking data is also an area of increasing interest in the machine learning community. Behavior analysis, especially human behavior analysis, is a topic of great interest across a wide range of disciplines from psychology to personal robotics. Recent work in machine learning has focused on addressing challenges

such as inferring and anticipating intent [27], identifying “hotspots” or strategic locations [28], and coordination between different agents [29].

Automated computer vision systems that can understand complex human behavior are not yet available and are the object of intense research in the computer vision community. Such systems offer, in principle, significant advantages over using human observers with pencil and clipboard [30], [31]. First: automated systems can deal with a much larger volume of data, permitting the scaling of the size of the typical study from tens of hours of video to thousands of hours. Second, automated systems are not subject to the vagaries of human attention. Third, computer vision analysis is reproducible, while the analysis by human observers is intrinsically subjective. Fourth, it is much easier to maintain privacy when the data is not handled by people. Fifth, behavior traces across thousands of instances may be compared to each other to detect repeated patterns and inconsistencies, which human observers easily miss. Sixth, it is much easier to benchmark the performance of computer systems than humans and, if the system is found to be wanting, it is possible to improve it with systematic collection and annotation of data representing difficult cases and the improvements may be quickly propagated to all instances of the system, unlike humans who need to be retrained individually. Seventh, a single system may often benefit from multiple cameras that can observe the scene simultaneously from a diverse set of viewpoints, while a human observer only has one viewpoint. Lastly, the cost and the timeliness of analysis by automated systems is much improved with respect to using human experts.

The growing availability of recorded behavior in the OR is crucial to the development of new techniques for behavior analysis, especially related to surgical flow disruptions, in machine learning and artificial intelligence more broadly [32]. Furthermore, the new challenges posed by analyzing behavior in the OR will generate new research questions downstream for the machine learning community. Ultimately, computerized vigilance in the OR could lead to more timely and nuanced recognition of deviations from standard behaviors, providing the opportunity for individuals and teams to correct these behaviors and thereby avoid the commission of a patient-threatening medical error.

III. COMPUTER VISION AND ENDOSCOPIC VIGILANCE

Surgeons’ manipulation of patients’ anatomy remains central to surgical care. Indeed, recent evidence suggest that intraoperative technical performance predicts patient’s outcomes and correlates with non-technical skills both at an individual and team level [33]. Understanding the intraoperative course of events can thus offer insights into how to improve surgical safety and efficiency. Furthermore, a reliable representation of the intraoperative context and workflow is necessary to design intelligent computer assistance systems and surgical robots with various degrees of autonomy [34]. Surgical videos are the most informative data source to study intraoperative technical performances. In minimally invasive and robotic surgery, endoscopic cameras natively guide procedures, and videos can be easily recorded. Video recording of open surgery is not similarly intuitive given the lack of ad hoc cameras; however, videos using cameras mounted on the surgeon’s head or on the OR light can still be acquired with modest efforts.

Computer vision offers a unique opportunity to understand, quantify, and provide feedback on surgeon-patient interactions captured by endoscopic videos (Fig. 1). Technically, machine-learning pipelines are set to extract visual features from endoscopic images and then classify those using techniques such as support vector machines, random forests, and hidden Markov models. More recently, the field was boosted by deep-learning approaches based on multilayer architectures capable of progressively extracting higher-level features from raw inputs [35]. Semantically, low-level features (e.g. tool usage and anatomy) are usually extrapolated to represent higher-level surgical concepts (e.g. phase of the procedure) with the final aim of embodying the "language of behavior" or, more specifically, the "language of surgery", a meaningful representation of surgical activities [36].

Detecting which tool is present in each given image—one of the first tasks algorithms need to master in order to understand the surgical context—has been achieved with average areas under the ROC curve above 0.99 with models simultaneously deploying convolutional neural networks (CNNs) to extract visual features and recurrent neural networks (RNNs) to add the temporal context [37]. Furthermore, temporal dependencies of tool usage can be harnessed to infer the phase of a surgical procedure, as successfully achieved with a bi-directional long short-term memory (Bi-LSTM) [38]. Similarly, tracking tool movements in the surgical field can be used to extrapolate surgical skills metrics such as tool usage patterns, range, and economy of movements [39]. These works train models in a supervised fashion, i.e. using datasets of endoscopic images annotated with the information we want the model to output.

A key step towards the advent of surgical data science is collecting and publishing large and well-annotated datasets of surgical video, such as datasets of robotic and endoscopic images like Cholec80 [40] and M2CAI16 [41]. Indeed, fully supervised approaches are difficult to scale since manual annotations of large surgical databases is time-consuming and costly, requiring physicians to review and consistently annotate large sets of images. To overcome this limitation, approaches using less supervision to train models are being investigated. For instance, deep-learning algorithms capable of reliably localizing [42] and tracking [43] tools in endoscopic scenes have been trained on videos with only tool binary presence annotation. These models were trained in a weakly supervised manner since they learned to output the tool's position in space (i.e. tool coordinates frame by frame) by looking at videos on which only the tool's presence or absence was indicated, a much easier feature to annotate. Weakly or semi-supervised training approaches have also shown promising results in tool segmentation [44], prediction of remaining surgery duration [45], [46], and phase detection [47].

If adequately built and translated in the OR, these machine-learning information-extraction methods have the potential to substantially benefit surgical care in a number of ways. The most intuitive application of these extraction methods is in providing recommendations for alternative surgical instruments dependent upon the current surgical phase detected. Trainees, including surgical trainees and OR nurses, may not appreciate subtle distinctions in surgical instruments. Enhanced guidance at an early stage of learning could foreseeably minimize frustrations associated with teaching loads for attending surgeons and reduce total

procedure duration. In the case of an expert, such recommendations could be helpful reminders to correct an otherwise undetected mistake.

Offline, automated multimedia content analysis could enable easy browsing of videos for intraoperative “near miss” events analysis [48] and video-based surgical coaching [49], video summarization of surgical procedures for objective postoperative documentation and patient’s briefing [50], and sensor-free surgical technical skills assessment in the OR [51]. Real-time prediction algorithms could be used to assist surgeons and OR staff intraoperatively. For example, one could build a model capable of notifying the surgeon about risky deviations from the normal course of events or safety boundaries [52] and suggest implementation of best practices and evidences in surgery [53] by merging a thorough understanding of intraoperative decision-making processes, computer vision quantification of surgical workflow, and patient outcomes. Furthermore, continuous automated monitoring of the procedure and OR status could help inform staff of the optimal timing to call the next patient, allowing better resource allocation and scheduling to decrease costs. A mock-up of the above and other possible intraoperative feedback types is shown in Fig. 2. In addition, computer vision methods could enable the intraoperative use of advanced imaging technologies [54]. For instance, the spectral signature acquired through multi- and hyperspectral imaging can be analyzed for automated intraoperative tissue recognition [55] and residual tumor identification [56]. Altogether, these methods could feasibly contribute to the realization of a surgical “control tower” to monitor and augment surgical care with the final aim of improving patient outcomes [35], [57].

IV. COMPUTER VISION DETECTING AND AFFECTING BEHAVIOR

Behavioral sensing and video recording in the OR have many potential applications for research, quality improvement, and education [58], [59]. As opposed to reliance on individuals to report safety issues during a procedure, video technology can passively and objectively record how healthcare personnel perform their jobs. The faster intraoperative data can be analyzed, interpreted and presented to the clinician, the more useful it can be. Armellino described third-party remote video auditing and real-time feedback to evaluate healthcare personnel’s hand hygiene [60]: during the 16-week pre-feedback period, hand hygiene rates were less than 10% (3,933/60,542) and in the 16-week post-feedback period it was 81.6% (59,627/73,080). These data suggest that video auditing combined with real-time feedback can produce significant and sustained improvement in human behavior.

Guerlain and colleagues at the University of Virginia developed a customizable digital recording and analysis system for studying human performance in the operative environment [58]. Their Remote Analysis of Team Environment (RATE) tool is a digital audiovisual data collection and analysis system that automates the ability to digitally record, score, annotate and analyze team performance. In ten laparoscopic cholecystectomy cases, the RATE tool allowed real-time, multi-track data collection of all aspects of the operative environment, while permitting digital recording of the objective assessment data in a time-synchronized and annotated fashion during the procedure. Interestingly, measures of situational awareness, an important non-technical skill, varied widely among team members, with the attending

surgeon typically the only team member having comprehensive knowledge of critical case information [58].

Similarly, Grantcharov and colleagues have developed a multi-source acquisition system, the OR Black Box™, to capture intraoperative data from patients, providers, and the environment [59]. Analysis of these data to date has highlighted the high frequency of intraoperative distractions, including the observation of 138 auditory distractions per case [61]. Nara measured the position of medical personnel in the OR using ultrasonic sensor markers affixed on them [62] with the intent of estimating automatically the main patterns of motion (e.g. two nurses exchanging places), and found that the trajectories of the personnel were clustered.

Capturing behaviors of human agents through video recordings, however, may not reflect their typical behavior. The Hawthorne effect, first observed by French in 1953, posits that awareness of being observed induces behavioral change, often in the desired direction of the intervention and thereby conflating observed outcomes. Inconclusive evidence of the Hawthorne effect, however, has been documented through mixed results in fields such as car accidents as a result of traffic cameras, police aggression with body-worn cameras [63], and physician consultation behavior [64]. A recent systematic review investigating the presence and characteristics of the Hawthorne effect in health sciences research concluded that while evidence does support existence of the Hawthorne effect, evidence of its effect size and magnitude is largely inconclusive [65].

Despite the uncertainties surrounding the impact of the Hawthorne effect in the healthcare field, measures can be taken to overcome potential confounding effects. As an example, the research group implementing the OR Black Box™ built in a one-year lead-in period where cameras were installed in the OR, but analysis was not conducted on these data. Authors state that this was an attempt to habituate participants to new technology and minimize the potential influence of the Hawthorne effect on results collected during the project period [61]. Furthermore, the presence of video monitoring equipment is already ubiquitous in the many hospital settings and the recordings are protected by the Health Insurance Portability and Accountability Act (HIPAA) law that requires only those who need to know to have access. According to Section 164.514(a) of the HIPAA Privacy Rule, "*Health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual*".

To address the issue of potential patient and staff privacy concerns, video recordings may be processed with several methods of de-identification. Silas recorded OR videos using the Kinect 2.0 system (Microsoft Corp., Redmond, WA) through regular red green blue (RGB) video recording and infrared depth sensing. They used three de-identification modalities: (a) "blurred faces" was RGB video altered using a post-processing algorithm to blur out skin tone color spectra, while keeping the rest of the footage unaltered; (b) "infrared" was unedited data collected with the infrared depth-sensing camera, and (c) "point cloud" was the infrared data processed with an algorithm that assigns a point cloud and skeleton structure for each individual [66]. An additional technique involves replacing or merging the individual's face with a generic face to remove their identity [67].

An alternative approach, referred to as ambient sensing, provides the opportunity to detect behavior and behavioral deviations by triangulating data sources which do not necessarily include capturing identifiable information through video streams [68]. This approach capitalizes on the convergence among incoming data sources, including inputs from depth, thermal, radio, and acoustic sensors to detect human behavior [69]. One specific application in healthcare has included monitoring and intervening with hand hygiene behavior [70]. This approach could readily be extended to detect unanticipated patient states derived from technical missteps (e.g. bleeding due to suboptimal anastomosis) or to notify OR staff of potential surgical flow disruptions (e.g. detection of missing equipment) pre-emptively.

A. People Detection and Tracking

The state of the art in detecting people and objects may be currently achieved with deep networks [71], [72]. Key issues are detection rates and processing speed. Current detection rates approach 90% in cluttered environments, while processing speeds range between 10Hz and 100Hz.

The state of the art in tracking is achieved either by direct matching [73] or by concatenating the detector with a combinatorial evaluation of the optimal continuation of each trajectory [74]. Person identification may be carried out by face analysis [75], [76] or by exploiting differences in clothing and ad-hoc individual markings in top-down views where the face may not be visible [77]. Accurate detection, tracking and identification have recently become available as inexpensive cloud-based services from a number of commercial providers (AWS, Google, Microsoft and others).

B. Pose Analysis and Action Detection

The state of the art in pose analysis may be achieved by hourglass deep networks [26]. The state of the art in action detection and classification in continuous video is currently achieved by time-series classifiers that are applied to the position and pose trajectories of the actors' bodies [78].

OpenPose [79] is an open-source, deep-learning enabled computer vision system capable of detecting multiple humans and labeling up to 25 key body points using 2D video input from conventional cameras. The architecture of this system uses a two-branch multi-stage CNN in which each stage in the first branch predicts confidence 2D maps of body part locations, and each stage in the second branch predicts Part Affinity Fields (PAF) which encode the degree of association between parts. Training and validation of the OpenPose algorithms were evaluated on two benchmarks for multi-person pose estimation: the MPII human multi-person dataset and the COCO 2016 keypoints challenge dataset. Both datasets had images collected from diverse real-life scenarios, such as crowding, scale variation, occlusion and contact. The OpenPose system exceeded previous state-of-the-art systems [79].

C. Team Interactions

The dynamics of robotic surgical team activities may provide relevant information for understanding the multitude of factors that impact surgical performance and patient safety outcomes [80]. Measures extracted through computer vision, such as robotic team centrality,

team proximity, and face-to-face communication may provide important insights into team coordination and dynamics metrics, with the advantage of being automatically generated, instead of human-annotated [81].

The integration of team motion analysis through computer vision with psychophysiological data (e.g. heart rate variability) has also been proposed as measure of team dynamics in the context of cognitive load monitoring, a valuable component of cognitive engineering approaches in the surgical context [5]. Fig. 3 shows an integrated visualization of motion tracking data and cognitive load (heart rate variability) during a real-life cardiac surgery.

D. Situational Awareness (SA)

Through computer vision, not only low-level semantics can be extracted from the OR (e.g. workflow segmentation, instrument identification) but also contextual information can be inferred from body motion, eye movement data, and head poses [82]–[84]. For instance, a surgeon's gaze entropy and velocity have been shown to be related to surgical procedure complexity and can differentiate expert from novice surgeons [85], [86]. However, although computer vision may provide useful inferences about the perception component (e.g. visual attention via eye gaze) of the SA construct, the other SA components (comprehension and projection) are more difficult to be inferred based only on visual data [87]. Multi-source approaches that integrate visual, audio and physiological data may be more successful in extracting automated measures of SA in the OR. Although computer vision per se has some limitations in measuring SA, it may provide useful information to support the surgical team during situations that require a high SA. For example, if a computer vision application can detect bleeding in the surgical site that was not perceived by the surgeon, it can alert the surgical team and even pinpoint the exact site of bleeding, enhancing the SA in the OR. This is an example of augmented cognition in a distributed cognitive system composed by human (surgical team) and non-human (computer vision application) teammates [88]. While the computer vision application can augment human performance, however, it cannot automatically achieve OR safety in its current state.

V. CHALLENGES OF COMPUTER VISION IN SURGERY

Computer vision in surgery is still in its infancy and, even if highly promising, a number of challenges need to be addressed to fully exploit the potential benefits coming from the intersection of these disciplines (Table I).

First, granular OR data need to be acquired, moved, stored, annotated, and queried in an efficient way. Though medicine and surgery are data-intensive disciplines, high-quality, structured and diverse information is rarely available. Various groups are proposing data acquisition systems, shared standards for device integration, and scalable infrastructures for data transmission and information generation. For example, the CONDOR (Connected Optimized Network and Data in Operating Rooms) project has promoted the DICOM-RTV (Digital Imaging and Communications in Medicine Real-Time Video) standard for surgical videos, an over-IP system for data streaming and methods for knowledge extractions (https://condor-project.eu/). Similarly, the OR Black Box™ (Surgical Safety Technologies Inc, Toronto, ON) enables synchronized acquisition through cameras, microphones, and digital

monitors providing valuable insights to optimize surgical safety [59]. Once acquired, these data need to be processed and consistently labeled, a burdensome task especially if surgical knowledge needs to be annotated. To take this duty off of surgeons' shoulders and decrease cost, in addition to the above discussed methods needing less supervision to train, crowds' annotation is being investigated with promising results [89]. Crowdsourcing methods based on trained workforces are increasingly becoming commercially available.

When it comes to methods development, computer vision in surgery faces domain-specific challenges. Environmental and endoscopic cameras don't always capture all the meaningful information needed to answer a specific question. Endoscopy-specific limitations include fast changing scenes due to rapid camera movements, changes in luminosity, and organ deformation. Furthermore, patient-specific anatomical variations and the act of performing surgery itself significantly contribute to inter- and intra-endoscopic scenes variability. In environmental OR videos, the staff is typically in scrubs and wearing masks, making recognition tasks more challenging, a limitation that can be overcome by identifying and tracking distinguishing features. Additionally, the state of the art in human detection, pose computation, tracking, and action classification are far from perfect when humans are crowded together [90], which is common in the OR. Thus, current methods will have to be validated in the specific scenarios of interest for surgeons and further technical progress is likely to be needed for accurate performance.

Factors including the enthusiasm surrounding artificial intelligence in medicine, the inherent complexity of machine learning models, and the naïve surgical audience targeted could hamper the critical appraisal of the literature on computer vision in surgery and lead to hyped claims. A framework for evaluating computer vision in surgery has not yet been elaborated, however lessons can be drawn from white papers and guidelines developed for related fields [91]–[93]. Main indications include careful description of the data the model has been built upon, selection of performance metrics meaningful to the expected clinical use, evaluation of clinical value of the models' success and error cases, and validation on external dataset sets to verify generalizability of findings. It must be stressed that with current machine-learning approaches large, diverse, and well-annotated datasets are crucial to success. Aside from the more obvious pitfalls of data, including failing to represent reality and be inclusive, datasets may contain subtle subclasses affecting model performances [94] or may incorporate unexpected elements to infer outcomes. For example, a deep-learning melanoma detection model was found to recognize skin markers as signs frequently used to flag worrisome lesions instead of melanomas, with a 40% false-positive rate in this subset of images [95].

Another domain-specific challenge is to more fully understand the model's logic [96]. Explainable machine-learning algorithms, capable of outputting both predictions and the logic behind those, follow various approaches; for example, disease markers can be sequentially removed from the model function to evaluate the weight of each single one, segmentation could be used to highlight region of interest in images, and verbal explanations may be generated automatically [97].

Finally, patients, surgical teams, computer scientists, and healthcare systems alike need to set up a constructive communication strategy to define values, bridge needs, and cross cultural barriers. For this to happen, privacy, legal liability, and ethical concerns must be resolved [98]. Patients and surgeons should be at the center through reciprocal education in order to define valuable use-cases, share data, and critically understand models' performances. Similar to what happens in aviation, computer vision vigilance to improve surgical safety should be deployed in a non-punitive but constructive environment for widespread adoption. Furthermore, deep understanding and consistent definition of team dynamics [99] and intraoperative events [100], collaboration, and consensus are needed to scale the approach and impact surgical practices.

To realize the ultimate goal of enhanced patient safety, it is critical that researchers and clinicians across disciplines work closely to overcome these engineering and clinical challenges. Resolving technical obstacles related to data acquisition and algorithm development is not sufficient to improve workflows and instill positive changes in safety-enhancing behavior. To achieve optimal OR safety, computer vision models must also be interpreted appropriately and adopted reliably by end users.

VI. CONCLUSIONS

Effectiveness of computer vision approaches and techniques has been widely demonstrated through a number of applications, both within and outside of the realm of healthcare. The opportunity to apply these principles to surgery specifically is underexplored, but has the potential for significant public health, patient safety, and economic impact. As the fields contributing to the interdisciplinary domain of computer vision continue to advance, collaboration and coordination will be paramount to ensuring the highest levels of success.

Acknowledgments

This study is funded by R01HL126896 from the National Heart, Lung, and Blood Institute of NIH (PI Zenati). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. N. Padoy acknowledges the support from French State Funds managed by the Agence Nationale de la Recherche (ANR) through the Investissements d'Avenir Program under Grant ANR-16-CE33-0009 (DeepSurg), Grant ANR-11-LABX-0004 (Labex CAMI), Grant ANR-10-IDEX-0002-02 (Idex Unistra) and Grant ANR-10-IAHU-02 (IHU Strasbourg) and by BPI France through Project CONDOR.

VII. References

[1]. Howell A-M, Panesar SS, Burns EM, Donaldson LJ, and Darzi A, "Reducing the Burden of Surgical Harm," Ann. Surg, vol. 259, no. 4, pp. 630–641, 4. 2014. [PubMed: 24368639]

[2]. Calland JF, Guerlain S, Adams RB, Tribble CG, Foley E, and Chekan EG, "A systems approach to surgical safety," Surg. Endosc, vol. 16, pp. 1005–1014, 2002. [PubMed: 12000985]

[3]. Wahr JA et al., "Patient safety in the cardiac operating room: Human factors and teamwork: A scientific statement from the american heart association," Circulation, vol. 128, no. 10, pp. 1139–1169, 2013. [PubMed: 23918255]

[4]. Rosen MA, Dietz AS, Yang T, Priebe CE, and Pronovost PJ, "An integrative framework for sensor-based measurement of teamwork in healthcare," J. Am. Med. Informatics Assoc, vol. 22, no. 1, pp. 11–18, 2014.

[5]. Zenati MA, Kennedy-Metz L, and Dias RD, "Cognitive Engineering to Improve Patient Safety and Outcomes in Cardiothoracic Surgery," Semin. Thorac. Cardiovasc. Surg, pp. 1–7, 2019.

[6]. Gawande A and Weiser T, “WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives,” 2009.

[7]. Norman S, “Google car takes the test,” Nature, vol. 514, no. 7253, p. 528, 2014.

[8]. Gomes JFS and Leta FR, “Applications of computer vision techniques in the agriculture and food industry: A review,” Eur. Food Res. Technol, vol. 235, no. 6, pp. 989–1000, 2012.

[9]. Gruen A, “Development and Status of Image Matching in Photogrammetry,” Photogramm. Rec, vol. 27, no. 137, pp. 36–57, 2012.

[10]. de Mello LSH and Lee S, Computer-Aided Mechanical Assembly Planning. 2012.

[11]. Jones GA, Paragios N, and Regazzoni CS, Video-Based Surveillance Systems: Computer Vision and Distributed Processing. 2012.

[12]. Davies ER, “Machine vision in the food industry,” in Robotics and Automation in the Food Industry: Current and Future Technologies, 2013.

[13]. Fuchs TJ and Buhmann JM, “Computational pathology: Challenges and promises for tissue analysis,” Comput. Med. Imaging Graph, vol. 35, no. 7–8, pp. 515–530, 2011. [PubMed: 21481567]

[14]. Campanella G et al., “Clinical-grade computational pathology using weakly supervised deep learning on whole slide images,” Nat. Med, vol. 25, no. 8, pp. 1301–1309, 2019. [PubMed: 31308507]

[15]. Mansoor A et al., “Segmentation and image analysis of abnormal lungs at CT: Current approaches, challenges, and future trends,” Radiographics, vol. 35, no. 4, pp. 1056–1076, 2015. [PubMed: 26172351]

[16]. McKinney SM et al., “International evaluation of an AI system for breast cancer screening,” Nature, vol. 577, no. 7788, pp. 89–94, 1. 2020. [PubMed: 31894144]

[17]. Väyrynen JP, Vornanen JO, Sajanti S, Böhm JP, Tuomisto A, and Mäkinen MJ, “An improved image analysis method for cell counting lends credibility to the prognostic significance of T cells in colorectal cancer,” Virchows Arch, vol. 460, no. 5, pp. 455–465, 2012. [PubMed: 22527018]

[18]. Doukas C, Stagkopoulos P, Kiranoudis CT, and Maglogiannis I, “Automated skin lesion assessment using mobile technologies and cloud platforms,” in 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012, pp. 2444–2447.

[19]. Esteva A et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115–118, 2. 2017. [PubMed: 28117445]

[20]. Perona P and Malik J, “Scale-Space and Edge Detection Using Anisotropic Diffusion,” IEEE Trans. Pattern Anal. Mach. Intell, vol. 12, no. 7, pp. 629–639, 1990.

[21]. Misawa M et al., “Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience,” Gastroenterology, vol. 154, no. 8, pp. 2027–2029.e3, 2018. [PubMed: 29653147]

[22]. Luo H et al., “Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study,” Lancet Oncol, vol. 20, no. 12, pp. 1645–1654, 2019. [PubMed: 31591062]

[23]. Sapiro G, Hashemi J, and Dawson G, “Computer vision and behavioral phenotyping: an autism case study,” Curr. Opin. Biomed. Eng, vol. 9, pp. 14–20, 2019.

[24]. Murino V, Cristani M, Shah S, and Savarese S, Group and crowd behavior for computer vision. Academic Press, 2017.

[25]. Benitez-Quiroz CF, Srinivasan R, and Martinez AM, “EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, vol. 2016-Decem, pp. 5562–5570, 2016.

[26]. Newell A, Yang K, and Deng J, “Stacked hourglass networks for human pose estimation,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9912 LNCS, pp. 483–499, 2016.

[27]. Monfort M, Liu A, and Ziebart BD, “Intent Prediction and Trajectory Forecasting via Predictive Inverse Linear-Quadratic Regulation,” Proc. Twenty-Ninth AAAI Conf. Artif. Intell, pp. 3672–3678, 2015.

[28]. Eyjolfsdottir E, Branson K, Yue Y, and Perona P, “Learning recurrent representations for hierarchical behavior modeling,” in ICLR 2017, 2016, pp. 1–12.

[29]. Le HM, Yue Y, Carr P, and Lucey P, “Coordinated multi-agent imitation learning,” 34th Int. Conf. Mach. Learn. ICML 2017, vol. 4, pp. 3140–3152, 2017.

[30]. Anderson DJ and Perona P, “Toward a science of computational ethology,” Neuron, vol. 84, no. 1, pp. 18–31, 2014. [PubMed: 25277452]

[31]. S. R. Datta, D. J. Anderson, K. Branson, P. Perona, and A. Leifer, “Computational Neuroethology: A Call to Action,” Neuron, vol. 104, no. 1, pp. 11–24, 2019. [PubMed: 31600508]

[32]. Loftus TJ et al., “Artificial Intelligence and Surgical Decision-making,” JAMA Surg, vol. 155, no. 2, p. 148, 2. 2020. [PubMed: 31825465]

[33]. Fecso AB, Kuzulugil SS, Babaoglu C, Bener AB, and Grantcharov TP, “Relationship between intraoperative non-technical performance and technical events in bariatric surgery,” Br. J. Surg, vol. 105, no. 8, pp. 1044–1050, 2018. [PubMed: 29601079]

[34]. Panesar S, Cagle Y, Chander D, Morey J, Fernandez-Miranda J, and Kliot M, “Artificial Intelligence and the Future of Surgical Robotics,” Ann. Surg, vol. 270, no. 2, pp. 223–226, 2019. [PubMed: 30907754]

[35]. Padoy N, “Machine and deep learning for workflow recognition during surgery,” Minim. Invasive Ther. Allied Technol, vol. 0, no. 0, pp. 1–9, 2019.

[36]. Vercauteren T, Unberath M, Padoy N, and Navab N, “CAI4CAI : The Rise of Contextual Artificial Intelligence in Computer Assisted Interventions,” in Proceedings of the IEEE, 2019, pp. 1–17.

[37]. Al Hajj H, Lamard M, Conze PH, Cochener B, and Quellec G, “Monitoring tool usage in surgery videos using boosted convolutional and recurrent neural networks,” Med. Image Anal, vol. 47, pp. 203–218, 2018. [PubMed: 29778931]

[38]. Mondal SS, Sathish R, and Sheet D, “Multitask Learning of Temporal Connectionism in Convolutional Networks using a Joint Distribution Loss Function to Simultaneously Identify Tools and Phase in Surgical Videos,” pp. 1–15, 2019.

[39]. Jin A et al., “Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks,” Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018, vol. 2018-Janua, pp. 691–699, 2018.

[40]. Twinanda AP, Shehata S, Mutter D, Marescaux J, De Mathelin M, and Padoy N, “EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos,” IEEE Trans. Med. Imaging, vol. 36, no. 1, pp. 86–97, 2017. [PubMed: 27455522]

[41]. Twinanda AP, Mutter D, Marescaux J, de Mathelin M, and Padoy N, “Single- and Multi-Task Architectures for Surgical Workflow Challenge at M2CAI 2016,” 2016.

[42]. Vardazaryan A, Mutter D, Marescaux J, and Padoy N, “Weakly-supervised learning for tool localization in laparoscopic videos,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11043 LNCS, pp. 169–179, 2018.

[43]. Nwoye CI, Mutter D, Marescaux J, and Padoy N, “Weakly supervised convolutional LSTM approach for tool tracking in laparoscopic videos,” Int. J. Comput. Assist. Radiol. Surg, vol. 14, no. 6, pp. 1059–1067, 2019. [PubMed: 30968356]

[44]. Fuentes-Hurtado F, Kadkhodamohammadi A, Flouty E, Barbarisi S, Luengo I, and Stoyanov D, “EasyLabels: weak labels for scene segmentation in laparoscopic videos,” Int. J. Comput. Assist. Radiol. Surg, vol. 14, no. 7, pp. 1247–1257, 2019. [PubMed: 31165349]

[45]. Rivoir D, Bodenstedt S, von Bechtolsheim F, Distler M, Weitz J, and Speidel S, “Unsupervised Temporal Video Segmentation as an Auxiliary Task for Predicting the Remaining Surgery Duration,” in OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging, 2019, pp. 29–37.

[46]. Yengera G, Mutter D, Marescaux J, and Padoy N, “Less is More: Surgical Phase Recognition with Less Annotations through Self-Supervised Pre-training of CNN-LSTM Networks,” 2018.

[47]. Yu F et al., “Assessment of Automated Identification of Phases in Videos of Cataract Surgery Using Machine Learning and Deep Learning Techniques,” JAMA Netw. open, vol. 2, no. 4, p. e191860, 2019. [PubMed: 30951163]

[48]. Bonrath EM, Gordon LE, and Grantcharov TP, “Characterising ‘near miss’ events in complex laparoscopic surgery through video analysis,” BMJ Qual. Saf, vol. 24, no. 8, pp. 516–521, 2015.

[49]. Greenberg CC et al., "A Statewide Surgical Coaching Program Provides Opportunity for Continuous Professional Development," Ann. Surg, vol. 267, no. 5, pp. 868–873, 2018. [PubMed: 28650360]

[50]. Van De Graaf FW et al., "Comparison of Systematic Video Documentation with Narrative Operative Report in Colorectal Cancer Surgery," JAMA Surg, vol. 154, no. 5, pp. 381–389, 2019. [PubMed: 30673072]

[51]. Vedula SS, Ishii M, and Hager GD, "Objective Assessment of Surgical Technical Skill and Competency in the Operating Room," Annu. Rev. Biomed. Eng, vol. 19, no. 1, pp. 301–325, 2017. [PubMed: 28375649]

[52]. Hashimoto DA et al., "Surgical procedural map scoring for decision-making in laparoscopic cholecystectomy," Am. J. Surg, vol. 217, no. 2, pp. 356–361, 2019. [PubMed: 30470551]

[53]. Smith AB and Brooke BS, "How Implementation Science in Surgery is Done," JAMA Surg, vol. 154, no. 10, 2019.

[54]. Mascagni P et al., "New intraoperative imaging technologies: Innovating the surgeon's eye toward surgical precision," J. Surg. Oncol, vol. 118, no. 2, pp. 265–282, 2018. [PubMed: 30076724]

[55]. Baltussen EJM et al., "Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery," J. Biomed. Opt, vol. 24, no. 01, p. 1, 2019.

[56]. Panasyuk SV et al., "Medical hyperspectral imaging to facilitate residual tumor identification during surgery," Cancer Biol. Ther, vol. 6, no. 3, pp. 439–446, 2007. [PubMed: 17374984]

[57]. Mascagni P and Padoy N, "OR Black Box and Surgical Control Tower: recording and streaming data and analytics to improve surgical care," J. Visc. Surg, 2020.

[58]. Guerlain S et al., "Assessing team performance in the operating room: Development and use of a 'black-box' recorder and other tools for the intraoperative environment," J. Am. Coll. Surg, vol. 200, no. 1, pp. 29–37, 2005. [PubMed: 15631917]

[59]. Goldenberg MG, Jung J, and Grantcharov TP, "Using Data to Enhance Performance and Improve Quality and Safety in Surgery," vol. 152, no. 10, pp. 972–973, 2017.

[60]. Armellino D et al., "Using high-technology to enforce low-technology safety measures: The use of third-party remote video auditing and real-time feedback in healthcare," Clin. Infect. Dis, vol. 54, no. 1, pp. 1–7, 2012. [PubMed: 22109950]

[61]. Jung JJ, Juni P, Lebovic G, and Grantcharov T, "First-year Analysis of the Operating Room Black Box Study," Ann. Surg, vol. XX, no. Xx, pp. 1–6, 2018.

[62]. Nara A, Allen C, and Izumi K, "Surgical phase recognition using movement data from video imagery and location sensor data," Adv. Geogr. Inf. Sci, pp. 229–237, 2017.

[63]. Ariel B et al., "Wearing body cameras increases assaults against officers and does not reduce police use of force: Results from a global multi-site experiment," Eur. J. Criminol, vol. 13, no. 6, pp. 744–755, 2016.

[64]. Pringle M and Stewart-Evans C, "Does awareness of being video recorded affect doctors' consultation behaviour?," Br. J. Gen. Pract, vol. 40, no. 340, pp. 455–458, 1990. [PubMed: 2271278]

[65]. McCambridge J, Witton J, and Elbourne DR, "Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects," J. Clin. Epidemiol, vol. 67, no. 3, pp. 267–277, 2014. [PubMed: 24275499]

[66]. Silas MR, Grassia P, and Langerman A, "Video recording of the operating room - Is anonymity possible?," J. Surg. Res, vol. 197, no. 2, pp. 272–276, 2015. [PubMed: 25972314]

[67]. Dale K, Sunkavalli K, Johnson MK, Vlasic D, Matusik W, and Pfister H, "Video face replacement," ACM Trans. Graph, vol. 30, no. 6, pp. 1–10, 2011.

[68]. Yeung S, Downing NL, Fei-Fei L, and Milstein A, "Bedside computer vision: Moving artificial intelligence from driver assistance to patient safety," N. Engl. J. Med, vol. 378, no. 14, pp. 1271–1273, 2018. [PubMed: 29617592]

[69]. Haque A, Milstein A, and Fei-Fei L, "Illuminating the dark spaces of healthcare with ambient intelligence," Nature, vol. 585, no. February, pp. 193–202, 2020. [PubMed: 32908264]

[70]. Haque A et al., “Towards Vision-Based Smart Hospitals: A System for Tracking and Monitoring Hand Hygiene Compliance,” in Proceedings of Machine Learning for Healthcare 2017, 2017, pp. 1–13.

[71]. Girshick R, Donahue J, Darrell T, and Malik J, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.

[72]. Kirillov A, He K, Girshick R, Rother C, and Dollar P, “Panoptic segmentation,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, vol. 2019-June, pp. 9396–9405, 2019.

[73]. Feichtenhofer C, “Detect to Track and Track to Detect Christoph,” Z. Vgl. Physiol, vol. 14, no. 4, pp. 709–736, 1931.

[74]. Burgos-Artizzu XP, Hall D, Perona P, and Dollár P, “Merging pose estimates across space and time,” BMVC 2013 - Electron. Proc. Br. Mach. Vis. Conf. 2013, pp. 1–11, 2013.

[75]. Taigman Y, Ranzato MA, Aviv T, and Park M, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014.

[76]. Grother PJ, Ngan M, and Hanaoka KK, Face Recognition Vendor Test (FRVT) Part 2: Identification. US Department of Commerce, National Institute of Standards and Technology, 2019.

[77]. Su C, Li J, Zhang S, Xing J, Gao W, and Tian Q, “Pose-Driven Deep Convolutional Model for Person Re-identification,” Proc. IEEE Int. Conf. Comput. Vis, vol. 2017-Octob, pp. 3980–3989, 2017.

[78]. Yeung S, Russakovsky O, Jin N, Andriluka M, Mori G, and Fei-Fei L, “Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos,” Int. J. Comput. Vis, vol. 126, no. 2–4, pp. 375–389, 2018.

[79]. Cao Z, Simon T, Wei S-E, and Sheikh Y, “Realtime multi-person 2d pose estimation using part affinity fields,” Proc. IEEE Int. Conf. Comput. Vis, pp. 7291–7299, 2017.

[80]. Tiferes J et al., “The Loud Surgeon behind the Console: Understanding Team Activities during Robot-Assisted Surgery,” J. Surg. Educ, vol. 73, no. 3, pp. 504–512, 2016. [PubMed: 27068189]

[81]. Kozlowski SWJ, Ghao GT, Chang C-H, and Fernandez R, “Team Dynamics: Using Big Data to Advance the Science of Team Effectiveness,” in Big Data at Work: The Data Science Revolution and Organizational Psychology, Tonidandel S, King EB, and Cortina JM, Eds. New York, NY: Routledge, 2015, pp. 273–309.

[82]. Bardram JE and Nørskov N, “A context-aware patient safety system for the operating room,” in UbiComp 2008 - Proceedings of the 10th International Conference on Ubiquitous Computing, 2008, no. 5, pp. 272–281.

[83]. Twinanda AP, Alkan EO, Gangi A, de Mathelin M, and Padoy N, “Data-driven spatio-temporal RGBD feature encoding for action recognition in operating rooms,” Int. J. Comput. Assist. Radiol. Surg, vol. 10, no. 6, pp. 737–747, 6. 2015. [PubMed: 25847670]

[84]. Kellnhofer P, Recasens A, Stent S, Matusik W, and Torralba A, “Gaze360: Physically unconstrained gaze estimation in the wild,” in Proceedings of the IEEE International Conference on Computer Vision, 2019, vol. 2019-Octob, pp. 6911–6920.

[85]. Erridge S, Ashraf H, Purkayastha S, Darzi A, and Sodergren MH, “Comparison of gaze behaviour of trainee and experienced surgeons during laparoscopic gastric bypass,” Br. J. Surg, vol. 105, no. 3, pp. 287–294, 2. 2018. [PubMed: 29193008]

[86]. Diaz-Piedra C, Sanchez-Carrion JM, Rieiro H, and Di Stasi LL, “Gaze-based Technology as a Tool for Surgical Skills Assessment and Training in Urology,” Urology, vol. 107, pp. 26–30, 2017. [PubMed: 28666793]

[87]. Endsley MR, “Toward a Theory of Situation Awareness in Dynamic Systems,” Hum. Factors, vol. 37, no. 1, pp. 32–64, 1995.

[88]. Dias RD, Yule SJ, and Zenati MA, “Augmented Cognition in the Operating Room,” in Digital Surgery, Atallah S, Ed. Springer Nature Switzerland AG, 2020.

[89]. Créquit P, Mansouri G, Benchoufi M, Vivot A, and Ravaud P, “Mapping of crowdsourcing in health: Systematic review,” J. Med. Internet Res, vol. 20, no. 5, pp. 1–23, 2018.

[90]. Ronchi MR and Perona P, “Benchmarking and Error Diagnosis in Multi-instance Pose Estimation,” Proc. IEEE Int. Conf. Comput. Vis, vol. 2017-Octob, pp. 369–378, 2017.

[91]. Collins GS, Reitsma JB, Altman DG, and Moons KGM, “Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement,” BMC Med., vol. 13, no. 1, pp. 1–10, 2015. [PubMed: 25563062]

[92]. El Naqa I et al., “Machine learning and modeling: Data, validation, communication challenges,” Med. Phys, vol. 45, no. 10, pp. e834–e840, 2018. [PubMed: 30144098]

[93]. Luo W et al., “Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view,” J. Med. Internet Res, vol. 18, no. 12, pp. 1–10, 2016.

[94]. Oakden-Rayner L, “Exploring large scale public medical image datasets,” 2019.

[95]. Winkler JK et al., “Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition,” JAMA Dermatology, pp. 1135–1141, 2019.

[96]. Gordon L, Grantcharov T, and Rudzicz F, “Explainable Artificial Intelligence for Safe Intraoperative Decision Support,” JAMA Surg, pp. E1–E2, 2019.

[97]. Antol S et al., “VQA: Visual question answering,” Proc. IEEE Int. Conf. Comput. Vis, pp. 2425–2433, 2015.

[98]. O’Sullivan S et al., “Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery,” Int. J. Med. Robot. Comput. Assist. Surg, vol. 15, no. 1, pp. 1–12, 2019.

[99]. Dias RD et al., “Dissecting Cardiac Surgery: A Video-Based Recall Protocol to Elucidate Team Cognitive Processes in the Operating Room,” Ann. Surg, pp. 1–8, 2019.

[100]. Mascagni P et al., “Formalizing video documentation of the Critical View of Safety in laparoscopic cholecystectomy: a step towards artificial intelligence assistance to improve surgical safety,” Surg. Endosc, no. 0123456789, 2019.

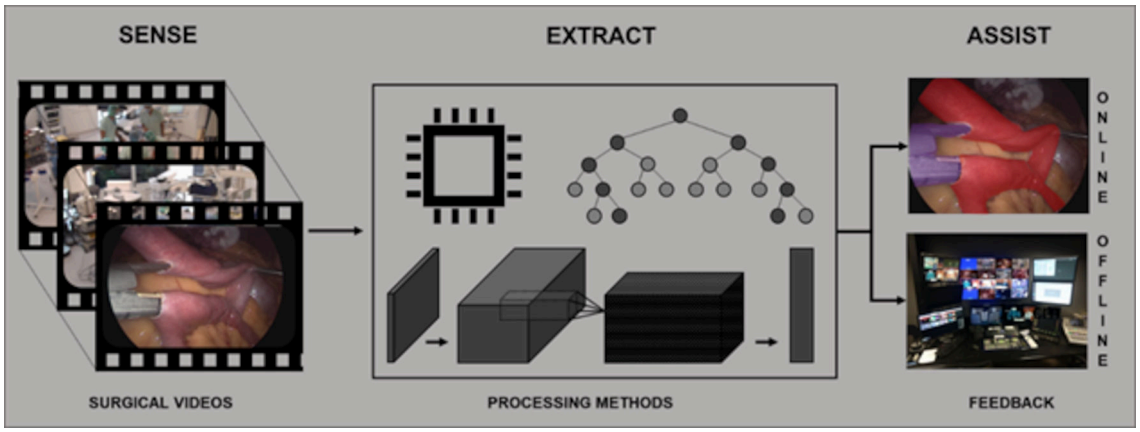

Fig. 1.

Workflow schematization. Surgical videos are recorded and processed to extract meaningful information that can be fed back to surgeons in real-time (i.e. online) or postoperatively (i.e. offline).

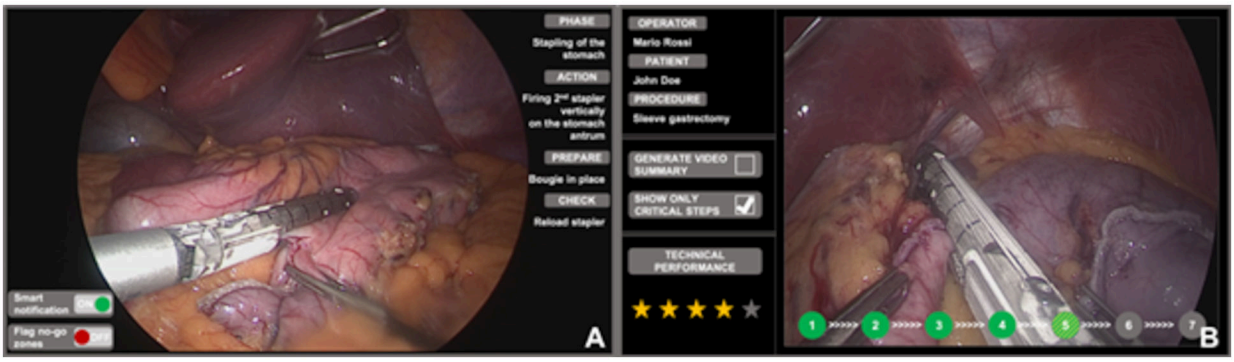

Fig. 2.

Mockup of intraoperative (A) and postoperative (B) assistance. Intraoperatively, surgeons can enable smart notifications to foster OR staff awareness and readiness (A, bottom left of the screen). Furthermore, surgeons and trainees can benefit of intraoperative guidance by overlying no-go areas on endoscopic images. Postoperatively, surgeons and trainees can review critical steps, generate video summaries and have their technical performance assessed.

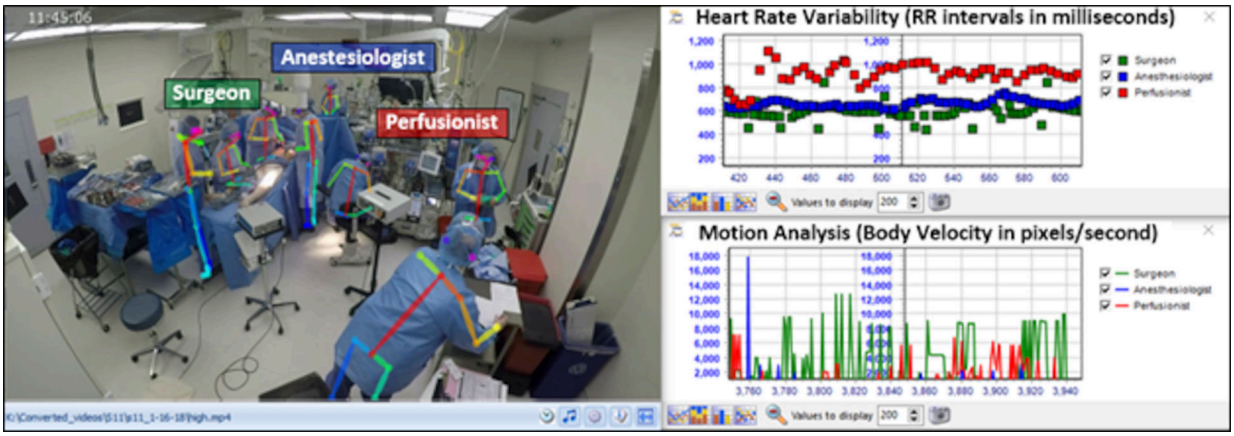

Fig. 3.

Integrated visualization of team motion tracking, cognitive load, and team dynamics in cardiac surgery.

TABLE I

Challenges of Computer Vision in Surgery

Domain	Challenges	Examples of proposed solutions
Data		
Capture	Systematic acquisition of endoscopic and OR videos	OR Black Box™[55]
Access	Transmission, storage and access of surgical images from multiple terminals	DICOM-RTV protocol, CONDOR infrastructure, in-hospital data centers
Label	Consistent annotation of relevant information	Annotation ontologies and software, crowdsourcing [75], automatic annotations
Models		
Surgical contingencies	Tissue manipulation, scrubbed staff, patient's variability, occluded views	Large, shared datasets [41], novel methods
Critical appraisal	Description of training data, evaluation of performances, case-studies, generalizability, validity	Guidelines on how to evaluate and read medical ML papers[79]
Interpretability	Understand factors influencing predictions, user interface design, correlation/causation dependencies	Explainable AI [82], causal inference
Cultural		
Legal	Health data privacy, regulatory clearance, medical liability	Legislation [84], ad-hoc protocols
Collaboration	Effective clinical-technical communication, definition of use-cases	Reciprocal education, interdisciplinary conferences and journals
Acceptance	Patients' and surgeons' acceptance of AI technologies, reimbursement frameworks	Surveys, incentives, long-term surveillance

An overview of factors contributing to current barriers in the implementation of computer vision approaches to surgery.

OR = operating room; DICOM-RTV = digital imaging and communications in imaging real-time video; CONDOR = Connected Optimized Network and Data in Operating Rooms; ML = machine learning; AI = artificial intelligence