HAL open science

Surgical Video Summarization: Multifarious Uses, Summarization Process and Ad-Hoc Coordination

Ignacio Avellino, Sheida Nozari, Geoffroy Canlorbe, Yvonne Jansen

▶ **To cite this version:**

Ignacio Avellino, Sheida Nozari, Geoffroy Canlorbe, Yvonne Jansen. Surgical Video Summarization: Multifarious Uses, Summarization Process and Ad-Hoc Coordination. Proceedings of the ACM on Human-Computer Interaction , 2021, 5 (CSCW1), pp.140:1-140:23. 10.1145/3449214 . hal-03160860

HAL Id: hal-03160860
https://hal.science/hal-03160860

Submitted on 5 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Surgical Video Summarization: Multifarious Uses, Summarization Process and Ad-Hoc Coordination

IGNACIO AVELLINO, Sorbonne Université, CNRS, INSERM, ISIR, France
SHEIDA NOZARI, Sorbonne Université, CNRS, INSERM, ISIR, France
GEOFFROY CANLORBE, Sorbonne Université, INSERM UMR_S_938, CRSA. AP-HP, Department of Gynecological and Breast Surgery and Oncology, Pitié-Salpêtrière University Hospital, France
YVONNE JANSEN, Sorbonne Université, CNRS, ISIR, France

While surgical videos are valuable support material for activities around surgery, their summarization demands great amounts of time from surgeons, resulting in the production of very few videos. We study the practices involving surgical video to motivate and inform the future design of tools for their summarization. Through interviews and observations in a field study, we find that (1) video summaries provide an important support for surgery, being used for self-improvement, education, discussing cases, scientific research, patient communication and as legal resources; (2) video summarization follows a process hindered by the loss of knowledge that originates during recording; and, (3) surgeons develop ad-hoc coordination strategies which involve using the video itself for articulation work, making it both the field of work and coordination artifact. We discuss ways in which tools can facilitate capturing knowledge during live action using these strategies.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: video summary, video summarization, coordination mechanisms, endoscopic video, minimally-invasive surgery

ACM Reference Format:
Ignacio Avellino, Sheida Nozari, Geoffroy Canlorbe, and Yvonne Jansen. 2021. Surgical Video Summarization: Multifarious Uses, Summarization Process and Ad-Hoc Coordination. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 140 (April 2021), 23 pages. https://doi.org/10.1145/3449214

1 INTRODUCTION

Surgery involves multiple actors working in complex collaboration: schedules constantly change and require re-coordination [8], coordination breakdowns when using imaging systems can demand pauses in surgery [44], and coordinating rooms requires complex visualizations [32]. Much of CSCW work on coordination in health involves "daily operation" objects like electronic patient records [9], non-digital artifacts such as whiteboards or post-its [7], paper operation schedules [6], paperwhat records [31], and even the layout of hallways and rooms [59], with the goal of understanding the underlying collaborative processes and increasing efficiency and quality. In this work, we bring to light another artifact central to the work of many surgeons: surgical video recordings.

Authors' addresses: Ignacio Avellino, ignacio.avellino@sorbonne-universite.fr, Sorbonne Université, CNRS, INSERM, ISIR, Paris, 75006, France; Sheida Nozari, shnozari@gmail.com, Sorbonne Université, CNRS, INSERM, ISIR, Paris, France; Geoffroy Canlorbe, geoffroy.canlorbe@aphp.fr, Sorbonne Université, INSERM UMR_S_938, CRSA. AP-HP, Department of Gynecological and Breast Surgery and Oncology, Pitié-Salpêtrière University Hospital, Paris, France; Yvonne Jansen, yvonne.jansen@sorbonne-universite.fr, Sorbonne Université, CNRS, ISIR, Paris, France.

© 2021 Association for Computing Machinery.
This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the ACM on Human-Computer Interaction*, https://doi.org/10.1145/3449214.

Videos are essential in Minimally-Invasive Surgery (MIS), as surgeons depend on a video feed to see their surgical instruments which were previously inserted into the patient through small holes on the skin. The unavoidable use of live video makes it easy to record surgery. However, for efficient reviewing, surgeons need to summarize video recordings as much of an hours-long surgery consists of routine activities. While other application areas, such as sports [22], have seen good results with automatic summarization tools, the context of surgical videos remains a challenge [37].

During informal conversations with surgeons, we identified that editing video is highly time-consuming, and therefore surgeons divide tasks among colleagues to mitigate this, along with some of the latent inefficiencies in the now unavoidable coordination work for summarization. We thus study the practices of surgeons surrounding video, from production to summarization and use, through interviews and observations of both surgery and the summarization process. Our goal is to motivate and inform the design of future systems that can support these practices and its coordination, which, as we find, is currently highly improvised and a manual craft.

The results elucidate more practices than initially thought. First, video summaries play an important role as they support surgical practice in multiple ways, including: self-improvement, education, discussing cases, scientific research, patient communication and as legal resources. Second, raw video recordings are summarized in a process with clear steps tailored to surgical content. Still, the underlying protocol is loosely formalized, and the process is inefficient as it is shaped by video-editing tools rather than latent domain needs, with breakdowns due to the loss of key information from the live activity to the summarization stage. Third, surgeons develop ad-hoc coordination strategies where they imprint coordination information on the video itself, a special case of a coordination mechanisms [60] where the field of work is used as a coordination artifact.

We conclude with implications for the design of tools that support coordination work in this particular setting. As this activity has two temporally-disjoint stages, first live surgery and later video summarization, effective coordination depends on the capture of situated knowledge (particular events to a procedure) and domain knowledge (general surgery) during the live stage, as well as process knowledge during the summarization stage. Different layers of information need to be associated through simple mechanisms that do not tax cognition during the live activity.

2 SURGERY AND THE ROLE OF VIDEO

Operating rooms are codified environments where different actors play specific roles during surgery, including: surgeons, residents who assist the surgeon, sterile nurses, circulating nurses and anesthesiologists. Sterility is a requirement for actors in close proximity to the patient, mostly the surgeon, assistant and sterile nurse. In Minimally-Invasive Surgery, instruments and a camera, called endoscope, are inserted into the patient through small orifices. MIS can also be performed through a robotic platform, where a multi-armed robot is controlled by a surgeon from a non-sterile console located inside the operating room. Residents train during surgery, observing surgeons work and performing parts of surgery under senior supervision [49]. The complex parallel activities impose heavy cognitive and physical demands on surgeons [15, 19, 68].

Surgeries are oftentimes recorded for later review, a practice that can reduce medical errors [9]. Surgeons have a vast domain knowledge of surgery in general and also of the procedures they carry out, as they know the particular case and they bear the decision-making responsibility during surgery. This makes them the most qualified people to summarize a video, yet they are the ones with the least amount of time to do it, as their schedules prioritize providing patient care through surgeries and consultations. As a consequence, surgeons engage in collaboration to summarize video recordings, as we learned through informal conversations, which requires coordination work.

3 COORDINATION WORK

Coordination is the management of dependencies among activities [41], a concept CSCW has largely developed including coordination mechanisms [58, 60], conceptualizing information systems as reading and writing artifacts [9], studying the temporal [6, 55], and physical [59] dimensions, and even how avoiding work relates to coordination [31].

In this article, we focus on the concept of coordination mechanisms: *"[..] a protocol, encompassing a set of explicit conventions and prescribed procedures and supported by a symbolic artifact with a standardized format, that stipulates and mediates the articulation of distributed activities so as to reduce the complexity of articulating distributed activities of large cooperative ensembles"* [60]. It is important for our research to unpack Simone et al.'s definition. First, *explicit* conventions and *prescribed* procedures give the notion of protocols being thought out and defined beforehand, even if the authors acknowledge that protocols work as plans for situated action, beacons that guide actors, that do not necessarily determine their course [64]. Second, the use of *symbolic* means that an artifact is different from the domain material: *"The state of the artifact is* ***de-coupled*** *from the state of the field of work"* [60], and thus changes to the state of one are not reflected on the other. Here, artifacts are conceptualized as material objects that objectify a coordinative construct, the protocol, giving place to *artifactually imprinted protocols* [58]. The material form has benefits: it provides communicative properties lacking in voice coordination, such as making them visible and accessible to the various actors, functioning as a checklist of information, making coordination communication persistent, and making the process flow explicit [14]. Third, the protocol stipulates and mediates the *articulation* of distributed activities. This is called *articulation work* [63], the "meta-work" on top of the domain work to establish division, allocation and scheduling of labor. CSCW scholars have argued that, instead of focusing on computer support for groups, the field should be concerned with how systems can improve actors' ability to articulate their activities [57].

3.1 Coordination Studies in Health

Coordination is a central aspect of the medical field and especially surgery. CSCW in the health context has focused on coordination of "daily operation" work, for example studying the coordinating role of medical records [9] as well as health practitioners coordinating equipment, patient preparedness, staffing, rooms, and schedules [48], with much of the interaction happening informally thanks to staff moving and working within a specific range area that increases chances of casual meetings [43]. Previous work has also identified that hospitals need to be reformulated, as surgery wards are constantly reorganizing [6], multiple coordinative artifacts are interwoven at a hospital ward, including artifacts such as whiteboards, work schedules, examination sheets, care records and post-it notes [7]. This complexity relates to the high amounts of information hospitals manage and its complex distribution, leading to "information hotspots" that act as physical spaces that facilitate coordination [59]. Finally, coordination in intra- and inter-departmental activities of transferring patients [1], and paper records in a radiology department [31].

In this research, we aim at achieving a better understanding of the articulation needs between actors involved in an object less central to daily operations but important for supporting activities: video recordings. We thus pave the way for an evidence-driven development of collaborative technologies tailored to this rich problem space.

4 VIDEO SUMMARIZATION

We provide an overview of the techniques used to summarize videos automatically or semi-automatically, to highlight the complexities of these approaches and the challenges for applying them to surgical videos. Approaches for summarization can be broadly divided into three categories: those using internal information sourced from the video to summarize, those using external information not coming from the video, and hybrid approaches combining internally and externally sourced information [46]. The final use of the summarized video also plays a role in selecting a summarization approach; for example, a video summarized to reduce its size aims to remove frames showing redundant information, whereas a video of a sports event should use more semantic information to identify and include all the highlights viewers may want to see.

Internal approaches depend exclusively on the actual video content to automatically generate a summarized video. The inference of meaning from low-level video cues is non-trivial and thus many approaches are specific to domains such as sports, music, or news. Approaches for extracting key frames from a video include detecting gestures [34, 35], color [25], motion activities [12, 20] and specific events, e.g., goals in sports or differences between frames in surveillance videos [16]. Internal approaches are challenging for surgical videos which generally do not contain any easily detectable shot changes, events nor differences in color.

External approaches use either information coming from an annotator or contextual information collected from the environment at the time of recording. This can be either explicitly by prompting human annotation, for example, to annotate sports events, or through sensors which collect metadata associated with video frames such as GPS coordinates [2], or floor pressure sensor readings [18]. Algorithms can then be parameterized to use these events when generating summarized videos. However, relying on human annotators is only semi-automatic and thus much slower and more costly than automatic approaches.

Minimally-invasive surgery videos in particular differ in a few important ways from much of the "regular" video content: (1) they are image only, i.e., no audio; (2) they are fundamentally a single long shot, i.e., there are no cuts or shot changes; (3) the colors are somewhat uniform, i.e., most of the video has some hue of red or pink except for out-of-patient frames which are often blue; (4) they are usually quite long, i.e., most surgeries last several hours. The literature proposes two different approaches for automatic summarization of surgical videos: (1) static summaries, that is, a static image summarizing a long video through a selection of the most relevant frames shown as small multiples [38], and (2) video summaries, that is, shortened versions of the full video. The caveat is that automatic summarization methods are currently not capable of reducing an hour-long video to a meaningful 5-minute summary since they exclusively focus on removing blurry frames [3, 51] or detecting out-of-patient frames [62]. Only Fan et al. claim a well-performing semi-automatic editing method based on human annotation of a 5–10% subset of frames [23]. However, the article only includes a technical evaluation with self-defined quality metrics, but does not offer any insights from surgeons concerning the quality of the generated video summary. In summary, it still remains unclear how useful an expert would find these semi-automatic summaries.

5 RESEARCH RATIONALE AND METHODOLOGY

Surgeons record, summarize and use videos, a practice that remains understudied, perhaps because of its secondary purpose. Our goals are to gain understanding of the practices surrounding video and to understand the coordination strategies during the summarization process. Ultimately, we aim at proposing mechanisms for supporting articulation work when summarizing videos, increasing the efficiency of the process and realizing the potential of this practice by increasing the amount of video summaries produced.

5.1 Participants

Nine surgeons from three different public hospitals located in the region of Paris, France, participated in the field study, and 31 surgeons in the survey. Figure 1 shows details of how each participant contributed. The participants' tight schedule highly influenced the sample size and the modalities on which each participant contributed. We were opportunistic when participants had an opening in their schedule. Our recruiting strategy was to contact surgeons who use video in our network through e-mail, asking them for recommendations of colleagues who also use video.

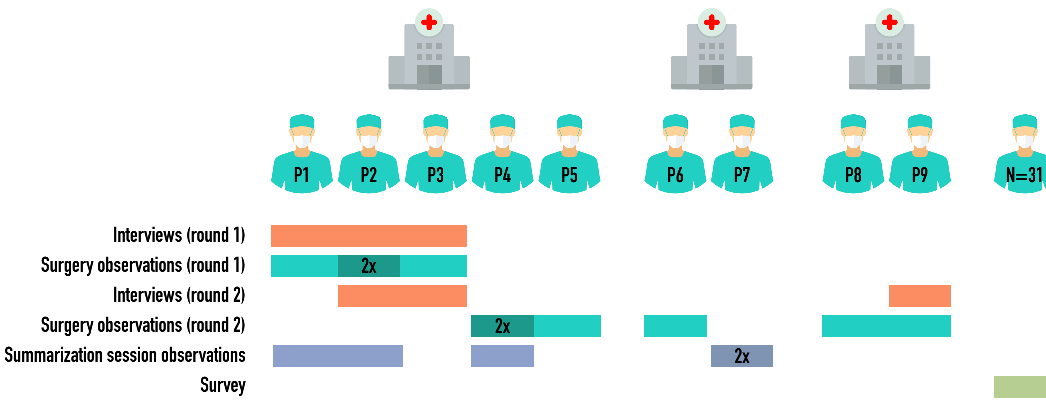

Fig. 1. Overview of participants and hospitals, and in which research phases they participated.

5.2 Ethical Considerations

At the time we conducted this study, the main institution leading it did not have an ethics board equipped to evaluate risks to participants of non-biomedical research. Nonetheless, all co-authors, except for one junior co-author, have received training on the ethical treatment of human subjects. Since the subject of study are the surgeons, and not the patients, we did not record any data concerning the patients' clinical history, nor any data or image that could identify them, only the endoscopic video. All participants read and signed (1) consent forms, which explicitly gave them the option to drop out at any time with no reason, and (2) image release forms, where they had consent choices for using their images at a granular level (e.g., in a class setting, for a publication, during a conference presentation). The data is stored in two encrypted hard drives to which only two of the researchers have access, and which cannot be linked to participant identities. The list matching participant identities with their participant id exists only in an encrypted location different from the data, to which two of the authors have access. Data was anonymized while coding when appropriate (e.g., removing the name of participants when referring to each other).

5.3 Method and Data Collection

Two initial surgery observations and numerous informal conversations in hospital hallways and break rooms informed the study design. Here, we realized that videos have a complex life cycle worthy of study from various perspectives in an ethnographic study. We conducted two rounds of interviews interleaved with two rounds of surgery observations and of video summarization sessions. Although we did not set strict criteria for determining data saturation, we scheduled further sessions until no new uses of video were mentioned and the principal investigator (first author) saw a complete picture of the summarization processes. We provide open access to supplementary materials through this link: https://osf.io/kjpu2.

5.3.1 Semi-Structured Interviews. We conducted in total six semi-structured interviews, lasting about 4 hours. We recorded audio and took written notes. The focus was on: (1) general aspects of surgery (interruptions, constraints to interacting with a system), (2) how residents learn during and after surgery and (3) the life cycle of videos, from their role during surgery to how they are summarized and used afterwards. We were initially also interested in how surgeons use live video during surgery, as the interview questions in the supplementary material show. However, we abandoned this direction after the first round of interviews as surgeons had difficulties thinking around this concept since they do not use live video for other purpose than performing surgery.

5.3.2 Surgery Observations. We observed in total ten surgeries, lasting about 16 hours, both classic and robotic using the *da Vinci* surgical robot (Intuitive Surgical, Mountain View, CA, USA) [28]. We recorded the endoscope video feed, wrote down notes, and recorded on paper: the intervention type and, for each person in the room, their role, if they are sterile, and whether they moved in the room or adopted a fixed position. We did not video-record patients from the outside, neither noted any medical information. The focus of our observation was the process of recording videos, the challenges of interacting with a system while performing surgery and the situated knowledge that determines what moments are relevant during surgery. While observing surgeries in round 1, we realized that our research would benefit from collecting data about the moments that impact the summarization process. Thus, for the surgery observations in round 2, we encouraged surgeons to indicate to us which moments were important as they unfolded, which we recorded in a paper grid (see supplemental material) noting down the time, surgeon's comments, the constraints both at the time the surgeon addressed us and when the moment actually took place, and the state of the members of the surgical team. We refer to these as *Key Moments*.

5.3.3 Video Summarization Observations. We observed in total five surgeons summarizing surgical videos, lasting about 4.5 hours. We video-recorded the sessions and took written notes. We asked surgeons to explain their actions and intentions while summarizing the videos, asking clarifying questions during the process. Almost all participants spoke in English, but none of them were native English speakers. Some spoke in French, in which case quotes have been translated to English.

5.3.4 Validation Survey. After analyzing the interviews and observations, we conducted a short survey to collect data on the usefulness and frequency of uses we found in our qualitative analysis. We inquired about (1) the time spent for summarization, (2) the frequency of summarization, and (3) for each of the uses described in section 6, the perceived usefulness and the frequency with which they summarize videos for that particular use case (a) given the summarization tools available today and (b) if they had more efficient tools at their disposal. The survey questions are included in the supplemental material. Regarding recruiting, we sent the survey through personal e-mail invitations to the nine surgeons who participated in the interviews and observations, and to further surgeons in our personal network who are interested in surgical video recordings. We received 31 responses out of 42 invitations (about 73%). The survey was open for 6 days.

5.4 Data Analysis

Qualitative Data. We performed Thematic Analysis [11] with the following considerations: initially, we did not rely on a particular theoretical positioning, but rather approached the data in an exploratory way, with an inductive approach, bounded to a constructivist epistemology [47]. During the coding process described below, the collaborative aspect of video summarization and the coordination intricacies became more and more salient. Thus, after finishing a first round of coding and analysis, we re-analyzed the data using the theoretical framing of coordination work, identifying articulation work, coordination strategies and protocols. Our coding varied between the latent and semantic level. The uses of video summaries was mostly at the semantic level, whereas the creation of video summaries and its collaborative nature was towards the latent level. The rationale behind these choices is that we do not intend to claim universal generalizations of our findings, but rather interpret particular examples of video summarization, bounded to the multiple factors that shape the ways this process is particularly performed as we observe it. We construct meaning around the challenges when coordinating this process, as we acknowledge that our observations are shaped by cultural influences of the particular participating hospitals, the country where they are located and the multiple factors of the participants (e.g., experience, race and gender).

Regarding data coding, two of the researchers independently listened to interviews, went over their notes, and watched the video-summarization sessions. They performed open coding on each data source by noting down the interesting parts linked to the time (for media data sources), the participant, and the interviewer. Codes amounted to about 43 different ones, focusing on the benefits and challenges of using and creating video summaries, for instance *"summarization process steps"*, *"summarization process - finding important moments"*, *"video summary uses - teaching"* or *"coordination mechanisms for summary"*. One of the researchers then created diagrams that described the data at different levels of abstraction, first grouping codes, then showing the codes and finally for each one concrete examples. Three of the researchers met and, together, discussed and clarified differences, making sense of the data. We did not seek inter-rater agreement as this would not be consistent with our analysis method [42]. We *constructed* meaning through discussion and evolved the codes when there were differences in understanding. These three are HCI researchers, part of an interaction & robotics lab with ongoing projects in surgery and rehabilitation for the past decade. One of them is western, with two years experience working with surgeons and has conducted previous studies both in the operating room and in simulation environments. Another one is middle-eastern, new to HCI. The third one is western and has 10 years of experience running qualitative and quantitative studies in HCI and visualization. A fourth and last researcher is a medical doctor and a university professor in a teaching hospital, with more than 10 years of experience in MIS.

Quantitative Data. Regarding the survey data, we present responses as histograms, without performing statistical analyses. Their main purpose is to illustrate on which questions most surgeons agree and on which opinions and habits vary considerably.

Before diving into the results, we clarify that we chose a logical order that puts the importance of videos at the forefront, starting by the uses of video summaries, which chronologically would correspond to the last part of their life cycle.

6 THE IMPORTANCE OF VIDEO SUMMARIES: MULTIFARIOUS USES

Video summaries play an inconspicuous but extensive role in the activities of surgeons. Their uses are manifold, which can justify the long time it takes to produce them in a complex editing process (described in the next section). We identified six different uses, which we now detail.

6.1 Personal: Self-Improvement

In preparation for a surgical procedure they have not performed in a long time, surgeons sometimes watch video summaries of a previous similar surgery. P2 mentioned a time when he checked how to exactly place surgical instruments during the critical phase of a procedure, and P3 recalled that earlier in the week he performed a surgery he had not done in a long time—a partial resection of the bladder—so the night before he searched his computer for videos of the procedure, while also searching the internet to compare against other surgeons' strategies. Moreover, he oftentimes checks the videos reviewed the night before also during the surgery. Participants mentioned searching videos in general-purpose repositories such as YouTube or specialized ones such as *WebSurg*[1], as well as sharing them among colleagues. Additionally, they also review materials before surgery by referring to a textbook containing an overview of the procedure with text and images, which is usually in their phones as a PDF, downloaded from the internet or shared among colleagues.

6.2 Education: Teaching Residents

Video summaries are widely used as teaching material. All surgeons in our study pointed out that surgical videos are useful for teaching, for instance to show strategies (P3), different ways to perform a particular gesture (P2), including how *not* to do a certain gesture (P1), or complications that can be used to trace mistakes (P3). P3 emphasized the importance of videos specifically for robotic surgery, where the techniques to operate are currently not taught in surgical school. Surgeons share videos with residents not only individually in one-to-one meetings, but also collectively in teaching sessions. P2 explained he had set up in his hospital a teaching program, where he and his team taught new residents using videos. Residents found this program useful and it had success, although the initiative was eventually abandoned as preparing the videos took too much time.

Our observation is consistent with literature on medical training: surgeons teach residents through a process of active learning, where students perform parts of surgeries during real cases [49], while complementing their knowledge through well-structured university teaching [30]. Also, literature has shown the advantages of using videos for teaching. First, debriefing videos after surgery has been shown to reduce technical errors in future surgeries [29]. Second, this benefits the learning process itself: students go over the material faster than with text and perform fewer attempts before successfully completing gestures [67].

6.3 Professional: Discussing Cases

Videos are also used in weekly staff meetings to discuss cases and make collective decisions based on concrete evidence. One way of presenting the severity of a disease in these meetings is through clinical scores, which are standardized measures that quantify severity. However, these scores can be biased, resulting in different values when computed by different people. Videos on the other hand can objectively convey the extent of a disease, such as cancer.

> *"In case of oncological surgery, we are performing sometimes what is called 'laparoscopic exploration' to have an idea of the spread of the disease inside the peritoneal cavity, for example,* ***before*** *beginning chemotherapy, or taking a decision of treatment. Today, the surgeon is performing the surgery and we will have some clinical score to describe the spread of the disease, then afterwards there is a staff meeting and the surgeon explains it to the others [..]. The scores are [..] the least subjective possible [measure], but, when we* ***see****, we all* ***know*** *how it is."* — P1

6.4 Scientific: Conferences and Studies

Conference submissions are one of the most notable uses of video summaries for scientific purposes in medicine. Several participants explained that a common format for such videos is a summary that shows *"the 10 steps to"* (P3) a procedure. Even if a video summary in this format can seem artificial, as it may hide routine steps, it facilitates conveying both the standard procedure steps and the unexpected moments. Submission tracks generally have a specific focus, such as rare procedures, particular anatomies (e.g., a large uterus (P3)), particular complications (P9), or even showing compilations of your worst moments (P3); their length can typically vary between 2 to 10 minutes. When invited to submit to a conference, surgeons in our study said that they look in their

[1] https://www.ircad.fr/e-learning/websurg/

computers for videos they have already recorded that fit the submission topic, as frequently raw videos are stored for long periods before they are summarized. P9 and P8 highlighted however that, although they remember having performed a surgery that fits, it happens that they simply cannot find the video and are even unsure if they have actually recorded it.

Participation in large randomized clinical studies is another use case for summarized videos, where each surgery constitutes a data point. P2 explained that videos can be used for two purposes here: first, to certify a surgeon's skills to the principal investigator and meet the criteria for participating in the study, and second, once a surgeon participates in a study, videos can be used to report the outcome of a surgery instead of reporting a numerical score.

Lastly, video summaries can replace live broadcasting of surgery during a congress, known as Live Surgical Broadcast (LSB). P3 pointed out that more and more conferences replace LSB with longer, 20-minute video summaries. The reason being that LSB has questionable ethics towards patients, as it increases risk, exacerbates care access inequality, and compromises respect and dignity from breaching confidentiality [54]. For instance, patients might feel financially coerced to consent to a surgeon's request for broadcasting their case live, against their will, only because surgery cost might be waived [53]. Moreover, patients might undergo unnecessary risk if surgeons perform an innovative technique to increase their reputation, as doing so in front of selected peers increases stress [21]. Video recordings are considered a viable alternative for LSB [61], as they can be appropriately edited, narrated live, and paused for discussions in a panel or with the audience.

6.5 Public: Information and Communication

Patients and the general public could also be a potential audience of video summaries. P2 suggested using videos to explain patients the procedure they will undergo, or have already undergone. This practice is not common today, and perhaps the reason relates to potential challenges on viewers' sensitivity, although this could be addressed by using filters to reduce affective responses [10].

> *"Maybe we can show to patients the procedure just to explain [..] it may be a little bit frightening for patients to see blood and flesh and bones [..] so maybe the system could make a specific image, just to have a more beautiful image [..] into a more understanding thing for non medical people"* — P2

6.6 Legal: Litigation

While there are currently no legal obligations to record surgeries in the hospitals included in our study, this may very well be the case in the future. Three participants foresee that videos might be used as medico-legal traces for surgeries. P3 recalled a personal experience where a post-operative report included a particular procedure being performed, for which he found no traces while exploring the anatomy at the beginning of a subsequent surgery. Here, a video could have provided a ground truth to this inconsistency.

P9 keeps some video records for legal purposes, in spite of there not being an obligation. As a head-and-neck surgeon, she could damage facial nerves and cause facial paralysis, a devastating aesthetic risk of surgery. Thus, she records muscle stimulation at the end of surgeries by sending an electrical charge through the nerve, which contracts facial muscles and is detected by an electrode, proving that the nerve communicates with the muscle and is not damaged. The video recording system lets her overlay the live video (in this case from a microscope) with the result of the electro-stimulation test. In this way, she can record herself sending the electrical charge with the nerve's response, constituting a more solid proof than writing in the surgical report that the test showed no nerve damage. P2, however, warned about an unintended effect of having videos as part of surgical reports, as surgeons can misjudge other surgeons' work only because they have a different style.

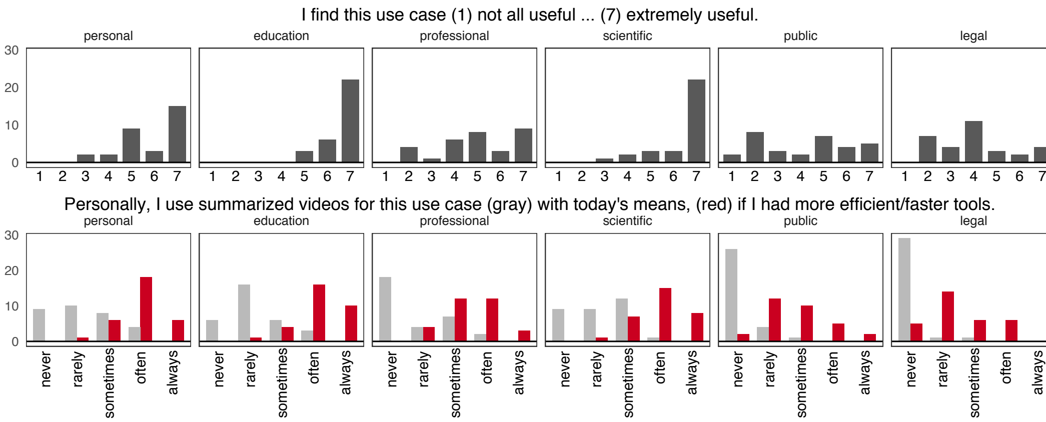

Fig. 2. Survey responses regarding the usefulness of the identified video summary uses (top row), and their current (gray) and future (red) frequency of use (bottom row).

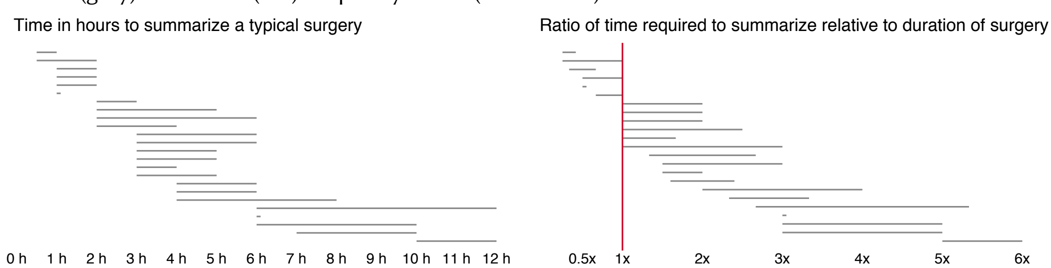

Fig. 3. Survey responses regarding summarization time. Responses on the amount of time spent for video summarization, for respondents who answered they summarize videos at least "rarely" (N=23/31). Each line represents the response of one participant as interval (between x and y hours). Left: time spent to summarize the video of a typical surgery in hours. Right: time spent to summarize such video, as multiple of the duration of the surgery. The red line indicates the same time spent performing a surgery and summarizing it.

6.7 Video Summaries Usefulness, Frequency of Summarization and of Use

To further understand the identified uses of surgical video summaries, we conducted a survey inquiring about (1) their usefulness, and (2) the frequency of use today and in a future scenario where tools would be available to reduce the time needed for summarizing videos. As Figure 2 illustrates, respondents most strongly agree on the extremely high usefulness for *education* and *scientific* purposes. *Personal* was rated by a third of respondents as somewhat useful and by almost half as extremely useful. Opinions vary more on the use cases *professional*, *public*, and *legal* where respondents selected (almost) all possible response options. Interestingly, we find a clear trend for interest in increasing the use of summarized videos if respondents had more efficient and faster tools at their disposal: the second row shows a clear offset between the responses shown in gray (today's tools) and those shown in red (assuming more efficient tools were available). Figure 3 shows data on the time spent for summarization, for respondents who answered they summarize videos at least "rarely". We see how the time required to summarize video, on the left in hours, on the right as a multiple of the duration of a typical surgery, can quickly become a great burden. Note that values that seem "low" in this ratio, such as 2, can quickly translate into non-negligible amount of work: for surgery lasting between 2 and 6 hours, summarization can take from 4 to 16 hours.

7 THE SUMMARIZATION OF SURGICAL VIDEOS

In this section, we detail the complex multi-step process surgeons carry out for summarizing surgical videos, where technology taints the way in which they work in a way that is not aligned with their goals and does not consider the temporal aspect of the process.

7.1 Summarization Process

At a glance, we find that summarization involves watching the video several times, and iteratively removing frames until obtaining a summary of the desired length. The summarization practices we observed were personal, where participants developed methods of their own, most likely as video summarization is essentially self-taught without formal training during medical school. Nonetheless, we observed that there were common strategies across participants we present as a series of steps. The steps in the summarization process involve: first, overviewing the whole video, then, cleaning out irrelevant segments, creating sequences and, finally, contracting them (see Figure 4 for an illustration). We now expand on each step using examples to illustrate them.

(1) *Overviewing* consists of seeking through the whole video by hovering over the timeline with the mouse. The purpose is to get a sense of the entire surgery (P2), and thus not always performed, essentially when the surgery was performed a long time ago or by someone else.
(2) *Cleaning* consists of removing irrelevant segments during which nothing useful is visible in the video feed. Examples include when the image is blurry, the camera is dirty from debris (P2), the endoscope is outside the patient—easily identified by the distinctive blue color of the drapes—or, when the camera moves a great deal (P7).
(3) *Coarse Sequencing* consists of identifying the broad strokes of a surgery. These are long sequences, which tend to correspond with the steps of the surgery. Surgeons usually split the video into parts in the editing timeline, although we observed one surgeon additionally placing empty title cards to reify this separation. As P1 points out, this step is easier when the person summarizing the video also performed the surgery and less so otherwise.
(4) *Fine Sequencing* consists of trimming long sequences which often contain multiple repetitions of the same gesture. The goal of this step is to select the one execution that best exemplifies it. During this step, we consistently observed surgeons doing the same sequence of actions: they create a candidate sequence (short, about 5 to 30 seconds) by splitting the video to mark the beginning and then seeking forward with the mouse, splitting again to mark the end. They iteratively identify these fine sequences until they are able to pinpoint which one best exemplifies a surgical gesture, at which point they eliminate all the other candidates. Then, they continue until the coarse sequence is finished, moving on to the next one.
(5) *Contracting* consists of removing *"lost time"* (P2) inside a sequence. The goal is to remove all uninformative frames and convey only that a surgical gesture took place. For example, when coagulating tissue, the surgeon grabs it, starts the coagulation (which takes about five to ten seconds) and then lets go of the tissue. In the summarized video, the surgeon will keep the beginning of the sequence (tool grabs tissue and starts coagulation) and its end (tool lets go of the coagulated tissue), removing the time of coagulation where the tissue progressively changes color. P1 explained that in this way, a 5-minute sequence of a suture can be reduced to 20 seconds.

It is important to note that steps are not silos of action, but they can actually overlap. We observed times when, during steps that require navigating through the whole video, surgeons opportunistically performed parts of other steps as they found, by chance, information relevant for those steps. For example, while *Cleaning* or *Overviewing*, a surgeon might find a gesture that she wants to highlight, so she already leaves a mark for later.

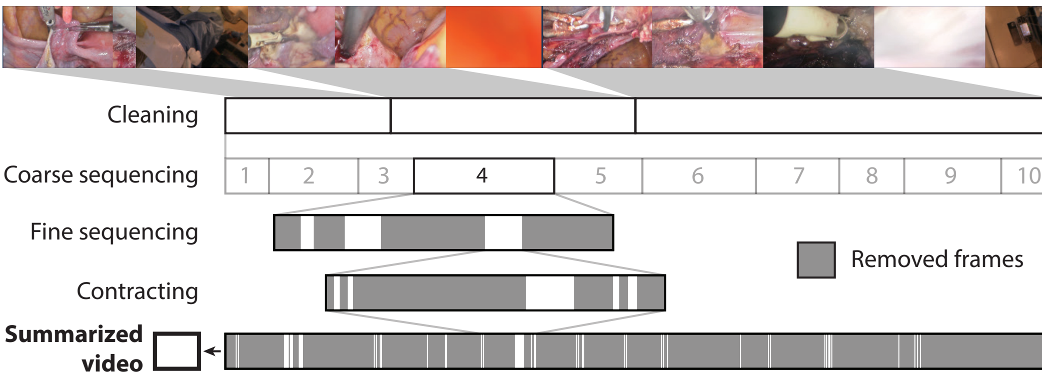

Fig. 4. Schematic of the summarization steps. The top line represents the raw video footage, then each line shows the removed (gray) and retained (white) frames during each step. The last line gives an indication of how frames that are included in the final summarized video are distributed throughout the raw footage.

7.2 Digital Tools Impose a Working Paradigm

We were surprised that all participants adopted a strategy consisting of *removing* sequences and frames with the goal of *keeping* only what they consider most relevant, as shown in Figure 5. We believe this is a consequence of the technologies used. All participants summarized videos using off-the-shelf video-editing tools such as *iMovie*[2] or *OpenShot*[3]. The underlying paradigm of these editors is to delete unwanted sequences by trimming them, which results in costly consequences when used to summarize surgical videos. First, the segments present in the editor track are the current version of the video: once a sequence has been deleted, it is gone and cannot be brought back easily. Second, even if the best sequence in step *Fine Sequencing* has been identified, the remaining repetitions of the gesture have to be first split and only then can be deleted. Third, numerous sequences that will not make the cut must be reviewed, which requires a lot of time.

Fig. 5. Video summarization where several short sequences compose a final video of 4m 05s

Contrary to their strategy of *removing* sequences, the participants' intention was to select sequences to *retain*. The language they use to describe the process reflects this, for instance as P3 mentioned in an interview while showing a summarized video: *"What is difficult is to* ***select the good moments****, which can show you the interesting strategy"*. Having to remove almost all of the raw footage requires revising and selecting a great deal of uninteresting material, a time-heavy process. For a video of, say, 5 hours, surgeons have to delete 4 hours 55 minutes of footage, when all they want is to indicate the 5 minutes that remain. We hypothesize that it should be much faster to instead be able to *select what to keep* without the need to delete the rest.

[2]iMovie is a video editor shipped with MacOS, see www.apple.com/imovie.
[3]OpenShot is an open-source, cross-platform video editor, see https://www.openshot.org/.

7.3 Knowledge Is Lost From the Live Activity to the Summarization

We find that the knowledge about which parts to include in a video summary takes form during live surgery, and it is not transferred to the summarization process; regardless of whether this happens a few hours or years after the surgery, or if the person summarizing is the one who performed the surgery. As a result, surgeons need to watch the entire video and recreate this information to identify the important moments. This is especially important when the surgery was not performed by the person summarizing the video, even if they know the standard surgical steps (P3).

7.3.1 Key Moments: Frequency and Types. During the second round of surgery observations, we encouraged participants to indicate us any moments which they would like to be able to identify later when summarizing the video recording. We collected 35 *Key Moments* during these surgeries.

Frequency. The *Key Moments* surgeons indicated suggest that their frequency varies across procedures, in our study between 1 and 11 (Table 1). Surprisingly, we find a contrast with the interviews where participants mentioned a greater amount and wider variety of important moments to record. We believe this is due to three reasons. First, the observed surgeries were routine interventions without unexpected moments, which in the current workflow would rarely, if ever, be summarized. While we had the opportunity to observe one "special" surgery which would be a good candidate for a later conference submission, it happened during the first round of observations. Second, surgeons directed their attention to the surgery and many times forgot to indicate *Key Moments*. Third, during debriefing, two participants mentioned that, while they are used to talking to residents during surgery for teaching, they are not used to the additional task of "tagging".

Participant	Tags Recorded	
P4	2	||
P4	11	|||||||||||
P5	6	||||||
P6	11	|||||||||||
P8	1	|
P9	4	||||

Table 1. Number of tags reported during the respective surgeries.

Nonetheless, participants manifested in hindsight an interest in having recorded moments for summarizing video, during summarization observations. For example, while observing P7 summarize a surgery by P6, P7 mentioned that it would be good if the surgeon could leave traces of moments to keep, such as anatomical anomalies or a good view of a nerve, to make sure not to remove them. However, participants remained prudent about recording moments by sheer force of unfamiliarity. P1, P2 and P3 mentioned that it is hard to envision leaving traces, as it is something they have never experienced before, and P1 acknowledged that, although he frequently notices moments that he knows must be part of a later video summary during complex surgeries, he does not know how to capture them.

Types. Participants indicated consistent types of *Key Moments* related to the video summary uses we report in section 6, as well as to the steps of the summarization process we report in this section. We distinguish between four types of *Key Moments*.

(1) *Surgical Steps.* Surgical interventions follow protocols, with distinct steps that define a procedure. For example, P4 reported during a hysterectomy: salpingectomy (01:25), opening of the right ligament (15:20), vesico-uterine ligament dissection (22:23), end of dissection (31:23), coagulation right uterine artery (45:55), start coagulation left uterine artery start (51:17), colpotomy start (1:01:10), and, colpotomy start for real (1:11:00). Surgical steps are important for education (6.2), as they embody the protocol of a particular type of procedure, as well as for conference submission (6.4), given the popular *"the 10 steps to"* format for conference submissions, where surgeries are showcased in 10 steps. These steps correspond to *Coarse Sequencing* during video summarization.

(2) *Patient Anatomy.* A patient's anatomy is the starting point of surgery, it shows for example the extent of a disease, the state of tissue subsequent to a previous surgery, unusual characteristics such as size or color and even parts missing or out of place (e.g., arteries or nerves). Surgeons consistently reported moments concerning the patient's anatomy at the beginning and end of the surgery, at the end of certain gestures (e.g., sutures), and anytime they ran into an anatomy out of the ordinary. For example, when P6 found a particular anatomy 11 minutes into the recording, he took a photo of it and left the robotic console to fetch a colleague operating in a nearby room. Unique anatomies are interesting for conference submission and teaching (6.4, 6.2), as they exemplify how protocols are adapted. For legal purposes (6.6), showing the anatomy constitutes hard evidence of the state before surgery, of the outcome, and it explains why certain decisions were made in the case of rare anatomies. This type of moment does not relate to specific summarization steps, but it constitutes important segments always present in summaries.

(3) *Particular Surgical Gestures.* Surgeons notified us to note a moment while performing certain gestures due to their importance, complexity, or because the surgeon performed it very well. P5, 16 minutes into a hysterectomy, reported a moment and said *"pull to expose well"*, referring to pulling on tissue to make way for placing the tools behind it and working in that area. These moments are important mostly for teaching or self-improvement (6.2, 6.1), as they depict gestures performed in a particularly good way (e.g., while being well exposed, fast, without any complications) or bad way (e.g., applying too much pressure to a delicate structure, dropping a needle and possibly even losing it). Particular gestures do not relate to specific summarization steps, and they can account for the highlights of a summary.

(4) *Unexpected or Critical Moments.* Surgeries have unexpected events, even if they happen rarely [66]. We did not encounter any during the routine surgeries we observed, but surgeons emphasized their importance during interviews. P3 mentioned a time where massive bleeding occurred while performing robotic surgery, which was quickly controlled as the team converted to open surgery. The surgeon's hypothesis was that the robotic arm holding the endoscope tore a main artery as he moved it up, and this happened off-screen. The lack of haptic feedback from the robotic platform did not let the surgeon feel that the artery was in the way, and thus to stop pulling before causing damage. As the video was not recording, P3 could not go back to trace at what moment the endoscope shaft was placed in the wrong position. These moments are important for teaching and self-improvement (6.2, 6.1), as they enable to analyze critical, rare, moments and the steps that led to that point. They can also be relevant for conference submissions (6.4), which sometimes focus solely on errors, as P3 recalled a particular video session called *"what is your worst surgical nightmare?"*. Critical moments can also be used for discussing a case among colleagues (6.3) and have legal value in case of post-operative complications. Their scarcity is what makes them highly valuable, today it is rare that surgeons are recording when they happen. Unexpected moments are not related to specific summarization steps, but they can certainly be the main reason for a video.

Consequently, we find that the *Key Moments* reflect two types of intertwined knowledge, especially relevant for summarizing surgical videos: *situated* knowledge, to recall and understand what was going on in the operating room, and *domain* knowledge, to determine how to use situated knowledge as well as its relevance for this particular procedure and later use of the video summary. Situated knowledge includes both events related to the procedure as well as external events in the operating room. We observed several times during summarization sessions how surgeons paused to reflect on the video, trying to recall what was happening. P2, for example, stumbled upon footage of the tools not moving during a long sequence, recalling after long reflection that the robotic arms were not properly placed so he had to move them, which happens outside of the body and thus is not visible in the laparoscopic video. Slow movements of the tools also triggered surgeons to think what had happened during surgery, such as P2 who remembered that he was explaining the resident how to perform the motion by doing it slowly, or P4, recalling that she had passed the tools to the resident. While such sequences are unlikely to be included in videos for conference submissions, they may be relevant for teaching purposes, for example, to help residents analyze their technique outside of surgery and help them improve their technique.

8 THE COLLABORATIVE NATURE OF SUMMARIZATION: AD-HOC COORDINATION

Surgeons in our study performed video summarization in collaboration with colleagues mainly for efficiency reasons: the work is divided as it is very time-consuming. During the summary observations and interviews, we noticed that, although they collaborate in different ways, surgeons follow a similar principle: first, one does an initial summary, oftentimes a less-experienced surgeon or a resident, taking the raw footage and reducing it to about 30 minutes; then, another one finishes the summarization. For example, P6 and P7 have developed a workflow where P6 performs the surgery and delegates to P7, a resident, the first pass of the summary which consists of *Cleaning*, *Coarse Sequencing* and some parts of *Fine Sequencing*, until reaching a 30-minute version. P7 leaves several instances of the same gesture during *Fine Sequencing*, and P6 selects the best ones, finishing by *Contracting*. P3 mentioned another example during an interview where he had performed a 4-hour surgery and then did a first 1-hour summary by performing *Cleaning* and *Coarse Sequencing*, and later P1 finished the summary to obtain a 6-minute video for a conference submission. To collaboratively summarize videos, participants did not rely on functionalities in video-editing tools, simply because these tools do not provide any, they are not designed to support collaboration. They developed instead ad-hoc strategies to reify information used as articulation work when handing off a partially-done summarization to another surgeon. We observe three strategies in particular.

8.1 Strategy 1: Segmenting the Live Recording

The first strategy involves cutting the video footage itself while recording it to signal relevance. In the case of P6 and P7, P6 has the habit of starting and stopping the video recording with the purpose of both (1) already removing useless sequences (essentially "live *Cleaning*") and (2) signaling sequences that are interesting so that P7 keeps them during *Coarse Sequencing* in the first iteration. In one particular surgery, P6 regularly asked a nurse to push the stop/start recording button, and also to verify that the recording was active in multiple occasions. While this was a robotic surgery where the surgeon is not sterile and could in principle push the recording button, the button was far from the robotic console, thus P6 delegated the task to the nurse. The nurse later mentioned that it would be better if the surgeon could be autonomous, as this interrupts her work. P9 and P8 adopt a similar strategy with the difference that their endoscope has buttons to stop/start recording, accessible during surgery, as they are placed under the sterile plastic sleeve.

This strategy, which has been observed in previous work [13], is certainly not as effective as desired. P7 reported during a summarization session that he usually cannot rely on the segments that P6 produces live as it is hard to interpret their meaning. The end and start of a file do not convey semantic information and can exist for different reasons, for instance because a relevant event is about to start or has just finished, but also because adjustments on the surgical robot were necessary and thus the surgeon stopped recording to reduce efforts when *Cleaning*.

8.2 Strategy 2: Capturing Moments on External Media

A second strategy is to use external media to record moments during surgery. P2 explained how he uses his phone to take pictures of the screen showing the endoscopic video feed. P6 uses the video-recording interface in his operating room as it can take snapshots while recording video. These images are then passed on to the person doing the summary as memorandums of important moments, which requires additional work. P7 explained that many times P6 forgets to include them. What is more, the person doing the summary has to spend much time finding the isolated frame within the long video, to then assess why it is important. Automatic approaches to find snapshot images within their original surgical video can lighten the search load [13], but still deciding why this instant is relevant remains unaddressed.

8.3 Strategy 3: Engraving Marks on a Partial Video Summary

Fig. 6. P1 runs into an empty title card and says: *"I don't know if I will keep it in the end, but [P3] put this title here to.."* as he makes a chopping motion with his hand. *"Separate?"* *"yes, but I don't know what he wanted to..."*. P1 then leaves this section for later and moves to summarize the next sequence.

The third strategy takes place after surgery, the intention is to facilitate the hand-off from the surgeon starting the summary to the one finishing it. It involves leaving footprints engraved in the video summary file before passing it along. For example, P3 typically leaves empty title cards made with *iMovie* to point out sequences, as we observed in the summarization session by P1. As with the first strategy, the limitation is that it is hard to make sense of these cuts without additional information. During a summarization session (Figure 6), P1 ran into an empty card and thought that P3 tried to signal the start of a new sequence, but he hesitated, reflecting on whether P3 meant something else. Surprisingly, although they could add text to express meaning, they usually did not.

In general, all three ad-hoc coordination strategies served their purpose poorly since they seemed to neither increase efficiency in work (saving time) nor efficacy (finding the important moments to keep). What is more, process knowledge was lost when participants sent video for collaborative summarization. As they exported video, they rendered the project "flattening" their work into a single file, which means their colleagues can no longer infer the status of the process, signaled by the amount of processed segments, their length, and the ones still unedited (see Figure 5).

9 DISCUSSION AND IMPLICATIONS

Our study exposes the intricacies when video summarization is performed collaboratively. While there are rare examples in previous work of video summarization through the contribution of multiple human coders (e.g., [17]), they have not explored notions of coordination work from Malone & Crowstone's framework [41]. In our study, we identified two of these notions: summarization with breakdowns in the *division of tasks and assignment* (who does *Cleaning* and who *Coarse Sequencing*?) and the *management of dependencies* (is *Coarse Sequencing* finished so can I start *Fine Sequencing*?). As video editing systems are not designed to support these notions, there are inefficiencies from breakdowns in *articulation work* [58], as mechanisms of coordination are not mutually understood nor ubiquitously followed, thus they do not enable coordination of interdependent work [56]. We posit there are two particular characteristics of coordination work in this setting, that can lead to more efficient collaboration, if considered.

First, coordination happens across **two stages disjoint in their temporality and aim**: a data-generation stage during a domain-specific activity, and later a data-processing stage during a non domain-specific activity (summarization). We are not the first to identify temporality in coordination at CSCW. Bardram et al. [6] define *temporal coordination*, referring to the activities that ensure distributed actions happen in a timely manner, something that is not much of a concern in our work as surgical video can actually spend months gathering digital dust before it is summarized. Reddy et al. [55] propose a framework through which we can formalize the life-cycle of surgical video as its *temporal trajectory*: video is recorded during surgery, then summarized, and finally used. In our work, we are concerned with the transferring of information that supports articulation work from one stage to the next as a component of this trajectory.

Second, our study shows a particular case for the conceptualization of coordinating artifacts in CSCW. Coordinating artifacts, as conceptualized by Simone, Schmidt and colleagues are *symbolic artifacts*, where *"the state of the artifact is decoupled from the state of the field of work"* [60], a notion reinforced in a later proposition: *"The artifact of a coordination mechanism is distinct from the field of work in the sense that changes to the state of the field of work are not automatically reflected in the state of the artifact [..]"* [58]. In video summarization, **the coordination artifact *is* the common field object**. We believe this inseparability is a consequence of the lack of tools to deal with it, resulting in a protocol with *implicit* conventions and *unprescribed* procedures, contrary to the classic notion: *"[..] a protocol, encompassing a set of explicit conventions and prescribed procedures[..]"* [60], as they cannot be materialized. Still, they exist as a collective construct, participants followed summarization steps (section 7) and followed coordination work conventions (section 8) in similar ways. Taking a step back, we note that video summarization is not taught in surgical school, and thus it is understandable that the procedure is blurry, contrary to strictly protocolar surgery. The lack of a clear protocol results in inefficient collaboration as, first, it cannot be used in articulation work [56], meaning surgeons cannot refer to it when coordinating their actions, and, second, it cannot be used as a "map" [64] that guides situated action, thus it cannot be referenced for orientation.

9.1 Implication: Making Video a Communicative Artifact

The ability to associate information to video, the field of work, can make video a communicative artifact, making coordination more efficient. This information should exist at different layers, as not only it should support various types of *Key Moments*, but also as it can be associated to different types of uses for the video once summarized. Users can crystallize different types of knowledge towards articulation work in the video, necessarily time-stamped according to the recording, at the two stages: during recording (live action) for situated and domain knowledge, and during summarization for process knowledge.

Capturing Situated and Domain Knowledge During Live Action. The person performing the live action will collaborate in the future with another person, or with themselves, during summarization, creating a need to transfer knowledge from the live stage to the editing stage. The main challenge is the cognitive and physical load that performing a complex task demands of an expert. As we observed, surgeons were concentrated on surgery and frequently forgot to stop and start the recording, and when taking a picture of the video screen they had to break sterility by removing their gloves and pause surgery. Based on this, the following should be taken into consideration when capturing *Key Moments* during live action.

- *Temporal Offset.* The moment when knowledge is captured is not necessarily the same as when an event happens. This temporality can be captured as *"terminus post quem"* and *"terminus ante quem"*, the known limits to the actual moment the event happened.

- *Outsourced Interaction.* External actors can act as surrogates and capture knowledge the expert generates when performing a task. These can be people or intelligent systems, such as digital assistants.
- *Multiple Interaction Possibilities.* Interaction multimodality for capturing moments increases the chance they will be captured, for instance including the use of voice, hands, feet or their combination. These choices will depend on what the constraints of the task at hand.
- *Flexibility.* As capturing knowledge is a secondary task, it will happen that some of the key moments are omitted, and thus this should not be penalized in a system. Absence of evidence is not evidence of absence: missing a moment does not mean it did not matter.

We highlight for the case of surgery the need to capture situated knowledge, which can include an unusual anatomy of a patient or unexpected bleeding, as well as domain knowledge, such as the expected anatomy or surgical steps in the protocol. What is more, it is important to include also situated knowledge that originates in the operating room, not only inside the body, as events external to surgical gestures can influence factors such as communication and learning that impact the result of surgery, especially in robotic surgery [5]. This can be achieved by adding the feed of an external camera as picture-in-picture in the endoscopic video feed. Regarding the interaction technique to capture information during surgery, HCI research has explored voice and gesture-based systems [45, 52] that are compatible with the sterility requirements. Touch can also be used for instance by using a sterile cover around buttons on the endoscope, although oftentimes surgeons are busy holding the surgical tools and cannot simply switch to the endoscope to press buttons [4].

Capturing Process Knowledge During Summarization. One of Tang's findings while studying collaborative work [65] was that the process of one person working on an object conveys a lot of information that should be available to collaborators. Our observations of how surgeons use ad-hoc strategies to leave traces is an indication that they would benefit from seeing each other's process. In the case of surgical video, this would indicate how far a segment has been processed, showing for instance if this iteration is closer to *Fine Sequencing* than *Coarse Sequencing*.

Design Implication. In the field of surgery, endoscopic video annotation has been proposed for example by taking snapshots [36] or adding information using the voice [27]. The technical implementation of these approaches relies on using a separate log file to keep annotations, similarly to work from HCI such as *ChronoViz* [24] for recorded footage, or *EVA* [39] and later *DIVA* [40] for live video while it is being recorded. This lets adding semantic information and time stamps, but the video and the annotations remain separate. *GoPro* [26] cameras recently incorporated functionality to create marks during live recording that actually become part of the final mp4 file, but without associating semantics. To achieve both associating annotations to the video and including semantic information, we take advantage of the inseparability of both the coordination artifact and field of work in our study. The strategy we observed of leaving marks on a partial video summary or segmenting the live recording is a mechanisms for dealing with this inseparability, by imprinting, on the common object itself, instructions, referencing steps of the process. Although this "stains" the final object, it eliminates the disadvantage of desynchronization between the objects, as changes to the field of work (e.g., adding an empty title card) are now automatically reflected in the state of the coordination artifact (e.g., a boundary between sequences has been set).

Inspired by this, we envision mechanisms that take advantage of the visual aspect of video data. For example imprinting, literally, codified information on the footage itself in the form of colored shapes in a predefined area of pixels, that can be interpreted during summarization to reconstruct the knowledge by a system or a person. What is more, this process can be completely moved from a technological implementation to a social protocol, where actors agree in advance a protocol that involves creating salient segments on the video that carry meaning. In the case of surgery, surgeons could associate information *using their surgical tools*, by placing them in a particular location (e.g., the corner of the screen) or doing a particular action that does not interrupt the main task (e.g., two consecutive circles). This involves changing the domain practice, and can have consequences that must be considered. We believe that recent toolkits for video in HCI such as *Videostrates* [33] can provide the necessary infrastructure to prototype these strategies with a low cost.

9.2 Broader Benefits

We expect that formalizing knowledge will bring broader benefits, beyond coordination.

Templates for the Domain Task. The key moments captured for summarization can also serve as a template for the domain task, to be used as a checklist. Just as Bardram & Bossen [7] find that artifacts used to achieve coordination in hospitals, such as whiteboards or patient records, form *templates*, videos can also serve a similar purpose. During surgery, this checklist can formalize a tacit knowledge on the standard surgical procedure steps, including possible adaptations of individual surgeons, useful for their reuse during subsequent surgeries. More interestingly, they can be used for formalizing pending activities during the live activity. An example for the case of surgery, is when counting objects (e.g., needles) inserted into the patient, to make sure they are removed and none stay inside after surgery. Although miscounting rarely happens [66], when it does, it can have a significant impact, including stopping the surgery momentarily so that the surgeon can look for the object inside the patient while nurses outside, including in the trash bins. These checklist can close the loop on inserted objects making sure they have been removed.

Support the Learning Process. Having a soon-to-be-expert in the learning phase use the knowledge for summarization can support learning in the process. Similar to Nardi & Miller's [50] study of how collaborative use of spreadsheet and their observation of how users "exchange *domain* knowledge" in the process of spreadsheet development, we envision a similar impact when working with video.

More Material is Generated. Regarding our use case in particular, we believe that increasing the efficiency of surgical video summaries can be expected to increase the overall amount of video summaries generated. The accumulated knowledge and the semantics gathered during surgery, combined with how surgeons then use these to summarize a video, carries immensely rich knowledge that can be used in the future, along with the summary result, to move forward the development of automatic summarization methods.

10 LIMITATIONS AND OUTLOOK

The study results are bound to the population we observed, which contextualizes their practice to the cultural, political and social aspects of where they live, as well as their experience. The specific procedures we observed can also limit our findings, as certain events characteristics of other surgeries might expand our understanding of the summarization process. Regarding the collaborative aspects, pairs normally are bound to hierarchical constraints, or the personal relations they developed. This can result in mutual knowledge or implicit rules that escaped our understanding. In future work, we think it is necessary to study potential new uses of live video and we believe that using technology probes can be a promising strategy, as they allow to introduce and refine a new techniques while studying them through several iterations. Lastly, the exploration of systems for tagging live video adapted to the context of surgery.

ACKNOWLEDGMENTS

This research was supported by the Labex CAMI (project ANR-11-LABX-0004) operated by the ANR as part of the program "Investissement d'Avenir". We thank all the participants of this study for their valuable time, as well as Gilles Bailly, Elodie Bouzbib and Oleksandra Vereschak for their feedback on early versions of this article.

REFERENCES

[1] Joanna Abraham and Madhu C. Reddy. 2013. Re-coordinating activities: an investigation of articulation work in patient transfers. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. Association for Computing Machinery, San Antonio, Texas, USA, 67–78. https://doi.org/10.1145/2441776.2441787

[2] Kiyoharu Aizawa, Datchakorn Tancharoen, Shinya Kawasaki, and Toshihiko Yamasaki. 2004. Efficient retrieval of life log based on context and content. In *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*. ACM, New York, NY, USA, 22–31.

[3] Selen Atasoy, Diana Mateus, Joe Lallemand, Alexander Meining, Guang-Zhong Yang, and Nassir Navab. 2010. Endoscopic Video Manifolds. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010 (Lecture Notes in Computer Science)*, Tianzi Jiang, Nassir Navab, Josien P. W. Pluim, and Max A. Viergever (Eds.). Springer, Berlin, Heidelberg, 437–445. https://doi.org/10.1007/978-3-642-15745-5_54

[4] Ignacio Avellino, Gilles Bailly, Mario Arico, Guillaume Morel, and Geoffroy Canlorbe. 2020. Multimodal and Mixed Control of Robotic Endoscopes. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, Honolulu, HI, USA, 1–14. https://doi.org/10.1145/3313831.3376795

[5] Ignacio Avellino, Gilles Bailly, Geoffroy Canlorbe, Jérémie Belgihti, Guillaume Morel, and Marie-Aude Vitrani. 2019. Impacts of Telemanipulation in Robotic Assisted Surgery. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 583:1–583:15. https://doi.org/10.1145/3290605.3300813

[6] Jakob E. Bardram. 2000. Temporal Coordination –On Time and Coordination of CollaborativeActivities at a Surgical Department. *Computer Supported Cooperative Work (CSCW)* 9, 2 (May 2000), 157–187. https://doi.org/10.1023/A:1008748724225

[7] Jakob E. Bardram and Claus Bossen. 2005. A web of coordinative artifacts: collaborative work at a hospital ward. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work (GROUP '05)*. Association for Computing Machinery, Sanibel Island, Florida, USA, 168–176. https://doi.org/10.1145/1099203.1099235

[8] Jakob E. Bardram and Thomas Riisgaard Hansen. 2010. Why the plan doesn't hold: a study of situated planning, articulation and coordination work in a surgical ward. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*. Association for Computing Machinery, Savannah, Georgia, USA, 331–340. https://doi.org/10.1145/1718918.1718977

[9] Marc Berg. 1999. Accumulating and Coordinating: Occasions for Information Technologies in Medical Work. *Computer Supported Cooperative Work (CSCW)* 8, 4 (Dec. 1999), 373–401. https://doi.org/10.1023/A:1008757115404

[10] Lonni Besançon, Amir Semmo, David Biau, Bruno Frachet, Virginie Pineau, El Hadi Sariali, Rabah Taouachi, Tobias Isenberg, and Pierre Dragicevic. 2018. Reducing Affective Responses to Surgical Images Through Color Manipulation and Stylization. In *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch-Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering (Expressive '18)*. ACM, New York, NY, USA, 11:1–11:13. https://doi.org/10.1145/3229147.3229158 event-place: Victoria, British Columbia, Canada.

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101.

[12] Eyuphan Bulut and Tolga Capin. 2007. Key Frame Extraction from Motion Capture Data by Curve Saliency. *Computer Animation and Social Agents* 20 (2007), 5.

[13] Jennifer Roldan Carlos, Mathias Lux, Xavier Giro-i Nieto, Pia Munoz, and Nektarios Anagnostopoulos. 2015. Visual information retrieval in endoscopic video archives. In *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*. IEEE, New York, NY, USA, 1–6. https://doi.org/10.1109/CBMI.2015.7153618

[14] Peter H. Carstensen and Morten Nielsen. 2001. Characterizing modes of coordination: a comparison between oral and artifact based coordination. In *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '01)*. Association for Computing Machinery, Boulder, Colorado, USA, 81–90. https://doi.org/10.1145/500286.500301

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

[16] U. Damnjanovic, V. Fernandez, E. Izquierdo, and J. M. Martinez. 2008. Event Detection and Clustering for Surveillance Video Summarization. In *2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*. IEEE, New York, NY, USA, 63–66. https://doi.org/10.1109/WIAMIS.2008.53

[17] Kaveh Darabi and Gheorghita Ghinea. 2014. Video summarization by group scoring. In *2014 International Conference on Multimedia Computing and Systems (ICMCS)*. IEEE, New York, NY, USA, 112–116. https://doi.org/10.1109/ICMCS.2014.6911240

[18] Gamhewage C De Silva, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2005. Evaluation of video summarization for a large number of cameras in ubiquitous home. In *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, New York, NY, USA, 820–828.

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

[20] Ajay Divakaran, Kadir A. Peker, and Huifang Sun. 2001. Video summarization using motion descriptors. In *Storage and Retrieval for Media Databases 2001*, Vol. 4315. International Society for Optics and Photonics, 517–522. https://doi.org/10.1117/12.410963

[21] Brian Duty, Zhamshid Okhunov, Justin Friedlander, Zeph Okeke, and Arthur Smith. 2012. Live Surgical Demonstrations: An Old, but Increasingly Controversial Practice. *Urology* 79, 5 (May 2012), 1185.e7–1185.e11. https://doi.org/10.1016/j.urology.2011.12.037

[22] A. Ekin, A.M. Tekalp, and R. Mehrotra. 2003. Automatic soccer video analysis and summarization. *IEEE Trans. on Image Process.* 12, 7 (July 2003), 796–807. https://doi.org/10.1109/TIP.2003.812758

[23] Zhiwei Fan, Congliang Chen, and Tingting Jiang. 2017. A semi-automatic editing method for surgery videos. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, New York, NY, USA, 115–120.

[24] Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D. Hollan. 2011. ChronoViz: a system for supporting navigation of time-coded data. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*. ACM Press, Vancouver, BC, Canada, 299. https://doi.org/10.1145/1979742.1979706

[25] K. Fujimura, K. Honda, and K. Uehara. 2002. Automatic video summarization by using color and utterance information. In *Proceedings. IEEE International Conference on Multimedia and Expo*, Vol. 1. IEEE, New York, NY, USA, 49–52 vol.1. https://doi.org/10.1109/ICME.2002.1035715

[26] GoPro. 2019. What is HiLight Tagging & How Does it Work? https://gopro.com/help/articles/Question_Answer/What-is-HiLight-Tagging-and-How-Does-it-Work

[27] Mario Guggenberger, Michael Riegler, Mathias Lux, and Pål Halvorsen. 2014. Event understanding in endoscopic surgery videos. In *Proceedings of the 1st ACM International Workshop on Human Centered Event Understanding from Multimedia*. ACM, Association for Computing Machinery, New York, NY, USA, 17–22. https://doi.org/10.1145/2660505.2660509

[28] Gary Guthart and John Kenneth Salisbury. 2000. The Intuitive/sup TM/ telesurgery system: overview and application. *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)* 1 (2000), 618–621 vol.1.

[29] Giselle G. Hamad, Matthew T. Brown, and Julio A. Clavijo-Alvarez. 2007. Postoperative video debriefing reduces technical errors in laparoscopic surgery. *The American Journal of Surgery* 194, 1 (July 2007), 110–114. https://doi.org/10.1016/j.amjsurg.2006.10.027

[30] D. A. Hill and R. S. A. Lord. 1991. Complementary value of traditional bedside teaching and structured clinical teaching in introductory surgical studies. *Medical Education* 25, 6 (1991), 471–474. https://doi.org/10.1111/j.1365-2923.1991.tb00099.x

[31] Naja Holten Møller and Paul Dourish. 2010. Coordination by avoidance: bringing things together and keeping them apart across hospital departments. In *Proceedings of the 16th ACM international conference on Supporting group work (GROUP '10)*. Association for Computing Machinery, Sanibel Island, Florida, USA, 65–74. https://doi.org/10.1145/1880071.1880081

[32] Peter F. Hu, Yan Xiao, Danny Ho, Colin F. Mackenzie, Hao Hu, Roger Voigt, and Douglas Martz. 2006. Advanced Visualization Platform for Surgical Operating Room Coordination: Distributed Video Board System. *Surgical Innovation* 13, 2 (June 2006), 129–135. https://doi.org/10.1177/1553350606291484 Publisher: SAGE Publications Inc.

[33] Clemens N. Klokmose, Christian Remy, Janus Bager Kristensen, Rolf Bagge, Michel Beaudouin-Lafon, and Wendy Mackay. 2019. Videostrates: Collaborative, Distributed and Programmable Video Manipulation. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 233–247. https://doi.org/10.1145/3332165.3347912

[34] Jarrod Knibbe, Sue Ann Seah, and Mike Fraser. 2015. Videohandles: Searching through action camera videos by replicating hand gestures. *Computers & Graphics* 48 (2015), 99–106.

[35] D. I. Kosmopoulos, A. Doulamis, and N. Doulamis. 2005. Gesture-based video summarization. In *IEEE International Conference on Image Processing 2005*, Vol. 3. IEEE, New York, NY, USA, III–1220. https://doi.org/10.1109/ICIP.2005.1530618

[36] Andreas Leibetseder, Bernd Munzer, Klaus Schoeffmann, and Jorg Keckstein. 2017. Endometriosis Annotation in Endoscopic Videos. In *2017 IEEE International Symposium on Multimedia (ISM)*. IEEE, Taichung, 364–365. https://doi.org/10.1109/ISM.2017.69

[37] Constantinos Loukas. 2018. Video content analysis of surgical procedures. *Surgical endoscopy* 32, 2 (2018), 553–568.

[38] Mathias Lux, Oge Marques, Klaus Schöffmann, Laszlo Böszörmenyi, and Georg Lajtai. 2010. A novel tool for summarization of arthroscopic videos. *Multimedia Tools and Applications* 46, 2-3 (2010), 521–544.

[39] W. E. Mackay. 1989. EVA: An Experimental Video Annotator for Symbolic Analysis of Video Data. *SIGCHI Bull.* 21, 2 (Oct. 1989), 68–71. https://doi.org/10.1145/70609.70617

[40] Wendy E. Mackay and Michel Beaudouin-Lafon. 1998. DIVA: Exploratory Data Analysis with Multimedia Streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 416–423. https://doi.org/10.1145/274644.274701

[41] Thomas W. Malone and Kevin Crowston. 1994. The interdisciplinary study of coordination. *Comput. Surveys* 26, 1 (March 1994), 87–119. https://doi.org/10.1145/174666.174668

[42] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 72:1–72:23. https://doi.org/10.1145/3359174

[43] David A. Mejia, Alberto L. Morán, and Jesus Favela. 2007. Supporting Informal Co-located Collaboration in Hospital Work. In *Groupware: Design, Implementation, and Use (Lecture Notes in Computer Science)*, Jörg M. Haake, Sergio F. Ochoa, and Alejandra Cechich (Eds.). Springer, Berlin, Heidelberg, 255–270. https://doi.org/10.1007/978-3-540-74812-0_20

[44] Helena M. Mentis. 2017. Collocated Use of Imaging Systems in Coordinated Surgical Practice. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 78:1–78:17. https://doi.org/10.1145/3134713

[45] Helena M. Mentis, Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Robert Corish, Antonio Criminisi, Rikin Trivedi, and Pierre Theodore. 2015. Voice or Gesture in the Operating Room. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 773–780. https://doi.org/10.1145/2702613.2702963

[46] Arthur G Money and Harry Agius. 2008. Video summarisation: A conceptual framework and survey of the state of the art. *Journal of Visual Communication and Image Representation* 19, 2 (2008), 121–143.

[47] Katie Moon and Deborah Blackman. 2014. A Guide to Understanding Social Science Research for Natural Scientists. *Conservation Biology* 28, 5 (2014), 1167–1177. https://doi.org/10.1111/cobi.12326

[48] Jacqueline Moss and Yan Xiao. 2004. Improving Operating Room Coordination: Communication Pattern Assessment. *JONA: The Journal of Nursing Administration* 34, 2 (Feb. 2004), 93–100. https://journals.lww.com/jonajournal/Abstract/2004/02000/Improving_Operating_Room_Coordination_.8.aspx

[49] B. R. Nair, J. L. Coughlan, and M. J. Hensley. 1998. Impediments to bed-side teaching. *Medical education* 32, 2 (March 1998), 159–162. https://doi.org/10.1046/j.1365-2923.1998.00185.x

[50] Bonnie A. Nardi and James R. Miller. 1990. An ethnographic study of distributed problem solving in spreadsheet development. In *Proceedings of the 1990 ACM conference on Computer-supported cooperative work (CSCW '90)*. Association for Computing Machinery, New York, NY, USA, 197–208. https://doi.org/10.1145/99332.99355

[51] JungHwan Oh, Sae Hwang, JeongKyu Lee, Wallapak Tavanapong, Johnny Wong, and Piet C de Groen. 2007. Informative frame classification for endoscopy video. *Medical Image Analysis* 11, 2 (2007), 110–127.

[52] Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, and Tom Carrell. 2014. Touchless Interaction in Surgery. *Commun. ACM* 57, 1 (Jan. 2014), 70–77. https://doi.org/10.1145/2541883.2541899

[53] Jaydeep Palep. 2011. Commentary: live telecast surgery on shaky ground. *Indian Journal of Medical Ethics* 8, 4 (Oct. 2011), 245. https://doi.org/10.20529/IJME.2011.092

[54] Joanna Philip-Watson, Shahid A. A. Khan, Marios Hadjipavlou, Abhay Rane, and Thomas Knoll. 2014. Live surgery at conferences – Clinical benefits and ethical dilemmas. *Arab Journal of Urology* 12, 3 (Sept. 2014), 183–186. https://doi.org/10.1016/j.aju.2014.04.002

[55] Madhu C. Reddy, Paul Dourish, and Wanda Pratt. 2006. Temporality in Medical Work: Time also Matters. *Computer Supported Cooperative Work (CSCW)* 15, 1 (Feb. 2006), 29–53. https://doi.org/10.1007/s10606-005-9010-z

[56] Yvonne Rogers. 1992. Coordinating computer-mediated work. *Computer Supported Cooperative Work (CSCW)* 1, 4 (Dec. 1992), 295–315. https://doi.org/10.1007/BF00754332

[57] Kjeld Schmidt and Liam Bannon. 1992. Taking CSCW seriously. *Computer Supported Cooperative Work (CSCW)* 1, 1 (March 1992), 7–40. https://doi.org/10.1007/BF00752449

[58] Kjeld Schmidt and Carla Simonee. 1996. Coordination mechanisms: Towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work (CSCW)* 5, 2 (June 1996), 155–200. https://doi.org/10.1007/BF00133655

[59] Peter G. Scupelli, Yan Xiao, Susan R. Fussell, Sara Kiesler, and Mark D. Gross. 2010. Supporting coordination in surgical suites: physical aspects of common information spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, Atlanta, Georgia, USA, 1777–1786. https://doi.org/10.1145/1753326.1753593

[60] Carla Simone, Monica Divitini, and Kjeld Schmidt. 1995. A notation for malleable and interoperable coordination mechanisms for CSCW systems. In *Proceedings of conference on Organizational computing systems (COCS '95)*. Association for Computing Machinery, Milpitas, California, USA, 44–54. https://doi.org/10.1145/224019.224024

[61] Arthur Smith. 2012. Urological live surgery - an anathema. *BJU international* 110, 3 (Aug. 2012), 299–300. https://doi.org/10.1111/j.1464-410X.2012.11314.x

[62] Sean R Stanek, Wallapak Tavanapong, Johnny Wong, Jung Hwan Oh, and Piet C De Groen. 2012. Automatic real-time detection of endoscopic procedures using temporal features. *Computer methods and programs in biomedicine* 108, 2 (2012), 524–535.

[63] Anselm Strauss. 1985. Work and the division of labor. *The Sociological Quarterly* 26, 1 (March 1985), 1–19. https://doi.org/10.1111/j.1533-8525.1985.tb00212.x Publisher: John Wiley & Sons, Ltd.

[64] Lucy A. Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication.* Cambridge University Press, Cambridge, UK.

[65] John C. Tang. 1991. Findings from Observational Studies of Collaborative Work. *Int. J. Man-Mach. Stud.* 34, 2 (Feb. 1991), 143–160. https://doi.org/10.1016/0020-7373(91)90039-A

[66] Cornelius A. Thiels, Tarun Mohan Lal, Joseph M. Nienow, Kalyan S. Pasupathy, Renaldo C. Blocker, Johnathon M. Aho, Timothy I. Morgenthaler, Robert R. Cima, Susan Hallbeck, and Juliane Bingener. 2015. Surgical never events and contributing human factors. *Surgery* 158, 2 (Aug. 2015), 515–521. https://doi.org/10.1016/j.surg.2015.03.053

[67] Patrick Yeung, Tiffany Justice, and Resad Paya Pasic. 2009. Comparison of Text versus Video for Teaching Laparoscopic Knot Tying in the Novice Surgeon: A Randomized, Controlled Trial. *Journal of Minimally Invasive Gynecology* 16, 4 (July 2009), 411–415. https://doi.org/10.1016/j.jmig.2009.02.011

[68] Denny Yu, Bethany Lowndes, Cornelius Thiels, Juliane Bingener, Amro Abdelrahman, Rebecca Lyons, and Susan Hallbeck. 2016. Quantifying intraoperative workloads across the surgical team roles: room for better balance? *World journal of surgery* 40, 7 (2016), 1565–1574.

Received June 2020; revised October 2020; accepted December 2020