Techniques for Semantic Search and Comparison for Robotic Surgery Videos

JEONGYEON KIM, School of Computing, KAIST, South Korea
DAEUN CHOI, School of Computing, KAIST, South Korea
MATTHEW BEANE, Technology Management, University of California Santa Barbara, USA
JUHO KIM, School of Computing, KAIST, South Korea

Videos are a major learning material for surgical training, presenting intricate details of techniques and tacit knowledge of surgeons. However, it is tedious to consume videos in a linear manner, which often requires scrubbing the video timeline back and forth to find video segments of interest. Our formative studies with six surgeons and thirty residents reveal two primary challenges of video-based surgery learning: (1) limited access to semantic information in video search and (2) lack of support for cross-video reference and comparison. The residents seek to search surgery videos based on phase structures and compare multiple procedures to be prepared for various cases. We introduce an interface prototype to support semantic search and cross-video referencing. Our prototype provides a procedural diagram through which users can understand the landscape of procedures and filter videos based on their structural information. It also enables users to compare multiple videos and synchronize the video playback using graph-based interactions. Finally, we discuss design implications for interfaces to support search and comparison for procedural videos.

CCS Concepts: • **Human-centered computing → Interactive systems and tools**; **Empirical studies in interaction design**.

Additional Key Words and Phrases: Video-Based Surgery Learning; Online Surgical Training; Video Interaction Design

ACM Reference Format:
Jeongyeon Kim, DaEun Choi, Matthew Beane, and Juho Kim. 2018. Techniques for Semantic Search and Comparison for Robotic Surgery Videos. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Procedural knowledge is knowledge about how to do something, for example, how to carry out sequences of operations or actions [23, 36]. Video is an effective medium to learn procedural knowledge, which presents tacit knowledge and a continuous flow of procedures that are difficult to depict through text or static images. Surgery recordings, in particular, have been primary learning materials for residents and medical students in learning surgical procedures [27, 39], though such materials are still not formally integrated into surgical curricula [7]. Learners have turned to video because it shows techniques that are difficult to explain verbally [3, 10]. For example, applying traction (i.e., pulling on tissue) before making incisions into fascia leads to better outcomes, but the details of this technique are rarely explicitly explained or documented. Videos demonstrate such subtle nuances in surgical technique, which allows residents to build the capability required to perform in real surgical situations, going beyond simply learning anatomy, the order of surgical steps and other conceptual knowledge [21, 26].

We conducted iterative interviews with six surgeons and thirty residents from three top-tier US hospitals specializing in robotic surgery to investigate the current practices and challenges of video-based surgery learning. The interview results revealed that search and comparison were the most challenging tasks associated with learning surgery through video, as contrasted with navigation, note-taking, bookmarking, and reviewing. The tedium and time cost of search and comparison derived from the fact that video is a linear format, making it difficult to understand the semantic structure of procedural knowledge (e.g., sequences, relations between phases).

Limited Access to Semantic Information in Video Search. The residents we interviewed wanted a semantic video search by surgical phases and approaches (e.g., anterior or posterior approach in robotic prostatectomy). However, current video platforms only support basic search based on non-semantic information such as video length and upload date. No publicly available system automates the extraction of semantic or structural information from raw surgery recordings with similar video frames and doing so is difficult for humans, as this often requires expertise in the medical domain and significant effort to label such information.

Lack of Support for Cross-Video Reference and Comparison. Residents were usually under time pressure with excessive academic and patient workloads. They most commonly watched surgical video to prepare for a surgical case scheduled for the same or the next day. In this context, residents wanted to expose themselves to a wide range of surgical cases in a limited time to be prepared for various situations. They then sought the most feasible and well-performed procedure by comparing the techniques across multiple videos. Performing this comparative analysis was quite difficult as existing video interfaces (e.g., YouTube) did not support cross-video referencing interaction. Residents, as a consequence, ended up comparing multiple videos through time inefficient linear browsing, sometimes across multiple windows.

To summarize the results of multiple rounds of interviews, residents' main pain point was the limited access to semantic and structural information of surgeries while searching and comparing videos. Based on these findings, we designed a prototype to support structural search and cross-video referencing for surgery video. We especially focused on the recordings of robot-assisted radical prostatectomy, the most common robotic procedure, which provides us with abundant video data [16, 34]. Given that most robotic laparoscopic abdominal surgical procedures share phases, steps, and challenges (e.g., maintaining intraabdominal visibility, blunt dissection of fascia, retraction of organs and tissues, dissection of cancerous masses), much of our approach should be generalizable across procedures and disciplines using this tool (e.g., OB-Gyn, Colorectal, General surgery).

Our prototype consists of two main pages: video search and comparison interface. The search interface provides an interactive representation of surgical procedures. The visual representation helps residents understand the landscape of the surgery and supports graph-based interaction for filtering videos. For instance, surgical phases are visualized as nodes in graphs, and residents can click the nodes to watch the videos that contain certain surgical phases that correspond to the selected nodes. On the other hand, the comparison interface supports cross-video referencing. Residents can quickly reference and switch across videos by watching different clips in multiple windows. They can also synchronize the videos at a phase level by using graph-based interactions. To be specific, clicking a surgical phase that is included in two surgical procedures in common synchronizes the video play bars, thereby enabling an easy comparison for the same phase in multiple videos. Finally, we discuss the design implications of search and comparison tools for videos with procedural knowledge.

2 RELATED WORK

Video is a primary medium for surgical training [2, 4, 33]. According to Mota et al. [27], 98.6% of residents and surgical specialists are using videos for surgery preparation. Surgery video recordings provide key learning points in surgical training, such as anatomical landmarks and surgical maneuvers [1, 25]. They also accelerate the learning curve for surgical training [19]. As video becomes a predominant medium in learning surgery, several researchers designed supporting tools to enhance video-based surgery learning. The main two activities supported by the existing work are video navigation and collaborative interaction. To enhance video navigation in surgical recordings, Hudelist et al. [18] developed a video interface that displays clickable and zoomable keyframes of endoscopic videos. Munzer et al. [28] introduced EndoXplore, a video player that supports content-based video navigation based on phases of surgical operations. Their system automatically extracts the surgical phases from endoscopic videos and displays the clickable thumbnails that allow navigation to each surgical phase. On the other hand, surgeons and residents co-construct understanding on surgery through collaborative interaction such as telementoring [6, 11]. Mentisa et al. [25] investigated how a shared workplace is shaped by remote experts over surgery videos. Avellino et al. [5] introduced a system design for a collaborative video summary tool. They suggested that the requirements for a collaborative surgery summary tool include enabling appropriate division of tasks and management of dependencies between tasks.

However, semantic video search and comparison, the two main pain points revealed by our formative study, remained unsupported by the existing systems or techniques mainly focused on supporting video navigation and collaborative interaction. The importance of efficient video search and comparison tasks is further underlined as the massive volume of surgical recordings is piling up, driven by the prevalence of endoscopic video recordings and advances in camera technologies [30, 32].

3 FORMATIVE INTERVIEWS

To understand the current practices and needs surrounding video-based surgery learning, we conducted multiple rounds of need-finding interviews over a year with residents and surgeons. We used an iterative interviewing method [12, 38] that is appropriate for exploratory research and discovery of new research directions, allowing holistic investigation of research problems.

3.1 Method

3.1.1 Participants. We recruited six surgeons and thirty residents in urology in the U.S. by contacting surgeons through email, asking them to recommend colleagues who specialize in robotic surgery. The residents consisted of 6 PGY1, 6 PGY2, 7 PGY3, 3 PGY4, 7 PGY5, 1 PGY6 students. All of the participants have prior experience of learning surgery through videos. We refer to six surgeons as S1 through S6 and thirty residents as R1 through R30.

3.1.2 Procedures. We conducted remote semi-structured interviews using ZOOM and recorded the interviews under consent. The interviews lasted about 50 minutes with two main sessions: (1) current practice of video-based surgery learning and (2) challenges and needs. Finally, we presented a low-fidelity prototype of our video interface for surgery learning and collected feedback during the last batch of interview. All of the authors have received training on the ethical treatment of human subjects, and we did not collect any data from patients, including video recordings, clinical history, nor medical images. We used a saturation method [8] to determine the number of participants. We conducted several batches of interviews with 3-4 participants per each interview session, and performed a preliminary analysis of the transcripts for each batch. We stopped conducting more interviews when the analysis stopped revealing new insights.

3.2 Analysis

The interviews were video recorded and transcribed using a transcription service [1]. Two of the authors performed thematic analysis [17]. They independently made a codebook for half of the transcripts, using an inductive approach. They then merged and refined the codebook until they reached a consensus. The two authors then coded two randomly selected transcripts using the codebook. To validate the qualitative coding, we computed Cohen's kappa. The average Cohen's kappa score across the entire code was 0.81 with a standard deviation of 0.06. Each author then coded the rest of the interviews independently. After the coding, they met to discuss discrepancies in applying the code set and adjusted the coded data.

3.3 Findings

3.3.1 Context and Practice of Video-Based Surgery Learning.

The residents were using three main surgery video platforms: YouTube, Michigan Urological Surgery Improvement Collaborative (MUSIC), and American Urological Association (AUA). Most of our interviewees watched surgery videos the night before their operating room training. Their purpose of video watching was to remind themselves of certain surgical techniques or the flow of surgical phases. They stressed that they usually watch videos under tight schedules. Meanwhile, video-based surgery learning involved multiple activities, which includes video search, navigation, comparison, bookmarking, note-taking, and reviewing.

3.3.2 Challenges and User Needs.

Limited access to semantic information in video search. Residents' main consideration when searching surgery videos was the feasibility of procedures. They wanted to watch videos with feasible procedures that they can perform in practice. The feasibility of a certain procedure was determined by semantic information such as approaches and techniques. Residents pointed out that the existing video platforms do not support the semantic search so that they needed to tediously navigate to the video segments they want to watch. They wished for a semantic video search by which they can filter and find videos based on surgical approaches and phases.

Unavailability of structural information on procedural knowledge. Residents wanted to search videos using structural information of surgical procedures, for example, the compositions and orders of phases. The needs for the structural information differed depending on the expertise of residents. In the case of junior residents, they expressed a need for finding videos that contain a standard procedure to learn the common structure of phases. In addition to searching standard procedures, several novice residents wanted to find videos that do not miss or skip any surgical phase for a certain surgery. Several surgeons (S1, 2, 3) noted that surgical phases are one of the important learning points for novice residents. On the contrary, senior residents wished to search videos that can teach them variations and details of each phase. For example, R30 stated that "several surgery videos show interesting portion of surgery which is complex, but I'm a junior level and want to learn the basic steps.". R29, a PGY5 resident, also mentioned that "as a junior resident, you need to stick to your attending's approach, but as you become more independent, you might need to find your own technique." and "I want to go beyond. I would want to see many different versions of that specific variation.".

Lack of support for cross-video reference and comparison. We could observe a clear user need for comparing multiple videos. Residents wanted to compare different procedures, approaches (e.g., anterior approach or posterior approach in prostatectomy), and techniques (e.g., dissection using thermal energy or scissors). Meanwhile, most of our interviewees pointed out a surgical phase as a unit of comparison. R26, for example, said that "I want to know different

[1] https://otter.ai/.

Fig. 1. Users can search videos based on semantic and structural information of procedures using our prototype's video search UI. (a) Clustering Visualization: clustering results of surgical procedures are displayed, in which a dot represents a video clip. (b) Procedural Diagram: users can see the overall structure of procedures and filter the videos by clicking the nodes or edges of a diagram. (c) Filtered Video List: users can see a list of videos that are filtered by the procedural diagram.

techniques for the same phase.". R30 wished to compare a bunch of videos for the same phase, stating that "it'd be helpful if I can compare 20 different clips for a specific phase. Doctors do it slightly differently and patients' situations can be different.". S3 also mentioned that building links across multiple videos would enhance residents' learning. However, comparing videos for the same phase involves multiple steps. Residents needed to search a video, move back and forth to navigate to a phase they are interested in, play a clip, and repeat this process for the next video. They mentioned the inefficient process of comparing multiple techniques for the same phase using current video platforms.

Inaccessible surgery information. Some residents wanted additional information around surgery, such as operating room situations (e.g., port placement in robotic surgery) and anatomical landmarks, although they were not mentioned by many interviewees. The residents wanted to have audio narrations to know the anatomical landmarks and distinctions between phases. S1 explained that "understanding anatomy is one of the first steps of learning as residents.". They also wanted to know surgeons' expertise (e.g., surgeon volume) and surgery outcomes (e.g., restoring sexual function in prostatectomy), which are not accessible in most surgery recordings.

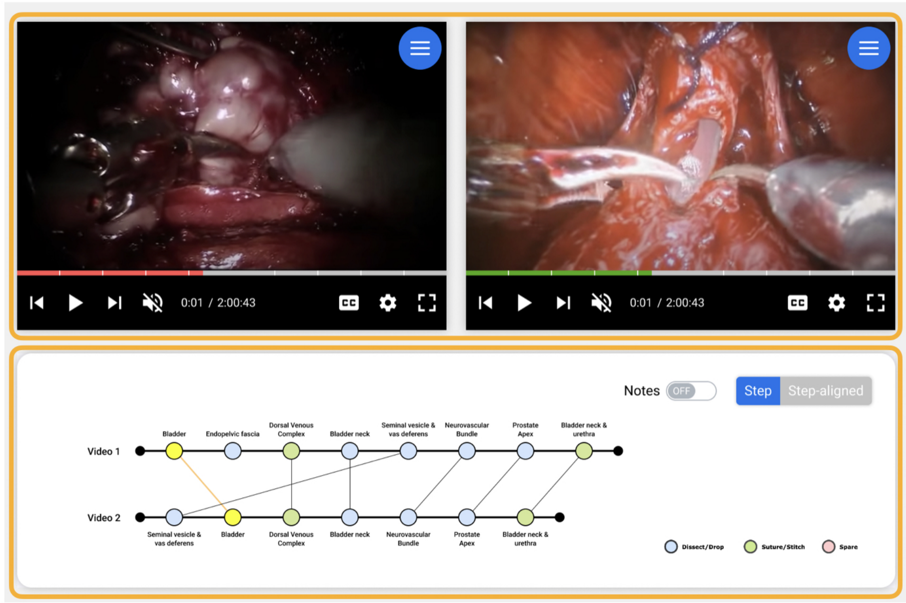

Fig. 2. Users can compare procedures across videos using our prototype's video comparison UI. (a) Multi-Window Videos: users can watch multiple videos through the juxtaposed video windows. (b) Phase-Based Video Synchronization: users can synchronize the video playback by selecting the phases on the interactive procedural diagram.

4 PROTOTYPE DESIGN

We designed a prototype depicted in Fig. 1 and Fig. 2 using Figma [2]. Our prototype introduces two interaction techniques to augment video search and comparison for surgery learning: semantic video search and cross-video comparison.

On the video search page, the prototype provides a clustering visualization (Fig. 1 (a)) and a procedural diagram (Fig. 1 (b)). It displays the clustering results of surgery videos, in which a dot corresponds to a video clip. Different colors represent clusters, and the distances between dots imply the similarity between procedures in videos. In other words, the farther the dots, the more dissimilar the procedures in videos are (e.g., a larger difference of compositions or orders of phases).

On the other hand, the procedural diagram shows the structure of surgical procedures, using a graph representation with nodes and edges. The nodes represent phases in procedures, and edges are for paths between phases. The thickness of edges reflects the commonness of paths. For instance, a thick edge implies that this path is included in many videos, indicating that the path is commonly performed by surgeons. The nodes are color-coded according to surgical actions, blue for dissection, green for suturing, and red for sparing (e.g., nerve-sparing). Meanwhile, users can also filter videos by clicking the nodes or edges. If they click an edge, our system displays a list of videos that contain a certain path (Fig. 1 (c)). This graph-based search filter enables video search based on structural information of procedures.

[2] https://www.figma.com/.

After referencing the clustering visualization and procedural diagram, users can select videos of interest. The selected videos are highlighted on the clustering visualization window. After selecting the videos, clicking the "Start Analysis" button leads users to a video comparison page (Fig. 2). The video comparison page supports multi-window videos (Fig. 2 (a)) and phase-based video synchronization (Fig. 2 (b)). Users can watch multiple videos through juxtaposed video windows. Meanwhile, they can synchronize the videos' playback using the interactive diagrams for the surgical phases. The connected lines between nodes indicate the common phases included in both videos. Users can synchronize the two videos by clicking nodes. For example, they can navigate both videos to the "bladder neck dissection" phase by clicking the first node in the upper line and the second node in the lower line.

5 DISCUSSION AND ONGOING WORK

5.1 Generality and Domain Specificity

5.1.1 User Needs for General Procedural Videos. During the interviews with surgeons and residents, we could observe user needs that overlap with those for general videos containing procedural knowledge, which includes cooking, make-up, and assembly. Procedural knowledge describes the steps in the process of performing a task or skill [13, 14]. Therefore, it is required to segment the video content into semantic phases to make the information of step structures available to viewers. Such information enhances video watching experience in learning make-up [37] and software tools [20], and so will the surgery videos [29]. According to the interviews, residents perceived surgical phases as an important unit of the stepwise structure, and wanted to search, navigate, and compare videos based on the phases.

In addition to the step structure, another component of procedural knowledge learning is acquiring tacit knowledge. The tacit knowledge contained in procedural knowledge includes, for example, cyclists' breathing methods and carpenters' hammering skills that they unconsciously perform [31, 35]. It is also one of the important learning points in the surgery domain [24]. The tacit knowledge residents carefully look at in videos includes how much traction is given for grabbing or whether the bleeding part is burnt or not. Our interviewees emphasized the importance of comparing multiple cases to see the differences in these intricate nuances.

5.1.2 User Needs Specific to Surgical Domain. On the other hand, we found user needs specific to surgical education. First, there were clearly distinct needs for video search and comparison depending on the expertise of residents. Junior residents (PGY1, 2, 3) expressed their need of understanding the overall structure of phases by watching videos with standard procedures. On the other hand, senior residents (PGY4, 5, 6) focused on variations and details of each phase. They wanted to compare multiple techniques and approaches for the same surgical phase.

Meanwhile, another distinction of learning surgery compared to general procedural knowledge is the complexity and intricacy of the task. Assume the differences between the two tasks, cutting the pork belly into strips and dissecting the bladder neck. The latter task involves more complicated considerations such as identifying anatomical landmarks, giving traction, and cleaning up bleeding. This complexity of task leads to residents' need for exposure to multiple techniques and approaches to be prepared for various situations.

Lastly, the unique characteristics of surgery recordings pose challenges of video search and navigation. Most surgery videos are long in length with several hours on average, contain similar video scenes of colors alike, and lack audio narrations. This lack of semantic indicators for video content deteriorates video-based learning experiences, making video search and comparison time-consuming.

5.2 Technical Challenges and Design Opportunities

Designing a system for surgery video search comparison, we identified several challenges and opportunities in building a technical pipeline and user interface. One challenge is to build clustering algorithms that can capture important features of sequences of surgical phases. Most existing clustering algorithms are black-box models whose internal structures of workings are not transparent and understandable to humans [15]. Thus, the clustering results can be uninterpretable even to medical experts.

The characteristics of black-box techniques can also lead to an issue of balancing findability and discoverability [22]. Findability refers to the ease of finding content or feature that users already know or assume. On the other hand, discoverability is the ability to unexpectedly encounter new content or feature that users are not aware of [9]. It is important to support goal-direct seeking, enabling residents to find the videos that contain approaches or phases they want to watch. At the same time, a system can add pedagogical value by allowing a loosely-directed exploration of unknown patterns or trends of procedures. We plan to support both findability and discoverability by providing affordances for advanced search features on demand. This on-demand support for an advanced feature will benefit users with a wide range of expertise and experience. Junior residents can focus on basic information such as learning a standard procedure, and senior-level residents can opt to compare variations of the standard procedure by comparing multiple videos.

Meanwhile, segmenting videos into surgical phases is a building block that enables phase-based search and comparison. However, as far as we know, there is no large-scale dataset nor a phase detection model for robot-assisted radical prostatectomies (RARP). Cross-video interactions can be another design opportunity for surgery video interfaces. Displaying multiple videos can cause cognitive load and overwhelm users. This unique challenge motivates the need for interaction support for seamless referencing and attention switching across multiple videos.

We plan to develop an interface that supports semantic video search and cross-video comparison. Finally, we envision designing a system to enhance search and comparison for videos with procedural knowledge.

REFERENCES

[1] Jad M Abdelsattar, TK Pandian, Eric J Finnesgard, Moustafa M El Khatib, Phillip G Rowse, EeeLN H Buckarma, Becca L Gas, Stephanie F Heller, and David R Farley. 2015. Do you see what I see? How we use video as an adjunct to general surgery resident education. *Journal of surgical education* 72, 6 (2015), e145–e150.

[2] Akgul Ahmet, Kus Gamze, Mustafaoglu Rustem, and Karaborklu Argut Sezen. 2018. Is video-based education an effective method in surgical education? A systematic review. *Journal of surgical education* 75, 5 (2018), 1150–1158.

[3] David Alderson. 2010. Developing expertise in surgery. *Medical teacher* 32, 10 (2010), 830–836.

[4] Knut Magne Augestad, Khayam Butt, Dejan Ignjatovic, Deborah S Keller, and Ravi Kiran. 2020. Video-based coaching in surgical education: a systematic review and meta-analysis. *Surgical endoscopy* 34, 2 (2020), 521–535.

[5] Ignacio Avellino, Sheida Nozari, Geoffroy Canlorbe, and Yvonne Jansen. 2021. Surgical Video Summarization: Multifarious Uses, Summarization Process and Ad-Hoc Coordination. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–23.

[6] Garth H Ballantyne. 2002. Robotic surgery, telerobotic surgery, telepresence, and telementoring. *Surgical Endoscopy and Other Interventional Techniques* 16, 10 (2002), 1389–1402.

[7] Matthew Beane. 2019. Shadow learning: Building robotic surgical skill when approved means fail. *Administrative Science Quarterly* 64, 1 (2019), 87–123.

[8] H Russell Bernard and Harvey Russell Bernard. 2013. *Social research methods: Qualitative and quantitative approaches*. Sage.

[9] James H Creager and Douglas J Gillan. 2016. Toward understanding the findability and discoverability of shading gradients in almost-flat design. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 339–343.

[10] Martin Dawes and Marko Lens. 2007. Knowledge transfer in surgery: skills, process and evaluation. *The Annals of The Royal College of Surgeons of England* 89, 8 (2007), 749–753.

[11] Simon Erridge, Derek KT Yeung, Hitendra RH Patel, and Sanjay Purkayastha. 2019. Telementoring of surgeons: a systematic review. *Surgical innovation* 26, 1 (2019), 95–111.

[12] Sam Fujisaka. 1986. Informal Structured Iterative Interview: A Method and Short Training Course for the central Visayas Research Consortium. *Philippine quarterly of culture and society* 14, 3 (1986), 263–274.

[13] Robert M Gagne. 1988. Mastery learning and instructional design. *Performance Improvement Quarterly* 1, 1 (1988), 7–18.

[14] Robert M Gagne, LJ Briggs, and WW Wager. 1988. Instructional design. *New York: Holt, Rinehart, Winston* (1988).

[15] Diansheng Guo, Donna Peuquet, and Mark Gahegan. 2002. Opening the black box: interactive hierarchical clustering for multivariate spatial patterns. In *Proceedings of the 10th ACM international symposium on advances in geographic information systems*. 131–136.

[16] S Duke Herrell and Joseph A Smith Jr. 2005. Robotic-assisted laparoscopic prostatectomy: what is the learning curve? *Urology* 66, 5 (2005), 105–107.

[17] Daniel J Hruschka, Deborah Schwartz, Daphne Cobb St. John, Erin Picone-Decaro, Richard A Jenkins, and James W Carey. 2004. Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field methods* 16, 3 (2004), 307–331.

[18] Marco A Hudelist, Sabrina Kletz, and Klaus Schoeffmann. 2016. A tablet annotation tool for endoscopic videos. In *Proceedings of the 24th ACM international conference on Multimedia*. 725–727.

[19] Andrew M Ibrahim, Oliver A Varban, and Justin B Dimick. 2016. Novel uses of video to accelerate the surgical learning curve. *Journal of Laparoendoscopic & Advanced Surgical Techniques* 26, 4 (2016), 240–242.

[20] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J Guo, Robert C Miller, and Krzysztof Z Gajos. 2014. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 4017–4026.

[21] Yu-Wei Lin, Holger A Volk, Jacques Penderis, Andrea Tipold, and Jan P Ehlers. 2015. Development of learning objectives for neurology in a veterinary curriculum: part I: undergraduates. *BMC veterinary research* 11, 1 (2015), 1–9.

[22] Stephann Makri, Yi-Chun Chen, Dana McKay, George Buchanan, and Melissa Ocepek. 2019. Discovering the unfindable: The tension between findability and discoverability in a bookshop designed for serendipity. In *IFIP Conference on Human-Computer Interaction*. Springer, 3–23.

[23] Robert McCormick. 1997. Conceptual and procedural knowledge. *International journal of technology and design education* 7, 1 (1997), 141–159.

[24] PJ McLeod, Yvonne Steinert, Tim Meagher, Lambertus Schuwirth, Diana Tabatabai, and AH McLeod. 2006. The acquisition of tacit knowledge in medical education: learning by doing. *Medical Education* 40, 2 (2006), 146–149.

[25] Helena M Mentis, Yuanyuan Feng, Azin Semsar, and Todd A Ponsky. 2020. Remotely Shaping the View in Surgical Telementoring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[26] George E Miller. 1990. The assessment of clinical skills/competence/performance. *Academic medicine* 65, 9 (1990), S63–7.

[27] Paulo Mota, Nuno Carvalho, Emanuel Carvalho-Dias, Manuel Joao Costa, Jorge Correia-Pinto, and Estevão Lima. 2018. Video-based surgical learning: improving trainee education and preparation for surgery. *Journal of surgical education* 75, 3 (2018), 828–835.

[28] Bernd Muenzer, Klaus Schoeffmann, and Laszlo Böeszöermenyi. 2017. EndoXplore: A Web-Based Video Explorer for Endoscopic Videos. In *2017 IEEE International Symposium on Multimedia (ISM)*. IEEE, 366–367.

[29] Tahmina Nazari, Floyd W van de Graaf, Mary EW Dankbaar, Johan F Lange, Jeroen JG van Merriënboer, and Theo Wiggers. 2020. One step at a time: step by step versus continuous video-based learning to prepare medical students for performing surgical procedures. *Journal of Surgical Education* 77, 4 (2020), 779–787.

[30] Maya Peled-Raz, Nadav Willner, Dan Shteinberg, Keren Or-Chen, and Tova Rainis. 2019. Digital recording and documentation of endoscopic procedures: physicians' practice and perspectives. *Israel journal of health policy research* 8, 1 (2019), 1–12.

[31] Kjeld Schmidt. 2012. The trouble with 'tacit knowledge'. *Computer supported cooperative work (CSCW)* 21, 2 (2012), 163–225.

[32] Klaus Schoeffmann, Heinrich Husslein, Sabrina Kletz, Stefan Petscharnig, Bernd Muenzer, and Christian Beecks. 2018. Video retrieval in laparoscopic video recordings with dynamic content descriptors. *Multimedia Tools and Applications* 77, 13 (2018), 16813–16832.

[33] Naomi M Sell, Douglas J Cassidy, Sophia K McKinley, Emil Petrusa, Denise W Gee, Mara B Antonoff, and Roy Phitayakorn. 2021. A needs assessment of video-based education resources among general surgery residents. *Journal of Surgical Research* 263 (2021), 116–123.

[34] NL Sharma, NC Shah, and David Edgar Neal. 2009. Robotic-assisted laparoscopic prostatectomy. *British journal of cancer* 101, 9 (2009), 1491–1496.

[35] John—Christopher Spender. 1993. Competitive Advantage from Tacit Knowledge? Unpacking the Concept and Its Strategic Implications.. In *Academy of Management Proceedings*, Vol. 1993. Academy of Management Briarcliff Manor, NY 10510, 37–41.

[36] Jason Stanley and Timothy Willlamson. 2001. Knowing how. *The Journal of Philosophy* 98, 8 (2001), 411–444.

[37] Anh Truong, Peggy Chi, David Salesin, Irfan Essa, and Maneesh Agrawala. 2021. Automatic Generation of Two-Level Hierarchical Tutorials from Instructional Makeup Videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[38] Daniel W Turner III. 2010. Qualitative interview design: A practical guide for novice investigators. *The qualitative report* 15, 3 (2010), 754.

[39] Andrew Yee, William M Padovano, Ida K Fox, Elspeth JR Hill, Amanda G Rowe, L Michael Brunt, Amy M Moore, Alison K Snyder-Warwick, Lorna C Kahn, Matthew D Wood, et al. 2020. Video-based learning in surgery: establishing surgeon engagement and utilization of variable-duration videos. *Annals of surgery* 272, 6 (2020), 1012–1019.