Objective Assessment of Surgical Technical Skill and Competency in the Operating Room

S. Swaroop Vedula[1], **Masaru Ishii**[2], and **Gregory D. Hager**[1]

[1]Malone Center for Engineering in Healthcare, Department of Computer Science, The Johns Hopkins University Whiting School of Engineering, Baltimore, Maryland 21218

[2]Department of Otolaryngology—Head and Neck Surgery, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21287

Abstract

Training skillful and competent surgeons is critical to ensure high quality of care and to minimize disparities in access to effective care. Traditional models to train surgeons are being challenged by rapid advances in technology, an intensified patient-safety culture, and a need for value-driven health systems. Simultaneously, technological developments are enabling capture and analysis of large amounts of complex surgical data. These developments are motivating a "surgical data science" approach to objective computer-aided technical skill evaluation (OCASE-T) for scalable, accurate assessment; individualized feedback; and automated coaching. We define the problem space for OCASE-T and summarize 45 publications representing recent research in this domain. We find that most studies on OCASE-T are simulation based; very few are in the operating room. The algorithms and validation methodologies used for OCASE-T are highly varied; there is no uniform consensus. Future research should emphasize competency assessment in the operating room, validation against patient outcomes, and effectiveness for surgical training.

Keywords

surgical technical skill; surgical technical competence; objective skill assessment; objective computer-aided surgical skill evaluation; OCASE; surgical data science

1. SURGICAL TECHNICAL SKILL AND COMPETENCY

In 2012, 312.9 million major surgical procedures were performed annually worldwide (36 million in the United States alone), which is a 38% increase since 2004 (1, 2). It has been well established that poor-quality surgical care is not only ineffective but also increases the risk of death and other severe complications, resulting in wasteful health care (3–8). In short, access to high-quality surgical care is integral to preserving global public health and is a major determinant of health care costs (1).

Efficiently training and credentialing competent surgeons is critical to ensure global access to high-quality surgical care. Although there is no consensus definition for surgical competency, most existing definitions refer to the level of skill required to safely and independently perform a procedure (9–11). This includes both technical and nontechnical skills as well as the knowledge and judgment required to complete new or familiar procedures (9–13). As a result, assessment of surgical skill and competence is an essential aspect of training and certifying surgeons (14–16).

Technical skill, which is the focus of this review, is a key aspect of training that affects the safety and effectiveness of surgical care. A considerable body of research suggests that poor technical skill is associated with severe adverse outcomes in patients, including death, reoperation, and readmission (3–8). In a study of medical malpractice claims in the United States, technical errors, most of which were manual in origin, were implicated in permanent disability or death of patients two-thirds of the time (17). Technical skills are also correlated with nontechnical skills (18), which, in turn, are associated with fewer technical errors per operation (19).

Assessment of technical skill is thus critical to train competent surgeons. However, in an era in which technology for surgical care has made rapid advances, surgical education and assessment have not kept pace (Figure 1). Traditionally, subjective assessment of technical skill was embedded within the Halstedian apprenticeship model for surgical training. More recently, global rating scales (GRS), such as the Objective Structured Assessment of Technical Skills [OSATS-GRS (20)], have been developed for objective and standardized assessment (21–24). Although GRS accurately distinguish surgeons with different levels of experience (22), integrating them into surgical training curricula has been limited, in part due to resource constraints because they require an expert observer (21, 25). Recent innovations using crowdsourcing may be an efficient alternative to observation by experts, but their role in surgical training curricula is still uncertain (26–30). Other, less resource intensive approaches, such as in-training evaluations and examinations, or surrogates, such as surgical volume (31–33), do not correlate with objective assessments of surgical technical skill (34).

With increasing attention to the efficiency and effectiveness of health care, multiple factors are converging to make traditional methods of training combined with structured manual assessment untenable. First, health care is being driven by a goal to maximize value, namely outcomes achieved per dollar spent (35). The cost and safety implications of teaching in the operating room severely limit opportunities to systematically impart and assess intraoperative technical skill (36, 37). Second, there is increasing awareness of the harmful consequences to patients from preventable technical errors and an emphasis on minimizing them, placing great pressure on efficient and accurate quantification of surgical competence. Third, surgical trainees must acquire skill and competency with an increasingly wide spectrum of procedures and techniques such as laparoscopic, open, robotic, and endoscopic surgery, within limited work hours and access to educators.

The confluence of these pressures, together with the growing availability of quantitative data documenting surgical performance and recent developments in machine learning methods,

has led to a rapidly growing literature on objective computer-aided skill evaluation (OCASE) of surgical technical skill (OCASE-T). OCASE-T offers a "surgical data science" approach that can provide low-cost, high-value feedback to surgeons in training and practicing surgeons to improve their acquisition and maintenance of skills (38). Developing data science methods for surgery is challenging, and depends on creating platforms to capture data, algorithms to provide assessments from data, and mechanisms to leverage these tools within surgical training curricula. Despite these challenges, research progress on OCASE-T has been rapid over the past decade. However, the literature describing this progress is scattered through a variety of publication venues and research fields. In this review, we bring together the disparate literature and provide a synopsis of the state of the art in OCASE-T research and applications.

This review is intended to elucidate the conceptual problem space around the emerging area of OCASE-T, and to provide a comprehensive synthesis of research in this area. Research on OCASE-T is highly multidisciplinary, ranging from fundamental technical work in computer science and electrical engineering to applications of methods addressing the needs of surgeons and surgical educators. In the interest of addressing a broad audience, we note that tutorial material on different analytical techniques and algorithms used for OCASE-T is beyond the scope of this review. Such details can be found in other resources (e.g., (39–41)). Instead, we present a conceptual discussion of different aspects of OCASE-T with reference to the validity and utility of the algorithms.

We have chosen to focus on data analytics algorithms that use any combination of tool motion, video images, surgeons' eye gaze, and other data available in the surgical context for objective skill assessment. To identify the specific papers reported herein, we searched Inspec, Pubmed, Google Scholar, the Proquest® database of dissertations and theses, and references listed in articles and prior reviews describing algorithms for OCASE-T (42). Studies on summary metrics of time and motion efficiency for technical skill assessment were excluded; these metrics are discussed elsewhere (43–48). Technical skills required for interventions such as central vein catheterization, biopsies, and airway management are also not considered here (49). Finally, we note that this review focuses on skill assessment, but does not consider the complementary problem of automatic detection of surgical activities or phases/workflow. These methods are discussed elsewhere (50–53).

We identified 45 publications from our literature search that encompassed OCASE-T for various surgical techniques and platforms such as traditional open surgery and minimally invasive techniques such as laparoscopy, robotic surgery, endoscopy, natural orifice transluminal endoscopic surgery, and endovascular interventions. Only two studies described OCASE-T for procedures in the operating room; the rest focused on benchtop or virtual reality simulations (Table 1).

The remainder of this article reviews the available literature as follows. Research questions for OCASE-T are described in Section 2, sources of data for OCASE-T in Section 3, different representations of the data in Section 4, algorithms for OCASE-T in Section 5, and feedback based on OCASE-T in Section 6. We present our conclusions in Section 7.

2. RESEARCH QUESTIONS FOR OCASE-T

Defining the problem space for OCASE-T is complicated by the fact that surgical skill is a multidimensional construct with cognitive and motor components, and there is little consensus on dimensions constituting technical skill (22). Furthermore, technical skill and its assessment are affected by several factors, including surgeons' stress and fatigue (98), ergonomics (99), environment in the operating room (e.g., lighting), surgeons' nontechnical skills, and case complexity. Thus, algorithms for OCASE-T must be developed and evaluated in the context of the factors affecting them.

It is also essential to recognize that the ultimate goal of assessment is to improve the process of acquiring technical skill, which relies on deliberate practice and appropriate feedback. Deliberate practice refers to directed repetition of specifically chosen activities tailored to improve an individual's performance (100–102). Such repeated practice can be assisted by immediate feedback, typically from a teacher or a coach. Coaching by an expert surgeon effectively improves technical skill acquisition (102).

Taken together, these data suggest that the end goal for OCASE-T should not be simply assessment but also automated coaching for efficient acquisition and maintenance of technical skill through deliberate practice. This goal can be attained when technology for OCASE-T not only measures skill but also enables feedback through targeted assessments (i.e., how to do components of a surgical task) and diagnosis of skill deficits or errors in performance (i.e., how not to do a surgical task). Figure 2 illustrates the resulting "closed-loop" system enabled by OCASE-T. Data captured in the skills training laboratory and in the operating room are at the core of OCASE-T. These data transform into an intermediate representation to extract pertinent information, which is then modeled using various data analytics algorithms for OCASE-T. The endpoint for OCASE-T is then feedback to surgeons through data products that rely on the algorithms, and the process repeats.

More concretely, raw data for OCASE-T are usually a time series (Figure 2). The raw data, $\mathbf{T}$, may be complicated by many factors unrelated to technical skill, often necessitating a transformation of the data into a more refined or succinct representation, $\mathbf{X}$, by suppressing unwanted variation. We refer to both $\mathbf{T}$ and $\mathbf{X}$ as observations or observed data. The analytical goal of OCASE-T is to learn a mapping, Φ, from $\mathbf{X}$ to a ground-truth assessment, $\mathbf{Y}$, that may be discrete (e.g., beginner versus expert) or continuous. $\mathbf{Y}$ is typically acquired from an expert panel. Numerous algorithms may be chosen for Φ; the choice of method and its performance are heavily influenced by how input data are represented ($\mathbf{X}$) and the form of $\mathbf{Y}$. The relative importance of choosing a data representation versus an algorithm for OCASE-T is still an open question. Finally, feedback based on OCASE-T relies on providing some diagnostic or illustrative feedback as to what the learner may do to improve their performance.

The rest of this review is organized around the four key steps in this process model:

1. How may data relevant for OCASE-T be routinely captured during surgical training and in the operating room?

2. How may data be represented or transformed to maximize performance and robustness of algorithms for OCASE-T and to optimize feedback based on OCASE-T?

3. How may data be mapped to valid assessments of surgical technical skill and competence using data analytical algorithms?

4. How may algorithms for OCASE-T be used to enable effective feedback for technical skill acquisition through diagnosis of skill deficits, targeted assessments, and detection of errors in performance?

Before proceeding with specifics on the key steps in the process model, we note that reporting valid measures of technical skill is an explicit end goal of most algorithms for OCASE-T. The validity of the algorithms for OCASE-T may be assessed through several sources (103), including relation to pertinent external variables, generalizability of assessment, structural and content aspects, and consequences of assessment. Existing research on OCASE-T has focused exclusively on validation through relation to external variables, namely validity against criteria considered ground-truth skill assessments. Such validation was limited to surgeons' appointment status, years in training or practice, experience, or OSATS-GRS as the ground truth. In addition, all of the studies we identified uniformly recruited convenience samples of participants, which raises the potential for selection bias and limits the generalizability of the findings. None of the studies evaluated consequences of assessment for training or patient care.

Although nearly all of the 45 publications discussed in this review reported some type of assessment of algorithm performance (e.g., classification accuracy), the diversity of approaches, surgical tasks studied, data sets, and validation methodologies precluded a comprehensive comparative assessment or compilation of the validity of algorithms used for OCASE-T. Consequently, the below discussion is limited to those instances in which the surgical tasks and metrics for validation are directly comparable.

3. SOURCES OF DATA FOR OCASE-T

Surgery is characterized by several rich sources of fine-grained data throughout the patient care process, including preoperative patient data, video of the surgical field, tool use and movement data, video and other data of the operating room environment and personnel, physiological status of the patient, and postoperative outcomes. However, to date most of these data have not been systematically captured, presenting a major challenge to OCASE-T. Table 2 summarizes data collected for OCASE-T in existing research.

The ideal data collection system for OCASE-T is transparent, pervasive, complete, integrated into workflows, scalable, and automated. Transparency of the system, meaning that it is invisible to the user, is essential to secure buy-in from clinical teams, patients, and ethics review boards. Pervasiveness means that the data are always collected, whether in the surgical skills training laboratory or in the operating room. Complete data collection systems capture all relevant manifestations of technical skill and the multitude of factors that affect surgical performance and its assessment (Figure 3). Data collection systems must also have a small footprint and be automated (require minimal space and participant input) for effective

integration into surgical training and patient care workflows. Finally, automation is also essential for scalable data collection to efficiently build large data corpora for OCASE-T. Full automation of integrated data collection systems for OCASE-T may also require additional technology for context awareness using activity recognition (50–53).

Existing integrated data collection systems for OCASE-T are far from ideal and are diverse in terms of both the environments in which they may be deployed and the data they capture. Integrated systems developed for laparoscopic training platforms are able to capture tool motion, arm and head pose, eye gaze, electrocardiogram data, and audio recordings (104– 106). These systems have used either infrared or color markers that are visually tracked and thus require an uninterrupted line of sight. Another integrated system, which may be applicable across different surgical platforms (107), used electromagnetic sensors and potentiometers attached to instruments to capture tool motion and pose in addition to video and other environmental signals. Ostler et al. (108) developed a centralized framework for acquisition of data on laparoscopic procedure workflow in the operating room. Finally, Zhao (109) developed a general-purpose system, called the surgical data collection, integration, storage, and retrieval system, to capture intraoperative data for laparoscopic surgery including endoscopic video, speech, system events such as usage of the foot pedal, and patients' physiological data.

The surgical technique or platform affects the ease and feasibility of capturing data for OCASE-T. Surgeons' hands must be tracked during open surgery, in contrast to surgical tools in various minimally invasive surgical techniques (e.g., laparoscopic, robotic, endoscopic, and microscopic surgery). In addition, video images of the surgical field can be recorded more easily from endoscopic cameras that are an integral part of minimally invasive techniques than with open surgery. Many different systems for tracking laparoscopic tools have been developed, which are reviewed elsewhere (110). More recently, such systems for tracking laparoscopic tools have been developed for use over the worldwide web (75). However, data collection from robotic surgery continues to be the closest to ideal, as complete, transparent, scalable, and automated data capture is relatively easy to achieve (111).

The surgical environment also influences the types of data available for OCASE-T and the difficulty with which they can be captured. Data in the training laboratory are easier to capture than in the operating room, and may also be more extensive. At one extreme, virtual reality or computer simulation readily provides data on all aspects of tool and tissue manipulation. Benchtop simulation is amenable to instrumentation, which provides data on both surgical tools and manipulated tissue (46, 83). Data for OCASE-T in the operating room has so far been limited to surgeons' eye gaze and movement of surgical tools (54, 55).

Surgical tools are a major source of data for OCASE-T. In the case of robotic surgery, rich stereo video and tool movement data can be captured completely transparently in both the training laboratory and the operating room (76, 78–80, 85, 86, 111). In nonrobotic cases, tool motion data in the training laboratory can be captured by affixing a variety of sensors to the instruments, for example, accelerometers, strain gauges, and electromagnetic or optical trackers (46, 66, 67, 72, 73, 86, 87, 112, 113).

The need to preserve a sterile surgical field is a major factor affecting the scope and feasibility of data that may be captured in the operating room. Tool motion in the operating room may be captured by attaching sterilizable sensors to the instruments (54). Wireless sensors to track tools would be ideal for use in the operating room because they would avoid clutter in the surgical field, but such sterilizable sensors have yet to be developed (114–116).

Tool tracking using approaches that require a direct line of sight, such as optical markers or tracking based on video images, is also harder to implement in the operating room than in the training laboratory (68, 117). More generally, although alterations of surgical tools and surgical workflow for OCASE-T are feasible (54), whether such methods are scalable across procedures is unknown.

The extent of tissue manipulation is more easily captured in the training laboratory than in the operating room. Tissue manipulation not only is a measure of technical skill but also quantifies the magnitude of surgical intervention, enables assessment of errors in performance (75), and serves to explain variation in outcomes of intervention across patients (118). In the training laboratory, forces applied on the tissue may be captured using sensorized models such as a "smart skin" with an array of sensing elements or a skin dummy with a photo interrupter–based mechanism to measure tissue deformation during suturing (46, 119). Force sensors may also be built into platforms holding the inanimate model in benchtop simulation (71, 83, 87). In the operating room, video images of the surgical field are the primary source of data on tissue manipulation. Consequently, the granularity at which deformation of tissues during surgery can be measured is limited.

Some data can be captured in the training laboratory and in the operating room with equal ease. For example, surgeons' eye gaze may be captured using trackers placed on wearable head mounts or eyeglasses, or remotely on high-resolution monitors (55, 56, 120–122). Similarly, data on surgeons' posture or physiological function may be obtained in both environments (99, 113). Such data include surgeons' posture, muscle activation, electrodermal activity, temperature, heart rate, respiratory rate, blood pressure, and functional magnetic resonance images (113, 123). Although advances in wearable technologies allow easy capture of variety of data from the surgeon, their utility for OCASE-T has yet to be demonstrated.

Some sources of data for OCASE-T are pertinent only in the operating room context, for example, the environment (e.g., lighting), interaction with other personnel, patients' physiological status, medications administered, and outcomes of surgical intervention. Data on patients' physiological status may include heart rate, blood pressure, breathing (ventilation settings), temperature, and fluid inflows and outflows. Although data on the patient's physiological state may not inform technical skill assessment, they may independently affect outcomes of care. These data are therefore relevant to understand heterogeneity in how technical skill affects patient outcomes. For example, technical skill is associated with postoperative adverse outcomes (3), but technical skill may have a negative correlation with patient-centered or functional outcomes (124).

There are significant systemic and logistical barriers that must be addressed in order to develop universal data collection systems for OCASE-T. First, surgical workflow is extremely diverse across different procedures and surgical teams. Data collection systems must be flexible to enable integration into customized patient care workflows. Second, privacy and confidentiality of data on surgical performance and patient outcomes must be rigorously protected, particularly in contexts in which such information has substantial medicolegal implications. Third, generating structured data sets from raw data captured in the surgical context is challenging. In addition to common standards for data formats and ontologies, structuring raw surgical data requires extensive curation such as annotations for surgical phase, activity, skill, or other descriptors of performance or ground-truth assessments and outcomes for validation of algorithms. Finally, the surgical community itself must demand that such data collection become pervasive and lead its incorporation into training and patient care in order to establish a culture of data-driven continuous improvement.

4. DATA REPRESENTATIONS FOR OCASE-T

Data for OCASE-T that emanate from different sensors (**T**; see Figure 2) are complex, noisy, sampled at high frequencies; include numerous highly correlated dimensions; and incorporate several sources of variation. As such, preprocessing and transforming the raw sensor data into more compact representations (**X**; see Figure 2) achieve the following goals to varying extents (125):

1. Remove noise and irrelevant sources of variation in the data.

2. Reduce data dimensionality.

3. Enhance or extract information relevant for skill and/or feedback.

The type of raw data captured from sensors and the target algorithms for OCASE-T determine the relevance and utility of different transformations (Ψ; see Figure 2). The data transformations used for OCASE-T in the existing literature may be grouped into computation of fixed-dimensional summary features, conversion into time-series models, and transformation into dictionaries or histogram-based representations. Table 3 lists the representations used for different types of data for OCASE-T in the literature.

4.1. Preprocessing Data for OCASE-T

Raw data for OCASE-T are usually preprocessed to achieve objectives such as smoothing, normalizing dimensions, removing coordinate dependence, and extracting information from the data. Techniques such as a moving average or a median filter are used for data smoothing (73). Standardization or normalization of dimensions (centering by the mean and scaling by the variance) is useful for subsequent analytics that rely on computation of distances or maximizing variances (70, 71). The Fourier transform (FT) is a standard preprocessing technique for time-series signals and is used to identify the frequency components that make up a continuous signal (126). FT within short discrete time windows [short-time FT (STFT)] simultaneously yields time and frequency information in the signal (76, 95).

Some types of data, such as tool position, velocity, acceleration, and forces, are embedded within a coordinate system. Preprocessing techniques may be used to transform time-series data into consistent coordinates, using an origin either defined by the time series or external to it. Leong et al. (94) and Zhang & Lu (127) computed the centroid distance to induce rotational and translational invariance in surgical tool positions. Descriptive curve coding (DCC) is another technique that relies on an origin within the time series, which is specified as a Frenet frame along different points on the motion trajectory (57). DCC results in a time series of discrete symbols corresponding to the accumulated change in direction of the trajectory within a window specified over time or space. In an example of inducing coordinate invariance using an origin external to the time series, Jog et al. (97) transformed data with respect to trajectory space of an expert performance.

Preprocessing video image data for OCASE-T may involve any number of computer vision techniques, most of which are beyond the scope of this review. Existing studies on OCASE-T using video images have employed spatiotemporal interest points (STIPs) (128) to identify repeatable locations. A small volume of pixels around each STIP is used to compute features such as histograms of orientation gradients (HoG) and optical flow (HoF) as input into algorithms for OCASE-T (64, 65, 67).

4.2. Summary Features to Represent Data

Computing summary features for OCASE-T involves collapsing time-series data into a fixed-dimensional vector. Summary features have distinct advantages in spite of information loss during transformation. First, summary features enable the application of a variety of algorithms that take fixed-dimensional vectors as the input. Second, they can be informative about specific aspects of the surgical task or its performance, and can therefore form the basis for feedback. Third, summary features may improve the performance of algorithms for OCASE-T because they extract specific and pertinent information from the time-series data.

Summary features for OCASE-T may be nonspecific derivatives from data preprocessing techniques, generic descriptors of motion efficiency, or hand-engineered features. Summary features are more easily computed for tool motion and eye-gaze data, whereas video images must be extensively preprocessed to extract features. Principal components that explained the majority of variation in the data were used as nonspecific summary features for OCASE-T in some studies (69, 70, 73, 79). Generic descriptors of tool motion efficiency, such as time, path length, or movements, which may be easily computed on any surgical platform, were used for OCASE-T in several other studies (78, 79, 81, 82, 84, 85, 93, 97). Descriptors of other measures of tool motion, such as force, were used as features for OCASE-T in a few studies (54, 70, 71, 83, 88). Finally, generic features may also refer to frequency characteristics of the signal, such as the power density spectrum of speed measured at joint angles (73), and frequency-spectrum profile specified as magnitudes of coefficients from a fast FT (46, 54, 76).

Hand-engineered summary features are designed on the basis of prior knowledge of the structure of surgical activity as well as an expert understanding of how to skillfully perform the task and specific technical skill deficits. Such features are typically defined with respect to anatomical structures in the surgical field. For example, Sewell (89) specified velocities

and forces near the facial nerve during mastoidectomy. Similarly, Ahmidi et al. (54) defined features to describe strokes made with the surgical tool to elevate the mucosal flap overlying the nasal septal cartilage in septoplasty.

More recently, Ershad et al. (129) developed a novel crowdsourcing approach to generate summary features using semantic labels that describe quality of tool or hand movement. The semantic labels were specified in contrasting pairs, which were then mapped to corresponding data-driven metrics. For example, high/low values for mean jerk (derivative of acceleration) correspond to a label pair of crisp/jittery, respectively. Whereas metrics such as jerk may be computed with data on tool motion, metrics such as relaxed/tense and calm/ anxious require other measurements, such as electromyography and galvanic skin response, respectively. Semantic representation of surgical performance based on quality of movement is an innovative frontier that harnesses human insights, but their validity for OCASE-T and utility for automated feedback have yet to be determined.

4.3. Time-Series Data Representations

Time-series representations of raw data from sensors for OCASE-T capture inherent sequential information, but algorithms that capture sequential information in the time-series data are often sensitive to the underlying coordinate system. Thus, preprocessing the data to induce coordinate invariance is routinely performed before time-series data are used for OCASE-T.

A common transformation of time-series data for OCASE-T is conversion to a time series of discrete symbols. Discretization reduces the complexity and dimensionality of the data by grouping together similar points in the time-series signal in terms of a small set of prototypes. Several different clustering techniques have been used for OCASE-T, including vector quantization (VQ) with algorithms such as K-means, piecewise linear approximation (PLA) (66), and DCC (57). K-means was used to discretize raw time-series data from sensors (56, 58, 60, 62, 72, 87, 90, 91, 95) and to cluster coefficients from STFT (76). Other techniques, such as symbolic aggregation approximation (SAX), aligned cluster analysis (ACA), Persist (130), and methods based on affine velocities and the two-thirds power law (131), have been explored to transform time-series data for surgical activity detection but not for OCASE-T.

4.4. Dictionaries or Histogram-Based Representations

Dictionaries or histograms of data for OCASE-T are computed either as class-specific summaries or as features from time-series data. In one approach using data on tool position, following a discrete transformation with DCC, Ahmidi et al. (77) applied a common substring model (CSM). Class-specific CSMs were computed by applying a longest common substring algorithm to the time series of DCC symbols, presumably capturing motifs within surgical tool motion. Similarity metrics to each class-specific CSM were used for OCASE-T. In a similar approach, class-specific histograms of motion increments were used to represent tool motion data and similarity metrics were used for subsequent skill classification (59). In a different approach, Khan et al. (66) computed features to describe

transition matrices obtained from PLA-derived discrete representations of surgical tool accelerations.

Bag-of-words (BoW) dictionaries are a common way to represent data from video images for OCASE-T (64, 65, 67, 69, 74). In this approach, visual features (STIPs in this case) are clustered using K-means applied to HoG and HoF features. Thus, each image is described as a histogram of cluster labels for the STIPs (BoW). Traditional BoW dictionaries ignore the order of cluster labels. Sequential information in video images may be captured by augmenting BoW dictionaries with n-grams or randomly sampled regular expressions (132). In addition, features corresponding to the texture of motion may be extracted from the histogram representation of images (64, 65, 67).

Research on the best way to structure data for OCASE-T continues, and may evolve as OCASE-T methods mature. The data representation techniques discussed above, including DCC, K-means, BoW, and augmented BoW (A-BoW) differ in the extent to which they preserve the geometry or inherent structure of surgical motion. DCC encodes changes in the direction of tool motion trajectories and thus preserves the geometry of tool motion in a coordinate-invariant manner. By contrast, BoW and K-means encode only the occurrence of local patches of features, not their relative geometry.

5. ALGORITHMS FOR OCASE-T

The optimal algorithm for OCASE-T is determined by the type of data available, how the data are represented, and the type of task (classification or regression) to be performed (Table 3). Existing algorithms for OCASE-T may be categorized by how input data are represented:

1. methods in which X_i is a fixed-dimensional vector of summary features, including support vector machines (SVMs), linear discriminant analysis (LDA), discriminant function analysis (DFA), nearest-neighbor classifiers (NNC), logistic regression (LogReg), fuzzy models (FMs), linear regression (LinReg), and neural networks (NNets) such as learning vector quantization (LVQ);

2. methods in which X_i is a time series, including Markov models (MMs), hidden Markov models (HMMs), and NNets; and

3. methods in which X_i is a dictionary or histogram-based representation and related metrics.

Well-designed algorithms for OCASE-T should be valid, generalizable across use cases, and robust to factors affecting technical skill, and should enable effective targeted feedback. In addition, the algorithms should be efficient to implement at scale and intuitive to use. Scalability of algorithms is heavily affected by the level of manual annotations required for data curation and other manual input needed to provide a result. Curation of the data may involve segmentation into constituents, such as activities or instances of performances, coordinate transformations, or other preprocessing, before the algorithms can be trained or implemented. Algorithms for OCASE-T must also be easily retargetable to be scalable

across procedures, application domains (e.g., training laboratory to the operating room), surgical techniques, and surgical sites.

5.1. Methods That Use Summary Features

Algorithms for classification or regression are typically classed as discriminative or generative. Discriminative models involve learning a function to directly map $\mathbf{X}$ to $\mathbf{Y}$; Φ: $\mathbf{X}$ $\rightarrow$ $\mathbf{Y}$. Generative models involve first learning the class-specific joint distribution of observations ($\mathbf{X}$) and ground truth ($\mathbf{Y}$), $P(\mathbf{X}, \mathbf{Y})$, which is then used to predict class labels $P(\mathbf{Y}|\mathbf{X})$ with the Bayes rule. These models are referred to as generative because synthetic data can be produced by sampling from the joint distribution (39).

Methods that use summary features as data representation for OCASE-T are uniformly discriminative. Some algorithms for OCASE-T that use summary features identify linear decision boundaries (LinReg, LogReg, LDA, DFA), whereas others identify nonlinear decision boundaries (NNC, SVM). The methods also differ in how the decision boundaries are modeled and derived. In one approach, which includes LogReg, LDA, and DFA, a discriminant function is modeled for each class label in the ground truth (46, 73, 89). Each observation is assigned to the class with the largest value for its discriminant function. In another approach, boundaries to optimally separate classes in the ground truth are directly modeled (e.g., SVM) (54, 57, 78, 79, 81, 82, 84, 97). In a third approach, including what may be considered prototype-based methods (NNC and LVQ NNets), each observation is assigned a class label corresponding to its nearest prototype (55, 88).

Summary features derived from simple data preprocessing such as components from principal components analysis that explain majority of variance in the data are also useful for OCASE-T. For example, Horeman et al. (70, 71) reported 100% accuracy in classifying performances as either expert or novice. More complex signal processing, such as multivariate autoregression (MVAR) (96), yielded higher sensitivity and specificity than HMMs trained using time-series data to classify task performances into expert or novice (86–96% versus 64–87%). Generic features of tool motion efficiency yield moderately accurate algorithms to assign one of two class labels for skill (expert versus novice). Reported measures of accuracy in classification were similar across two studies for a suturing task (64–79%), with minimal differences among linear, nonlinear, and prototype-based classifiers (78, 84, 85). However, one study using LVQ NNets reported a low accuracy of 50%; performance in this study was lower for distinguishing intermediates versus experts than for novices versus experts (88). A few studies reported very high measures of accuracy with simple methods such as LogReg (83) and LDA (70, 71) by using features from tool forces and vibrations, suggesting that these features capture information discriminative of skill. Finally, two studies describing hand-engineered summary features for OCASE-T reported high classification accuracies of at least 87.5% (54, 89). In addition, hand-engineered features appeared to yield greater accuracy for a two-class classification (90.9%) compared with simple signal-based features such as the frequency-spectrum profile (71.2%) (54).

5.2. Methods That Use Time-Series Data

Methods described in the existing literature that use time-series data for OCASE-T may be discriminative (NNets) or generative (MMs and HMMs). Nearly half of the studies for OCASE-T we identified in the literature evaluated MMs or HMMs (Table 3). Technical details about MMs and HMMs can be found elsewhere (133). For OCASE-T using MMs or HMMs, separate class-specific models are trained using data from experts and novices. A test performance is then compared with each class-specific model by using a distance metric. The test performance is assigned the class label corresponding to the model to which it is closer on the basis of a similarity metric, or the distance to all class-specific models is input to an additional classifier (e.g., SVM). Several metrics of similarity of a performance to a class-specific model have been described (60, 62, 63, 76, 95). By contrast, the existing literature on NNets for OCASE-T using time-series data is limited to a single study (55). In this study, a high classification accuracy (expert versus novice) using eye-gaze data was reported in a virtual reality simulation (93%) and in the operating room (91%).

Two methods that use time-series data for OCASE-T, namely MMs and HMMs, have been studied across all surgical techniques and environments (simulations, animal and cadaver models, and the operating room), and to assess a variety of surgical tasks. In addition, different types of data have been modeled using MMs and HMMs, including tool motion, eye gaze, and video images (Table 3). HMMs for OCASE-T are specified in several different ways in accordance with the data representation that was used. For example, Gaussian mixture models or Dirichlet processes, multinomial probability distributions, or sparse dictionaries of motion words may be specified to model emissions from the HMMs (54, 60, 61, 69, 80).

The performance of HMMs for OCASE-T is sensitive to the surgical environment and type of data. HMMs yielded only low to moderate accuracy (29% to 71%) for skill classification for nasal septoplasty in the operating room (expert versus novice), using tool position, orientation, and velocities (54). In a cadaver model, higher sensitivity in classifying skill was reported when surgeons' eye-gaze positions were used in addition to position and orientation of surgical tools (83% versus 73%) (56, 58). Although several studies used discrete representations of tool motion and other data such as manual annotations of tool–tissue interactions, whether such discretization translates into improved performance of time-series methods such as MMs or HMMs is unknown (Table 3).

The performance of MMs and HMMs for OCASE-T is also sensitive to the coordinate system in which the time series are embedded. For example, the position of the tool relative to that of the endoscopic camera in sinus surgery appeared to yield higher sensitivity than absolute tool position (77% versus 73%). Similar findings were reported in other studies (89, 92), in which slightly higher skill classification accuracy was observed when features with relative information, such as distance between instruments, were used to train HMMs (77% versus 81%). Although transformations such as DCC (77) and the centroid distance function (94, 127) have been found to induce coordinate invariance in tool motion data, direct comparisons to ascertain the impact of such transformations on performance of HMMs for OCASE-T are lacking.

Finally, the performance of HMMs for OCASE-T is sensitive to new surgeons (i.e., surgeons whose data are not used to train the HMM) (54, 63, 80). The performance of other algorithms based on dictionary representations of data after transformation for coordinate invariance may be robust to data from new surgeons (54).

5.3. Methods That Use Dictionaries or Histogram-Based Representations

Dictionaries or histogram-based representations of data for OCASE-T are used to compute metrics of similarity to class-specific models, or to further extract time-series representations or a vector of summary features. These derivative data representations are subsequently used as input for appropriate algorithms, for example, HMMs for time-series data and SVMs for vectors of similarity metrics or summary features. In a specific example, Ahmidi et al. (59) computed the Hellinger distance between the test-performance and class-specific histograms and assigned the label corresponding to the class to which the distance was lower. In another example, a customized metric was specified (77) to include not only similarity to a class-specific dictionary (CSMs) of longest common substrings but also the location and length of the substrings. As noted above, this algorithm yielded high accuracy in classifying expert versus novice performances both in the operating room (54) and in the training laboratory (77).

Dictionary representations of video images for OCASE-T have been used to extract a time series of histograms (67, 69, 74) and a vector of summary features (64, 65, 67), which then served as input to appropriate algorithms for classification of a skill class or regression on a skill score. Algorithms based on a dictionary representation of video images have been used to predict assessments for components within OSATS-GRS with accuracies ranging from 84% to 100% (64, 65, 67, 69, 74, 75).

6. FEEDBACK BASED ON OCASE-T

Feedback is essential for learning. In one study, verbal feedback and instruction by an expert surgeon led to improved retention after 1 month (134), although the effect did not persist after 6 months of learning (135). Feedback based on OCASE-T has been described in terms of summary features for OCASE-T, assessment of errors in performance, or skill assessment within low-level activity segments. However, the effectiveness of feedback based on targeted assessment of activity segments of interest has yet to be determined.

Several studies have suggested that summary features used for OCASE-T can also serve as feedback for learning. Rosen et al. (61) reported descriptive analyses showing differences between summary features of forces/torques applied on the surgical tool by expert and novice surgeons. On the basis of the observed differences, Rosen et al. proposed that features used as input to MMs evaluated in the study may serve as feedback for skill acquisition (61). In a second study, Rhienmora et al. (91) developed automated feedback based on the magnitude and direction of forces applied on instruments in different segments of the task. The automated feedback generated in this study was optimized using data from expert surgeons performing the task. In addition, Rhienmora et al. asked expert surgeons to rate the acceptability of feedback generated from observations regarding forces applied on tools during tooth preparation for crown replacement (91). The resulting acceptability

scores, with an average of 4.15 on a Likert scale of 1 to 5, suggest high face and content validity of the feedback generated through OCASE-T. Finally, hand-engineered features that were differently distributed for expert and novice surgeons were proposed as feedback for skill acquisition (54). Evaluation of the effectiveness of feedback based on summary features for skill acquisition is limited and remains inadequate.

A single study proposed feedback based on detection of errors in performance (i.e., diagnosis of skill deficits) using OCASE-T. Islam et al. (75) used simple computer vision techniques such as object detection to identify errors in different tasks within the Fundamentals of Laparoscopy (FLS) curriculum. They used object detection algorithms to determine the number of missed attempts for the peg transfer task, assess deviation of a Penrose drain from its original position for an intracorporeal suturing task, and perform image processing such as smoothing and Hough transformation to assess imperfection in cutting a circle for a shape-cutting task (75). However, the effectiveness of feedback based on errors in performance detected using OCASE-T has not been ascertained.

The final approach for providing feedback based on OCASE-T involves targeted assessments for activity segments of interest. This approach provides feedback by localizing skill deficits to specific areas of a task. Ahmidi et al. (77) used DCC, which preserves geometry of motion, to discretize tool motion data, demonstrating an accuracy of 70–98% in classifying skill for individual gestures as expert, intermediate, or novice. Reiley et al. (76), by contrast, discretized tool motion data using K-means and attempted a similar three-class skill classification. They reported sensitivities of 75%, 59%, and 76% to classify gestures performed by expert, intermediate, and novice surgeons, respectively. Finally, Vedula et al. (85) used summary features computed using tool motion data to illustrate their utility to classify skill for tasks, maneuvers, and gestures. They reported similar area under the receiver operating characteristic curves (AUROCs) to classify skills into expert or novice using maneuver-level (0.78) and task-level (0.79) features, and a lower AUROC (0.7) using gesture-level features. The differences were not statistically significant. Furthermore, OCASE-T for segments within a task may be extrapolated to task-level skill assessment. Malpani et al. (81, 82) used summary features computed from tool motion data to train a SVM to determine which among a pair of maneuvers was better performed. The pairwise preferences were used to rank maneuvers, compute a percentile score for each maneuver, and use linear regression to map maneuver-level percentile scores to task-level OSATS-GRS; the authors reported a moderate (0.47) correlation between the two scores.

7. CONCLUSIONS

Assessment and feedback play a central role in training skillful and competent surgeons. Our survey of the literature indicates that technology is now available to enable methods for OCASE-T that can supplant traditional, subjective, resource-intensive training and assessment methods. However, the field is by no means mature. Although several promising methods for OCASE-T have been developed, their comparability, generalizability, application in real-world settings, validity against patient outcomes, and effectiveness for surgical skills training remain largely unexplored.

Additional research in this area can have a major impact on surgical training and maintenance of surgical skill. Successful development of OCASE-T technologies can potentially transform technical skill training through uniform objective assessment and individualized, targeted feedback for automated coaching. Furthermore, such automation would ensure efficient access to individualized and lifelong coaching to support surgeons' continuous improvement during both training and practice. Widespread availability of technology for OCASE-T could also minimize disparities in technical skill and, consequently, access to high-quality surgical care.

There are challenges that must be addressed for OCASE-T to achieve this vision. First, a large number of stakeholders with diverse interests are involved in surgical training and care. Research on OCASE-T should be carefully designed to align with the interests of these stakeholders and to demonstrate the value of OCASE-T. Second, funding and regulatory agencies must allocate resources to enable the development of integrated data capture systems and valid analytics, which are sufficiently mature to incorporate into training and certification. Third, OCASE-T should be adopted as an integral component of assessment during training and a resource to support continuous improvement throughout the duration of surgeons' professional practice.

Looking ahead, technology development for OCASE-T should be targeted toward high-value contexts, identified by surgical educators, where there are few barriers on the path from data collection to deployment within training curricula. For example, developing technology for OCASE-T for the FLS curriculum or robotics surgery, where data collection is easily accomplished; creating a robust competency assessment; and validating it in multi-institutional studies could have a considerable impact on graduate surgical education. However, such progress will be possible only if surgeons and engineers join forces to develop valid, relevant technology that will be readily adopted, and if funding agencies actively support such cross-disciplinary, translational research.

Acknowledgments

LITERATURE CITED

1. Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, et al. Size and distribution of the global volume of surgery in 2012. Bull World Health Organ. 2016; 94:F201–9.

2. Weiser TG, Regenbogen SE, Thompson KD, Haynes AB, Lipsitz SR, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. Lancet. 2008; 372:139–44. [PubMed: 18582931]

3. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med. 2013; 369:1434–42. [PubMed: 24106936]

4. Nathan M, Karamichalis JM, Liu H, del Nido P, Pigula R, et al. Intraoperative adverse events can be compensated by technical performance in neonates and infants after cardiac surgery: a prospective study. J Thorac Cardiovasc Surg. 2011; 142:1098–107. [PubMed: 21840545]

5. Nathan M, Karamichalis JM, Liu H, Emani S, Baird C, et al. Surgical technical performance scores are predictors of late mortality and unplanned reinterventions in infants after cardiac surgery. J Thorac Cardiovasc Surg. 2012; 144:1095–101. [PubMed: 22939862]

6. Shuhaiber J, Gauvreau K, Thiagarajan R, Bacha E, Mayer J, et al. Congenital heart surgeons technical proficiency affects neonatal hospital survival. J Thorac Cardiovasc Surg. 2012; 144:1119–24. [PubMed: 22421402]

7. Parsa CJ, Organ CH Jr, Barkan H. Changing patterns of resident operative experience from 1990 to 1997. Arch Surg. 2000; 135:570–75. [PubMed: 10807282]

8. Smith R. All changed, changed utterly. BMJ. 1998; 316:1917–18. [PubMed: 9641922]

9. Szasz P, Louridas M, Harris KA, Aggarwal R, Grantcharov TP. Assessing technical competence in surgical trainees: a systematic review. Ann Surg. 2015; 261:1046–55. [PubMed: 25119118]

10. Bhatti NI, Cummings CW. Viewpoint: competency in surgical residency training: defining and raising the bar. Acad Med. 2007; 82:569–73. [PubMed: 17525542]

11. Satava RM, Gallagher AG, Pellegrini CA. Surgical competence and surgical proficiency: definitions, taxonomy, and metrics. J Am Coll Surg. 2003; 196:933–37. [PubMed: 12788431]

12. Sharma B, Mishra A, Aggarwal R, Grantcharov TP. Non-technical skills assessment in surgery. Surg Oncol. 2011; 20:169–77. [PubMed: 21129950]

13. Yule S, Flin R, Paterson-Brown S, Maran N. Non-technical skills for surgeons in the operating room: a review of the literature. Surgery. 2006; 139:140–49. [PubMed: 16455321]

14. Reznick RK. Teaching and testing technical skills. Am J Surg. 1993; 165:358–61. [PubMed: 8447543]

15. Pradarelli JC, Campbell DA Jr, Dimick JB. Hospital credentialing and privileging of surgeons: a potential safety blind spot. JAMA. 2015; 313:1313–14. [PubMed: 25849177]

16. Roman H, Marpeau L, Hulsey TC. Surgeons experience and interaction effect in randomized controlled trials regarding new surgical procedures. Am J Obstet Gynecol. 2008; 199:108. [PubMed: 18456226]

17. Regenbogen SE, Greenberg CC, Studdert DM, Lipsitz SR, Zinner MJ, Gawande AA. Patterns of technical error among surgical malpractice claims: an analysis of strategies to prevent injury to surgical patients. Ann Surg. 2007; 246:705–11. [PubMed: 17968158]

18. Hull L, Arora S, Aggarwal R, Darzi A, Vincent C, Sevdalis N. The impact of nontechnical skills on technical performance in surgery: a systematic review. J Am Coll Surg. 2012; 214:214–30. [PubMed: 22200377]

19. Mishra A, Catchpole K, Dale T, McCulloch P. The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. Surg Endosc. 2007; 22:68–73. [PubMed: 17479324]

20. Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997; 84:273–78. [PubMed: 9052454]

21. Ghaderi I, Manji F, Park YS, Juul D, Ott M, et al. Technical skills assessment toolbox: a review using the unitary framework of validity. Ann Surg. 2015; 261:251–62. [PubMed: 24424150]

22. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. Am J Surg. 2011; 202:469–80.e6. [PubMed: 21798511]

23. Jelovsek JE, Kow N, Diwadkar GB. Tools for the direct observation and assessment of psychomotor skills in medical trainees: a systematic review. Med Educ. 2013; 47:650–73. [PubMed: 23746155]

24. Middleton RM, Baldwin MJ, Akhtar K, Alvand A, Rees JL. Which global rating scale? J Bone Joint Surg Am. 2016; 98:75–81. [PubMed: 26738906]

25. Anderson DD, Long S, Thomas GW, Putnam MD, Bechtold JE, Karam MD. Objective Structured Assessments of Technical Skills (OSATS) does not assess the quality of the surgical result effectively. Clin Orthop Relat Res. 2015; 474:874–81.

26. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. Acad Med. 1996; 71:1363–65. [PubMed: 9114900]

27. Rooney DM, Hungness ES, DaRosa DA, Pugh CM. Can skills coaches be used to assess resident performance in the skills laboratory? Surgery. 2012; 151:796–802. [PubMed: 22652120]

28. Lendvay TS, White L, Kowalewski T. Crowdsourcing to assess surgical skill. JAMA Surg. 2015; 150:1086–87. [PubMed: 26421369]

29. Ghani KR, Miller DC, Linsell S, Brachulis A, Lane B, et al. Measuring to improve: peer and crowdsourced assessments of technical skill with robot-assisted radical prostatectomy. Eur Urol. 2016; 69:547–50. [PubMed: 26755338]

30. Powers MK, Boonjindasup A, Pinsky M, Dorsey P, Maddox M, et al. Crowdsourcing assessment of surgeon dissection of renal artery and vein during robotic partial nephrectomy: a novel approach for quantitative assessment of surgical performance. J Endourol. 2015; 30:447–52. [PubMed: 26597352]

31. Rutegård M, Lagergren J, Rouvelas I, Lagergren P. Surgeon volume is a poor proxy for skill in esophageal cancer surgery. Ann Surg. 2009; 249:256–61. [PubMed: 19212179]

32. Sheikh F, Gray RJ, Ferrara J, Foster K, Chapital A. Disparity between actual case volume and the perceptions of case volume needed to train competent general surgeons. J Surg Ed. 2010; 67:371–75.

33. Snyder RA, Tarpley MJ, Tarpley JL, Davidson M, Brophy C, Dattilo JB. Teaching in the operating room: results of a national survey. J Surg Ed. 69:643–49.

34. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D, Fried GM. Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. J Am Coll Surg. 2004; 198:105–10. [PubMed: 14698317]

35. Porter ME. What is value in health care? N Engl J Med. 2010; 363:2477–81. [PubMed: 21142528]

36. Bell RH Jr. Why Johnny cannot operate. Surgery. 2009; 146:533–42. [PubMed: 19789010]

37. Chikwe J, de Souza AC, Pepper JR. No time to train the surgeons. BMJ. 2004; 328:418–19. [PubMed: 14976074]

38. Vedula, SS., Ishii, M., Hager, GD. Perspectives on surgical data science. Presented at Worksh. Surg. Data Sci.; June 20; Heidelberg, Ger.. 2016.

39. Bishop, CM. Pattern Recognition and Machine Learning. New York: Springer; 2009.

40. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning. New York: Springer; 2001.

41. Steyerberg, EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York: Springer;

42. Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. Surg Endosc. 2010; 25:356–66. [PubMed: 20607563]

43. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. Br J Surg. 2010; 97:972–87. [PubMed: 20632260]

44. Oropesa P, Sánchez-González P, Lamata P, Chmarra MK, Pagador JB, et al. Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. J Surg Res. 2011; 171:e81–95. [PubMed: 21924741]

45. Dosis, A., Bello, F., Rockall, T., Munz, Y., Moorthy, K., et al. Proceedings of the 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine. Piscataway, NJ: IEEE; 2003. ROVIMAS: a software package for assessing surgical skills using the da Vinci telemanipulator system; p. 326-29.

46. Aizuddin, M., Oshima, N., Midorikawa, R., Takanishi, A. Proceedings of the 1st IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics. Piscataway, NJ: IEEE; 2006. Development of sensor system for effective evaluation of surgical skill; p. 678-83.

47. Rosen, J. Surgeon-tool force/torque signatures—evaluation of surgical skills in minimally invasive surgery. In: Westwood, JD.Hoffman, HM.Robb, R., Stredney, D., editors. Medicine Meets Virtual Reality Convergence of Physical and Informational Technologies: Options for a New Era in Healthcare. Lansdale, PA: IOS; 1999. p. 290-96.

48. Hattori, M., Egi, H., Tokunaga, M., Suzuki, T., Ohdan, H., Kawahara, T. Proceedings of the 2012 ICME International Conference on Complex Medical Engineering. Piscataway, NJ: IEEE; 2012. The integrated deviation in the HUESAD (Hiroshima University endoscopic surgical assessment device) represents the surgeon's visual–spatial ability; p. 316-20.

49. Wang, C., Noh, Y., Ishii, H., Kikuta, G., Ebihara, K., et al. Proceedings of the 2011 IEEE International Conference on Robotics and Biomimetics. Piscataway, NJ: IEEE; 2011.

Development of a 3D simulation which can provide better understanding of trainees performance of the task using airway management training system WKA-1RII; p. 2635-40.

50. Ahmidi, N., Tao, L., Sefati, S., Gao, Y., Lea, C., et al. A dataset and benchmark for segmentation and recognition of gestures in robotic surgery. IEEE Trans Biomed Eng. 2017. https://doi.10.1109/TBME.2016.2647680

51. Lea C, Hager GD, Vidal R. An improved model for segmentation and recognition of fine-grained activities with application to surgical training tasks. IEEE Winter Conf Appl Comput Vis. 2015; 2015:1123–29.

52. Gao Y, Vedula SS, Lee GI, Lee MR, Khudanpur S, Hager GD. Int J Comput Assist Radiol Surg. 2016; 11:987–96. [PubMed: 27072835]

53. Lalys F, Jannin P. Surgical process modelling: a review. Int J CARS. 2014; 9:495–511.

54. Ahmidi N, Poddar P, Jones JD, Vedula SS, Ishii L, et al. Automated objective surgical skill assessment in the operating room from unstructured tool motion in septoplasty. Int J CARS. 2015; 10:981–91.

55. Richstone L, Schwartz MJ, Seideman C, Cadeddu J, Marshall S, Kavoussi LR. Eye metrics as an objective assessment of surgical skill. Ann Surg. 2010; 252:177–82. [PubMed: 20562602]

56. Ahmidi N, Hager GD, Ishii L, Fichtinger G, Gallia GL, Ishii M. Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery. Med Image Comput Comput Assist Interv. 2010; 13:295–302. [PubMed: 20879412]

57. Ahmidi N, Hager GD, Ishii L, Gallia GL, Ishii M. Robotic path planning for surgeon skill evaluation in minimally-invasive sinus surgery. Med Image Comput Comput Assist Interv. 2012; 15:471–78. [PubMed: 23285585]

58. Ahmidi N, Ishii M, Fichtinger G, Gallia GL, Hager GD. An objective and automated method for assessing surgical skill in endoscopic sinus surgery using eye-tracking and tool-motion data. Int Forum Allergy Rhinol. 2012; 2:507–15. [PubMed: 22696449]

59. Ahmidi N, Hager GD, Ishii M. Towards surgical skill assessment based on fractal patterns in minimally-invasive surgeries. Int J CARS. 2012; 7:185–200.

60. Rosen J, Solazzo M, Hannaford B, Sinanan M. Objective laparoscopic skills assessments of surgical residents using hidden Markov models based on haptic information and tool/tissue interactions. Stud Health Technol Inform. 2001; 81:417–23. [PubMed: 11317782]

61. Rosen J, Hannaford B, Richards CG, Sinanan MN. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE Trans Biomed Eng. 2001; 48:579–91. [PubMed: 11341532]

62. Rosen J, Brown JD, Chang L, Sinanan MN, Hannaford B. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. IEEE Trans Biomed Eng. 2006; 53:399–413. [PubMed: 16532766]

63. Chen J, Yeasin M, Sharma R. Visual modelling and evaluation of surgical skill. Pattern Anal Appl. 2003; 6:1–11.

64. Sharma, Y., Bettadapura, V., Plötz, T., Hammerld, N., Mellor, S., et al. Video based assessment of OSATS using sequential motion textures. Proceedings of the 5th Workshop on Modeling and Monitoring of Computer Assisted Interventions. 2014. https://smartech.gatech.edu/handle/1853/53651

65. Sharma, Y., Plötz, T., Hammerld, N., Mellor, S., McNaney, R., et al. Proceedings of the 11th IEEE International Symposium on Biomedical Imaging. Piscataway, NJ: IEEE; 2014. Automated surgical OSATS prediction from videos; p. 461-64.

66. Khan, A., Mellor, S., Berlin, E., Thompson, R., McNaney, R., et al. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York: ACM; 2015. Beyond activity recognition: skill assessment from accelerometer data; p. 1155-66.

67. Zia A, Sharma Y, Bettadapura V, Sarin EL, Clements MA, Essa I. Automated assessment of surgical skills using frequency analysis. Med Image Comput Comput Assist Interv. 2015; 11:430–38.

68. Speidel S, Zentek T, Sudra G, Gehrig T, Müller-Stich BP, et al. Recognition of surgical skills using hidden Markov models. Proc SPIE. 2009; 7261:25.

69. Zhang, Q., Li, B. Proceedings of the 2011 International ACM Workshop on Medical Multimedia Analysis and Retrieval. New York: ACM; 2011. Video-based motion expertise analysis in simulation-based surgical training using hierarchical Dirichlet process hidden Markov model; p. 19-24.

70. Horeman T, Rodrigues SP, Jansen FW, Dankelman J, van den Dobbelsteen JJ. Force parameters for skills assessment in laparoscopy. IEEE Trans Haptics. 2012; 5:312–22. [PubMed: 26964129]

71. Horeman T, Dankelman J, Jansen FW, van den Dobbelsteen JJ. Assessment of laparoscopic skills based on force and motion parameters. IEEE Trans Biomed Eng. 2014; 61:805–13. [PubMed: 24216633]

72. Kowalewski, TM. PhD thesis. Univ. Wash.; Seattle: 2012. Real-time quantitative assessment of surgical skill.

73. Lin Z, Uemura M, Zecca M, Sessa S, Ishii H, et al. Objective skill evaluation for laparoscopic training based on motion analysis. IEEE Trans Biomed Eng. 2013; 60:977–85. [PubMed: 23204271]

74. Zhang Q, Li B. Relative hidden Markov models for video-based evaluation of motion skills in surgical training. IEEE Trans Pattern Anal Mach Intell. 2015; 37:1206–18. [PubMed: 26357343]

75. Islam G, Kahol K, Li B, Smith M, Patel VL. Affordable, web-based surgical skill training and evaluation tool. J Biomed Inform. 2016; 59:102–14. [PubMed: 26556643]

76. Reiley CE, Hager GD. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. Med Image Comput Comput Assist Interv. 2009; 12:435–42. [PubMed: 20426017]

77. Ahmidi N, Gao Y, Bjar B, Vedula SS, Khudanpur S, et al. String motif–based description of tool motion for detecting skill and gestures in robotic surgery. Med Image Comput Comput Assist Interv. 2013; 16:26–33. [PubMed: 24505645]

78. Kumar R, Jog A, Malpani A, Vagvolgyi B, Yuh D, et al. Assessing system operation skills in robotic surgery trainees. Int J Med Robot. 2012; 8:118–24. [PubMed: 22114003]

79. Kumar R, Jog A, Vagvolgyi B, Nguyen H, Hager G, et al. Objective measures for longitudinal assessment of robotic surgery training. J Thorac Cardiovasc Surg. 2012; 143:528–34. [PubMed: 22172215]

80. Tao, L., Elhamifar, E., Khudanpur, S., Hager, GD., Vidal, R. Lecture Notes in Computer Science, vol 7330: Information Processing in Computer-Assisted Interventions. New York: Springer; 2012. Sparse hidden Markov models for surgical gesture classification and skill evaluation; p. 167-77.

81. Malpani, A., Vedula, SS., Chen, CCG., Hager, GD. Lecture Notes in Computer Science, vol 8498: Information Processing in Computer-Assisted Interventions. New York: Springer; 2014. Pairwise comparison-based objective score for automated skill assessment of segments in a surgical task; p. 138-47.

82. Malpani A, Vedula SS, Chen CCG, Hager GD. A study of crowdsourced segment-level surgical skill assessment using pairwise rankings. Int J CARS. 2015; 10:1435–47.

83. Gomez ED, Aggarwal R, McMahan W, Bark K, Kuchenbecker KJ. Objective assessment of robotic surgical skill using instrument contact vibrations. Surg Endosc. 2015; 30:1419–31. [PubMed: 26201410]

84. Fard, MJ. PhD thesis. Wayne State Univ.; Detroit, Mich.: 2016. Computational modeling approaches for task analysis in robotic-assisted surgery; p. 1449

85. Vedula SS, Malpani A, Ahmidi N, Khudanpur S, Hager G, Chen CCG. Task-level versus segment-level quantitative metrics for surgical skill assessment. J Surg Ed. 2016; 73:482–89.

86. Gomez ED, Aggarwal R, McMahan W, Bark K, Kuchenbecker KJ. Objective assessment of robotic surgical skill using instrument contact vibrations. Surg Endosc. 2016; 30:1419–31. [PubMed: 26201410]

87. Rafii-Tari, H., Payne, CJ., Liu, J., Riga, C., Bicknell, C., Yang, GZ. Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE; 2015. Towards automated surgical skill evaluation of endovascular catheterization tasks based on force and motion signatures; p. 1789-94.

88. Kramer, BD., Losey, DP., O'Malley, MK. Advances in Intelligent Systems and Computing, vol 428: Advances in Self-Organizing Maps and Learning Vector Quantification. New York: Springer;

2016. SOM and LVQ classification of endovascular surgeons using motion-based metrics; p. 227-37.

89. Sewell, C. PhD thesis. Stanford Univ.; Stanford, CA: 2007. Automatic performance evaluation in surgical simulation.

90. Rhienmora, P., Haddawy, P., Suebnukarn, S., Dailey, MN. Providing objective feedback on skill assessment in a dental surgical training simulator. In: Combi, C.Shahar, Y., Abu-Hanna, A., editors. Lecture Notes in Computer Science, vol 5651: Artificial Intelligence in Medicine. New York: Springer; 2009. p. 305-14.

91. Rhienmora P, Haddawy P, Suebnukarn S, Dailey MN. Intelligent dental training simulator with objective skill assessment and feedback. Artif Intell Med. 2011; 52:115–21. [PubMed: 21641781]

92. Liang H, Shi MY. Surgical skill evaluation model for virtual surgical training. Appl Mech Mater. 2010; 4041:812–19.

93. Hajshirmohammadi, I. PhD thesis. Simon Fraser Univ.; Burnaby, Can.: 2006. Using fuzzy set theory to objectively evaluate performance on minimally invasive surgical simulators.

94. Leong JJH, Nicolaou M, Atallah L, Mylonas GP, Darzi AW, Yang GZ. HMM assessment of quality of movement trajectory in laparoscopic surgery. Med Image Comput Comput Assist Interv. 2006; 9:752–59. [PubMed: 17354958]

95. Megali G, Sinigaglia S, Tonet O, Dario P. Modelling and evaluation of surgical performance using hidden Markov models. IEEE Trans Biomed Eng. 2006; 53:1911–19. [PubMed: 17019854]

96. Loukas C, Georgiou E. Multivariate autoregressive modeling of hand kinematics for laparoscopic skills assessment of surgical trainees. IEEE Trans Biomed Eng. 2011; 58:3289–97. [PubMed: 21908250]

97. Jog, A., Itkowitz, B., Liu, M., DiMaio, S., Hager, G., et al. Proceedings of the 2011 IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE; 2011. Towards integrating task information in skills assessment for dexterous tasks in surgery and simulation; p. 5273-78.

98. Arora S, Sevdalis N, Nestel D, Woloshynowych M, Darzi A, Kneebone R. The impact of stress on surgical performance: a systematic review of the literature. Surgery. 2010; 147:318–30. [PubMed: 20004924]

99. Lee G, Lee T, Dexter D, Klein R, Park A. Methodological infrastructure in surgical ergonomics: a review of tasks, models, and measurement systems. Surg Innov. 2007; 14:153–67. [PubMed: 17928614]

100. Ericsson, KA.Charness, N.Feltovich, PJ., Hoffman, RR., editors. The Cambridge Handbook of Expertise and Expert Performance. New York: Cambridge Univ. Press; 2006.

101. Crochet P, Aggarwal R, Dubb SS, Ziprin P, Rajaretnam N, et al. Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. Ann Surg. 2011; 253:1216–22. [PubMed: 21516035]

102. Bonrath EM, Dedy NJ, Gordon LE, Grantcharov TP. Comprehensive surgical coaching enhances surgical skill in the operating room: a randomized controlled trial. Ann Surg. 2015; 262:205–12. [PubMed: 25822691]

103. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol. 1995; 50:741–49.

104. Anderson F, Birch DW, Boulanger P, Bischof WF. Sensor fusion for laparoscopic surgery skill acquisition. Comput Aided Surg. 2012; 17:269–83. [PubMed: 23098188]

105. Snaineh, STA., Seales, B. Proceedings of the IEEE International Symposium on Signal Processing and Information Technology. Piscataway, NJ: IEEE; 2015. Minimally invasive surgery skills assessment using multiple synchronized sensors; p. 314-19.

106. Sonnadara R, Rittenhouse N, Khan A, Mihailidis A, Drozdzal G, et al. A novel multimodal platform for assessing surgical technical skills. Am J Surg. 2012; 203:32–36. [PubMed: 22075120]

107. Ruda K, Beekman D, White LW, Lendvay TS, Kowalewski TM. SurgTrak—a universal platform for quantitative surgical data capture. J Med Devices. 2013; 7:030923.

108. Ostler, D., Kranzfelder, M., Stauder, R., Wilhelm, D., Feussner, H., Schneider, A. Proceedings of the 17th International Conference on E-Health Networking, Application Services. Piscataway, NJ: IEEE; 2015. A centralized data acquisition framework for operating theatres; p. 1-5.

109. Zhao, X. PhD thesis. Va. Commonw. Univ.; Richmond: 2005. Acquisition, storage and reconstruction of multidimensional surgical information in a digital operation room environment.

110. Chmarra MK, Grimbergen CA, Dankelman J. Systems for tracking minimally invasive surgical instruments. Minim Invasive Ther Allied Technol. 2007; 16:328–40. [PubMed: 17943607]

111. DiMaio, S., Hasser, C. The da Vinci research interface. MIDAS J. 2008. http://hdl.handle.net/10380/1464

112. Trejos AL, Patel RV, Naish MD, Lyle AC, Schlachta CM. A sensorized instrument for skills assessment and training in minimally invasive surgery. J Med Devices. 2009; 3:041002.

113. Rutherford DN, D'Angelo ALD, Law KE, Pugh CM. Advanced engineering technology for measuring performance. Surg Clin N Am. 2015; 95:813–26. [PubMed: 26210973]

114. Kirby GSJ, Guyver P, Strickland L, Alvand A, Yang GZ, et al. Assessing arthroscopic skills using wireless elbow-worn motion sensors. J Bone Joint Surg Am. 2015; 97:1119–27. [PubMed: 26135079]

115. King RC, Atallah L, Lo BPL, Yang GZ. Development of a wireless sensor glove for surgical skills assessment. IEEE Trans Inf Technol Biomed. 2009; 13:673–79. [PubMed: 19726263]

116. Saggio, G., Santosuosso, GL., Cavallo, P., Pinto, CA., Petrella, M., et al. Proceedings of the IEEE International Workshop on Medical Measurements. Piscataway, NJ: IEEE; 2011. Gesture recognition and classification for surgical skill assessment; p. 662-66.

117. Kodama, H., Tercero, C., Ooe, K., Shi, C., Ikeda, S., et al. Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ: IEEE; 2012. 2-D optical encoding of catheter motion and cyber-physical system for technical skills measurement and quantitative evaluation in endovascular surgery; p. 3565-70.

118. Conditt MA, Noble PC, Thompson MT, Ismaily SK, Moy GJ, Mathis KB. A computerized bioskills system for surgical skills training in total knee replacement. Proc Inst Mech Eng H. 2007; 221:61–69. [PubMed: 17315769]

119. Lu J, Kowalewski TM. Flexible, stretchable skin sensors for two-dimensional position tracking in medical simulators. J Med Devices. 2015; 9:020927.

120. Atkins MS, Tien G, Khan RSA, Meneghetti A, Zheng B. What do surgeons see: capturing and synchronizing eye gaze for surgery applications. Surg Innov. 2012; 20:241–48. [PubMed: 22696024]

121. Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang GZ, Darzi A. Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. Surg Endosc. 2015; 29:405–13. [PubMed: 25125094]

122. Pandya A, Reisner LA, King B, Lucas N, Composto A, et al. A review of camera viewpoint automation in robotic and laparoscopic surgery. Robotics. 2014; 3:310–29.

123. Morris MC, Frodl T, D'Souza A, Fagan AJ, Ridgway PF. Assessment of competence in surgical skills using functional magnetic resonance imaging: a feasibility study. J Surg Ed. 2015; 72:198–204.

124. Evans AW, Leeson RMA, Petrie A. Correlation between a patient-centred outcome score and surgical skill in oral surgery. Br J Oral Maxillofac Surg. 2005; 43:505–10. [PubMed: 15908083]

125. Ahmidi, N. PhD thesis. Johns Hopkins Univ.; Baltimore, Md.: 2015. Activity detection and skill assessment for dexterous motions in robotic and minimally-invasive surgery.

126. Bergland GD. A guided tour of the fast Fourier transform. IEEE Spectrum. 1969; 6:41–52.

127. Zhang D, Lu G. Review of shape representation and description techniques. Pattern Recognit. 2004; 37:1–19.

128. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B. Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE; 2008. Learning realistic human actions from movies; p. 1-8.

129. Ershad, M., Koesters, Z., Rege, R., Majewicz, A. Meaningful assessment of surgical expertise: semantic labeling with data and crowds. In: Ourselin, S.Joskowicz, L.Sabuncu, MR.Unal, G.,

Wells, W., editors. Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer; 2016. p. 508-15.

130. Sant'Anna, A., Wickström, N. Proceedings of the 4th International Congress on Image and Signal Processing. Piscataway, NJ: IEEE; 2011. Symbolization of time-series: an evaluation of SAX, Persist, and ACA; p. 2223-28.

131. Shafiei, SB., Guru, KA., Esfahani, ET. Proceedings of the 2015 ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference: 39th Mechanisms and Robotics Conference. Boston: ASME; 2015. Using two-third power law for segmentation of hand movement in robotic assisted surgery; p. V05CT08A014

132. Bettadapura, V., Schindler, G., Ploetz, T., Essa, I. Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE; 2013. Augmenting bag-of-words: data-driven discovery of temporal and structural information for activity recognition; p. 2619-26.

133. Jelinek, F. Statistical Methods for Speech Recognition. Cambridge, MA: MIT Press; 2001.

134. Porte MC, Xeroulis G, Reznick RK, Dubrowski A. Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. Am J Surg. 2007; 193:105–10. [PubMed: 17188099]

135. Bjerrum F, Maagaard M, Sorensen JL, Larsen CR, Ringsted C, et al. Effect of instructor feedback on skills retention after laparoscopic simulator training: follow-up of a randomized trial. J Surg Ed. 2015; 72:53–60.

SUMMARY POINTS

1. The availability of data and the need for efficient surgical training provide a unique opportunity for OCASE-T to augment surgical education.

2. Major research goals for OCASE-T include developing technology for data capture in the surgical environment; techniques for transforming the data; and algorithms to assess skill and enable automated coaching/feedback through diagnosis of skill deficits, targeted assessments, or both.

3. Algorithms for OCASE-T have been developed mostly for benchtop and virtual reality simulation settings; research in the operating room has been limited to only two studies.

4. Existing methods for OCASE-T have produced promising results, but the approaches are highly varied. Establishing a consensus on the best approaches to OCASE-T will require additional replication and validation in large multicenter studies.

5. Algorithms for OCASE-T have focused only on technical skill and not on competency. The effectiveness of feedback based on OCASE-T for skill acquisition remains to be ascertained.

FUTURE ISSUES

1. Algorithms for OCASE-T should focus on high-stakes assessment, such as technical competency.

2. OCASE-T should emphasize assessment of learning curves in the training laboratory and in the operating room for integration into graduate surgical education.

3. Large, curated, open access data sets must be compiled for rapid development and comparative evaluation of algorithms for OCASE-T.

4. Systematic assessment of validity should be built into the development of algorithms for OCASE-T, potentially using consistent evaluation measures.
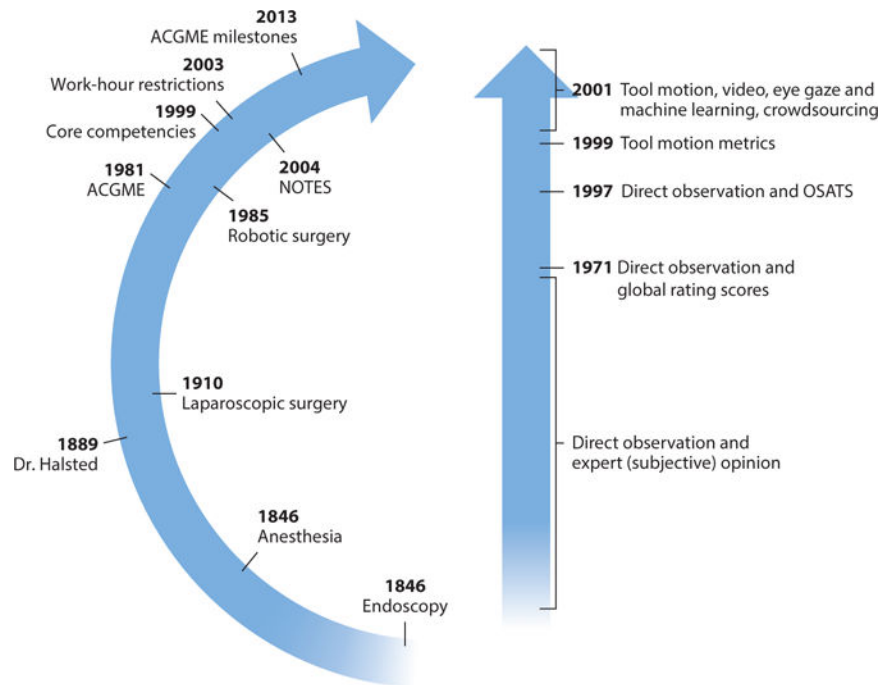
Figure 1.
Timelines showing major advances in surgical technique, education, and technical skill assessment. Major innovations in surgical technique are illustrated on the inside of the arc on the left. Landmarks in graduate surgical education in the United States are shown on the outside of the arc. Abbreviations: ACGME, Accreditation Council for Graduate Medical Education; NOTES, natural orifice transluminal endoscopic surgery; OSATS, objective structured assessment of technical skills.
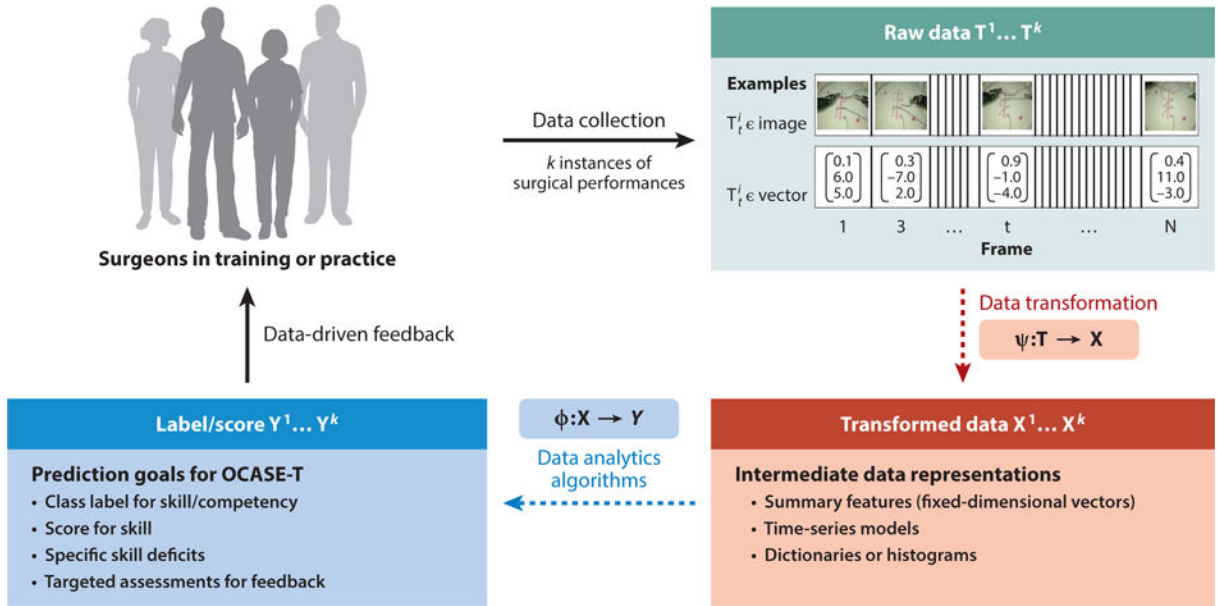
Figure 2.
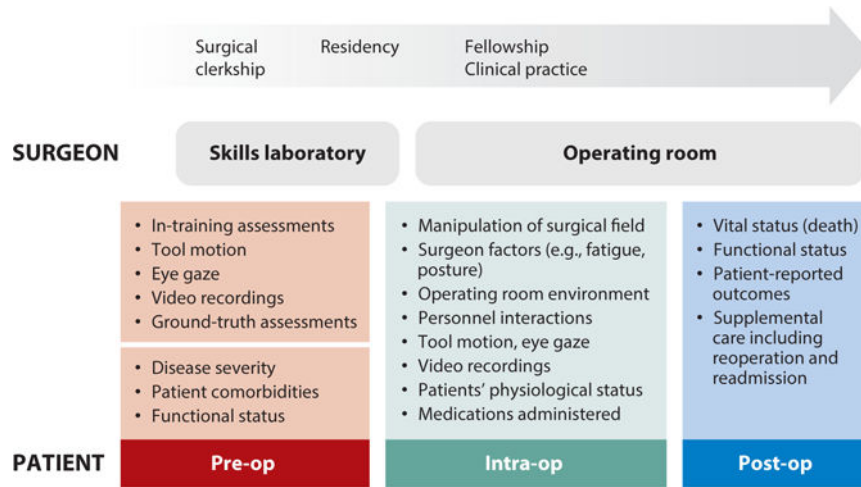Research questions for objective computer-aided skill evaluation (OCASE-T).

Figure 3.
Data for objective computer-aided skill evaluation (OCASE-T) to support continuous improvement in surgical skill and patient care.

Table 1

Studies on algorithms for objective computer-aided technical skill evaluation (OCASE-T)

Surgical site	Surgical technique	Reference(s)	Data sources	Surgical tasks studied	Study sample size
Operating room	Open (freehand) surgery	(54)	Electromagnetic sensors attached to tools	Mucosal flap elevation task in nasal septoplasty	Unclear
	Laparoscopic surgery	(55)	Head-mounted eye tracker	Colon mobilization task in nephrectomy	11
Cadaver	Endoscopic surgery	(56–59)	Electromagnetic sensors attached to instruments	Reaching anatomical targets in paranasal sinuses	11–20
Pig	Laparoscopic surgery	(60, 61)	Force/torque sensors attached to instruments	Cholecystectomy, Nissen fundoplication	8–10
	Robotic surgery	(62)	Direct capture from robot	Intracorporeal knot-tying	30
Benchtop simulation	Open (freehand) surgery	(46, 63–67)	Video images, colored markers on gloves, accelerometers attached to instruments	Suturing, knot-tying	4–20
	Laparoscopic surgery	(68–75)	Video images, optical tracking of instruments, force sensors on platform holding benchtop models, sensors attached to instruments and placed on surgeon	Suturing, knot-tying, needle passing, peg transfer, intracorporeal suturing, shape cutting	4–52
	Robotic surgery	(76–85)	Direct capture from robot, accelerometers attached to instruments	Suturing, knot-tying, needle passing, peg transfer, intracorporeal suturing, shape cutting, dissection	8–18
	Endovascular surgery	(87, 88)	Motion and force sensors on platform holding model; magnetic tracker at tool tip	Aortic arch cannulation and other unspecified tasks	15–23
Virtual reality simulation	Open (freehand) surgery	(89–91)	Virtual reality: known environment	Mastoidectomy, tooth preparation for crown replacement	10–16
	Laparoscopic surgery	(69, 92–96)	Virtual reality: known environment	Needle passing, reaching objects, knot-tying, peg transfer	6–28
	Robotic surgery	(97)	Virtual reality: known environment	Ring walk and other unspecified tasks	17

Table 2

Types of raw data captured in different surgical environments for objective computer-aided technical skill evaluation (OCASE-T)

Type of raw data	Benchtop simulation	Virtual reality simulation	Cadaver and animal models	Operating room
Surgical tools or surgeons' hands using sensors	(66, 68, 70–72, 76–85, 88)	(89–97)	(56–62)	(54)
Surgical tools or surgeons' hands in video images	(63)	No data	No data	No data
Surgeons: eye gaze, stress, fatigue, posture, movement	(73)	(55)	(56, 58)	(55)
Surgical field: extent of tissue manipulation	(46, 87)	No data	No data	No data
Surgical field: video images	(64, 65, 67, 69, 74, 75)	No data	No data	No data
Patients: disease severity, co-morbidities, anatomical complexities, physiological status, medications, outcomes of surgical intervention	Not applicable	Not applicable	Not applicable	No data
Operating room: environmental factors, other personnel	Not applicable	Not applicable	No data	No data

Table 3

Data representations and algorithms for objective computer-aided technical skill evaluation (OCASE-T)

Raw data	Summary features			Time series			Dictionaries or histogram-based	
	Generic motion efficiency	Signal processing	Hand-engineered	Continuous values	Discrete: preserving geometry of motion	Discrete: ignoring geometry of motion	Dictionaries or histograms	Features from dictionaries
Tool motion, forces, vibration, acceleration	LinReg (86); LogReg (84, 85); NNC (84); SVM (78, 79, 81, 82, 84, 97); FM (93)	DFA (46); SVM (54, 96); LDA (70, 71); LogReg (83); NNets (88)	SVM (54)	HMM (57, 59, 60, 80, 87, 92, 94, 96); MM (58, 61)	SVM (54, 57)	MM (62); HMM (56, 68, 72, 76, 95)	nhMI (59), CSM/SVM (54, 57, 77), PLA/SVM (66)	No data
Tool motion relative to anatomy	No data	No data	LogReg (89)	HMM (89)	No data	HMM (90,91)	No data	No data
Eye gaze	No data	No data	No data	NNets (55); LDA (55)	No data	HMM (56)	No data	No data
Surgeon's posture or movement	No data	LDA (73)	No data	No data	No data	No data	No data	No data
Video images	LinReg (75)	No data	No data	HMM (63)	No data	No data	HMM (69, 74); NNC (64, 67)	NNC (64, 65, 67)

Abbreviations: CSM, common substring model; DFA, discriminant function analysis; FM, fuzzy models; HMM, hidden Markov models; LDA, linear discriminant analysis; LinReg, linear regression; LogReg, logistic regression; MM, Markov models; nHMI, normalized histogram of motion increments; NNC, nearest-neighbor classification; NNets, neural networks; SVM, support vector machine.