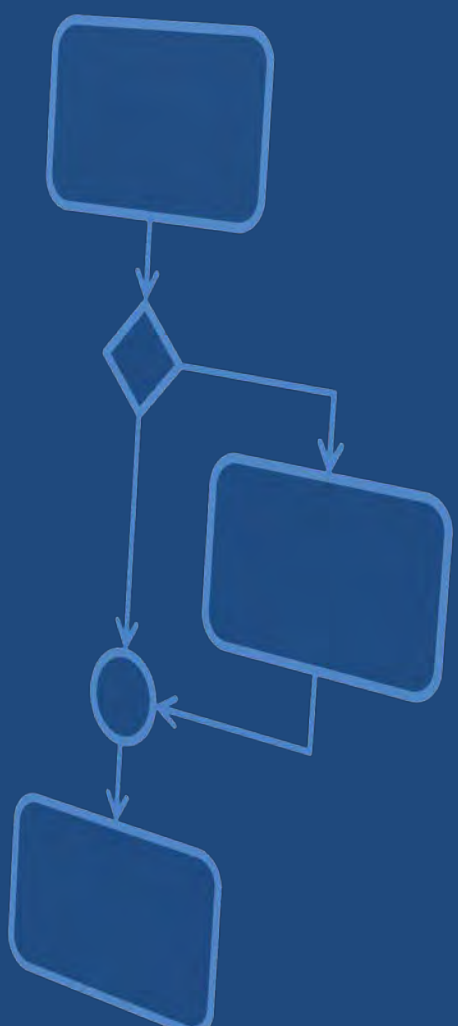
Surgical Process Modeling

- Theory, Methods, and Applications -



Thomas Neumuth

Für meine Familie.

Thomas Neumuth is senior researcher at the Innovation Center Computer Assisted Surgery (ICCAS) at the Universität Leipzig. His research topics include the modeling of surgical processes, workflow management systems, sensor system engineering for situation-dependent information systems in the operating room, operating room infrastructure, and surgical ontologies. After completing an apprenticeship in Business and Administration at Siemens AG, Germany (1997), he studied Electrical Engineering Management (2001) and Automation Engineering (2004) at the University of Applied Sciences Leipzig. He received his PhD in medical engineering (2009) and his habilitation in medical informatics (2012) from the Universität Leipzig.

ISBN 978-3-00-038630-5

Foreword

Dear Readers,

Surgery is a challenging, responsible, and complex task that requires considerable experience and skill. The surgical process involves highly complex procedures and depends on a multitude of factors that entail the surgeon's awareness and attention to patient-specific abnormalities in anatomy and pathology, or to technical resources. With respect to the multifaceted and highly variable processes of surgical interventions, a modeling and behavior observation strategy is necessary to account for this complexity. Until today, a concise form of report that is able to reproduce surgical process evolution in a detailed and accurate way is missing. This lack of explicitly represented knowledge constitutes a “blank spot” on the map of patient treatment in hospitals and prevents a straight-forward development of surgical techniques and technical resources, such as information and communication technology in the operating room for an improved patient care.

This book has the objective to fill the mentioned “blank spot” by providing methods and example applications to describe the sensitive area between the first cut and the last suture at the patient. The articles that are part of this book result from my research at the Innovation Center Computer Assisted Surgery (ICCAS) at the Universität Leipzig between 2005 and 2011 and represent my habilitation treatise. This work includes the following articles:

Neumuth T, Jannin P, Strauß G, Meixensberger J, Burgert O. Validation of knowledge acquisition for surgical process models. Journal of the American Medical Informatics Association. 2009; 16(1): 72-80.

Neumuth T, Trantakis C, Riffaud L, Strauss G, Meixensberger J, Burgert O. Assessment of technical needs for surgical equipment by surgical process models. Minimally Invasive Therapy and Allied Technologies. 2009; 18(6):841-849.

Neumuth T, Kaschek B, Neumuth D, Ceschia M, Meixensberger J, Strauss G, Burgert O. An observation support system with an adaptive ontology-driven user interface for the modeling of complex behaviors during surgical interventions. Behavior Research Methods. 2010; 42:1049-58.

Neumuth T, Wiedemann R, Foja C, Meier P, Neumuth D, Wiedemann P. Identification of surgeon-individual treatment profiles to support the provision of an optimum treatment service for cataract patients. Journal of Ocular Biology Diseases and Informatics. 2010; 3(2):73-83.

Neumuth T, Jannin P, Schlomberg J, Meixensberger J, Wiedemann P, Burgert O. Analysis of surgical intervention populations using generic surgical process models. International Journal of Computer Assisted Radiology and Surgery. 2011; 6(1):59-71.

Neumuth T, Krauss A, Meixensberger J, Muensterer O. Impact quantification of the daVinci telemanipulator system on the surgical workflow using resource impact profiles. International Journal of Medical Robotics. 2011; 7(2):156-64.

Neumuth D, Loebe F, Herre H, Neumuth T. Modeling Surgical Processes: A four-level translational approach. Artificial Intelligence in Medicine. 2011; 51(3):147-161.

Neumuth T, Loebe F, Jannin P. Similarity metrics for surgical process models. Artificial Intelligence in Medicine. 2012; 54(1):15-27.

Neumuth T, Meißner C. Online recognition of surgical instruments by information fusion. International Journal of Computer Assisted Radiology and Surgery. 2012; 7(2):297-204.

Neumuth T, Liebmann P, Wiedemann P, Meixensberger J. Surgical workflow management schemata for cataract procedures: Process model-based design and validation of workflow schemata. Methods of Information in Medicine. 2012; 51(5):371-382.

The reader might notice that the articles in this work do not follow a time-based sequence, but rather a logical sequence from setting the theory to its clinical application. However, this structure was chosen to facilitate the access to the research topic. Additionally, the 'level' of mathematical formalization was oriented at the audience of the respective journals where the single research topics were published. Therefore the work contains different 'levels' of formalization.

Of course, such endeavor cannot be undertaken without the support of many people that worked alongside me at the ICCAS. These people assisted me in bridging interdisciplinary gaps between clinicians and technicians, participated in many scientific discussions, or provided resources for performing the research. Thus, I hereby would like to extend my gratitude to Jürgen Meixensberger, Pierre Jannin, Frank Loebe, Heinrich Herre, and Oliver Burgert. Additionally, many people accompanied the research over the years and helped perform the studies. I am especially grateful to Marcello Ceschia, Michael Czygan, Caroline Elzner, Bernadett Kaschek, Maik Müller, Sandra Schumann, and Michael Thiele.

Finally, I especially thank my wife Dayana for her continuing support and encouragement over the many years.

Most of this research was funded by the German Federal Ministry for Education and Research (BMBF) within the scope of the program "Unternehmen Region" and parts were funded by the European Regional Development Fund (ERDF) and the state of Saxony.

If you have questions about this work, please do not hesitate to contact me.

Leipzig, December 2012
Thomas Neumuth (thomas@neumuth.de)

Thomas Neumuth

Surgical Process Modeling
- Theory, Methods, and Applications –

(Habilitation thesis)

Table of Contents

1 Introduction to the topic

1.1 Subject and motivation

Surgical work is highly influenced by a number of challenges originating from clinical, social, practical, financial, or qualitative aspects. Focal points, such as the growing number of patients [Deutsche Krankenhaus Verlagsgesellschaft GmbH 2009], the perpetually augmenting complexity of surgical interventions [Stitzenberg and Sheldon 2005], and the increasing need to justify surgical interventions [McPherson and Bunker 2007; Sangha et al. 2002], as opposed to decreasing staff expenditures [Schuhmann 2008; Offermanns 2002] and lump compensations [Alberty 2004], less time exposure per patient [Barmer GEK 2010], ever growing requirements concerning quality management [Thüsing 2005; Weiler et al. 2003] and shortened means of acquisition and/or replacements of surgical equipment [Jakisch 2007].

At the same time, the hospital as service contractor has to face the growing pressure exerted by the public: to achieve more treatments of success with less time expenditure, to increase the patient satisfaction factor and patient safety, while, at the same time, completely eliminating medical malpractices [Kreyher 2001].

With this in mind, the operating room (OR) in particular, as the most cost-intensive unit of patient treatment [Archer and Macario 2006; Geldner et al. 2002], surely is a toehold when it comes to examine processes within the hospital system. The aim of this examination is to render an amelioration of these processes possible. This consideration, however, presupposes a comprehensive and thorough understanding of the present processes. Furthermore, it requires innovative methods for process modeling and optimization. To reach this aim, the hospitals have to provide and implement an appropriate general framework for the surgeons.

The processes of the perioperative phase provide ample scope for improvement measures. Due to some limitations concerning technical, ergonomic, and process-related deficiencies, the turnover rate of ORs might be improved [Friedman et al. 2006; Dexter et al. 2003a; Dexter et al. 2003b]. This fact influences the costs of the OR as functional unit. The reasons for this are manifold. The existing information and communication technology (ICT) of the hospitals presently supports the clinical operational sequence only very inadequately [Sandberg et al. 2003; Sunyaev et al. 2006; Mauro et al. 2011; Documet et al. 2010]. Breaches in the workflow and media disruptions are the result. Information that is already available in digital format is not used any further, for instance, to support systems for the management of the ORs. In addition, patients, material, and information should be at the right place at the right time.

Aside from the points mentioned so far, the accelerated pace of developing new information technology procedures in surgery re-creates the task of the surgeon itself: from an as yet rather mechanically oriented and performing occupation to a rather monitoring and steering task [Stahl et al. 2005; Baumgart et al. 2010]. The surgeon is confronted with an ever growing multitude of information, which needs to be inspected, interpreted, selected (or discarded), and applied to the surgical situation at hand. For this, methods for a consistent and process-oriented ICT-support of the surgeon is needed, which, again, is based on a comprehensive describability of processes.

The launch of a technological innovation or modification, however, is only accepted by its users, when it adds quantitative benefits to surgical efficiency or leads to an improved quality of treatment [Sackett et al. 1996; Wente et al. 2003]. As a consequence, the measurability of improvements within the course of treatment is a requirement for the success of new technical systems, because decisions concerning the development and implementation of new technology should solely be based on evidence. This, however, is hampered by lacking information and missing sources of knowledge concerning surgical *modi operandi.* Thus, the starting points for technological improvements and developments to be examined and elaborated are: the transparency of surgical processes know-how, the availability of information concerning these processes, and an insight into intraoperative processes.

1.2 General problem

The present standard method for process modeling is either based on the process modelers' experience and/or on the results of interviews with domain experts or clients [Scheer 1997; Scheer 1999]. Predominantly, the principle of refinement is used where, starting from the top abstraction level, more and more details are being modeled. This approach is therefore also called the *top-down modeling approach* [Gadatsch 2002; Rosenkranz 2005].

However, this method has been stretched to its limits due to the fact that surgical procedures are characterized by a very high variability. In addition, this strategy provides, for some use cases, only a very insufficient resolution. Furthermore, top-down modeling has various drawbacks:

Firstly, the approach is cost- and time-consuming. To reduce the expenses, one possibility is the focusing on a small number of instances of processes. However, this strategy does not account sufficiently for the high variability.

Secondly, top-down models can be biased because the subjective perception of observers can be deficient. Thus, the ensuing models are insufficiently quantifiable due to their subjective character. No resilient, empirically observable assertions can be made concerning the frequency of process variants, the duration of single surgical work steps or the relevant intervention phases. Conversely, this is especially necessary for the evaluation of the employment of surgical assist systems (SAS).

Next, if the processes are highly dynamic and thus subjected to frequent modifications, the ensuing models are obsolete after short spans of time and have to be updated very frequently, causing renewed expenditures of cost and time.

And, last but not least, process models are mostly being modeled from the point of view of either business processes or workflow systems. This leads to an exclusion of relevant aspects of the model currently not in use, such as the participants of a process, influencing variables (such as the prompting of rules or principles, or quality aspects), localization, and relevant resources. The resulting process models are mostly used for documentation purposes or in workflow engines.

1.3 Relevance of the approach and objectives

Against the backdrop of such inadequate, inflexible, or expensive possibilities for the modeling of surgical processes, the relevance of the presented work emerges. The availability of ICT methods for description, acquisition, and abstraction is a substantial prerequisite for modeling surgical processes and for their convenience in clinical as well as technical use. This cannot be accomplished by existing, conventional methods, because of the lack of describability and measurability to provide concrete results. Furthermore, no methods for the generalization of previously acquired process models are in existence at present, neither are the possibilities to employ the attained models in information technology for the support of the surgical tasks and the surgeon. In view of these facts, the development of innovative and revised methods for the definition, compilation, and documentation of surgical interventions is needed. This would result in an improved possibility to document and evaluate surgical work.

The availability of such methods would lead to the possibility to identify objective topics that influence the course of the surgical intervention, such as to assess the significance of newly introduced technologies or surgical assist systems. This could, in turn, lead to new comprehensions concerning the development and improvement of technical systems. As currently no appropriate approaches for the modeling of such process models are available, these need to be developed. Therefore, the goal of the presented work is to provide methods for the systematic development, implementation, evaluation, and validation of the individual methods for surgical process modeling. Figure 1.3.1 gives an overview of the chapters of this work.

The major objectives are:

(1) the development of ICT-based methods for the description, data acquisition, abstraction, and utilization of surgical process models and
(2) the evaluation of the developed methods for surgical process models in clinical applications.

Secondary objectives of this work are the comprehensive applicability of the methods and its use-case based evaluation or validation.

(3) The methods developed in the course of this work should be comprehensively applicable. This is to be demonstrated with the help of studies concerning different use cases for various intervention types from different surgical disciplines.
(4) All hypotheses and milestones will be supported by corresponding scientific evaluation and valuation methods. Accordingly, the ICT methods have been rather more strongly formalized, while the aspects of greater importance for surgeons have been formalized to a lesser extent.

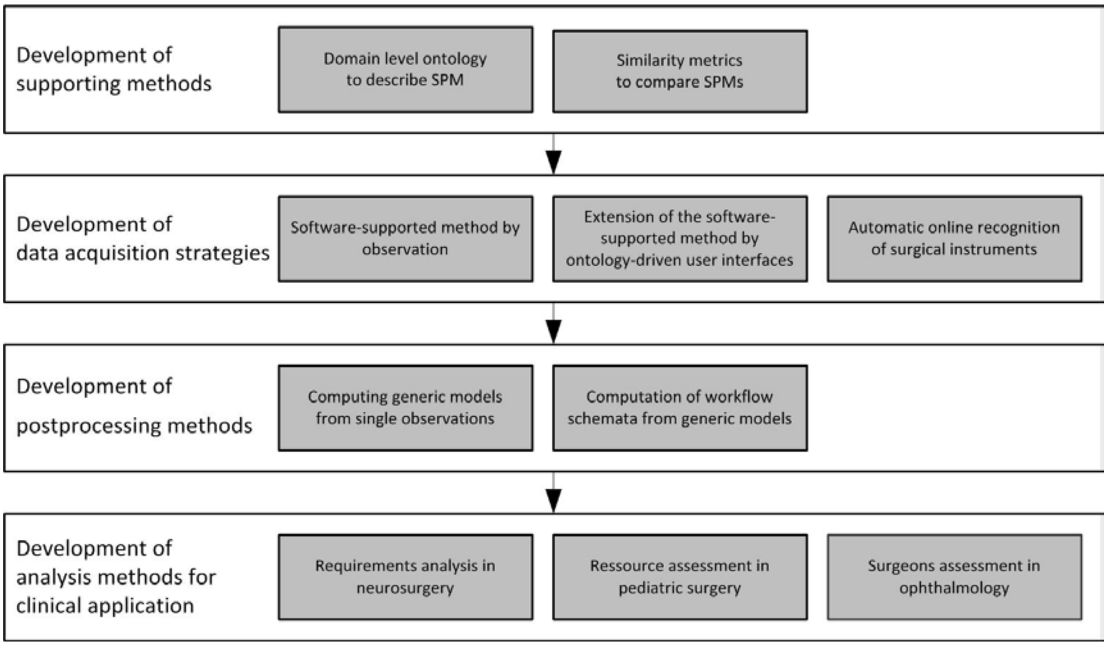


Figure 1.3.1: General overview of the main chapters.

This work has different target audiences. On the one hand, surgeons are to be provided with validated and evaluated instruments for the modeling of surgical processes, and the ensuing advantages, such as the evaluation of the advantageousness of new surgical assist systems or of alternative surgical strategies considering process-oriented aspects. On the other hand, medical engineers and medical technology enterprises will benefit from this work immediately. By applying the proposed methods, requirements analyses for the development of new surgical assist systems can be compiled and supported, for instance, to enable an objective assessment of the utilization benefit of these systems after the completion of their development.

1.4 Hypotheses and milestones

The following hypotheses and milestones are basic for this work:

Hypothesis 1: Supporting methods for surgical process modeling

Supporting methods are needed for the modeling of surgical processes. An appropriate formal and ontological basis for the modeling and exchange of information concerning surgical processes is necessary. In addition, validated metrics for the assessment of the quality of the data acquired with the help of observers are desirable.

Milestone 1a: Development of a process ontology for surgical processes

The milestone aims at the development of a domain-level ontology for the support of the modeling and the exchange of surgical process models (SPMs). This domain-level ontology should be represented by a generic framework. Until now, there is no generic framework that is well adjusted to the field of surgical processes and which has the possibility to integrate significant levels of abstraction into one single and comprehensive system. However, the main target group to use this framework has mostly medical or engineering background, and is thus typically unfamiliar with logical formalisms. Therefore, tools based on ontologically supported semantics to transform natural language specifications of processes into mathematical models should be made available by this work.

Milestone 1b: Design of similarity metrics for surgical process models

The major goal of the milestone is the development of a set of metrics. This metrics set can be applied to assess the performance of observers of surgical processes. Until today, such metrics are not available.

Hypothesis 2: Data acquisition strategies for surgical process modeling

Different strategies can be used for the bottom-up modeling of surgical process models. New methods have to be devised which avoid the disadvantages of top-down modeling approaches.

Milestone 2a: Observer-based data acquisition with observation support software

Observer-based data acquisition for surgical process models can be supported by appropriate software tools. This observation support software allows for an accurate acquisition of complex surgical process models.

Milestone 2b: Observer-based data acquisition with adaptive user interfaces

The observer-based data acquisition of surgical process models can be further supported by extending the observation support software with adaptive user interfaces. This will result in an increased accuracy of the modeling result. The goal of this milestone is the design and validation of a suchlike observation support system with an adaptive user interface.

Milestone 2c: Sensor-based data acquisition

The milestone aims at the conception of a sensor-based data acquisition strategy for SPMs. As it seems recognizable that not one single sensor-based strategy for data acquisition has the ability to acquire complete and exhaustive surgical process models, an architecture will be devised. By means of this architecture, the integration of different sensor signals on various levels of abstraction will be rendered possible.

Hypothesis 3: Model generalization and surgical workflow management

The analysis of surgical process models obtained by bottom-up modeling needs to be based on a sample of processes to draw conclusions with increased validity. This requires the computation of a generalized model.

Milestone 3a: Computation of generic surgical process models

The objective of this milestone is the preparation of an approach for the construction and computation of a generic model on the basis of single process models. This approach needs to be clinically applicable and assessable.

Milestone 3b: Process model-based generation of workflow schemata

The milestone investigates the application of generic surgical process models to generate workflow schemata for workflow management in the OR. Special attention is paid on how many single process models are needed to compute a generic model that is able to track the course of a process.

Hypothesis 4: Clinical Applications of surgical process models

The surgical process models devised in this work are applicable to various surgical disciplines, different types of surgical interventions, surgical strategies and for miscellaneous clinical use cases.

Milestone 4a: Deriving requirements for a surgical assist systems in neurosurgery

This milestone targets at the implementation of surgical process models for the determination and prediction of quantitative implementation parameters of a surgical assist system in neurosurgery. Furthermore, a number of iSPMs will be analyzed to identify time and work step requirements that need to be considered for the development of the intended system.

Milestone 4b: Evaluation of a surgical assist system in pediatric surgery

This milestone investigates the applicability of surgical process models to quantify the impact of a surgical assist system on a surgical procedure from pediatric surgery. Conclusions concerning the advantageousness of the system for the presented use case will be drawn based on the analyses of gSPMs.

Milestone 4c: Assessment of surgeons' strategies in ophthalmology

The objective of this milestone is the quantification of surgeon-specific working strategies in ophthalmology with the help of gSPMs. A gSPM is computed for different surgeons and the results are then compared to each other to identify different working strategies.

2 Supporting methods for surgical process modeling

Some requirements are necessary to support the modeling and processing approaches in the subsequent sections. The first requirement to be dealt with is the specification of the use of expressions for the description of surgical processes. This description is usually performed by surgeons using natural language. However, to be able to handle such processes in an ICT system, they need to be describable with the help of (semi-) formal methods, such as, for instance, ontologies. Furthermore, it is necessary to validate the data acquisition strategies for SPMs objectively by measurements. While methods for the evaluation of sensor systems are typically conducted using binary classification, this approach is inappropriate for the assessment of observer-based data acquisition strategies. Therefore, particular metrics need to be developed to assess observer-based data acquisition.

The two supporting methods introduced in this section are of relevance for process modelers with a technical background; they provide assistance for the implementation of the modeling and the assessment of the measurement methods.

The publication

> *Neumuth D, Loebe F, Herre H, Neumuth T. Modeling surgical processes: A four-level translational approach. Artificial Intelligence in Medicine. 2011;51(3):147-161.*

describes the development of a process ontology approach that has the objective to use particular natural language expressions, as might be used by surgeons, to represent surgical processes using formal representations. The applicability of this process ontology is evaluated by using various examples for interview-based, observer-based, and sensor-based approaches for surgical process modeling. Thus, at the same time, an ontology-based meta-language is developed that allows for a mapping between these different approaches. As a conclusion it can be said that a unifying, ontologically, and mathematically founded framework for the modeling of surgical processes has been developed and its capacities were demonstrated by applying it to for different contemporary approaches to model surgical processes.

The publication

> *Neumuth T, Loebe F, Jannin P. Similarity metrics for surgical processes. Artificial Intelligence in Medicine. 2012; 54(1):15-27.*

describes the development of a set of process metrics for the assessment of observer-based process observations. Until now, there is no suitable means for the evaluation of observer-based protocols available. Existing approaches [Reneman et al. 2005; Baglio et al. 2004] for the determination of the recording quality, such as used, for instance, in the field of behavior research, put out a ratio that is too general, as single sections of processes or sequences of activities are rather inadequately considered. Therefore, this publication introduces a set of metrics for the described purpose, proves it formally and evaluates it experimentally.

Thus, five similarity metrics have been devised: one each for the granularity of a process, the content, time, order, and frequency of surgical activities. These metrics were then proven mathematically and validated experimentally by a simulation of clinical data sets from different surgical disciplines, such as cataract interventions from ophthalmology, craniotomy interventions, and supratentorial tumor removal interventions from neurosurgery. In addition, the metrics were evaluated concerning the education of observers for the acquisition of SPMs. Subsequently, the metrics were evaluated in the article by showing the learning progress of freshly trained observers.

2.1 Development of a process ontology for surgical processes

Title

Modeling surgical process: a four-level translational approach

Authors

Dayana Neumuth[1], Frank Loebe[2,3], Heinrich Herre Heinrich[3], Thomas Neumuth[1]

Departments

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] Universität Leipzig, Department of Computer Science, Intelligent Systems, Leipzig, Germany

[3] Universität Leipzig, Institute of Medical Informatics, Statistics and Epidemiology (IMISE), Research Group Ontologies in Medicine and Life Sciences (Onto-Med), Leipzig, Germany

Citation

Neumuth D, Loebe F, Herre H, Neumuth T. Modeling surgical process: a four-level translational approach. Artificial Intelligence in Medicine 2011; 51(3):147-161.

Keywords

Models, Theoretical; Linguistics; Concept Formation; Practice Guideline; Surgical Procedures, Operative

Abstract

Motivation: The precise and formal specification of surgical interventions is a necessary requirement for many applications in surgery, including teaching and learning, quality assessment and evaluation, and computer-assisted surgery. Currently, surgical processes are modeled by following various approaches. This diversity lacks a commonly agreed-upon conceptual foundation and thus impedes the comparability, the interoperability, and the uniform interpretation of process data.

Objective: However, it would be beneficial if scientific models, in the same context, shared a coherent conceptual and formal mathematical basis. Such a uniform foundation would simplify the acquisition and exchange of data, the transition and interpretation of study results, and the transfer and adaptation of methods and tools. Therefore, we propose a generic, formal framework for specifying surgical processes, which is presented together with its design methodology.

Methods: The methodology follows a four-level translational approach and comprises an ontological foundation for the formal level that orients itself by linguistic theories.

Results: A unifying framework for modeling surgical processes that is ontologically founded and formally and mathematically precise was developed. The expressive power and the unifying capacity of the presented framework are demonstrated by applying it to four contemporary approaches for surgical processes modeling by using the common underlying formalization.

Conclusions: The presented four-level approach allows for capturing knowledge of the surgical intervention formally. Natural language terms are consistently translated into an implementation level to support research fields where users express their expert knowledge about processes in natural language, but, in contrast to this, statistical analysis or data mining needs to be performed based on mathematically formalized data sets. The availability of such a translational approach is a valuable extension for research regarding the operating room of the future.

Introduction

In the domains of medical informatics and medical engineering, surgical workflows and time-action-analyses are gathering momentum. These broadly applicable concepts [Neumuth et al. 2009b] have been explored from the points of view of many surgical disciplines [Neumuth et al. 2007a] and for various reasons, including the evaluation of surgical-assist systems [Lemke and Vannier 2006], the control of surgical robots [Münchenberg et al. 2001a], instrument assessments [Mehta et al. 2002], and requirement engineering [Neumuth et al. 2009c]. Clinical work has also focused on surgical workflows for reengineering [Casaletto and Rajaratnam 2004], assessing human reliability [Malik et al. 2003], or comparing substitutive surgical strategies [den Boer et al. 1999]. A consolidated view of all of these factors indicates that there is a stable and growing demand for these kinds of studies and analyses.

What is quite salient, however, is that all of the mentioned approaches show an inclination to a disordered growth with regard to their basic concepts; only two of them use explicit models or ontologies [Jannin et al. 2003; Neumuth et al. 2006b]. Instead of a formal basis, the respective authors have used a variety of self-defined description 'languages'. This situation raises the question whether it is possible to find a common set of concepts that can be captured formally and that is applicable to every approach.

The advantages of such a formal basis would be manifold; we believe that it would enrich the research fields of medical computer-science and surgical workflow analysis. It would enhance the comparability, measurability, interoperability, and communicability of findings, statistical interpretations, and data-mining operations, as well as software applications (e.g., the construction of exchange platforms for surgical process models (SPMs) and study results). These may also be of increasing interest for medical personnel, who could use them to gather knowledge, plan interventions, or teach their craft.

The goal of this paper is to present a four-level framework that is ontologically founded and can serve as a basis for a formal representation of surgical processes. This framework will make different scientific approaches comparable and a mapping onto other languages possible. These 'other languages' comprise, amongst others, modeling languages for business process modeling [White and Miers 2008] and languages used for the modeling of discrete system behavior (e.g., automata, Petri nets, or execution languages for workflow schemas, such as structured Petri nets or Business Process Execution Language [van der Aalst and van Hee 2002]).

There is no generic framework for process modeling and analysis available that is adjusted to the medical field of surgical workflows and which specifies and integrates all relevant levels of abstraction into one coherent system. Such a framework should close the gap between individual data and the knowledge expressing abstract patterns about the data [Adlassnig et al. 2006]. Since the intended users are typically not familiar with logical formalisms, due to their mostly medical or engineering background, this framework should include a natural language level for communication. Then, this framework should provide means to transform natural language specifications of processes into mathematical models based ontologically based semantics. None of the existing formalisms has this as focus.

We will present a framework and its methodological basis to represent particular process models (corresponding to 'cases' in workflow terminology, in most instances). The methodology follows a four-level translational approach. Here, the term ‘translational’ conveys three different meanings: it refers to a translation between different levels of description specified and founded by this methodology, it relates to a translation between models associated to the corresponding levels, and, finally, it expresses the idea of a translation between theories from different fields of research. Further, the framework is related to existing approaches to modeling surgical workflows in order to demonstrate its applicability as the lowest common denominator between different approaches.

This article provides an introduction to the background of surgical process modeling, domain-specific terminology and abbreviations, and presents related approaches. The Methods section expounds basic methodological principles and the mathematical framework. The latter focuses on modeling patient-specific surgical processes, among other purposes for their electronic recording and analysis, e.g. regarding clinical questions, and the experimentally justified derivation of surgical workflows. The Application section demonstrates the implementation of the framework. Several aspects of the framework and its application are discussed, and prospects on future developments are given, finally followed by the conclusion.

Background

Terms and definitions for surgical process modeling

The term surgical process (SP) denotes a concept whose instances are individual surgical procedure courses. An SP is specified, in an adaptation of the definition of a business process in [Workflow Management Coalition 1999a], by a set of one or more linked procedures or activities whose instances (are intended to) collectively realize surgical objectives within the context of an organizational structure defining functions, roles, and relationships [Neumuth et al. 2009b].

The surgical objective is to achieve a normal, or at least ameliorated, state of the patient's body, and a surgical process changes an abnormal condition of the human body into a normal or better state. A procedure is performed in the organizational structure of a hospital which defines the functions, roles, and relationships of the participants within the operating room (OR).

In order to handle surgical processes in information systems, they must be represented as models. According to the general limitations of models – they exhibit reductions and simplifications of the domain [Frigg and Hartmann 2006] – we define a *surgical process model* (SPM) as a simplified pattern of a surgical process that reflects a predefined aspect of interest in a formal or semi-formal representation [Neumuth et al. 2009b]. Furthermore, we take on different types of SPMs: individual SPMs (iSPMs) and generic SPMs (gSPMs) [Neumuth et al. 2011b]. The term iSPM refers to individual, patient-specific models of SPs, thus representing the model of a single surgical case, while the term gSPM refers to a model of several surgical cases, such as a 'mean' treatment. The methods presented herein are applicable for iSPMs.

Introduction to pertinent literature

In computer science, there is a vast number of approaches, languages, and communities regarding process specifications in general. Constraining this to the present context, a considerable amount of work remains that deals with the formalization of workflow systems [Cicekli and Cicekli 2006]. However, the available methods and languages mainly share the ability to represent workflows on a formal basis. Apart from that, they are best suited to different tasks in connection with workflows: graph-based approaches (e.g., Petri nets and state-and-activity charts) are powerful tools with respect to visualizing workflows, as well as regarding the specification and verification of workflow properties [van der Aalst and van Hee 2002]. There is a large number of analysis methods and implemented tools for Petri nets.

Another broad line of workflow-related research comprises logic-based approaches, e.g., employing concurrent transaction logic for workflow analysis [Davulcu et al. 1998] or event calculus for specifying and executing workflows [Cicekli and Cicekli 2006]. Moreover, other process models have been proposed in connection with workflows, but they are more limited in scope (e.g., process algebras or event-condition-action rules (ibid.)). Temporal aspects of workflows, if supported at all, are dealt with mainly in the form of temporal constraints. Ignoring immediate relations to the field of workflows, numerous logic-based process formalisms have been presented in artificial intelligence (AI), where we just name situation calculus [Reiter 1991] and event calculus [Kowalski and Sergot 1986] as well-known representatives, and the unifying action calculus [Thielscher 2010] as a more recent, integrative approach.

There are three main problems with the mentioned approaches. Firstly, according to our knowledge, all mentioned approaches are designed for other purposes than naturally and efficiently supporting statistical analysis and data mining, for which they are not well-suited. Instead, logical approaches, for instance, obviously support reasoning as a core task and can be applied for, e.g. automated treatment planning and in decision support systems. Secondly, approaches applied in the workflow area in most cases assume or employ a top-down modeling of workflows in terms of manually devised models, in order to provide precise specifications, to verify their properties and schedules, to compute workflow executions, etc. Note that this holds true for medical guidelines, also, cf. [ten Teije et al. 2008; Hendler and Nau 1994; Anselma and Montani 2008; Terenziani et al. 2008; de Clercq et al. 2008; Mulyar et al. 2007]. These approaches are directed at normative processes rather than capturing and recording actual process information and are therefore not suitable for the retrospective analysis of individual processes. However, in the domain of surgical workflows no explicit knowledge exists that might be cast into formal models in a top-down manner. A high variability of patient properties, surgical skills and experience, as well as of available surgical technologies results in models showing high diversity [Neumuth et al. 2011b]. Furthermore, top-down models are usually equipped with few or no temporal measurements, which are in turn needed for many applications of surgical workflows, such as quantitative requirement analyses [Neumuth et al. 2009c]. Consequently, we require a formal model that also supports the bottom-up generation of workflows by observing iSPMs, as well as detailed time measurements within those recordings. We are not aware of any corresponding workflow-formalization approach. Thirdly, especially logical formalisms are not intelligible to and comprehensible for our intended users, as mentioned above. Logical representations cannot be easily communicated to medical staff, and they are hard to use in evaluations that are to be run by medical engineers or computer scientists without an appropriate background.

The most closely related resource in computer and information science that focus explicitly on surgical processes are terminological resources, for instance, national procedure classifications. In this particular context, the European norm EN 1828 [European Committee for Standardization (CEN) 2002] provides a minimal computer-based concept system for surgical procedures in order to “support the exchange of meaningful surgical procedure information between different national classifications or coding systems (…)”. The resulting level of granularity is coarse because such classifications are mainly used in connection with electronic health-care records and accounting systems. Moreover, temporal relationships are not covered. Modeling the temporal structure of interventions is therefore beyond the scope of EN 1828.

There is another large branch of related work that pertains to AI, with influences from linguistics, cognitive science and philosophy. A few corresponding approaches were named above as representatives of logic-based process representations [Reiter 1991; Kowalski and Sergot 1986; Thielscher 2010]. Indeed, the AI subdomain of theories and reasoning about action and time has been an active field of research for several decades. Frequently drawing on linguistic and philosophical inspiration, it includes works like the development of formalisms for reasoning about actions [Allen 1984] and the deployment of temporal constraints between causes and effects of causal relations [Terenziani and Torasso 1995]. Due to the close relationship between processes and time, there is a further large intersection with the AI subfield of temporal representation and reasoning, cf. [Adlassnig et al. 2006; Fisher et al. 2006]. For our purposes, the dynamic aspects of logical representations like reasoning and its further applications (e.g. for planning) are not yet immediately applicable. Adopting logical formalisms as a declarative form of representation is appropriate for some parts of our framework, but plays a minor role for the mathematical model to be presented below, due to its intended application cases. Therefore, currently the main connection to these fields in AI resides in the theories of time and processes that are presented there for their adoption and extension as conceptual or ontological basis of formal models.

Indeed, processes and time form important classes of entities that have been studied in ontology research, including philosophical investigations [Ma 2007], knowledge representation [Schlenoff et al. 1999; Seibt 2007], and computer-science ontologies [Herre et al. 2007]. The category of processes is at the most general level of abstraction of concrete individuals and, hence, is usually included in top-level ontologies. Top-level or foundational ontologies apply to every area of the world, in contrast to the various *generic*, *domain core*, or *domain ontologies*, which are associated with more restricted fields of interest. The category of processes is contained in the top-level ontologies DOLCE [Masolo et al. 2003], GFO [Herre et al. 2007], and ISO 15926-2 [West et al. 2003], each of which represents a different approach to processes. In DOLCE, objects (endurants) and processes (perdurants) are disjoint classes of entities that are connected by certain relations. ISO 15926-2 contains processes as the only basic category, whereas GFO provides three kinds of concrete basic entities (perpetuants, presentials, and processes), which are fully integrated into a unified system. The basic integration axiom says that for every perpetuant (presenting the notion of enduring object), there exists a corresponding process such that the snap-shots of that process coincide with the presentials associated with (“exhibited by”) the perpetuant [Herre et al. 2007; Herre 2009].

Process modeling in the framework of top-level ontologies is a new research field, and there are few papers or investigations related to this topic [Green and Rosemann 2000; Evermann and Wand 2001; Evermann 2009] with respect to our focus on process descriptions with detailed temporal information. The closest related effort is the ISO Standard 18629 on the process specification language (PSL, [Schlenoff et al. 1999]). PSL consists of a core that exhibits the following four kinds of entities: activities, activity occurrences, time-points, and objects. The underlying ontology of PSL pertains (to some extent) to the top-level ontology of DOLCE [Masolo et al. 2003]. In particular, the notion of activity occurrence relates to the notion of perdurant in DOLCE, whereas objects in PSL correspond to endurants in DOLCE. Additionally, there are several extensions of the PSL core, treating relevant aspects of processes. PSL can be interpreted and mapped into the GFO, providing an ontological foundation of the PSL semantics. PSL is formalized in machine-readable formats covering first-order logic. Alongside the resulting descriptions themselves, the main purpose of that representation is to support automated reasoning over them. The relation between process characterizations in natural language and PSL formalizations has not been established. The purposes of declarative representation and of reasoning also differ from goals such as the statistical evaluation and data mining of surgical processes, which can be more easily supported by broader, more general mathematical machinery than first-order logic.

In this paper, we present the first application of process ontologies in the surgical domain, where no process-related ontology has yet been developed or applied.

Methods

Basic methodology: a four-level approach

From the methodological point of view, we propose a modeling strategy that considers four different levels: the *natural language level*, the *conceptual* or *ontological level*, the *formal* or *mathematical level*, and the *implementation* itself. Certain relations connect these four levels. The natural language level is linked to the ontological level by ontological analyses through a process called *ontological reduction* [Herre and Heller 2006; Herre and Loebe 2005], whereas the mathematical level results from a translation of ontological categories at the second level into mathematics (e.g., set-theoretical structures). In this section, we will introduce the single levels and describe their relations to the subsequent sections.

Characterization of the levels

Level one, the natural language level, is related to the user. In our case, the assumed users are mostly surgeons and medical engineers. The former, especially, are not accustomed to dealing with formal representations or using formal methods to analyze surgical concepts. For this reason, the natural language level is required in order to include the implicit knowledge and experience of the clinical users into our model. The natural language level further provides an interface for communicating the results of analyses, which are carried out in terms of the remaining levels, back to the users.

The second level, the conceptual or ontological level, deals with the ontological analysis of domain knowledge, which is significantly based on natural language expressions. Because natural language expressions usually allow for distinct interpretations depending on context, distinct ontologies may be derived from them. Linguistic patterns can be employed for ontological analysis, and existing bodies of real-world knowledge might be reused, for example, as represented in pre-existing ontologies. In particular, top-level ontologies can be used as a basis for developing domain-specific ontologies. This is the primary field of application of the method of ontological reduction.

The third and formal level provides for mathematical formalizations of domain knowledge dedicated to determinate purposes. Such formalizations must rest on the second – the conceptual – level, where different formalizations based on a single ontology may be useful for distinct purposes. Maintaining the link to the conceptual level allows for interoperability and comparability of different models, making cross-modeling approaches possible and thus the gathering of knowledge from different sources and from different points of view.

Finally, the implementation level is concerned with the realization of formalizations from the previous level in languages with a practical orientation, primarily machine-processable languages. Here, another multiplication of representations arises due to multiple different implementations of a single formal model. Distinct implementations occur for different languages as well as for a single language. Several implementations may encode a formalization in progressively complex ways. The four levels can be seen in Figure 2.1.1.

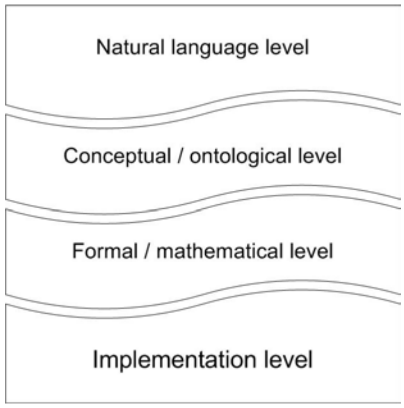


Figure 2.1.1: The four levels of the methodology.

Coverage of the levels

The goal of this article is to present a formal framework for specifying surgical processes. Accordingly, the formal level is expounded in detail in the section Mathematical formalization. The purpose of this formalization is to share datasets of surgical processes for extended analysis, data mining, and processing. The remaining levels are only partially covered or completely elided in the case of the natural language level. For the implementation level, an example implementation is depicted in terms of the unified modeling language (UML) [Booch et al. 2005; Rumbaugh et al. 2005] and eventually defined as a dialect of the extensible markup language (XML) [Bray et al. 2008] in the section Implementation.

Regarding the ontological level, an elaborate ontological analysis of surgical processes is outside the scope of this paper and has not yet been completed. However, it is also not required at this stage of our work. One fundamental premise for the formal framework presented is the separation of concepts into those captured by representational structures and others referring to specific content. This division provides for a generic and uniform syntactic representation on an abstract, minimized conceptual basis, whereas further specificities must be encapsulated. This is desirable because of the different purposes of SPMs, on the one hand, and the high degree of dependence on natural language of detailed content on the other hand. The distinction between structure and content draws on an analogy to the relationship of top-level and domain-specific ontologies. Top-level ontologies provide a basic structure that can be refined by domain-specific concepts. Similarly, the primitives of the framework introduced below are implicitly based on an abstract ontology to which SPMs may commit by adopting the framework.

Another ontology-related aspect is to consider classifications of entities of basic types. In particular, we expect that a classification of processes will prove useful for the proposed framework. For instance, classifications can be utilized to tailor process analyses to specific kinds of processes. Therefore, we restrict the exposition regarding the ontological level herein mainly to an outline of the established theory of eventualities from the domain of linguistics, which is adopted for the classification of processes. Notably, further analyses of that theory should be conducted with respect to top-level ontologies. Initial results suggest that the classification of processual structures in GFO [Herre et al. 2007] can be used for this step, which remains for future work. The next section describes the classification system adopted in the present work. In addition, comments on the part-whole relation and granularity with respect to processes close the treatment of the ontological level herein.

Conceptual level

Theory of eventualities

In connection with the use of natural language in many present-day SPMs, as well as the level of abstraction in which the framework is based, we decided to rely on a basic classification of processes originating primarily from linguistics, but based on philosophical approaches (see [Casati and Varzi 2008]). In linguistics, and more specifically in the organization of the grammar of natural languages, eventualities have played a major role for more than 30 years. Linguists (e.g. [Dowty 1979; Bach 1986]) rely heavily on philosophical works (e.g. [Ryle 1984; Kenny 2003; Vendler 1967]), which in turn refer to Aristotle [Barnes 1995]. Moreover, there is a fruitful mutual influence with process-related branches of AI, cf. [Ma 2007; Moens and Steedman 1988].

Herein, Bach's term 'eventualities' [Bach 1986] will be used to refer to the topmost category of (linguistically speaking) verbs or (from the modeling perspective) of processes and processual entities. We distinguish four 'classical' main types of eventualities that are mainly based on Vendler's theories [Vendler 1967]: *states* (processes without change), *activities* (unbounded processes), *accomplishments* (bounded processes), and *achievements* (point events).

The presented classification examines three semantic properties of verbs, some of which are inherent in the verb itself, while others are conveyed by the interaction of the verb and its arguments. It is important to note that the distinction between eventualities is not strict in the sense that in natural language linguistic features, such as the use of progressive or adverbials, can result in a change of eventuality [Moens and Steedman 1988]. The semantic properties are the following: whether or not an eventuality has a natural endpoint [± telic], whether it can be analyzed as being constructed of phases that can be different [± dynamic], and whether it continues for a period of time or is limited to a point of time [± durative]. Following [Bach 1986; Ryle 1984; Kenny 2003; Vendler 1967; Barnes 1995; Moens and Steedman 1988; Carlson 1981; Parsons 1994] these three properties suffice to differentiate between all four eventualities, as shown in Figure 2.1.2. Note that in this article, semelfactives [Comrie 1976] are excluded for simplicity.

In the remainder of this section, we further characterize the four eventuality types for better comprehension, as shown in Figure 2.1.3. *States*, for example, 'scalpel is used', are classified as [+durative, -dynamic, -telic]. They carry on for some time, and one can ask for how long a state lasts. However, it is not reasonable to ask how long a state takes or whether it culminates because states are regarded as non-developing (there are no changes within a state with respect to its defining conditions), and, therefore, they cannot have natural endpoints. Two special characteristics of states are that they are cumulative and strongly homogenous. The former characteristic allows one to infer from the statements 'This scalpel was used from 10:00 a.m. to 10:10 a.m.' and 'This (indicating the same) scalpel was used from 10:10 a.m. to 10:15 a.m.' that 'This scalpel was used from 10:00 a.m. until 10:15 a.m.' is true. Homogeneity is concerned with parts of an eventuality. In the example above, given a state from 10:00 a.m. to 10:15 a.m., homogeneity dictates that the scalpel was used at any given point of time within this interval.

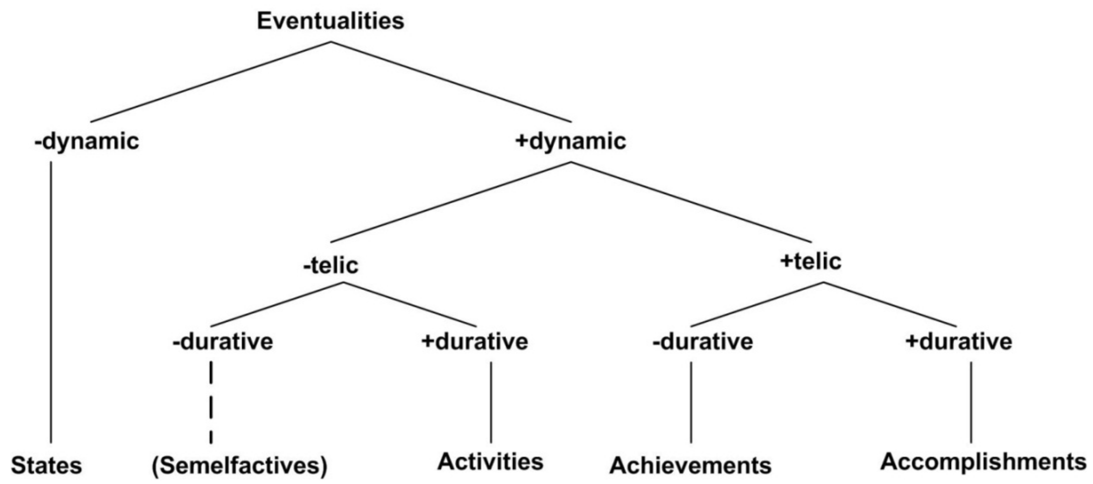


Figure 2.1.2: Eventuality classification according to [Bach 1986] and [Carlson 1981].

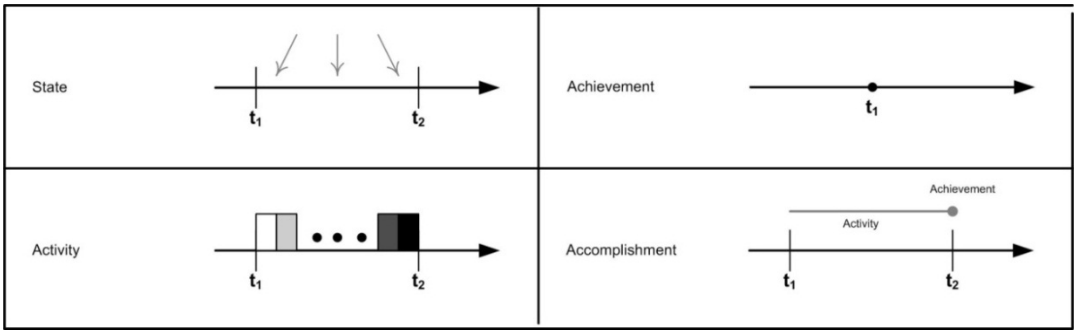


Figure 2.1.3: Schematic representation of eventuality types.

Activities share with states the properties of being extended and having no inherent endpoints [+durative, -telic], but they are [+dynamic]. An example is 'The surgeon cuts (sth.)', with the connotation that he is moving the scalpel. Although the cutting will stop at some point, the point of time *at which* the cutting will end cannot be determined from the type of eventuality given in the sentence. Activities report progress and exhibit an inner structure, for instance, by being composed of phases or by some inherent development. In terms of homogeneity, activities can be homogeneous up to a certain degree, but they need not be. Moreover, activities may be interrupted and continue later on.

The sentence 'The surgeon cuts off the thread' reports an *accomplishment* (note again that in addition to the verb 'to cut' being of relevance for the corresponding eventuality type, the verb-argument interaction may be involved as well). Like activities, accomplishments are temporally extended and have a certain structure; in addition, they have an inherent endpoint [+durative, +dynamic, +telic]. Within accomplishments, as can be seen in Figure 2.1.3, an activity is present, which is often referred to as a *preparation phase*. In addition, there is a natural condition characterizing the end (or the beginning) of an accomplishment, its *culmination point*, which can also be regarded as achievement. In 'cuts off the thread', the preparation phase covers all of the cutting while the thread is still attached. The transition to 'thread is severed' and 'cutting stopped' necessarily yields the culmination point. Accomplishments may be interrupted, analogously to activities. In addition, it is possible that an accomplishment remains unfinished (without culmination). In [Dowty 1979], this fact is called the *imperfective paradox*.

Finally, *achievements* denote eventualities that have no duration (though linguists also allow for a very limited amount of time) and no internal structure [-durative, +dynamic, +telic]. Thus, asking how long an achievement takes or lasts is irrational. What is important for achievements is the change that they incur. 'The surgeon turns off the endoscope.' is an achievement example, addressing the change of the endoscope's status from 'being on' to 'being off'. Piñon [Pinon 1997] argues that some achievements can be treated as the beginnings or ends of other eventualities, such as 'The surgeon starts to cut.' This view can be combined with a variation of the understanding of accomplishments reported in [Rothstein 1999], namely that an accomplishment is composed of an activity and an achievement. This links directly with considerations of the part-whole relation (regarding processes) and different levels of granularity with respect to that relation.

Mereology and granularity

Existing surgical process models are often specified using different levels of part-whole granularity, which is easily visible in the approaches discussed in the Application section. For instance, there is the overall surgical procedure, which may be divided into phases. Phases might be split into work steps, and those in turn may comprise particular tasks. As indicated above, herein we cannot provide in-depth accounts of neither mereology (the theory of the part-whole relation) nor granularity, even if limiting ourselves to surgical processes.

These issues are aspects of extending the ontological analysis of surgical processes, which we will pursue in future work, cf. Discussion section below. Moreover, this relates directly to the top-level ontological foundation of such an analysis, because the category of processes as well as the part-whole relation are commonly based on top-level ontologies. As mentioned in the Introduction, processes in general require further treatment from an ontological point of view. Nevertheless, there are numerous works that include mereological or granularity issues of processes and that are therefore expected to affect the mentioned ontological analysis. This starts from general mereology [Ridder 2002] in formal ontology, spans over detailed treatments in linguistics [Moens and Steedman 1988; Carlson 1981; Parsons 1994; Pietroski 2005] and reaches into artificial intelligence in general, cf. [Fisher et al. 2006] and AI in medicine in particular, see [Adlassnig et al. 2006].

According to this situation, herein we restrict ourselves to sketching some types of constraints for processes in terms of a number of examples for which wide agreement can be expected. These are collected in natural language in this section, whereas selected formal equivalents are presented in the following subsection. Note that, based on the GFO theory of processes [Herre et al. 2007], we distinguish two basic kinds of part-whole relations for processes: *temporal part-of* (for temporal parts of processes, which may involve all participants of the process) and *layer part-of* (for parts of processes that encompass less participants or aspects than the original process, but may share its temporal extension with the original process).

1.) The temporal position and extension of every temporal part of an eventuality E must be temporally constrained by the temporal position and extension of E, i.e., every temporal part of E must happen during E. This entails that achievements, as eventualities without temporal extension, cannot have proper temporal parts.

2.) All participants within every layer part of an eventuality E must be parts of participants in E. Analogously, aspects covered by a layer part of E must be "justified" by E, i.e., those aspects must pertain to either participants in E or to parts of E -participants.

3.) For every eventuality E there is a coherent eventuality C [Herre et al. 2007] that E is a temporal or layer part of, or from which E can be derived.

4.) Every temporal part and every layer part of an eventuality E is finer grained than or at maximum at the same level of granularity as that determined by E itself.

Moreover, the linguistic and philosophical literature discusses the interplay between the part-whole relation regarding eventualities and the classification of eventualities. As discussed in the previous section, accomplishments are frequently considered to be composed of an activity and an achievement [Moens and Steedman 1988; Parsons 1994; Ridder 2002]. Further statements of this kind are available and might be included in an elaborate mereology for surgical processes or even processes in general. Note, however, that we see this literature primarily as a starting point, whereas an integrated mereological account for processes is expected to require an extensive amount of further work.

Mathematical formalization

In this section, we present the definition and description of structural representations of surgical processes and their components, corresponding to the formal level of our methodology. This provides an abstract, general framework and terminology for the specification of surgical processes. Moreover, it serves as a basis for scientific description and usage. This framework is capable of representing, formally and mathematically, recordings of individual surgical interventions and some of their generalized patterns.

The framework is introduced in an arrangement that progresses from simple to complex. Ultimately, it is based on classical mathematical representations, mainly set theory and real-valued functions. For the specification of granularity, we introduce three functions λ, μ, and ι. λ (local granularity) is based solely on the parts/components of a process; μ (model granularity) adapts λ measures to a reference process. Whereas these two are formally captured, ι is content-oriented, referring to "global" levels of granularity. Hence, we do not assign concrete values to ι applications, but will use it merely comparatively. The domains of all three functions are the eventualities (attributive and processual) that are introduced below. The ranges of λ and μ are $\mathbb{N}$, where smaller values of $\mathbb{N}$ indicate finer granularity. To highlight that a number is to be interpreted as a granularity value, we may write λ_i or μ_i for an $i \in \mathbb{N}$.

Following the theory of eventuality types as introduced above, requires a formalization adapted to the presented problems. Firstly, a formal foundation is defined in terms of attributes and values. Secondly, the eventualities are formally described at distinct levels of granularity.

Attributes and values

The basic data elements of the framework approximately follow the attribute-value model [Ziarko and Shan 1996]. Measurements provide knowledge of situations that are present in a surgical process. We refer to *attributes* as representations of measurable phenomena of a surgical process in great generality. More precisely, an attribute represents a range of conditions to which a particular element may apply at a time, called a *value*. Values may refer to qualities, relations, and complex situations (each represented by a single value). From the perspective of processes, attributes characterize processes (which themselves reside at a certain level of part-whole granularity).

Formally, an attribute $A = (L_A, V_A)$ is understood as a labeled set of possible values or a value space, cf. also Gärdenfors' *conceptual spaces* [Gärdenfors 2000]. Labeling is necessary because attributes are "linked" with the phenomena they measure. Therefore, two attributes may formally refer to the same set of values yet be definitely distinct and not mutually exchangeable.

The set of all attributes used for describing processes is denoted by ATT. T denotes a special attribute for temporal values, $T \in ATT$. Its values comprise time stamps, $t_i \in V_T$. For the elements of T, a strict linear order is assumed, denoted by $<$ ($\leq$ for the reflexive variant). The elements of T are at least ordinal, if not scalable. We assume that there is no smallest and no greatest element for T.

With attributes, values, and a dedicated time attribute, the basics for recording temporal information of processes are available.

Attributive dynamics

The phenomena underlying attributes (apart from T) develop in the course of a surgical process, more precisely of the part that the attribute characterizes. For a single attribute A, the course of development of its values can be reflected formally in terms of a partial function $f: T \rightarrow V_A$. We require that those functions are total over an interval of T and call such fragments a *stage* of development of an attribute A over time (an A-*stage*). For reference and for flexibility in specifying temporal relations, we further add an optional identifier (S_A in the subsequent example, "_" if omitted herein) for the particular stage and the attribute label L_A for readability. Hence, an A-stage is represented as a quintuple (S_A, L_A, f, t, t') such that $[t, t']$ is the interval over which f is defined. Hence, for any stage, $t \leq t'$. Note that V_A may comprise a specific value $undefined_A$ in order to ensure f being total over $[t, t']$. For a given time stamp t'', (S_A, L_A, f, t, t') *covers* t'' iff $t \leq t'' \leq t'$.

Already at this level, three types of eventualities, as introduced in the Basic Methodology section, can be borrowed. Stages with constant functions correspond to states because they do not exhibit changes (of the phenomenon behind the attribute). S_A is a *stative stage* (or *state*) iff for every t_1 and t_2 covered by S_A: $f(t_1) = f(t_2)$. In contrast, variable functions reflect the dynamic aspect of activities and accomplishments. Grasping the further distinction between those two types according to telic aspects is harder. In some cases of accomplishments, the final value of the function may be distinguished in relation to the course of f over T. However, there are other cases that necessitate implicitly accounting for subsequent stages. Following many approaches in linguistics, we adopted the view that accomplishments are composed of "an activity and a resultant change of state, where the change of state gives the natural stopping point for the activity" [Rothstein 1999]. Hence, we call a stage (S_A, L_A, f, t, t') an *activity* iff there are $t_1, t_2 \in [t, t']$ such that $f(t_1) \neq f(t_2)$. This entails that every stage is either a state or an activity.

For greater flexibility in modeling, we allow several stages of the same attribute (possibly with temporal gaps) to characterize a process, instead of only a single stage. This is captured by the notion of an *admissible stage set*, which enforces an ordering of all stages in the set. Admissible means that the functions of any two stages in that collection overlap, at most, in their interval boundaries, and at most two stages overlap at every time stamp. Accordingly, a set of stages S^*_A is admissible iff the following conditions are satisfied:

- for each pair of stages (S_A, L_A, f, t_1, t_2), $(S'_A, L_A, f', t'_1, t'_2)$, it is the case that
  - $[t_1, t_2]$and $[t'_1, t'_2]$ are disjoint *or*
  - $t_2 = t'_1$or $t'_2 = t_1$
- for every t such that an $S_A \in S^*_A$ covers t, there is at most one further $S'_A \neq S_A$ that covers t.

The conjunction of these two conditions yields the effect that two stages can, at most, "touch" each other temporally on two of their boundaries (i.e., the stages of an admissible stage set can be completely temporally ordered). As for time stamps, the symbol $<$ may be used to describe the temporal ordering of stages.

It remains to capture "changes of state", including achievements. An *achievement* refers to a stage (S_A, L_A, f, t_1, t_2) and its temporally closest successor $(S'_A, L_A, f', t'_1, t'_2)$. Formally, given an admissible stage set, S'_A is the successor of S_A iff $t_2 \leq t'_1$, and there is no stage $(S''_A, L_A, f'', t''_1, t''_2)$ with $t_2 \leq t''_1 \leq t'_1$ in the stage set. An achievement between S_A and S'_A with identifier set to E_A^{ach} is captured by any of the tuples $(E_A^{ach}, L_A, S_A, S'_A, t_2, t'_1)$ and $(E_A^{ach}, L_A, f, f', t_2, t'_1)$. The latter case allows for achievements without recording their surrounding stages. For generality, we do not establish restrictions at this point , i.e., $f = f'$ and $t_2 = t'_1$ remain possible. This would correspond to instantaneous atelic eventualities, cf. semelfactives in Figure 2.1.2, although in our experiments we have not yet observed and thus not yet arranged to record eventualities of this type. They could be useful in connection with abstractions among attributes. Nevertheless, the default assumption for an achievement is that f and f' or t_2 and t'_1 are distinct, thus involving a "change". We can distinguish general achievements from proper ones, where the latter are temporally defined as those satisfying the condition $t_2 = t'_1$ (i.e., a proper achievement has equal time stamps). For example, a transition from a state of attribute value v to another state of v' at one moment t yields a proper achievement.

It follows immediately from the above definitions that stages and achievements are complementary to each other, which gives rise to an integrated view. An *eventuality system* E_A^* for an attribute A consists of an admissible stage set S_A^* and a set of $E_A^{ach,*}$ i.e., $E_A^* = S_A^* \cup E_A^{ach,*}$. In accordance with the above remark on *accomplishments*, those can now be understood as an eventuality system $\{(\_, L_A, f, t_1, t_2), (\_, L_A, f, f', t_2, t_2)\}$ (corresponding to reaching the telic conditions at the end; a change at the beginning is analogously handled).

An eventuality system E^* can be incomplete in the sense that it need not contain an achievement for every pair of successive stages, and it need not comprise two fitting stages for an achievement. However, it must be completable. We postulate as eventuality system completion requirement that for every eventuality system

$E^* = S^* \cup E^{ach,*}$ there is a corresponding completed eventuality system $E'^* = S'^* \cup E^{ach,*}$ with an admissible stage set $S'^* \supseteq S^*$", where S'^* contains two fitting stages for every achievement in $E^{ach,*}$. Note that this prevents any two achievements from sharing a time stamp, t, because otherwise t would entail the existence of three states covering t (e.g., one for the shared later stage, plus two for both achievements, which have distinct earlier stages). This appears reasonable for a single attribute. The system E^* is said to cover an interval $[t_1, t_2]$ iff t_1 is the starting point of the first stage or achievement in E^* and t_2 is the end point of the last stage or achievement.

Regarding granularity assignments, only local granularity measures are applicable to attributes, where they do not "add" part-whole granularity. Thus, for every attribute stage or achievement E^{att} and eventuality system E^*, $\lambda(E^{att}) = 0$ and $\lambda(E^*) = 0$. Alternatively, we say that an eventuality (system) is in the local granularity level λ_0.

Basic and high-level dynamics

With the notions of attributes, values, and (attribute-level) eventuality systems in the previous section, we introduced basic modeling elements in order to record the temporal development of individual observables (attributes) with respect to processes. However, processes must also be represented, and most descriptions of surgical processes require a higher level of aggregation (or several levels). In general, we adopt a common representation scheme for all *processes*, which is similar to attribute stages. Tuples of the form $(P, L_P, AS, CS, t_1, t_2)$ represent processes by an identifier P, a type label L_P, a set of attribute eventuality systems AS, a set of parts/components CS (which are themselves processes), and time stamps for the beginning and end, t_1 and t_2 (examples are given in the following section). Identifier and type labels are optional, and AS and CS arguments that are singleton sets may omit set notation. Moreover, a mereological theory for processes may be adopted and should then be reflected as far as possible in terms of corresponding formal constraints, e.g. on the nesting of processes. For instance, the first sample condition in section Mereology and granularity would translate into the formalization presented here by requiring that, for every pair of a process $(P, L_P, AS, CS, t_1, t_2)$ and one of its components $(P', L'_P, AS', CS', t'_1, t'_2)$, where $P' \in CS$, it holds that $t_1 \leq t'_1$ and $t'_2 \leq t_2$.

Achievements between processes arise in an analogous manner to attributive achievements (i.e., as transitions between two processes referred to via their identifiers). Similar to ATT, EVT denotes the set of all corresponding tuples representing eventualities in a certain context (which are subject to further conditions introduced below).

Granularity concerning processes is modeled as follows: the local granularity of a process is determined by the (maximum of the) granularities of its components: $\lambda(P) = \max_{C \in CS}(\lambda(C)) + 1$ if $CS \neq \emptyset$, otherwise $\lambda(P) = 1$. A process is homogeneous with respect to the level if all components have the same λ value. For a fixed set of processes (closed with respect to their nested components), the model granularity μ is defined by determining the maximal local granularity μ_{max}, which is assigned to all processes not contained in any other process (called root processes). The μ value of all remaining processes is $\mu_{max} - \mu_{path}$, where μ_{path} is the length of the shortest containment path from the process to any root process.

It remains to describe admissible compositions, starting with the basic level λ_1. This first step unites attributes describing different aspects of a single process. The latter is not subject to further 'part-whole analysis', and it is typically rather limited in temporal extent. A *basic-level process* is of the form $(P, L_P, AS, \emptyset, t_1, t_2)$, meaning that it consists only of attributive eventuality systems $AS = \{E^*_{A,1}, \dots, E^*_{A,n}\}$ and the interval $[t_1, t_2]$ that it covers. The $E^*_{A,i}$ are defined over distinct attributes $A_1, \dots, A_n$. All $E^*_{A,i}$ over a common attribute must form an eventuality system. All these systems must cover $[t_1, t_2]$ (the latter condition may be weakened to allow for fuzzy boundaries). Basic-level processes can again be classified as the three durative eventuality types, which yield *basic-level states*, *activities*, and *accomplishments*. In some cases, this classification can be derived from the constituents $E^*_{A,i} \in AS$. For instance, if there is only one accomplishment and the remaining attribute-level systems comprise only single states each, the resulting basic process may be considered an accomplishment. Another case is one where all attribute-level systems correspond to a single state, which most reasonably leads to a basic-level state. However, in other cases, the nature of a basic process may be hidden due to the unavailability of a corresponding observable or to not measuring it. Another observation is that basic-level activities and accomplishments are commonly aggregated from several attributes, whereas states may reasonably be lifted to the basic level as singleton sets. Accordingly, the classification of basic-level processes should be considered from case to case.

Further levels of aggregation can provide useful levels of abstraction. Given basic-level eventualities, this kind of aggregation is more focused on finding *high-level processes*, $\lambda(P) \geq 2$, whose components are temporally distinct processes. Currently, in most cases, proximate components are considered. This kind of aggregation sometimes involves further abstraction beyond temporal summarization, representable in terms of new attributes. The formal account for high-level processes is strictly analogous to basic-level processes, in terms of the common representation with attribute and component sets and time limits. Again, the four-fold classification can be considered at all levels. Because the sets of higher levels differ considerably among existing approaches, we do not introduce any particular account into the general model.

Implementation

Regarding the fourth level of our method, the implementation level, we first represent the mathematical model as a UML class diagram [Booch et al. 2005; Rumbaugh et al. 2005] (cf. Figure 2.1.4), as this form of representation paves the way for software applications. Moreover, it serves as an intermediate representation for SPMs in XML [Bray et al. 2008]. For the latter purpose, we provide an XML-Schema file[1] that defines an XML dialect for exchanging SPMs according to the presented framework. The file can be used for validating models.

A few final remarks about the UML diagram may prove useful. Attributes and their dynamics are covered in the lower half, whereas processes reside in the upper half of the diagram. In both cases, the UML associations 'isFrom' and 'leadsTo' express achievements based on explicitly available stages or processes. Attributive achievements may be specified alternatively by two functions only. Time is central to both attributes and processes. Because time points must occur in a pair wise fashion

[1] Available from: http://www.onto-med.de/software/spm.xsd

(if they occur at all), the UML class 'Time Specification' is as appropriate as individual associations to start and stop times would be. The doubly named 'associations' link, by order, with the UML attributes 'initial' and 'final'. For example, a process starts at the 'initial' time and ends at the 'final' time. A processual achievement transitions from a process ending at an 'initial' time to another starting at a 'final' time.

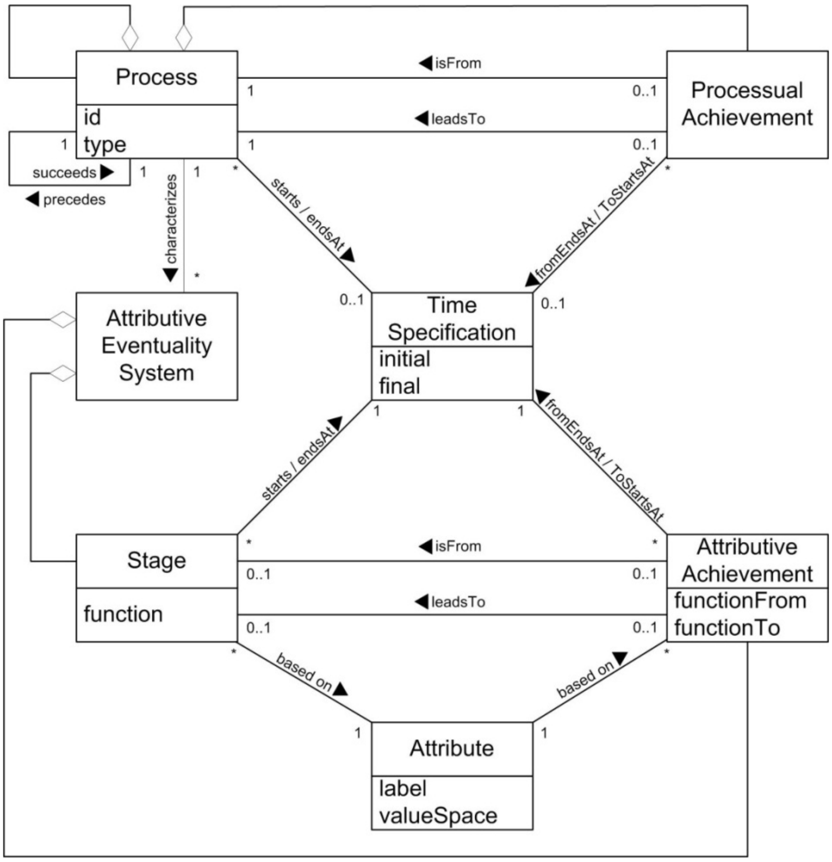


Figure 2.1.4: UML class diagram of the presented formalization.

Application of the four-level translational approach

Next, the proposed framework is compared with, applied to, and evaluated for several recent approaches to surgical workflows and SPMs. These works were chosen, because each of them established the base for clinically useful applications and explicitly published SPMs.

A Model for laparoscopic Nissen fundoplications

MacKenzie et al. [MacKenzie et al. 2001; Cao et al. 1996; Ibbotson et al. 1999] have published ergonomic studies based on videotaped laparoscopic training workshops for Nissen Fundoplications. Laparoscopic interventions are 'keyhole' interventions, a minimally invasive kind of surgical approach. The intention of the mentioned research was to assess the skills of surgical residents and to develop a hierarchical framework for assessment based on plans and the structure of goal-directed human behavior. Their approach was the first attempt to decompose a complete procedure to the level of simple instrument motions.

The authors identified surgical activity types and proposed a semi-formal, hierarchical decomposition of laparoscopic interventions. The modeling was performed iteratively, using both top-down and bottom-up approaches.

Figure 2.1.5 shows a cutout of the resulting procedure model. Within this model, a surgical procedure is divided into six granularity levels: *surgical procedure*, *step*, *sub-step*, *task*, *sub-task*, and *tool motion*. Five basic motion elements were identified. This decomposition includes only one kind of relation, namely part-of relations, and attribute values as natural language expressions.

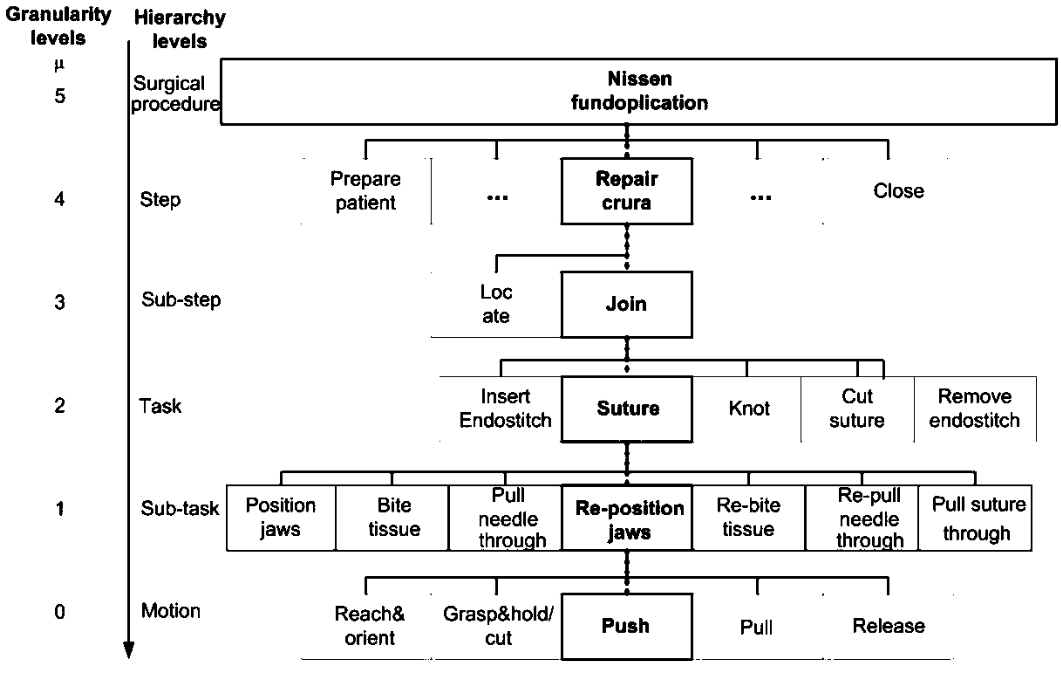


Figure 2.1.5: Procedure model proposed for Nissen fundoplications (cutout) [MacKenzie et al. 2001], with an example aggregation from the tool-motion level to the complete surgical procedure.

This approach, however, has some limitations. Concurrencies and iterations were disregarded by the researchers and therefore not treated. In addition, the notion *surgical event*, from which the model is derived, is not clearly defined. There are also some minor slips within the model, such as accounting for the insertion of instruments but not their removal. An ontology would seem a sensible addition to this model.

On the other hand, the approach did not aim at presenting a generic model and has some strong points: in terms of modeling, the authors tried to find basic patterns to decompose single surgical maneuvers.

In order to transform the model of MacKenzie et al. into an instance of the model presented herein, we consider only their lowest level of tool motions as attribute stages, whereas the components at all other levels amount to processes in our framework. Specifying the highlighted components in Figure 2.1.5 in a top-down manner yields the following result, where t_{init} and t_{fin} denote postulated start and end times of the overall procedure:

- $\left(E_1, NissenFundoplication, \emptyset, \{E_2, \dots, E_8\}, t_{init}, t_{fin}\right)$
- $(E_5, RepairCrura, \emptyset, \{E_{18}, E_{19}\}, E_4, E_6)$
- $(E_{19}, Join, \emptyset, \{E_{54}, \dots, E_{58}\}, E_{18}, \_)$
- $(E_{55}, Suture, \emptyset, \{E_{83}, \dots, E_{89}\}, E_{54}, E_{56})$
- $\left(E_{86}, RepositionJaws, \{E^*_{ro}, E^*_{ghc}, E^*_{push}, E^*_{pull}, E^*_{rel}\}, \emptyset, E_4, E_6\right)$

Further components are merely omitted but would be specified completely analogously. Only the attributive components cannot be described because the original articles do not provide corresponding data. However, for a valid model, there must be a set of attributive eventuality systems over the five motion attributes *reach & orient* (ro), *grasp & hold/cut* (ghc), *push*, *pull*, and *release* (rel) such that each attribute-specific set yields an eventuality system over its defining attribute. For illustration purposes, Figure 2.1.6 graphically represents the hypothetical stage set $E^*_{push} = \{(A_1, push, 1, t_1, t_2), (A_2, push, 0, t_2, t_3), (A_3, push, 1, t_3, t_4)\}$.

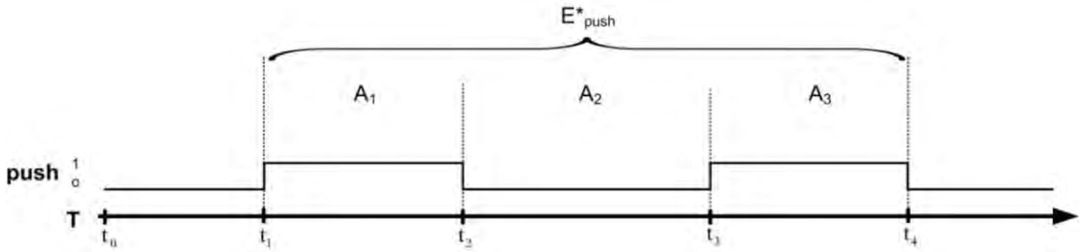


Figure 2.1.6: Hypothetical stage set E^*_{push} (attribute level).

The attributes of E_{86} exhibit local level λ_0 and $\lambda(E_{86}) = 1$. The model level is also $\mu(E_{86}) = 1$, whereas the lowest numbered model level of all neighbors of 'Suture' (E_{55}) is μ_2, which corresponds to the task level in [MacKenzie et al. 2001]. Regarding the conceptual levels, the model is mostly but not completely uniform compared to the model levels. For instance, $\iota(E_2) < \iota(E_5)$, where E_2 is the step "Prepare Patient", $\iota(E_5)$ as depicted above.

Altogether, this demonstrates that the formal account introduced in this work is fully applicable to the approach of MacKenzie et al.

A model for cerebral tumor surgery

Jannin et al. [Jannin et al. 2003; Raimbault et al. 2005; Jannin and Morandi 2007] proposed a hierarchical procedure model for cerebral tumor surgeries in the context of image-guided surgery. The objective of their work was to provide enhanced support for surgical planning with the help of a generic model of surgical procedures, which consists of a mixture of classes of different kinds. There are hierarchical classes for decomposing the procedure, limited to two granularity levels: *surgical procedure* and *step/action* (step and action are in a one-to-one correspondence). Furthermore, informational classes were enclosed to indicate supplemental image-related data (e.g., image entities or pathological, functional, or anatomical concepts) (cf. Figure 2.1.7). The data was acquired offline pre- and post-operatively with the help of questionnaires and assessed afterwards. The proposed procedure model represents a top-down approach and accounts for the differentiation of planned and performed work steps.

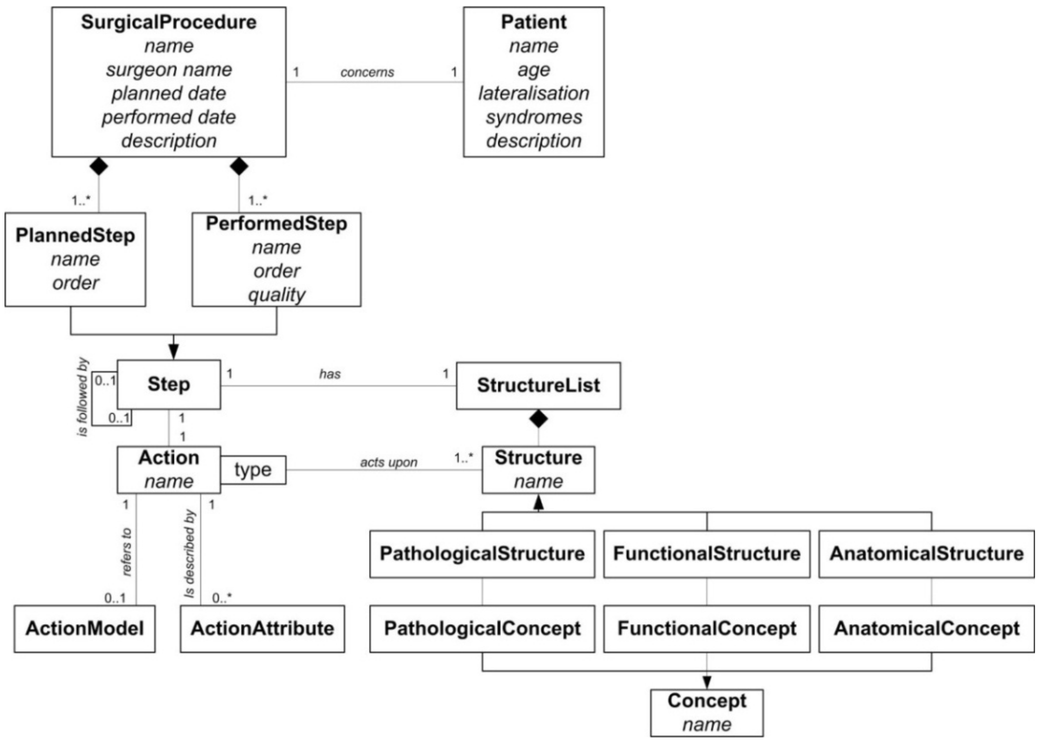


Figure 2.1.7: Procedure model proposed for cerebral tumor surgeries (cutout) [Jannin et al. 2003].

However, the model does not allow for expression of parallel or iterative surgical work steps and does not include temporal information. Also missing are specific relation cardinalities, and mandatory and optional entities cannot be distinguished from one another. Nevertheless, this is the first approach to include surgical expertise and an ontological foundation, represented as a UML model, and used as the basis for a database. This fact distinguishes this approach from the work of MacKenzie et al. (above) and Ahmadi, Padoy et al. (below).

The process-related parts of the model of Jannin et al. are easy to "translate" into the formal framework suggested in this paper, as there is no real hierarchical order. At first glance, there seem to be three levels of granularity that can be perceived: the *surgical procedure* itself (as the highest), the *step* (as the middle), and the *action* (as the lowest level). However, as mentioned above, only two hierarchical levels can be employed.

The surgical procedure is broken down into a sequential list of surgical steps (see Figure 2.1.8). Each of these steps is then associated with a single action. Regarding the examples given by Jannin et al., it becomes clear that each *action* is actually a generalization of the corresponding *step*. For instance, the *steps* "transgyral approach" and "transsulcal approach" are both associated with the *action* "to approach". This is a sensible solution for the authors' specific purpose. For our purposes, however, we proceed on the assumption that this model has two granularity levels, namely *surgical procedure* and *surgical step*, including *action* into the latter by modeling it as an attribute of the step. The two granularity levels can be compared conceptually to λ_5 (surgical procedure) and λ_4 (steps) in the model by MacKenzie et al.

Other elements of Jannin's model can be adequately represented as attributes in our approach, in particular ActionModel, ActionAttribute, PlannedStep, PerformedStep, and Structure and its Subclasses. *Incident* is not described in detail in the available publications, but the name suggests that our achievements should be used. While some specific elements (such as ImageEntity) cannot be covered reasonably, we still conclude that the formalization developed in this paper could well be applied to this model. This should even apply to extensions with temporal information and new granularity levels.

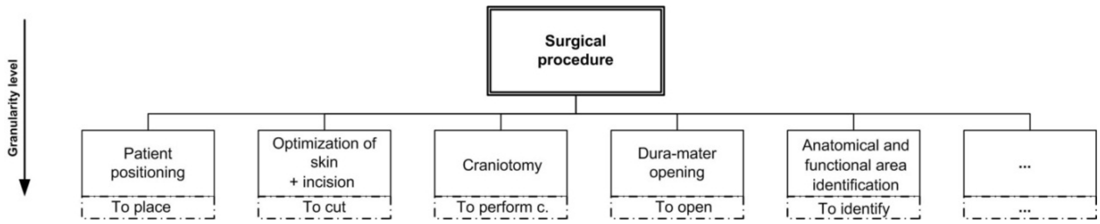


Figure 2.1.8: The two granularity levels, 'surgical procedure' and 'surgical step', identified from the model by Jannin et al. [Jannin and Morandi 2007].

A model applicable to multiple surgical disciplines

The approach described by Neumuth et al. [Neumuth et al. 2009b; Neumuth et al. 2006a] is aimed at developing surgical-assist systems and integrated operating room control systems based on SPMs. Neumuth et al. describe concepts and technologies for the acquisition of surgical process models by monitoring surgical interventions. Furthermore, they subdivide surgical interventions into work steps at different levels of granularity and propose a recording scheme for the acquisition of manual surgical work steps from interventions in progress.

Trained observers record the surgical interventions live in the OR. They are supported by a software architecture, backed by dedicated ontologies, that has been devised by the authors. Live and offline recordings are possible with this method.

The drawbacks of this approach are that the attention span and the reaction time of human observers are limited in live observation settings; consequently, many rapid, consecutive, or simultaneous work steps are hard to keep track of. In addition, information that is not in the field of view cannot be recorded properly. However, this detriment is partly compensated for by the software.

The advantages of this approach are that it includes temporal information and is knowledge-based; in addition, ontologies from different domains, such as the foundational model of anatomy (FMA) [Rosse and Mejino 2003], can be integrated into the model. Additionally, sensor signals can be included. In contrast to the other approaches presented here, the work of Neumuth et al. can be applied independent of the surgical discipline, school, or intervention type and allows for a universal adoption for observer or sensor system based data acquisition.

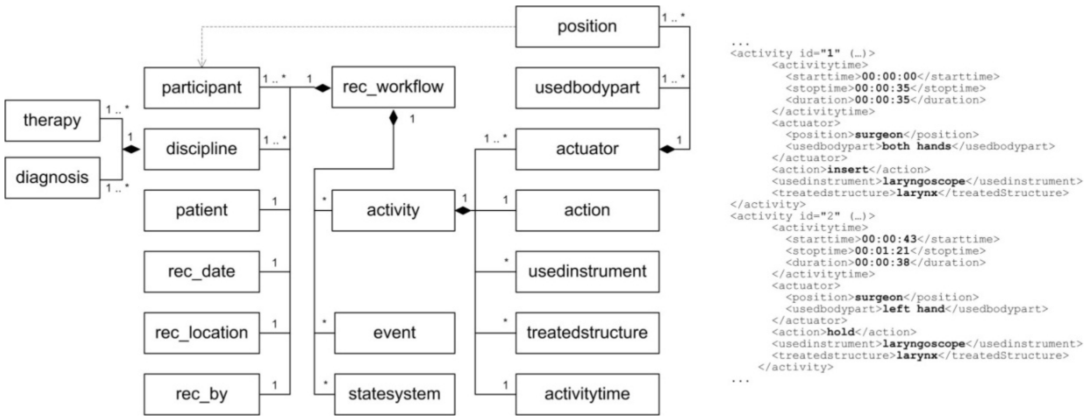


Figure 2.1.9: Generic procedure model proposed by Neumuth et al. [Neumuth et al. 2006a] (left, cutout) and example of activities in an SPM.

Processes at the lowest level of granularity, as identified in these works, are described in terms of attributes and clustered into *perspectives*. There are five possible perspectives, namely the *organizational*, *functional*, *operational*, *spatial*, and *behavioral* perspectives. The behavioral perspective captures explicit temporal information of processes in terms of start and stop times. All other attributes of the perspectives are situated at λ_0 and determine a basic process at λ_1. For instance, a partial specification of activity 1 in the XML-fragment of Figure 2.1.9 amounts to:

- $(L_{P1}, insert/laryngoscope/larynx, \{A_1, A_2, A_3\}, \emptyset, 00{:}00{,}00{:}35)$
- $(A_1, action, insert, 00{:}00{,}00{:}35)$
- $(A_2, instrument, laryngoscope, 00{:}00{,}00{:}35)$
- $(A_3, treatedStructure, larynx, 00{:}00{,}00{:}35)$

Figure 2.1.10 shows an example of several steps of aggregation. These can also be captured in terms of the proposed framework; e.g., $(P_1, cutting, \{E^*_{A,1}\}, \{C_1, C_2, C_3\}, t_1, t_4)$ represents a cutting procedure composed of two cuts, with a period of no cutting in between (the C_2), to be carried out by the surgeon using his/her right hand (A_1). This indicates that the presented framework is sufficient to handle the approach of Neumuth et al.

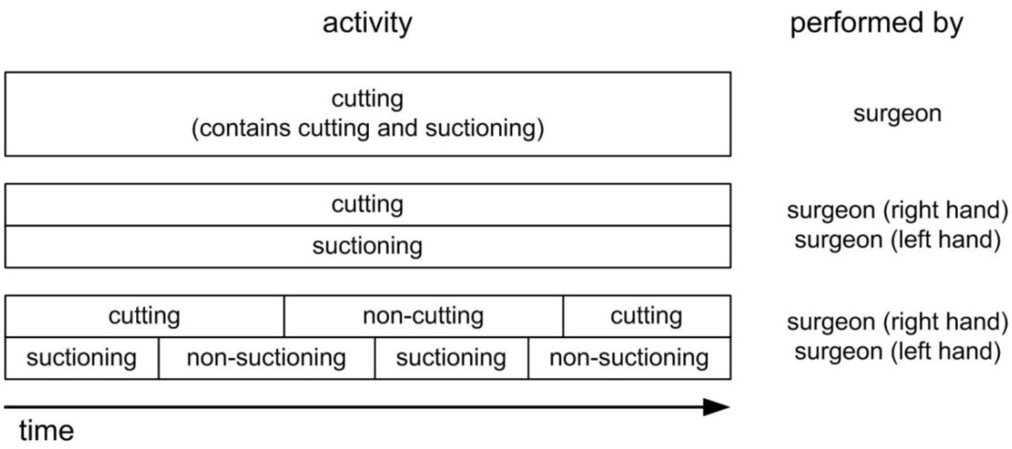


Figure 2.1.10: Example of different granularity levels.

A model for laparoscopic cholecystectomies

Ahmadi and Padoy [Ahmadi et al. 2006; Padoy et al. 2007; Padoy et al.] proposed a method for the determination of surgical phases based on information obtained from sensor signals. In contrast to the previously described works, this represents a bottom-up modeling approach, segmenting surgical workflows into 14 interventional phases by a temporal synchronization of multidimensional state vectors. 17 surgical instruments were used to record a binary model for instrument usage. Every instrument can acquire two states:

$$u(t) = \begin{cases} 1 \text{ if the instrument is used at the time } t \\ 0 \text{ if the instrument is not used at the time } t \end{cases}.$$

Thus, the approach aims at the automatic detection of phases within a surgical procedure by assessing instrument usage. Until now, the work has only been applied to laparoscopic cholecystectomies, a method to remove the gallbladder. However, it requires a reference model for synchronization that yields segmentation, and the sensor signals are not obtained automatically. Information about treated structures and the detection of performed actions are not included in the approach.

This approach is the first to use live signals from the OR to detect intervention phases, later also supported by color and clip detection and an endoscopic camera signal. However, according to the authors themselves, it detects some phases according to the two previous phases and the upcoming phase [Padoy et al. 2007]. The fact that a "future" phase is used speaks against live detection. In addition, the overall number of surgeries processed is not very high.

The overall approach of Ahmadi and Padoy yields three relevant model elements, namely individual signals, phases, and the surgical intervention itself, and it can be described profitably by the means presented in this paper.

The recording of the binary values of 'instrument used' or 'instrument not used' (the graph displays a baseline or a peak of variable height), as shown in Figure 2.1.11, is understood to provide attribute stages for the overall intervention from the very beginning to the end. The course of the signal values is reflected in the function argument of an attributive stage in our framework. Subsequently, the authors calculated phase boundaries and aggregated temporal fragments of these attributes into phases. Clearly, the temporal attribute T is immediately applicable to this approach. Given the phase allocation, the surgical intervention can easily be described in our framework with those sequential phases as components, and each phase can be characterized by all instrument attribute stages. Note that these phases form basic processes at a random ($\iota -$) level of granularity, which is mainly due to the characteristics of the recording methods. They also involve many more entities and aspects than do processes described in a top-down fashion.

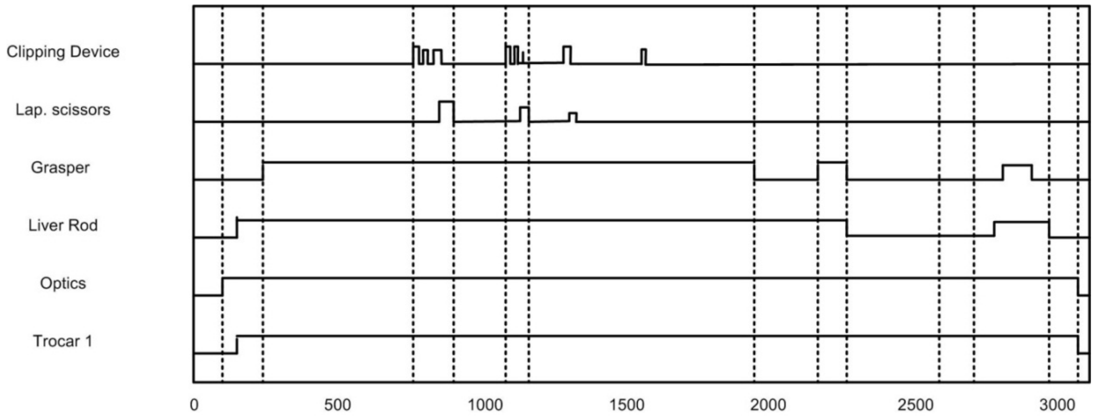


Figure 2.1.11: Instrument usage diagram for cholecystectomies [Ahmadi et al. 2006; Blum et al. 2008a] (cutout).

Discussion

The goal of this paper is to present a four-level framework that is ontologically founded and can serve as a basis for the formal representation of surgical processes models for statistical analysis and data mining. The approach closes the gap between data and knowledge in the domain by using a linguistic approach.

In recent years, several different approaches to structuring and modeling surgical interventions have been proposed. Each of these attempts uses its own constructs, and the variety of different underlying conceptual systems impedes the comparability of results, the exchange of data, and their unified interpretation. We developed the approach to allow a data exchange between different groups working in the respective domain, which was not available before. The focus was not to provide a general applicable framework, but rather to provide a formalism that can be applied to the few existing approaches for modeling SPMs. The expressive power and the representational capacity of our framework were demonstrated by applying it to four recent and frequently cited approaches to surgical procedure modeling. Further related approaches could not be taken into account due to spatial constraints. Regardless, our approach is well adapted for the domain of surgical process modeling, as the evaluation has shown where all four approaches could be reconstructed within our framework.

To bridge the gap between data and knowledge, we used verbs as process categories for the selective representation of knowledge over time-distributed data of surgical processes. By using this ontological view, we can cover the entire processes. By measuring different aspects of these processes we select distinguished attributes for subsequent mathematical modeling. However, the origin of our approach is based on the ontological view, where the attributes need to have meanings that can be ontologically derived.

The reconstructions capture in detail the temporal structures of processes, offering a high degree of expressiveness. Arbitrary relationships between the temporal parts of an intervention can be represented, for instance, including concurrency and branching. This could already be achieved by requiring time stamps for all temporal entities (stages and processes), but that temporal information is not available for all approaches. For instance, Jannin et al. [Jannin et al. 2003] lack time-stamped data (due to differing goals for their work). In order to cover approaches without explicit temporal data, a sequential ordering of processes was added to our framework. The expressiveness of our method suffices to cover the application cases but could be further extended to partial orders or preorders. An even greater extension would be to allow for temporal variables in the formalism, which further implies the need for a constraint language on these variables (e.g. in order to cover relationships such as the temporal ordering of time stamps ($t_1 \leq t_2$)). However, this further increases the complexity of the model. For data acquisition purposes, we would recommend restricting the dataset to explicitly time-stamped data. An extension by variables is a step towards explicitly modeling gSPMs, as well as patterns to express iterations and concurrencies, even though the latter remain implicit in the time-stamped data thus far. Alternatively and depending on the respective purpose, the adoption of other process representation formalisms is worth being reconsidered, when experimentally observed gSPMs are derivable from the data collected on the basis of the presented model.

The second consideration concerning the temporal structure of surgical processes is to assign to them different levels of granularity. Three of the four application cases define a fixed hierarchy of granularity levels. First, we believe that the notion of granularity as such, and in its particular combination with processes, requires further (ontological) analysis. One central question is whether discrete levels of granularity can be assumed or whether they should be regarded as the discretization of a continuous notion of granularity. The presented framework follows two strategies. The formal granularity functions λ and μ account for variable hierarchical levels of granularity that are oriented at the modeling primitives, which are chosen when the framework is applied. This is suitable for representing the four application cases. In contrast to λ and μ, the content-oriented granularity relation ι is intended to serve as a simple, preliminary means of comparison across different models and application cases. In future work, terminologies of surgical procedures may be employed as a basis for global granularity comparison. Another future goal related to granularity is to better understand the principles of distinct methods of process modeling, in particular of (and between) the poles of top-down and bottom-up modeling.

The non-temporal aspects of surgical processes are covered uniformly in terms of attributes in the presented framework. This abstraction is fairly strong and exhibits some similarity with the process specification language [Schlenoff et al. 1999], which also offers only one basic non-temporal modeling element. This choice promotes the uniform analysis of temporal patterns of surgical processes. On the other hand, some distinctions regarding non-temporal aspects are available in other approaches (including the norm EN 1828 [European Committee for Standardization (CEN) 2002]), e.g., between anatomical and instrumental participants in an intervention. If these distinctions are to be maintained during conversion into the mathematical framework, additional, unambiguous guidelines and conventions must be established that specify their encoding. It is reasonable to incorporate a more detailed model of non-temporal aspects in order to extend the framework. At the present stage, however, further analysis concerning which of these aspects are universally applicable to arbitrary interventions is required.

Besides the formal level, at which the proposed framework is located, our basic methodology comprises three additional levels, including the second (i.e., the ontological/conceptual) level. This requires further development and refinement. A particularly important future task is the ontological analysis and declarative formalization of the theory of eventualities. Eventualities can generally be interpreted as process categories that are related to verbs. The precise ontological foundation, also in connection with top-level ontologies, is not yet complete. We encountered some complexities that have their origins in linguistics and hence in the usage of natural language. One question remaining, for instance, is: ‘How can gradual developments and the precise moment when a goal is reached (achievement) be expressed?’ In the pertinent literature, there are no complete solutions to this problem, as natural language cannot be described as a clearly framed set of rules.

From the given formal representation of eventuality types, a similar situation arises as was discussed for granularity above. Many intervention models are constructed top-down and rely only on natural language labels for the phenomena that are captured (in addition to temporal aspects). Those labels are reflected in the mathematical framework as stages with constant functions only. Given the built-in formalization of eventuality types, this leads to many states in intervention models. From a content perspective, such states may well be of a different nature -

accomplishments or achievements, for instance - where that nature is hidden in the original natural language expression. This must be taken into account for evaluations according to eventuality types. In the context of our application cases, the distinction between different eventuality types is rare in current models. Nevertheless, we believe that the expressiveness of our framework will allow for more fine-grained statistical analyses and/or data mining of surgical procedure records.

The presented approach needs further development to meet requirements of future applications, such as a mapping onto established logical formalisms that allow reasoning for treatment planning or decision support systems, the ontological basis of the delineated formal-mathematical elements is to be extended and can be linked with a top-level ontology, such as the general formal ontology (GFO, [Herre et al. 2007]). An explicit model of the eventualities and of the overall model in the web ontology language (OWL) [European Committee for Standardization (CEN) 2002; Horrocks et al. 2003; van Harmelen and Antoniou 2003] would be useful for the context of the semantic web. In addition, the explicit definition of semantic relations between the basic entities on a linguistic grounding is conceivable, as well as the integration into the surgical ontologies for computer assisted surgery (SOCAS) [Mudunuri et al. 2007], which was developed in a related project. Finally, an incorporation of the framework into an interactive knowledge base will be attempted.

Conclusion

This work presents an attempt at developing a unifying framework for generating surgical process models (SPMs) that is ontologically founded and formally and mathematically precise. Our aim is to create a common basis for the different and varying approaches in this field. With the help of sample instantiations, it was demonstrated that the proposed framework applies to four major approaches. Thereby, we have shown that it is possible to syntactically adapt very different approaches and to render them interoperable and comparable. In effect, the value of data from surgical processes can be increased by using this framework. Its ontological foundation arises within a novel four-level methodology. The well-established theory of eventualities is initially adopted for process classification.

The growing number of recent studies based on surgical workflows and time-action analyses shows the rising interest in this subject area. That interest can be accounted for by the multitude of possible applications from both the technical and the clinical points of view. Some examples are the evaluation of surgical-assist systems or surgical skills, the design of technical support systems for the operating room, the conception of surgical knowledge bases and the generation of knowledge from them, the planning of interventions, requirement analyses, and so forth. For all of these applications, surgical process models could be more useful if they were designed according to a common basis. The formal framework and the embedding methodology presented here provide a coherent and rigorous contribution towards this end.

Acknowledgements

The Innovation Center Computer Assisted Surgery (ICCAS) at the Faculty of Medicine/Universität Leipzig is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the scope of the initiative Unternehmen Region with grant numbers 03 ZIK 031 and 03 ZIK 032 and by funds of the European Regional Development Fund (ERDF) and the state of Saxony within the frame of measures to support the technology sector.

Symbols

Symbol	Meaning
_	blank argument position
$<, \leq$	(linear) ordering of time stamps
A	attribute, $A = (L_A, V_A)$
ATT	set of all attributes for describing processes
AS	set of attribute eventuality systems characterizing a process
C	process (in connection with being a component of another process)
CS	set of parts/components, which are processes themselves
E	eventuality
E^*	eventuality system
E_A^*	eventuality system for an attribute A
$E_{A,i}^*$	i-th attribute eventuality system (within a collection)
E^{act}	eventuality: activity
E^{acc}	eventuality: accomplishment
E^{ach}	eventuality: achievement
E_A^{ach}	eventuality: achievement based on attribute A
$E_A^{ach,*}$	set of A-achievements
E^{sta}	eventuality: state
EVT	set of all tuples representing eventualities
λ	local granularity, $\lambda: EVT \rightarrow \mathbb{N}$
λ_i	i used as local granularity value
ι	global granularity
$(P, L_P, AS, CS, t_1, t_2)$	representation scheme for processes
$(P, L_P, AS, \emptyset, t_1, t_2)$	representation scheme for basic level process
L_P	type label of process P
L_A	label of attribute A
μ	model granularity, $\mu: EVT \rightarrow \mathbb{N}$
μ_{max}	maximum local granularity
μ_{path}	length of the shortest containment path from a part to its root process
P	process
S_A	stage based on attribute A
S_A^*	set of A -stages
(S_A, L_A, f, t, t')	A -stage (structured form) development of an attribute A with label L_A over time interval

	$[t, t']$
t, t', t'', t_i	time stamps
$[t_1, t_2]$	time interval; $t \in [t_1, t_2]$ iff $t_1 \leq t \leq t_2$
T	special attribute for temporal values
$u(t)$	instrument usage in [74]
$undefined_A$	special A -value without interpretation regarding contents
V_A	value space of attribute A
V_T	value space of temporal attribute T

2.2 Design of similarity metrics for surgical process models

Title

Similarity metrics for surgical process models

Authors

Thomas Neumuth[1], Frank Loebe[2,3], Pierre Jannin[4,5,6]

Departments

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] Universität Leipzig, Department of Computer Science, Intelligent Systems, Leipzig, Germany

[3] Universität Leipzig, Institute of Medical Informatics, Statistics and Epidemiology (IMISE), Research Group Ontologies in Medicine and Life Sciences (Onto-Med), Leipzig, Germany

[4] INSERM, U746, Faculty of Medicine, Rennes, France

[5] INRIA, VisAGeS Unit/Project, Rennes, France

[6] University of Rennes I, CNRS, UMR 6074, IRISA, Rennes, France

Citation

Neumuth T, Loebe F, Jannin P. Similarity metrics for surgical process models. Artificial Intelligence in Medicine. 2012; 54(1):15-27.

Keywords

Medical Informatics; Weights and Measures; Workflow; Decision Making, Computer-Assisted; Surgery, Computer-Assisted; Surgical Process Model

Abstract

Objective: The objective of this work is to introduce a set of similarity metrics for comparing surgical process models (SPMs). SPMs are progression models of surgical interventions that support quantitative analyses of surgical activities, supporting systems engineering or process optimization.

Methods and Materials: Five different similarity metrics are presented and proven. These metrics deal with several dimensions of process compliance in surgery, including granularity, content, time, order, and frequency of surgical activities. The metrics were experimentally validated using 20 clinical data sets each for cataract interventions, craniotomy interventions, and supratentorial tumor resections. The clinical data sets were controllably modified in simulations, which were iterated ten times, resulting in a total of 600 simulated data sets. The simulated data sets were subsequently compared to the original data sets to empirically assess the predictive validity of the metrics.

Results: We show that the results of the metrics for the surgical process models correlate significantly ($p<0.001$) with the induced modifications and that all metrics meet predictive validity. The clinical use of the metrics was exemplarily, as demonstrated by assessment of the learning curves of observers during surgical process model acquisition.

Conclusion: Measuring similarity between surgical processes is a complex task. However, metrics for computing the similarity between surgical process models are needed in many uses in the field of medical engineering. These metrics are essential whenever two SPMs need to be compared, such as during the evaluation of technical systems, the education of observers, or the determination of surgical strategies. These metrics are key figures that provide a solid base for medical decisions, such as during validation of sensor systems for use in operating rooms in the future.

Introduction

Objectives and motivation

The objective of this work is to introduce a set of similarity measures for comparing SPMs and to describe an approach for experimental validation of these measures.

Surgical process models (SPMs) are progression models of surgical interventions that are used in a variety of cases, including the optimization and evaluation of computer-assisted surgery systems and requirements engineering. Because surgical interventions, which can be more abstractly defined as surgical processes, cannot be analyzed directly, SPMs are used as computable process models that allow for quantitative analysis. As formal representations of surgeons' activities collected from patient data, SPMs have a great deal of potential in surgical education and training.

Similarity metrics are required for quantitatively describing the similarities among multiple SPMs. Such comparisons can serve as an advanced method for representing surgical treatment strategies [Neumuth et al. 2011b; den Boer et al. 1999], a retrospective evaluation of surgical assist systems [Neumuth et al. 2011d], assessment of a surgeon's expertise [Riffaud et al. 2010], the basis for requirements engineering [Neumuth et al. 2009c], or the evaluation of the accuracy of SPM acquisition systems, as shown in this paper with human observers [Neumuth et al. 2009b; Neumuth et al. 2010].

The research question is "How can we quantify the similarity between two surgical process models?" To answer this question, we introduced a measurement system that aims to study different SPM dimensions, such as granularity, content, time, order, and frequency. For each of the metrics, mathematical proofs and an experimental validation were performed. Validation studies were based on simulations of real clinical cases from three different types of surgeries from different surgical disciplines. In addition to experimental validation, we demonstrated the value of our similarity measures by using them to report on the learning progress of clinical observers that recorded the SPMs. These topics have not yet been considered in the literature and may serve as a basis for future work in this field and in affiliated sectors of research.

Modeling surgical processes is a complex task. Established sources of intraoperative knowledge that may provide information about surgical processes, primarily surgical textbooks or clinical guidelines [AHRQ-Agency for Health Care Research and Quality 2010a; AWMF-Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. 2010a], have major constraints that engender the use of SPMs. They are either designed for representing expert knowledge in a top-down-modeling approach, but they do not cover concepts that are necessary for quantification, such as temporal constraints, or they are not able to cope with high inter-patient and inter-surgeon variability in the surgical processes. Because no suitable models exist, no metrics exist.

Approaches for modeling surgical processes have gained recent interest in the literature. Entire SPMs have been modeled by several groups in several application contexts, such as surgical education and surgeon training [MacKenzie et al. 2001; Cao et al. 1996], image-guided surgery [Jannin et al. 2003; Raimbault et al. 2005; Jannin and Morandi 2007], context-driven user interfaces [Sudra et al. 2007; Padoy et al.], the evaluation of surgical instruments [den Boer et al. 2002a; den Boer et al. 2002b], the performance of requirements analyses [Neumuth et al. 2009c], and other assessments of surgical strategies or auxiliaries [Neumuth et al. 2009b; den Boer et al. 1999; Sjoerdsma et al. 2000; Minekus et al. 2003]. However, none of the approaches dealt with SPM metrics. None of these models used similarity metrics to compare different surgical processes, even though some of them used statistical approaches that are indirectly related to similarity metrics, such as hierarchical clustering [Jannin and Morandi 2007]. Recently, similarity measures have been introduced by Combi et al. [Combi et al. 2009] in the context of clinical workflows. However, the similarity measures were restricted to temporal information and did not consider other dimensions, such as granularity.

In the framework of information systems theory, several studies have been presented in recent years. Bae et al. [Bae et al. 2006b] presented the measurements of similarities between binary trees in business process models. They introduced δ-comparability and a structural comparison of process blocks. In another work, similarities in processes were assessed by subtracting network matrices [Bae et al. 2006a; Bae et al. 2007]. Furthermore, van Dongen et al. [van Dongen et al. 2008] derived predecessor-successor relations as "causal footprints" from event-driven process chains (EPCs) and introduced similarity measures for these relationships. The measures were judged by facial validity.

Van der Aalst introduced number-based fitness measures between traces of event logs for EPCs [van der Aalst 2005] and, in later works, equivalence structures for Petri nets [van der Aalst et al. 2006a; van der Aalst et al. 2006b]. A review of metrics related to process mining can be found elsewhere [Rozinat et al. 2007]. However, none of these existing metrics is focused on surgical process models. Furthermore, they do not consider inputs that differ because of changes in patient-specific properties or treatments that differ because of variation in surgical experience or available surgical technology. Finally, a metric for SPMs must be clinically meaningful because the results need to be used by the surgeon as a clinical end user [Jannin et al. 2006], who must interpret the SPM results with a clinical perspective.

In this paper, we start with a brief introduction of contextual terms and definitions and then present the metrics, first on a general level and then on a formal level. Subsequently, metric properties are mathematically proven and experimentally validated using 20 clinical data sets each from cataract interventions, craniotomy interventions, and supratentorial tumor resections. Finally, we demonstrate the clinical utility of the metrics by assessing the learning curves of observers during surgical process modeling acquisition.

Contextual terms and definitions

A surgical process model represents a surgical process (SP) in the real world as a set of eventualities, which is a general term for (parts of) processes and processual entities [Neumuth et al. 2009b; Neumuth et al. 2011a]. Here, we focus on surgical work steps in SPs and define their representations in SPMs as *activities*. Thus, each activity in an SPM is associated with a surgical work step in the underlying SP. When aligning our terminology to that of the Workflow Management Coalition, our “activities” correspond to their “activity instances” [Workflow Management Coalition 1999a].

Surgical processes that are performed with the same surgical objectives and the same strategies have high variability. This variability is caused by the use of different surgical technologies or by patient-specific properties in anatomy and pathology. To represent this variability, we introduce the concept of perspectives into our SPMs. Jablonski and Bussler [Jablonski and Bussler 1996] introduced the use of perspectives to differentiate among several aspects of activities for workflow management systems. We use perspectives in our application context to decompose activities into more fine-grained entities. Five different perspectives are distinguished (cf. also Figure 2.2.1):

- the functional perspective (*FUN*) describes what is done in a work step,
- the organizational perspective (*ORG*) describes who performs the work step,
- the operational perspective (*OPR*) describes which instruments or devices are used to perform the work step,
- the spatial perspective (*SPA*) describes the location on the patient's body where a work step is performed, and
- the behavioral perspective (*BHV*) describes time information.

The perspectives express who was doing what, where, how and when (cf. Table 2.2.I) for each work step in the surgical process. A number of activities together constitute the surgical process model.

Formally, SPMs, activities and perspectives are captured as follows, starting from the set of perspectives relevant to our purposes. That is, $PE = \{FUN, ORG, OPR, SPA, BHV\}$, where BHV is called the behavioral perspective and all remaining perspectives are defined as nominal perspectives, i.e., those in $PE \setminus \{BHV\}$. Each perspective is associated with a set of possible values, denoted by PE_p^{val} for each perspective $p \in PE$. These value sets are required to be mutually pairwise disjoint. Perspectives span the space of possible activities and the set of which $AC = PE_{FUN}^{val} \times PE_{ORG}^{val} \times PE_{OPR}^{val} \times PE_{SPA}^{val} \times PE_{BHV}^{val}$. Finite sets of activities, in turn, form the surgical process model, $SPM = \wp^{fin}(AC)$, where $\wp^{fin}(x)$ yields the set of all finite subsets of x.

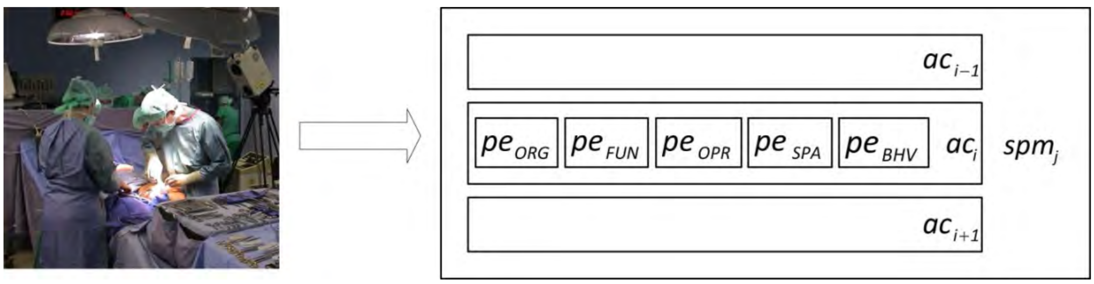


Figure 2.2.1: General principle of using SPMs to present surgical work steps: an SPM spm_j consists of activities ac_i, and each activity consists of several perspectives $pe_p \in PE_p^{val}$ ($p \in PE$).

Some further technical machinery will be useful below. For an activity a, projection functions yield the value of individual perspectives within a, denoted by $\pi_p(a)$ for a perspective $p \in PE$. Two activities a, b are *nominally equivalent* ($a \equiv^{nom} b$) iff $\pi_p(a) = \pi_p(b)$ for every $p \in PE \setminus \{BHV\}$. Nominal equivalence over AC is an equivalence relation (reflexive, symmetric, and transitive), with equivalence classes called *activity types*. The type of an activity is denoted by $type(a)$. Note that the *type* of an activity basically captures its nominal fragment, i.e., for $a = (p^{FUN}, p^{ORG}, p^{OPR}, p^{SPA}, p^{BHV})$, the activity type would be the tuple $(p^{FUN}, p^{ORG}, p^{OPR}, p^{SPA})$. However, the introduction of terms of nominal equivalence yields types as sets of activities and, thus, the membership of each activity in its type, i.e., we have $a \in type(a)$ for every $a \in AC$.

Table 2.2.I: Overview of the perspectives used to describe surgical work steps as activities, with examples.

Perspective	Symbol	Point of view	Example
functional	*FUN*	what is done in the work step	dissect, disinfect, position
organizational	*ORG*	who is performing the work step	surgeon, assistant, scrub nurse
operational	*OPR*	which instruments, devices, or resources are used	forceps, scalpel, trephine
spatial	*SPA*	where the step is performed	dura, cranial nerve, or septum nasi
behavioral	*BHV*	when the step is performed	activity starts at 00:00:30, activity ends at 00:02:40

Concerning the behavioral perspective, its values are expected to be ordered pairs (t^s, t^e) of real numbers encoding the start and the stop time of an activity. Therefore, $t^e \geq t^s$ can be assumed for every activity for which both components are recorded. Moreover, for every activity a where $\pi_{BHV}(a) = (t^s, t^e)$, we define the duration of a by $duration(a) \stackrel{\text{def}}{=} t^e - t^s$.

Metrics for measuring SPM similarity

Outline of similarity metrics

In this section, we introduce five different similarity metrics: **granularity similarity**, **content similarity**, **temporal similarity**, **transitional similarity**, and **transition frequency similarity**. Each of these metrics represents one or more of the following dimensions: granularity, content, time, and order of surgical activities in SPMs.

Prior to formulating the intended similarity metrics in general and technical forms, let us provide some illustrations for each of them by means of Figure 2.2.2. As a preliminary simplifying assumption, an SPM can be considered to be a sequence of activities.

The first metric, **granularity similarity**, is concerned with the presence and structure of activities in the two compared SPMs. In the presence of an alignment between fragments (sub-sequences) of both SPMs that reflect corresponding parts of the underlying SPs, differences in the numbers of activities in such aligned fragments are utilized for quantifying process granularity differences.

Content similarity, the second metric, focuses on aligned pairs of single activities in each SPM and considers the differences in values between their nominal perspectives. Two nominally equivalent activities will exhibit a content distance value of 0. Complementary to content similarity, **temporal similarity** is concerned with differences in the durations of aligned pairs of activities.

The final two similarity measures focus on the sequences of types of activities, which are derived from direct transitions between two (or more) activities. In the corresponding parts of Figure 2.2.2, capital letters indicate types of activities. Accordingly, in SPM_1, there are three activity transitions (of length 2, indicated by the vertical lines), for which the type sequences read as AB, BA, and AB. **Transitional similarity** determines the extent to which such type sequences overlap in both SPMs without considering the frequency of the corresponding transitions in activities. The latter is taken into account by the **transition frequency similarity**, which is indicated by the numbers in the right-most component of Figure 2.2.2 and specifies the number of occurrences of a type of transition.

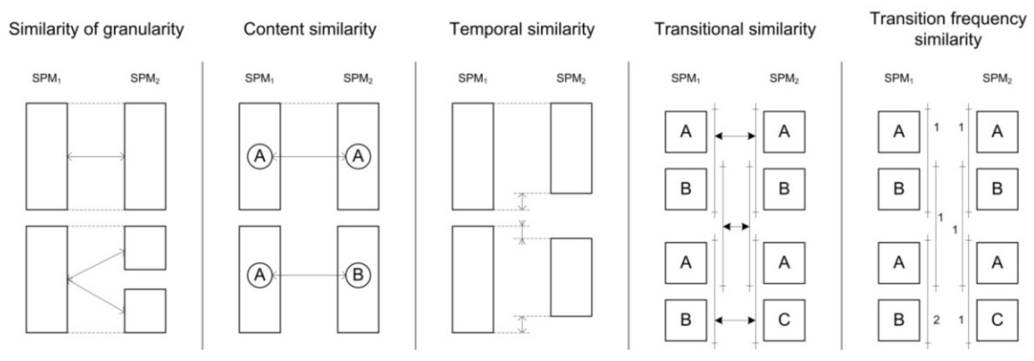


Figure 2.2.2: Schematic principles of the similarity metrics.

Definitions of similarity metrics

All similarity metrics are defined in terms of corresponding distance functions, namely granularity, content, temporal, transitional, and transition frequency distance. By derivation from the purpose of measuring similarity, we refer to these functions as distance and similarity functions (or measures) interchangeably. Technically, a number of prerequisites are commonly assumed in surgical process modeling. Moreover, these prerequisite assumptions are necessary to prove that the measures defined obey the properties of a metric. These assumptions are presented next, in preparation for subsequent definitions.

The first major assumption is the existence of *registration mappings* that link the activities of one SPM to those of another. A mapping between the two SPMs s and t is formally captured as the relation $m \subseteq s \times t$. These mappings are necessary for defining granularity, content and temporal distance measures. The property of these three measures as a metric then applies with respect to an overall set of SPMs, where for any two members a corresponding registration mapping is required. That family of mappings must obey a number of technical properties, as well, which are specified in detail in the "Metric Properties" section (cf. the definition of "registration context"). Intuitively, the self-mappings of any SPM must be reasonably close to the identity mapping, the mappings between SPMs s and t must be inverses of each other, and all mappings must be "transitive" or compatible with each other, such that following two mappings consecutively from s over t to u yields the same result as following the mapping from s to u directly.

The precise versions of these properties in the general case are actually more complex, especially for transitivity, as demonstrated in the "Metric Properties" section. Here, it is sufficient to note that the characterizations involve viewing a registration mapping $m \subseteq s \times t$ (by reinterpretation of m as a set of *unordered* pairs) as a registration graph $g = (s \cup t, m)$. This step allows us to refer to connection components of a registration graph where $cc(g)$ denotes the set of connection components of g. The connection components are highly relevant for the notion of granularity distance. In particular, remembering the ordering in the pairs of m, the notion of a *type* of a connection component c of g is introduced as the pair $k{:}\,l$, where k is the number of activities of s involved in c and l is the number of t-activities in c. Note that in general, types $k{:}\,l$ for arbitrary non-negative integer values of k and l, including 0, are possible in this setting. Types with a 0 component result from a situation in which a protocol s may lack activities that would correspond to certain activities in the comparison SPM t. Moreover, the properties mentioned above may lead to viewing a group of several activities as corresponding to a single gap in another SPM.

The question of how registration mappings are established among SPMs is irrelevant from the perspective of defining distance functions and proving them to be metrics. Nevertheless, the question of the motivation for those properties required for mappings and the question of whether there are sets of SPMs and corresponding mapping sets that satisfy these properties are natural. In our case, such sets of SPMs and mutual mappings arise as follows. The original intended application of all similarity measures was to compare SPMs to a pre-determined reference SPM. The latter can act as a crystallization point for its own metric space in the following sense: all SPMs are aligned to the reference SPM, e.g., through manual alignment. All further mappings between any two SPMs s and t are then derived from the mapping of s and t to the reference SPM. By this construction, all properties of a registration context (see the "Metric Properties" section) are implicitly satisfied.

Granularity similarity

The first measure indicates the granularity compliance of two SPMs. The metric refers to the identification of the number of activities a given real-world process is resolved into in an SPM. The registration mapping is of great importance here because it provides a relational comparison when no objective measures are available. From a mapping graph of two SPMs and, more precisely, from each of its connection components, two parts of the SPMs can be identified as the sets of activities involved in exactly that connection component. We assume that those two parts of the distinct SPMs are "equivalent" (or at least comparable) in the sense that they capture mutually corresponding parts of the original surgical processes.

Accordingly, the type of connection components serves as the basis for measuring granularity differences. For instance, a 1:1 connection component means that, in both SPMs, one activity is used to capture a certain part of the processes. $0{:}\,l$ or $k{:}\,0$ components represent the case for which one SPM lacks parts that the other exhibits. If $k < l$ in the general case of type $k{:}\,l$ of a component, the first SPM seems to cover equivalent process parts in less detail than the second.

A simple approach for measuring granularity along these lines is to start by defining granularity distances of the connection components. This distance is well defined because the type of a component is uniquely determined. Summation yields a measuring function $DG{:}\,SPM \times SPM \rightarrow \mathbb{R}^{+}$ at the level of the overall SPMs. The notation $|x|$ refers to the absolute value of x if x is an arithmetical expression, whereas it denotes the cardinality in case of x being a set.

Def. (*granularity distance*): Let s and t be two SPMs with registration graph g. We define:

$DG^{CC}(c) \stackrel{\text{def}}{=} |k-l|$ for every connection component $c \in cc(g)$, assuming $type(c) = k{:}\,l$.

$$DG(s,t) \stackrel{\text{def}}{=} \frac{1}{|cc(g)|} \Sigma_{c \in cc(g)} DG^{CC}(c)$$

Content similarity

Content similarity compares the two SPMs for the contents of each of the nominal perspectives $p \in$ {FUN,ORG,OPR,SPA} described in Table 2.2.I for activities. This measure is limited to (parts of) SPMs that exhibit registration mappings where all connection components are of type $1{:}\,1$.

The definition of content distance proceeds in three steps. The basis is the content distances of single perspectives, which are aggregated into content distances of activities and SPMs.

Def. (*perspectival content distance of activities*): For every perspective $p \in PE \backslash \{BHV\}$ and for arbitrary activities $a, b \in AC$, we define the function $DC_p^{PE}{:}\,AC \times AC \rightarrow \{0,1\}$, where $\pi_p(x)$ yields the value of perspective p in an activity x (see the "Contextual Terms and Definitions" section):

$$DC_p^{PE}(a,b) \stackrel{\text{def}}{=} \begin{cases} 0 & \mathit{if}\, \pi_p(a) = \pi_p(b) \\ 1 & \mathit{otherwise} \end{cases}$$

Consolidating the perspectival content distances of activities involves a weight vector for all nominal perspectives, $w^{PE} = (w_{FUN}^{PE}, w_{ORG}^{PE}, w_{OPR}^{PE}, w_{SPA}^{PE}) \in [0,1]^4$. The sum of these weight values must be 1. This assignment makes it possible to emphasize some perspectives over others, giving more scope of influence to the user. If, for instance, the correct indication of values for the organizational perspective is more important than other perspectives, the user can simply assign a higher value for this perspective. Another useful case is the ability to focus an analysis solely on one perspective, e.g., the use of surgical instruments and devices. All weights are then set to 0 except for the weight of the perspective of interest (in this example, the functional perspective FUN). By default, however, equal weighting factors are assigned to all perspectives.

Def. ((*aggregated*) *content distance of activities*): Let $w^{PE} = (w_{FUN}^{PE}, w_{ORG}^{PE}, w_{OPR}^{PE}, w_{SPA}^{PE})$ a perspectival weight vector. The content distance based on w^{PE} is the function $DC^{AC}: AC \times AC \to [0,1]$ such that, for two arbitrary activities $a, b \in AC$:

$$DC^{AC}(a,b) \stackrel{\text{def}}{=} \sum_{p \in PE \setminus \{BHV\}} w_p^{PE} \cdot DC_p^{PE}(a,b)$$

Similar to weighting perspectives within an activity, it is sometimes useful to differently weight, for example, the temporal positions of activities within an SPM. This weighting is captured formally in terms of a set of aligned total orderings of the SPM activities. Note that all registration mappings must be bijective functions because of the limitation of a 1:1 connection component, as stated above. Thus, all mappings exhibit the same number of mapping pairs, and consequently, all SPMs have the same number of activities, represented by n. Then, one can determine sets of aligned strict total orderings over all SPMs.

Def. (*aligned ordering*): Let $s, t \in SPM$ be two SPMs and m_{s-t} be their registration mapping. The ordering relations $\prec_s$ and $\prec_t$ are aligned if for arbitrary activities $a', a'' \in s$ and $b', b'' \in t$, such that $(a', b') \in m_{s-t}$ and $(a'', b'') \in m_{s-t}$ meet the condition $a' \prec a''$ if and only if $b' \prec b''$ is satisfied.

On the basis of such aligned orderings, one can meaningfully refer to a set of global selection functions for activities, $\{\sigma_i: SPM \to AC \mid i \in \{1,2,\dots,n\}\}$, where $\sigma_i(s)$ denotes the i^{th} activity in SPM s according to the chosen orderings. Moreover, we introduce a weight vector $w^{OD} = (w_1^{OD}, \dots, w_n^{OD}) \in [0,1]^n$ for activities (n is the number of activities in each SPM; the label OD is derived from "ordering"). Again, the primary goal is to rank temporal phases of a surgical process as being more important than others. Activities in the surgical preparation stages can be emphasized as compared with activities during patient scheduling, for instance. The default for the general case is to assign equal values to all w_i^{OD}, where $i \in \{1, 2, \dots, n\}$.

Def. (*content distance of SPMs*): Let $\{\sigma_i: SPM \to AC \mid i \in \{1,2,\dots,n\}\}$ a set of global selection functions (based on a corresponding set of pairwise aligned orderings), w^{PE} a perspectival weight vector, and w^{OD} an ordering-based weight vector. On that basis, the content distance of SPMs is a function $DC: SPM \times SPM \to [0,1]$, defined as follows. Let $s, t \in SPM$ with $|s| = |t| = n$, then

$$DC(s,t) \stackrel{\text{def}}{=} \sum_{1 \le i \le n} w_i^{OD} \cdot DC^{AC}\big(\sigma_i(s), \sigma_i(t)\big)$$

Temporal similarity

Whereas the behavioral perspective is out of the scope of content similarity, temporal similarity focuses only on that perspective and aims to measure temporal deviations between two SPMs. The limitation of applicability is equal to that of content distance, namely, registration mappings must contain only connection components of type $1:1$. One should also remember the consequences of this restriction in the context of the general preconditions, in particular, a common number of activities in each SPM or registration mapping elements results from these preconditions.

The difference of the durations of activities defines their temporal distance, which is accumulated into the temporal distance of SPMs.

Def. (*temporal distance of activities*): For activities, temporal distance is the function $DTp^{AC}: AC \times AC \to \mathbb{R}^+$ such that for arbitrary $a, b \in AC$:

$$DTp^{AC}(a,b) \stackrel{\text{def}}{=} |duration(a) - duration(b)|$$

Def. (*temporal distance of SPMs*): For SPMs, temporal distance is the function $DTp: SPM \times SPM \to \mathbb{R}^+$ such that, for arbitrary $s, t \in SPM$ with registration mapping m:

$$DTp(s,t) \stackrel{\text{def}}{=} \frac{1}{|m|} \sum_{(x,y)\in m} DTp^{AC}(x,y)$$

Transitional similarity

Transitional similarity aims to provide a key figure for two SPMs on the basis of temporal transitions between their activities. To design this metric, we oriented ourselves with the metric behavioral precision/recall presented previously [van der Aalst et al. 2006a; van der Aalst et al. 2006b]. The main idea here is to compare the number of predecessor-successor relations within activities (or more generally, chains of succession) that are similar in two SPMs.

The formal definition does not rely on registration mappings between SPMs but does require a few additional comments beforehand. First, the notion of "transition" shall be captured formally.

Def. (*n-transition; length*): Let $s \in SPM$ and $\{a_1, \dots, a_n\} \subseteq s$. The sequence $(a_1, \dots, a_n) \in AC^n$ of activities in s is an n-transition of s if, for every i, j with $1 \le i < j \le n$, a_j follows a_i immediately within s, i.e., if $\pi_{BHV}(a_i) = (t_i^s, t_i^e)$ and $\pi_{BHV}(a_j) = (t_j^s, t_j^e)$ then $t_j^s \ge t_i^e$, and there is no activity $b \in s$ starting between t_i^e and t_j^s, i.e., for every $b \in s$ with $\pi_{BHV}(b) = (t_b^s, t_b^e)$, either $t_b^s < t_i^e$ or $t_b^s \ge t_j^s$. n represents the length of the transition.

The set of all n-transitions of an SPM s is denoted by $trs_n(s)$, and the set of all transitions of s (of arbitrary length) is denoted by $trs(s)$. Note that these sets are finite because SPMs are finite sets of activities. Additionally, the notion of transition type is introduced based on activity types (cf. the "Contextual Terms and Definitions section"). The latter are associated intuitively with activities having their behavioral perspective removed.

Def. (*n-transition type*): A subset $\tau \subseteq AC^n$ is an *n-transition type* iff these conditions are satisfied:

- there is a sequence of activity types $(\alpha_1, \ldots, \alpha_n)$ such that $x_i \in \alpha_i$ for every $(x_1, \ldots, x_n) \in \tau$ and $i \in \{1, 2, \ldots, n\}$
- there is no $\tau' \supset \tau$ that satisfies the previous condition applied to τ' instead of τ.

Note that the type of an individual transition $\chi = (x_1, \ldots, x_n) \in AC^n$ is uniquely determined and can thus be denoted $type(\chi)$. In connection with the definition of activity types, the following equation is true: $type(\chi) = type(x_1) \times \ldots \times type(x_n)$. Intuitively, a transition type captures all transitions with a common scheme of activity types.

Def. (*set of n-transition types*): For $s, t \in SPM$, the set of n-transition types occurring in s and t is $\theta_n(s,t) \stackrel{\text{def}}{=} \{type(\chi) | \chi \in trs_n(s) \cup trs_n(t)\}$

Def. (*frequency*): Let $s \in SPM$ and τ an n-transition type. The frequency of occurrence of τ within the SPM s is:

$$freq_\tau(s) \stackrel{\text{def}}{=} |trs(s) \cap \tau|$$

Now, we are prepared to introduce the transitional distance between two SPMs, with respect to single and arbitrary n-transition types.

Def. (*transitional distances*): Let $s, t \in SPM$. The transitional distance of s and t with respect to a single n-transition type τ is the function $DTr_\tau : SPM \times SPM \rightarrow \{0,1\}$ with

$$DTr_\tau(s,t) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } freq_\tau(s) > 0 \text{ and } freq_\tau(t) > 0 \\ 1 & \text{otherwise} \end{cases}$$

The transitional distance regarding all n-types occurring in s and t, $DTr_n : SPM \times SPM \rightarrow [0,1]$, is

$$DTr_n(s,t) \stackrel{\text{def}}{=} \frac{1}{|\theta_n(s,t)|} \sum_{\tau \in \theta_n(s,t)} DTr_\tau(s,t)$$

Transition frequency similarity

The final distance measure extends transitional distances by considering not only the existence of an occurrence of a transition type but also the equality of frequency values of occurring transition types. This extension is captured in close analogy to transitional distance.

Def. (*transition frequency distances*): Let $s, t \in SPM$, τ an n-transition type. The transition frequency distances are defined by the functions $DTf_\tau : SPM \times SPM \rightarrow \{0,1\}$ and $DTf_n : SPM \times SPM \rightarrow [0,1]$:

$$DTf_\tau(s,t) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } freq_\tau(s) = freq_\tau(t) \\ 1 & \text{otherwise} \end{cases}$$

$$DTf_n(s,t) \stackrel{\text{def}}{=} \frac{1}{|\theta_n(s,t)|} \sum_{\tau \in \theta_n(s,t)} DTf_\tau(s,t)$$

Metric properties

The main similarity measures introduced in the previous sections focus on granularity, content and temporal aspects and on transitions and transition frequency of activities (DG, DC, DTp, DTr_n and DTf_n). Mathematically, the measures all satisfy the conditions of a pseudometric over either sets of mutually registered SPMs (in the case of $DG, DC,$ and DTp)or the set of all SPMs (for DTr_n and DTf_n). When formulated for an arbitrary set S with arbitrary elements $s, t, u \in S$ and a distance measure $\Delta: S \times S \rightarrow \mathbb{R}$, the following conditions are satisfied:

- self-distance of 0: $\Delta(s,s) = 0$
- non-negativity: $\Delta(s,t) \geq 0$
- symmetry: $\Delta(s,t) = \Delta(t,s)$
- triangle inequality: $\Delta(s,t) \leq \Delta(s,u) + \Delta(u,t)$

As is well known, non-negativity is a consequence of the other three conditions. Nevertheless, a short argument for the validity of the property by direct reference to distance measures is given. Detailed proofs for the self-distance of 0, symmetry and triangle inequalities are presented below for the interested reader, after the technical prerequisites are specified in detail. Before these detailed explanations, we outline a number of central ideas from these proofs. The content of this section rounds out the formal introduction of the similarity functions, but it is not essential for the remainder of the paper.

First, all distance functions are introduced by rather simple basic functions, e.g., activities, which are aggregated into a normalized sum for entire SPMs. In the cases of DG, DC, and DTp, the pseudometric properties are easy to see for basic functions. The lifting of the functions to SPMs rests on requirements imposed on the mutual registrations within the given set of registered SPMs. Note that these requirements are naturally satisfied in our application context, as demonstrated in the beginning of the previous section. In a nutshell, bijections between SPMs exist with respect to their activities (for DC and DTp) or subsets of their activities (for DG) for each pair of SPMs (in such a set). On that basis, the pseudometric properties of the normalized sums are straightforwardly derived from the validity of those properties for the basic functions.

The proofs for DTr_n and DTf_n are basically equivalent and are similar to those for DG, DC, and DTp in regard to the lifting of the properties shown for the basic functions to their validity at the level of SPMs. Whereas the first three properties are easy to see from the definitions of DTr_n and DTf_n, the triangle inequality can be proven in terms of cardinality and combinatorial considerations for the involved transition types.

Detailed descriptions of preconditions

For the proofs, the preconditions that are roughly outlined at the beginning of the "Outline of Similarity Metrics" section must be developed in greater detail. Classical set theory and mathematical notation is assumed. Concerning unconventional general notation, for any sets S, T and binary relation $r \subseteq S \times T$, $dom(r) \stackrel{\text{def}}{=} \{x \in S | \exists y. (x, y) \in r\}$ denotes the domain of r, and $rg(r) \stackrel{\text{def}}{=} \{x \in T | \exists y. (y, x) \in r\}$ denotes its range. The disjoint union of two sets S, T is denoted by $S \dot{\cup} T$. Remember that $|S|$ denotes the cardinality of S, whereas for an arithmetical expression E the notation $|E|$ refers to the absolute value of the evaluation of E.

Registration mappings

The distance functions DG, DC and DTp rely on registration mappings between SPMs.

Def. (*registration mapping*): Let $s_1, s_2 \in SPM$. A registration mapping is a relation $m \subseteq s_1 \times s_2$ such that $dom(m) = s_1$ and $rg(m) = s_2$.

Frequently, we indicate the domain and range of m as an index, e.g., $m_{s_1-s_2}$.

In mapping SPMs recorded from real-world surgical processes, it is necessary to cover the situation in which an activity in one SPM cannot be mapped to one in a different SPM. For formal simplicity, we assume that registration mappings (and graphs, see below) are completed by pseudo-activities prior to considering or establishing their mapping. In detail, one would have to consider the joint sets of all possible activities and pseudo-activities (say, AC'), $AC^+ = AC \mathbin{\dot\cup} AC'$, and derive a corresponding new set of SPMs, SPM^+. For readability, however, we will keep the distinction between AC/AC^+ and SPM/SPM^+ implicit, except for the definition of the types of connection components below.

Registration graphs

Each registration mapping can be equivalently viewed as a graph.

Def. (*registration graph*): Let $m \subseteq s_1 \times s_2$ be a registration mapping over $s_1, s_2 \in SPM$. The graph $g = (s_1 \cup s_2, m)$, where m is considered to be a set of *unordered* pairs, is referred to as the registration graph corresponding to m.

As notational variants, the SPMs on which g is based can be indicated as either an index, $g_{s_1-s_2}$ or a registration mapping m, g_m.

The functional expression $cc(g)$ denotes the set of connected components of g. Restoring the directedness in the mapping underlying g, we use $dom(g) \stackrel{\text{def}}{=} dom(m)$ $[= s_1]$ and $rg(g) \stackrel{\text{def}}{=} rg(m)$ $[= s_2]$. Likewise, for $c \in cc(g)$ with $c = (v, e)$, we define $dom(c) \stackrel{\text{def}}{=} dom(e)$ $[= v \cap s_1]$ and $rg(c) \stackrel{\text{def}}{=} rg(e)$ $[= v \cap s_2]$. Moreover, each connection component $c \in cc(g)$ can be assigned a type of the form $k{:}\,l$. This assignment indicates the number of activities involved in c (here, explicitly not counting pseudo-activities) and in either of the SPMs.

Def. (*type* of a connection component): Let g_m be a registration graph and $c \in cc(g_m)$.
Let $k = |dom(c) \cap AC|$ and $l = |rg(c) \cap AC|$. Then, the *type* of c is defined as $type(c) \stackrel{\text{def}}{=} k{:}\,l$.

The partitioning of g_m into its set of connection components $cc(g_m)$ yields further partitions of the involved domain and range SPMs, such that registration mappings can be "lifted" to mappings between those components.

Def. (*domain/range components*): Let m be a registration mapping and g_m be its corresponding registration graph. Then, we define:

- the domain components of m: $dcp(m) \stackrel{\text{def}}{=} \{dom(c) \mid c \in cc(g_m)\}$
- the range components of m: $rcp(m) \stackrel{\text{def}}{=} \{\ rg(c) \mid c \in cc(g_m)\}$

Def. (*registration graph component mapping*): Let m be a registration mapping. The relation $m^* \subseteq dcp(m) \times rcp(m)$ defined by

$m^* \stackrel{\text{def}}{=} \{ (dom(c), rg(c)) \mid c \in cc(g_m) \}$ is called the registration graph component mapping of m.

Lemma 1. Let m be a registration mapping, g_m be its corresponding registration graph, and m^* be the registration graph component mapping based on m. Then:

- $|dcp(m)| = |rcp(m)| = |cc(g_m)|$
- $dom(c) = dom(c')$ implies $c = c'$ for all $c, c' \in cc(g_m)$
- $rg(c) = rg(c')$ implies $c = c'$ for all $c, c' \in cc(g_m)$
- m^* is a bijective function.

Proof: All of this follows immediately from the definitions and the fact that $cc(g_m)$ is a partitioning of g_m and every connection component has a uniquely determined domain and range (with regard to m). For instance, assume $c \neq c'$ for arbitrary $c, c' \in cc(g_m)$. Then, c and c' are disjoint graphs, i.e., the sets of their nodes and (thus) their edges are disjoint. Hence, $dom(c) \neq dom(c')$, and $rg(c) \neq rg(c')$.

Registration context

Def. (*registration context*): A registration context is a pair (S, M) of a set $S \subseteq SPM$ and a set of registration mappings M, such that the following properties are satisfied:

(1) for every $s, t \in S$, there is exactly one $m \in M$ such that $dom(m) = s$ and $rg(m) = t$.
(2) for every $m \in M$, $dom(m) \in S$ and $rg(m) \in S$.
(3) for all $m, m' \in M$, $dom(m) = dom(m')$ implies $dcp(m) = dcp(m')$.
(4) for all $m, m' \in M$, $rg(m) = rg(m')$ implies $rcp(m) = rcp(m')$.
(5) if $m \in M$ and $dom(m) = rg(m)$, then for every $c \in cc(g_m)$ there is a $k \in \mathbb{N}$ such that $type(c) = k{:}k$.
(6) for every $m, m' \in M$ where $dom(m') = rg(m)$ and $rg(m') = dom(m)$, $m' = m^{-1}$.
(7) for all mappings m_{s-t}, m_{u-t}, $m_{s-u} \in M$, the corresponding registration graph component mappings satisfy $m^*_{u-t} \circ m^*_{s-u} = m^*_{s-t}$.

These conditions will be referred to as the registration context (RC) conditions in the following section.

Given our usage scenario, these conditions are fairly natural restrictions, as compared with allowing for arbitrary mappings. Note that RC condition (5) is a weakened requirement for identity mapping for self-mappings, and RC condition (7) is a weakening of "transitivity" directly applied to the mappings. The weakening of these requirements is useful for uniformly handling SPMs with or without pseudo-activities. Concerning the latter, note further that the question of how many pseudo-activities are introduced into an SPM within a registration context may be determined by other SPMs, and the aim is to satisfy the RC conditions. In any case, there is only a single enrichment with pseudo-activities for each $s \in S$.

We present several lemmas for the subsequent main proofs.

Lemma 2. Let (S, M) be a registration context. Then, for all $m, m' \in M$, $|cc(g_m)| = |cc(g_{m'})|$.

Proof: Let $s, t, u, v \in S$, and assume $m = m_{s-t}, m' = m_{u-v}$. Then, the following equalities hold: $|cc(g_{s-t})| \underbrace{=}_{\text{Lemma 1}} |rcp(m_{s-t})| \underbrace{=}_{(6)} |dcp(m_{t-s})| \underbrace{=}_{(3)} |dcp(m_{t-u})| \underbrace{=}_{\text{Lemma 1}} |rcp(m_{t-u})| \underbrace{=}_{(6)} |dcp(m_{u-t})| \underbrace{=}_{(3)} |dcp(m_{u-v})| \underbrace{=}_{\text{Lemma 1}} |cc(g_{u-v})|$.

Lemma 3. Let (S, M) be a registration context. For every $s \in S$, there is a unique partitioning $P(s)$ such that $P(s) = dcp(m)$ for all $m \in M$ where $dom(m) = s$, and $P(s) = rcp(m')$ for all $m' \in M$ where $rg(m') = s$.

Proof: First, note that for an arbitrary relation $s, t \in S$, $dcp(m_{s-t})$ and $rcp(m_{s-t})$ yield unique partitionings of s and t, respectively. Because RC condition (3) determines that all mappings with equal domain share their domain components and because RC condition (4) determines that all mappings with equal range have the same range components, it suffices to find two mappings $m_{s-t}, m_{u-s} \in M$ for every $s \in S$ by further reference to any $t, u \in S$ that satisfy $dcp(m_{s-t}) = rcp(m_{u-s})$. For this purpose, let $s, v \in S$ be arbitrary, and consider $dcp(m_{s-s}) \underbrace{=}_{(3)} dcp(m_{s-v}) \underbrace{=}_{(6)} rcp(m_{v-s}) \underbrace{=}_{(4)} rcp(m_{s-s})$, i.e., for $t = s$ and $u = s$, the required condition applies to every $s \in S$.

Lemma 4. Let (S, M) be a registration context with $s, t, u \in S$, and let $m_{s-t}, m_{s-u}, m_{u-t} \in M$. Then, the functions

- $\xi_s: cc(g_{s-t}) \rightarrow cc(g_{s-u})$, defined by $\xi_s(c) = c'$ iff $dom(c) = dom(c')$
- $\xi_t: cc(g_{s-t}) \rightarrow cc(g_{u-t})$, defined by $\xi_t(c) = c'$ iff $rg(c) = rg(c')$

are bijections that satisfy $rg(\xi_s(c)) = dom(\xi_t(c))$ for every $c \in cc(g_{m_{s-t}})$.

Proof: To see that ξ_s is an injective function, let $\xi_s(c) = \xi_s(c')$ for arbitrary $c, c' \in cc(g_{s-t})$. Then, $dom(c) \underbrace{=}_{\text{by def.}} dom(\xi_s(c)) \underbrace{=}_{\text{by ass.}} dom(\xi_s(c')) \underbrace{=}_{\text{by def.}} dom(c')$, and Lemma 1 yields $c = c'$. For surjectivity, observe that, according to Lemma 2, ξ_s is a total injective function between two sets of equal cardinality. Altogether, these observations prove that ξ_s is a bijection. The proof of this property for ξ_t proceeds in exact analogy by replacing dom with rg.

It remains to be shown that $rg(\xi_s(c)) = dom(\xi_t(c))$. Let $c \in g_{s-t}$ be arbitrary, and let $\xi_s(c) = c'$. $c' \in g_{s-u}$; hence, $(dom(c'), rg(c')) \in m^*_{s-u}$. By Lemma 3, there must be a $c'' \in g_{u-t}$ such that $dom(c'') = rg(c')$, i.e., $(rg(c'), rg(c'')) \in m^*_{u-t}$. By RC condition (7), $(dom(c'), rg(c'')) \in m^*_{s-t}$. Because m^*_{s-t} is bijective by Lemma 1 and $dom(c') = dom(c)$ by definition of ξ_s and $c' = \xi_s(c)$, $rg(c'') = rg(c)$. This equality in turn justifies $c'' = \xi_t(c)$ by definition of ξ_t, such that the property $dom(c'') = rg(c')$ turns into $dom(\xi_t(c)) = rg(\xi_s(c))$.

Proofs for granularity similarity

A general assumption for all proofs for DG is the given registration context of (S, M). m_{s-t} denotes the uniquely determined mapping between $s, t \in S$ in M, g_{s-t} or g_m, which is its corresponding registration graph. RC condition (1), ensuring the existence of these mappings in M, is left implicit when using m_{s-t} or g_{s-t}.

Non-negativity

Let $s, t \in S$. For every $c \in cc(g_{s-t})$, $DG(c) \geq 0$ because $type(c) = k{:}\,l$, $k, l \geq 0$. Consequently, $DG(s,t)$ is a sum of non-negative integers divided by a positive integer, i.e., $DG(s,t) \geq 0$.

Self-distance of 0

Let $s \in S$. By RC condition (5), $m_{s-s} \in M$ such that for every $c \in cc(g_{s-s})$, there is a $k \in \mathbb{N}$ such that $type(c) = k{:}\,k$. Hence, $DG^{CC}(c) = 0$. Therefore, the sum over all $c \in cc(g)$ is 0, and $DG(s,s) = 0$.

Symmetry

Let $s, t \in S$. Symmetry follows from RC conditions (4) and (6) because $m_{s-t} = m_{t-s}^{-1}$ entails that $cc(g_{s-t}) = cc(g_{t-s})$ (viewing the graphs as unordered pairs), and for every $c \in cc(g_{s-t})$, we see that $type(c) = k{:}\,l$ (with regard to m_{s-t}) iff $type(c) = l{:}\,k$ (with regard to m_{t-s}). Because $|k - l| = |l - k|$, and by writing c^{s-t} for viewing c as member of g_{s-t} (c^{t-s} with regard to g_{t-s}), $DG^{CC}(c^{s-t}) = DG^{CC}(c^{t-s})$ for all $c \in cc(g_{s-t})$. Hence, the sums in $DG(s,t)$ and $DG(t,s)$ are equal as well as the divisors, and thus, $DG(s,t) = DG(t,s)$.

Triangle inequality

We show that $DG(s,t) \leq DG(s,u) + DG(u,t)$ for arbitrary $s, t, u \in S$. First, consider the level of connection components. Justified by Lemma 3, let $c' \in cc(g_{s-u})$ *and* $c'' \in cc(g_{u-t})$ be arbitrary but defined such that $rg(c') = dom(c'')$. Assume $type(c') = k{:}\,l$ and $type(c'') = l{:}\,m$ ($k, l, m \in \mathbb{N}$); then, $DG^{CC}(c') = |k - l|$, and $DG^{CC}(c'') = |l - m|$. Now, for g_{s-t}, there must be a $c \in cc(g_{s-t})$ such that $dom(c) = dom(c')$ and $rg(c) = rg(c'')$, due to RC conditions (3), (4), and (7). Then, $type(c) = k{:}\,m$. Now, the triangle inequality regarding absolute values of integers passes on to DG^{CC}, i.e., $|k - m| \leq |k - l| + |l - m|$ applies to integers, which then proves that $DG^{CC}(c) \leq DG^{CC}(c') + DG^{CC}(c'')$.

This property transfers to $DG(s,t) \leq DG(s,u) + DG(u,t)$, largely due to Lemma 4. First, let $n \overset{\text{def}}{=} |cc(g_{s-t})| = |cc(g_{s-u})| = |cc(g_{u-t})|$ (equalities that hold according to Lemma 2). Then, the inequality resolves into

$$\frac{1}{n} \sum_{c \in cc(g_{s-t})} DG^{CC}(c) \;\leq\; \frac{1}{n} \sum_{c' \in cc(g_{s-u})} DG^{CC}(c') \;+\; \frac{1}{n} \sum_{c'' \in cc(g_{u-t})} DG^{CC}(c'')$$

Given the functions ξ_s and ξ_t from Lemma 4 (and multiplying by n), it suffices to show

$$\sum_{c \in cc(g_{s-t})} DG^{CC}(c) \;\leq\; \sum_{c \in cc(g_{s-t})} DG^{CC}\big(\xi_s(c)\big) \;+\; \sum_{c \in cc(g_{s-t})} DG^{CC}\big(\xi_t(c)\big)$$

This relationship is now obvious from the triangle inequality for DG^{CC}, and the latter is applicable because $\xi_s(c)$ and $\xi_t(c)$ satisfy the necessary requirement $rg\big(\xi_s(c)\big) = dom\big(\xi_t(c)\big)$ for every $c \in cc(g_{s-t})$.

Proofs for content similarity

(Note the additional preconditions for content distance found in the main text, in particular the restriction to registration mappings with components that are only of type 1:1.)

Non-negativity

$DC(s,t) \geq 0$ is obviously satisfied because DC_p and DC^{AC} yield only non-negative integers and all weights are also non-negative.

Self-distance of 0

By definition, $DC_p^{PE}(a,a) = 0$ for every $p \in PE$ due to $\pi_p(a) = \pi_p(a)$. Hence, $DC^{AC}(a,a) = 0$ because all DC_p^{PE} factors are 0. Thus, $DC(s,s) = 0$ for the same reasons applied for DC^{AC}.

Symmetry

By definition, $DC_p(a,b) = DC_p(b,a)$ for every $p \in PE$. Moreover, $DC^{AC}(a,b) = DC^{AC}(b,a)$ because of the symmetry of DC_p and the independence of the perspective weights w_p^{PE} of a, b; i.e., the summands of $DC^{AC}(a,b)$ and $DC^{AC}(b,a)$ are equal. Along the same lines, $DC(s,t) = DC(t,s)$ follows from the equal summands for $DC(s,t)$ and $DC(t,s)$, which result from the symmetry of DC^{AC}.

Triangle inequality

First, we prove the triangle inequality for $DC_p(a,b)$ for arbitrary $p \in PE$ by contradiction.
Assume $DC_p(a,b) > DC_p(a,c) + DC_p(c,b)$. The only possible case for this inequality is $DC_p(a,b) = 1$ and $DC_p(a,c) = DC_p(c,b) = 0$. The former implies that $\pi_p(a) \neq \pi_p(b)$, and the latter that $\pi_p(a) = \pi_p(c) = \pi_p(b)$. However, both implications together are contradictory.

Next, the triangle inequality for DC^{AC} is shown. As the summands of $DC(a,b), DC(a,c)$ and $DC(c,b)$ are in one-to-one-correspondence (with the nominal perspectives functioning as a shared index set), it is sufficient to justify that every triple of the summands satisfies a triangle inequality. For this purpose, let $p \in PE \setminus \{BHV\}$ be arbitrary, and consider

$$w_p^{PE} \cdot DC_p(a,c) + w_p^{PE} \cdot DC_p(c,b) = w_p^{PE} \cdot \left(DC_p(a,c) + DC_p(c,b)\right).$$

The triangle inequality for DC_p yields $w_p^{PE} \cdot \left(DC_p(a,c) + DC_p(c,b)\right) \geq w_p^{PE} \cdot DC_p(a,b)$.

It remains to lift this expression to the aspired property for DC. Analogously to DC^{AC}, the summands in DC are in one-to-one-correspondence, and we can focus on the case of a single index i to immediately entail the triangle inequality of the overall sum. For every i with $1 \leq i \leq |m|$, the i^{th} summand of $DC(s,u)$ is $w_i^{Ord} \cdot DC^{AC}(\sigma_i(s), \sigma_i(u))$, and the i^{th} summand of $DC(u,t)$ is $w_i^{Ord} \cdot DC^{AC}(\sigma_i(u), \sigma_i(t))$. On the basis of the associativity and commutativity of addition, the sum of these two expressions can be seen as the i-th summand of $DC(s,u) + DC(u,t)$. The weight can be factored out, and the triangle inequality for DC^{AC} can be applied, which completes the proof with the following true condition:

$$w_i^{Ord} \cdot \left(DC^{AC}\left(\sigma_i(s), \sigma_i(u)\right) + DC^{AC}\left(\sigma_i(u), \sigma_i(t)\right)\right) \geq w_i^{Ord} \cdot DC^{AC}\left(\sigma_i(s), DC^{AC}\left(\sigma_i(t)\right)\right).$$

Proofs for temporal similarity

(Note the additional preconditions for temporal distance found in the main text, in particular the restriction to registration mappings with components only of type 1:1.)

Non-negativity

$DTp(s,t) \geq 0$ is obvious from $DTp^{AC}(a,b) \geq 0$ by definition, and hence, only non-negative summands occur.

Self-distance of 0

The restriction to 1:1 mappings means that every mapping between two SPMs is a bijection, as in the corresponding registration graph component mappings. Moreover, RC conditions (1), (6), and (7) enforce the condition that the self-mapping of each SPM can only be the identity mapping. To demonstrate this condition, let $s \in S$, and assume $(a,b) \in m_{s-s}$ such that $a \neq b$. Then, $(b,a) \in m_{s-s}^{-1}$ by RC condition (6), and by RC condition (1), $(b,a) \in m_{s-s}$. RC condition (7) would thus require that $(a,a) \in m_{s-s}$, which contradicts $a \neq b$ if m_{s-s} is a bijection.

Accordingly, all mapping elements of m_{s-s} have the form (a,a) and yield a temporal distance of 0; hence, their sum can also be expressed as $DTp(s,s) = 0$.

Symmetry

First, $DTp^{AC}(a,b) = DTp^{AC}(b,a)$ is obvious from the definition of an arbitrary $a,b \in AC$. This connection can be lifted to DTp. RC condition (6) implies that $(a,b) \in m_{s-t}$ iff $(b,a) \in m_{t-s}$. According to this condition and the symmetry of DTp^{AC}:

$$\frac{1}{|m_{s-t}|} \sum_{(x,y)\in m_{s-t}} DTp^{AC}(x,y) = \frac{1}{|m_{s-t}|} \sum_{(x,y)\in m_{s-t}} DTp^{AC}(y,x)$$
$$= \frac{1}{|m_{t-s}|} \sum_{(x,y)\in m_{t-s}} DTp^{AC}(x,y) .$$

Triangle inequality

The proof that $DTp(s,t) \leq DTp(s,u) + DTp(u,t)$ for arbitrary $s,t,u \in S$ proceeds in strong analogy to that of the triangle inequality for granularity distance. At the level of connection components, justified by Lemma 3, let $c' \in cc(g_{s-u})$ and $c'' \in cc(g_{u-t})$ such that, employing the additional restriction on the types of the connection components, the only edge of c' is (a,a'') and that of c'' is (a'',a'). This assumption satisfies the condition that $rg(c') = dom(c'')$. By RC conditions (3), (4), and (7), there must be $c \in cc(g_{s-t})$ with the only edge (a,a'). Clearly, for all c,c',c'' in these relationships, $DTp^{AC}(a,a') \leq DTp^{AC}(a,a'') + DTp^{AC}(a'',a')$ due to $|duration(a) - duration(a')| \leq |duration(a) - duration(a'')| + |duration(a'') - duration(a')|$, according to the triangle inequality regarding absolute values of integers. This expression suffices to prove the intended triangle inequality from Lemma 4 in exact analogy to the final part of the corresponding proof for granularity distance, which we have therefore omitted. Overall, the normalized sums adhere to the triangle inequality if each triple of the components in that summation satisfies the inequality.

Proofs for transitional and transition frequency distances

These distance measures do not presuppose any given registration context, but they do apply to arbitrary pairs of SPMs. The proofs proceed in strict analogy for DTr_n and DTf_n and are independent of the particular length n of the transitions under consideration. Therefore, the symbol DT stands for either DTr_n or DTf_n in this section. n is arbitrary but assumed to be fixed.

Non-negativity

$DT(s,t) \geq 0$ is obvious from $DT_\tau(s,t) \geq 0$ by definition; hence, only non-negative summands occur, which are normalized by a positive value.

Self-distance of 0

$DT(s,s) = 0$ is immediately clear because every transition type of s occurs with a unique frequency. Hence, all summands of $DT(s,s)$ are 0.

Symmetry

$DT_\tau(s,t) = DT_\tau(t,s)$ follows from the independence of the condition in the definition from the order of s and t. Concerning DT, the summation index as well as the normalization factor are also independent from the order of s and t because $\theta_n(s,t) = \theta_n(t,s)$ by definition.

Triangle inequality

Let $s,t,u \in SPM$, and define $\theta_n^0(s,t) = \{\tau \in \theta_n(s,t) \mid DT_\tau(s,t) = 0\}$ and $\theta_n^1(s,t) = \{\tau \in \theta_n(s,t) \mid DT_\tau(s,t) = 1\}$. Let $|\theta_n(s,t)| = k + l$, where $k = |\theta_n^0(s,t)|$ and $l = |\theta_n^1(s,t)|$. Then, $DT(s,t) = \frac{l}{k+l}$. u may contain transitions of types that appear neither in s nor in t. Let the number of these types be m. These m types are members of $\theta_n(s,u)$ as well as of $\theta_n(u,t)$. Moreover, all types in $\theta_n^0(s,t)$ must occur in both s and t and, thus, in both $\theta_n(s,u)$ and $\theta_n(u,t)$. Hence, $k + m \leq |\theta_n(s,u)| \leq k + l + m$ and $k + m \leq |\theta_n(u,t)| \leq k + l + m$, such that $\frac{1}{k+l+m}$ can serve as a lower bound for the normalization factor in both $DT(s,u)$ and $DT(u,t)$, i.e.

$$\frac{1}{k+l+m}\left(\sum_{\tau\in\theta_n(s,u)} DT_\tau(s,u) + \sum_{\tau\in\theta_n(u,t)} DT_\tau(u,t)\right) \leq DT(s,u) + DT(u,t)$$

A lower bound for the sum of the sums arises as follows. Clearly, transitions of all m types that occur solely in u lead to a distance value of 1 for both $DT_\tau(s,u)$ and $DT_\tau(u,t)$ (assuming τ to be any such type). $2m$ and, for simplicity later, m, thus represent a first lower bound.

Next, consider transition types where s and t differ, i.e., let $\tau \in \theta_n^1(s,t)$. (1) If $DT_\tau(s,u) = 1$, τ contributes at least 1 to the sum of sums. (2) If $DT_\tau(s,u) = 0$, then $DT_\tau(u,t) = 1$ because τ then satisfies the condition of DT equally for s and u. Hence, u must behave like s in relation to t, i.e., differ with respect to the condition of DT. Accordingly, τ contributes at least 1 to the sum of the sums in this case as well. The same cases can be made for $DT_\tau(u,t)$, such that every member of $\theta_n^1(s,t)$ contributes at least 1 to the sum of the sums (i.e., not just those in $\theta_n(s,u)$).

In combination with the first lower bound, which features types that are fully independent from those of the second consideration, we get $l + m$ as lower bound of the sum of sums (remember that $l = |\theta_n^1(s,t)|$):

$$\frac{l+m}{k+l+m} \leq DT(s,u) + DT(u,t).$$

Because $DT(s,t) = \frac{l}{k+l} \leq \frac{l+m}{k+l+m}$, apparent from $l^2 + (k+m)l \leq l^2 + (k+m)l + mk$, the triangle inequality is proven.

Experimental validation of similarity metrics

In addition to the mathematical proof of the (pseudo-)metric properties, we define a similarity metric as being validated if it meets predictive validity [Cronbach and Meehl 1955]. Predictive validity is defined as assessment of the metric's ability to predict something that it should theoretically be able to predict. To demonstrate that the metrics meet predictive validity, we defined an experimental validation protocol based on simulations (Figure 2.2.3) derived from [Jannin et al. 2008].

Experimental setup

For the validation of a similarity metric Δ, a modified version of a given SPM s is produced based on one or two parameters that determine the degree of the modifications. Those parameters are denoted by μ for simplicity here, and they are vectors in the general case. The transformations result in an SPM $T_\mu^\Delta(s)$. The latter is understood as a simulated SPM and is compared to the original s by means of Δ. Notably, the required registration mappings for the similarity metrics concerning granularity and content and for temporal similarity arise naturally from the kind of the transformation (Table 2.2.II). Finally, the correlation between the similarity value $\Delta\left(s, T_\mu^\Delta(s)\right)$ and the modification factor(s) μ is determined. We assume that the similarity metric Δ meets predictive validity if it strongly correlates with the modification factor(s). A strong correlation means that metric predicts the modification factor(s).

The changes to the original SPM s are primarily achieved by deleting or modifying a percentage of randomly selected activities (modification factor μ^{AC}) in 10% intervals in regard to the total number of activities in s. For instance, if $\mu^{AC} = 20$, 20% of the activities of s were randomly selected and modified. These random modifications were repeated ten times for each value of $\mu^{AC} \in \{10, 20, \ldots, 100\}$.

The specific kinds of transformations were chosen depending on of the particular focus targeted by each metric. These transformations are presented in Table 2.2.II. For similarity of content and temporal similarity, an additional modification factor was used in each case. These factors were μ^{PE} (number of modified perspectives) and μ^{BHV} (change in activity durations in seconds), respectively. In those cases, similarity values were calculated for different combinations of μ^{AC} and μ^{PE} and μ^{AC} and μ^{BHV}, respectively.

Three clinical data sets from different surgical disciplines were used as original SPMs for the experimental validation. The first validation data set consisted of SPMs of 20 cataract procedures. Within a cut-suture time of approximately 20 minutes, approximately 30 different activities were performed. The cataract procedures were recorded by observation by trained medical students according to ICCAS methodology [Neumuth et al. 2009b] at the Ophthalmology Department of the University Hospital Leipzig in 2006.

The second validation data set consisted of 20 SPMs of craniotomies, a neurosurgical intervention. Craniotomies have an approximate duration of 120 minutes and consist of approximately 200 activities. The interventions were acquired by trained observers according to ICCAS methods in the Neurosurgery Department at the University Hospital Leipzig in 2007 and 2008.

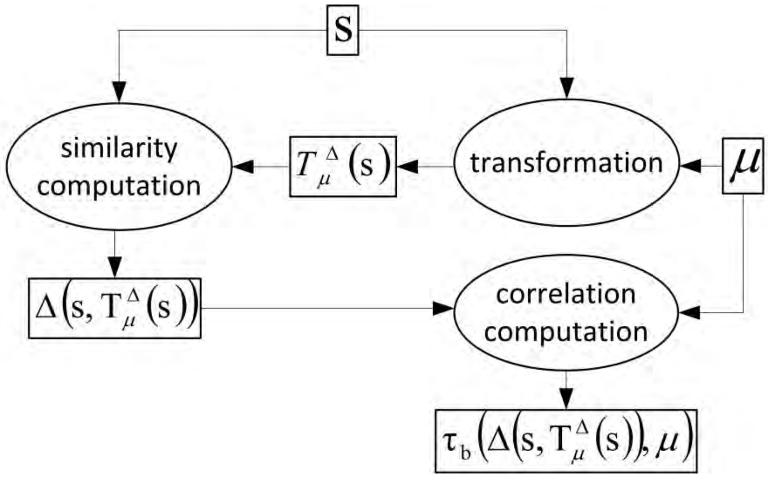


Figure 2.2.3: Experimental design for validation of the metrics: the result $\Delta\left(s, T_\mu^\Delta(s)\right)$ of the similarity metric is compared to the introduced degree of modification μ.

The third data set consisted of 20 SPMs of supratentorial tumor removals, also neurosurgery procedures. These data sets were acquired by structured interviews according to the methods developed at the University of Rennes [Jannin et al. 2003] between 2003 and 2005. The data sets comprised a maximum of 7 surgical steps and contained no time stamps.

The processing of Δ and T_μ^Δ was performed by modifying the clinical data sets in a PostgreSQL® database [PostgreSQL Global Development Group 2009]. Statistics were performed using SPSS® [SPSS Inc. 2008]. The correlations between $\Delta\left(s, T_\mu^\Delta(s)\right)$ and μ were calculated using bivariate non-parametric Kendall Tau-b coefficients. The correlations for transformations of metrics with two modification factors were calculated using the product of the modified percentage of activities μ^{AC} and the respective second parameter (cf. Table 2.2.II).

Table 2.2.II: Modification procedures in preparation for experimental validation: for each metric and an arbitrary but fixed original SPM s, the modifications and the parameter ranges are shown.

Metric	**Modification / simulation procedure and similarity computation for validating the metric**	**Parameter ranges**
Granularity similarity	$T^{DG}_{\mu^{AC}}(s)$: delete μ^{AC}% randomly selected activities of s Δ: calculate $DG\left(s, T^{DG}_{\mu^{AC}}(s)\right)$	$\mu^{AC} \in \{10, 20, \ldots, 100\}$
Content similarity	$T^{DC}_{\mu^{AC},\mu^{PE}}(s)$: 1. randomly select μ^{AC}% of the activities of s 2. modify μ^{PE} randomly selected nominal perspectives of the selected activities Δ: calculate $DC\left(s, T^{DC}_{\mu^{AC},\mu^{PE}}(s)\right)$	$\mu^{AC} \in \{10, 20, \ldots, 100\}$ $\mu^{PE} \in \{1, 2, 3, 4\}$
Temporal similarity	$T^{DTp}_{\mu^{AC},\mu^{BHV}}(s)$: 1. randomly select μ^{AC}% of the activities of s 2. modify activity times by a random value of μ^{BHV} seconds at each selected activity Δ: calculate $DTp\left(s, T^{DTp}_{\mu^{AC},\mu^{BHV}}(s)\right)$	$\mu^{AC} \in \{10, 20, \ldots, 100\}$ $\mu^{BHV} \in \{0, 1, \ldots, 5\}$
Transitional similarity	$T^{DTr_2}_{\mu^{AC}}(s)$: delete μ^{AC}% randomly selected activities of s Δ: calculate $DTr_2\left(s, T^{DTr_2}_{\mu^{AC}}(s)\right)$	$\mu^{AC} \in \{10, 20, \ldots, 100\}$
Transitional frequency similarity	$T^{DTf_2}_{\mu^{AC}}(s)$: delete μ^{AC}% randomly selected activities of s Δ: calculate $DTf_2\left(s, T^{DTf_2}_{\mu^{AC}}(s)\right)$	$\mu^{AC} \in \{10, 20, \ldots, 100\}$

Results of the experimental validation

The resulting correlations are shown in Table 2.2.III. The correlation coefficients were higher than 0.8 for all metrics. For granularity similarity, content similarity, and temporal similarity, the correlation coefficients were close to 1. All metrics showed a two-sided significance of p<0.001.

The 95% confidence intervals for the similarity metrics with one modification parameter are shown in Table 2.2.IV.

Table 2.2.III: Correlation coefficients (τ_b) from Kendall's Tau-b test for similarity metrics, with p<0.001 for two-sided significance at the 0.01 level for all metrics.

Metric	**Cataract interventions**	**Craniotomy interventions**	**Supratentorial tumor removal interventions**
Granularity similarity	0.966	0.964	0.935
Content similarity	0.988	0.987	0.968
Temporal similarity	0.982	0.982	-
Transitional similarity	0.931	0.888	0.831
Transition frequency similarity	0.947	0.962	0.841

Table 2.2.IV: Means and standard deviations of $\Delta\left(s, T_{\mu}^{\Delta}(s)\right)$ for metrics with one modification parameter. Results are given as the mean ± SD [95% CI].

	μ^{AC}	Cataract interventions	Craniotomy interventions	Supratentorial tumor removal interventions
Granularity similarity	10	10.21±0.88 [10.10;10.34]	9.98±0.06 [9.97;9.99]	11.00±7.04 [10.04;11.96]
	20	20.03±0.58 [19.95;20.11]	20.00±0.06 [19.99;20.01]	16.71±2.31 [16.39;17.02]
	30	29.80±0.77 [29.69;29.91]	29.98±0.05 [29.97;29.99]	33.42±4.63 [32.79;34.05]
	40	40.17±0.70 [40.07;40.27]	40.00±0.06 [39.99;40.01]	38.87±4.09 [38.31;39.42]
	50	50.40±0.89 [50.28;50.52]	50.01±0.06 [59.99;60.01]	49.86±6.94 [48.91;50.80]
	60	59.83±0.70 [59.73;59.92]	59.99±0.06 [59.99;60.01]	61.13±4.09 [60.58; 61.69]
	70	70.20±0.77 [70.09;70.31]	70.01±0.06 [70.00;70.02]	72.29±5.29 [71.57;73.01]
	80	79.97±0.58 [79.89;80.05]	80.00±0.06 [79.99;80.01]	83.29±2.31 [82.97;83.60]
	90	90.35±0.83 [90.24;90.47]	90.03±0.05 [90.02;90.04]	89.00±7.04 [88.04;89.96]
Transitional similarity	10	14.13±4.74 [13.92;14.33]	8.85±6.07 [8.58;9.11]	19.88±15.00 [19.23;20.52]
	20	27.58±5.41 [27.34;27.82]	16.90±7.63 [16.57;17.23]	30.13±11.15 [29.65;30.60]
	30	40.16±6.24 [39.88;40.43]	24.22±8.30 [23.85;24.58]	57.29±14.63 [56.67;57.92]
	40	53.27±6.53 [52.98;53.55]	31.55±8.70 [31.17;31.93]	65.26±13.87 [64.67;65.86]
	50	64.99±6.38 [64.71;65.27]	38.49±8.68 [38.12;38.89]	77.59±14.53 [76.96;78.20]
	60	74.53±6.11 [74.27;74.80]	46.01±8.79 [45.62;46.39]	88.82±11.42 [88.33;89.30]
	70	84.09±5.34 [83.86;84.32]	54.25±8.97 [53.85;54.64]	95.48±7.98 [95.14;95.82]
	80	91.37±4.31 [91.18;91.56]	64.46±8.23 [64.10;64.83]	0**
	90	97.33±2.56 [97.22;97.44]	78.37±6.75 [78.07;78.66]	0**
Transition frequency similarity	10	7.80±1.72 [7.73;7.86]	5.64±0.31 [5.62;5.65]	11.84±9.24 [11.44; 12.23]
	20	15.90±2.23 [15.80;16.00]	11.91±0.52 [11.89;11.93]	18.82±7.86 [18.48;19.15]
	30	24.88±3.10 [24.88;25.01]	18.86±0.77 [18.83;18.90]	38.35±14.34 [37.73;38.96]
	40	35.91±3.94 [35.74;36.08]	26.69±1.01 [26.64;26.73]	46.36±14.88 [45.72;46.99]
	50	47.37±5.23 [47.15;47.61]	35.46±1.21 [35.40;35.51]	60.90±18.36 [60.12;61.69]
	60	59.09±6.01 [58.83;59.35]	45.33±1.47 [45.27;45.40]	78.27±18.35 [77.48;79.06]
	70	72.11±6.39 [71.84;72.39]	56.58±1.61 [56.50;56.65]	90.05±14.49 [89.43;90.67]
	80	84.40±6.03 [84.13;84.66]	69.34±1.62 [69.27;69.41]	0**
	90	95.21±4.14 [95.03;95.38]	83.98±1.29 [83.92;84.04]	0**

Evaluation study

We performed an additional study to demonstrate the added value of the proposed metrics for estimating the learning curve of human observers during SPM acquisition. Human observation is currently the most flexible solution for acquiring data for surgical process modeling in the OR. Automatic acquisition in the OR is currently not feasible because there is a lack of efficient sensors available for use in the operating room. These observers need to be trained to ensure that the observed SPMs meet the quality requirements preset by the study context. An unequivocal criterion for this acquisition is the achievement of certain values for the metrics. In this evaluation study, we trained three medical students who did not have previous recording experience to acquire SPMs according to the ICCAS Surgical Workflow Editor. The students were asked to record different variations of actions during functional endoscopic sinus surgery (FESS), an otorhinolaryngology intervention, which was performed by actors based on simulation scripts. The students recorded three variants of each variation, three times each, in a randomized order. For each recording, similarity measures were taken for the SPMs by the observers, and the SPMs were defined by the simulation scripts. Figure 2.2.4 shows the results of the learning progress and the trend lines for granularity similarity, content similarity, and temporal similarity among the three observers in generating SPMs.

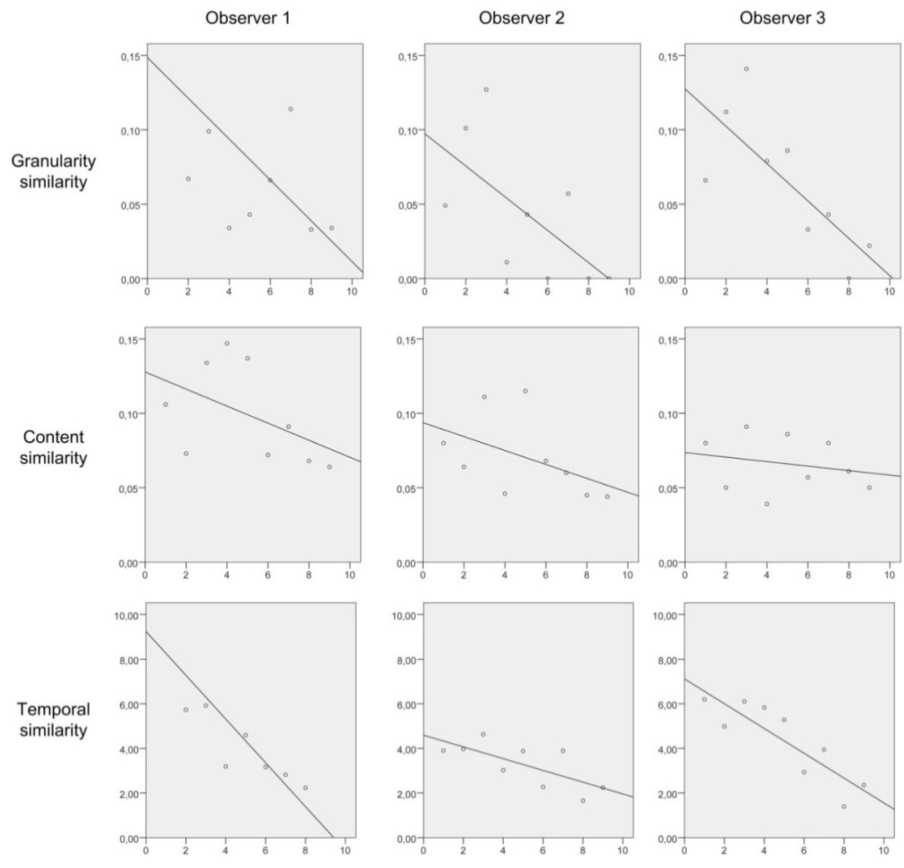


Figure 2.2.4: Example application of the metrics to indicate the learning curves of newly trained observers for granularity similarity, content similarity, and temporal similarity. Dots indicate the achieved results, and the regression line indicates the learning progress.

Discussion

To the best of our knowledge, there are no similarity metrics available in the research field of surgical process modeling and surgical workflows. Similarity metrics for workflows have been introduced for business process models [Bae et al. 2006b; de Medeiros et al. 2008; Rozinat and van der Aalst 2008] and for clinical workflows [Combi et al. 2009]. Because some of these metrics, such as those regarding temporal similarity, are trivial and easy-to-use, we have included them in our metrics set. However, most of the existing metrics are not adapted to the specificities of SPMs because they cover only isolated aspects. For instance, no available metrics consider concepts such as granularity or perspective. Additionally, for most of the available metrics, predictive validity has not yet been demonstrated.

The first metric, granularity similarity, estimates the compliance of granularity. Even though a definition of 'correct' granularity cannot be derived objectively, the compliance with a granularity or resolution value that previously defined as correct can be measured. The results showed a high correlation between the introduced modification and the measures. Content similarity includes measures for describing the correctness of perspective contents for an activity. Perspectives might be weighted according to their importance within the study context. The correlation of content accuracy with the modified values was significant, and the metric decreased linearly. Temporal similarity assesses temporal compliance among activities and estimates potential time errors in a protocol. The metric is strongly linear, as was shown in the validation section. Here, correlations with modifications were also highly significant, and the metric behaved as expected. Transitional similarity and transition frequency similarity are metrics based on information about the order of activities. Whereas the former focuses on the presence of activities, the latter also considers their frequency. Both metrics showed the poorest correlations amongst all presented metrics. However, they were still highly significant. Although both metrics showed linear behavior for each of the surgical intervention types, the measurements between the SPMs for the intervention types showed high differences in their 95% confidence intervals. This discrepancy is caused by the different number of activities in the SPMs. For example, there were a mean number of 7 activities during supratentorial tumor removals, 20 activities in cataract interventions, and 200 activities in craniotomies. This difference suggests that both metrics need a minimum number of activities to be applicable. The similarity of this metric is strongly dependent on the available number of transitions and the chosen length of transition n. The greater the chosen length, the greater the expected distance values because the probability that two transitions differ in at least one activity increases. With the results of the validation of the metrics, we show that an experimental validation of metrics should be done in addition to mathematical proofs to reveal unexpected behavior that cannot be obviously discovered by the formal proof. However, this point needs further investigation.

Generally, it was shown that the metrics meet predictive validity independent of a certain type of surgery. It was also realistic to iterate each of the simulations ten times because this is equivalent to simulating certain degrees of randomized noise in the SPMs.

The value of the metrics for clinical users was demonstrated by presenting a case in which the metrics were used to train SPM observers. The metrics provided a useful instrument 1) to estimate skill level and learning curves during SPM acquisition and 2) to evaluate when an observer has received enough training and can be employed to acquire data for clinical studies. Indirect clinical value can be derived by the availability of accurate SPMs that facilitate the potential uses mentioned above, in addition to facilitating the quantitative evaluation of new surgical assist systems or determining the performance of requirements analysis for new systems that improve surgical procedures. Both of these tasks rely on accurate SPMs.

From the definition and mathematical point of view, many options remain to be explored in future research. This research starts with definitional variants of the presented metrics, e.g., by considering alternatives in comparing activities with regard to their temporal aspects. Moreover, despite basic properties of (pseudo)metrics being proven for all measures, further properties that could be deemed mathematically desirable do not apply to the present metrics. One could search for variants of the first three metrics with reduced or even no dependence on registration mappings, for instance. Eventually, future uses for similarity metrics may uncover aspects that need to be taken into account but have been neglected thus far. Nevertheless, the proposed metrics form a suitable starting point for quantifying differences between SPMs.

Furthermore, the design of a general metric that summarizes the multiple dimensions of SPMs (granularity, content, time, order, and frequency of surgical activities) can be discussed. From our experience, this discussion is currently not useful because each aspect would cause a different dimension that would then influence the metric. This situation would cause problems during interpretation because one would never know which dimensions caused bad values. For this reason, we chose to provide a set of metrics.

The introduced set of metrics is intended for comparing SPMs. One important future use for these metrics is the design of an infrastructure for context-aware operating rooms. Context-awareness in operating rooms is a field with growing interest among medical engineers [Lemke and Vannier 2006]. These ORs aim to have an awareness concerning the behavior of the medical personnel and of the technical systems involved in the patient's surgical treatment. A future OR infrastructure requires "knowledge" of the current situation in the operating room, which can only be obtained by observation or by sensor systems. Therefore, the similarity metrics and the validation methodology we introduced in this paper can be used to study the predictive validity of such systems.

Conclusion

Measuring the similarity of surgical processes is a complex task. However, the availability of metrics for the computation of similarity between surgical process models is needed for many uses in medical engineering. These metrics are the key for providing a solid base for decisions, such as those necessary when validating observers or sensor systems for use in the operating rooms of the future. Generally, it was shown that the metrics meet predictive validity independent of a certain type of surgery.

We have introduced and validated a set of five different similarity metrics that deal with several dimensions of process compliance, including granularity, content, time, order, and frequency of surgical activities. We introduced a set of metrics for SPMs and demonstrated their validity since most of the existing metrics for processes were not adapted to the specificities of SPMs. Some of the new metrics like granularity similarity or content similarity were even not available in other domains such as business process management. Here we provided an instrument to measure the compliance with a process granularity or resolution value objectively and without the availability of a global definition of a 'correct' process granularity. Additionally, new concepts such as perspectives in content similarity allow for the description of perspective contents, for instance process resources.

The presented metrics are beneficial for clinical users, such as was demonstrated by presenting a use case in which the metrics were used to train SPM observers. The metrics provided a useful instrument 1) to estimate skill level and learning curves during SPM acquisition and 2) to evaluate when an observer has received enough training and can be employed to acquire data for clinical studies. Indirect clinical value can be derived by the availability of accurate SPMs that facilitate the potential uses mentioned above, in addition to facilitating the quantitative evaluation of new surgical assist systems or determining the performance of requirements analysis for new systems that improve surgical procedures. Both of these tasks rely on accurately acquired SPMs.

Acknowledgements

We are very grateful to the anonymous reviewers for their detailed comments that helped us improve the manuscript. Many thanks to Ringo Baumann for valuable comments and discussions, primarily on formal aspects of the definitions of the similarity metrics.

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the scope of the Unternehmen Region with grant numbers 03 ZIK 031 and 03 ZIK 032.

3 Data acquisition strategies for surgical process modeling

Current descriptions of surgical processes have limitations. Sources, such as clinical guidelines [AWMF-Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. 2010a; AHRQ-Agency for Health Care Research and Quality 2010a] or surgical textbooks are plausible references to apply the top-down modeling strategy. However, the applicability of these sources concerning analytical purposes has limitations because of the lacking attention to detail, the missing objective quantifiability, and the subjective point of view of the modeler. Also, the variability of the process is insufficiently represented. Due to these facts it is important and sensible to develop and implement new approaches for the acquisition of information about surgical processes, providing a basis to overcome the current limitations.

Data acquisition strategies are methods that are of interest to the process modeler and can be employed for the realization of the surgical and technical use cases that are depicted in the following chapters on model generalization and clinical applications.

Generally speaking, there are two general strategies available for bottom-up modeling: data acquisition based on observers or on sensors. The first two publications

> *Neumuth T, Jannin P, Strauß G, Meixensberger J, Burgert O. Validation of knowledge acquisition for surgical process models. Journal of the American Medical Informatics Association. 2009; 16(1): 72-80.*

and

> *Neumuth T, Kaschek B, Goldstein D, Ceschia M, Meixensberger J, Strauss G, Burgert O. An observation support system with an adaptive ontology-driven user interface for the modeling of complex behaviors during surgical interventions. Behavior Research Methods. 2010; 42:1049-58.*

describe the development and validation of observation support systems. The first publication presents the general approach and shows the validity of surgical process models that were acquired using this strategy. The second publication extends the approach of the observation support software by applying the concept of adaptive and situation-dependent user-interfaces.

The assessment of both approaches is accomplished by utilizing the similarity metrics described in the previous chapter. Also, the applicability of the approach is evaluated. Due to the fact, that the method of applying adaptive user interfaces is new even on the traditional field of behavior research, the second publication was published in a journal focusing on this scientific branch, rather than in a journal concerning medical engineering or medical informatics, as the other original research presented here has been.

The resulting recording accuracies achieved by the observers were >90% for the process granularity and content accuracy, the temporal accuracy being 1.8s. There were no observed significant differences between the results of live observation as compared to video-based observation. Thus it can be concluded, that live observations are easily available and are an accurate mean to acquire surgical process models. However, there were significant differences between the observers themselves, depending on their background, technical or medical.

As concerns the ontology support, the results of the observation support system, concerning the differences between live and video-based observation, were confirmed. In addition, the observation workload for the users was significantly decreased when the ontologically supported user interface system was employed.

An additional publication

> *Neumuth T, Meißner C. Online recognition of surgical instruments by information fusion. International Journal of Computer Assisted Radiology and Surgery. 2012; 7(2):297-304.*

delineates the acquisition of information of surgical process models by means of sensor systems. Available sensor-based approaches [Padoy et al. 2007; Ahmadi et al. 2010; Lalys et al. 2010] cannot be employed comprehensively concerning the type or discipline of the surgical process. This is mainly due to two facts: These methods are not flexible enough to engender different intervention types or have a limited description range for the compilation of distinct procedure stages, comprising lacking possibilities to correctly relate applied surgical instruments, performed surgical work steps, or treated anatomical structure. To overcome these challenges, the employment of information-fusion strategies is proposed in this article. Thus, various sensor types are combined to optimize the comprehensiveness of the description of the surgical process, and the overall description of the process itself. In the course of this work, different information-fusion strategies are presented, and their respective effects on the process model are being evaluated using the example of instrument detection based on radio frequency identification.

The article resulted in a statistical evaluation of redundant, complementary, and cooperative sensor signal fusion strategies. It was shown that all three fusion strategies proved significant ($p<0.001$) to increase the recognition accuracy.

3.1 Observer-based data acquisition with observation support software

Title

Validation of knowledge acquisition for Surgical processs

Authors

Thomas Neumuth[1], Pierre Jannin[2,3,4], Gero Strauss[1,5], Jürgen Meixensberger[1,6], Oliver Burgert[1]

Departments

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] INSERM, U746, Faculty of Medicine, Rennes, France

[3] INRIA, VisAGeS Unit/Project, Rennes, France

[4] University of Rennes I, CNRS, UMR 6074, IRISA, Rennes, France

[5] University Medical Center Leipzig, Department of Otorhinolaryngology, Leipzig, Germany

[6] University Medical Center Leipzig, Department of Neurosurgery, Leipzig, Germany

Citation

Neumuth T, Jannin P, Strauß G, Meixensberger J, Burgert O. Validation of knowledge acquisition for surgical process models. Journal of the American Medical Informatics Association. 2009; 16(1): 72-80.

Keywords

Validation; Validation Studies as Topic; Observation; Surgical Procedures, Operative

Abstract

Objective: Surgical process models (SPMs) are models of surgical interventions. The objectives of this study are to validate acquisition methods for surgical process models and to assess the performance of different observer populations.

Design: The study examined 180 SPM of simulated Functional Endoscopic Sinus Surgeries (FESS), recorded with observation software. About 150,000 single measurements in total were analyzed.

Measurements: Validation metrics were used for assessing the granularity, content accuracy, and temporal accuracy of structures of SPMs.

Results: Differences between live observations and video observations are not statistically significant. Observations performed by subjects with medical background gave better results than observations performed by subjects with technical background. Granularity was reconstructed correctly by 90%, content by 91%, and the mean temporal accuracy was 1.8 s.

Conclusion: The study shows the validity of video as well as live observations for modeling surgical processes. For routine use, we recommend live observations due to their flexibility and effectiveness. If high precision is needed or the SPM parameters are altered during the study, video observations are the preferable approach.

Introduction

Surgery is a clinical specialty with a long history, but surgical techniques are learned in an apprentice-master model that leads to several surgical schools treating the same disease in different ways.

There is no explicit methodology available, which prevents an objective comparison of surgical strategies at a fine-grained level. Using process models with fine-grained descriptions of surgical interventions as the processes, surgeons get a powerful tool for the discussion of different surgical approaches and scientifically sound process models of their surgical work steps.

A detailed surgical process model (SPM) may help in understanding a procedure, especially in difficult cases. Such a detailed model must be available for a broad variety of similar interventions to cover all clinically relevant deviations from the standard procedure.

Furthermore, a collection of verified and valid SPMs of surgical processes, especially for rare cases, could help in the implementation of new surgical techniques (e.g., minimally invasive surgery or computer assisted interventions) that require a detailed understanding of the intervention course in order to optimally assist the surgeon.

Surgical process models may be used to facilitate the development of technical components for surgical assist systems (SAS) [Lemke and Vannier 2006; Cleary et al. 2005] and to support standardization efforts for desired functionalities of SAS, such as future extensions of Digital Imaging and Communications in Medicine (DICOM) for surgery [Lemke 2007; Burgert et al. 2007]. The ultimate purpose of SPMs is the generation of these descriptions for technical requirement analysis, evaluation, and systems comparison.

For the modeling, data must be at an adequate level of granularity. The modeling must address behavioral, anatomical, and pathological aspects and surgical instruments [Jannin et al. 2003].

Accuracy is crucial. This is why the modeling must be rigorously validated. The objective of our study was the validation of data acquisition for SPMs. The research question was, “How accurate are observations of surgical processes by human observers?” We designed a rigorous validation strategy that assessed the accuracy of SPMs that were acquired from simulated interventions in a controlled environment. We studied several validation criteria: granularity, content accuracy, and temporal accuracy, using video and live observation as data acquisition strategies and using medical and technical students as the observer populations. For assessing the validation criteria, metrics have been defined and applied to the SPMs. Secondary questions of interest included time to complete observations, the subjective workload estimation of observers, and the level of surgical knowledge required by observers.

Background

The amount of information available from surgical processes is large and complex, although the knowledge of the surgeon is mostly implicit and hidden from formal assessment. Data may be acquired by using two main strategies: sensor systems or, in a more classical way, human observation.

Only a few sensor technologies are available for application in the sensitive operating room (OR) environment. These technologies are not suitable for uniform acquisition of data such as work-step information, inter-device communication, human-device behavior or inter-human behavior for modeling due to missing information models, network communication, and interfaces. It is necessary to use human recognition and perception capabilities for parts of the data acquisition, which is a common strategy in biomedicine [Payne et al. 2007] and empirical social sciences [Kromrey 2006].

Only a few approaches for modeling surgical processes are described in the literature. MacKenzie et al. [MacKenzie et al. 2001] performed iteratively top-down and bottom-up analyses for assessing laparoscopic Nissen fundoplications for training residents. The data acquisition was performed based on video observations. Münchenberg et al. [Münchenberg et al. 2001a] modeled surgical procedures of Frontal Orbital Advancements to treat craniosynostosis for the purpose of planning and technical intra-operative support for the surgeon; the data acquisition methodology was not mentioned. Jannin et al. [Jannin et al. 2003] modeled surgical procedures in the context of multimodal image-guided neurosurgery. Data were acquired pre- and post-operatively via questionnaires. None of the previous work validated the data acquisition process. Validation of data acquisition in the clinical domain has been performed by Vawdrey et al. [Vawdrey et al. 2007], who assessed the data quality of ventilators operated by respiratory therapists. Data were acquired by electronic medical records. Rosenbloom et al. [Rosenbloom et al. 2006; Rosenbloom et al. 2008] evaluated the interface terminologies of clinical interfaces. These studies were adequate for medical patient records, but they did not provide an overall measure of the accuracy.

Working definitions used in this article are strongly related to business process modeling and workflow management systems [Workflow Management Coalition 1999a]. By analogy, we define a *surgical process (SP) as a set of one or more linked procedures or activities that collectively realize a surgical objective within the context of an organizational structure defining functional roles and relationships*. The surgical objective is the correction of an undesirable state of the patient's body, which is performed in the organizational structure of a hospital. The responsible surgeon coordinates the performance of the surgical procedure. We define a s*urgical process model (SPM) as a simplified pattern of a surgical process that reflects a predefined subset of interest of the SP in a formal or semi-formal representation* [Neumuth et al. 2007b]. The working definitions are also provided to clarify the relationship to the frequently used term surgical workflow, which relates to the performance of a surgical process with support of a workflow management system [Jannin and Morandi 2007].

The objective of this work was to perform a validation study for assessing data acquisition results of SPMs by human observers with specialized software. The SPs consisted of simulations of Functional Endoscopic Sinus Surgeries (FESS).

Methods

First, the data acquisition software and its underlying ontological concepts are introduced. Then, the experimental setup and post-processing are described. The notion of variables that might influence a validation study for SPMs is discussed in a separate section. These variables were divided into three groups: extraneous variables that need to be held constant, independent variables that were manipulated according to the experimental design, and dependent variables that were affected by the manipulation of the independent variables. Finally, the validation metrics quantified the manipulation effects.

Data acquisition software and fundamental concepts

The data were acquired with a JAVA software application, the surgical workflow editor [Neumuth et al. 2006b; Neumuth et al. 2006a]. The objective of the software is to devise ontological concepts used for describing the SP to the observer and to ask him or her for the instantiation of these concepts to create an observation protocol. A screenshot of the surgical workflow editor is shown in Figure 3.1.1.

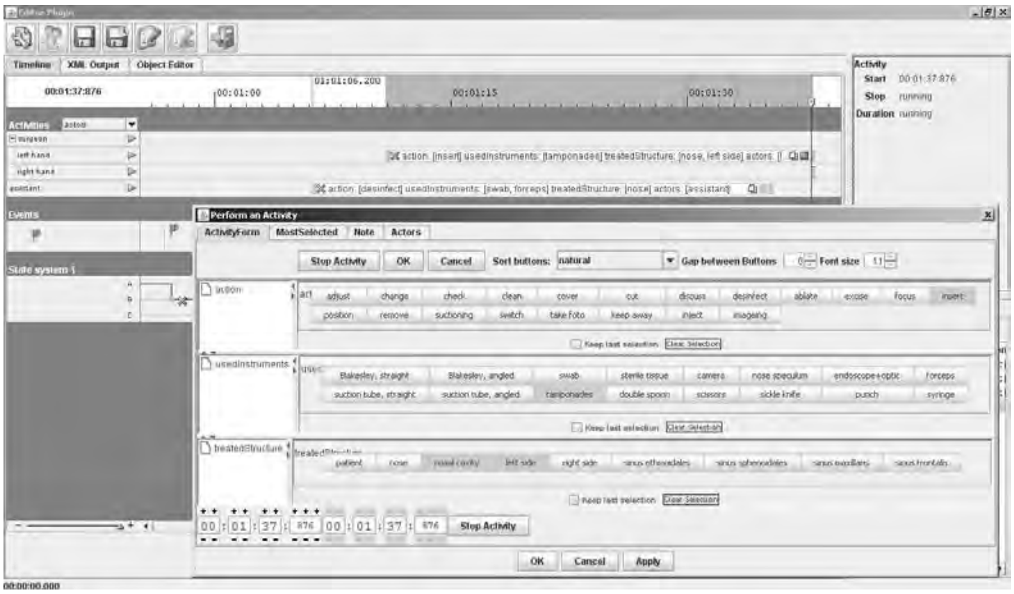


Figure 3.1.1: A screenshot of the surgical workflow editor.

The data acquisition process begins with the definition of the structure of the SPM. The structure is described by the *structural ontology* and specifies how information of the SP is represented in the SPM. During actual data acquisition, specific concepts of the observed SP, described by the *content ontology* (e.g., surgical actions, participants, or instruments) are instantiated by the observer.

Our structural ontology contains three types of flow objects [White 2004]: *activities*, *state transitions*, and *events*. Each SPM consists of these flow objects.

Activities represent manual work steps performed during the interventions. To structure their content, we used the factual perspectives for workflow schema proposed in [Jablonski and Bussler 1996], modified them, and added the spatial perspective. An activity consists of five perspectives, which decompose the observer's view into various viewpoints:

- the functional perspective describing *what* is done in a surgical work step;
- the organizational perspective describing *who* is performing a work step;
- the operational perspective to describe *instruments* used in performing a work step;
- the spatial perspective describing *where* a work step is performed;
- and the behavioral perspective describing *when* a work step is performed.

Perspectives are extended by perspective attributes. They decompose perspectives further (e.g., indicating that a surgeon is performing a work step with his right hand, where both perspective attributes belong to the organizational perspective). More examples may be found in Table 3.1.I.

For recording work steps with no measurable time extension, we defined the concepts of *state transitions* and *events*. State transitions are changing variables between predefined values, e.g. observable on monitors in the operating room or the phases of an intervention. Events might describe the content of messages, such as the surgeon's instruction to administer a drug. State transitions and events each include the functional and behavioral perspective.

The purpose of the content ontology is to determine the correct intervention-specific relations between perspective contents, e.g., for *suctioning* (functional perspective) only a *suction tube* (operational perspective) may be used. The development of the content ontology is based on expert knowledge.

Table 3.1.I: Example flow object pattern used for gold standard terminology.

perspective	perspective attribute	example activity	example activity	example activity	example event	example state transition
functional	action	disinfect	dissect	insert	event 1	A -> B
organizational	participant	assistant	surgeon	surgeon	-	-
	used bodypart	-	right hand	left hand	-	-
operational	main instrument	swab	Blakesley	nose speculum	-	-
	supporting instrument	forceps	-	-	-	-
	property of main instrument	-	straight	-	-	-
spatial	anatomical structure patient	patient	patient	patient	-	-
	anatomical structure nose	nose	nasal cavity	nasal cavity	-	-
	anatomical structure side	-	right side	right side	-	-
	anatomical structure nasal cavity	-	c. ethmoidales	-	-	-
behavioral	starttime	00:00:00	00:00:00	00:00:00	00:00:00	00:00:00
	stoptime	00:00:20	00:00:20	00:00:20	-	-

Experimental setup

The validation procedure consisted of recording the simulated SP and comparing the resulting SPM to a reference afterward. The main steps for the experiments are shown in Table 3.1.II.

Table 3.1.II: Step descriptions for the experimental design.

Experiment preparation	1 Select structural ontology of the SPM 2 Define concepts of content ontology 3 Design terminology patterns for Gold Standards 4 Design Gold Standards 5 Speak and record audio representation of Gold Standards 6 Perform simulations without observers and video record them for later use in the video observations and to serve as the Bronze Standards for video observations 7 Code these videos as XML-protocols
Data acquisition sessions	8 Perform simulations with live observations, record these simulations on video as the Bronze Standards for the live observations 9 Perform observations of video simulations (recorded in Step 6)
Post processing	10 Register Bronze Standard protocols to respective Gold Standard protocols 11 Register observation protocols to Bronze Standard protocols 12 Extract Bronze Standard terminology pattern and observation terminology pattern 13 Register Bronze Standard terminology pattern to Gold Standard terminology pattern 15 Register observation terminology pattern to Bronze Standard terminology pattern
Observation validation	16 Calculate vm_i for observation protocols by comparing observation protocols to corresponding Bronze Standards as references 17 Perform statistical analysis
Simulation validation	18 Calculate vm_i for Bronze Standards by comparing them to corresponding Gold Standards as references 19 Perform statistical analysis

The validation was performed on simulated Functional Endoscopic Sinus Surgeries (FESS) as SPs. A FESS intervention has the objective of removing polyps from nasal cavities. During the core part of the intervention, the surgeon holds an endoscope with one hand while his or her other hand performs the actual work steps.

The processes to be simulated were built based on real FESS intervention recordings. For the study, the FESS-specific content ontology contained concepts of two participants, two used body parts, twelve actions, 13 surgical instruments, three instrument attributes, and seven treated structures. The concepts were chosen based on routine daily clinical terminology.

In preparation for defining the Gold Standards for the study, flow object patterns of the structural ontology and work step information of the content ontology were used to construct FESS-specific terminology. This was composed of flow object patterns for 41 different activities, which represented surgical work steps, three state transitions, and three events. Pattern examples are shown in Table 3.1.I. The patterns of the Gold Standard terminology were used to design three different Gold Standards as simulation scripts that served as references for assessing the accuracy of the simulations. First, the prototype Gold Standard SPM was generated. It contained a typical sequence of work steps with predefined timestamps. From this, two more simulation scripts, the second and the third Gold Standard, were derived by adding noise. The noise additions included modifying the treatment order of nasal cavities, increasing work speed, and switching the surgeon and assistant roles temporarily. The created simulation scripts were checked by two ENT-surgeons for clinical realism. Each of the simulations was 21 minutes in length and was limited to 60 to 90 activities.

The three different Gold Standards were spoken and recorded as audio files, containing detailed instructions for the work steps to be performed by the actors. One simulation for each Gold Standard was performed without observers, recorded with multiple video cameras, synchronized, and cut as a video representation of the simulation for later use in video observations. These protocols were coded in XML-format and named as “Bronze Standards”. They served as reference SPMs for allowing quantitative validation of the simulations against the Gold Standard simulation scripts and of the video observations by medical and technical observers.

The data acquisition sessions were performed in the ICCAS-demonstration OR in Leipzig and consisted of one educational session day for the uniform training of the observers and three data acquisition sessions days for each observer group. The educational session introduced the purpose of data acquisition for SPM, the surgical workflow editor software, the surgical objectives of FESS procedures, the typical intervention course, and the content ontology to the observers to establish a common context of use. The objective of this session was to simulate the situation for observing real surgical processes, where the observer needs to understand the procedure in depth before he or she begins to record data.

Ten observers performed nine observations for each data acquisition strategy with tablet-PCs. Video observations were conducted based on the performance of the simulation scripts without observers during the experimental preparation. The live observations were based on live simulations by the actors. The work steps of the live simulations were recorded by endoscope and two video cameras. After the recorded videos were synchronized, they served as the Bronze Standards for the live observations.

After each observation, the observers performed a workload assessment, the Task Load Index (TLX) test [Hart and Staveland 1988] of the National Aeronautics and Space Administration (NASA), for describing their subjective workload feeling, and they continued with acquiring data by the respective other data acquisition strategy of video and live observation. Additionally, the observers were required to pass a knowledge test twice per data acquisition day.

Post-processing

Before analysis, post-processing was required to link each SPM to its reference. Post-processing started with the manual association of each flow object of an observer protocol to its corresponding reference flow object in a Bronze Standard protocol. By performing this association between flow objects, registration matrices of the protocols were created.

Subsequently, the protocols and registration matrices were transferred to a database, where the terminology patterns of the Bronze Standard terminologies and the terminologies of the observations were extracted from the respective protocols and automatically compared to each other.

Finally, the validation metrics were calculated, and the statistical analysis was performed. The statistical analysis was done using multivariate Generalized Linear Models (GLM) for the data acquisition strategy and the observer population. The simulated Gold Standard and the repetition of the measurements were considered as covariates. All statistical tests were performed with a significance level of $\alpha = 0.05$ and computed with the SPSS software (SPSS Inc., Chicago, IL).

Analysis

Preliminary identification of factors that may influence such a validation is required. Inspired by Shah and Darzi [Shah and Darzi 2003], we classified the influence factors for SPs by distinguishing *surgeon-specific factors* S, *technology-specific factors* T, and *patient-specific factors* P (see Table 3.1.III for an overview of used symbols). Generally, we consider a surgical treatment to be a surgical process SP, which is a function of the outlined factors.

Technically, a surgical process SP is recorded by a measurement system, influenced by measurement system factors M. The *measurement system factors* M therefore influence the representation of a surgical process SP by a surgical process model SPM: $M: SP \rightarrow SPM$.

Additionally, we arranged the influence factors and the validation metrics into three groups: extraneous, independent, and dependent variables.

Table 3.1.III: Symbol overview.

Symbol	Meaning
$s_i \in S$	surgeon-specific factors
$p_i \in P$	patient-specific factors
$t_i \in T$	technology-specific factors
$m_i \in M$	measurement systems factors
$sp_i \in SP$	surgical process
$spm_i \in SPM$	surgical process model
$vm_i \in VM$	validation metrics

Extraneous variables

Surgeon-specific factors $s_i \in S$ that influence a surgical process are mainly the human factors of surgeons [Shah and Darzi 2003] and the staff in the OR. Two actors performed the simulations of our study: one played the role of the surgeon, and the other played a combined role of assistant and scrub nurse. The surgeon-specific factors were not considered separately because the actors were directed to follow the work steps of the audio representations of the Gold Standards closely.

Surgical Processes vary due to the use of different surgical tools, instruments, and devices. The technology factors $t_i \in T$ were also considered as extraneous variables, not separated, and constant for the study due to the predefinition of instrument names, usage times, and order by the simulation scripts.

We introduced the patient-specific factor group $p_i \in P$ to indicate the patient's current situation, his or her history or future, and his/her specific anatomical and pathological circumstances. We considered the patient-specific factors group as an extraneous variable and constant because the simulations were performed on 3D-Rapid Prototyping models, which all use the same template.

For the study, we focused on data acquisition by human observers, supported by the surgical workflow editor. We classified the measurement system into influence factors $m_i \in M$. We considered m_1 as structural ontology, m_2 as content ontology, and the surgical workflow editor as observation support software m_3. For the observer, we opted for the factors m_4 as the observation workload and m_5 as the knowledge level of the observer. We considered $m_1, \ldots, m_5$ as extraneous variables, assuming them to be constant.

Independent variables

The focus of this study was the validation of accuracy differences in SPM resulting from different data acquisition strategies as factor m_6, and different observer populations as factor m_7. Data acquisition by observers may be performed intra-operatively as live observation or post-operatively from videos. The observer populations (m_7) consisted of ten individuals: five medical students (4th-6th semester) and five technical students. None of them was experienced in SPM recording. Each of them performed nine observations each for video and live situations (3 observations for each of the 3 Gold Standards) in random order. Live simulations were performed 18 times because only 5 observers could observe simultaneously due to space limitations. Each Live simulation was recorded on video to serve as the Bronze standard for observations recorded in that particular session.

Dependent variables

We defined six different metrics for validation within the context of surgical process modeling: $vm_i \in VM$. The six metrics were designed to cover the facets that characterized the quality of data acquisition for SPM and were complementary to each other. For an overview of the computational order of the validation metrics, the reader is referred to Figure 3.1.2.

- The measurement of the structural outliers of an observation (vm_1) focused on the compliance of granularity guidelines of an observation compared to its reference. Structural outliers are measured as the percentage of outlier flow objects to all flow objects in the observation. *Structural outliers are defined in the context of the study as structural parts of a surgical process model, spm_1, contradicting the structural parts of the reference surgical process model, spm_2, by assuming that both $spm_i \in SPM$, are triggered by the same structural ontology and represent the same instance of a surgical process $sp_i \in SP$.* In Figure 3.1.3, the various interpretations for structural outliers are shown. Flow objects registered in a 1:1 relationship to their referential flow objects reflected the *correct granularity*. Flow objects that appeared in the observation, but not in the reference, were denoted as *additional observations*. Flow objects in the reference that were not recorded were *missing observations*. Flow objects represented as multiple activities in the reference, but represented as one activity in the observations, were denoted as *decreased granularity*. One flow object of the reference represented as multiple flow objects in the observations represented *increased granularity*. A mixture of multiple flow objects in the reference and multiple flow objects in the observations represented *mixed granularity*. Before applying the validation metrics vm_2 and vm_3, all flow objects not representing the *correct* granularity were removed because only similar granularities may be compared.

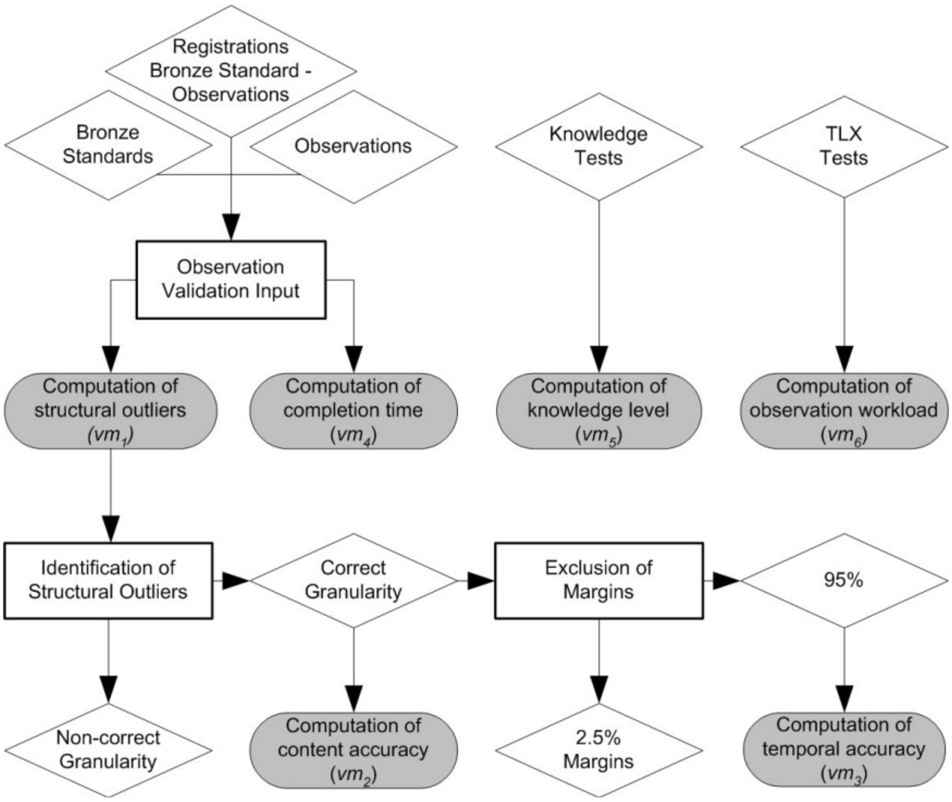


Figure 3.1.2: Computation of validation metrics.

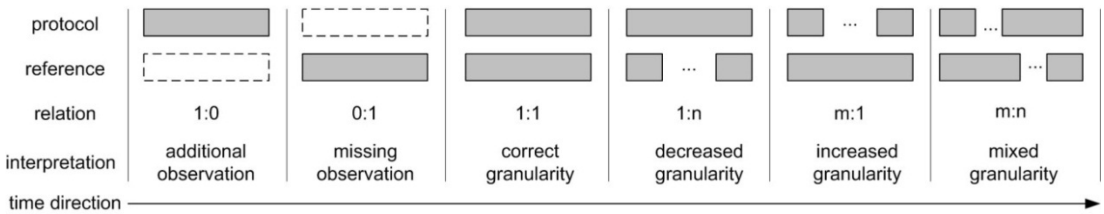


Figure 3.1.3: Types of structural outliers.

- The validation metric vm_2 estimates the validation criterion of content accuracy of an observation. Content accuracy was defined as the *distance of conceptual instances in a surgical process model,* spm_1*, compared to the conceptual instances in the reference surgical process model,* spm_2*, assuming that both* $spm_i \in SPM$*, are triggered by the same content ontology and represent the same instance of a surgical process* $sp_i \in SP$. The metric vm_2 calculates a similarity measure for the content accuracy of perspective attributes of an observation compared to the corresponding perspective attributes of the reference. Based on the perspective attributes, a content accuracy value for each perspective was calculated. Subsequently, this procedure was repeated for activities, state transitions, and events as flow objects of the overall protocol.

- The measurement of the temporal accuracy of an observation (vm_3) indicates the temporal distance between durations of activities that was calculated based on the start timestamps and stop timestamps of registered flow objects. The calculation has been done after the rejection of temporal outliers corresponding to abnormal excessively large time deviations due to hardware failures.
- The measurement of the completion time of an observation (vm_4) is calculated as a ratio based on the time needed to create the observation protocol with respect to the duration of the reference. It begins when the first activity is set and ends when the final protocol is saved after review by the observer.
- Experimental conditions were controlled by the knowledge level of an observer (vm_5) and the assessment of workload observation from observer feedback (vm_6). vm_5 was used to check the learning curve of the observers. This parameter is expressed as the percentage of correct answers on the knowledge tests. The software feedback vm_6 assessed the workload of the observation task by subjective ratings of the criteria of the NASA Task Load Index [Hart and Staveland 1988].

Results

Detailed results for structural outliers are presented in Table 3.1.IV. Medical students recorded granularity correctly 92.3% (±5.7%) of all activities in the reference in live observations and 92.5% (±5.2%) in video observations, as opposed to 86.6% (±6.8%) in live observation and 91.2% (±6.7%) in video observation for technical students. The mode of data acquisition was significant. Video observations were more accurate in terms of correct granularity. Missing activities and activities with decreased granularity were more prevalent in the live observations. The observer population also had a significant influence on structural outliers. For instance, medical student observers were more likely than technical students to record granularity correctly.

Table 3.1.IV: Study results for structural outliers (vm_1).

	m_6 live		m_6 video		Significance	Significance
(granularity) [%]	m_7 medical	m_7 technical	m_7 medical	m_7 technical	m_6	m_7
additional observation of activities	1.5±0.9	1.9±1.7	1.3±1.6	3.6±5.9		F=11.4, p=0.001
missing observation of activities	2.1±2.1	3.7±3.1	1.2±1.3	2.5±4.2	F=6.0, p=0.02	F=9.1, p=0.001
correct granularity of activities	92.3±5.7	86.6±6.8	92.5±5.2	91.2±6.7	F=7.4, p=0.01	F=16.2, p<0.001
decreased granularity of activities	0.4±1.1	1.3±1.9	1.7±2.8	0.8±2.2	F=14.5, p<0.001	F=7.1, p=0.01
increased granularity of activities	4.9±4.8	8.6±4.9	4.2±4.4	4.1±4.4		
mixed granularity of activities	0.03±0.02	0.1±0.3	-	0.3±1.3		
correct granularity of events	57.1±19.3	51.9±19.7	23.5±35.0	37.2±41.6	F=31.4, p<0.001	
correct granularity of state transitions	87.4±10.5	89.1±7.3	68.2±39.5	74.0±35.2	F=18.4, p<0.001	

The overall content accuracy for activities is 91.5% (±5.4%) in live and 91.5% (±5.3%) in video observation by medical observers. Content accuracy for activities was 88.9% (±2.6%) for live and 87.4% (±8.9%) for video observations by technical students (cp. Table 3.1.V). The data acquisition type had no significant influence on content accuracy for activities, but video observations produced significantly lower content accuracy for events.

Table 3.1.V: Study results for content accuracy (vm_2).

	m_6 live		**m_6 video**		**Signifi-cance**	**Signifi-cance**
(content accuracy) [%]	**m_7 medical**	**m_7 technical**	**m_7 medical**	**m_7 technical**	**m_6**	**m_7**
functional perspective of activities	93.1±6.7	87.1±5.6	92.9±7.8	82.7±14.1		F=36.7, p<0.001
organizational perspective of activities	97.8±3.0	97.7±2.5	98.3±2.4	96.9±6.5		
operational perspective of activities	88.3±7.3	84.5±4.4	89.2±6.2	84.8±13.2		F=11.8, p=0.001
spatial perspective of activities	70.8±9.0	70.8±8.8	70.8±10.6	68.9±9.8		F=12.5, p<0.001
total content accuracy of activities	91.5±5.4	88.9±2.6	91.5±5.3	87.4±8.9		F=15.1, p<0.001
total content accuracy of events	93.2±20.2	96.8±10.7	72.8±41.8	71.9±38.7	F=25.3, p<0.001	
total content accuracy of state transitions	98.1±5.1	99.1±3.5	99.0±4.6	98.2±8.7		

The mean absolute value for temporal accuracy was less than 2 s. for all factors. The data acquisition type had only low significant influence on temporal accuracy (cp. Table 3.1.VI). The observer population had a significant influence on temporal accuracy.

Data acquisition from videos required 80 % more time than data acquisition for live observations. No significant differences were found in completion time between medical and technical observers.

Table 3.1.VI: Study results for temporal accuracy (vm_3) and completion time (vm_4).

	m_6 live		m_6 video		significance	significance
	m_7 medical	m_7 technical	m_7 medical	m_7 technical	m_7 medical	m_7 technical
temporal accuracy [s]	1.7±0.4	1.9±0.5	1.5±0.3	1.8±0.8	F=4.9, p=0.03	F=12.1, p=0.001
completion time	1.6±0.1	1.2±0.2	2.3±0.5	2.1±0.4	F=389.7, p<0.001	

Nearly all workload criteria, and also the estimation of one's own performance, were rated higher for live observations (cp. Table 3.1.VII). All workload criteria were rated more demanding by the technical observer population.

The Gold Standards had a significant influence only on the number of structural outliers. Medical students scored 94.1 % correct answers on the knowledge tests, while technical students scored 78.3 %.

Table 3.1.VII: Study results for observation workloads (vm_5).

	m_6 live		m_6 video		significance	significance
(criteria)	m_7 medical	m_7 technical	m_7 medical	m_7 technical	m_7 medical	m_7 technical
Effort	60.0±21.2	69.6±11.3	51.6±24.5	67.2±11.5	F=6.5, p=0.01	F=19.7, p<0.001
Frustration	49.0±18.0	46.8±11.7	38.2±19.2	42.3±14.0	F=14.6, p<0.001	
Mental Demand	58.7±19.6	69.6±15.0	54.9±20.6	70.0±16.9		F=18.9, p<0.001
Performance	47.1±16.0	48.8±13.4	37.8±17.0	46.5±16.8	F=8.7, p=0.004	F=5.3, p=0.02
Physical Demand	41.5±23.7	61.0±12.7	35.4±20.4	57.6±10.1		F=52.3, p<0.001
Temporal Demand	73.7±17.2	77.4±14.0	44.0±19.0	56.8±16.5	F=39.2, p<0.001	F=11.0, p=0.001
Total Workload	62.4±14.2	67.1±10.3	49.1±15.3	60.3±10.0	F=39.2, p<0.001	F=17.8, p<0.001

Discussion

Significance of the work

To the best of our knowledge, this is the first extensive validation of knowledge acquisition for surgical process models in the medical domain; no previous comparisons of live and video observations were found in the literature. Based on a rigorous control of influence factors [Jannin et al. 2008] affecting surgical processes and the definition of validation metrics, a complex and rigorous study has been designed and conducted.

Former studies validated observations based on inter-observer agreements [Reneman et al. 2005; Baglio et al. 2004] and used correlations as indirect metrics to quantify the agreements. For valid observations, a threshold of 85% inter-observer agreement is reported [Baglio et al. 2004]. Our results were calculated based on direct comparison of observation results with the observed process as a reference.

We found that observers generally record accurately, robustly, and reproducibly. The accuracy of data acquisition for live or video observation was comparable.

The results for structural outliers give a measurement for the assessment of the granularity of an SPM. Nearly all of the activities were observed with correct granularity. In contrast to the observer population, the influence of the data acquisition type had low significance. We may conclude that differences between video and live observations of activities regarding the validation criterion of structural outliers are not statistically significant.

The observations for state transitions and events were unacceptable. Seemingly, the concentration of the observers was focused on the interventional site and on the monitor displaying the endoscope view, not on the monitor displaying the state transitions and the events. This might be compensated by introducing acoustic signals that highlight them for the observers or perhaps even for the surgeons themselves in the operating room.

Content accuracy showed no significant differences between the data acquisition strategies. Thus, we conclude that live and video observations may be considered similar regarding the validation criterion of content accuracy. The medical observers recorded the activity content significantly better than the technical observers. Low accuracy occurred mainly because students could not properly assess the spatial perspective. None of the perspectives showed a significant difference by data acquisition strategy. However, there is still work to do to develop a method for direct global content accuracy comparison that accounts for the positive and negative variation in granularity.

The completion time was far longer when recording from videos than from live simulations. This result is especially interesting when considering the comparable outcomes of live and video observations for granularity, content and temporal accuracy.

The small increases of the measured ratios of the knowledge tests during the data acquisition sessions showed the effectiveness of the training sessions. We trained the technical observers in a similar manner to the medical observers, but they were not able to attain the same level of knowledge. Values for all workload criteria were lower for video observations. Technical observers rated all workload criteria more demanding than medical observers. This may have influenced the lower granularity, content accuracy, and temporal accuracy of the technical observers (compared to the medical observers).

The validity of the simulation was checked by comparing the Bronze Standards to the Gold Standards. In the context of the study, the Gold Standards were held as the objective and unequivocal models that were, by definition, the simulation scripts. The Bronze Standards were viewed as the best results that the observers could achieve. The simulation validation was used to cross-check the validity of the simulations. For instance, the mean ratio of correct granularity of the Bronze Standards was 97.2%, and the mean content accuracy was 96.3%. Thus, the actors introduced only a very few simulation errors.

The realism of the Gold Standards did not comprise each possible aspect of a FESS but worked as a robust simulation base for a continuous repetition of the surgical processes. Furthermore, the experimental design was used to hold constant the influences P, S, and T of the patient, the surgeon, and the technology on the surgical process. To study s_i, for instance, one would evaluate differences resulting from different personal 'styles' of surgeons or different education levels that may result in different procedure courses (if the same intervention was performed twice on the same patient by two surgeons).

We referred to the reasons for variation in surgical processes caused by using different surgical instruments or devices as technology-specific factors. The limited number of instruments representing the technology-factors T represents a restriction, but this was ignored to facilitate the work of the actors.

Advantages and disadvantages of live and video observations are shown in Table 3.1.VIII. The choice of the data acquisition strategy does not only depend on the objectives of clinical studies to be performed, but also on the available resources for observation.

Table 3.1.VIII: Advantages and disadvantages of video and live observation.

	Video observation	Live observation
Advantages	- temporal resolution can be increased by pausing the video - knowledge acquisition can be repeated, if the structural ontology or the content ontology need to be altered - workload of the observers is less than in live observations	- instantaneous access to information and possibility to ask for hidden information, e.g. surgical decisions - dynamic repositioning of the observer in the OR, e.g. if line of sight is blocked
Disadvantage s	- not all information for the SPM can be captured on video - field of view can be blocked by intervention participants - high costs of time for data acquisition	- loss of information due to distraction or increased workload of the observer - limited temporal resolution

Limitations of the present study

Limitations to our work include:

- The observations were based on simulated surgical processes. Of course, simulations are not 100% realistic. Ideally, the study would have used real surgical cases, but that would have prevented control of many factors that could affect results.
- The validation metrics used for assessing the quality of data acquisition for surgical process models need to be validated in additional studies.

Implications for future work

In this study, we proposed an innovative experimental design for the validation of knowledge acquisition for surgical process models. This validation method may be extended and modified, and it may be used to validate modifications of S, T, P, or M. The actual design could be proposed as validation support for other more technical approaches such as those described in [Sudra et al. 2007] or [Padoy et al.]. We are unaware of any research results that delineate which measures of observations for SPMs are acceptable and which are not; we plan to address this topic in future work. Additionally, knowledge bases could be developed and validated to support and facilitate observations for SPMs. If a knowledge base were used that contained information about which actions can be performed with a specific surgical instrument, e.g., Blakesley, the surgical workflow editor could propose the action *dissect* to the observer and ask for confirmation, as soon as *Blakesley* is chosen as instrument.

Conclusions

The results of this study can provide useful guidance for the design of other studies to acquire knowledge for SPMs. We demonstrated the validity of video as well as live observations for modeling SPMs and that trained human observers generally record accurately, robustly, and reproducibly. We also outlined the areas where human observations were less accurate; future work should concentrate on these areas. Live observations of state transitions and events should be supported by a technical sensor system with intra- or post-observation synchronization to the observer protocol or an acoustic signal that draws the attention of the observer to the displaying device. For routine use, we recommend live observations due to their relative speed, flexibility, and effectiveness. If high precision is needed or SPM parameters, such as the ontologies used, are altered during the study, video observations are preferable. Trained medical students can be highly accurate observers.

This study also provided an estimate of the expected accuracy of modeling surgical processes by observation. We identified influence factors that can serve as basis for designing similar studies, in which, for example, the work of surgeons with varying levels of experience or the effect of the use of different surgical instruments might be compared. Our validation metrics can be applied to studies with comparable reference standards, but producing such references is a significant challenge.

Modeling surgical processes is undoubtably a challenge for the observers. Special advance training is required, for example, for live observations in the operating room. The study setup, of course in a narrower context, as well as the validation metrics, can be used to benchmark the level of observers in training. For instance, if it were important for the observers to achieve a certain degree of content accuracy before they can participate in clinical studies, the methods used in this study could be used to measure their proficiency.

Acknowledgements

We thank the team that supported the study at the Innovation Center Computer Assisted Surgery, University of Leipzig: T. Baumgärtner, M. Ceschia, M. Czygan, F. Eckhardt, J. Georgi, D. Goldstein, A. Hoffmeier, K. Hornoff, B. Kaschek, N. Ritter, and S. Schumann. We would also like to thank I. Hertel and M. Hofer of the Department of ENT-Surgery, University Hospital Leipzig for their support.

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) within the scope of the Unternehmen Region with grant numbers 03 ZIK 031 and 03 ZIK 032.

3.2 Observer-based data acquisition with adaptive user interfaces

Title

An observation support system with an adaptive ontology-driven user interface for the modeling of complex behaviors during surgical interventions

Authors

Thomas Neumuth[1], Bernadett Kaschek[1], Dayana Goldstein[1], Marcello Ceschia[1], Jürgen Meixensberger[1,2], Gero Strauß[1,3], Oliver Burgert[1]

Departments

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] University Medical Center Leipzig, Department of Neurosurgery, Leipzig, Germany

[3] University Medical Center Leipzig, Department of Otorhinolaryngology, Leipzig, Germany

Citation

Neumuth T, Kaschek B, Goldstein D, Ceschia M, Meixensberger J, Strauss G, Burgert O. An observation support system with an adaptive ontology-driven user interface for the modeling of complex behaviors during surgical interventions. Behavior Research Methods. 2010; 42:1049-58.

Keywords

observation support software; ontology; validation study; surgery; process description; human behavior

Abstract

The field of surgical interventions emphasizes knowledge and experience; explicit and detailed models of surgical processes are hard to obtain by observation or measurement. However, in medical engineering and related developments, such models are highly valuable.

Surgical process modeling (SPM) deals with the generation of complex process descriptions by observation. This places high demands on the observers, who have to use a sizeable terminology to denominate surgical actions, instruments, patient anatomies, or describe processes unambiguously. Here, we present a novel method, employing an ontology-based user interface that adapts to the actual situation and describe the principles of the system.

A validation study showed that this method enables observers with little recording experience to reach a recording accuracy of >90%. Furthermore, this method can be used for live and video observation.

We conclude that the method of ontology-supported recording for complex behaviors can be advantageously employed when modeling surgical processes.

Introduction

Surgery requires considerable experience and skill. The surgical process involves highly complex procedures and depends on a multitude of factors that entail the surgeon's awareness and attention to patient-specific abnormalities in anatomy and pathology or technical resources. With respect to the multifaceted and highly variable processes of surgical interventions, a modeling and behavior observation strategy is necessary to account for this complexity. Until today, a concise form of report that is able to reproduce surgical process evolution in a detailed and accurate way is missing. A comprehensive observation and modeling tool that could record the progress of processes such as surgical interventions would be a novel and insightful approach to this challenge.

The availability of expressive process models supports a multitude of other application areas in medical engineering, aside from medical training, most notably the performance of requirements analyses [Neumuth et al. 2009c] or the introduction and evaluation of new therapeutic standards and surgical assist systems [Strauß et al. 2006a]. Evolutionary models of surgical processes, surgical process models (SPM), can further the understanding, reproduction, analysis, training, and teaching of surgery. However, to be used in the aforementioned application areas, the recordings and observations must be accurate and comprehensive.

In the literature, various works have focused on the modeling of process sequences in surgery. Some research groups have used interviews as main basis for their investigations [Jannin et al. 2003; Raimbault et al. 2005], while others propose the use of dedicated measuring systems, which record partial information about surgical process steps [Ahmadi et al. 2006; James et al. 2007; Padoy et al. 2007]. However, the most frequently used techniques are based on observers and range from data acquisition without software support [Mehta et al. 2002; Malik et al. 2003; den Boer et al. 1999; Strauß et al. 2006a] to observer-based methods that are supported by software systems [Neumuth et al. 2009b]. Moreover, several software solutions [Castellano et al. 2008; Hänninen and Pastell 2009; MacLin and MacLin 2005] and combined hard- and software solutions [Held and Manser 2005; Sarkar et al. 2006] have been proposed in the context of behavior analysis.

The procedure of modeling surgical processes by observation, especially with the objective of performance assessment in the context of surgical training [Leong et al. 2007; Richards et al. 2000; Rosen et al. 2006; Megali et al. 2006] or on comparing strategies for surgical treatment [den Boer et al. 1999; Strauß et al. 2006a] generally focuses on two measurement strategies: high-resolution, sensor based measurement of the performance of a limited number of surgical actions, as for instance instrument movement trajectories while placing knots, or low-resolution and simpler observer based measurements for surgical interventions or interventional phases, such as [Schuster et al. 2007; Archer and Macario 2006], without reference to specific surgical process steps. In previous work, we proposed an approach that allows for a medium level of granularity to be used for the decomposition of surgical process steps into categories, as described in the following section, in order to accommodate the complexity and diversity of information and the high variability of surgery. An observation support software, known as the surgical workflow Editor [Neumuth et al. 2006a; Neumuth et al. 2006b; Neumuth et al. 2009a; Neumuth et al. 2009b], was developed for such observation on a medium level of granularity.

Due to the diverse information that can be acquired with the observation approach for surgical processes on a medium level of granularity, high demands are put on observers, because, apart from being under continuous time-pressure during the observation, they also need to deal with extensive surgical and anatomical terminologies. For this reason, we introduce the usage of ontologies as knowledge bases to support the observer by means of an adaptive user interface, a situation-dependent edition of the observation software interface. The use of a knowledge based system is necessary to overcome initially contradictable objectives posed by the new observation strategy: On the one hand, due to the high variability of surgical processes, the recording system needs to be able to deal with a large number of surgical terms to achieve expressiveness, and, on the other hand, this information needs to be declarative to abstract higher-resolution information and to allow for usability by clinical and non-technical users.

To our knowledge, there is currently no method available that deals with the support of the observer using knowledge based software, although adaptive software interfaces are in use in several applications in computer science [Berrais 1997; Kuehme 1993; Love et al. 2008]. With the help of such interfaces, the amount of terms available in observation situations can be considerably decreased and thus reduce the observers' workload. An implementation of the proposed method could also be employed to great advantage in other fields of science that rely on data gathering as the basis for observation or monitoring, such as behavioral or educational science.

This work presents the working principle of the observation support system with the adaptive user interface, shows its technical implementation, and presents the results of a validation study that demonstrates the accuracy of the system. Within the validation study it is shown that the system is applicable to live and video based observation. Furthermore, the study introduces several measurements to assess granularity, content accuracy and temporal accuracy of observed surgical process steps, as well as observer workload. The study is based on simulated interventions from the field of Otorhinolaryngology; an application to other surgical disciplines or to use cases in behavior research is possible, but is not discussed explicitly.

Materials and methods

Development of ontology and materials used for testing

Some specific terminology will be used and explained in this section. Furthermore, the software which has been used will be presented, and its functionality explained.

Surgical process models

Following the Workflow Management Coalition's definition of a Business Process [Workflow Management Coalition 1999a], we define a surgical process (SP) as *a set of one or more linked procedures or activities whose instances (are intended to) collectively realize surgical objectives within the context of an organizational structure defining functions, roles, and relationships*. A surgical process is transformed into a surgical process model (SPM), a model representing the surgical intervention. The SPM as core concept of our approach is derived via an observation protocol and represents partial aspects of the original surgical process in a formal or semiformal way.

The surgeon’s work is rendered as temporally extended process steps in the SPM, called activities, and consist of various perspectives [Jablonski and Bussler 1996]. All perspectives conjointly form an activity. Each of the activities describes a different point of view on a surgical process step:

- the organizational perspective describes who performs a process step, for instance 'surgeon' or 'assistant';
- the functional perspective describes what is done in a process step, for example 'cutting' or 'suturing';
- the operational perspective indicates the technical resources that are used to perform a process step, such as 'scalpel' or 'needle';
- the spatial perspective describes at which location at the patient's body the process step is performed, for instance at the 'sinus maxillaris', and
- the behavioral perspective indicates at which point of time a process step takes place; this perspective is represented by time stamps.

Observation support software system with an adaptive user interface

The established way of composing surgical process models is to record them with the help of specially trained observers. The observer relates terms concerning the several perspectives to each other, for example, actions are assigned to single members of the operating room staff, in order to create a description of a situation. Herein, he is supported by the observation support software, the surgical workflow editor, which generates the surgical process model as observation protocol. The surgical workflow editor is a JAVA based application used for the recording and analysis of SPMs. It stores such information as is acquired by the observer, with the help of an interface which displays all possible entities needed to record surgical interventions, such as surgical instruments and activities.

However, there are some general difficulties in recording SPMs on this medium level of granularity. Firstly, a single surgical intervention can consist of up to 300 single process steps, each of which has different perspectives. Secondly, the observers need to choose the proper terms for the different perspectives out of large repositories: the designated terminology for the operational perspective might comprise about 50 different surgical instruments, about the functional perspective includes about 30 different actions, and about 20 different anatomical structures are used for the spatial perspective. Thirdly, a preliminary definition of the different observational codes, as peculiar to the realm of behavioral research, is not possible here due to the sheer number of potential combinations of terms from each of the perspectives. And, lastly, the display size of the computer screen limits the choice of representable items from the terminology.

One possible approach to overcome these problems is the implementation of an ontological knowledge base support. An ontology is an "explicit specification of conceptualization" [Gruber 1993], a formal representation of concepts and their relations. The perspectives used to describe process steps in our approach, instruments, activities, and anatomical structures were defined as ontological concepts by domain experts and logical conjunctions between different elements were described as relations that link terms to one another.

Various tools are available to design these logical constructions. We used Protégé [Stanford University 2009] to formalize the concepts, the semantic relationships between the concepts, and that allows for an implementation of enhancing applications, such as logical reasoners that test the relations for formal correctness [Racer Systems 2009]. In the proposed system, the generated Protégé ontology in OWL-format (Web Ontology Language) is saved into the ontology server and loaded when the observation support software is started.

This knowledge base support was designed to facilitate observations for the recorder and to allow the choosing of the appropriate term that best describes the most recent process step without browsing extensive terminology lists. Depending on the entity selected from one perspective, only terms from the other perspectives that can be sensibly combined with the preselected perspective are shown. For example, if 'cutting' were chosen as the functional perspective for a surgical activity, the list of surgical instruments would be restricted to those that are 'able to cut', for instance 'scalpel' and 'scissors'. An illustration is provided in Figure 3.2.1.

A) B)

Figure 3.2.1: Functionality of the ontology based adaptive user interface: After selecting the action 'suck' (A) the terminology list for surgical instruments is restricted automatically to 'suction tubes' (B). In contrast to the static interface of the conventional system the user has to choose the correct instrument from a smaller variety of items.

More precisely, our technical solution consists of three main components (see Figure 3.2.2): the ontology server that contains the knowledge base, the adaptive user interface, and the editor engine. The knowledge base contains the concepts and the relations between the concepts that are necessary to record a surgical intervention. The adaptive user interface is the input mask for the instantiation of items by the user and represents the contents of the perspectives. It adapts automatically to the current situation according to the actual user input and the knowledge base response. The editor engine administers and delegates the central business logic, as the communication management with the ontology server, for example.

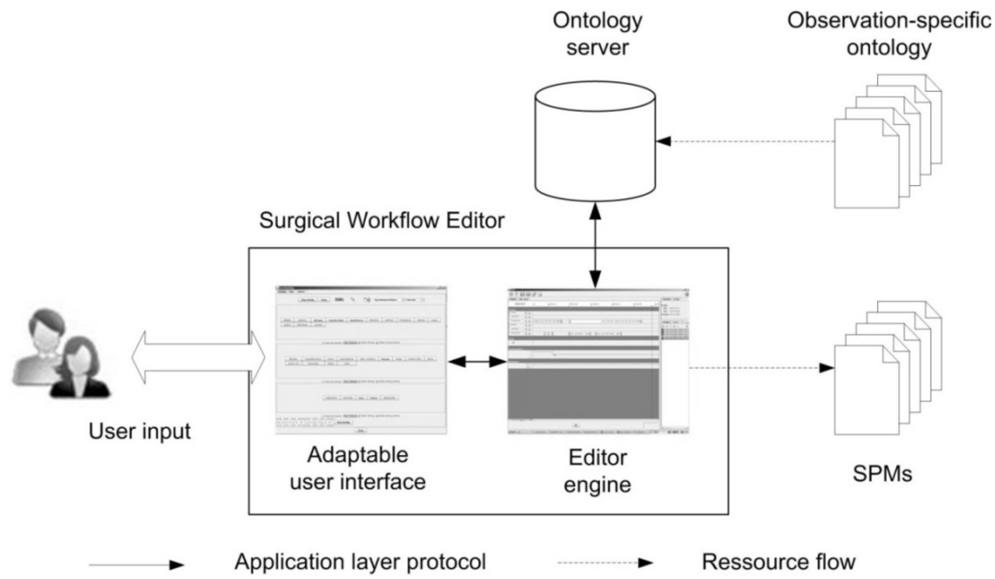


Figure 3.2.2: Software component infrastructure and data flow for the adaptive knowledge based observation support system.

Surgical application case and simulation scripts

System validation used Functional Endoscopic Sinus Surgeries (FESS) as an example. In Germany, it is one of the most frequently performed interventions in surgical Otorhinolaryngology: about 50,000 such procedures are performed annually [German Federal Statistical Office 2008a]. This minimally invasive, endoscopic intervention involves examining nasal cavities, such as the ethmoid sinuses, and removing diseased or obstructive tissue or growths, such as nasal polyps. The goal of the FESS is to purposefully remediate diseased areas and restore natural drainage and ventilation paths. During the intervention, the surgeon generally handles the endoscope with one hand, while the other hand performs surgical activities, such as the removing of tissue with forceps or exhausting liquids.

The basic purpose of the validation study was to design and act out FESS interventions according to simulation scripts, as is shown in Figure 3.2.3. The simulation scripts were developed by ICCAS Institute in close cooperation with surgeons from the Department for Otorhinolaryngology at the University Hospital Leipzig. Three simulation scripts were developed, each the variant of a different typical type of FESS surgery. These simulation scripts contained detailed patterns for process steps of FESS interventions for two actors (‘surgeon’ and ‘assistant’), 16 surgical instruments, 13 functional tasks, such as ‘cut’ and ‘clean’, and 7 anatomical and pathological structures. Figure 3.2.3 shows examples of devised process steps.

Each simulation script had a total duration of about 20 minutes and contained 60-90 single process steps. As preparation for the validation of our system, the three simulation scripts were read aloud and thus recorded as audio file instructions.

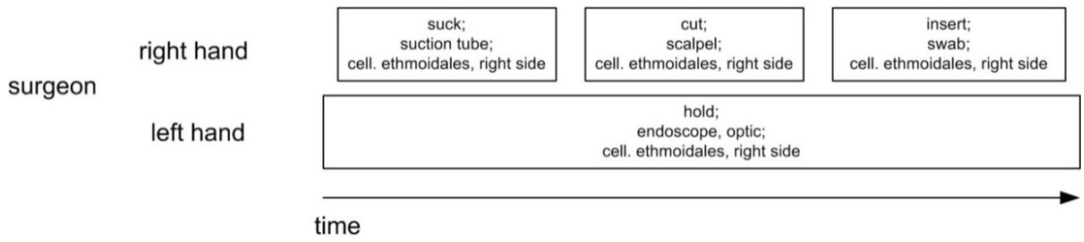


Figure 3.2.3: Cut-out procedure of simulation script for the FESS simulation with detailed instruction for the actors.

Participants

For the present study six observers, referred to as novices, were recruited to record the SPMs. None of the observers had any previous experience with recording or describing surgical interventions, nor had they used the software before. The six novices were then trained to collect data for SPMs in an introductory workshop. Additionally, they were introduced to the topic of FESS interventions, the intervention type's typical progression, its surgical goal, the instruments used, and the anatomical and pathological structures involved. Furthermore, the introductory workshop was used to present the predefined ontology and its coherent use during the observation to the participants.

Out of the six novices, three had a medical background, studying human medicine, while the other three had an engineering or computer science background. This distinction was made to test whether observers without medical background can achieve the same quality of results as the medical students, as our experience from the last years has shown that non-medical students have a less tight curriculum and therefore more time for extra-curricular activities.

Study design

The presented recording system has been validated in a complex study setup. On the basis of simulated surgical procedures, observers were asked to generate SPMs as precisely as possible using the observation software. Subsequently, the accuracy of observation protocols was compared to the simulation scripts (see Figure 3.2.4).

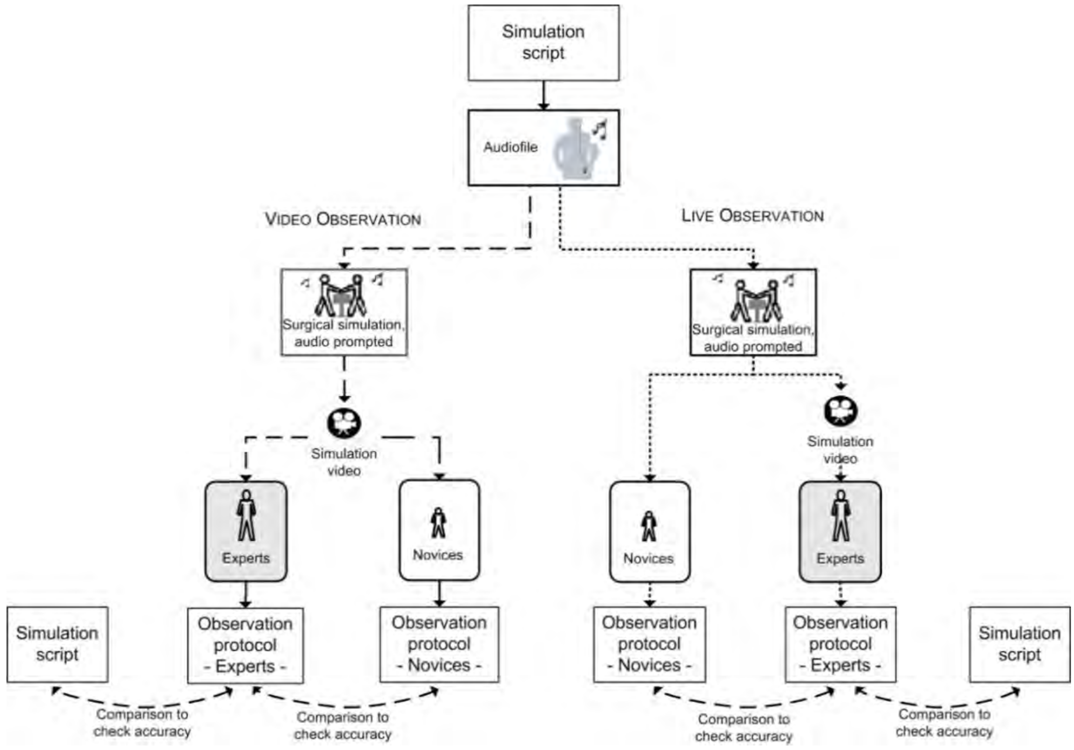


Figure 3.2.4: Overview of the system validation study showing the data flow from the design of the simulation scripts to the protocols.

Preparation of data acquisition

The simulation scripts were performed by actors that received instructions about the process steps from an audio file by means of mp3 players and earphones. The simulation was performed on anatomically correct and detailed paranasal models, as shown in Figure 3.2.5.

For the video observations, each of the three simulation scripts was performed once, while being recorded by multiple video cameras. The resulting videos were then cut short and rendered into presentations. Two observation experts having significant experience with the surgical workflow editor generated the SPMs for the filmed simulations and converted the protocols to XML format. These observations were compared to the simulation scripts as references for the validation of the simulation itself.

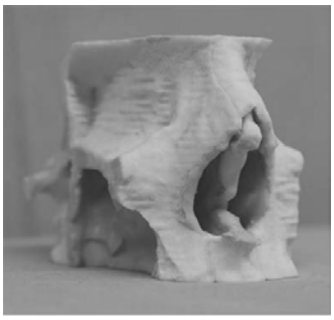


Figure 3.2.5: 3D-print model of the nasal cavities of a real patient as used in the system validation study.

Figure 3.2.6: Data acquisition example. Novices watching the simulation scripts performed by the actors (A) and operating a Tablet-PC (B) to record the surgical process. The surgical process step 'dissection of material with a Blakesley' (a specific type of surgical pliers) is seen as performed by the actors and on the endoscope screen (C). The observer selects 'Blakesley' or 'dissect' in the observation support software and the interface adapts according to the relation between 'Blakesley' and 'dissect' (D).

Data acquisition

Live observation was performed with the help of live-acted simulations, which were journalized by the novices. In addition, these simulations were also documented on video. After the recording, the two experts compiled their observations from the video documentation of the simulations as references. These were again compared to the simulation scripts for the validation of the simulation.

Data acquisition was performed in the demonstrator OR at the ICCAS of Leipzig University. The overall study took four days within a time frame of two weeks. Following an introductory workshop, all observers had to complete live and video observations with tablet PCs thrice for each of the simulation scripts (see Figure 3.2.6). The succession of live and video observations alternated and the sequence of the simulations was randomized. To summarize it can be said, that each observer recorded all three simulations thrice, resulting in 9 protocols. This holds true for the live and the video recordings, accounting for a total number of 18 protocols. Thus, considering all 6 observers, an overall amount of 108 protocols was achieved.

Post-processing and analysis

After data collection, the SPMs of the simulation scripts, the observations by the experts, and the observations by the novices were transferred to a PostgreSQL 8.3 data base [PostgreSQL Global Development Group 2009]. The appraisal followed the three-stage concept of the study. In preparation for validation, the observations by the experts and the observations by the novices were compared to validate the accuracy of the observations. Furthermore, each of the observations by the experts was compared to its corresponding simulation script to validate the simulation accuracy. This was accomplished with the help of a special software tool, which represented each recorded activity in relation to its respective reference. The experts could then decide manually whether or not the novices had recorded the right action, instrument, and anatomical structure. The observations by the experts were employed to retain differences between simulations and the observations by the novices, which occurred when actors made mistakes during the simulated FESS interventions.

The data acquisition method, live or video recording, the background of the observers, medical or engineering, and the number of the simulation script were regarded as independent variables. As for the validation of the accuracy of the knowledge based observer support system, five different dependent variables were analyzed: granularity, content accuracy, temporal accuracy, completion time, and workload for the observers.

The goal of measuring the granularity of the recorded process steps was to determine the ratio of structurally correctly recorded activities in the observation by the novices in reference to the respective observation by the experts. A correct granularity would be a $1{:}1$-relationship between the recorded activity and the respective activity in the reference. Incorrect activities were regarded as additional observations that did not appear in the reference ($0{:}1$-relation) or missing observations that were missed in the observations by the novices but appeared in the observation by the experts as reference ($1{:}0$-relation). Other possibilities were the logging of single activities as multiple activities (increased granularity, $1{:}m$, $m > 1$), the subsumption of different activities into a single activity (decreased granularity, $n{:}1$, $n > 1$) or mixed granularity of activities ($m{:}n$, $m \neq n$, $n > 1$).

The content accuracy was determined by a comparison of the content of perspectives in the observation protocol with the perspectives in the particular references. A total correlation was appraised as 1 and a deviation as 0. Subsequently, the percentage of correctly observed activities for the respective perspective was determined.

The temporal accuracy identified the absolute value of temporal deviation between the duration of activities in the observation and reference. The measurement of the completion time is expressed as ratio between the duration of the observation and the duration of the simulation. The observation duration measured the expenditure of time needed to complete the whole protocol by indicating the temporal margin between the start of the recording and the final release of the workflow protocol by the observer.

In addition to examining the outcome of the observation, we have also tested on the usability of the observation support system with the adaptive user interface. For this goal, the NASA task load index (TLX) [Hart and Staveland 1988; Cao et al. 2009] was employed, which was specifically intended for operators of human-machine systems and which gathered subjective information about physical workload. The TLX includes six subscales: Mental Demands, Physical Demands, Temporal

Demands, Performance, Effort, and Frustration. The novices were asked to fill out a questionnaire after each observation.

The statistical analysis was conducted with the help of a GLM (Generalized Linear Model) for the independent variables live or video observation, medical or engineering observers, and the simulation script. All statistical tests were conducted with SPSS 15 [SPSS Inc. 2008] at a significance level of $\alpha = 0.05$.

Results

In the context of the validation study, the novices had to record SPMs in live and video observations according to the methods described in the previous section. Data acquisition was performed using the adaptive user interface of the surgical workflow editor. Results are presented in Table 3.2.I.

The observers reached a mean correct granularity of 92.8% ± 7.3% (mean ± standard deviation). With regard to granularity, video observation was 8.6% more accurate than live observations. Furthermore, medical students recorded surgical behavior with 8.4% more accurate granularity than engineering students. In contrast, decreased granularity was more frequent in live observation protocols. No mixed granularity was observed.

The total content accuracy was 92.3% for the medical group; whereas the engineering group achieved 93.5%. Arguably, the main differences originated from the operational perspective. The temporal fidelity of the activities in the observation protocols showed a comprehensive mean error of 2.0 s ± 1.6 s and no statistically significant difference between the groups.

Only slight disparities were observed between the observer populations with regard to completion time, as is shown in Figure 3.2.7. However, the difference between the data acquisition strategies live and video observation in this respect was highly significant (p<0.001).

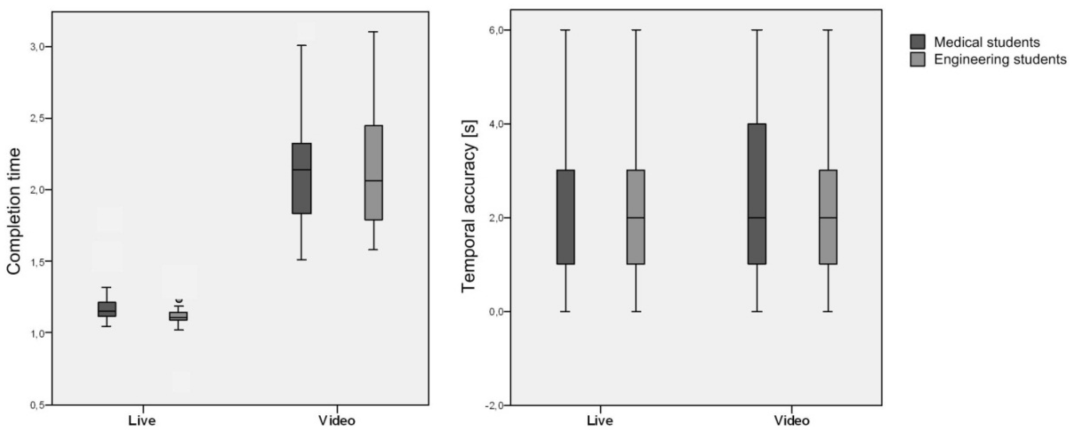


Figure 3.2.7: Study results for the completion time (left) and temporal accuracy in seconds (right). The boxplots provide medians and sample range for observations by medical and engineering students in live and video settings.

Interestingly, all workload criteria were rated as more demanding for live observations than for video observations. Nevertheless, the differences were not significant, with the sole exception of Temporal Demand (p<0.001). Observers with an engineering background rated the overall workload higher than medical observers, in all regards.

The results for the different simulation scripts showed small differences between the means. The observation of simulations for one of the scripts showed slightly less accuracy concerning granularity and content accuracy.

Between-subject effects, such as the combination of data acquisition strategy and observer population, were not significant.

Table 3.2.I: Study results for granularity, content accuracy, and observer workload (df=106). Granularity and Observation accuracy values are given as percentages. Observation workload is a unitless value on a scale between 0 (no demand) and 100 (highest demand).

		Data acquisition strategy						**Observer population**						**Gold Standard**							
		live		**video**		**significance**		**medical**		**engineering**		**significance**		**#1**		**#2**		**#3**		**significance**	
		M	**SD**	**M**	**SD**	**F**	**sig.**	**M**	**SD**	**M**	**SD**	**F**	**sig.**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**F**	**sig.**
Granularity	Additional observation	0.5	0.8	0.7	1.0	F= 3.1	p= 0.08	1.1	1.1	0.1	0.3	**F= 54.3**	**p<0.001**	0.6	1.0	0.8	1.0	0.4	0.8	F= 2.9	p= 0.06
	Missing observation	3.6	7.3	2.2	3.9	F= 2.1	p= 0.15	2.0	2.9	3.8	7.7	F= 3.3	p= 0.07	2.3	5.1	0.6	1.1	5.8	7.9	**F= 8.7**	**p<0.001**
	Correct granularity	90.5	10.1	94.5	5.9	**F= 7.6**	**p= 0.007**	94.4	5.3	90.6	10.5	F= 6.9	p= 0.01	92.9	9	95.6	4.8	89	9.6	**F= 6.5**	**p= 0.002**
	Decreased granularity	0.6	1.3	0.8	1.9	F= 0.3	p= 0.59	1.1	2.1	0.3	0.8	**F= 8.6**	**p= 0.004**	1.0	2.0	0.1	0.7	1	1.7	F= 3.8	p= 0.03
	Increased granularity	5.7	6.9	3.1	4.8	F= 4.8	p= 0.03	2.6	4.2	6.2	7.0	**F= 10.5**	**p= 0.002**	4.6	7.2	3.9	4.9	4.7	5.9	F= 0.2	p= 0.84
Content accuracy	Functional Perspective	97.9	1.8	97.5	2.1	F= 4.6	p= 0.04	97.6	2.2	97.8	1.8	F= 2.4	p= 0.12	98.3	1.6	98.1	1.6	96.7	2.4	**F= 27.2**	**p<0.001**
	Organizational Perspective	98.1	2.8	98.8	1.6	F= 1.7	p= 0.20	98.7	2.2	98.2	2.3	F= 0.1	p= 0.73	99.2	1.4	99.5	0.9	96.7	2.9	**F= 6.3**	**p= 0.003**
	Operational Perspective	86.5	6.6	85.6	5.1	F= 0.9	p= 0.33	84.3	6.7	87.8	4.2	**F= 12.8**	**p= 0.001**	86.5	3.9	89.3	5.8	82.3	5.6	**F= 15.9**	**p<0.001**
	Spatial Perspective	75.5	5.6	73.7	5.9	F= 5.0	p= 0.03	73.8	4.5	75.5	6.8	F= 3.7	p= 0.06	73.6	4.4	80.0	3.7	70.3	4.5	**F= 47.9**	**p<0.001**
	Total content accuracy	92.9	2.9	92.7	2.5	F= 0.2	p= 0.67	92.2	2.6	93.5	2.7	F= 6.7	p= 0.01	93.2	2.2	94.1	2.3	91.1	2.7	**F= 13.7**	**p<0.001**

Observation Workload	Effort	56.4	18.0	51.2	20.4	F= 2.8	p= 0.10	42.1	15.9	65.5	14.9	**F= 58.6**	**p< 0.001**	55.0	19.0	52.0	20.7	54.3	18.7	F= 0.5	p= 0.62
	Frustration	48.8	18.0	41.1	17.8	F= 5.1	p= 0.03	39.6	15.8	50.1	19.0	**F= 9.2**	**p= 0.003**	42.6	16.7	44.0	17.9	48.2	20.1	F= 1.0	p= 0.36
	Mental Demand	51.2	18.5	46.3	20.1	F= 3.5	p= 0.07	35.2	14.5	62.2	13.5	**F= 95**	**p< 0.001**	46.	±19.1	49.6	19.8	49.9	19.8	F= 0.4	p= 0.66
	Performance	43.7	17.6	38.3	15.9	F= 2.5	p= 0.12	40.1	12.0	41.8	20.8	F= 0.2	p= 0.65	40.9	17.6	40.6	16.7	41.3	16.8	F= 0.0	p= 0.99
	Physical Demand	48.3	15.7	45.8	16.9	F= 0.6	p= 0.45	46.1	16.6	47.9	16.1	F= 0.3	p= 0.57	45.8	15.6	47.0	17.6	48.4	16.0	F= 0.2	p= 0.80
	Temporal Demand	56.1	17.5	45.6	18.6	**F= 11.9**	**p= 0.001**	41.1	15.2	60.4	17.0	**F= 38.7**	**p< 0.001**	50.1	17.9	51.0	19.2	51.2	19.6	F= 0.1	p= 0.94
	Total Workload	54.8	13.6	49.6	14.4	F= 5.1	p= 0.03	44.2	10.6	60.1	12.8	**F= 47.1**	**p< 0.001**	51.4	13.4	51.7	14.7	53.4	14.7	F= 0.3	p= 0.74

Discussion

In the research field of medical engineering, the employment of observational strategies is highly relevant, for instance to obtain surgical process models as base for performing requirements analyses for surgical assist systems or the evaluation of newly developed surgical instruments.

The objective of this work was the implementation of a knowledge base driven adaptive user interface for the surgical workflow editor observation support software that provides assistance to observers who need to deal with large terminologies. As we have shown, this method constitutes a robust and expressive basis for the observation of highly variable surgical behavior.

As the sample application for the proposed methodology, a validation study for the modeling of surgical interventions was designed and performed. The study results have shown that even inexperienced observers were able to attain good results by means of the new tools presented. The utilization of this method was nearly equivalent for live and video observations. Even lay users and anatomically or medically inexperienced persons attained good results with this method. However, utilizing this methodology in other fields, such as behavioral science, is also very likely and recommendable.

The results of the validation study for the proposed system showed that it is eligible for both, video and live observations, because only a few significant differences were found between these data acquisition strategies. A slightly higher correct granularity of activities was achieved in video observations. Differences between medical and engineering observers were significant for several criteria. However, for the accuracy criteria, these differences were mostly less than 5%.

Observers that have to deal with large terminologies can be adequately supported by engineering systems for recording surgical processes. This work presented a methodology for the application of a knowledge base driven adaptive user interface for observation support software. This adaptive user interface represents terminology information depending on the current situation and the system is employed to support the user in modeling surgical interventions.

Due to the refinement of surgical activities into different perspectives, which is required because of the variability of surgical processes, a multitude of possible combinations of terms emerges. In the moments of recording, the observer combines these terms from different perspectives into sensible activities that describe the surgical process steps. For this, the implementation of ontological relations between the terms of the different perspectives, such as surgical instruments and actions that can be performed with them, is reasonable. With the employment of ontological relations, large terminologies can be pre-rendered for observation tasks. This can be very informative, especially with respect to complex or variable environments, as demonstrated here, using the example of surgical interventions. Thus, the diversity of observable entities is no longer limited by engineering factors, such as the account of a large number of observation categories on a small display.

Our observation approach and the analysis are based on an approach with medium level of granularity. Granularity for surgical work, which is hard to capture in formal ways, strongly relies on subjective estimation. Due to the nature of our study, we defined the granularity of the process steps described in the simulation scripts as 'correct' and instructed our observers accordingly, in the training workshops prior to the study.

None of the novices had any previous experience in observing surgical processes or recording with the presented system. In addition, the observers with an engineering background neither had any surgical background knowledge, nor medical training in advance. Nevertheless, as shown by the performed validation study, both observer populations reached a correct granularity of about 90% and a total content accuracy of also approximately 90%. This result was achieved despite of the lacking previous experience in dealing with the observation software and with the subject matter of the observation itself.

Only the recordings of the spatial perspective showed some peculiarities. Here, neither engineering nor medical observers attained acceptable results. A possible reason for this could be the complex anatomic features of the nasal cavities which make observations highly demanding, even for medical observers. Another reason for the difficulties in determining the correct anatomical locations could be the ontology itself. Even though ontological relations between the surgical actions and the instruments are relatively explicit, the triangle among the functional, operational, and spatial perspectives has to be rather loosely interpreted because most combinations of surgical actions and instruments can be applied to anatomical structures.

It can also be presumed from the results that the observer training and the performance of the surgical simulations were appropriate. The duration of the training workshop was mostly adequate, as attested by the average accuracy of about 97% for each perspective, with the sole exception of anatomical structures, as previously mentioned. Seemingly, more experience in this field is advisable. The simulation accuracy of the surgical simulations was validated by comparing the observations by the experts and the simulation scripts. Here it became apparent that the surgical simulation was accurate, achieving 99.9% ± 0% in granularity and 99.5% ± 1.6% in content accuracy. The statistical results for the simulation scripts showed that all three scenarios were commensurable. All results for the metrics had the same magnitude.

The analysis of the NASA TLX index has shown that the observation support system with the adaptive user interface is an appropriate means for both data acquisition strategies. As expected, live observations placed higher workload demands on the observers than video observations. This disadvantage of live observations might be compensated by the significantly lower completion time for the live observations, which decreases study costs. Nevertheless it was also shown, that the observational workload was significantly higher for non-medical observers. However, we attribute these results to the fact that the engineering students had to cope with two new challenges, the observation support software and the medical background they needed to acquire, whereas the medical students had only the software as new challenge while the medical background was known to them beforehand.

The method for the generation of SPMs presented in this paper provides a well-founded basis for the observation of surgical behavior as we have shown with the help of a clinical example. The validation study was conducted based on an intervention in otorhinolaryngological surgery. Similar accuracy is expected if the results of this study were generalized to other types of surgical interventions, which should be experimentally substantiated.

Additionally, an adaptation of the surgical workflow editor software for non-surgical application fields that require structured observation of behavior is conceivable. Due to the configurability of the software and the possibility of adjusting the ontology, the implementation of this tool can be extended to include a wide range of possible future study fields from diverse areas, such as sociobiology and psychology. Sociobiology or, more specifically, ethology, deals with ethograms, catalogues of discrete behavior, which could be described with explicit reference to their purpose with the help of this tool. As for psychology, for instance, usages in the fields of behavioral psychology and educational research can be conceived of as well as industrial or organizational psychology. More specifically, Applied Behavioral Analysis, organizational learning, problem solving, the development of educational technology, and scientific management could benefit from the implementation of the ontology and the software presented here.

Acknowledgements

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the scope of the Unternehmen Region by grant numbers 03 ZIK 031 and 03 ZIK 032 and by funds of the European Regional Development Fund (ERDF) within the framework of measures supporting the technology sector.

3.3 Sensor-based data acquisition

Title

Online recognition of surgical instruments by information fusion

Authors

Thomas Neumuth[1], Christian Meißner[1]

Departments

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

Citation

Neumuth T, Meißner C. The impact of information fusion strategies for online recognition of activities for surgical process modeling. International Journal of Computer Assisted Radiology and Surgery. 2012; 7(2):297-304.

Keywords

Surgical Procedures, Operative; Workflow; Decision Making, Computer-Assisted; Information System ; Surgery, Computer-Assisted; Process Assessment (Health Care); Surgical Process Model

Abstract

Purpose: Automatic online recognition of surgical instruments is required to monitor instrument use for surgical process modeling. A system was developed and tested using available technologies.

Methods: A recognition system was developed using RFID technology to identify surgical activities. Information fusion for online recognition of surgical process models was conceived as a layer model to abstract information from specific sensor technologies. Redundant, complementary, and cooperative sensor signal fusion was used in the layer model to increase the surgical instrument recognition rate. Several different information fusion strategies were evaluated for situation recognition abilities in a mock-up environment based on simulations of surgical processes.

Results: This information fusion system was able to reliably detect, identify, and localize surgical instruments in an interventional suite. A combination of information fusion strategies was able to achieve a correct classification rate of 97% and was as effective as observer-based acquisition methods.

Conclusion: Different information fusion strategies for the recognition of surgical instruments were evaluated, showing that redundant, complementary, and cooperative information fusion is feasible for recognition of surgical work steps. A combination of sensor- and observer-based modeling strategies provides the most robust solution for surgical process models.

Introduction

In the last few years, information and communication technology (ICT) has become increasingly available in the operating room (OR). The increasing availability of information is expected to support better, faster, and less cost-intensive surgical interventions in the future OR [Cleary et al. 2005; Lemke and Vannier 2006]. ICT systems provide many different types of information. However, not every piece of information is useful to the surgeon in every situation. Some information is useful only during certain work steps. For instance, the visualization of spatial coordinates is only useful during navigation; it is unnecessary during the other parts of the surgical process. Therefore, an information selection process needs to be implemented.

The ICT system must include autonomous detection and recognition of surgical work steps and a priori knowledge about the context of the recognized surgical work step in the surgical process to enable situational awareness of OR technology for automatic provision, management, and presentation of information to the surgical team. The ICT system can be compared to a car navigation system. A priori knowledge is represented by the street map; autonomous detection of the car's location is performed by the Global Positioning System (GPS) receiver and registered to the map.

While approaches for generating the a priori knowledge, the "street maps" of surgical interventions, are available as generic surgical process models [Neumuth et al. 2011b; Blum et al. 2008b], few systems are available for automatic online recognition of surgical activities. Therefore, the objective of this study is the development of an automatic online recognition system to monitor activities during surgical processes.

The development of automatic situation recognition for surgical process models is a relatively new field of research. Information is gathered by sensor systems that make direct or indirect measurements of signals from surgical activities. This information is then used to construct the surgical process model. Direct methods for gathering information include recognizing surgical instruments or motions based on video recordings [Padoy et al. 2012; Sudra et al. 2009; Lalys et al. 2010; Bouarfa et al. 2010; Lin and Hager 2009], kinematic data recordings from telemanipulators [Varadarajan et al. 2009] or from virtual environments [Darzi and Mackay 2002], recognizing surgical actions based on force/torque signatures [Rosen et al. 2001], or using acceleration sensors [Ahmadi et al. 2010]. Indirect methods include interpreting the patients' vital parameters [Xiao et al. 2005], the motions of surgical instruments [Lin et al. 2006], the surgeons' eye movements [James et al. 2007], the locations of OR staff measured using ultrasound [Nara et al. 2010], or a general utilization of the operating room [Bhatia et al. 2007].

In recent years, a number of approaches to modeling surgical processes have been developed. In 2001, MacKenzie et al. presented a top-down model for the structuring of laparoscopic Nissen fundoplications [MacKenzie et al. 2001]. The top-down approach, however, only allows for a very general representation of these fundoplications. Jannin et al. [Jannin et al. 2003; Jannin and Morandi 2007] have ontologically examined the discrepancies between planned and implemented surgical processes. They used questionnaires to determine and describe the series of single surgical work steps in brain tumor surgery. In 2009, Neumuth et al. presented a generic method for observer-based modeling of patient-individual surgical processes

[Neumuth et al. 2009b] and for the computation of generic and statistically averaged surgical process models [Neumuth et al. 2011b]. Furthermore, the authors have shown that these models can be used to establish requirements for new medical engineering products [Neumuth et al. 2009c; Neumuth et al. 2011d]. Although these approaches form good foundations for the modeling, description, and analysis of surgical processes, they have not been designed to recognize work steps during surgical interventions.

While some approaches to recognizing surgical activities are currently available, they have some limitations: they strongly depend on specific sensor technologies, work on a low granularity level of surgical phases rather than on the higher level of surgical activities, or do not consider data from a variety of sources to employ information fusion strategies for multimodal situation recognition.

We present the design, implementation, and evaluation of an online situation recognition system that uses RFID data to identify surgical activities. Our design is based on multi-layer processing that provides information fusion from different sensors and enables the abstraction of the information from a specific sensor technology by using an ontological approach. In this context, the application of information fusion strategies to optimize the recognition abilities of the system is of particular interest. The system was implemented as a mock-up. The operability of the recognition system and its use in monitoring process operations was evaluated during simulated surgical interventions.

Information fusion for instrument recognition in surgical processes

Information fusion is defined as an information process that interrelates data and information from a variety of sources [Hall and LLinas 1997; Xiong and Svensson 2002; Llinas et al. 2004; Kokar et al. 2004]. Data are matched, correlated, and combined to create an abstract, but nevertheless appropriate and precise, likeness of the world. The classification scheme by Durrant-Whyte differentiates fusion strategies according to the type of sensors used: redundant, complementary, and cooperative information fusion [Durrant-Whyte 1988; Luo et al. 2002]. All three strategies were implemented in our information fusion system design (see Figure 3.3.1 and Table 3.3.I).

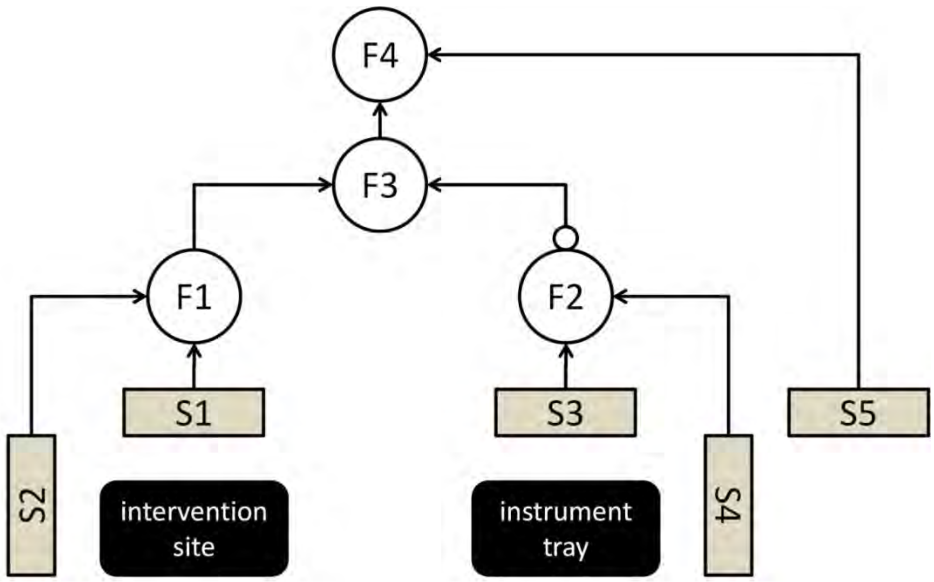


Figure 3.3.1: Information fusion scenarios in the operating room for automatic instrument recognition for surgical process models.

Antennas S1 and S2 were positioned at the interventional site, while antennas S3 and S4 were positioned at the instrument tray. While S1 and S3 were mounted horizontally, S2 and S4 were mounted vertically. All sensors registered the presence or absence of surgical instruments which were equipped with RFID tags. Additionally, a human observer worked as "sensor" S5 and recorded the simulated surgical process by observation as described in Neumuth et al. [Neumuth et al. 2009b]. The information fusion involved a series of hierarchical steps, which were implemented using four fusion sites F1 to F4.

We used a standard notebook computer to run the detection framework with software written in C# because manufacturers of most products we used delivered their drivers in C#. The RFID readers were Sirit-InfinityTM 510 UHF antennas, connected to our notebook by Ethernet. Four circular polarized patch antennas were placed in the operating room to detect the presence or absence of surgical instruments in different recognition zones in the OR. The average distance of the antennas to the intervention site and to the instrument tray was between one and two meters.

Table 3.3.I: Overview of symbols used for sensors and fusion sites.

	Symbol	Remarks
Sensors	S1	Sensor 1 - horizontal antenna at interventional site
	S2	Sensor 2 - vertical antenna at interventional site
	S3	Sensor 3 - horizontal antenna at instrument tray
	S4	Sensor 4 - vertical antenna at instrument tray
	S5	"Sensor" 5 - observer with tablet PC recording surgical actions, used instruments, and treated anatomical structure
Fusion sites	F1	Fusion site 1 for redundant fusion of site zone information, $\mathrm{F1} = \mathrm{S1} \vee \mathrm{S2}$
	F2	Fusion site 2 for redundant fusion of instrument tray zone information, $\mathrm{F2} = \mathrm{S3} \vee \mathrm{S4}$
	F3	Fusion site 3 for complementary information fusion of F1 and F2, $\mathrm{F3} = \mathrm{F1} \vee \overline{\mathrm{F2}}$
	F4	Fusion site 4 for cooperative information fusion of F3 and S5

The aim of *redundant information fusion* was to combine information from similar sensors to obtain a complete representation of the world. Sensors were considered redundant if every sensor detected the same parameters of the same object, independent of other sensors. This strategy was used to enhance the robustness and the margin of error of the overall system. In Figure 3.3.1, the fusion sites F1 and F2 performed a redundant fusion of information from sensors S1 and S2, with respect to S3 and S4.

Complementary information fusion was performed by combining different sensors with non-overlapping data about the presence of surgical instruments in the interventional site. The sensors were independent from one another and were used to obtain a more complete representation of the current situation. In Figure 3.3.1, the fusion site F3 represents an example of complementary information fusion that merges the results from the monitoring of the site and the instrument tray. To strengthen object recognition, the information gathered at sites F1 and F2 was treated as new sensor information and was amalgamated at F3, the complementary fusion site. F3 surveys the situs, where an object or instrument is rated as recognized if the following holds true: $F3 = F1 \vee \overline{F2}$.

Cooperative fusion is needed to derive information that could not be obtained with the use of a single sensor. This is achieved by using two independent sensors with information concerning different objects. This reduces general uncertainty and enhances the system's robustness by combining different information sources or different features of a single source. Cooperative fusion was performed at fusion site F4. This site combined the object recognition information from F3 with additional information generated by the observer working as "sensor" S5. S5 was used to record the surgical actions and the treated anatomical structure.

Layer model for sensor-technology abstraction

The information fusion system was implemented based on a layer model with multiple abstraction levels. The objectives of this layer model were to separate higher-layer information from the specific sensor technology and to allow for the application of multiple fusion strategies. The sensors were combined to form logical sensors and virtual zones and were analyzed with the help of axioms. We implemented our axioms by using the Web Ontology Language (OWL). The overall layer model consisted of a hardware-abstraction layer, a sensor-pooling layer, a virtual-zone layer, a roaming layer, a cooperative-fusion layer, and an application layer (see Table 3.3.II).

The first layer, the *hardware-abstraction layer*, provided an interface to the sensor hardware to access the RFID readers. This included the use of a common data format to represent actual events and sensor activities.

The dedicated sensors were combined into logical sensor groups in the *sensor pooling layer*. A confidence level for the presence of a tag at an antenna was computed for basic noise reduction. Tag IDs were translated to the names of the instruments to use the abstract notions of the objects.

The *virtual-zone layer* interpreted the data from the sensor pools and implemented redundant fusion at F1 and F2. Each object was given a binary condition to express whether it was within a zone or not within a zone, facilitating logical evaluation in the roaming layer (described below).

The presence of objects in the situs zone or the instrument tray zone was resolved in the *roaming layer*. OWL axioms were applied in this layer to perform cooperative fusion at F3. Subsequently, the presence of an instrument in the interventional site was passed on to the *cooperative-fusion layer* for combination with information gathered by the observer (S5) at F4.

Finally, the *application layer* provided the resulting surgical process model to other applications.

Table 3.3.II: Layer model for sensor-technology abstraction-independent implementation of information fusion strategies including implementation examples.

Layer	Objective
Application layer	Providing recognized information to other applications
Cooperative-fusion layer	Performance of cooperative fusions of instrument information with data of the observer to generate the surgical process model
Roaming layer	Determination of object location based on OWL-axioms
Virtual-zone layer	Implementation of sensor pools to virtual zones that represent logical spaces
Sensor-pooling layer	Logical sensor-pooling and object identification
Hardware-abstraction layer	Provision of generic hardware interface to sensor systems

Evaluation study for online recognition of surgical instruments

Study design

The objective of the study was to evaluate if and how each individual fusion strategy contributes to the automatic recognition of activities in surgical process models. Paranasal surgical interventions were simulated by actors as surgical processes based on predefined scripts. The scripts contained work step sequences that were based on observation protocols of real surgical interventions. Each script contained a number of surgical activities that were performed by two actors, one acting as the surgeon and one as the assistant. The actors performed typical surgical work steps for this intervention type, such as dissection of polyps in different nasal cavities with Blakesleys, suctioning with suction tube, insertion of nose speculum, or disinfection of the patient with swabs. An example portion of the simulation protocol and the surgical instruments that were equipped with RFID tags is shown in Figure 3.3.2.

The scripts were defined based on observations of real paranasal interventions and verified by experienced Ear-Nose-Throat (ENT) surgeons during study preparation. Three script variants were defined: a typical procedure with main activities by the surgeon, a procedure course with faster working speed, and a procedure course with a temporal switch of roles between surgeon and assistant. Each script variation had an average performance duration of approximately 20 minutes and contained between 60 and 90 surgical work steps. Each script was repeated three times. Thus, a total of nine measurement series were recorded, and approximately 650 activity measurements were taken.

The patient was simulated using a rapid-prototyping model. The nasal cavities were printed as 3D-models based on CT scans of a real patient. Using this approach, a non-varying model of a typical patient could be used for the study. An endoscopic view from the rapid-prototyping model is shown in Figure 3.3.2.

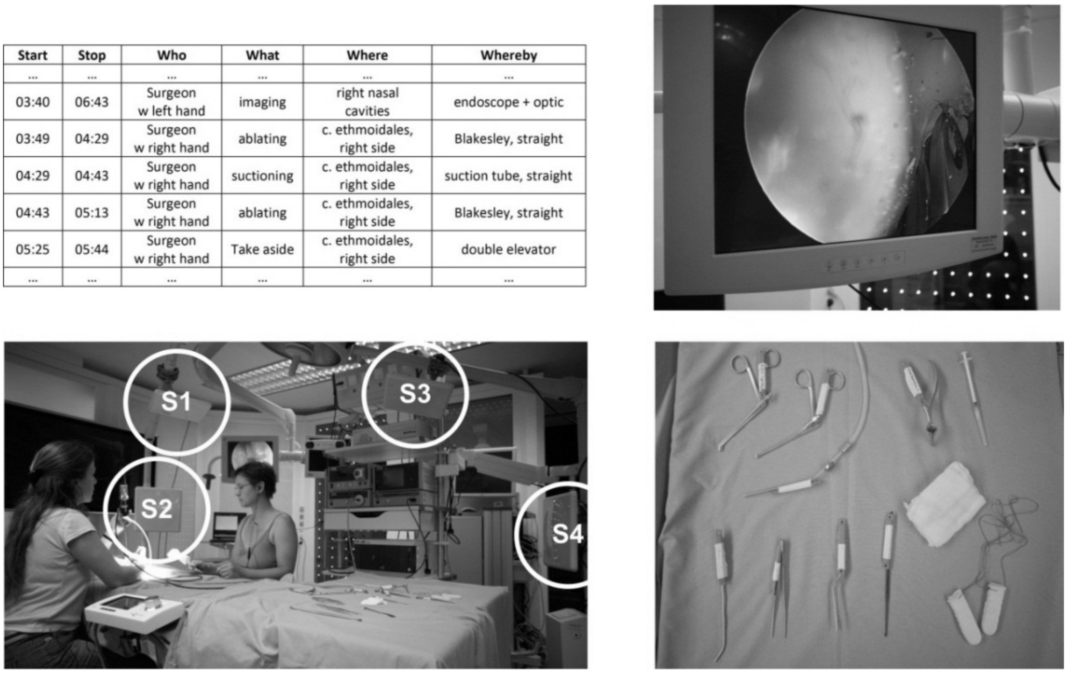
The actors were trained beforehand to render the scene and received detailed work step information from the scripts via audio input. The reenactment took place in the demonstration OR of the Innovation Center for Computer Assisted Surgery (ICCAS) at the Medical Faculty of the Universität Leipzig.

The data was obtained in three steps: First, the S1-S4 data was recorded using the delineated RFID system. Simultaneously, a surgical process model was prepared by means of conventional intra-operative observation by the experienced observer S5 as described in [Neumuth et al. 2009b]. Finally, the respective position of the instruments, the ‘localization reference’, was defined by analyzing video records of the scene.

Measurements of sensitivity, specificity, and correct classification of the RFID system were calculated using time steps with duration of 1/20 s. A comparison of the measurement results from sensors S1 to S4 with the results of the localization protocol showed that the measurement system generated reliable results and could be used as accurate input for the lowest fusion step. The results from fusion sites F1 and F2 were further used to determine the complementary fusion F3: $\mathrm{F3} = \mathrm{F1} \vee \overline{\mathrm{F2}}$.

To determine the results of cooperative fusion at F4, the results from F3 and the observer S5 were compared to the scripts. F3, which contained the instrument information, was compared to the instrument information in the scripts. S5, which gathered observed information about surgical actions, instruments, and treated anatomical structures, was compared to the full information set in the scripts. To compare F4 to the scripts, the instrument information from S5 was deleted from the model and replaced by that of F3. The resulting model was then compared to the scripts, meaning that the online recognition performed by the RFID system and additional information gathered by the human observer were fused to generate the surgical process model.

The post-processing and rehashing of the measurement results were done using MATLAB 2009a [MathWorks 2009]. The statistical analyses of the recognition capacities of the different instruments were performed using SPSS 15 [SPSS Inc. 2008]. The Friedman test was run on the three dependent samples, the Wilcoxon test on the two dependent random tests, and the Mann-Whitney U test on the independent samples, each with a significance level of $\alpha = 0.05$.



Start	Stop	Who	What	Where	Whereby
...	...	...	...	...	...
03:40	06:43	Surgeon w left hand	imaging	right nasal cavities	endoscope + optic
03:49	04:29	Surgeon w right hand	ablating	c. ethmoidales, right side	Blakesley, straight
04:29	04:43	Surgeon w right hand	suctioning	c. ethmoidales, right side	suction tube, straight
04:43	05:13	Surgeon w right hand	ablating	c. ethmoidales, right side	Blakesley, straight
05:25	05:44	Surgeon w right hand	Take aside	c. ethmoidales, right side	double elevator
...	...	...	...	...	...

Figure 3.3.2: Surgical script portion (top left); endoscopic view inside the rapid-prototyping model used as patient simulator (top right); general setting with mounted RFID antennas (bottom left) and surgical instruments equipped with RFID tags (bottom right).

Study results

An overview of the results of the information fusion strategy evaluation is presented in Table 3.3.III for redundant, complementary, and cooperative-information fusion. The redundant fusion site F1 exhibited an average sensitivity of 0.95 ± 0.13, a specificity of 0.86 ± 0.23, and a correct classification rate of 0.92 ± 0.11. Generally, the redundant fusion site F1 performed significantly better than the individual sensors without fusion. For instance, the combination of vertical and horizontal sensors produced highly significant enhancements ($p<0.001$) to sensitivity and specificity when compared to using only a single sensor. Even though the correct classification rates of the two sensors were not significantly different from each other, the combination of the sensors showed a slightly decreased correct classification rate. Generally, the recognition performance of the horizontal sensor determined the overall combined performance, while the vertical sensor changed the overall combined performance by only a small percentage.

The redundant fusion site F2 showed an average sensitivity of 0.93 ± 0.14, a specificity of 0.71 ± 0.16, and a correct classification rate of 0.92 ± 0.10. Again, this redundant fusion showed significantly better results than the individual sensors. For instance, redundant fusion of vertical and horizontal sensors showed highly significant ($p<0.001$) sensitivity and specificity improvements. The recognition rates of the horizontal and vertical sensors were comparable, though the horizontal sensor again had slightly better results.

The study results for the complementary fusion site F3 were significant for every combination. Sensitivity was significantly improved ($p<0.001$) to 0.96 ± 0.12 by complementary fusion, while specificity was significantly ($p<0.001$) reduced to 0.81 ± 0.27. Specificity declined because the lower specificity of F2 influenced the result strongly.

Cooperative fusion at F4 did not perform significantly differently in sensitivity, specificity, and correct classification than the observer S5. A total correct classification of 0.97 ± 0.02 was achieved for cooperative fusion.

Table 3.3.III: Results of the evaluation study for redundant, complementary, and cooperative information fusion.

	[mean ± sd]	S1	S2	F1	p	S1 vs. S2	S1 vs. F1	S2 vs. F1
Redundant Fusion	Sensitivity	0.93 ± 0.13	0.77 ± 0.24	0.95 ± 0.13	p<0.001	p<0.001	p<0.001	p<0.001
	Specificity	0.92 ± 0.12	0.88 ± 0.22	0.86 ± 0.23	p<0.001	p>0.05	p<0.001	p<0.001
	Correct classification	0.93 ± 0.09	0.91 ± 0.09	0.92 ± 0.11	p<0.001	p>0.05	p>0.05	p<0.001
	[mean ± sd]	S3	S4	F2	p	S3 vs. S4	S3 vs. F2	S4 vs. F2
	Sensitivity	0.87 ± 0.21	0.56 ± 0.45	0.93 ± 0.14	p<0.001	p<0.001	p<0.001	p<0.001
	Specificity	0.74 ± 0.15	0.49 ± 0.37	0.71 ± 0.16	p<0.001	p<0.001	p<0.001	p>0.05
	Correct classification	0.88 ± 0.16	0.59 ± 0.44	0.92 ± 0.10	p<0.001	p<0.001	p>0.05	p<0.001
Complementary fusion	[mean ± sd]	$F1$	$\overline{F2}$	F3	p	$F1$ vs. $\overline{F2}$	$F1$ vs. $F3$	$\overline{F2}$ vs. $F3$
	Sensitivity	0.95 ± 0.13	0.51 ± 0.25	0.96 ± 0.12	p<0.001	p<0.001	p<0.001	p<0.001
	Specificity	0.86 ± 0.23	0.91 ± 0.25	0.81 ± 0.27	p<0.001	p<0.001	p<0.001	p<0.001
	Correct classification	0.92 ± 0.11	0.91 ± 0.1	0.91 ± 0.12	p<0.001	p>0.05	p<0.001	p<0.001
Cooperative fusion	[mean ± sd]	F3	S5	F4	p	F3 vs. S5	F3 vs. F4	S5 vs. F4
	Sensitivity	0.91 ± 0.24	0.80 ± 0.28	0.80 ± 0.28	p<0.001	p<0.001	p<0.001	p>0.05
	Specificity	0.74 ± 0.29	0.98 ± 0.03	0.98 ± 0.03	p<0.001	p<0.001	p<0.001	p>0.05
	Correct classification	0.85 ± 0.13	0.97 ± 0.03	0.97 ± 0.02	p<0.001	p=0.004	p<0.001	p>0.05

Discussion

Existing approaches to monitoring surgical processes have some limitations. 'Knowledge-based applications are based on manual observations, which are considered to be universal with regard to the type of surgical intervention. However, the data are too imprecise for technical applications, and the methods are typically too costly for broad application. 'Sensor-based' technical approaches are too specific, are suitable only for very special cases, or were presented for the detection of surgical phases and not for a more fine-grained detection of surgical work steps.

Our approach combines the advantages of both knowledge- and sensor-based applications. Our concept can be used for multiple surgery types because the RFID tags can easily be mounted on every surgical instrument; it also provides high temporal resolution. Additionally, we proposed the concept of information fusion for online recognition of surgical process models and provided a layer model to abstract information from specific sensor-technologies.

This study has shown that the system can reliably detect, identify, and localize the presence of instruments in the interventional site. The system was able to significantly increase sensitivity by implementing redundant and complementary fusion. Additionally, the rate of correct classification of surgical activities was increased to 97% by sequential application of information fusion strategies.

However, this study has also shown that direct recognition of objects leads to distinctly better results than indirect recognition. The overall sensitivity of the cooperative fusion was predictably subpar because the analysis was based on time steps, and the observer had a slight delay in recording the surgical activities.

The design of the presented system was based on a layered approach, which allowed the integration of various sensors in the OR. We have shown how the layered approach was related to an overall information concept. By using this concept, we expect that sensor-technology-specific properties, such as sampling rates and sampling features, could be decoupled, and virtual sensor clusters could be formed.

The RFID-based mock-up implementation of the system was assessed in an evaluation study. The study was based on simulations of real surgical processes on rapid prototyping models. The simulation scripts were clinically verified surgical process models. Generally, the recognition results were sensitive to simulation errors by the actors. However, previous studies have shown that actor-based simulations of processes are a reliable approach for reenacting surgical interventions [Neumuth et al. 2009b]. In [Neumuth et al. 2009b] it was also shown that actor-based simulations are performed with 97% accuracy to the predefined simulation scripts, and actors introduce very few errors. Observer errors were also represented in the measured process model. The accuracy of data acquisition by observers was also validated previously [Neumuth et al. 2009b].

The use of the RFID approach introduced some challenges. An RFID tag was likely to be undetectable by one of the antennas because the field was shielded by the physiological properties of the actor's hand. In many cases the application of redundant fusion in our system overcomes this problem. As the results show, it was less likely for a tag to be undetectable by both sensors because the sensors were mounted at different orientations. Generally, the 'visibility' problem of the exact position of RFID tags is less difficult than in other sensor technologies such as infrared sensors used in navigated surgery because the objective is the detection of a tag in a volume and not the detection of its exact position.

Because most surgical instruments were made of metal, the metal properties also influenced the performance of the tags. Additionally, the specific shapes and structures of the surgical instruments might have influenced the radio characteristics. Because we used standard tags mounted on standard instruments, this was an expected consequence. In our study the major detection errors came from instruments with solid metal structures, such as Blakesleys, which influenced the radiation characteristics of their tags. In contrast, the double elevator with a low complex metal structure permitted good recognition rates. We also neglected soiling problems that could influence the recognition rate because the patient phantom that we used in our study did not contain liquids.

Because the RFID approach is merely a subsystem of the overall object-recognition scheme, the results can be compared to available generic acquisition methods, such as observer-based acquisition, but only to a limited extent. Although manual and automatic recordings were performed on two different levels of abstraction, the two systems are comparable because they have the same objectives: accurate recognition of the situation and modeling of the surgical process.

Future work needs to be performed to assess the application of the system in the course of real surgical interventions. Furthermore, a thorough examination of the influence of the OR equipment and personnel on electromagnetic compatibility in sensitive environments is necessary, as shown in other studies [van der Togt et al. 2008]. These challenges could be overcome by reducing radiation power or using an alternative portion of the electromagnetic spectrum less affected by the specific environment.

Situations such as the performance of several surgical activities using the same instrument cannot be detected by the RFID system. Therefore, other sensor technologies may be integrated, such as wireless LAN-based localizations, accelerometers, or time-of-flight sensors. Furthermore, automatic recognition needs to be extended to other surgical actions and treated anatomical structures.

Situation recognition applies to many use cases. Combining situation recognition with a priori knowledge from generic surgical process models [Neumuth et al. 2011b] enables applications for surgical management and procedure guidance. These applications include the timely and automatic intraoperative presentation of preoperatively acquired information from radiological images or histological examination results by using augmentation in monitors or endoscopic/microscopic views for adaptive user interfaces. Situation recognition can also be used for automatic and timely parameterization of surgical assist systems and triggering of intraoperative measurements of key indicators for quality management. Further applications in the context of surgical quality management include documentation of the surgical process for legal purposes as well as guidance systems for learning surgeons, e.g., to indicate necessary and uncompleted surgical work steps.

Conclusion

The localization of objects in the operating room is an essential source of data for automatic recognition of surgical process models. In our study, we evaluated different information fusion strategies for the recognition of instrument recognition and showed that redundant, complementary, and cooperative information fusions contribute to the recognition of surgical work steps. With the presented RFID-based information-fusion approach, we were able to identify instrument usage during surgical processes and recognize surgical activities with a correct classification rate of up to 97% in a simulation environment.

Acknowledgements

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the scope of the Unternehmen Region with grant numbers 03 ZIK 031 and 03 ZIK 032 and by funds of the European Regional Development Fund (ERDF) and the state of Saxony within the frame of measures to support the technology sector.

4 Model generalization and surgical workflow management

After completion of the data acquisition, technically processible surgical process models are available. These models describe the intervention courses on single patients as (patient-) individual surgical process models (iSPMs). To be able to make an assertion concerning a set of iSPMs, it is an option to generate a new model that represents the mean of all considered iSPMs, a generic surgical process model (gSPM).

gSPMs are of interest for nearly every group of users in the field of surgery and surgical research. The computation of such generic surgical process models paves the way for a host of technical, as well as clinical, application scenarios, as will be presented in chapter 5.

The first publication

> *Neumuth T, Jannin P, Schlomberg J, Meixensberger J, Wiedemann P, Burgert O. Analysis of surgical intervention populations using generic surgical process models. International Journal of Computer Assisted Radiology and Surgery. 2011; 6(1):59-71*

describes how gSPMs are generated from iSPMs. This step is necessary for the computation of a model having at least similar expressional power than models gained from top-down modeling. Based on a number of iSPMs, the computation of the gSPM as statistically averaged intervention procedure is being proposed. In amending the weaknesses of top-down models, the gSPMs possess features such as quantifications of performance times and branch probabilities. For the calculation, a simple method is used and the applicability of this method is evaluated by comparing two gSPMs of different surgical treatment strategies from ophthalmology.

The developed research method to compute generic surgical process models as statistical 'mean' procedure courses used the computation of a gSPM for each, an ambulatory and an inpatient patient sample. The assessment of the gSPMs and an ensuing statistical analysis was then employed to identify differences in treatment strategies. For the most part, the clinical evaluation by review showed that the resulting gSPMs correspond to the clinical guidelines.

The second publication,

> *Neumuth T, Liebmann P, Wiedemann P, Meixensberger J. Surgical workflow management schemata for cataract procedures: Process model-based design and validation of workflow schemata. Methods of Information in Medicine. 2012; 51(5):371-382.*

delineates the processing of generic surgical process models for their application in surgical workflow management in the operation room. To be able to apply the generated models as basis for the behavior of a workflow management system (WFMS) during an intervention, gSPMs can be converted into workflow schemata. This method is shown in the presented publication and evaluated with the help of a surgical workflow management system that is being designed. In addition, the number of iSPMs needed to create a robust working workflow scheme is being investigated.

The results of this study were the design of a surgical workflow management system and the development of a method for the automatic computation of workflow schemata from gSPMs from ophthalmology to guide highly variable surgical processes. With the help of a computed workflow schemata, this was then validated. In addition, it has been investigated, how many iSPMs are needed to generate a workflow schema to guide a randomly selected iSPM from a complementary set, resulting in 10 iSPMs to guide 70% of all surgical processes and 50 iSPMs to guide 90% of all surgical processes.

4.1 Computation of generic surgical process models

Title

Analysis of surgical intervention populations using generic surgical process models

Authors

Thomas Neumuth[1], Pierre Jannin[2,3,4], Juliane Schlomberg[5], Jürgen Meixensberger[1,6], Peter Wiedemann[5], Oliver Burgert[1]

Departments

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] INSERM, U746, Faculty of Medicine, Rennes, France

[3] INRIA, VisAGeS Unit/Project, Rennes, France

[4] University of Rennes I, CNRS, UMR 6074, IRISA, Rennes, France

[5] University Medical Center Leipzig, Department of Ophthalmology, Leipzig, Germany

[6] University Medical Center Leipzig, Department of Neurosurgery, Leipzig, Germany

Citation

Neumuth T, Jannin P, Schlomberg J, Meixensberger J, Wiedemann P, Burgert O. Analysis of surgical intervention populations using generic surgical process models. International Journal of Computer Assisted Radiology and Surgery. 2011; 6(1):59-71.

Keywords

Surgical Workflow; Surgical Process Model; Health Care Evaluation Mechanisms; Cataract surgery

Abstract

Purpose: According to differences in patient characteristics, surgical performance, or used surgical technological resources, surgical interventions have high variability. No methods for the generation and comparison of statistical 'mean' surgical procedures are available. The convenience of these models is to provide increased evidence for clinical, technical, and administrative decision making.

Methods: Based on several measurements of patient individual surgical treatments, we present a method of how to calculate a statistical 'mean' intervention model, called generic Surgical Process Model (gSPM), from a number of interventions. In a proof-of-concept study we show how statistical 'mean' procedure courses can be computed and how differences between several of these models can be quantified. Patient individual surgical treatments of 102 cataract interventions from eye surgery were allocated to an ambulatory or inpatient sample and the gSPMs for each of the samples were computed. Both treatment strategies are exemplary compared for the interventional phase Capsulorhexis.

Results: Statistical differences between the gSPMs of ambulatory and inpatient procedures of performance times for surgical activities and activity sequences were identified. Furthermore, the work flow that corresponds to the general recommended clinical treatment was recovered out of the individual surgical process models.

Conclusion: The computation of gSPMs is a new approach in medical engineering and medical informatics. It supports increased evidence, e.g. for the application of alternative surgical strategies, investments for surgical technology, optimization protocols, or surgical education. Furthermore, this may be applicable in more technical research fields, as well, such as the development of surgical workflow management systems for the operating room of the future.

Introduction

Surgical process models (SPMs) are models of surgical interventions. By modeling surgical processes, we have drawn attention in previous works [Neumuth et al. 2009b] to the fact that there exists no explicit methodology that can be used to objectively model surgical strategies at a detailed level. An availability of such models makes the knowledge about surgical processes, which was previously inaccessible, explicit. This is an essential step to facilitate e.g. quality management of surgeries, evaluation studies, or requirements studies, and may encourage discussions among clinicians and technicians.

The existence of SPMs results in a new layer of interest: since the previous generation of SPMs (introduced below) was able to represent only a single individual surgical intervention course, what new or additional possibilities would a generic SPM provide? A more comprehensive model could include and combine multiple individual courses into a statistically 'mean' model that exhibits a more generic character. Such a generic model could be valuable for the quantification, statistical assessment, and visualization of surgical knowledge and techniques with the purpose of quality management in health care.

Specific application cases could include a comparison of two generic Surgical Process Models to elucidate differences in surgical strategies or to clarify the use of certain instruments or devices. The approach may be useful to assess skill levels, or it could serve as the basis of a detailed extrapolation of intervention costs. Further applications, e.g. the comparison of patient individual surgical process models (iSPMs) with generic surgical process models (gSPMs), may include an investigation of the reasons why a single surgical intervention course may have deviated from the mean procedure course.

Currently, very few approaches have been proposed to evaluate individual patient or generic models of surgical processes. Recently, the use of SPMs in Medical Engineering and Medical Informatics has been discussed by several authors.

Jannin et al. [Jannin et al. 2003; Jannin and Morandi 2007] introduced a method for acquiring patient-individual SPMs using an ontological approach. They applied data mining-based methods to a database of 159 iSPMs, describing surgical procedures on the brain in order to predict certain features of these procedures (called *predicted variables*) from characteristics of the patient and the associated pathology (called *predictive variables*). They used the same methods to classify the data into main families based on the predictive variables and they manually allocated the values of the predicted variables to each family. However, even though computing gSPMs was one objective of their work, they failed to compute such models.

Other authors have modeled surgical processes in the context of medical engineering for several purposes, such as the automatic identification of interventional phases [Ahmadi et al. 2006; James et al. 2007], control of surgical robots [Münchenberg et al. 2001a], and instrument assessments [Mehta et al. 2002]. Clinical work has also focused on surgical processes for reengineering [Casaletto and Rajaratnam 2004], assessing human reliability [Malik et al. 2003], comparing substitutive surgical strategies [den Boer et al. 1999], and analyzing requirements for surgical assist systems [Strauß et al. 2006a].

However, all of these approaches either do not deal with the generation of a generic SPM or provide information only at the level of interventional phases rather than at the level of surgical work steps.

Some authors have presented approaches for computing gSPMs [MacKenzie et al. 2001; Meng et al. 2005; Blum et al. 2008b]. These methods, however, do not consider variations of several relevant procedures [MacKenzie et al. 2001]. They were applied only at the conceptual level of intervention modeling without quantification of measurement parameters [Meng et al. 2005], or they featured a low level of granularity with poor expressiveness [Blum et al. 2008b; Westbrook and Ampt 2009].

The notion of *Workflow Mining* in business informatics is closely related to our presented approach. In 1995, Cook and Wolf [Cook and Wolf 1995] published the first algorithms to determine process models from software event logs. The preparatory work, namely the use of process mining to explore business process models, was initiated by Agrawal et al. [Agrawal et al. 1998]. The process mining community has been actively working over the past five years to formalize the discovery of process models based on event logs, e.g. [Schimm 2004; de Medeiros et al. 2005a]. For a survey of this area, see van der Aalst et al. [van der Aalst et al. 2003]. Methods described herein are not applicable to the computation of gSPMs and comparisons between intervention samples, because they do not include multiple perspectives or concurrencies, such as parallel left- and right-handed surgical work steps.

Furthermore, existing sources of information related to surgical procedures, such as clinical guidelines [AHRQ-Agency for Health Care Research and Quality 2010a; AWMF-Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. 2010a] or surgical textbooks, describe surgeries at a very general level; their goals are not to describe interventions in detail, but rather to give treatment recommendations. However, this general level cannot be used for quantification since it usually consists of free text and does not rely on a formal numerical structure, whereas quantification is necessary to perform detailed needs assessment or evaluation studies in order to derive technical requirements for surgical assist systems. With the methods utilized in the present work, it becomes possible to base such measurement parameters, such as most probable intervention courses, on real clinical data.

In this paper, we introduce methods for computing generic surgical process models (gSPMs). It is shown that it is feasible to use gSPMs to quantify differences in surgical workflows of two intervention samples retrospectively. As a proof-of-concept study we use clinical data from 102 cataract interventions that were divided into two samples according to the application of different treatment strategies. gSPMs were then calculated as 'mean' treatments for each of the samples and the results were subsequently compared across the entire data set.

The research questions addressed in this article include: '*How can generic surgical process models be generated from a population of individual surgical process models?*' and '*How can two gSPMs be utilized to compare two different intervention samples?*'

Methods

This section introduces methods for generating generic surgical process models (gSPMs) from a population of individual surgical process models (iSPMs). Pertinent terms will be introduced, an overview of the model development process will be given, and the example application will be presented.

To compute a generic SPM, several stages must be processed (cf. Figure 4.1.1). Mandatory stages include: data acquisition for iSPMs, Inter-iSPM registration, and computation of the gSPM. Optionally, additional stages involving feature selection, segmentation, and filtering can be employed to decrease the visual complexity of the resulting models.

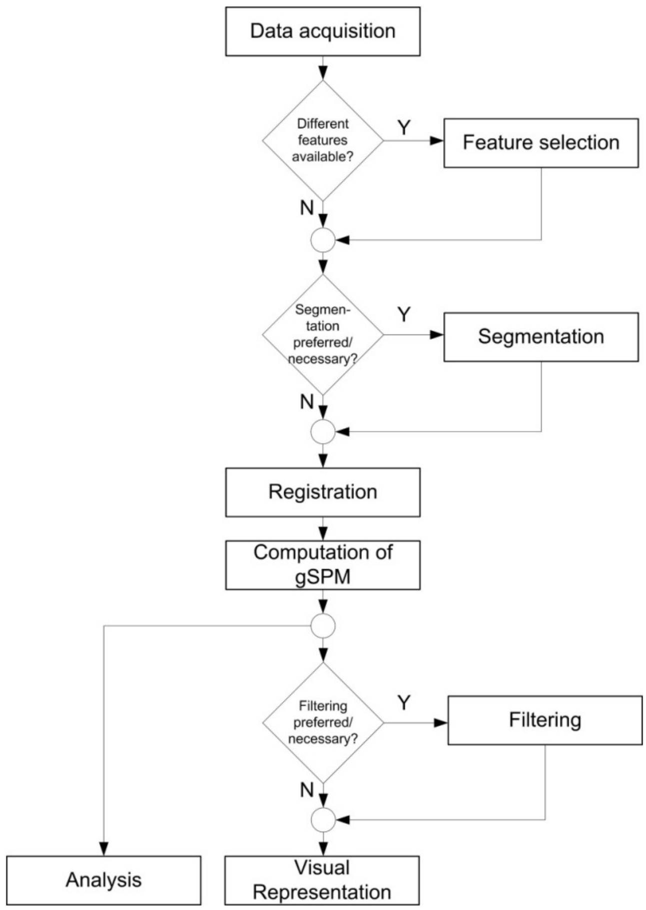


Figure 4.1.1: Overview of the model development process for gSPMs.

Terms

Essential in this context is the definitions of terms and concepts related to this approach (cf. [Neumuth et al. 2009b]): the surgical treatment performed on one specific patient is denoted a s*urgical process* (SP), and a model of the surgical process, e.g. in an information system, is called a s*urgical process model* (SPM). SPMs appear in two forms: *patient individual SPMs* and *generic SPMs*. The term *patient individual SPM* (iSPM) is used to refer to an SPM of a surgical process that was performed on one patient and thus represents one surgical case. The term *generic SPM* (gSPM) is used to represent the 'mean' surgical treatment of a theoretical patient. Generic SPMs are computed from different samples of iSPMs.

Data acquisition for iSPMs

Data acquisition deals with the mapping of the surgical procedure from a surgical process (SP) to a surgical process model (SPM). To store and process iSPMs, an appropriate data model is required. This data model describes, how entities of the surgical process are structured and presented within a given information system.

In this study, surgical work steps during the SP are represented as *activities*. Each iSPM consists of a number of activities that corresponds to the surgical work steps performed on the patient. Each activity is comprised of information about the work steps, termed *perspectives* (see Table 4.1.I). Examples of perspectives include: actions performed (e.g. *suctioning*, *cutting*); the surgical tool used (e. g. *scalpel*, *hook*); anatomical regions treated during the current work step; and start/stop times. An activity, therefore, describes *who* is doing *what*, with what *instrument*, *where*, and *when* during the surgical intervention. Activity examples are shown in Table 4.1.VI.

States symbolize status information and define the context in which activities were performed. Examples of states might be the different intervention phases of a procedure. A system of states acquired concurrently to activities implicitly relates activities to the interventional phases. An example that associates activities A, B, and C with intervention phase #1 is shown in Figure 4.1.2.

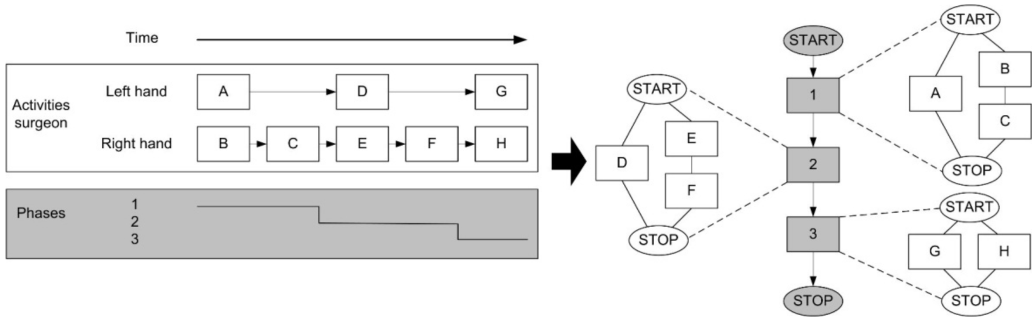


Figure 4.1.2: Segmentation of the iSPM according to the relations of activities and states.

Before gathering iSPM data, we had to define our terminology, especially for interventional phases or work steps. The former is crucial to segment the intervention into parts and thereby reduce the complexity of the resulting gSPM. The latter ensures a consistent naming of information entities across all relevant surgical cases. Table 4.1.IV and Table 4.1.V show examples of the interventional phases, surgical instruments, actions, and treated structures as used for the clinical case example in the next section.

During the live observation sessions iSPMs were recorded by trained medical observers, who were physically present in the operating room and recorded the performed surgical work steps of the intervention in the iSPM protocol. Data acquisition relied on a specially developed observation support software package, the surgical workflow editor (cf. Figure 4.1.3, [Neumuth et al. 2006a]). The software, running on a conventional tablet PC, presented terminology lists to the observer and asked for a description of the current surgical work step. Temporal information was added automatically. After each observation, the observer saved the protocol in extensible markup language (XML) format. The protocols that represented the iSPMs were then transferred to a database where further calculations were performed.

Table 4.1.I: Summary of concepts and subconcepts for SPMs.

<table>
<tr><th>Concept</th><th>Concept objective</th><th>Subconcepts / perspectives</th></tr>
<tr><td rowspan="5">activity</td><td rowspan="5">to represent surgical work steps in an SPM</td><td>what is done</td></tr>
<tr><td>who is doing it</td></tr>
<tr><td>wherby is it done</td></tr>
<tr><td>where is it done</td></tr>
<tr><td>when is it done</td></tr>
<tr><td rowspan="2">states</td><td rowspan="2">to represent surgical phases in an SPM</td><td>what is done, e.g. name of phase</td></tr>
<tr><td>when is it done</td></tr>
</table>

Feature selection

Data structures in iSPMs are comprised of various perspectives [Neumuth et al. 2009b]: organization, function, operation, and space. Each of these perspectives can be used to generate a gSPM with a different focus. The choice of perspective is termed *feature selection*. As features, perspectives can be chosen either exclusively or concurrently. An exclusive perspective choice results in a gSPM that is dedicated to the perspective in question, e.g. performed surgical actions, while a combination of perspectives results in a gSPM that has relevance for all chosen perspectives, e.g. the combination of actions performed and surgical instruments used. The more features that are included in building a gSPM, the more complex the resulting gSPM will be.

Segmentation

Splitting the iSPM into interventional phases is referred to as segmentation. The segmentation step was performed automatically according to the allocation of the activities to interventional phases as derived from the clinical guidelines and the time stamps of the activities. Consequently, all activities allocated to one interventional phase were selected across all iSPMs within a sample.

Inter-iSPM registration

The objective of the registration step was to associate reference points between iSPMs. In preparation of the generation of the gSPM, the iSPMs of the selected sample were registered to each other automatically, based on selected features from subsequent activities (cf. Figure 4.1.4). This registration step was performed for each interventional phase. Sequential activities represented transitions, expressed as predecessor-successor relationships. To include defined start and end nodes, artificial *START* and *END* features were added to each iSPM. *START* and *END* were included before the first predecessor and after the last iSPM successor respectively.

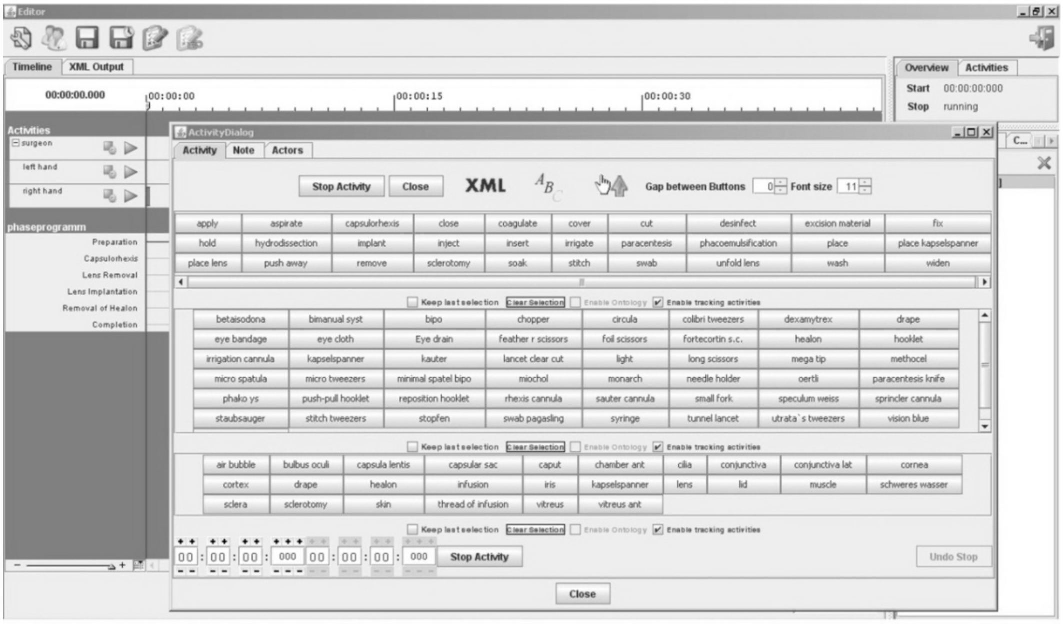


Figure 4.1.3: Screenshot of the surgical workflow editor user interface.

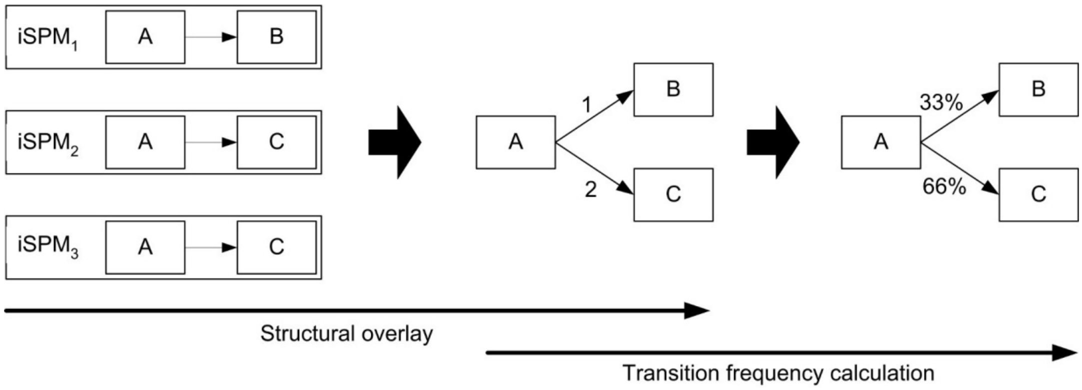


Figure 4.1.4: Simplified gSPM generation procedure example.

Computation of gSPMs

Computation of the gSPM structure

The transitions identified in the registration step were of relevance to the structural representation of the gSPM. For each interventional phase, all acquired transitions were registered based on the literals and were merged into one transition based on equal predecessor and successor activities. The result of this merging step was the gSPM structure.

Computation of the gSPM

The gSPM was subsequently annotated with global transition probabilities. The calculation was performed for all outgoing activity transitions in the iSPMs. The basis for the calculation was the number of sequential activities that each had the same predecessor. Subsequently, the local transition probability was calculated by normalizing the means of the global transition probabilities. The results quantify the transitions in the structural gSPM in terms of percentages (cf. Figure 4.1.4). Table 4.1.II shows how to compute the gSPMs.

Filtering for visual representation of gSPM

The resulting statistical gSPM can result in complex models that are not amenable to visual representation. For this reason, the optional step of filtering was included for the example data sets to improve visual accessibility. The filtering consisted of masking all transitions whose values were lower than a threshold defined by the user. Filtering did not affect the gSPM, but rather had the aim to increase the clarity of the visual representation.

Table 4.1.II: Computational algorithm for gSPM.

1. Calculation of activity values
    a. Grouping of all activities according to iSPM sample, interventional phase, and selected perspectives
    b. For the duration of each combination, the mean and the standard deviation is calculated
2. Calculation of activity transitions
    a. Grouping of all activities according to protocol-ID, interventional phases, and selected perspectives
    b. Computing the global transition probability for each selected transition in each protocol from the predecessor transition
    c. Calculating mean and standard deviations across all protocols of one iSPM sample
3. Joining activity durations with global transition probabilities, calculation of local transition probabilities, and creating the visualization

Example study

Objective of the example study

Cataract surgeries were chosen as a clinical example for the application of our methods. Based on clinical necessity, two treatment strategies are available for treating patients suffering from cataracts: ambulatory or inpatient treatments.

The objective of the example study was the retrospective assessment of the ambulatory and inpatient treatment strategies – our goal was to investigate differences in the gSPMs of both approaches. In addition to the assessment of more 'trivial' measures, such as total intervention times or durations of surgical phases, the example study showed how two gSPMs can be utilized to compare intervention samples.

iSPM-Samples

All cataract interventions were performed between March and September 2006 at the Eye Clinic of the University Hospital in Leipzig (Germany). The assignment of the patients to their respective treatment strategy was performed according to clinical necessity and expected complications.

The ambulatory, as well as the inpatient interventions, were conducted by three different, experienced surgeons: one surgeon performed inpatient treatments and two performed the ambulatory treatments.

Only patients with a cataract diagnosis were included in the study. The beginning of the first paracentesis and the end of the Healon® removal were chosen as unique criteria for defining the start and end of the interventional record (cf. Table 4.1.IV).

In total, 102 iSPMs of cataract surgery treatments were analyzed, 49 of which were performed as ambulatory and 53 as inpatient surgeries. The patient characteristics are presented in Table 4.1.III.

Cut-suture times were recorded from the Hospital Information System. One trained medical student was present in the operating room during the surgical procedures and acquired the data for the iSPMs through live observation with the aid of the surgical workflow editor [Neumuth et al. 2006a]. The validation of the accuracy of iSPM data acquisition has been published in a previous in-depth study [Neumuth et al. 2009b]. In the latter publication, observers were shown to acquire iSPMs accurately, robustly, and repeatable in both live and video observations, with a content accuracy of 92% and a temporal accuracy of <2 s. Examples of the terminologies used for the interventional phases and for describing perspective content are shown in Table 4.1.IV to Table 4.1.VI.

For all statistical analyses, Student's t-test with a significance level of $\alpha=0.05$ was used to calculate the p-values. Segmentation, registration, and gSPM calculation were performed in a PostgreSQL 8.3 database, and statistics were computed using SPSS.

Table 4.1.III: Patient Characteristics for the Example Study.

	Ambulatory	Inpatient
Number of cases	49	53
Age	73.7±7.8	68.0±11.2
Sex (m/f)	20/29	22/31
Treated eye (right/left)	27/22	23/30

Table 4.1.IV: Interventional phases for the cataract surgeries example.

Phase	Definition
Capsulorhexis	First paracentesis until end of material excision
Lens Removal	Hydrodissection until end of irrigation/aspiration of lens
Lens Implantation	Cut widening until beginning of irrigation/aspiration of Healon
Removal of Healon	Irrigation/aspiration of Healon

Table 4.1.V: Terminology list examples for the cataract surgeries example

WHO	WHAT	WHEREBY	WHERE
surgeon with left hand, surgeon with right hand, …	apply, aspirate, capsulorhexis, close, coagulate, cut, disinfect, hydrodissection, implant, inject, insert, irrigate, place, remove, widen, …	bipo, chopper, circula, colibri tweezers, drape, eye drain, foil scissors, hooklet, lancet clear cut, monarch, …	bulbus oculi, capsula lentis, capsular sac, caput, chamber ant, cilia, conjunctiva, cornea, cortex, …

Table 4.1.VI: Activity examples recorded by observation.

	Example activity 1	Example activity 2	Example activity 3
WHO	surgeon with right hand	surgeon with right hand	surgeon with left hand
WHAT	hydrodissection	wash	hold
WHEREBY	sauter cannula	sprinkler cannula	colibri tweezers
WHERE	cortex	conjunctiva	bulbus oculi
WHEN	00:05:30 – 00:06:10	00:02:30 – 00:02:40	00:03:35 – 00:05:05

Results

The general assessment of the cut-suture times showed a significant difference ($p<0.001$) between ambulatory and inpatient cataract procedures. Mean cut-suture times were 00:16:01±00:04:39 for ambulatory interventions and 00:25:16±00:15:34 for inpatient interventions (cf. Table 4.1.VII).

The interventional phases of *Capsulorhexis*, *Lens Removal*, *Lens Implantation*, and *Removal of Healon* constitute the surgical core of the intervention. In a second step, the total duration of these core phases was examined for both samples. The total duration for the interventional core phases was significantly different ($p<0.001$) and was 00:09:50±00:03:22 for ambulatory and 00:17:32±00:16:09 for inpatient interventions.

An investigation of the durations of the phases *Capsulorhexis*, *Lens Removal*, *Lens Implantation*, and *Removal of Healon* showed significant differences in the mean durations as compared to those of the phases *Capsulorhexis* ($p<0.001$; cf.Table 4.1.VII) and *Lens Removal* ($p=0.002$).

Example results are presented for the *Capsulorhexis* phase in

Table 4.1.VIII. The analysis revealed that during this phase in inpatient cataract interventions all activity performances took significantly longer than did the same activities in ambulatory interventions. Except for the activity *left hand hold(s) bulbus oculi (with) colibri tweezers*, the number of occurrences of the respective activities was not significantly different.

The surgeons' left hand used several different instruments. The micro spatula was not used at all in inpatient interventions.

Assessing the gSPMs for activity sequences revealed the most frequent transitions consistent with the surgical work sequences. The generic SPMs computed for the *Capsulorhexis* phase, using the example data, are shown for both samples in Figure 4.1.5. Both gSPMs were filtered with a threshold of 5%, and all transitions with a global probability of less than this threshold were deleted from the gSPM visualizations. Furthermore, the most probable paths were highlighted in each of the gSPMs (grey shaded activities). Due to the concurrent behavior of the surgeons' left and right hands, there are two main paths for each sample. As a simple criterion, all transitions connected to the main path that appeared in more than 50% of the respective iSPM sample were highlighted using bold lines. Solid lines symbolize the work flow of the surgeons' right hand, while dotted lines symbolize the work flow of the surgeons' left hand.

In

Table 4.1.IX, the significance of transitions between activities during the *Capsulorhexis* interventional phase is shown. Sample results are presented for all highlighted transitions along the main path of each hand. Both strategies were significantly different for the path of the surgeons' right hand for the *paracentesis* → *Healon injection* transition. This results from the existence of the alternative *paracentesis* → *Vision Blue® injection* → *irrigation* → *Healon injection* path that did not occur in ambulatory interventions. Additionally, the use of a different surgical instrument for left-handed *holding* is reflected in the gSPMs.

Table 4.1.VII: Phase durations.

mean±sd [CI95%]	**Ambulatory cataract interventions**	**Inpatient cataract interventions**	**p value**
Cut-suture time	00:16:01±00:04:39 [00:14:40,00:17:21]	00:25:16±00:15:34 [00:20:08, 00:25:37]	t(61.91)=-4.13, **p<0.001**
Begin *Capsulorhexis* until end *Removal of Healon*	00:09:50±00:03:22 [00:08:51, 00:10:48]	00:17:32±00:16:09 [00:12:15, 00:17:45]	t(56.90)=-3.39, **p=0.001**
Capsulorhexis	00:01:28±00:00:28 [00:01:20, 00:01:36]	00:02:48±00:01:11 [00:02:24, 00:03:00]	t(68.79)=-7.56, **p<0.001**
Lens Removal	00:05:42±00:02:24 [00:05:00, 00:06:23]	00:10:18±00:09:55 [00:07:18, 00:10:48]	t(58.61)=-3.28, **p=0.002**
Lens Implantation	00:00:42±00:00:58 [00:00:34,00:00:50]	00:00:58±00:01:05 [00:00:40, 00:01:16]	p>0.05
Removal of Healon	00:01:37±00:01:18 [00:01:15, 00:02:00]	00:01:41±00:02:20 [00:00:55, 00:02:14]	p>0.05

Table 4.1.VIII: Example durations of activities of the *Capsulorhexis* phase (in seconds, right hand activities shaded).

mean±sd [CI95%]	No. of occurrences in ambulatory sample	No. of occurrences in inpatient sample	No. of occurrences per intervention in ambulant interventions	No. of occurrences per intervention in inpatient interventions	p-value No. of occurrences per intervention	Average performance time ambulant cataract interventions in seconds	Average performance time inpatient cataract interventions in seconds	p-value Average performance time per activity
surgeon right hand paracentesis paracentesis knife cornea	49	34	1.00±0.00 [1.00; 1.00]	1.00±0.00 [1.00; 1.00]	-	6.06±1.92 [5.51; 6.61]	14.56±23.47 [6.37; 22.75]	t(33)=-2.11, **p=0.04**
surgeon right hand inject Healon chamber ant	47	52	1.09±0.28 [0,91;1,24]	1.27±0.6 [1.10; 1.44]	**p>0.05**	4.38±1.24 [4.01; 4.74]	6.15±1.26 [5.80; 6.50]	t(97)=-7.03, **p<0.001**
surgeon right hand capsulorhexis rhexis cannula capsula lentis	48	51	1.04±0.2 [0.98; 1.10]	1.18±0.56 [1.02; 1.33]	**p>0.05**	33.94±8.96 [31.33; 36.54]	64.37±23.43 [57.78; 70.96]	t(65.09)=-8.63, **p<0.001**
surgeon right hand cut lancet clear cut cornea	48	53	1.04±0.2 [0.98; 1.10]	1.02±0.14 [0.98; 1.06]	**p>0.05**	3.54±0.99 [3.25; 3.83]	4.75±1.25 [4.41; 5.10]	t(97.20)=-5.42, **p<0.001**
surgeon right hand excision material Utrata`s tweezers capsula lentis	48	53	1.21±0.41 [1.09; 1.33]	1.11±0.32 [1.03; 1.20]	**p>0.05**	4.38±1.92 [3.83; 4.93]	6.02±3.07 [5.17; 6.86]	t(88.43)=-3.26, **p=0.002**
surgeon left hand hold colibri tweezers bulbus oculi	13	53	1.08±0.28 [0.91; 1.24]	1.49±0.75 [1.28; 1.70]	t(53.75)=-3.22, **p=0.002**	33.92±14.73 [25.02; 42.82]	78.02±27.74 [70.37; 85.67]	t(35.73)=-7.89, **p<0.001**
surgeon left hand hold micro spatula bulbus oculi	38	0	1.16±0.37 [1.03; 1.28]	-	-	44.5±16.56 [39.08; 49.94]	-	-

Table 4.1.IX: Differences in *Capsulorhexis* activity transitions (for activities on the 'mean' path).

Start activity	Stop activity	global transition probability in ambulatory interventions (N=49)	global transition probability in inpatient interventions (N=53)	p value
start	surgeon right hand paracentesis paracentesis knife cornea	1.00±0.00 [1.00; 1.00]	0.64±0.48 [0.51; 0.77]	t(52)=5.39, **p<0.001**
surgeon right hand paracentesis paracentesis knife cornea	surgeon right hand inject Healon chamber ant	0.94±0.24 [0.87; 1.01]	0.49±0.50 [0.35; 0.63]	t(76)=5.78, **p<0.001**
surgeon right hand inject Healon chamber ant	surgeon right hand capsulorhexis rhexis cannula capsula lentis	0.81±0.38 [0.70; 0.92]	0.83±0.35 [0.73; 0.93]	p>0.05
surgeon right hand capsulorhexis rhexis cannula capsula lentis	surgeon right hand cut lancet clear cut cornea	0.92±0.26 [0.84; 0.99]	0.84±0.33 [0.75; 0.93]	p>0.05
surgeon right hand cut lancet clear cut cornea	surgeon right hand excision material Utrata`s tweezers capsula lentis	0.94±0.22 [0.88; 1.00]	0.99±0.07 [0.97; 1.00]	p>0.05
surgeon right hand excision material Utrata`s tweezers capsula lentis	end	0.86±0.27 [0.78; 0.93]	0.94±0.16 [0.90; 0.99]	p>0.05
start	surgeon left hand hold colibri tweezers bulbus oculi	0.24±0.43 [0.12; 0.37]	0.58±0.50 [0.45; 0.72]	t(99.689)=-3.68, **p<0.001**
surgeon left hand hold colibri tweezers bulbus oculi	end	0.23±0.42 [0.11; 0.36]	0.80±0.28 [0.72; 0.88]	t(81.514)=-7.944, **p<0.001**
start	surgeon left hand hold micro spatula bulbus oculi	0.65±0.48 [0.51; 0.79]	0.00±0.00 [0.00; 0.00]	t(48)=9.51, **p<0.001**
surgeon left hand hold micro spatula bulbus oculi	end	0.69±0.43 [0.57; 0.82]	0.00±0.00 [0.00; 0.00]	t(48)=11.28, **p<0.001**

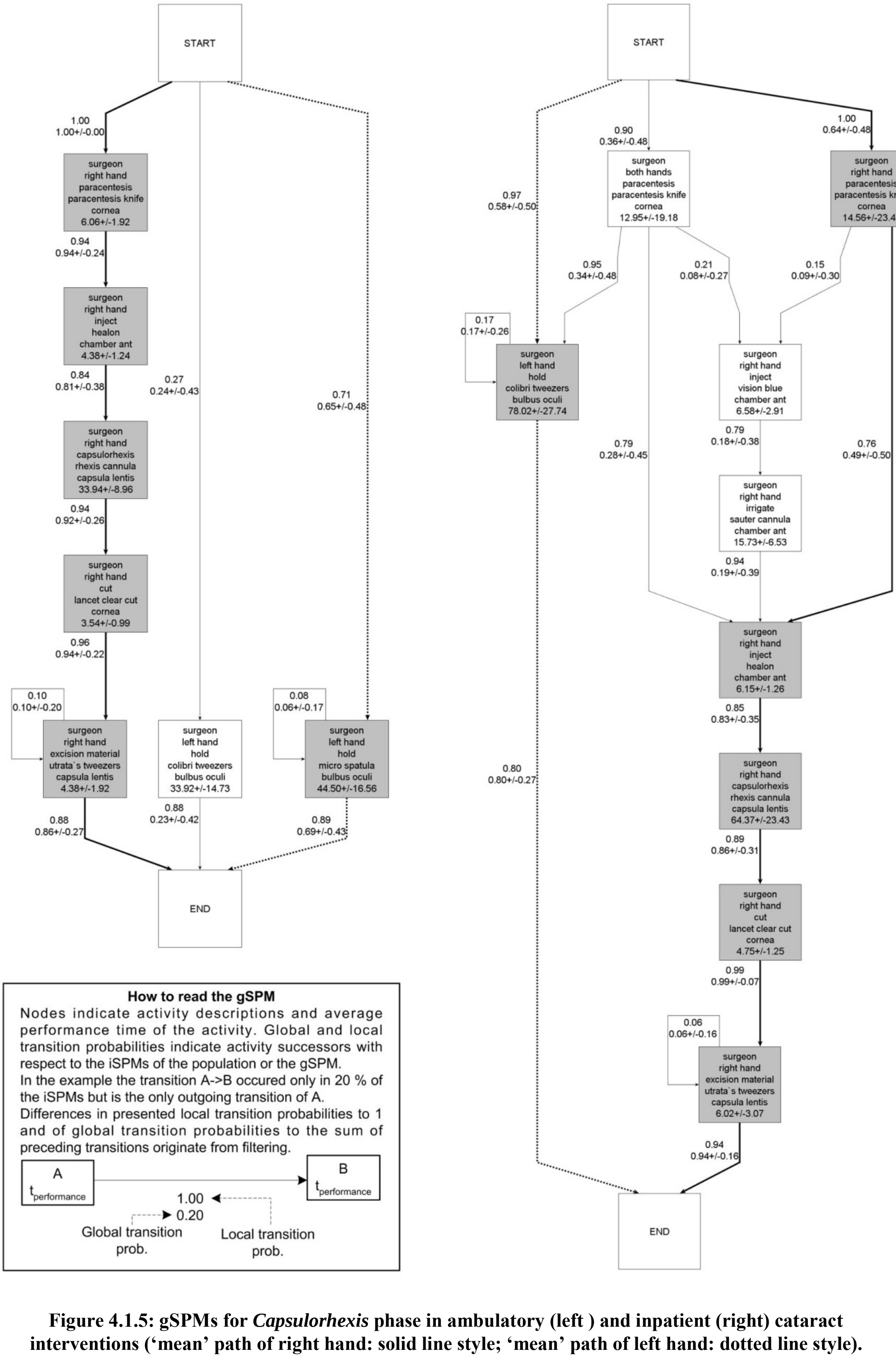


Figure 4.1.5: gSPMs for *Capsulorhexis* phase in ambulatory (left) and inpatient (right) cataract interventions ('mean' path of right hand: solid line style; 'mean' path of left hand: dotted line style).

Discussion

To the best of our knowledge, this is the first approach that presents the computation of a generic surgical process model – a statistical 'mean' surgical treatment based on extensive samples of detailed clinical data. As this work has shown, it is possible to create realistic gSPMs from real clinical data. The method presented in this paper showed the essential steps for building gSPMs and using them to assess the surgical work flow of intervention samples.

The example use case compared clinical data from ambulatory and inpatient cataract interventions and demonstrated that differences between two 'mean' treatments can be assessed and analyzed in detail. For the clinical example data, reasons for differences in procedure times of both surgical treatment strategies could be traced back to the work flow and to the performance of individual work steps in both samples. Our calculated gSPMs for the clinical use case data demonstrated several differences in treatment strategies, which could be expressed in terms of temporal information, as well as by workflow transition disparities. Our example use case showed that these transitional disparities can be clearly identified, quantified, and analyzed with the help of gSPMs.

From the clinical point of view, differences between gSPMs were investigated in detail by firstly analyzing differences in cut-suture times, secondly identifying the significant differences between the performance times of the interventional phases, and thirdly analyzing the differences in the surgical work flow on the activity level. However, the presented methods work for other intervention types as well, provided they have been recorded using the same methods as described here with technical expert knowledge. The application of the gSPM method is also feasible for tight budget system development projects and short-term clinical process improvement projects. The presented study was performed by the support of one medical student that recorded the iSPMs during a half year. However, data acquisition costs might be decreased further by the application of automatic recording systems.

We have considered the application of the overall method from the technical point of view and neglected possible biases and confounders from the clinical point of view, such as or the number of surgeons involved or the complexity of the surgical cases and therefore the allocation of the patients to the inpatient group, to show the feasibility of the approach. The cataract interventions in this article have not yet been interpreted from a clinical viewpoint. The differences between ambulatory and inpatient cataract interventions have been used only to provide a clinical example use case to present the idea of gSPMs and to illustrate the application of our methods. These decisions were reasonable because the method focuses on a proof-of-concept for generating the gSPM from any sample or population, not on building relevant, exhaustive and significant gSPMs. From the technical point of view, the method does not need to be free of confounders, because they are eliminated by calculating the means and subsequent filtering which are both inherent to the gSPM method. However, from the clinical point of view, confounder control is of course necessary.

The output of the gSPMs can be adapted to meet a given user's needs. Perspectives and activities can be chosen freely, resulting in models of higher or lower complexity. The more perspectives are concurrently selected, the greater the complexity of the resulting gSPM, and vice versa. Furthermore, a decrease in complexity, resulting in improved lucidity and a higher granularity, can be achieved by segmenting the iSPMs into parts, e.g. based on interventional phases.

Calculating the transitions between activities also had a side effect: the bottom-up identification of a 'mean' procedure course from the data. By following the transitions with the highest probabilities from the artificial *START* to the *END*, a statistical 'mean' procedure course was identified. It was possible to recover gSPMs that corresponded to recommended surgical treatments for both strategies. Clinical experts checked the resulting mean intervention courses to ensure they corresponded with the recommended cataract treatments.

The registration step between iSPMs in this study was based on the literal similarity of the features. This was appropriate in the context of gSPM calculations from the technical point of view, but it does not consider semantic similarities between work step descriptions. Computing the structural gSPM generated a purely logic-oriented model that only presents sequences of work steps.

To assess the transition probabilities between activities, only binary relations based on predecessor-successor relations were considered. Here, several other approaches could also be considered, such as data mining strategies [Blum et al. 2008b; de Medeiros et al. 2005a; van der Aalst et al. 2003]. Examining the sequence of transitions before the current predecessor might lead to a shift of probabilities. However, binary relations were chosen so that the models could be calculated with less complexity.

The objective of this work was to present a method to calculate gSPMs. However, further research is needed to investigate appropriate models from the clinical point of view, with a focus on the clinically relevant granularity of the gSPMs, the inclusion of several perspectives as features, and consideration of the 'history.' The models can also be improved by explicit indication of concurrent activities, a step that was neglected in our example use cases so as not to overload the visual representations.

Conclusion

Statistical 'mean' procedure courses can be computed as generic surgical process models (gSPMs) and differences between several of these models can be quantified. This is a new approach supporting increased evidence for clinical, technical, and administrative decision making.

Several clinical application cases for gSPMs emerge from the new methods in the context of quality management. Besides comparing surgical strategies, we could also quantify the use of different surgical technologies to achieve the same surgical goal. This makes an assessment of the influence of surgical assist systems possible. Using gSPMs to train residents allows for an assessment of their progress. Furthermore, intervention costs may be calculated in more detail to improve the hospitals' billing efficiency or other financial issues. For instance, operating rooms in hospitals command a vast amount of human resources, device resources, and materials. For this reason, they represent one of the most cost-intensive sectors in hospitals [Cleary et al. 2005; Archer and Macario 2006]. The use of these resources for individual patient treatments is usually estimated by measurement parameters such as cut-suture times or by derived parameters such as turnover rates [Schuster et al. 2007]. However, cut-suture times do not provide the level of detail of information about the statistical 'mean' treatment of an intervention population as generic surgical process models. This can be put to a multitude of possible uses such as the estimation of resource needs for surgical interventions or an examination of differences in surgical work flow, which may ultimately support administrative billing.

Consequently, the bottom-up identification of the 'mean' intervention course allows for a further application case: the comparison of an individual surgical process models (iSPMs) and generic surgical process models (gSPMs) that could be advantageous to investigate the reasons why a single surgical intervention course deviates from the mean procedure course.

In the future, detailed and rigorous analysis of gSPMs may serve as a powerful tool for surgeons to improve their work, for medical engineers to design support systems, for both to have a common, validated and standardized discussion base, and even for managing personnel to design better corporate structures, as illustrated in this article. Preclinical requirements analysis, retrospective analyses, or post-development evaluations of surgical strategies, surgical skill levels, or the use of new surgical instruments or devices are all use cases that could rely on models obtained from valid gSPMs. From the technical point of view, gSPMs can be also used as a pre-stage in developing workflow management support for the digital operating room of the future.

Acknowledgments

We thank the team that supported this work at the Innovation Center Computer Assisted Surgery, University of Leipzig and the Department of Eye Surgery, University Hospital Leipzig: M. Ceschia, D. Goldstein, S. Schumann, and M. Thiele. ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) within the scope of the Unternehmen Region with grant numbers 03 ZIK 031 and 03 ZIK 032.

4.2 Process model-based generation of workflow schemata

Title

Process model-based design and validation of surgical workflow management schemata for cataract procedures

Authors

Thomas Neumuth[1], Philipp Liebmann[1], Peter Wiedemann[2], Jürgen Meixensberger[1,3]

Departments

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] University Medical Center Leipzig, Department of Ophthalmology, Leipzig, Germany

[3] University Medical Center Leipzig, Department of Neurosurgery, Leipzig, Germany

Citation

Neumuth T, Liebmann P, Wiedemann P, Meixensberger J. Surgical workflow management schemata for cataract procedures: Process model-based design and validation of workflow schemata. Methods of Information in Medicine. 2012; 51(5):371-382.

Keywords

Operative surgical procedures; Workflow; Computer-assisted decision making; Information system; Computer-assisted surgery; Process assessment (Health Care); Surgical process model

Abstract

Objective: Workflow guidance of surgical activities is a challenging task. Because of variations in patient properties and applied surgical techniques, surgical processes have a high variability. The objective of this study was the design and implementation of a surgical workflow management system (SWFMS) that can provide a robust guidance for surgical activities. We investigated how many surgical process models are needed to develop a SWFMS that can guide cataract surgeries robustly.

Methods: We used 100 cases of cataract surgeries and acquired patient-individual surgical process models (iSPMs) from them. Of these, randomized subsets iSPMs were selected as learning sets to create a generic surgical process model (gSPM). These gSPMs were mapped onto workflow nets as workflow schemata to define the behavior of the SWFMS. Finally, 10 iSPMs from the disjoint set were simulated to validate the workflow schema for the surgical processes. The measurement was the successful guidance of an iSPM.

Results: We demonstrated that a SWFMS with a workflow schema that was generated from a subset of 10 iSPMs is sufficient to guide approximately 65% of all surgical processes in the total set, and that a subset of 50 iSPMs is sufficient to guide approx. 80% of all processes.

Conclusion: We designed a SWFMS that is able to guide surgical activities on a detailed level. The study demonstrated that the high inter-patient variability of surgical processes can be considered by our approach.

Introduction

Modern operating rooms are equipped with a variety of technical devices. The purpose of these devices is to support the surgeons' work, i.e., to achieve surgical efficiency by decreasing the invasiveness of the surgical strategy, reducing work complexity, and performing cost-effective treatments. Most of the technical equipment is stand-alone technology that fulfills a dedicated task during a dedicated surgical work step. Unfortunately, comprehensive cooperation between these devices is rarely possible due to of a missing 'global' guidance system that supports the overall surgical process on the one hand, and amends the lack of interoperability between the single devices [Lemke and Vannier 2006; Cleary and Kinsella 2005; Sandberg et al. 2003; Deinhardt 2003; Patkin 2003; Jolesz and Shtern 1992] on the other hand.

Currently, no workflow guidance has been developed for a "digital" operating room with extended connectivity and interoperability of devices that, for instance, displays context-sensitive information, depending on the current situation of the surgery, by augmenting microscopic views or surgical displays, that triggers and parameterizes technical devices and surgical assist systems, such as intraoperative measurements, that supports quality management by automatically documenting the surgical procedure, or that enhances the facilities for surgical training.

Management systems with global knowledge concerning the guided business process are well established for administrative business applications [Jablonski and Bussler 1996; van der Aalst and van Hee 2002]. These systems support the performance of standardized business processes by providing data and information to support the accomplishment of administrative processes and activities. Thus, resource use is optimized, and business operation costs are reduced. These systems should be transferred to the operating room, which is one of the most cost-intensive units in hospitals [Cleary and Kinsella 2005; Sutherland et al. 2005; Archer and Macario 2006].

However, until now, the application of business process modeling methods for surgical processes is hardly possible, due to the high variability of the latter. This high variability is caused by individual patient properties, such as anatomical characteristics, surgical capability and techniques, or the by the use of different technological resources [Neumuth et al. 2009b]. The standardized generation of process models based on expert knowledge, partly derived from clinical guidelines [AHRQ-Agency for Health Care Research and Quality 2010b; AWMF-Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. 2010b], is hardly applicable due to the high level of detail of workflow schemata that is required to support surgical activities with workflow management. The objective of this work is to demonstrate how to overcome these challenges using the example of cataract surgeries, having the largest proportional variability among specialties [Dexter et al. 2010].

In the pertinent literature, different approaches to workflow management in hospitals have been described. General requirements for workflow support in the health care domain were highlighted by Mans et al. [Mans et al. 2010a]. Approaches for workflow management support were presented to assist the performance of clinical guidelines, protocols, or clinical trials [Quaglini et al. 2001; Quaglini et al. 2000; Greiner et al. 2005; Haux et al. 2003; Latoszek-Berendsen et al. 2010]. More specifically, workflow management systems were used to support patient registration in hospitals [Kyriacou et al. 2006], to control the provision of supplies and instruments [Sutherland and van den Heuvel 2006], or to manage unscheduled health care processes [Mans et al. 2010b].

Workflow management systems were also used to support the work of clinical departments, such as emergency healthcare [Poulymenopoulou et al. 2003], radiology [Zhang et al. 2009; Halsted and Froehle 2008], or gynecology [Reichert et al. 2000]. Disease-related applications were published for stroke management [Panzarasa et al. 2006] and heart-disease identification [Jung et al. 2009]. Additionally, the work of surgical wards and nursing [Zai et al. 2008; Hansen and Bardram 2007; Agarwal et al. 2007; Prinyapol et al. 2009] and medical image processing [Fissell 2007; Krefting et al. 2010] has been supported.

Inside the operating room workflow management systems are considered as patient-safety critical systems [Bardram and Nørskov 2008]. Workflow support of surgical processes so far has examined two fields: anesthesia [Riedl 2003; Gebhard et al. 2003] and computer-assisted surgery [Münchenberg et al. 2001b; Dickhaus et al. 2004; Qi et al. 2006]. To provide high-level support for surgical processes, Dickhaus et al. have applied intraoperative workflow modeling to brachytherapy interventions [Dickhaus et al. 2004], and Münchenberg et al. have used a workflow management system for robot control in cranio-maxillo-facial surgery [Münchenberg et al. 2001b].

However, existing approaches have some limitations. An application of any of the mentioned approaches to our goal does not seem reasonable because they are either focused on related fields, such as radiology or nursing, with altogether different types of processes that require much less detailed process support than surgery, they consider superordinated processes, such as clinical guidelines, or they are specifically suited to one technology such as surgical robots.

However, none of the available approaches has dealt with the generation of workflow schemata from patient-individual surgical processes to encompass the high variability. In contrast to existing approaches, this work emphasizes the implementation and intraoperative application of a surgical workflow management system (SWFMS) that works with workflow schemata generated from individual process models. We provide an approach that, on the one hand, considers the high variability of surgical processes and, on the other hand, provides process models with a high level of detail.

In a broader sense, approaches for the generation of generic models from individual processes have been reported for simulated hospital process logs [Maruster et al. 2001] or for the modeling of peripheral processes in the operating room [Barkaoui et al. 2002]. Additionally, mining algorithms were used to discover process models in clinical pathways [Ceglowski et al. 2005; Mans et al. 2008; Mans et al. 2009; Lang et al. 2008; Zhou and Piramuthu 2010; Fernandez-Llatas et al. 2010]. These works did not use their models further to generate workflow schemata.

The SWFMS and the bottom-up generation of the workflow schema are described in the Methods section. The design and the experimental results of the system validation based on 100 example patient data-sets from cataract surgery are presented in the Results section. It will also be shown that high-resolution generation of the workflow schema is desirable to improve the system's ability to follow the surgical process. These schemata consider the high variability of surgical processes and can be used as basis for the development of workflow management systems in the operating rooms of the future.

Methods

Surgical process modeling and workflow schema generation

Analogously to the definition of a business process [Workflow Management Coalition 1999b], we define a surgical process as "... a set of one or more linked procedures or activities that collectively realize a surgical objective within the context of an organizational structure defining functional roles and relationships". Thus, we refer to the execution of an actual surgical procedure as a *surgical process*.

Furthermore, we define a surgical process model (SPM) as “a simplified pattern of a surgical process that reflects a predefined subset of interest of the surgical process in a formal or semi-formal representation” [Neumuth et al. 2009b]. These models are made up of activities that reflect the work steps of the surgeon during the surgical procedure. An SPM that represents a surgical process performed on a single patient is denoted as an *individual surgical process model* (iSPM), respectively a process instance.

To model an iSPM, we used the common method of modeling by observation as described by Neumuth et al. [Neumuth et al. 2009b; Neumuth et al. 2011a; Neumuth et al. 2012]. In previous works it has been shown that this method results in reliable iSPMs and is applicable to a wide variety of surgical disciplines and intervention types [Neumuth et al. 2009c; Krauss et al. 2009; Seeburger et al. 2012; Neumuth et al. 2011d]. Other approaches for automatic process model acquisition are available [Padoy et al. 2012; Lalys et al. 2010], but are not generic due to application of specific sensor systems.

To acquire data, a specially trained and experienced observer operated a modeling software, the Surgical Workflow Editor [Neumuth et al. 2006b], while he observed the surgical procedure in a live setting. The surgical process was described with the help of activities and states [Neumuth et al. 2009b]. Activities were used to describe surgical work steps and states were used to describe surgical phases. To identify activities and states, and to separate them from one another, we labeled them using a composite key. The elements of this key were called perspectives [Neumuth et al. 2009b]. The organizational perspective contained values on who performed a work step (e.g. the surgeon, the assistant), the functional perspective described *what* the acting person was doing (e.g. suctioning, grasping, cutting for activities, or the name of the surgical phase), the operational perspective described which tools were used (e.g. scalpel, hook), the spatial perspective described *where* the work step was performed (e.g. skin, muscle tissue x, bone y), and the behavioral perspective described *when* a work step or a surgical phase was performed.

To describe surgical phases, we used the functional and the behavioral perspective. To describe surgical work steps we used the functional, the organizational, the operational, the spatial, and the behavioral perspective (see Figure 4.2.1, upper part). The lower half of Figure 4.2.1 shows the activities of the surgeon during the surgical phase Capsulorhexis in cataract procedures.

The surgical activities and phases were composed by the observer during the preparation of the study and verified in discussions with a senior surgeon.

This approach of using information perspectives as composite keys was necessary to describe the activities of the surgical process in detail. The application of a simple key was unfavorable because this key would have to contain all possible combinations of the perspectives to provide the observer with a complete label set in the observation support software. By assuming mean numbers of 2-3 participants, 20 surgical actions, 20-30 instruments and supplies, and 10 anatomical and pathological structures of an average surgery, this would have result in several hundred simple keys that are not efficiently operable by the observer.

The Surgical Workflow Editor as observation support software was running on a tablet-PC and the observer selected the perspective information to create the appropriate key for the current situation. The software output was a file in the eXtensible markup language (XML) format. The iSPMs were stored in a database after having been acquired.

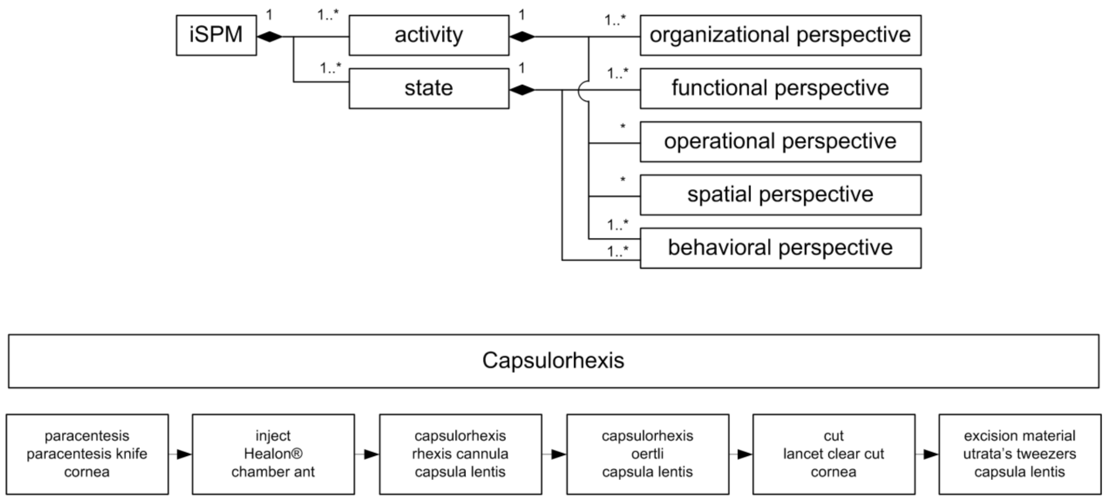


Figure 4.2.1: UML class diagram of iSPM-components (upper half) and cut-out of an iSPM. The activities of the surgeon during the surgical phase Capsulorhexis in cataract procedures (lower half, time information from behavioral perspective not depicted).

Cataract procedures from eye surgery were recorded as iSPMs in preparation of the technical study. In Germany, cataract surgery is the surgical procedure performed most often [German Federal Statistical Office 2008b]. The main surgical phases of the procedure are *Preparation*, *Capsulorhexis*, *Lens removal*, *Lens implantation*, and *Removal of Healon®*. During these phases, the capsule is sliced, the opacified lens is removed, a new lens is implanted and liquid is discharged that was used to support the procedure.

We acquired 100 clinical cases of cataract procedures. The number of cases was restricted prior to the study to limit the study costs by neglecting sensitivity. Only patients with cataract diagnosis were included in the study. The clinical data was recorded between March and September 2006 at the Department of Ophthalmology at the University Hospital of Leipzig. The clinical cases were performed by three experienced surgeons during their daily routine.

All iSPMs were acquired by one observer using the methodology described above. The observer was a medical student who had received a comprehensive training before the data acquisition. The observer received training from experienced surgeons about the characteristics of cataract procedures, including the typical intervention course and the clinical guidelines. He also received training concerning the names and applications of surgical instruments, materials, and supplies. Finally, she had to train the operation of the Surgical Workflow Editor to ensure a comprehensive handling of the observation support software.

Workflow schema generation

Subsequently, we merged several iSPMs from the iSPM database based on surgical phases and activity information to create a generic surgical process model (gSPM). gSPMs are statistically averaged models from multiple iSPMs and represent a "mean" statistical surgical procedure [Neumuth et al. 2011b].

The fusion started by splitting the iSPMs into surgical phases and by adding artificial *start* and *end* activities. Afterwards, all activities with corresponding perspective information in the same surgical phase of the different iSPMs were merged into one activity. Likewise, corresponding predecessor-successor relations between activities of the iSPMs were merged into one transition in the gSPM. Finally, transition probabilities were calculated based on the observed frequency in the iSPMs. For each activity, all subsequent transitions were labeled with the respective probability [Neumuth et al. 2011b]. Please note, that the gSPM did not contain behavioral perspective information. Behavioral perspective information was only used to determine which activities belong to which surgical phase and to identify the order of activities within a surgical phase. In previous works, it has been shown that this strategy results in reliable and clinically correct gSPMs that satisfy and adhere to the clinical guidelines [Neumuth et al. 2011b].

The resulting gSPM represented all of the activities of the respective surgical phase and their associated probabilities, where high probabilities indicated frequently occurring process model segments and mean process branches. The upper part of Figure 4.2.2 shows the UML class diagram of gSPM elements and an unfiltered example for the surgical phase *Capsulorhexis* (lower part).

Since the gSPM resulted in a model with many transitions, we also implemented a filter strategy to provide a facility to denoise the models from very infrequently occurring activities below a given filter threshold. The filter is then applied to eliminate low frequent transitions below the threshold. A filtered gSPM for the surgical phase *Capsulorhexis* and an applied filter of 1% is shown in Figure 4.2.3 (left).

Subsequently, the gSPMs of the surgical phases form the basis for the generation of the workflow schemata. The workflow schemata were used to drive the surgical workflow management system (SWFMS). The filtered gSPM structure was mapped onto workflow nets [van der Aalst 1997], a formal workflow language. Workflow nets are a petri net dialect and can be formally verified, which supports the subsequent application of the system in the sensitive OR environment. The workflow nets served as workflow schema for the open source YAWL (Yet Another Workflow Language) workflow management system (cp. Figure 4.2.3 right, [van der Aalst and ter Hofstede 2005]) which was used to perform the validation study by priming the execution of the workflows. To verify the proper functionality of the SWFMS, we checked that all 100 iSPMs could pass the workflow schema that was generated from them.

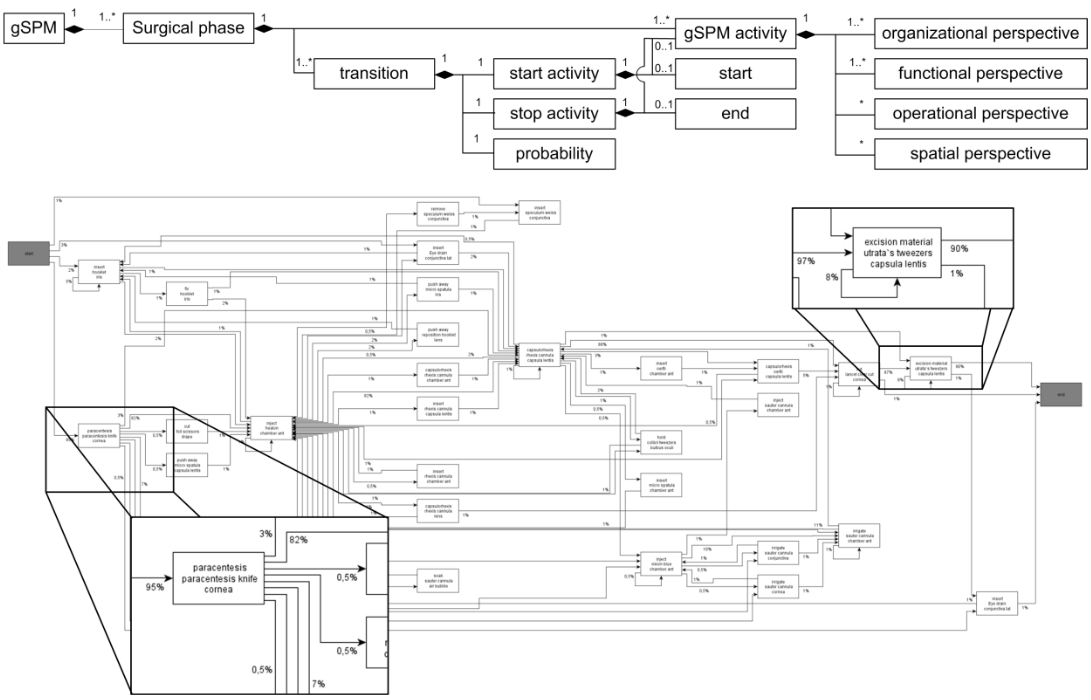


Figure 4.2.2: UML class diagram of gSPM-components (upper part) and gSPM for the surgical phase *Capsulorhexis* in cataract procedures aggregated from 90 iSPMs with activities (boxes) and transitions (edges). The value near the edges indicates the transition probability that was calculated based on the observation frequency.

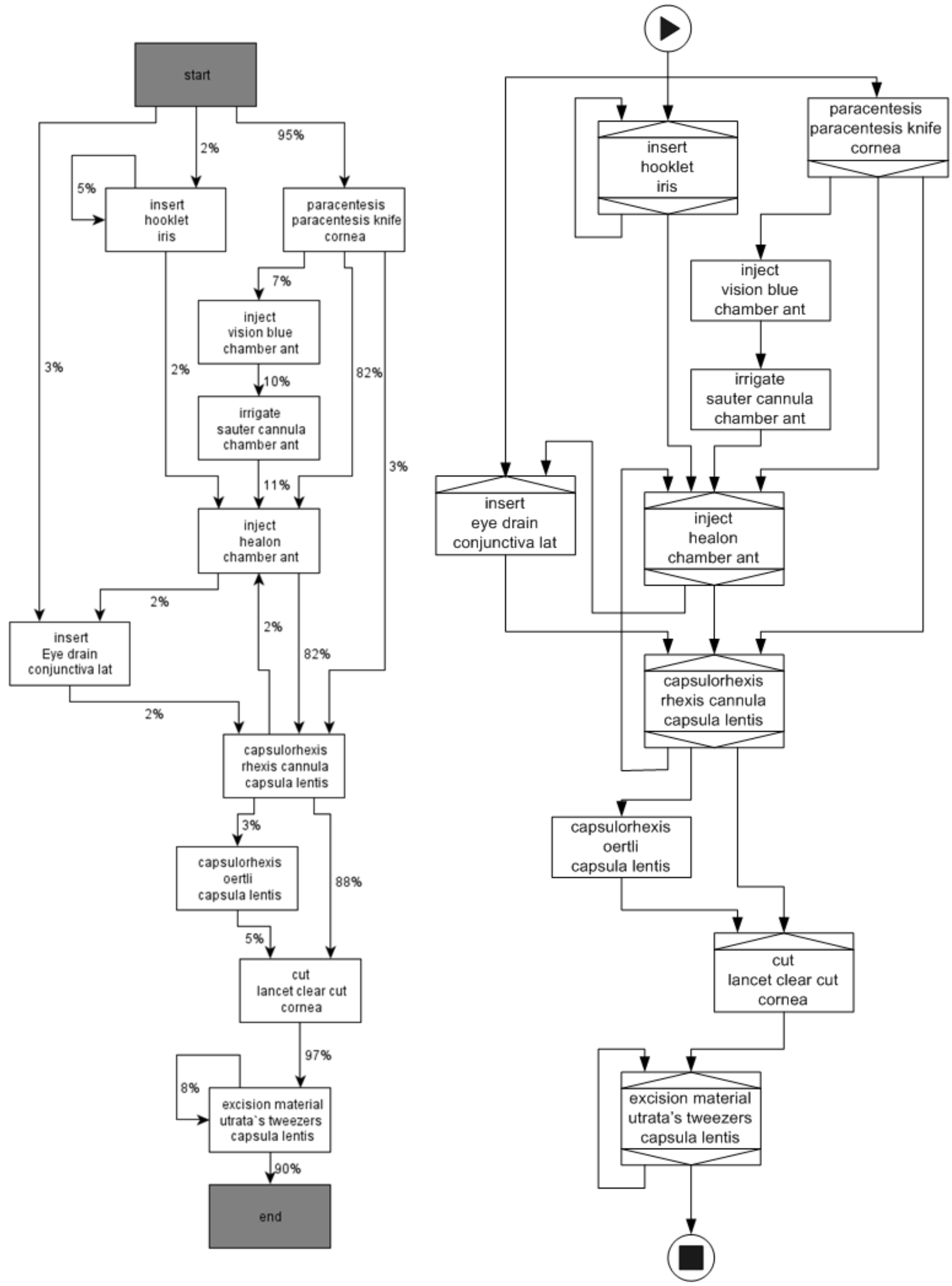


Figure 4.2.3: Generic Surgical Process Model (gSPM, numbers indicate mean transition probabilities, filtered with 1%) and derived workflow schema for the surgical phase *Capsulorhexis* (organizational perspective not depicted).

The surgical workflow management and simulation system

The surgical workflow management and simulation consisted of three parts: process model base, workflow management system, and analysis unit. In Figure 4.2.4, the structure of the simulation system is presented.

The process model base hosted the iSPM database (Postgres 8.2 database, [PostgreSQL Global Development Group 2010]), the gSPM generator, and the iSPM simulator. First, the gSPM generator randomly selected iSPMs from the database, generated a gSPM from them as described in the previous section, and denoised it according to the given filter threshold.

Next, a proprietary Java application was used to realize the mapping of a gSPM onto its corresponding workflow schema. This mapper transformed the artificial start and end activities into the corresponding elements of the YAWL language. Additionally, all perspective information of an activity in the gSPM was concatenated and mapped onto a YAWL atomic task. Afterwards, the transitions without labels were added in between the atomic tasks. Finally, the tasks were automatically transformed into XOR-joins and XOR-splits. In cases of multiple incoming transitions for one task, an XOR-join was added, and in cases of multiple outgoing transitions, an XOR-split was added to complete the workflow specification. The workflow schema was then sent to the process definition interface of the workflow management system for execution.

The iSPM process simulator randomly selected an iSPM from the database, concatenated its perspective information for each activity, and sent it, activity by activity, via web service to the process monitor interface of the SWFMS. The process monitor received the activities and forwarded them to the engine. If the perspectives of the activity matched with one of the name of the next scheduled tasks according to the workflow schema, the engine moved on to that task, logged successful execution of the task, and waited for the next task to be received. In the case of a missing transition or unscheduled task, an exception was caused, the execution of the workflow schema was terminated, and unsuccessful execution was logged. After the process simulation, the workflow log was transferred to the analysis unit for statistical analysis.

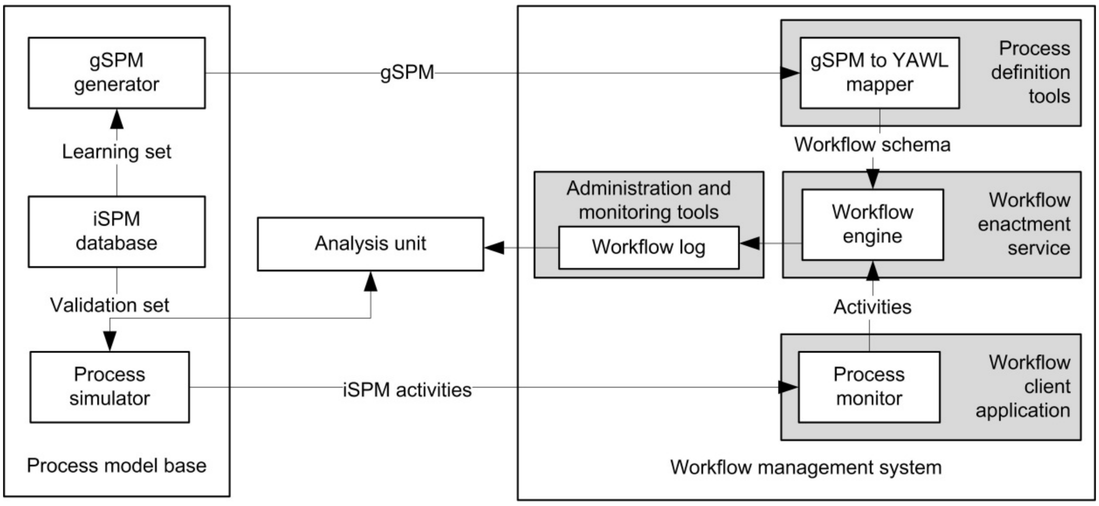


Figure 4.2.4: Design of the surgical workflow management system consisting of standard components of workflow management systems (shaded grey, [Workflow Management Coalition (WFMC) 1995]) and the simulation unit extension with the gSPM database, the gSPM generator, the process simulator, and the analysis.

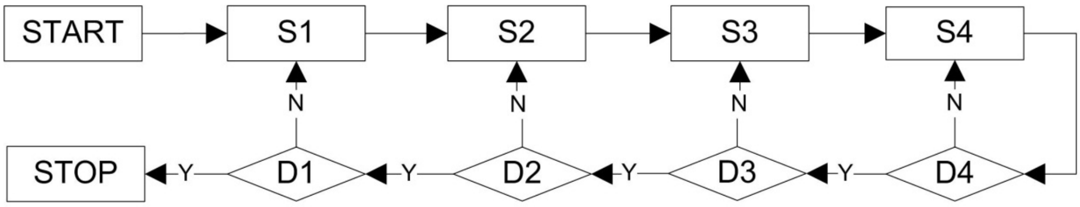
System validation study design

To determine whether the SWFMS can work with the help of a workflow schema generated from a gSPM for a surgical phase, a sophisticated study was designed.

The success of a workflow schema generated from a set of iSPMs was assessed using the success rate as a dependent variable. The success rate is a binary value that indicates whether or not the simulation of a randomly chosen iSPM was successfully finished without exception by the workflow schemata, i.e., if every surgical activity and every transition in the iSPM could be guided by the workflow schema.

To conduct the study, the design presented as a flow chart in Figure 4.2.5 was used. Initially, two disjoint subsets were generated from the whole set of all 100 iSPMs: the learning set was used to generate the gSPM and the workflow schema, and the validation set, containing 10 randomly chosen iSPMs for later testing against the workflow schema. The learning-set size increased from 10 to 90 iSPMs in steps of 10. Subsequently, the generated gSPM was filtered according to the filter value set $\{0\%, 1\%, 2\%, 3\%, 5\%, 7\%, 10\%\}$. The filtered gSPM was afterwards transformed into the workflow schema. Finally, all iSPMs of the validation set were simulated against the workflow schema. To incorporate randomization, the full study was repeated 1,000 times. We generated approx. 3,800,000 data sets based on the 6 surgical phases, 9 learning-set sizes, 7 filter levels, 1,000 repetitions during the validation study, and on 10 iSPMs in each validation set.

The statistical analysis was performed using the R-project system [R Foundation 2010] and SPSS statistics software [SPSS Inc. 2010]. Means and standard deviations were calculated for the results. A variability measure was not included, because the high correlation between surgical activities may not provide unbiased point estimates. Thus, results for standard errors or 95% confidence interval calculations might be misleading. Furthermore, a linear regression was performed to assess the influence of the independent variables of learning-set size and filter level on the success rate.



Step	Decision
(S1) Randomly select iSPMs and activities for the learning set and generate gSPM	(D1) Was the study repeated 1000 times?
(S2) Filter gSPM and generate workflow schema from gSPM	(D2) Were all filter values $\{0\%, 1\%, 2\%, 3\%, 5\%, 7\%, 10\%\}$ applied?
(S3) Select iSPM from the validation set	(D3) Were all iSPMs from the validation set simulated?
(S4) Sequentially test activity of the iSPM from the validation set	(D4) Were all activities of the iSPM simulated?

Figure 4.2.5: Test performance design for the validation study with respective steps and decisions.

Results

In our validation study we investigated how many cases of iSPMs were needed to create a robust gSPM as a basis for the workflow schema for complete surgical phases of cataract procedures.

The success rates in Table 4.2.I and Table 4.2.II indicate the number of iSPMs that were needed to generate a workflow schema that is able to complete the respective surgical phase and which filter level can be applied to limit the number of infrequently occurring transitions and activities.

Figure 4.2.6 shows the progression of the mean success rate *s* for all surgical phases. Table 4.2.I shows the rate of successful completion of unfiltered workflow schemes that were generated from a number of iSPMs given in the test set. For instance, a number of 20 iSPMs can be used to generate a workflow schema that guides 72.8% (sd=14.0%) of the cataract procedures and a test-set size of 50 iSPMs can be used to guide 79.6% (sd=12.2%) of the cataract procedures. The minimum success rate of 49.4% (sd=16.7%) for the surgical phase *Removal of Healon®* shows that even a number as low as 10 iSPMs can be used to guide 50% of the cataract procedures. Please note that the model represents a structural model and that the transitions do not include correlations to each other (see Discussion).

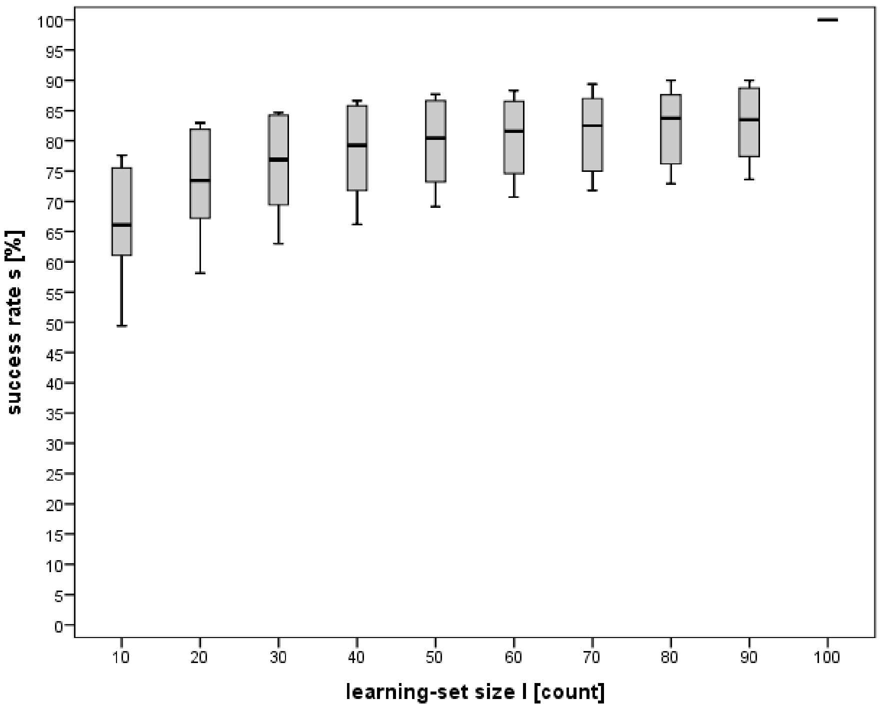


Figure 4.2.6: Progression of mean success rate for all interventional phases s depending on the number of iSPMs in the learning-set size.

Table 4.2.I: Success rates of workflow schemata with a filter level of 0% and different learning-set sizes (1st line: mean, 2nd line: standard deviation).

learning-set size (no. of iSPMs)	Preparation	Capsulorhexis	Lens removal	Lens implantation	Removal of Healon®	Completion	Mean
90	88.7 9.9	77.4 13.2	80.0 12.4	90.0 9.4	73.6 13.2	87.0 10.8	**82.8** **11.5**
80	87.6 10.5	76.2 13.3	80.3 12.2	90.0 9.0	72.9 13.5	87.2 10.8	**82.4** **11.6**
70	85.5 11.1	75.0 13.7	79.5 12.1	89.4 9.9	71.8 13.5	87.0 11.2	**81.4** **11.9**
60	84.4 11.6	74.6 12.9	78.8 12.4	88.3 10.2	70.7 14.4	86.5 11.0	**80.6** **12.1**
50	82.9 11.7	73.2 13.5	78.0 12.2	87.7 10.4	69.1 14.9	86.6 10.6	**79.6** **12.2**
40	81.9 12.5	71.8 13.5	76.6 12.8	86.6 11.2	66.2 14.7	85.8 11.1	**78.2** **12.6**
30	79.7 13.0	69.4 14.4	74.1 13.8	84.2 11.7	63.0 15.4	84.7 11.5	**75.9** **13.3**
20	75.3 14.0	67.2 14.0	71.5 14.5	81.9 12.6	58.1 15.4	83.0 13.5	**72.8** **14.0**
10	66.6 15.8	61.1 16.1	65.6 16.0	77.6 13.3	49.4 16.7	75.5 17.2	**66.0** **15.9**

Table 4.2.II shows the mean success rates for different filter levels for a learning-set size of 90 iSPMs. As expected, an increasing filter level reduces the chance for a successful completion of the workflow schema. For example, 82.8% (sd=11.5%) of the iSPMs were completely simulated in each surgical phase at a filter level of 1%, while only 74.5%±13.0% were successfully simulated at a filter level of 3%.

Table 4.2.II: Success rates of workflow schemata with a learning-set size of 90 iSPMs and different filter levels (1st line: mean, 2nd line: standard deviation).

Filter level	Preparation	Capsulorhexis	Lens removal	Lens implantation	Removal of Healon®	Completion	Mean
0.0	88.7 9.9	77.4 13.2	80.0 12.4	90.0 9.4	73.6 13.2	87.0 10.8	82.8 11.5
1.0	88.7 9.9	77.4 13.2	80.0 12.4	90.0 9.4	73.6 13.2	87.0 10.8	82.8 11.5
2.0	80.2 12.2	69.9 14.6	76.4 13.5	86.0 10.7	63.5 14.2	85.1 11.3	76.9 12.8
3.0	77.9 12.8	66.1 14.6	72.1 13.7	83.8 11.2	61.8 14.6	85.1 11.3	74.5 13.0
5.0	72.3 14.2	63.1 15.1	70.5 13.8	77.9 12.4	57.3 15.2	85.1 11.3	71.0 13.7
7.0	67.4 13.7	61.9 14.9	70.1 14.1	77.1 12.9	49.1 15.3	85.1 11.3	68.5 13.7
10.0	67.4 13.8	61.9 14.9	63.2 14.7	70.0 13.5	45.9 15.1	84.4 12.5	65.5 14.1

Combinations of the results of the two independent variables learning-set size and filter level are shown for two surgical phases as contour maps in Figure 4.2.7. The figures indicate the "success border" for the phase Capsulorhexis (left) and the for the phase Completion (right). According to the desired success rate, which is represented by the respective contour lines, learning-set size and filter level can be chosen. For instance, a desired successful workflow guidance of 75% of cataract cases for the phase Capsulorhexis can be achieved by using at least 70 iSPMs and applying a maximum filter level of 1%. In contrast, the same result may be achieved for the phase Completion by using only 40 iSPMs and applying a maximum filter level of 6%.

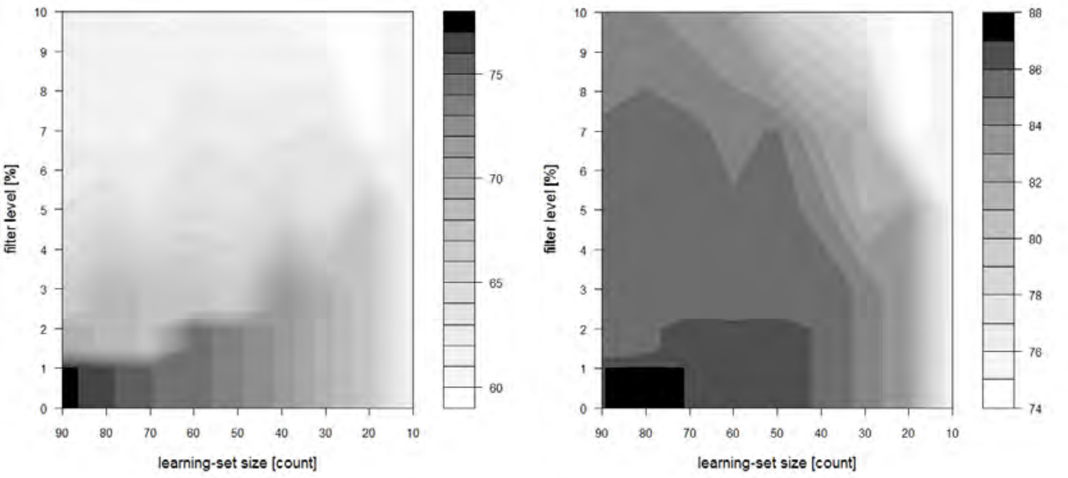


Figure 4.2.7: Contour plots for the surgical phase *Capsulorhexis* (left) and *Completion* (right) show the levels of success rate (percentages) depending on the learning-set size and the filter level.

The prediction of the success rate s by the independent variables using learning-set size l and filter level f was investigated by performing a linear regression analysis. The R-square values in Table 4.2.III indicate the coefficient of determination ranges between 0.631 for the phase *Completion* and 0.760 for the phase *Lens implantation*. The linear regressions showed intercepts between 60 and 84. All regression coefficients were significant (p<0.001).

Table 4.2.III: Linear regression equations and R^2 values of the regression equations for the success rate s depending on the learning-set size l and the filter level f.

Surgical phase	Linear Regression equation	R-square of regression
Preparation	$s = 76.16 + 0.11l - 1.69f$	0.699
Capsulorhexis	$s = 67.83 + 0.07l - 1.23f$	0.658
Lens removal	$s = 72.20 + 0.09l - 1.26f$	0.744
Lens implantation	$s = 83.99 + 0.06l - 1.42f$	0.760
Removal of Healon	$s = 60.44 + 0.13l - 2.23f$	0.734
Completion	$s = 80.19 + 0.10l - 0.49f$	0.631

Discussion

Usual methods for creating workflow schemata for the workflow management systems face the challenge of encompassing the high variability of surgical processes. Within our work, we presented a strategy to overcome this challenge. Due to the acquirement of gSPMs from iSPMs, it was possible to consider the high variability of observed surgical processes. Our method allowed for the estimation of the number of iSPMs that were required to compute a workflow schema for cataract surgeries. Based on two independent variables, the number of necessary iSPMs for a desired success rate could be estimated in advance to the system's design. The R-square values for the regression formulas suggested this dimension with a value of 70%. Although the presented approach was designed for cataract surgeries, we expect the methodology to be applicable to different surgery types, since the presented data acquisition strategy and the post-processing methods are not dependent on a specific type or kind of surgery.

It is not yet clear what a "good" percentage for a successful guidance of cataract surgeries is, and which success rates are clinically accepted. This needs to be investigated in future clinical studies with the objective to adhere to the right balance between clinical benefit and economic effort to design and execute SWFMS. However, we provided the necessary requirements for these investigations by introducing the method of workflow schema generation from iSPMs to consider high variability. On the other hand, the deployment of the filter threshold as mechanism to denoise the models provides a mean for the "fine-tuning" of the required success rate, e.g. with some practical applications for removing blur of visualization for a surgical decision support system that shows the next possible activities on a screen for a learning surgeon.

Our methods can be further improved by developing the computation strategy of the gSPM. The current gSPM generation approach used transitions between activities and concatenated them to build the model. Since transitions were considered locally and only between predecessor and successor activity, the model did not contain a history or trace. The computation of a gSPM that incorporates Bayesian analyses is a valuable future work topic. Although some more advanced approaches are available in current research in business information systems [van der Aalst et al. 2003; Cook and Wolf 1995; Agrawal et al. 1998; Schimm 2004; de Medeiros et al. 2005b], we applied the more elementary method to support the understanding and discussions with clinicians to verify the correctness of the models.

Further investigations also comprise the development of an effective exception handling with automatic recovery that needs to be implemented after an abortion of the workflow schema. Additionally, the system's behavior in response to inconsistent data needs to be investigated. Furthermore, it must be determined if and how fast the system can resume tracking. Finally, testing the system by automatically monitoring the process data could support SWFMS development, which could be done by replacing or enhancing the process generator with sensor system input.

The application of surgical workflow management to cataract interventions is of particular interest for many use cases. In combination with an accurate sensor strategy to automatically recognize the current surgical work step, such as proposed in [Lalys et al. 2011] or [Neumuth and Meißner 2012], and customized web services for command performances, several applications might emerge.

The first group of applications is the intraoperative presentation of preoperatively acquired information by augmentation in microscopic views or monitors. Preoperative examination results, such as corneal topography, wavefront aberrotomy, autorefraction, keratometry, pupillometry, automated assessments of cataracts [Cheung et al. 2011; Ligabue and Giordano 2007; Maeda 2009; Tanabe et al. 2011], or lens-power calculation [Jasman et al. 2010], can be visualized dependent on the situation by augmented reality to support the surgeon during critical work steps.

Secondly, technical devices and assist systems could be parameterized and intraoperative measurements could be triggered by the workflow management system. Examples for these applications comprise intraoperative dioptric power measurements [Carvalho et al. 2002], real-time intraocular pressure measurements during various stages of the cataract surgery [Kreutzer et al. 2010], measurements of incision quality by medical imaging [Leng et al. 2008; Schallhorn et al. 2008], and augmentation of the imaging models with the real microscopic view, or automatic image-capturing and transmission in the context of ophthalmologic telemedicine [Cuadros and Bresnick 2009].

Furthermore, surgical workflow management of cataract surgeries is of relevance for quality management, documentation, and patient scheduling. A trigger for the automatic transfer of intraoperative measurements, such as results of the calculation of the position of the implanted lens for later checking with postoperative follow-ups [Becker et al. 2004; Becker et al. 2006; Acharya et al. 2010], to the electronic patient record, might support the quality management of the surgery. Generally, the trace of the process instance through the model can be recorded for automatic documentation of the surgical activities during the treatment [Händel et al. 2002]. Additionally, this trace can be checked in real-time for completeness of all work steps, since studies have shown that especially novice surgeons are liable to forget single work steps [Webster et al. 2005]. In addition, the surgeon might be supported by a process navigation system that proposes next work steps until completion of the intervention. Based on the current progress of the intervention, the prediction of the completion might be calculated for the preparation of the next patient [Devi et al. 2012] and the generated workflow schema can also be used to simulate different variants of cataract surgeries and to simulate the effect of missing supplies etc. [Kubitz et al. 2001].

Finally, surgical education and training is an application field of interest. Since current virtual reality training systems for cataract procedures are mainly focused on individual parts, such as phacoemulsification or hydrodissection [Choi et al. 2009; Doyle et al. 2008; Henderson et al. 2010b; Khalifa et al. 2006; Oetting 2009; Privett et al. 2010; Santerre et al. 2007; Waqar et al. 2011], or support concepts like model-driven therapy [Cinquin and Troccaz 2003] in general, these systems might profit from the workflow schemata to enable the simulation of different variants of the surgery and surgical work steps in context to each other.

However, our approach focuses on the design, implementation, and validation of a robust surgical workflow management system from the technical point of view. Hereafter, the system can be tested in clinical practice to derive more clinical applications based on the approach.

Conclusion

The application fields of surgical workflow management systems in the digital operating room of the future are manifold. These systems might be used for situation- and context-dependent information visualization for the surgeon, such as timely presentation of previously acquired patient examination results, the automatic parameterization and control of surgical assist systems, or the provision of decision support for learning surgeons. Additionally, a communication with the hospital information system (HIS) can be employed, e.g. for an automatic and timely call for the next patient according to the predicted end time of the current intervention. All of these use cases could qualitatively ameliorate the process sequence of the surgeon and, therefore, be beneficial to the patients' safety.

The high model granularity that is required for the control of technical resources for the assistance of the surgeon causes a major challenge: The higher the granularity of the surgical process to be supported, the higher its variability. The term variability was used in this context as a general term to express the deviation of iSPMs from each other. Since no metrics exist to express the variability of surgical processes, quantification is part of the ongoing research.

The objective of this work was the design, implementation, and validation of a surgical workflow management system. It was shown that even a small learning set of 10 iSPMs can be used to generate a workflow schema for cataract surgeries that is able to guide 66% of the procedures. If higher success rates are desired, an increased number of 50 iSPMs can be used to achieve, for instance, 80% success rates.

The unique property of this system was the facilitation of a workflow schema that was generated from a number of individual surgical process models. We presented the system together with the approach, to generate the workflow schemata in a bottom-up manner from iSPMs with a comprehensive validation study on 100 patient data sets of cataract procedures from eye surgery.

The study demonstrated that the high variability of surgical processes can be considered with the presented approach, since a higher number of iSPMs can be guided by the SWFMS than the number of iSPMs that were necessary to generate the workflow schema.

Acknowledgements

The authors thank the team that supported the performance of the study and the preparation of the article at the Innovation Center for Computer Assisted Surgery, University of Leipzig: Caroline Elzner, Dayana Neumuth, Maik Müller, and Michael Thiele. The authors also thank Juliane Schlomberg from the Department of Ophthalmology, University Hospital of Leipzig, for her clinical support.

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the scope of the Unternehmen Region with grant numbers 03 ZIK 031 and 03 ZIK 032 and by funds from the European Regional Development Fund (ERDF) and the state of Saxony within the framework of measures to support the technology sector.

5 Clinical applications of surgical process models

The employment of surgical process models for the benefit of the clinical user is an imperative criterion for the successful application of the general approach. Additionally, the applicability of the methods presented in this work across intervention types and on different surgical disciplines is a key requirement. In this chapter, the application of the SPM theory is demonstrated in several use cases from different surgical disciplines.

The publication

> *Neumuth T, Trantakis C, Riffaud L, Strauss G, Meixensberger J, Burgert O. Assessment of technical needs for surgical equipment by Surgical Process Models. Minimally Invasive Therapy and Allied Technologies. 2009; 18(6):841-849.*

depicts the use of iSPMs to derive working condition parameters for a surgical assist system in neurosurgery. By means of these parameters, the requirements for the implementation of a navigated-control milling system to be used at the spine are predicted.

The outcome of the article is a method to derive requirements for a surgical assist system in neurosurgery from iSPMs. To achieve this aim, an analysis of 43 neurosurgical cases of discectomies was realized to predict temporal requirements for the automated milling system.

The second publication,

> *Neumuth T, Krauss A, Meixensberger J, Muensterer O. Impact quantification of the DaVinci telemanipulator system on the surgical workflow using resource impact profiles. International Journal of Medical Robotics. 2011; 7(2):156-64.*

demonstrates the implementation of generic surgical process models for the evaluation of surgical assist systems. In this work, the gSPMs of laparoscopic and telemanipulator-based Nissen fundoplications in pediatric surgery are compared. The notion of resource impact profile is introduced to tag gSPMs being computed with respect to the same resources employed; in this case the daVinci telemanipulator system. In the study, the system's impact on the process is investigated and quantified, resulting in the computation of gSPMs from 12 laparoscopic and 12 telemanipulator-based Nissen fundoplications, and the development of a method to quantify the impact of the telemanipulator on the surgical process by assessing the gSPMs as "resource impact profiles". Thus it was demonstrated and justified, why the use of a telemanipulator is not recommended for the presented clinical use case.

The third publication,

> *Neumuth T, Wiedemann R, Foja C, Meier P, Neumuth D, Wiedemann P. Identification of surgeon-individual treatment profiles to support the provision of an optimum treatment service for cataract patients. Journal of Ocular Biology Diseases and Informatics. 2011;3(2):73-83.*

describes the application of gSPMs to identify surgeon-specific working strategies in the context of surgical training.

In a use case from ophthalmology, the surgical process models of three different surgeons are acquired. For each of the surgeons, a gSPM is computed as *surgeon-individual treatment profile*. These treatment profiles are subsequently compared to one another to analyze differences in the working strategies, with the objective to promote the exchange of experiences between the surgeons.

The outcome of the computation of gSPMs for three different surgeons in ophthalmology as "surgeon individual treatment profiles" lead to the possibility to assess the three gSPMs to identify different working strategies of the surgeons, and, additionally, to the derivation of concrete recommendations from the treatment profiles for individual further education of the surgeons.

5.1 Deriving requirements for a surgical assist systems in neurosurgery

Title

Assessment of technical needs for surgical equipment by surgical process models

Authors

Thomas Neumuth[1], Christos Trantakis[1,2], Laurent Riffaud[3], Gero Strauß[1,4], Jürgen Meixensberger[1,2], Oliver Burgert[1]

Departments

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] University Medical Center Leipzig, Department of Neurosurgery, Leipzig, Germany

[3] University Medical Center Rennes, Department of Neurosurgery, Rennes, France

[4] University Medical Center Leipzig, Department of Otorhinolaryngology, Leipzig, Germany

Citation

Neumuth T, Trantakis C, Riffaud L, Strauss G, Meixensberger J, Burgert O. Assessment of technical needs for surgical equipment by surgical process models. Minimally Invasive Therapy and Allied Technologies. 2009; 18(6):841-849.

Keywords

Surgical Procedures, Operative; Engineering; Needs Assessment; Surgical Equipment; Electrical Equipment and Supplies

Summary

The presented approach introduces a method for estimating the potential benefit of a surgical assist system prior to its actual development or clinical use. The central research question is: *What minimal requirements must a future system meetso that its use would be more advantageous than a conventional or already existent method or system, and how can these requirements be obtained from routine clinical data?*

Forty-three cases of lumbar discectomies were analyzed with regard to activities related to bone ablation in order to predict the temporal requirements for an alternative strategy of using a surgical assist system for bone ablation. The study recorded and analyzed surgical process models (SPMs), which are progression models with detailed and exact-to-the-second representations of surgical work steps, as a sensible means for the detailed quantification of the temporal needs of the system.

The presented methods can be used for a systematic analysis of such requirements. Implementation of these methods will prove very useful in the future from a medical, technical, and administrative point of view. Manufacturers can use this analytical procedure to derive parameters for their systems that indicate success criteria. Additionally, hospitals can decide, before making actual capital expenditure decisions, if the system of interest is superior to the conventional strategy and therefore worth the investment.

Introduction

Surgical assist systems (SAS) support the work of surgeons, allowing them to work more accurately by providing support for orientation, improved access to information, and a more ergonomic environment in which to work [Cleary et al. 2005]. Appropriately-designed surgical assist systems allow for heightened surgical efficiency, and therefore have the potential to increase patient safety.

However, the development of SASs is more often inspired by emerging technologies than by the actual needs of the surgical end-user; therefore, SASs often fail to satisfy surgeons' needs [Lemke and Vannier 2006]. This leads to the question which requirements need to be fulfilled to justify expenditure of development costs for the manufacturer or investment costs for the customer.

Three primary perspectives should be considered: the medical, the technical, and the administrative points of view. From the point of view of the medical team, the prevailing concerns are for the benefit of the patient, regarding qualities such as minimal invasiveness, more precise surgery, and accelerated recovery time. Technicians, instead, focus on the accuracy, robustness, reliability, advanced usability, or ergonomics of the system. The administrative point of view emphasizes factors such as purchase price allocations, labor or overhead costs, [Cuschieri et al. 1997] and the impact of SAS on turnover rates for operating rooms.

The approach presented here introduces a method for estimating the benefit of a surgical assist system before its actual development or clinical use. The general research question is, "*What minimal requirements must a future system meet so that its use would be more advantageous than a conventional or already existent method or system, and how can these requirements be obtained from routine clinical data?*" The research question is answered exemplarily by analyzing time measures obtained from an observational study using time-action analysis, taken to estimate an upper time limit for a key function of the SAS, assuming that it should not prolong the procedure.

Customary existing parameters and reference values, such as cut-suture times, are too vague for use in deriving technical demands during the design process [den Boer et al. 2001]. The main disadvantage of these conventional criteria is that is not possible to break them down into single aspects: particular work steps, their durations, and their allocations cannot be considered using these metrics alone. To overcome these constraints, a new strategy is necessary.

The approach presented here is a novel strategy that answers this question from a scientific point of view. It puts special emphasis on parameters unveiled with the support of surgical process models (SPMs) [Neumuth et al. 2009b]. These SPMs are progression models of surgical procedures; they can depict the course of an intervention in a detailed and exact-to-the-second manner. Thus, their temporal resolution is distinctly higher than that of conventional parameters. Surgical process models can be used as fundamental sources of input for requirements engineering. These demand analyses provide the basis for a sensible prognosis of the developmental and applicative requirements that must be met by surgical assist systems. This issue is of great relevance from the clinical, the administrative, and the manufacturing point of view, as it enables direct extrapolation of the requirements of new systems from the clinical routine with the help of SPMs. Based on an appropriate logical data structure, they allow for inquiry and visualization of surgical processes.

The modeling of surgical procedures has gained focus and momentum in recent works, and various approaches to the methodology can be found in pertinent literature. Generally, it is assumed that temporal evolution of surgical tasks is important for quantitative and qualitative assessment, and most of the related work in literature is based on temporal information [MacKenzie et al. 2001; Cao et al. 1996; den Boer et al. 2002a; den Boer et al. 2002b; Strauß et al. 2006a; Fischer et al. 2006; Strauß et al. 2006b].

MacKenzie, Cao et al. [MacKenzie et al. 2001; Cao et al. 1996] suggested the use of SPMs for laparoscopic Nissen fundoplications. In the context of training junior surgeons, they created a hierarchical model that subclassifies the chosen interventional type into different levels. However, although the authors mentioned the possibility of deducing the system requirements, they did not demonstrate such deduction.

Furthermore, den Boer et al. [den Boer et al. 2002a; den Boer et al. 2002b] performed studies for laparoscopic procedures and, with the help of time-action analyses, presented an evaluation of surgical instruments and SASs. The systems were judged according to the efficiency of their applications. However, no temporal employment prognosis was deducible with this method.

A novel approach was attempted by Jannin et al. [Jannin et al. 2003; Raimbault et al. 2005; Jannin and Morandi 2007]. Within the framework of image guided surgery, the authors created models of supratentatorial tumor resections for neurosurgery. Their models, however, do not include information on the temporal evolution of the surgical procedure, and are thus inappropriate for quantitative needs assessment.

The work performed by Strauß et al. [Strauß et al. 2006a; Strauß et al. 2006b; Fischer et al. 2006] involved multiple demand analyses and postoperative evaluations for SASs for functional endoscopic sinus surgeries in Otorhinolaryngology. These early attempts, however, were conducted with a low amount of structured data and with a low level of detail.

Furthermore, other works have pursued the goal of comparing commonly applied surgical instruments, assist systems, surgical strategies, auxiliaries, such as prostheses, or the evaluation of these [den Boer et al. 1999; Sjoerdsma et al. 2000; Minekus et al. 2003].

This work demonstrates a method for a needs assessment based on surgical process models with the aid of a real surgical use case. With the help of SPM-based analyses, it is shown how detailed requirements for the employment of a surgical assist system can be extracted and formulated. An example is given related to the duration of surgical tasks, more advanced methods will be developed in future.

We demonstrate how temporal demands and parameters for the application of a surgical assist system based on navigated control [Strauß et al. 2005a] may be analyzed on the basis of forty-three real interventions. After a short introduction to the general approach, the case study, acquired data, and analysis parameters are presented in the application example section. The results section presents the measurement results for the selected parameters. Lastly, the results and validity of the concluded recommendations are discussed.

General approach

The general approach consists of several major steps, the first being the specification of the intervention type and the identification of the mission of the SAS. Subsequently, the data acquisition process and parameters need to be defined that help answering the research question by indicating criteria for the usefulness of the SAS. Finally, data acquisition, parameter analysis, and interpretation conclude the derivation of requirements.

Terms and definitions in the context of the presented work need to be defined: A surgical process (SP) is *a set of one or more linked procedures or activities that collectively realize a surgical objective within the context of an organizational structure defining functional roles and relationships*. Furthermore, a surgical process model (SPM) is *a simplified pattern of a surgical process that reflects a predefined subset of interest of the SP in a formal or semi-formal representation*.

The structure of SPMs consists of activities that represent the work steps of a surgical procedure. Activities have multiple perspectives: space, time, action, organization, and instrument [Neumuth et al. 2009b]. Each perspective represents a different aspect of the surgical process described. Put together, the activities make up the SPM.

Example analysis

Procedure description

For the presented study, we chose to consider lumbar discectomies as example application, as they are one of the most frequently performed surgical interventions in Germany: in 2007 144,100 such surgeries were performed [German Statistical Federal Office 2007]. The goal of this surgical procedure is the removal of herniated intervertebral discs and root nerve decompression with minimal invasion. The intervention is performed in close proximity to the spinal nerves in the loin area.

For our analysis, the typical course of the intervention was subdivided into three interventional phases. Phase I, *approach to the disc*, comprised all activities from the beginning of the procedure until the first dissection of the intervertebral disc. Phase II, *discectomy*, lasted from the first to the last dissection of the disc, and all activities by the surgeon after the last dissection were allocated to Phase III, *closure* (cp. Table 5.1.I).

The conventional strategy is distinct in that the course of action is non-linear; therefore, some of the work steps are iterative. This allows the surgeon to preserve minimal invasiveness, and conserve healthy vertebra material. The practice is to ablate as much bone as needed, but as little as possible. The surgeon thus constantly switches between steps related to bone ablation and supporting steps, such as, for instance, the repositioning of hooks or rinsing. The instruments for the former are, according to surgical necessity and personal preference, a punch, a trephine, or mallet and chisel. The goal of this work is to improve the conventional strategy with the help of a surgical assist system (SAS).

Table 5.1.I: Phase definitions.

Interventional phase	Definition
approach to the disc	activities performed in order to access the intervertebral disc
discectomy	activities occurring from the end of the previous phase to the end of removal of the intervertebral disc or prolapsed parts
closure	activities performed from the end of discectomy until the completion of suturing

System mission

The SAS to be examined is based on the principle of navigated control (NC) [Strauß et al. 2005a; Strauß et al. 2005b; Jank et al. 2006]. The system is intended to support the surgeon; this is accomplished by presenting an overlay of a previously segmented safe workspace with fluoroscopic images acquired by a C-arm during the intervention. All information, the segmented workspace and the acquired intraoperative position of the navigated mill, are collected to a single volume, registered to the patient, and used to either control the revolution speed of the mill or to stop it when the boundary of the safe workspace is violated. The work steps performed with the navigated milling system are intended to replace the manual bone removal work steps of the conventional intervention procedure.

There are numerous advantages of NC. From the surgeon's point of view, this support increases patient safety and ensures the invulnerability of sensitive areas and structures at risk. From the technical point of view, it ensures the spatial accuracy of the bone removal. As the volume of the bone material to be ablated is known in advance, the duration of the milling, and therefore of the overall invasion of the intervention, might be minimized. Moreover, the surgeon is spared repetitive loops and iterations and is subject to an improved ergonomic experience. From an administrative point of view, the shortened duration of the intervention is advantageous.

Patients, data acquisition and post-processing

For this analysis, data was acquired from 43 lumbar discectomies. The patients were chosen with regard to their diagnosis, either M51.1 (23 patients) or M51.2 (20 patients), according to the International Classification of diseases (ICD 10, [World Health Organization 2008]). The patient sample consisted of 23 females with a mean age of 50.4+/- 13.0 years, and 20 males with a mean age of 43.8+/- 13.2 years.

Specially-trained observers with medical background did a live recording of the operations. The work steps were timed and considered to be activities for inclusion in the SPM for this particular type of intervention. The observers were supported by the surgical workflow editor, a specialized, knowledge-based piece of software that is useful for modeling SPMs [Neumuth et al. 2006a; Neumuth et al. 2006b]. The editor was adjusted to apply to discectomies by adding preconfigured lists of surgical instruments, actions, and anatomical or pathological structures [Neumuth et al.

2009b]. It was the task of the observer to sample items into activity descriptions with detailed-to-the-second information according to the current situation. The surgical workflow editor was run on tablet-PCs and manipulated via touch screen. Thus, consistent data recording was ensured.

The output of the surgical workflow editor provided the initial data for post-processing analysis; this data was enhanced with information from the Hospital Information System to include patient data and cut-suture-times. The observation data was saved as a log in XML-format and transferred to a database (Postgres 8.3, [PostgreSQL Global Development Group 2009]) for further processing.

In order to obtain a consistent, uniform appraisal, the data was aggregated. This step was needed to conduct the requirements analysis of activities related to bone ablation. For the aggregation, the detailed information of the high granularity level of the recordings was reduced according to the correlation represented in Table 5.1.II. All activities in each of the SPMs were reviewed and associated to one of the activity groups. The analysis and evaluation of the results was carried out using SPSS [SPSS Inc. 2008].

Table 5.1.II: Aggregation strategy.

activity group	Definitions and Examples
access	surgical activities related to gaining access to the intervertebral disc (e.g. dissection of fascia or ligaments with scissors), not including bone ablation activities
support	activities that assist the surgeon, such as holding tissue with retractors, coagulation, or suctioning
use	setup and use of technical devices (microscope or x-ray devices), and the use of mircoscopes
bone ablation	activities that ablate bone material from the vertebra such as drilling or dissecting with punches, trephines, or mallets/chisels
discectomy	activities done to perform the dissection of the intervertebral disc
closure	activities related to closing the interventional area, such as sewing and disinfection
imaging	activities for image acquisition and review, such as making x-ray images, reviewing x-ray images or MRI images

Analysis and validation

The analysis was conducted on the basis of previously-defined parameters (cp.Table 5.1.III) for the activity definitions recorded in Table 5.1.II. These specifications can be subdivided into generic and specific parameters. Specific parameters were chosen to answer the specific objective of the requirements analysis. Therefore, these parameters are related to the task of the SAS, the bone ablation work steps. Generic parameters support the analysis by setting the context of the procedure or the interventional phases or activity groups in general.

Table 5.1.III: Parameter definitions.

Parameter type	Parameter	Definition
generic parameter for interventional phases or activity groups	cut-suture time	duration from the beginning of the first cut to the end of the last suture during the procedure
	phase duration	duration from the beginning of the first activity to the end of the last activity of the respective interventional phase according to the phase definitions in Table 5.1.I
	cumulated duration	summed durations of all activities within an activity group
	number of activities	number of activities in the aggregation group
specific parameter for deriving requirements	total bone ablation interval	interval from the beginning of the first bone ablation activity to the end of the last one during an intervention
	cumulated duration of bone ablation activities	summed durations of all activities within the bone ablation activity group of the resp. interventional phase
	number of bone ablation activities	number of activities in the bone ablation activity group of the resp. interventional phase

To answer the specific question of deriving the time limit of the SAS within the requirements analysis, an appropriate level of granularity is desired. According to the presented methodology of very fine grained observations, all activities in the SPMs need to be reviewed to decrease their granularity level. An example might be the handling of the surgeons' preference for using punches or trephines for bone ablation. For answering the study question, both activity options have been aggregated into the activity group "bone ablation", because for answering the study objective it is just worth to know if time was spent on bone ablation or not, not the exact instrument. Indeed, temporal activity evolution was preserved by this aggregation, because the activities were just renamed and not changed by their occurrence in time.

Generic parameters were parameters of interest in the context of the intervention, such as cut-suture-times and interventional phase duration (for each of the three phases: *approach to disc*, *discectomy*, *closure*). The cut-suture time indicates the time span from the very first cut until the end of the intervention. The durations of the individual phases were defined by the recorded start and stop times for a defined set of activities, as shown in Table 5.1.I. A cumulated duration was calculated for each of the aggregated activities by summing up its execution times. Likewise, the total number of occurrences of these activities was determined.

The parameter *total bone ablation* interval was raised as specific parameter indicating the duration of time over which bone ablation activities were performed. It begins upon commencement of the first bone ablation activity and ends with the conclusion of the last one. This interval was calculated across the interventional phases.

The validity of the data was checked by calculating the standard error of the mean and the 95 %-confidence intervals.

Results

The SPMs from the lumbar discectomies were analyzed in depth. The mean cut-suture time for the 43 cases was 01:27:21±00:27:24 (cp. Table 5.1.IV). The durations of the three interventional phases were 00:52:54±00:23:26 for the first, *approach to the disc*, 00:29:57±00:20:15 for the second, *discectomy*, and 00:16:26±00:08:10 for the third, *closure*. Thus, nearly half of the total intervention time was spent preparing access to the intervertebral disc: 53.1 %±12.9 %. The second most consuming activity group was the dissection of the intervertebral disc (29.2 %±11.6 %), and 17.7 %±8.1 % of the time was spent performing closure (cp. Figure 5.1.1).

Access and bone ablation activities were mainly performed in the first interventional phase *approach to the disc* (cp. Table 5.1.V). Discectomy and closure activities were the dominant activities in their respective phases. Supporting and equipment installation activities were prominent in every interventional phase.

The results of the specific parameters indicate requirements for assessing the feasibility of the SAS. Amongst others parameters, such as cumulated duration of bone ablation, activities are derived directly from the SPM. They are used to estimate technical parameters such as the upper limit of bone ablation time that shall be meet by the SAS. Results for the other activities might also be used for future feasibility estimations.

Table 5.1.IV: Results of intervention-related parameter.

[hh:mm:ss]	**Mean±SD**	**Standard Error of the Mean**	**95%-Confidence Interval for the Mean**
cut-suture-time	01:27:21±00:27:24	00:04:10	[01:18:55, 01:35:47]
phase duration *approach to the disc*	00:52:54±00:23:26	00:03:34	[00:45:41, 01:00:07]
phase duration *discectomy*	00:29:57±00:20:15	00:03:05	[00:23:45, 00:36:13]
phase duration *closure*	00:16:26±00:08:10	00:01:14	[00:13:50, 00:18:50]

Bone ablation activities had total interval durations of 00:40:35±00:22:57, CI [00:33:31, 00:47:39], across all interventional phases. Most of the time spent performing bone ablation activities occurred in the *approach to the disc* phase (0:13:36±0:08:35, CI [00:10:57, 00:16:14], cp. Table 5.1.V), but some repetitive bone ablation activities occurred in the *discectomy* phase (0:02:36±0:05:07, CI [00:01:01, 00:04:11]) as well.

Bone ablation work steps were performed in the *approach to the disc* phase with an average of 15.0±9.7 times, CI [12.0, 18.0], compared to 3.2±4.2 times, CI [1.9, 4.5], in the *discectomy* phase. The total cumulated activity durations of bone ablation across all interventional phases was 00:16:13±00:09:24 (CI: [14.6, 21.3]). The Q-Q plot of the cumulative duration of bone ablation for all three interventional phases that shows an approximate normality of the parameter is presented in Figure 5.1.2.

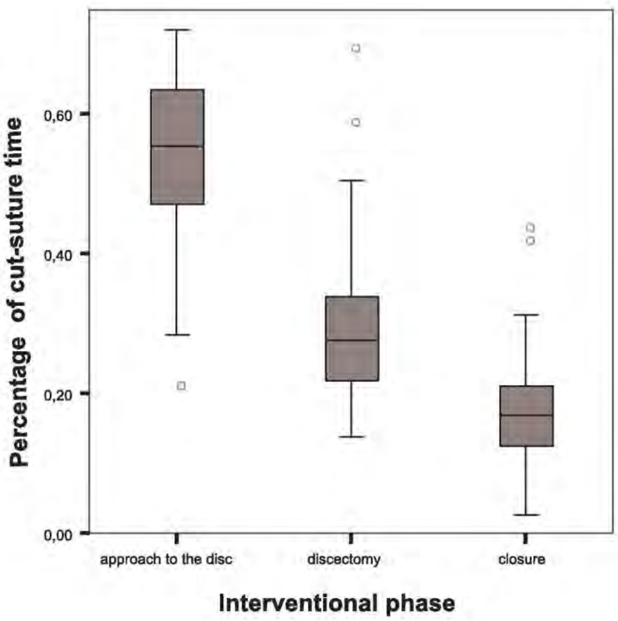


Figure 5.1.1: Percentages of total cut-suture time spent in each of the interventional phases.

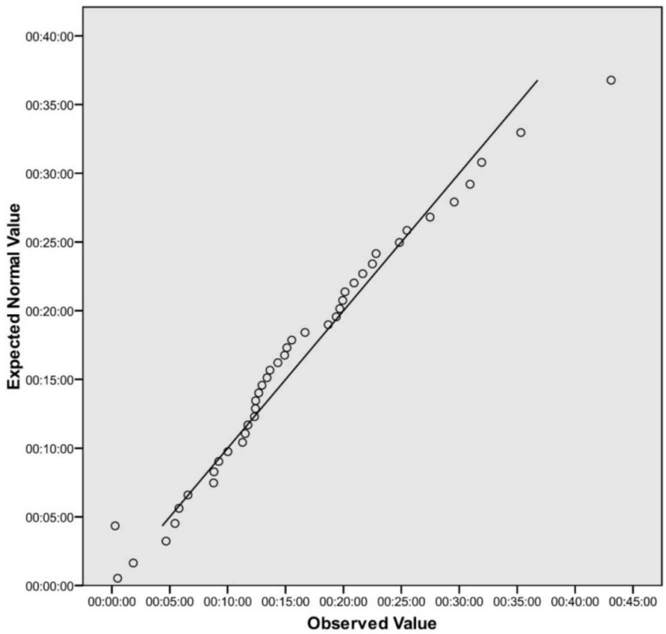


Figure 5.1.2: Normalized Q-Q plot of cumulated bone ablation durations.

Table 5.1.V: Aggregated activities: number of occurrences and cumulated durations.

		n			cumulated duration		
	aggregated activity	**Mean±Sd**	**StdError of the Mean**	**95% Confidence Interval for the Mean**	**Mean±Sd**	**StdError of the Mean**	**95% Confidence Interval for the Mean**
approach to the disc	Access	21.1±10.9	1.7	[17.8, 24.5]	0:11:26±0:07:14	00:01:06	[00:09:13,00:13:41
	bone ablation	15.0±9.7	1.5	[12.0, 17.9]	0:13:36 ±0:08:35	00:01:19	[00:10:58,00:16:15]
	Closure	0.1±0.3	0.0	[0.0, 0.2]	0:00:02±0:00:10	00:00:02	[-, 00:00:06]
	Discectomy	0.1±0.3	0.1	[0.0, 0.2]	0:00:07±0:00:49	00:00:08	[-, 00:00:23]
	Imaging	2.0±1.9	0.3	[1.4, 2.5]	0:02:12±0:01:46	00:00:16	[00:01:40, 00:02:46]
	Use	6.0±4.5	0.7	[4.6, 7.4]	0:21:27±0:18:47	00:02:52	[00:15:39, 00:27:16]
	Support	33.9±23.8	3.6	[26.5, 41.2]	0:44:06±0:25:20	00:03:52	[00:36:19, 00:51:55]
discectomy	Access	4.1±3.6	0.5	[3.0, 5.2]	0:01:57±0:02:35	00:00:24	[00:01:10, 00:02:45]
	bone ablation	3.2±4.2	1.7	[17.8, 24.5]	0:02:36 ±0:05:07	00:00:47	[00:01:02, 00:04:11]
	Closure	2.8±18	2.7	[-2.7, 8.3]	0:05:53±0:38:30	00:05:52	[-, 00:17:45]
	Discectomy	16.9±17.9	2.7	[11.4, 22.4]	0:12:49±0:11:17	00:01:43	[00:09:21, 00:16:18]
	imaging	0.1±0.4	0.1	[0.0, 0.2]	0:00:21±0:01:16	00:00:12	[-, 00:00:45]
	use	3.7±4.3	0.7	[2.4, 5.0]	0:15:54±0:15:06	00:02:18	[00:11:18, 00:20:35]
	support	23.9±29.9	4.6	[14.7, 33.1]	0:33:05±0:35:11	00:05:22	[00:22:24, 00:44:03
closure	access	0.3±1.1	0.2	[0.0, 0.7]	0:00:08±0:00:20	00:00:03	[00:00:02, 00:00:15]
	bone ablation	0.0±0.2	0.0	[0.0, 0.1]	0:00:00 ±0:00:00	00:00:00	-
	closure	9.6±7.2	1.1	[7.4, 11.8]	0:17:16±0:14:02	00:02:08	[00:12:57, 00:21:36]
	discectomy	0.1±0.3	0.0	[0.0, 0.1]	0:00:02±0:00:10	00:00:02	[- , 00:00:05]
	imaging	0.0±0.2	0.0	[0.0, 0.1]	0:00:09±0:00:44	00:00:07	[-, 00:00:24]
	use	1.6±2.6	0.4	[0.8, 2.5]	0:03:59±0:08:22	00:01:16	[00:01:25, 00:06:34]
	support	9.5±8.2	1.2	[6.9, 11.9]	0:08:11±0:12:36	00:01:55	[00:04:11, 00:11:57]

Discussion

We attempted to derive the minimal requirements of a needs assessment for a surgical assist system that would indicate the usefulness of the surgical assist system compared to a conventional method. It was shown that SPM-based analyses can be an asset for requirement engineering.

In this study, it was found that the technical requirements for an alternative approach to routine steps involved in lumbar discectomies -- the navigated control milling system -- can be derived based on the investigation of the distribution and number of bone ablation activities. The analysis revealed the mean amount of time spent performing bone ablation activities using the conventional approach. This duration, 00:16:13 in total for both, the *approach to the disc* and the *discectomy* interventional phases, gives an upper limit for the SAS that should not be exceeded. Also, the decrease in the number of repetitious bone removal activities can be used as an indirect indicator of the ergonomic improvement and experience of the surgeon's work.

The analysis of the SPM in this work focused on bone ablation activities in order to derive technical requirements for application of the system to lumbar discectomies. This was a reasonable undertaking, since the mission of the system is the removal of vertebral tissue. Predictions of the overall decrease in time or in the number of activities the system can support for other aspects of the discectomy process cannot be derived from the current data. Furthermore, certain outcomes resulting from application of the system, such as workspace segmentation or system setup time, could not be predicted. However, these might be evaluated in future studies by comparing SPMs from cases where the system was applied to the patient to the current data sets.

Furthermore, the system needs to undergo a more complete evaluation of other parameters such as ergonomics, safety, or complexity.

The validity of this method of data acquisition for SPMs was investigated in depth in a previous study for a similar application [Neumuth et al. 2009b]. It has been shown, that observers supported by the surgical workflow editor record SPMs robustly and accurately with a content accuracy of 92% and a margin of temporal error of less than 2 s.

The issue of deriving needs assessments from clinical data is of great relevance from the clinical, the administrative, and the manufacturing point of view; this work describes one such method of deriving this assessment. With the SPMs clinicians gain an instrument that enables them to validate hypotheses for the improvement of surgical interventions. The administrative departments of hospitals gain a method to evaluate their capital expenditure decisions before the actual expenditure takes place. Additionally, surgical technology manufacturers are provided with an apparatus that anticipates the demands for the development of new systems before the first investments in development have taken place.

The use of SPMs is essential for the needs assessment methods presented here. The advantages of the application of surgical process models are manifold. Amongst others, the high accuracy of the delineation of a surgical intervention allows retrospective analysis of surgeries within specific populations, yielding exact and valid assertions. Requirements for the systems can be extrapolated directly from the clinical routine. This makes it possible to specify development and employment decisions in a very precise way. For this and other reasons, surgical process models are a substantial, relevant, and supplemental basis for decision-making.

Conclusions

It was shown that the minimal requirements needed to ensure that a future system is more advantageous than a conventional or already existent method or system, as well as the method by which these requirements can be obtained from routine clinical data, can be derived through the use of surgical process models (SPMs). This analysis was performed for the identification of temporal requirements for a surgical assist system designed to aid bone ablation.

The presented methods can also be used for a systematic analysis of technical needs. Based on these analyses, manufacturers can derive the parameters for their systems that are indicative of success. Additionally, hospital administrations might use request that manufacturers demonstrate fulfillment of such requirements in order to prove that their system is superior to the conventional strategy and therefore worth the investment.

With regard to the overall goals of improving and planning novel systems based on SPMs, the method presented here contributes to the creation of better and more appropriate support systems for the surgeon by integrating the actual surgical workflow into development considerations for the system.

Acknowledgements

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the Unternehmen Region, grant numbers 03 ZIK 031 and 03 ZIK 032.

5.2 Evaluation of a surgical assist system in pediatric surgery

Title

Impact quantification of the daVinci telemanipulator system on surgical workflow using resource impact profiles

Authors

Thomas Neumuth[1], Alexandra Krauss[2], Jürgen Meixensberger[1,3], Oliver J. Muensterer[2,4]

Affiliations

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] University Medical Center Leipzig, Department of Pediatric Surgery, Leipzig, Germany

[3] University Medical Center Leipzig, Department of Neurosurgery, Leipzig, Germany

[4] University of Alabama at Birmingham, Division of Pediatric Surgery, Children's Hospital of Alabama, Birmingham AL, USA

Citation

Neumuth T, Krauss A, Meixensberger J, Muensterer O. Impact quantification of the DaVinci Telemanipulator system on the surgical workflow using resource impact profiles. International Journal of Medical Robotics. 2011; 7(2):156-64.

Keywords

Workflow; Process Assessment (Health Care); Health Care Evaluation Mechanisms; Surgical Equipment; Robotics; Fundoplication

Abstract

Background: It has yet to be determined whether surgical assist systems benefit surgical workflow. This question should be answered qualitatively and quantitatively and must be supported by evidence gathered from structured and rigorous analyses.

Methods: We present a method to quantify the benefits of the daVinci telemanipulator system on surgical workflow. Based on the modeling of surgical processes, we generated resource impact profiles (RIPs). RIPs are statistical mean intervention courses for a sample of surgical process models that were performed using a specific surgical assist system as a resource. A total of 12 laparoscopic and 12 telemanipulator-supported Nissen fundoplications were modeled and analyzed to quantify the impact of the surgical assist system.

Results: Few statistically significant benefits of the system on surgical workflow were found. It was found that the daVinci system is not superior to the conventional laparoscopic strategy if the surgeon follows the same workflow.

Conclusions: RIPs are a valuable method to estimate the impact of a surgical assist system on the surgical workflow. For the investigated use case, changes in workflow may be necessary to fully benefit from the advantages of using a telemanipulator in Nissen fundoplications. Conversely, the telemanipulator may only reach its full potential in more complex operations.

Introduction

Robots and telemanipulator systems play an increasing role as surgical assist systems (SAS). The goal of these systems is to ameliorate a surgeon's tasks by means of increasing surgical accuracy, ensuring minimal invasiveness, improving ergonomic handling and enhancing surgical effectiveness in general [Sandberg et al. 2003; Lemke and Vannier 2006; Cleary and Kinsella 2005; Deinhardt 2003]. The application of SAS makes sense if at least one of these points is beneficially satisfied and the others are not influenced in a negative way. The question of whether such a system is advantageous to the surgeon needs to be answered qualitatively and quantitatively and needs to be supported by evidence gathered from structured and rigorous analyses.

Usually, the assessment of surgical assist systems is performed by focusing on extrinsic factors of the surgical workflow, such as cut-suture-times [Schuster et al. 2007; Archer and Macario 2006], or with the help of single, very specific parameters, such as conversion rates [Hartmann et al. 2008]. However, a highly detailed and accurate contextual and temporal description of the intra-operative procedure of the surgical process would be more precise, sensible, and could grant a more variable perspective on the topic. At present, such direct measurements of surgical process models for the evaluation of surgical assist systems are not available. In recent studies, various methods of achieving a detailed description of surgical processes have been proposed [Neumuth et al. 2006a; Neumuth et al. 2009b]. This method, which is used to develop surgical process models (SPMs), [Neumuth et al. 2011b], can be employed to measure surgical procedures down to the level of seconds.

This work introduces resource impact profiles (RIPs), a method for computing and analyzing SPMs to check for benefits of SAS, and reports the results of a pilot study that had the objective of checking if and how RIPs can contribute to the evaluation of surgical assist systems. The feasibility study was performed by modeling and analyzing the surgical process models of 24 Nissen fundoplications, 12 of which were performed conventionally, i.e., laparoscopically, while the other 12 were performed with the help of a surgical assist system, the daVinci telemanipulator system. The feasibility study demonstrates how to explicitly quantify the impact of the daVinci system on the surgical workflow using the method of analyzing statistical 'mean' procedures as RIPs.

The use of the daVinci surgical system has been investigated in a number of surgical disciplines including urology [Thiel and Winfield 2008], orthopedic surgery [Bargar 2007], cardiovascular surgery [Mohr et al. 2006; Rodríguez et al. 2006; Chitwood Jr. et al. 2003], and gynecology [Bocca et al. 2007]. The daVinci surgical system has been compared with the conventional surgical strategy for prostatectomies [Trabulsi et al. 2008; Hu et al. 2006; Le and Gettman 2006; Ploussard et al. 2009], gastrectomies [Song et al. 2009], cholecystectomies [Breitenstein et al. 2008], cystectomies [Wang et al. 2008], low anterior resection in patients suffering from rectal cancer [Baik et al. 2009], diverticulectomies [Myer and Wagner 2007], adrenalectomies [Morino et al. 2004], nephrectomies [Nazemi et al. 2006; Kaul and Menon 2007], mitral valve replacements and sternotomies [Folliguet et al. 2006], and splenectomies [Bodner et al. 2005]. In the context of pediatric surgery, the system has been evaluated for pyeloplastic surgery [Yee et al. 2006; Bernie et al. 2005; Link et al. 2006]. However, these studies focused on parameters such as total intervention time, hospital stay duration, complication rate, blood loss and transfusion requirement, conversion rate and monetary intervention costs.

More specifically, intervention times were assessed for several urological interventions [Park et al. 2008], transoral robot-assisted surgeries [Weinstein et al. 2009], bypass interventions [Mishra et al. 2006; Bonatti et al. 2008], hysterectomies [Beste et al. 2005; Bell et al. 2009; Fanning et al. 2008], thoracoscopic surgeries [Bodner et al. 2005; Braumann et al. 2008], gynecological surgeries [Pitter et al. 2008; Nezhat et al. 2009], renal surgeries [Hubert and Siemer 2008], and mitral valve repairs [Rodríguez et al. 2006]. Finally, related parameters such as performance, security, efficiency, teachability, cost efficiency, set-up time, conversion rate, intra-operative complications, and perioperative morbidity and mortality rates have been investigated regarding the daVinci system [Hartmann et al. 2008; Sim et al. 2006; Villavicencio Mavrich et al. 2009; Patel et al. 2007; Artibani et al. 2008; John et al. 2007]. However, none of these studies concentrated on the impact of the system on the surgical process.

Other authors have modeled surgical processes in the context of medical engineering for several purposes, such as the automatic identification of interventional phases [Ahmadi et al. 2006], control of surgical robots [Münchenberg et al. 2001a], and instrument assessments [Mehta et al. 2002]. Clinical work has also focused on surgical processes for reengineering [Casaletto and Rajaratnam 2004], assessing human reliability [Malik et al. 2003], comparing substitutive surgical strategies [den Boer et al. 1999], and analyzing requirements for surgical assist systems [Strauß et al. 2006a]. However, none of these approaches dealt with the generation of a statistical mean treatment model for the SAS, or they provided only very low-resolution data for surgical phases in general.

This article is technically motivated, and only a limited clinical interpretation of the data is provided. The presented work concentrates on the evaluation of the SAS from the point of view of medical technology and on the introduction of SPM-based methods.

In the Material and Methods section, an overview of the surgical aspects of Nissen fundoplications, the conventional and the robot-assisted strategies, as well as the surgical cases will be given. Subsequently, the methods and the execution of the measurements for the generation of the surgical process model as resource impact profiles will be outlined jointly with the derivation of the analyzed parameters. The Results section will then present the outcomes of the evaluation, which will then be interpreted in the Discussion section.

Methods

Surgical processes: context and technology

The procedure type described here is the Nissen fundoplication [Nissen 1956; Nissen 1961]. This treatment is applied to patients suffering from gastroesophageal reflux disease or paraesophageal hiatus hernia. The surgical goal is the 360° wrapping or placating and subsequent fixation of the upper part of the gastric fundus around the lower end of the esophagus and the gastroesophageal junction. This is performed with the aim of suppressing the pathological reflux.

The typical conventional procedure consists of the laparoscopic performance of various surgical phases: the preparation, dissection, reconstruction (including hiatal repair and the actual fundoplication), and the conclusion phases. In the preparation phase of the intervention, the patient is positioned on his/her back on the operating table, immobilized, and trocars are inserted at the appropriate sites. In the next phase, the dissection, the gastrohepatic ligament and the diaphragmatic crurae are dissected. Furthermore, the esophagus is mobilized by a dissection of the esophagophrenic ligaments and the retroesophageal area. In the reconstruction phase, the esophageal hiatus is approximated with the help of two Collar stitches, and the fundus is wrapped around the gastroesophageal junction, suturing it in place with three interrupted stitches. The final phase covers the removal of the instruments and the closure of the abdominal incisions. The different phases of the intervention, as well as a short definition concerning their start and their end as used throughout the study, are described in Table 5.2.I.

This study was performed with the goal of evaluating whether the daVinci telemanipulator system is beneficial to the surgical process. The daVinci system is a four-armed, remote-controlled surgical assist system. It is operated by the surgeon using a spatially separated panel. Here, the panel can be regarded as a master system, while the telemanipulator system operating on the patient is the slave system.

The main goal and the expected advantages of this system are a more detailed and more precise execution of the surgical tasks owing to the multidimensional scaling of the movements and the heightened precision due to the filtering of the tremor.

All procedures were performed on infantile pigs. The data were collected at the University Hospital Leipzig. Of the 24 conducted interventions, 12 were performed conventionally and the other 12 were performed under non-sterile conditions with the help of the daVinci system. All of the interventions were performed by the same surgeon, who had significant experience in the performance of pediatric laparoscopic Nissen fundoplications. Although the surgeon had no intraoperative experience with the telemanipulator system, he had received basic training on the system and had acquired some dry-run experience in system usage.

Table 5.2.I: Surgical phases of Nissen fundoplications and phase definitions used in the study.

Intervention	Phase	Subphase
Nissen fundoplications From the beginning of the preparation of the laparoscopic tower until the end of the intervention, excluding the gastrostomic phase	Preparation From the beginning of the preparation of the laparoscopic tower until the end of the preparation of anesthesia, of the used technology and systems, and of the pig; including preparation time to dock the robot	
	Dissection From end of the preparation phase until the start of the introduction of the needle holder and the needle in the abdomen	
	Reconstruction From the beginning of the introduction of the needle holder and needle in the abdomen until the end of removal of the needle holder or Maryland-Dissector from the abdomen	Reconstruction hiatus suture Repairing the hiatus by a posterior hiatal suture
		Reconstruction collar stitches Placement of two Collar stitches between the esophagus and diaphragm
		Wrap creation Finding the passage point and retroesophageal pulling-through of the fundus
		Reconstruction fundoplication suture Placement of three fundus sutures for the fixation of the fundus cuff
	Conclusion From the end of the removal of the needle holder resp. Maryland-Dissector until the end of closure of all surgical wounds	

Process model generation and analysis

The impact of the daVinci system on the surgical workflow was investigated by the comparison of the differences between two statistical 'mean' treatment models. In general, surgical process models depict surgical processes in information systems in a formal or semiformal way [Neumuth et al. 2009b] and thus allow for a detailed behavioral analysis [Neumuth et al. 2009c]. Models of individual surgical cases are referred to as individual surgical process models (iSPMs), representing the evolution of a surgical case down to a level of seconds. Multiple superimposed iSPMs are referred to as generic surgical process models (gSPMs), which represent a statistically ‘mean’ surgical process for a population of iSPMs [Neumuth et al. 2011b]. The assessments of the changes within the surgical workflow are investigated based on resource impact profiles. These resource impact profiles were both gSPMs, one for each of the laparoscopic and telemanipulator-based samples.

SPMs consist of flow objects, the most important ones being the activities that depict surgical work steps. Each activity contains various perspectives, each of which represents a different point of view regarding the surgical process: the organizational perspective describes WHO performs a work step, the functional perspective characterizes WHAT is being performed, the operational perspective denotes the surgical instruments WITH WHICH a work step is performed, and the spatial perspective indicates WHERE on the patient’s body a work step is performed. The logical order of the single work steps and the temporal array are determined by the behavioral perspective. In addition to the work steps, an SPM contains information concerning the different phases of the interventions and the temporal allocation of the activities to their respective phases.

The data acquisition for the SPMs was performed by a comprehensively trained medical student. All recordings were performed based on video observations of the surgical procedures, using dedicated software to generate the iSPMs [Neumuth et al. 2009b; Neumuth et al. 2006a].

The observation support software, the surgical workflow editor (see Figure 5.2.1), is special, configurable software that has been developed to support the observer in his/her task of collecting data from surgical processes. It proposes concepts for the different perspectives and requests appropriate user selections, for instance concerning the surgical instrument used, or the current acting person. The observer chooses the items best representing the current situation within the surgical process and explicitly models the iSPM.

After the completion of the observation, the data were saved as XML files and transferred to a Postgres database [PostgreSQL Global Development Group 2009] for post-processing. The iSPMs were allocated to the laparoscopic- or to the telemanipulator-based strategy and subsequently segmented in time according to the observed interventional phases. Finally, a gSPM was computed as a statistical mean procedure for each of the surgical phases.

For the analysis of the workflows with and without the daVinci system, various parameters were defined. The application of the assist system was evaluated, and differences between the two models were quantified with the help of these parameters. By analyzing resource impact profiles, different parameters, such as execution time and duration, number of overall work steps, and probability of sequence, were compared and statistically analyzed.

We adjusted the parameter of the telemanipulator sample to consider the learning curve of the surgeon with the telemanipulator system. A regression analysis was performed for each parameter for the 12 samples. Subsequently, the regression line and the measurement values were adjusted to a zero slope. The computation of the differences between mean strategies was conducted with the help of the Mann-Whitney-U-Test concerning significance. The differences were regarded as statistically significant if the exact bilateral significance was higher than $\alpha=0.05$.

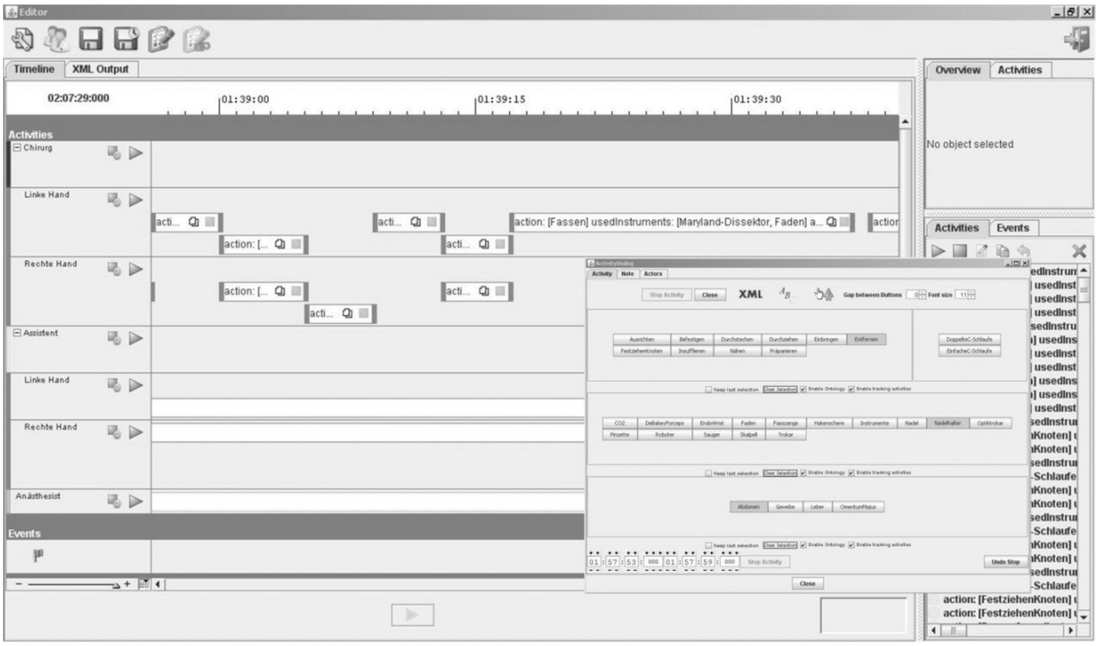


Figure 5.2.1: Screenshot of the surgical workflow editor with the activity window (bottom right).

Results

The mean durations of procedures were 01:41:05±00:21:56 for laparoscopic procedures and 01:36:47±00:17:08 for telemanipulator-supported procedures (p=0.05, cp. Table 5.2.II). The durations of the surgical phases (Preparation, Dissection, Reconstruction and Conclusion*)* showed no significant differences between laparoscopic- and telemanipulator-based procedures. The sub-phases of Reconstruction also showed no statistically significant differences, with the sole exception of the *Wrap creation* sub-phase (p=0.014).

Table 5.2.II: Procedure duration and duration of surgical phases and sub-phases for laparoscopic (LAP) and telemanipulator-based (TEL) procedures (exact significance for Mann-Whitney U and α=0.05).

		LAP [avg±sdev]	**TEL [avg±sdev]**	**p-value**
	Total intervention time	01:41:05±00:21:56	01:36:47±00:17:08	p=0.045
Phases	Preparation	00:42:31±00:16:56	00:38:44±00:07:34	p>0.05
	Dissection	00:18:49±00:10:44	00:13:18±00:06:40	p>0.05
	Reconstruction	00:33:25±00:10:08	00:37:32±00:07:44	p>0.05
	Conclusion	00:06:19±00:01:49	00:07:12±00:02:22	p>0.05
Subphases	Reconstruction hiatus suture	00:09:15±00:07:22	00:06:54±00:01:50	p>0.05
	Reconstruction collar stitches	00:08:25±00:02:13	00:10:49±00:03:32	p>0.05
	Wrap creation	00:03:00±00:00:44	00:04:45±00:05:10	p=0.014
	Reconstruction Fundoplication suture	00:12:43±00:02:31	00:15:32±00:03:13	p>0.05

For a detailed assessment of the impact of the system on the surgical workflow, the sub-phase *Reconstruction Fundoplication suture* with the longest duration was investigated. Exemplary results are reported for the surgeon's right-hand activities. Figure 5.2.2 shows the merged resource impact profiles for the activities of the surgeon and for both strategies as a process model. Activities that were not comparable between the resource impact profiles were combined to an artificial "supporting activity". This comprised the introduction of instruments such as hooks or the changing of instruments that was performed by the surgeon in the laparoscopic-based sample and by the assistant in the telemanipulator-based sample. For visual lucidity, all activities with an average count of less than 0.33 times per procedure were filtered from the process model.

Several surgical activities showed a significantly increased number of activity performances for the telemanipulator-based procedures (see Table 5.2.III), such as *Grasping needle and thread with forceps* (p<0.001) for right-hand activities. Additionally, some of the activities had longer total performance durations for the telemanipulator-based procedures than for laparoscopic procedures, such as *Pulling of needle and thread with forceps through tissue* (p=0.003) or *Grasping needle and thread with forceps* (p<0.001), and some had shorter total performance durations, e.g., *Construction of single C-loop with forceps, needle and thread* (p=0.02) or *Construction of double C-loop with forceps, needle and thread* (p=0.02). Most of the activity performances and durations were not significantly different.

Significant differences in the transition probabilities between activities were found (see Table 5.2.IV). For instance, the performance of the activity *Aligning forceps, needle and thread in the abdomen* after *Grasping needle and thread with forceps* with the right hand was performed significantly more often ($p<0.001$) in telemanipulator-based procedures and thus had a higher follow-up probability ($p<0.001$). In contrast to this, the activity *Construction of double C-loop with forceps, needle and thread* followed the activity *Pulling of needle and thread with forceps through tissue* significantly more often ($p=0.02$) in laparoscopic procedures than in telemanipulator-based procedures and thus had a lower follow-up probability ($p=0.001$). However, most of the transition probabilities were not statistically significantly different.

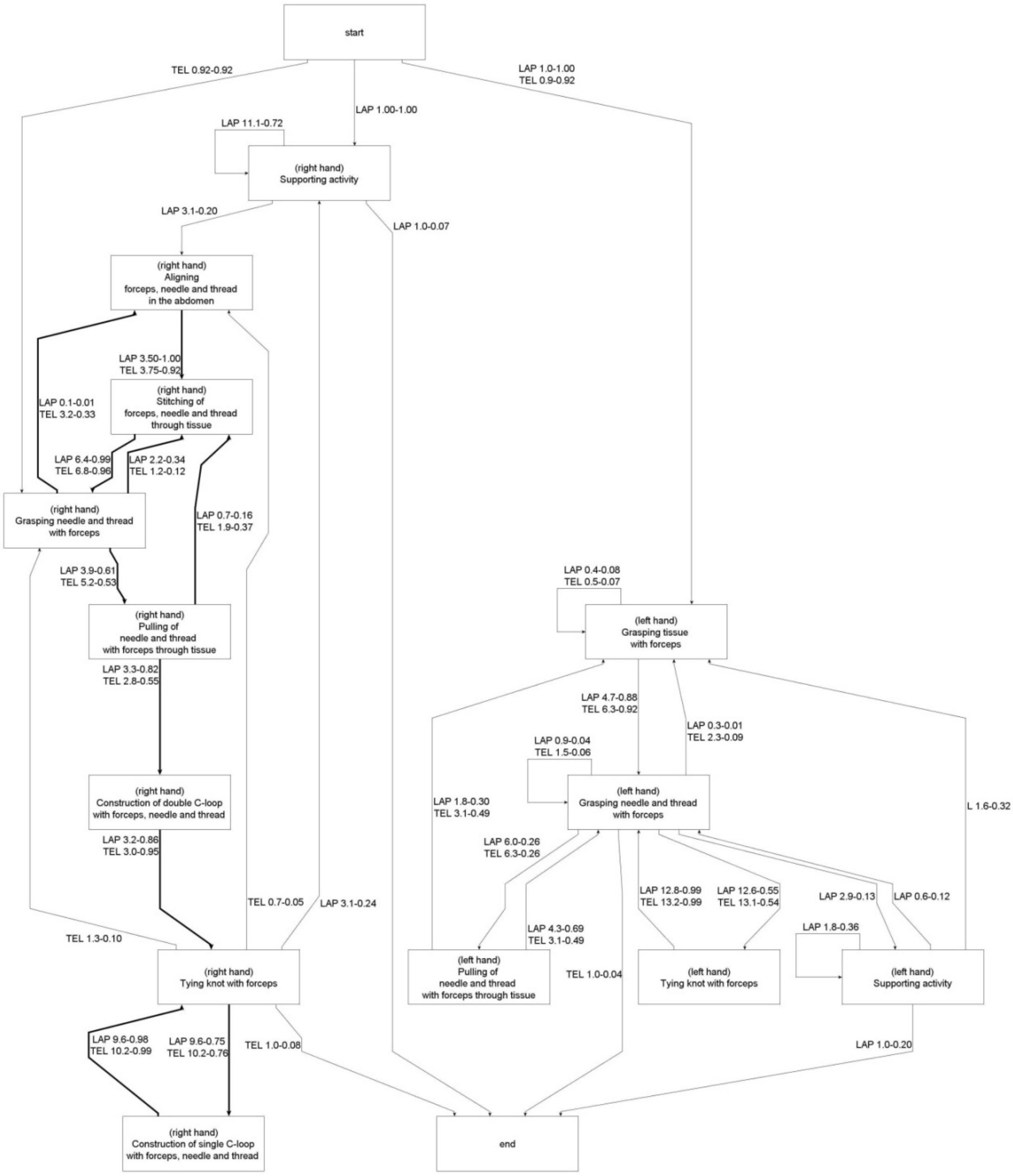


Figure 5.2.2: Merged resource impact profiles (RIP) of both strategies for the sub-phase *Reconstruction fundoplication suture*: laparoscopic (LAP) and telemanipulator-based procedures (TEL) as process models. Activity sequences are labeled with activity sequence counts and follow-up probabilities. The label TEL 3.0-0.95 means, for example, that the activity sequence was performed 3.0 times on average per intervention in telemanipulator-based procedures and had a transition probability of 95%.

Table 5.2.III: Count and durations of activities for the right hand of the surgeon for the sub-phase Reconstruction fundoplication suture for laparoscopic (LAP) and telemanipulator-based (TEL) procedures.

Activity	Number of activities			Duration of activities		
	LAP [avg±sdev]	TEL [avg±sdev]	p-value	LAP [avg±sdev]	TEL [avg±sdev]	p-value
Aligning forceps, needle and thread in the abdomen	3.5 ± 0.7	4.1 ± 1.2	p>0.05	00:00:54 ± 00:00:18	00:00:46 ± 00:00:18	p>0.05
Construction of double C-loop with forceps, needle and thread	3.7 ± 1.2	3.2 ± 0.4	p>0.05	00:00:57 ± 00:00:44	00:00:23 ± 00:00:09	p=0.02
Stitching of forceps, needle and thread through tissue	6.5 ± 1.8	7.0 ± 1.5	p>0.05	00:01:43 ± 00:00:25	00:01:21 ± 00:00:26	p>0.05
Pulling of needle and thread with forceps through tissue	4.1 ± 1.0	5.2 ± 0.9	p=0.02	00:00:51 ± 00:00:24	00:01:33 ± 00:00:31	p=0.003
Construction of single C-loop with forceps, needle and thread	9.8 ± 2.1	10.3 ± 0.8	p>0.05	00:01:11 ± 00:00:21	00:00:55 ± 00:00:19	p=0.03
Grasping of needle and thread with forceps	6.4 ± 1.7	9.7 ± 1.5	p<0.001	00:00:12 ± 00:00:04	00:00:32 ± 00:00:08	p<0.001
Tying knot with forceps	12.8 ± 2.2	13.3 ± 1.1	p>0.05	00:01:42 ± 00:00:22	00:02:07 ± 00:00:24	p=0.002

Table 5.2.IV: Counts of activity sequences and follow-up probabilities between activities of the right hand of the surgeon for sub-phase Reconstruction fundoplication suture for laparoscopic and telemanipulator-based procedures (analysis for bold printed transitions).

Start Activity	Stop Activity	Count of activity sequence			Follow-up probability of activity sequence		
		LAP [avg±sdev]	TEL [avg±sdev]	p-value	LAP [avg±sdev]	TEL [avg±sdev]	p-value
Aligning forceps, needle and thread in the abdomen	Stitching of forceps, needle and thread through tissue	3.5 ± 0.7	3.8 ± 0.9	p>0.05	1.00 ± 0.00	0.93 ± 0.10	p=0.025
Construction of double C-loop with forceps, needle and thread	Tying knot with forceps	3.2 ± 0.4	3.0 ± 0.0	p>0.05	0.91 ± 0.17	0.96 ± 0.10	p>0.05
Stitching of forceps, needle and thread through tissue	Grasping needle and thread with forceps	6.4 ± 1.7	6.8 ± 1.3	p>0.05	0.99 ± 0.04	0.97 ± 0.05	p>0.05
Pulling of needle and thread with forceps through tissue	Construction of double C-loop with forceps, needle and thread	3.3 ± 0.5	2.8 ± 0.3	p=0.02	0.85 ± 0.16	0.57 ± 0.16	p=0.001
Pulling of needle and thread with forceps through tissue	Stitching of forceps, needle and thread through tissue	0.7 ± 0.9	1.9 ± 0.9	p=0.002	0.13 ± 0.16	0.35 ± 0.14	p=0.003
Grasping needle and thread with forceps	Aligning forceps, needle and thread in the abdomen	0.1 ± 0.3	3.2 ± 1.0	p<0.001	0.01 ± 0.04	0.33 ± 0.11	p<0.001
Grasping needle and thread with forceps	Construction of double C-loop with forceps, needle and thread	0.0 ± 0.0	0.2 ± 0.4	p>0.05	0.00 ± 0.00	0.02 ± 0.04	p>0.05
Grasping needle and thread with forceps	Stitching of forceps, needle and thread through tissue	2.2 ± 1.3	1.2 ± 1.3	p>0.05	0.31 ± 0.17	0.11 ± 0.11	p=0.004
Grasping needle and thread with forceps	Pulling of needle and thread with forceps through tissue	3.9 ± 1.0	5.2 ± 0.9	p=0.009	0.64 ± 0.19	0.54 ± 0.1	p>0.05
Tying knot with forceps	Construction of single C-loop with forceps, needle and thread	9.6 ± 2.1	10.2 ± 0.9	p=0.05	0.74 ± 0.04	0.76 ± 0.02	p>0.05

Discussion

The computation of surgical process models is a promising method for the assessment studies of surgical devices. This is the first study that supports the impact evaluation of a surgical assist system on the surgical process with detailed work step assessments by calculating resource impact profiles (RIPs) as statistical mean models, so-called generic surgical process models. It was shown that by comparing two resource impact profiles, the impact of the surgical assist system on the surgical workflow could be investigated and quantified in detail. The proposed methodology provides appropriate quantitative support for investment decisions regarding surgical assist systems based on Surgical Process Models. It has been shown that a suitable evaluation of the influence of the daVinci system on the surgical process can be accomplished with the help of the presented methodology using gSPMs.

To assess the impact of the daVinci system on the surgical workflow, we started from the top-most resolution level, the surgical phases, and investigated the performance repetitions and performance durations of surgical activities in depth. Additionally, we calculated transition probabilities between activities that represent the surgical workflow. The impact of the daVinci system on the surgical process was found to not be significant. Although some differences between both resource impact profiles appeared, most differences in the performance repetitions, average performance durations, and transition probabilities were found to not be significantly different between laparoscopic and telemanipulator-based procedures. This was due to the application of the same surgical strategy in both scenarios. The surgeon followed the same procedure in the telemanipulator-based procedures as in the laparoscopic procedures. This evidence is suggested by the many transition probabilities that were found to not be statistically different between both RIPs. Furthermore, the differences in the performance repetitions and durations were too small to have an impact on the total duration time alone.

The study was limited to infantile pigs. This was due to ethical reasons, as the surgeon had no previous experience with the daVinci system, and it would have been ethically objectionable for him to perform such an unknown intervention on a real human patient. Nevertheless, he had extensive experience concerning the performance of conventional and laparoscopic Nissen fundoplications. To cope with his learning curve, we applied a statistical adjustment of the telemanipulator-based measurements. Furthermore, it could be considered as a restriction that only one surgeon was recorded, but on the other hand it minimized the influence factors on surgical processes [Neumuth et al. 2009b]. The data acquisition performed by an observer based on video recordings. Due to the performance of the study in a simulated environment, the cameras could optimally record the procedure. Another evaluation study already showed that trained observers reach a very high level of accuracy in recording surgical workflows [Neumuth et al. 2009b].

Sensible enhancements for future studies would be firstly an augmentation of the number of infantile pigs used or a repetition of the study for several surgeons to strengthen the results. Secondly, a study using real patients would be conceivable if an adequate number of surgeons with sufficient experience are available. Additionally, a variation of the general surgical strategy for telemanipulator-based procedures should be clinically investigated. However, this variation of the surgical strategy would adapt the surgical strategy to surgical technology.

Conclusion

The assessment of surgical assist systems is an important topic and is of great interest for medical device manufacturers and clinical users. The presented approach that is based on generic surgical process models facilitated the calculation of resource impact profiles to estimate the impact of a surgical assist system on the surgical process. It was demonstrated that the RIP approach is well suited for the evaluation of the application of the system and its influence on the overall procedure.

By using resource impact profiles, it was shown in a feasibility study that the daVinci telemanipulator system has only limited impact on the surgical workflow for Nissen fundoplications if the surgeon follows the same strategy as that of the conventional course. Superiority of the system is expected to be found in more complex procedures, which are mainly performed in anatomical areas without a direct line of view between the trocar position and the operation area, such as esophagectomies through the diaphragmatic hiatus, or interventions performed in the pelvis including prostatectomies.

With the help of such assessments, the system developers and vendors are able to identify application strengths of their systems on the one hand, and, of course, possible weaknesses on the other hand. The method of using RIPS has its benefits in the evidence-based assessment of the system’s impact on the workflow. In comparison to existing methods of impact quantifications, such as cut-suture time, or measures of usage counts or usage times of isolated surgical activities or incidents during the interventions, it is possible to cover and to analyze the systems impact on the overall workflow. By using this method, benefits that are based on the “coherence” of surgical activities can be identified, such as shortened performance durations for one surgical work step by the system, while other work steps are extended due to the use of the system.

The usage of RIPs provides the vendor with information about procedure course variations and gives hints on the development of surgical tools to redesign the workflow. An example for the latter might be the engineering question “How needs a surgical system be designed to prevent iterations of activity X?” Furthermore, manufacturers can employ the method of SPMs for the derivation of pre-development requirements

Benefits of system changes can also be investigated and analyzed by using RIPs. Here, it is necessary to compare RIPs that were computed from samples before and after the implementation of the system change. An example would be the calculation of a value for the “rectification” of a surgical process by a system by reducing variations, while coincidently situations are eliminated, where the system is not applicable.

.

Acknowledgements

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the scope of the Unternehmen Region with the grant numbers 03 ZIK 031 and 03 ZIK 032.

5.3 Assessment of surgeon's strategies in ophthalmology

Title

Identification of surgeon-individual treatment profiles to support the provision of an optimum treatment service for cataract patients

Authors

Thomas Neumuth[1], Renate Wiedemann[2], Christian Foja[2], Petra Meier[2], Juliane Schlomberg[2], Dayana Neumuth[1], Peter Wiedemann[2]

Affiliations

[1] Universität Leipzig, Medical Faculty, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany

[2] University Medical Center Leipzig, Department of Ophthalmology, Leipzig, Germany

Citation

Neumuth T, Wiedemann R, Foja C, Meier P, Neumuth D, Wiedemann P. Identification of surgeon-individual treatment profiles to support the provision of an optimum treatment service for cataract patients. *Journal of Ocular Biology Diseases and Informatics*. 2010;3(2):73-83.

Keywords

Education; Surgical Procedures, Operative; Ophthalmology; Process Assessment (Health Care); Workflow; Surgical Process Model

Abstract

One objective of ophthalmological departments is the optimization of patient treatment services. A strategy for optimization is the identification of individual potential for advanced training of surgeons based on their daily working results. The objective of this feasibility study was the presentation and evaluation of a strategy for the computation of surgeon-individual treatment profiles (SiTP).

We observed experienced surgeons during their standard daily performance of cataract procedures in the Ophthalmological Department of the University Medical Center Leipzig, Germany. 105 cases of cataract procedures were measured as surgical process models (SPMs) with a detailed-to-the-second resolution. The procedures were performed by 3 different surgeons during their daily work. Subsequently, SiTPs were computed and analyzed from the SPMs as statistical 'mean' treatment strategies for each of the surgeons.

The feasibility study demonstrated that it is possible to identify differences in surgeon-individual treatment profiles beyond the resolution of cut-suture times. Surgeon-individual workflows, activity frequencies and average performance durations of surgical activities during cataract procedures were analyzed. Highly significant ($p<0.001$) workflow differences were found between the treatment profiles of the three surgeons. Conclusively, the generation of SiTPs is a convenient strategy to identify surgeon-individual training potentials in cataract surgery. Concrete recommendations for further education can be derived from the profiles.

Introduction

The performance of surgical procedures is a complex interaction of manual skills and the experience that surgeons gather in the course of their professional life [Ezra et al. 2010; Ayanniyi et al. 2009]. The mediation of this accumulated knowledge to junior surgeons usually ensues by means of oral information, an one to one observation of the senior surgeons by the residents with subsequent practice, or other 'knowledge stores', such as clinical guidelines [AHRQ-Agency for Health Care Research and Quality 2010a; AWMF-Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. 2010a] or dedicated literature and videos for surgical education.

It is the objective of each clinic to ensure an optimal, but also homogenous, treatment service for every patient. However, due to patient-specific characteristics, different techniques favored by different seniors, to varying preferences, and the availability of technical resources to support the surgical task, a variation of surgical procedure courses results. To achieve a homogenous success of treatment, it is an option to minimize the variability in treatment service caused by different surgical preferences. With relation to the individual skill and experience of the surgeons, it is necessary to identify the individual potential for improvement of their surgical abilities and a subsequent targeted support with the help of advanced surgical training.

The goal of the presented feasibility study was the presentation of a method and its evaluation for the computation of the individual promotion potential for surgeons in their working life by advanced training. In the course of the study the surgical processes of 105 cataract procedures that were performed by three different surgeons were measured and analyzed. The promotion potential for each of the three surgeons was investigated and identified. The objective of the feasibility study was to answer the questions: (1) "How can surgeon-individual treatment service can be assessed?", (2) "How can a statistically averaged surgeon-individual treatment profile (SiTP) be identified?", and (3) "How can these SiTPs be used to derive advanced training strategies for the surgeon?"

The major focus of publications on education and training of cataract surgery was on the design and evaluation of training programs for residents, such as the training of residents by a virtual mentor system [Henderson et al. 2010a], the training of complication management [Prakash et al. 2009], the impact of the residents' curriculum design on complication rates [Rogers et al. 2009], or the use of models to teach residents surgical work steps in eye surgery [Hashimoto et al. 2001; Figueira et al. 2008] . Furthermore, a number of works have focused on the objective assessment of surgical skills with advanced methods [Fisher et al. 2006; Cremers et al. 2005; Taylor et al. 2007], or on data acquisition strategies in ophthalmology training [Saleh et al. 2007; Bhogal et al. 2010].

However, none of these works has used explicitly measured surgical processes to assess and support the education and training and to achieve detailed improvement results. The explicit measuring and modeling of surgical processes is a relatively new research topic. So far, application scenarios for surgical process models have mainly been the optimization of surgical treatment strategies [Neumuth et al. 2009c] by innovative computer-assisted technologies, the comparison of different intraoperative treatment strategies [Neumuth et al. 2011b], the evaluation of surgical mistakes [Malik et al. 2003], and the evaluation of the application of surgical tools and devices [Mehta et al. 2002], or surgical assist systems [Strauß et al. 2006a]. Furthermore, the approaches feature early engineering developments for the control of semi-automated surgical tasks [Münchenberg et al. 2001a], as well as process engineering [Casaletto and Rajaratnam 2004].

In the associated literature there is only a very limited number of approaches that explicitly deal with the acquisition and modeling of surgical processes. MacKenzie et al. [MacKenzie et al. 2001] presented a hierarchically organized model of a laparoscopic fundoplication procedure according to Nissen and Jannin et al. described a method of modeling supratentatorial tumor removals in the area of neurosurgery [Jannin et al. 2003; Jannin and Morandi 2007]. However, neither of these two considered the identification of surgical treatment profiles.

We present our feasibility study along with the measurement method that was used to model the surgical treatments and the strategy of computing surgeon-individual treatment profiles. To our best knowledge, a comparable approach is not available in the appertaining literature. Therefore, we believe that this work represents an important contribution to evidence-based surgery.

Methods and Materials

Patient sample and participating surgeons

In the course of the presented study, 105 cataract procedures were measured on a detailed-to-the-second work step level. All procedures were performed by experienced surgeons in their daily routine. The assignment of the patients to the respective surgeons was performed by an assistant medical director during the entry examination under exclusively clinical aspects with regard to the anticipated complications during the intervention and the health-related general condition of the patients. The decision concerning in- or outpatient intervention was made on the basis of the clinical guidelines of the German Ophthalmologic Society [Deutsche Ophthalomoligsche Gesellschaft e.V. DOG 2009]. Two surgeons (#1 and #2) performed outpatient procedures, while the third surgeon (#3) performed inpatient procedures. A further selection of patients, for instance with regard to age, gender or severity of affliction, did not take place, as this was irrelevant to demonstrate the applicability of the method. Table 5.3.I shows the patient characteristics.

Measurement of surgical processes

The computation of the SiTP was a multi-stage approach comprising definitions, measurements and data base computations (cp. Figure 5.3.1, left-hand side). An overview of relevant terms used to explain these methods is presented in Table 5.3.II.The computation of SiTPs is based on a detailed measurement of surgical activities ([Neumuth et al. 2009b], Figure 5.3.1, right-hand side): for every single surgical work step relevant data needs to be collected. This data encompasses information on what is being done, who performs the work step, whereby the work step is performed (meaning which instrument is being used), where at the patient's body the work step is carried out, and when it is performed. The collection of all surgical activities that were measured during the treatment of one patient is termed (patient-) individual surgical process model (iSPM).

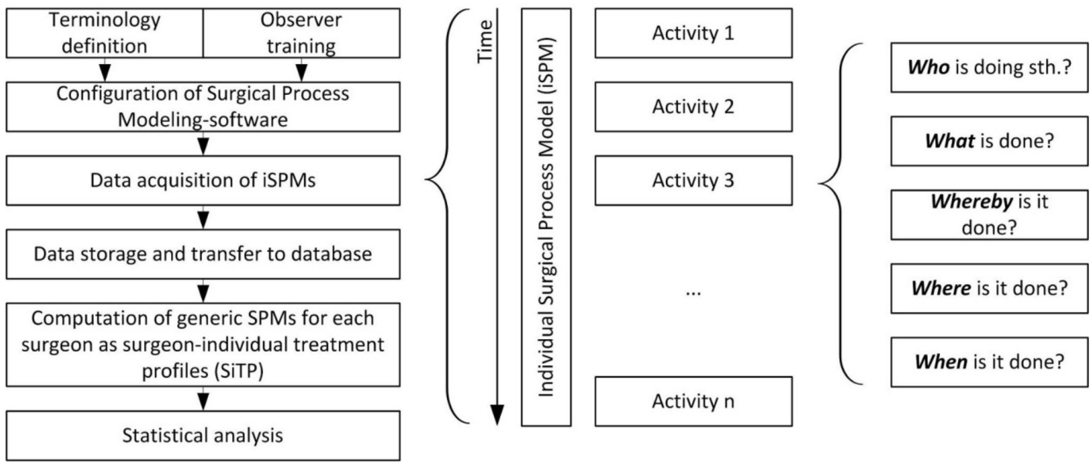


Figure 5.3.1: Stages for the measurement, computation, and analysis of surgeon-individual treatment profiles (SiTP, left-hand side) and principle of the data representation in the individual surgical process model (iSPM, right-hand side)

At the beginning of the measurement the clinical nomenclature needs to be defined. It comprises the names for the surgical phases of the procedure type, the surgical instruments and appliances needed, and the anatomical and pathological structures affected. This was performed jointly by a trained medical observer and an experienced surgeon from the department.

The terms of the clinical nomenclature were used to configure the measurement software, the surgical workflow editor [Neumuth et al. 2009b; Neumuth et al. 2006a]. The surgical workflow editor software was used by an observer to measure iSPMs (cp. Figure 5.3.2). It is able to describe the surgical process in detail and to the split second. During the actual intervention, the observer "translates" the observed work steps with the help of the software into a machine-readable format. The start of each beginning interventional phase is tagged and the data concerning *who*, *what*, *whereby*, and *where* is being collected (see Table 5.3.III for examples). The information concerning the *when* – the temporal information concerning the starting point and the breakpoint – is gathered automatically by the software.

All cataract procedures of the feasibility study were recorded by a trained medical student as observer. The observer had to complete a substantial training program before he started taking measurements. This program encompassed the dealing with the software and with the information concerning the procedure itself, such as the involved terms of the nomenclature. During data acquisition, the observer was present in the operating room during the actual intervention and operated the software on a touch-screen tablet-PC. The so recorded iSPMs were stored on the tablet-PCs in machine-readable extensible markup language (XML) format.

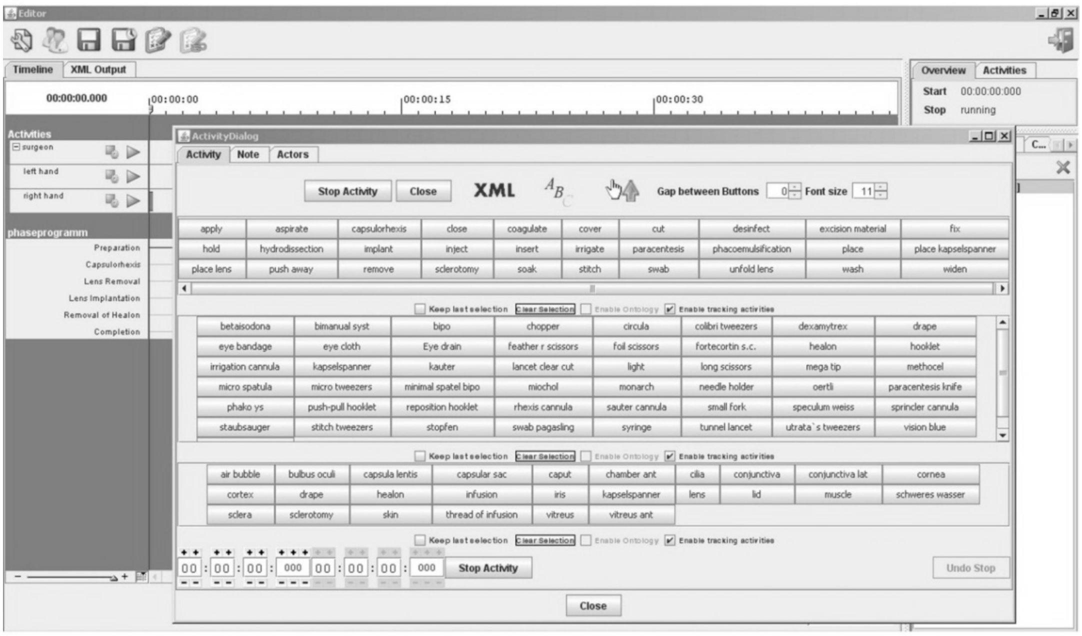


Figure 5.3.2: Screenshot of the tool kit for data acquisition: the surgical workflow editor

Generation of surgeon-individual treatment profiles and statistical analysis

After the completion of the data acquisition, the iSPMs were transferred into a data base for further processing. The activities in the iSPMs were grouped according their association to one of the surgical phases (cp. Table 5.3.IV). Subsequently, a generic surgical process model was computed for each of the surgeons and each of the surgical phases as surgeon-individual treatment profile. This gSPM contained the number and the average performance times of each activity and a probability for the following surgical activity. Finally, the gSPMs were filtered to delete infrequently occurring activities. The strategy has been described more specific in Neumuth et al. [Neumuth et al. 2011b].

To assess the differences between the SiTPs of the three surgeons, we calculated means and standard deviations for the occurrence number and the mean duration of each activity as well as for the probabilities. A statistical analysis using Bonferroni-tests with a significance level of α=0.05 was performed with the help of the statistics software SPSS [SPSS Inc. 2008] to check the means for statistical significance.

Table 5.3.I: Patient characteristics for the study.

	Surgeon #1	Surgeon #2	Surgeon #3
No. of cases	36	18	51
Mean patient age	70.1±9.6	63.5±13.3	73.7±7.8
Sex (m/f)	14/22	9/9	22/29
Treated eye (right/left)	18/18	6/12	29/22

Table 5.3.II: Terms and Definitions related to the method of measuring surgeon-individual treatment profiles.

Term	Definition
Surgical process (SP)	Surgical procedure, performed at one specific patient.
Surgical process model (SPM)	General term for a computer model of a surgical procedure course.
Activity	Representation of a surgical work step in the Surgical Process Model.
Individual surgical process model (iSPM)	Computer model of a surgical procedure course.
Generic surgical process model (gSPM)	Statistical averaged computer model of multiple surgical procedure courses.
Surgeon-individual treatment profile (SiTP)	gSPM that was computed for a number of patients that were treated by the same surgeon.

Table 5.3.III: Examples of activity descriptions for surgical work steps.

	Example activity 1	Example activity 2
Who	surgeon with right hand	surgeon with right hand
What	hydrodissection	wash
Whereby	sauter cannula	sprinkler cannula
Where	cortex	conjunctiva
When	00:05:30 – 00:06:10	00:02:30 – 00:02:40

Table 5.3.IV: Interventional core phases and their definitions.

Phase	Definition
Opening of the lens bag by rhexis cannula	From first paracentesis until end of material excision
Cataract Removal	Form hydrodissection until end of irrigation/aspiration of lens cortex
Posterior chamber intraocular lens implantation (PC-IOL)	From incision widening until beginning of irrigation/aspiration of Healon®
Removal of Healon®	Irrigation/aspiration of Healon® from anterior chamber

Results

The cut-suture-times showed averaged time spans of 00:23:01±00:11:59 for surgeon #1, 00:30:03±00:20:57 for surgeon #2, and 00:16:01±00:04:39 for surgeon #3 (cp. Table 5.3.V). The surgical core phase following the conclusion of the preparation, until the end of the Healon® removal was 00:15:17±00:12:20 for surgeon #1, 00:22:17±00:21:54 for surgeon #2, and 00:09:50±00:03:22 for surgeon #3. The differences were not statistically significant, with the exception of the difference between surgeons #2 and #3 (p<0.001).

Table 5.3.V: Cut-suture times and durations of the interventional phases (in hours:minutes:seconds and with mean ± standard deviation).

[mean±sd]	Surgeon #1	Surgeon #2	Surgeon #3	Between subject effects	p-value #1-#2	p-value #1-#3	p-value #2-#3
Cut-suture-time	00:23:01 ±00:11:59	00:30:03 ±00:20:57	00:16:01 ±00:04:39	F=12.4, p<0.001	p>0.05	**p=0.005**	**p<0.001**
Duration from begin of Opening the lens bag until end of Removal of Healon®	00:15:17 ±00:12:20	00:22:17 ±00:21:54	00:09:50 ±00:03:22	F=8.4, p<0.001	p>0.05	p>0.05	**p<0.001**
Opening of the lens bag by rhexis cannula	00:02:44 ±00:01:08	00:02:57 ±00:01:18	00:01:28 ±00:00:28	F=26.6, p<0.001	p>0.05	**p<0.001**	**p<0.001**
Cataract removal	00:09:34 ±00:10:29	00:11:50 ±00:08:41	00:05:42 ±00:02:24	F=9.5, p<0.001	**p=0.05**	p>0.05	**p<0.001**
Posterior chamber intraocular lens implantation	00:01:02 ±00:01:17	00:00:50 ±00:00:27	00:00:42 ±00:00:29	F=1.5, p>0.05	p>0.05	p>0.05	p>0.05
Removal of Healon®	00:01:10 ±00:00:42	00:02:45 ±00:03:51	00:01:37 ±00:01:18	F=4.1, p=0.02	**p=0.02**	p>0.05	p>0.05

For the surgical core phases highly significant differences (p<0.001) were determined; for the phase “Opening of the lens bag by rhexis cannula”, for instance, for the surgeons #1 and #3, as well as for #2 and #3. The differences in duration of the “Cataract removal” phase were highly significant (p<0.001) concerning surgeons #2 and #3, but only on a low significance level concerning surgeons #1 and #2 (p=0.05). The implantation of the lens showed no significant differences between the surgeons, while the removal of Healon® again showed a difference between surgeons #1 and #2, but again on a low significance level (p=0.02).

Figure 5.3.3 shows the individual progression course of the surgical process for each of the three surgeons for the interventional phase “Opening of the lens bag by rhexis cannula”. For a higher lucidity, all activity sequences occurring with a probability of less than 10% for all three surgeons were filtered. For the same reason, activity sequences with a probability of occurrence of more than 40% for all three surgeons were highlighted as main path by bold lines. For activities thus highlighted, the durations of the work steps have been computed exemplarily in Table 5.3.V.

Table 5.3.VI shows the average number of occurrences of a surgical activity during the phase and its average cumulated execution time. It turned out that the activity “Paracentesis with right hand” was performed significantly more often by surgeons #3 ($P<0.001$) and #2 ($p<0.001$) than by surgeon #1. However, the differences in averaged total execution times were not statistically significant.

The number of the occurrences of activity “Capsulorhexis with right hand”, on the other hand, was not significantly different; while the averaged cumulated durations were 00:01:15±00:00:24 (surgeon #1), 00:01:03 ±00:00:23 (surgeon #2), and 00:00:34±00:00:11 (surgeon #3) and therefore significantly different between surgeons #1 and #3 resp. #2 and #3 ($p<0.001$). The utilization of the different surgical instruments micro spatula and colibri tweezers to hold the Bulbus oculi by surgeon #3 is also visible in the analysis’ results.

The probability of sequence for the core activities of the surgical workflow and the respective differences in the SiTPs are represented in Table 5.3.VII and Figure 5.3.3. While most of the activity sequences in the SiTPs were not highly significantly different, the probability for the occurrence of the sequence “Paracentesis with right hand” to “Injection of Healon®” was highly significant for surgeons #1 and #3 on the one hand ($p<0.001$), and surgeons #2 and #3 on the other ($p<0.004$). The performance of both-handed paracenteses by surgeons #1 and #2 is represented in the results as statistically significant differences in the SPMs compared to the surgeon #3.

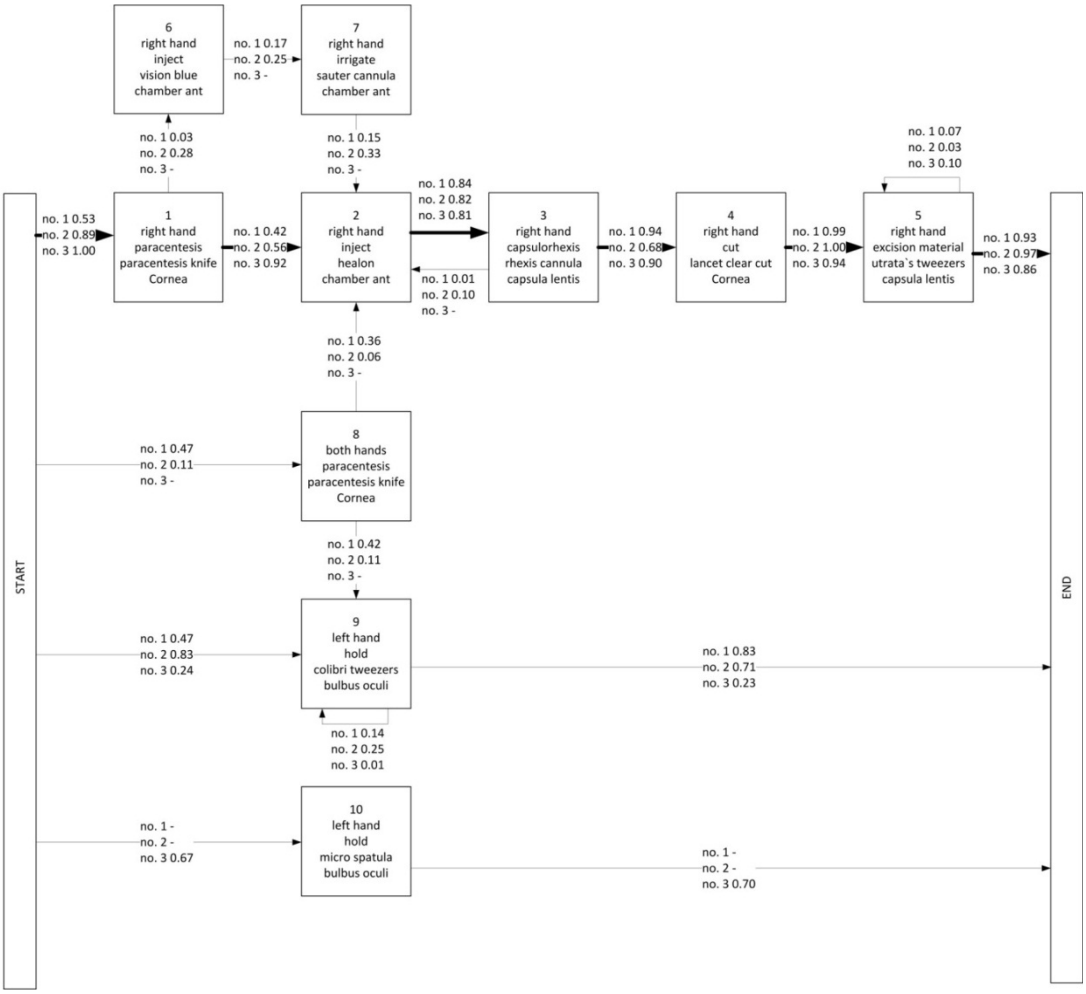


Figure 5.3.3: Visualization of the surgeon-individual treatment profiles (SiTPs) as generic surgical process models of the three surgeons for the intervention phase "Opening of the lens bag by rhexis cannula". The graph shows the most frequently measured activities. The probability of on activity following another one is indicated as percentage for each surgeon with the labels on the edges.

Table 5.3.VI: Average performance frequencies of surgical activities and average total performance times for each surgeon.

ID	Activity	Surgeon #1	Surgeon #2	Surgeon #3	p-value #1-#2	p-value #1-#3	p-value #2-#3	Surgeon #1	Surgeon #2	Surgeon #3	p-value #1-#2	p-value #1-#3	p-value #2-#3
1	right hand paracentesis paracentesis knife cornea	0,53 ±0,51	0,89 ±0,32	1.00 ±0.00	**p=0.001**	**p<0.001**	p>0.05	00:00:17 ±00:00:30	00:00:10 ±00:00:08	00:00:06 ±00:00:01	p>0.05	**p=0.01**	p>0.05
2	right hand inject Healon® chamber ant	1,17 ±0,45	1,50 ±0,92	1,04 ±0,34	p>0.05	p>0.05	**p=0.005**	00:00:06 ±00:00:02	00:00:10 ±00:00:05	00:00:04 ±00:00:01	**p<0.001**	**p=0.008**	**p<0.001**
3	right hand capsulorhexis rhexis cannula capsula lentis	1,11 ±0,52	1,28 ±0,83	1,02 ±0,24	p>0.05	p>0.05	p>0.05	00:01:15 ±00:00:24	00:01:03 ±00:00:23	00:00:34 ±00:00:11	p>0.05	**p<0.001**	**p<0.001**
4	right hand cut lancet clear cut cornea	1,03 ±0,17	1.00 ±0.00	1,02 ±0,24	p>0.05	p>0.05	p>0.05	00:00:05 ±00:00:01	00:00:04 ±00:00:01	00:00:03 ±00:00:01	p>0.05	**p<0.001**	p>0.05
5	right hand excision material Utrata`s tweezers capsula lentis	1,14 ±0,35	1,06 ±0,24	1,18 ±0,43	p>0.05	p>0.05	p>0.05	00:00:06 ±00:00:04	00:00:07 ±00:00:03	00:00:05 ±00:00:04	p>0.05	p>0.05	p>0.05
6	right hand inject vision blue® chamber ant	0,17 ±0,38	0,44 ±0,62	0.00 ±0.00	**p=0.01**	p>0.05	**p<0.001**	00:00:06 ±00:00:02	00:00:07 ±00:00:03	n.a.	p>0.05	-	-
7	right hand irrigate sauter cannula chamber ant	0,22 ±0,59	0,33 ±0,49	0,04 ±0,28	p>0.05	p>0.05	p>0.05	00:00:20 ±00:00:05	00:00:15 ±00:00:07	00:00:16 ±.00:00:00	p>0.05	p>0.05	p>0.05
8	both hands paracentesis paracentesis knife cornea	0,47 ±0,51	0,11 ±0,32	0.00 ±0.00	**p=0.001**	**p<0.001**	p>0.05	00:00:13 ±00:00:20	00:00:08 ±00:00:01	n.a.	p>0.05	-	-

9	left hand hold colibri tweezers bulbus oculi	1,42 ±0,69	1,83 ±1,15	0,27 ±0,49	p>0.05	**p<0.001**	**p<0.001**	00:01:42 ±00:00:29	00:01:48 ±00:00:52	00:00:37 ±00:00:21	p>0.05	**p<0.001**	**p<0.001**
10	left hand hold micro spatula bulbus oculi	0.00 ±0.00	0.00 ±0.00	0,92 ±0,59	p>0.05	**p<0.001**	**p<0.001**	n.a.	n.a.	00:00:48 ±00:00:13	-	-	-

Table 5.3.VII: Sequence probability for the work steps of the right hand (ideal progression course).

ID	Start activity	Stop activity	Surgeon #1	Surgeon #2	Surgeon #3	p-value #1-#2	p-value #1-#3	p-value #2-#3
S-1	START	right hand paracentesis paracentesis knife cornea	0.53 ±0.51	0.89 ±0.32	1.00 ±0.00	**p=0.001**	**p<0.001**	p>0.05
S-8	START	both hands paracentesis paracentesis knife cornea	0.47 ±0.51	0.11 ±0.32	0.00 ±0.00	**p=0.001**	**p<0.001**	p>0.05
S-9	START	left hand hold colibri tweezers bulbus oculi	0.47 ±0.51	0.83 ±0.38	0.24 ±0.43	**p=0.02**	p>0.05	**p<0.001**
S-10	START	left hand hold micro spatula bulbus oculi	0.00 ±0.00	0.00 ±0.00	0.67 ±0.48	p>0.05	**p<0.001**	**p<0.001**
1-2	right hand paracentesis paracentesis knife cornea	right hand inject Healon chamber ant	0.42 ±0.5	0.56 ±0.51	0.92 ±0.27	p>0.05	**p<0.001**	**p=0.004**
1-6	right hand paracentesis paracentesis knife cornea	right hand inject vision blue chamber ant	0.03 ±0.17	0.28 ±0.46	0.00 ±0.00	**p<0.001**	p>0.05	**p<0.001**
2-3	right hand inject Healon chamber ant	right hand capsulorhexis rhexis cannula capsula lentis	0.84 ±0.34	0.82 ±0.38	0.81 ±0.37	p>0.05	p>0.05	p>0.05
3-2	right hand capsulorhexis rhexis cannula capsula lentis	right hand inject Healon chamber ant	0.01 ±0.04	0.10 ±0.24	0.00 ±0.00	**p=0.004**	p>0.05	**p=0.001-**
3-4	right hand capsulorhexis rhexis cannula capsula lentis	right hand cut lancet clear cut cornea	0.94 ±0.22	0.68 ±0.4	0.90 ±0.28	**p=0.006**	p>0.05	**p=0.02**
4-5	right hand cut lancet clear cut cornea	right hand excision material utrata`s tweezers capsula lentis	0.99 ±0.08	1.00 ±0.00	0.94 ±0.22	p>0.05	p>0.05	p>0.05
5-5	right hand excision material utrata`s tweezers capsula lentis	right hand excision material utrata`s tweezers capsula lentis	0.07 ±0.18	0.03 ±0.12	0.10 ±0.20	p>0.05	p>0.05	p>0.05
5-E	right hand excision material utrata`s tweezers capsula lentis	END	0.93 ±0.18	0.97 ±0.12	0.86 ±0.27	p>0.05	p>0.05	p>0.05

6-7	right hand inject vision blue chamber ant	right hand irrigate sauter cannula chamber ant	0.17 ±0.38	0.25 ±0.43	0.00 ±0.00	p>0.05	**p=0.02**	**p=0.005**
7-2	right hand irrigate sauter cannula chamber ant	right hand inject Healon chamber ant	0.15 ±0.35	0.33 ±0.49	0.00 ±0.00	p>0.05	p>0.05	**p<0.001**
8-2	both hands paracentesis paracentesis knife cornea	right hand inject Healon chamber ant	0.36 ±0.49	0.06 ±0.24	0.00 ±0.00	**p=0.002**	**p<0.001**	p>0.05
8-9	both hands paracentesis paracentesis knife cornea	left hand hold colibri tweezers bulbus oculi	0.42 ±0.50	0.11 ±0.32	0.00 ±0.00	**p=0.004**	**p<0.001**	p>0.05
9-E	left hand hold colibri tweezers bulbus oculi	END	0.83 ±0.27	0.71 ±0.31	0.23 ±0.42	p>0.05	**p<0.001**	**p<0.001**
9-9	left hand hold colibri tweezers bulbus oculi	left hand hold colibri tweezers bulbus oculi	0.14 ±0.25	0.25 ±0.30	0.01 ±0.07	p>0.05	**p=0.009**	**p<0.001**
10-E	left hand hold micro spatula bulbus oculi	END	0.00 ±0.00	0.00 ±0.00	0.70 ±0.43	p>0.05	**p<0.001**	**p<0.001**

Discussion

The identification of surgeon-individual treatment profiles supports the provision of an optimum treatment service of ophthalmological departments. While the performance of a complete homogeneous treatment service is not possible due to patient-specific characteristics, the ophthalmological department can promote to come close to this objective by an advanced training of their surgical staff members for frequently occurring treatments, such as cataract procedures. We demonstrated with our feasibility study that it possible to measure and compute SiTPs. Based on this, options for individual advanced training can be identified for each surgeon.

The results of the example phase demonstrated that it is possible to measure detailed quantitative information concerning the ascertainable criteria frequency, duration, and sequence of surgical activities. With the help of the individual and generic surgical process models as used in this work, treatment profiles could be derived for each surgeon.

As was shown by means of the example of the interventional phases, a selective appraisal of different time spans is possible. For instance, the differences in cut-suture-times ($p<0.001$) and in the overall durations of the single core phases ($p=0.001$) from surgeon #1 to surgeon #2 and surgeon #3 were highly significant. A further investigation has shown that these differences are mainly due to the different time spans of the phases “Opening of the lens bag by rhexis cannula” ($p<0.001$) and “Cataract removal” ($p<0.002$).

The treatment profile of surgeon #3 appeared to be distinctly different from those of the other two surgeons. With him, work step repetitions occurred less frequently, for instance concerning the work step “Inject Healon® with right hand in chamber anterior”. His work steps had a shorter duration, for instance “Capsulorhexis with rhexis cannula”. Furthermore, he had a distinct preference concerning the instrument used to “Hold bulbus oculi” as compared to surgeons #1 and #2.

It was furthermore possible to identify differences in the activity sequences in the treatment profiles, as shown at the example of the sequence “Paracentesis with right hand” and “Injection of Healon®”. Surgeon #3 had a probability of occurrence of 92% concerning these activity sequences on his primary route, while the profile of surgeon #1 showed 56% and the profile of surgeon #1 42% probability. Here, the great differences between the three surgeons can be explained by the performance of additional surgical work steps, namely „Inject Vision Blue®“ and „Irrigate with sauter cannula“ as performed by surgeons #1 and #2. We cannot make a detailed analysis whether this surgical decision was necessary for the individual patient.

From the presented results a concrete recommendation for the advanced surgical training can be derived. Firstly, a profound exchange of experience would be very useful. Surgeon #3, for instance, had the least repetitions and usually shortest performance times. Also, a training of the work step “Paracentesis with right hand” could lead to an amelioration of results.

The presented data acquisition method allows for a detailed description of surgical processes and has been validated profoundly in previous studies [Neumuth et al. 2009b]. At the same time, the validity of the presented results was secured by the extensive training and instruction of the observers. In previous studies it has already been evaluated that it is possible to compute a gSPM from a number of iSPMs. This strategy allowed for representative computing of a treatment profile by preserving the variability of the surgical processes on the one hand and by eliminating infrequently occurring surgical activities on the other hand. It was also shown that gSPM generated from iSPMs lead to the procedure course that was recommended by clinical guidelines [Neumuth et al. 2011b].

An improvement of the presented study might be the control of the allocation of patients to the single surgeons. In this study, this allocation was not performed on the basis of age, gender, or severity of affliction, but it was regarded as randomized, because it can be assumed that the surgeon cannot chose his patients in his normal work either. Nevertheless an influence of different surgical severity cannot be excluded. The examination of these influences remains to be done in future clinical studies. Furthermore, the setting of one surgeon performing inpatient procedures and two surgeons performing outpatient procedures was chosen on purpose to demonstrate the feasibility of the approach to compare treatment profiles between surgeons performing the same strategy (surgeon #1 vs. #2) and surgeons performing different strategies (surgeon #1 resp. #2 vs. #3). Finally, we have intentionally refrained from discussing the reasons for the determined temporal differences in detail, because after the demonstration of the feasibility of the generation of surgeon-individual treatment profiles, the design of clinical studies can be featured.

Conclusion

The success of an ophthalmological department mainly depends on the capabilities of the surgical staff. We showed that it is possible to identify surgeon-individual treatment profiles (SiTP) to promote advanced training of surgical staff and therefore to support optimal patient treatment service. By means of the feasibility study it was shown that detailed profiles could be gathered with the help of surgical process modeling that provides an exact, validated, and objective decision base for the support of surgical teaching in the realm of evidence-based eye surgery.

Further extensions of treatment profiles are conceivable based on the presented results of this study. The computation of treatment profiles for using different surgical strategies or different surgical instruments is possible due to the availability of the method. These assessments could be done in cooperation with ophthalmologic societies to identify best-practice know-how for the optimal patient care.

Acknowledgements

The authors thank the team that supported the performance of the study and the preparation of the article at the Innovation Center Computer Assisted Surgery, University of Leipzig: Caroline Elzner and Michael Thiele. The authors also thank the surgeons that were subject to this study for their willingness to participate.

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the scope of the Unternehmen Region with the grant numbers 03 ZIK 031 and 03 ZIK 032 and by funds of the European Regional Development Fund (ERDF) and the state of Saxony within the frame of measures to support the technology sector.

6 Discussion and outlook

6.1 Achievements of milestones

Conclusively, the core statements of the milestones will be elaborated and completed with a discussion on remaining points of criticism and controversies. A repetition of the discussions presented in the single chapters and a discussion of the results in relation to the relevant literature will be omitted as far as possible, if these points have already been detailed in the original articles.

Milestone 1a: Development of a process ontology for surgical processes

Until today, no process ontology for surgical process models that integrates different modeling approaches into a single concept is in existence. In the recent past, various strategies to represent surgical processes from a technical point of view have been developed and presented. Even though all of these approaches have varying fields of application, they all share a common goal: to model the surgical process. Within the milestone, a generic framework for the mathematical and formal description of these processes has been designed and its applicability to a number of different approaches has been demonstrated. This process ontology constitutes a domain-level ontology with the capability to express differing strategies for the modeling of SPMs, including expert-based top-down modeling, observer-based modeling, and sensor-based data acquisition.

The modeling of surgical processes, however, has to overcome a prominent problem: the use of natural language. This usually entails unwanted characteristics such as ambiguity and equivocation. Nonetheless, the main fashion in which knowledge concerning surgical processes is expressed by its clinical users is by using natural language. Yet there remains a gap between the measurement of data and its interpretation as knowledge. With the methods in existence, this gap cannot be bridged and the creation and translation of further knowledge is hindered. Methods from the field of linguistics can help to reduce the gap between data and knowledge. Especially the use of verbs as process categories proves helpful, as verbs can be semantically defined and structured. Thus, verbs can be used to represent knowledge concerning the time-stamped data of surgical processes and these processes, represented in natural language, can be mapped onto mathematical descriptive pattern.

The concept of process granularity could not be conclusively resolved. The explicit description and formalization of different levels of granularity for surgical processes needs to be further developed. In the presented work, this problem has been circumvented by using formal granularity functions whose absolute integration remained open. It remains to be resolved, whether fixed granularity levels can be formulated.

Milestone 1b: Design similarity metrics for surgical process models

To assess observer-based recordings, special metrics are required and the observers need to be trained especially. In addition, the data acquisition needs to be evaluated, before studies can be performed based on the data thus generated.

Until now, no applicable metrics for the evaluation of observer-based recordings of surgical process models are available. Inter- and intra-observer assessments in nearby research fields are typically performed with the help of aggregated key figures such as correlations. However, metrics for a comparison of observer-based observation require a greater resolution to be able to detect and remedy the weaknesses of the observers comprehensively.

Additionally, an assessment of observer-based SPM-records is necessary with regard to several dimensions: they can be compared with regard to their granularity, content, time, order, and frequency. All of these dimensions represent different aspects, which are beneficial for the application-oriented studies in chapters 4 and 0. Prospectively, a further development of these metrics, for instance with regard to an independence from registrations, would be desirable.

Milestone 2a: Observer-based data acquisition with observation support software

Trained observers can be supported by observation support software. This software allows for the design of proper surgical process models by sampling complex surgical procedures as sequences of surgical activities. Surgical activities in turn are composed by the combination of perspectives that represent different views on the same surgical work step. The application of the software allows for data acquisition of freshly trained observers with an accuracy of more than 90 %.

Milestone 2b: Observer-based data acquisition with adaptive user interfaces

The observer-based acquisition of data for surgical process models can be further supported by employing adaptive user interfaces of the modeling support software. These adaptive user interfaces rely on a knowledge base and assist the observers in dealing with extensive and comprehensive terminologies. The knowledge base can serve to limit the choice of possible items for the observer. He has to choose from this with relation to the current situation to be able to describe a surgical activity correctly and exhaustively. The use of this knowledge base and the situation-dependent adaptation of the user interfaces enables even less experienced and non-specialist observers to attain observation results of quality.

Milestone 2c: Sensor-based data acquisition

The automatic online-recognition of partial information for SPMs with the help of sensor systems is feasible and desirable. Available approaches employ different strategies: the recognition of surgical instruments from videos, kinematic data acquisition, data compilation from virtual environments, or with the help of power or acceleration sensors. However, the weaknesses of such sensor-based strategies are mainly due to the complex implementation of technology. In addition, the detection capability for each different type of a surgical intervention needs to be evaluated afresh, as the sensor technologies mostly cannot be conferred from one type of intervention to another.

To achieve an input power similar to that of observer-based recordings, the implementation of multimodal fusion strategies is sensible. Nonetheless, existing approaches are not yet capable of recognizing multiple, different perspectives of SPMs. The increase of the input power via combination of various technologies is reasonable. This combination can be obtained by different fusion methods: using redundant, complementary, or cooperative fusion of partial information.

Milestone 3a: Computation of generic surgical process models

For the populations of surgical process models it is possible to compute a statistically averaged intervention course as a generic surgical process model. As shown in previous publications, iSPMs are valid for representing the procedure course of one surgical case. To be able to make broader assertions, two different strategies are conceivable: either, each model can be analyzed individually, as shown in section 5.1, or a new model, a generic surgical process model (gSPM), can be generated using generalization, each gSPM containing the information of all used iSPMs. These gSPMs can also be analyzed with respect to clinical aspects, such as shown in the sections 5.2 and 5.3.

To construct gSPMs, the emphasis was laid on the preparation of a simple and straightforward way to construct gSPMs. Even though considerably more complex algorithms are available in the context of business information systems, these were omitted since their results are not intuitively assessable by clinicians. The accommodation of simplicity and the use of a minimal model, however, also lead to a lower diversity of information. This results, for instance, in the partial loss of knowledge concerning preceding work step sequences.

By using the method of computing gSPMs, the recommended strategy for surgical interventions, as propagated in clinical guidelines, for instance, can be reconstructed. It was shown that the gSPMs generated from cataract interventions could be used to reconstruct the respective clinical guidelines. Thus, the clinical evaluation of the method is supported and a new possibility, namely to generate (semi-)automatic models of clinical guidelines or standard operating procedures by means of gSPMs, has been gained.

gSPMs represent an objective likeness of any surgical processes. By using measurement results and their conditioning, the objectivity of the process model is enhanced, especially when compared to top-down modeling strategies. Additionally, more detailed information is available to enable the computation of averaged execution times and frequentness of variants of the iSPMs. As has been shown, this information also is an appropriate basis for the analysis and comparison of surgical strategies.

Milestone 3b: Process model-based generation of workflow schemata

Generic surgical workflow models can be employed for the generation of workflow schemata for surgical workflow management systems (SWFMSs) for the operating room. These schemata, functioning as workflow specifications, are then used as supporting tools during the conduction of future surgical processes.

SWFMSs are well able to follow the course of a surgical process to a certain extent. The validity of the created models has been tested with the help of cataract surgery. This test has shown that a gSPM that has been computed from 50 iSPMs can track 90 iSPMs accurately. The establishing of a regression equation for the prediction of the success of a workflow scheme is possible and successful by means of various parameters.

The proposed study setup can be generalized beyond the specific intervention type of cataract surgery. Since the iSPMs as data input are universally applicable, and also the study setup does not rely on a specific intervention type, the overall approach can be used as a test bed to identify the minimum number of iSPMs that is required to generate a workflow schema for any intervention type. However, for practical application, the activities represented in the workflow schemata need to be enhanced by technical linking of appliances and information systems in the OR.

Milestone 4a: Deriving requirements for surgical assist systems

Requirements analysis performed with the aid of SPMs can be very beneficial for technical users. Aspects, such as design and implementation conditions of surgical assist systems, can be quantified by means of these models. Using the concrete example of lumbar discectomies in neurosurgery, it was shown that the selection and analysis of bone ablation work steps can result in predicting application conditions for surgical assist systems. However, qualitative properties, such as ergonomics or usability cannot be directly gathered from SPMs.

Milestone 4b: Evaluation of surgical assist systems

The investigation of SPMs allows for a quantification of the impact of resources on the surgical process, as demonstrated for the evaluation of the DaVinci telemanipulator system in pediatric surgery. To be able to quantify the deployment of the resource, the strategy is straightforward: from a set of iSPMs all those with and without the respective resource were chosen and assigned to different groups. These subsets were subsequently used to compute two different gSPMs which are afterward tested and evaluated against one another. Assuming that the general conditions remain consistent, the deviation within the gSPMs can then be traced to the assignment of the surgical assist system or resource. Using the gSPM approach, it was shown that the new resource had only a slight impact on the surgical process and that the use of the system does not yield objectively quantifiable benefits for the clinical user.

Milestone 4c: Assessment of surgical strategies

Finally, SPMs can be employed to compare practice modes of individual surgeons. Using gSPMs it was feasible to recognize differences in modi operandi and surgical strategies, to quantify the differences and to assess them in a clinically sensible way to facilitate discussions. At the same time, the SPMs can be used to homogenize the treatment success of the respective department.

6.2 Achievement of objectives and prospect

Achievement of objectives

The elicitation and analysis of surgical processes is of high relevance in connection with a vast number of applications in the fields of surgery and medical engineering. Furthermore, as has been shown in the previous chapters, acquisition, computation, and analysis of surgical processes is beneficial for both clinical users, such as surgeons, and technical users, such as medical engineers.

The problems concerning the modeling of processes ensuing from top-down modeling could be solved with the help of the development of this new approach, which differs fundamentally from the previously existing one. By means of the presented strategies in this work, the ICT-based acquisition, computation, and analysis of very detailed, objective, and quantifiable SPMs was rendered possible.

Particularly the computation of gSPMs as statistical mean process models represents a major new concept in computer assisted surgery. As demonstrated in the different sections of this work, gSPMs are essential to enable a multitude of process analysis strategies, such as strategy assessments, resource impact profiles or surgeon-individual treatment profiles. Additionally, the a priori functionality of an ICT system for the digital OR of the future will fundamentally rely on the concept of gSPMs.

The universal applicability of the designed methods was shown by employing them in different surgical disciplines and use cases. In addition, all achieved results have been evaluated and assessed following scientific procedures and strategies. The initially presented milestones have been satisfactorily fulfilled and the hypotheses basically proven.

Prospect

Beside the application areas of surgical process models presented in this work, such as requirements analyses, assessment of surgical instruments, systems, and strategies, the proposed methods could also trigger and be conductive to a great variety of further applications.

Integrated operating room management

With reference to the advance of the management of digital operating room, a trend towards the integration of different information classes into domain specific submodels is emerging. Other than the surgical process models introduced and elaborated on in this work, this holds true for a model which integrates all relevant information concerning the patient. Additionally, due to the growing technical interpenetration, connectivity, and communication between the single technical components used in surgical interventions, it may be expected that another model will prove necessary: a technical resource model which represents, amongst others, information such as the system states of the technical infrastructure. Similar concepts are already in existence in different economic branches, for instance the process monitoring used in process control engineering. The concepts of process models and workflow management thus represent an essential pillar of the integrated management of the digital operating room (see Figure 6.2.1).

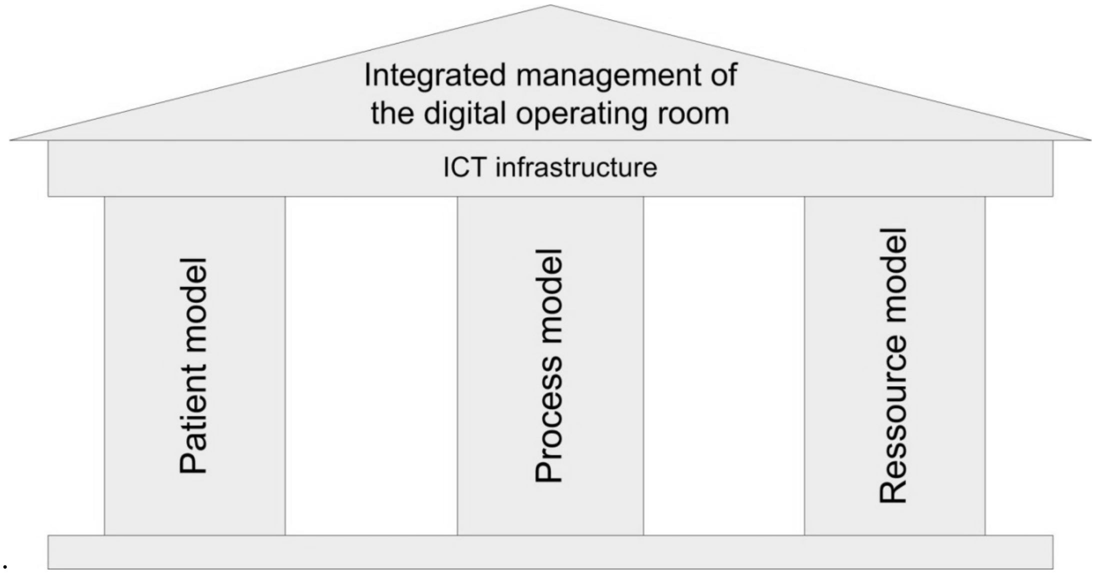


Figure 6.2.1: Integrated management of the digital operating room based on patient, process, and resource model.

Given this development as background, the need of an increasing, continuous, and interconnected acquisition of data, supported by standardized interfaces for the automatic detection of the conditions of the patient, of the process, and of the resources emerge. This concept might be realized by employing a central data recorder, the surgical black-box-system that works in analogy to a flight recorder used in aviation. This system should be able to receive and save all accruing data, persistently and legal. Subsequently, this data can be made available for other systems implementing clinical applications. Figure 6.2.2 shows the screenshot of a surgical black box prototype. However, due to the vast amount of expected data, further research and development efforts are necessary to structure the data that need to be stored, at which intervals this should happen, which lossless compensation methods are needed, and how the information storage can be technically accomplished.

For the persistent collection and storage of the data, with special regard to its sensitive background, it is necessary, on the one hand, to integrate existing systems as possible providers of information into the overall concept. On the other hand, it is also indispensable to create new systems for the bridging of technological information gathering gaps.

One of the more advanced areas of information gathering, as perceived from the information technology point of view, is the acquisition of data for the intraoperative patient model. Modern anesthesia systems, for instance, already provide a multitude of physiological patient data in digital form. Furthermore, the plurality of preoperative data concerning the patient, such as radiological or histological information, already exists in digital form, even though the partial inadequacies concerning an integration of the different systems need to be considered and tackled.

Figure 6.2.2: Prototype of the surgical black box system that might be used e.g. for automatic documentation of the intervention for quality management purposes.

As for the automatic capture of the surgical process models, some development works are still necessary, mainly due to the fact that here, in contrast to the data origins of the patient model, only limited methods for an automated data acquisition are available. In addition, it may be expected that the data needed for the elicitation of surgical process models is very heterogeneous. This is mainly owed to the fact to the wide variety of different information available, for instance videos, instrument trajectories, movement profiles of the surgical team, or the recalling of patient data from radiological systems resp. from the hospital information system (HIS). Thus, a single acquisition modality will possibly not be able to elicit all the data needed for the surgical process model. Rather, a fusion of information gathered by different sensor systems will be more apt for this task.

Taken even further, this notion also suggests, for instance, the review of the process data acquisition using a sensor-grid-network in the operating room. This sensor grid contains various peripheral sensors distributed in the OR, which gather information using different measuring procedures and consider various process perspectives, such as the location and movements of individuals or appliances within the room.

The challenge concerning the acquisition of process data continues with the recognition and abstraction of knowledge from the raw data. Here, the research task includes the finding of computing strategies, such as machine learning, that will be able to deduce information concerning the progression of the overall process from the physical raw data of the single sensors.

For the accumulation of data for the resource model, some preparatory work has been carried out in the last few years. Due to the rising number of available new generation of integrated operation rooms using interconnected components, a networking of essential data sources by the manufacturer of the integrated OR-systems is already well underway.

Integrated OR infrastructures, using distributed components, however, induce a growing complexity of the overall system, of the available data, and of the communication of the latter. This problem was tackled by Bohn et al. [Bohn et al. 2009] who introduced the monitoring of technical resources with the help of a supervisory control and data acquisition application (SCADA, see Figure 6.2.3). This approach proposes technical concepts for the observation, diagnosis, and control of hard- and software components. Also, this supervision concept for single components is necessary for the surgeon to be able to track and protocol the systems' status, such as availability or malfunction, with the help of the black-box-system. In addition, it is possible to present suitable strategic modifications to the surgeon and let him draw conclusions concerning the overall condition of the system.

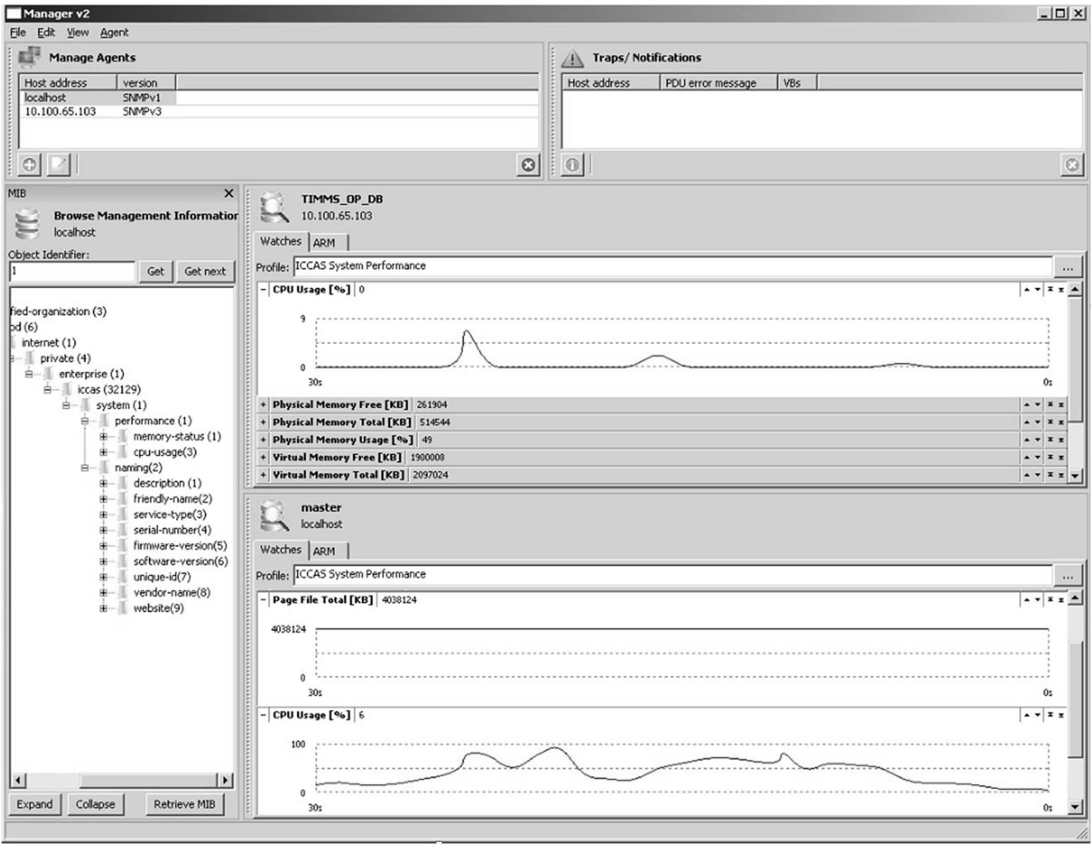


Figure 6.2.3: Acquisition of resources: a sequence of elicited performance parameters gathered from interconnected components.

For the functionality of the central monitoring and the processing of complex data of patient and process models it is reasonable to create a specialized functional unit with the ability to perform this purpose for several operating rooms at once. The concept of the surgical control center (SCC) for the establishment of a central control console seems predestined for this task (see Figure 6.2.4). This allows for a centralized processing of multiple tasks, such as the monitoring of the technical system components, the performance of resource-intensive functions to abstract workflow knowledge from raw sensor input, simulation or image processing, leading to a more optimal utilization of available resources, concerning both, technology and man-power. As a bonus, this saves the hospitals from the acquirement of (partly) redundant functional units, such as computation resources, for each OR separately. Likewise, this centralization would enable the creation and exchange of central repositories for both, patient and process model alike. According to the respective requirements, this central control console can be established for the use within a single hospital, for a hospital operator, or for manufacturers to supervise their integrated OR-systems at the customer sites.

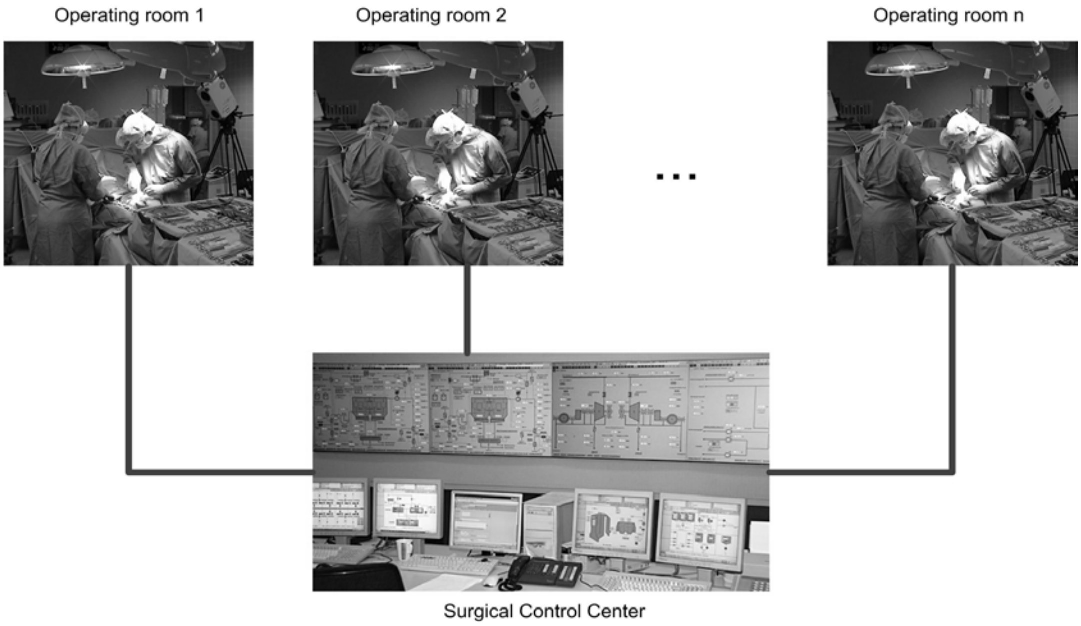


Figure 6.2.4: Concept of the surgical control center (SCC): Centralized supervision of patients, processes, and technical resources of multiple operating rooms.

Emerging applications for improved patient care based on surgical process models

A number of clinically relevant use cases are facilitated by the methods presented here. These use cases encompass various domains, such as the intraoperative technical assistance for the appropriate support of the surgical work, the management of the operating rooms, quality management, the adherence to clinical standards, documentation purposes, as well as surgical training and further education.

In the field of *intraoperative technical assistance* for the surgeon, the goal of future developments should be to establish a functional connection (see Figure 6.2.5) between the technical resources in the operating room and the surgical process, for instance to present surgically relevant data in every stage of a surgical intervention. Based on the knowledge about the process and the recognition of the current situation, surgeons could be supported by model guided interventions, e.g. to navigate along the surgical process an can be provided with an effective access to relevant information at exactly the time that it is needed, whether before, during or after the actual surgical intervention. Alongside the situation-dependent presentation of the preoperatively acquired data, such as the patient records or histological examination results, the intraoperative parameterization and control of the technical equipment, such as the prompting of intraoperative gathered information, for instance from automatic measurements, is also of interest.

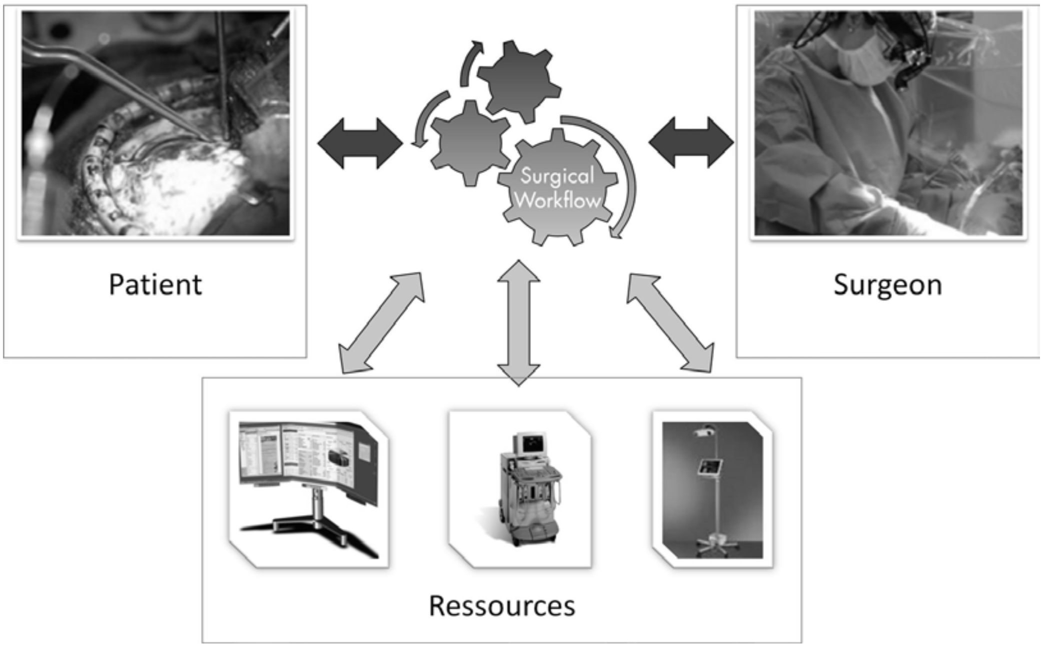


Figure 6.2.5: Intraoperative surgical assistance: The best possible support as appropriate for the situation.

These future systems need to be integrated seamlessly into the patient treatment process and various systems need to support the surgeon optimally without interfering with his usual work or adding to his workload. To achieve this goal, the existing knowledge, gathered by ICT-systems, needs to be presented at the best and the handling of the surgical user interfaces needs to take place intuitively. Figure 6.2.6 presents an interface concept for a "process navigation system" and Figure 6.2.7 shows a current prototype of the process navigation system.

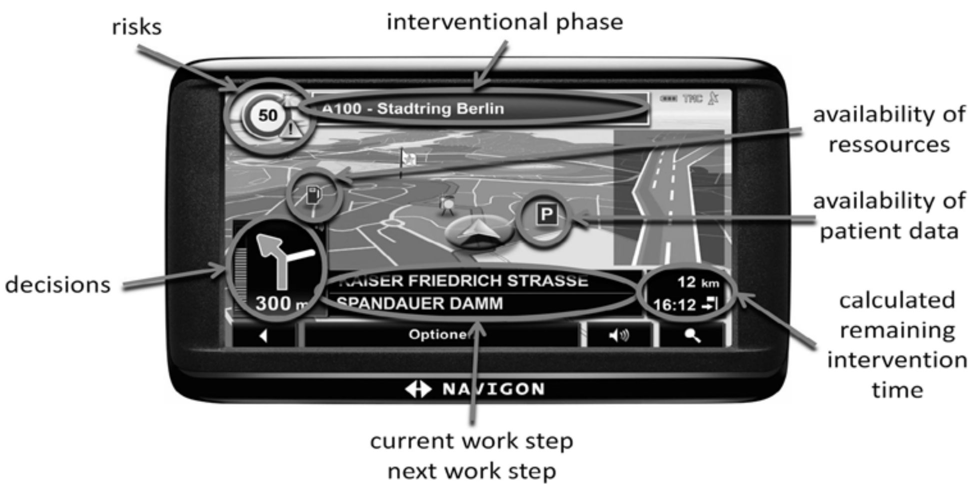


Figure 6.2.6: User interface concept for the surgical management and guidance system in analogy to a car navigation system.

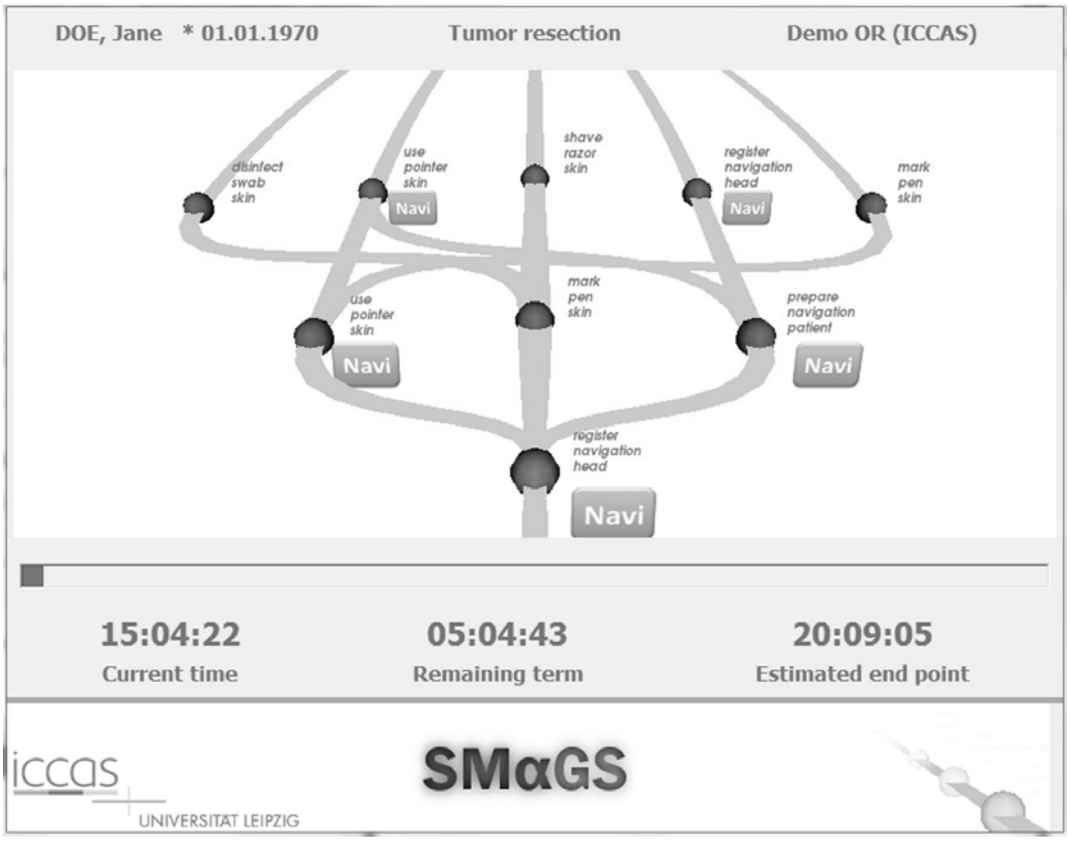


Figure 6.2.7: Prototype of the surgical management and guidance system user interface.

Concerning the notion of *quality management and documentation*, the results compiled in this work can contribute to an automatic documentation of clinically relevant parameters for the electronic patient record that can be extracted from the operating room black box system. Use cases are the automatic triggering of picture or video recordings to be used for documentation purposes. At the same time, whole sequences of surgical work steps could be automatically documented, leading to the machine-controlled generation of a process log of the surgical intervention or of parts thereof. Consequently, the obtained models can serve as a new basis for clinical standardization, e.g. to support the design of clinical guidelines.

In the case of complex and demanding interventions, an automatic analysis of the current SPM situation during the intervention according to specific rules, such as the accomplishment of indispensable and essential work steps, could be conceivable. This seems necessary, as studies have shown that especially novice surgeons tend to skip work steps (e.g. [Webster et al. 2005]). Finally, gSPMs could be used as basis for the standardization of clinical guidelines and standard operating procedures, and, as such, have a great economic potential, for instance for the cost-optimization in health services.

Another field that could greatly benefit from the presented work would be *operating room management*. Here, the intraoperatively gathered and analyzed models and information could be used for a pre- and postoperative control of patient treatment processes in the hospital. For instance, a prediction of the conclusion time of a surgical intervention based on the recognition of the current situation and of known remaining work steps in the surgical process, could be of interest for the timely admittance of the next patient to the OR. Moreover, the use of workflow schemata as basis for software simulations of different scenarios of surgical interventions could help with the identification of critical stages and with the optimization of perioperative strategies to streamline the work in ORs.

Also, SPMs could prove a valuable asset for *surgical education and training*. On the one hand, the identification and comparison of different working strategies, as presented in section Assessment of surgeon’s strategies in ophthalmology, could be useful in this area, and, on the other hand, gSPMs could be employed for the simulation of different variants of a surgical intervention. This would enable surgeons to train on real scenarios and to extend surgical training beyond isolated scenes or work steps.

The results of the work presented here is also of mediate use for various *adjacent user groups*. Firstly, the well-being of the patient can be optimized with the help of a continuous process optimization. Secondly, the hospital administration or that of health insurance companies can implement these methods to ensure an economic employment of resources. This engenders, amongst others, the employment of SPMs as benchmarking support for decisions concerning capital expenditures. Finally, society in general will benefit from the use of SPM-based methods. With the availability of this new instrument, the continuous development and amelioration of patient treatment can be achieved and evidence-based surgery employed. Furthermore, with the help of surgical process models, surgical expert knowledge can be facilitated and passed on in alternative form, distributed and used for optimal patient care.

7 Bibliography

VAN DER AALST, W. 1997. Verification of workflow nets. *Azema P, Balbo G (eds): Proceedings of the 18th International Conference on Application and Theory of Petri Nets*, Springer-Verlag, 407–426.

VAN DER AALST, W. 2005. Business alignment: using process mining as a tool for delta analysis and conformance testing. *Requirements Engineering Journal 10*, 3, 198–211.

VAN DER AALST, W., VAN DONGEN, B., HERBST, J., MARUSTERA, L., SCHIMM, G., AND WEIJTERS, A. 2003. Workflow mining: A survey of issues and approaches. *Data & Knowledge Engineering 47*, 2, 237–267.

VAN DER AALST, W. AND VAN HEE, K. 2002. *Workflow Management: models, methods, and systems*. MIT Press, Cambridge.

VAN DER AALST, W. AND TER HOFSTEDE, A. 2005. YAWL: yet another workflow language. *Information Systems 30*, 4, 245–275.

VAN DER AALST, W., DE MEDEIROS, A., AND WEIJTERS, A. 2006a. Process equivalence in the context of genetic mining. *Data 64*, 1, 55–74.

VAN DER AALST, W., DE MEDEIROS, A., AND WEIJTERS, A. 2006b. Process equivalence: comparing two process models based on observed behavior. In: *International Conference on Business Process Management (BPM 2006)*. Springer, Berlin, Heidelberg, 129–144.

ACHARYA, R., YU, W., ZHU, K., NAYAK, J., LIM, T., AND CHAN, J. 2010. Identification of cataract and post-cataract surgery optical images using artificial intelligence techniques. *Journal of Medical Systems 34*, 4, 619–628.

ADLASSNIG, K., COMBI, C., DAS, A., KERAVNOU, E., AND POZZI, G. 2006. Temporal representation and reasoning in medicine: Research directions and challenges. *Artificial Intelligence in Medicine 38*, 2, 101–113.

AGARWAL, S., JOSHI, A., FININ, T., YESHA, Y., AND GANOUS, T. 2007. A pervasive computing system for the operating room of the future. *Mobile Network Applications 12*, 2-3, 215–228.

AGRAWAL, R., GUNOPULOS, D., AND LEYMANN, F. 1998. Mining process models from workflow logs. In: I. Ramos, G. Alonso, H. Schek and F. Saltor, eds., *Advances in Database Technology - EDBT'98*. Springer, Berlin, Heidelberg, 469–483.

AHMADI, A., PADOY, N., RYBACHUK, K., FEUßNER, H., HEINING, S., AND NAVAB, N. 2010. Motif discovery in OR sensor data with application to surgical workflow analysis and activity detection. *MICCAI Workshop on Modeling and Monitoring of Computer Assisted Interventions (M2CAI)*.

AHMADI, S., SIELHORST, T., STAUDER, R., HORN, M., FEUSSNER, H., AND NAVAB, N.

2006. Recovery of surgical workflow without explicit models. In: R. Larsen, M. Nielsen and J. Sporring, eds., *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2006)*. Springer, Berlin, Heidelberg, 420–428.

AHRQ-AGENCY FOR HEALTH CARE RESEARCH AND QUALITY. 2010a. National Guideline Clearinghouse. *National Guideline Clearinghouse*. http://www.guideline.gov/.

AHRQ-AGENCY FOR HEALTH CARE RESEARCH AND QUALITY. 2010b. National Guideline Clearinghouse. *National Guideline Clearinghouse*. http://www.guideline.gov/.

ALBERTY, J. 2004. Erste Erfahrungen mit dem G-DRG-Fallpauschalensystem. *Laryngorhinootologie 83*, 11, 759–760.

ALLEN, J. 1984. Towards a general theory of action and time. *Artificial Intelligence 23*, 2, 123–154.

ANSELMA, L. AND MONTANI, S. 2008. Planning: supporting and optimizing clinical guidelines execution. *Studies in Health Technology and Informatics 139*, 101–120.

ARCHER, T. AND MACARIO, A. 2006. The drive for operating room efficiency will increase quality of patient care. *Current Opinion in Anaesthesiology 19*, 2, 171–176.

ARTIBANI, W., FRACALANZA, S., CAVALLERI, S., ET AL. 2008. Learning curve and preliminary experience with da Vinci-assisted laparoscopic radical prostatectomy. *Urologia Internationalis 80*, 3, 237–244.

AWMF-ARBEITSGEMEINSCHAFT DER WISSENSCHAFTLICHEN MEDIZINISCHEN FACHGESELLSCHAFTEN E.V. 2010a. Klinische Leitlinien. *Science-based guidelines for diagnostics and therapy issued by the scientific medical societies in Germany*. http://www.awmf-leitlinien.de/.

AWMF-ARBEITSGEMEINSCHAFT DER WISSENSCHAFTLICHEN MEDIZINISCHEN FACHGESELLSCHAFTEN E.V. 2010b. Science-based Guidelines for Diagnostics and Therapy issued by the Scientific Medical Socities in Germany. http://www.awmf-leitlinien.de/.

AYANNIYI, A., ADEPOJU, F., AND OWOEYE, J. 2009. Trainee ophthalmologists' opinions on ways to improve cataract surgical rate. *Annals of African Medicine 8*, 4, 276–280.

BACH, E. 1986. The algebra of events. *Linguistics and Philosophy 9*, 1, 5–16.

BAE, J., CAVERLEE, J., LIU, L., AND ROUSE, W. 2006a. Process mining, discovery, and integration using distance measures. In: *International Conference on Web Services 2006*. IEEE Computer Society.

BAE, J., CAVERLEE, J., LIU, L., AND YAN, H. 2006b. Process mining by measuring process block similarity. In: J. Eder and S. Dustdar, eds., *Business Process*

Management Workshops 2006. 141–152.

BAE, J., LIU, L., CAVERLEE, J., AND ZHANG, L. 2007. Development of distance measures for process mining, dicovery and integration. *International Journal of Web Services Research 4*, 4, 1–17.

BAGLIO, M., BAXTER, S., GUINN, C., THOMPSON, W., SHAFFER, N., AND FRYE, F. 2004. Assessment of interobserver reliability in nutrition studies that use direct observation of school meals. *Journal of the American Dietetic Association 104*, 9, 1385–1392.

BAIK, S., KWON, H., KIM, J., ET AL. 2009. Robotic versus laparoscopic low anterior resection of rectal cancer: short-term outcome of a prospective comparative study. *Annals of Surgical Oncology 16*, 6, 1480–1487.

BARDRAM, J. AND NØRSKOV, N. 2008. A context-aware patient safety system for the operating room. In: *Proceedings of the 10th international conference on Ubiquitous computing*. ACM, Seoul, Korea, 272–281.

BARGAR, W. 2007. Robots in orthopaedic surgery: past, present, and future. *Clinical Orthopaedics and Related Research 463*, 31–36.

BARKAOUI, K., DECHAMBRE, P., AND HACHICHA, R. 2002. Verification and optimisation of an operating room workflow. In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences 2002*. IEEE Computer Society, Los Alamitos, CA, USA, 210.

BARMER GEK. 2010. Arztreport 2010. http://www.barmer-gek.de/barmer/web/Portale/Versichertenportal/Presse-Center/Infomaterial/Arztreport/Arztreport.html?w-cm=CenterColumn_tdocid.

BARNES, J. 1995. Metaphysics. In: *The Complete Works of Aristotle: The Revised Oxford Translation*. Princeton University Press.

BAUMGART, A., SCHÜPFER, G., WELKER, A., BENDER, H., AND SCHLEPPERS, A. 2010. Status quo and current trends of operating room management in Germany. *Current Opinion in Anaesthesiology 23*, 2, 193–200.

BECKER, K., AUFFARTH, G., AND VÖLCKER, H. 2004. Measurement method for the determination of rotation and decentration of intraocular lenses. *Der Ophthalmologe 101*, 6, 600–603.

BECKER, K., HOLZER, M., REULAND, A., AND AUFFARTH, G. 2006. Accuracy of lens power calculation and centration of an aspheric intraocular lens. *Der Ophthalmologe 103*, 10, 873–876.

BELL, M., TORGERSON, J., AND KREADEN, U. 2009. The first 100 da Vinci hysterectomies: an analysis of the learning curve for a single surgeon. *South Dakota Medicine 62*, 3, 91, 93–95.

BERNIE, J., VENKATESH, R., BROWN, J., GARDNER, T., AND SUNDARAM, C. 2005. Comparison of laparoscopic pyeloplasty with and without robotic assistance. *Journal of the Society of Laparoendoscopic Surgeons 9*, 3, 258–261.

BERRAIS, A. 1997. Knowledge-based expert systems: user interface implications. *Advances in Engineering Software 28*, 1, 31–41.

BESTE, T., NELSON, K., AND DAUCHER, J. 2005. Total laparoscopic hysterectomy utilizing a robotic surgical system. *Journal of the Society of Laparoendoscopic Surgeons 9*, 1, 13–15.

BHATIA, B., OATES, T., XIAO, Y., AND HU, P. 2007. Real-time identification of operating room state from video. In: *Proceedings of the 19th national conference on Innovative applications of artificial intelligence*. AAAI Press, Vancouver, British Columbia, Canada, 1761–1766.

BHOGAL, M., ANGUNAWELA, R., AND LITTLE, B. 2010. Use of low-cost video recording device in reflective practice in cataract surgery. *Journal of Cataract and Refractive Surgery 36*, 4, 542–546.

BLUM, T., PADOY, N., FEUSSNER, H., AND NAVAB, N. 2008a. Modeling and online recognition of surgical phases using Hidden Markov Models. In: D. Metaxas, L. Axel, G. Fichtinger and G. Szekely, eds., *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2008)*. Springer, Berlin, Heidelberg, 627–635.

BLUM, T., PADOY, N., FEUßNER, H., AND NAVAB, N. 2008b. Workflow mining for visualization and analysis of surgeries. *International Journal of Computer Assisted Radiology and Surgery 3*, 5, 379–386.

BOCCA, S., STADTMAUER, L., AND OEHNINGER, S. 2007. Current status of robotically assisted laparoscopic surgery in reproductive medicine and gynaecology. *Reproductive Biomedicine Online 14*, 6, 765–772.

BODNER, J., KAFKA-RITSCH, R., LUCCIARINI, P., FISH, J., AND SCHMID, T. 2005. A critical comparison of robotic versus conventional laparoscopic splenectomies. *World Journal of Surgery 29*, 8, 982–985.

DEN BOER, K., BRUIJN, M., JASPERS, J., ET AL. 2002a. Time-action analysis of instrument positioners in laparoscopic cholecystectomy. *Surgical Endoscopy 16*, 1, 142–147.

DEN BOER, K., DANKELMAN, J., GOUMA, D., AND STASSEN, H. 2001. Time-action analysis of laparoscopic procedures - input for clinically driven instrument design. *Minimally Invasive Therapy & Allied Technologies 10*, 3, 139–144.

DEN BOER, K., DANKELMAN, J., GOUMA, D., AND STASSEN, H. 2002b. Peroperative analysis of the surgical procedure. *Surgical Endoscopy 16*, 3, 492–499.

DEN BOER, K., STRAATSBURG, I., SCHELLINGER, A., DE WIT, L., DANKELMAN, J., AND GOUMA, D. 1999. Quantitative analysis of the functionality and efficiency of three surgical dissection techniques: a time-motion analysis. *Journal of Laparoendoscopic & Advanced Surgical Techniques 9*, 5, 389–395.

BOHN, S., LESSNAU, M., AND BURGERT, O. 2009. Systems monitoring and diagnosis of digital operating room (DOR) equipment using supervisory control and data acquisition (SCADA). *International Journal of Computer Assisted Radiology and Surgery 4*, sup 1, 146–147.

BONATTI, J., SCHACHNER, T., BONAROS, N., ET AL. 2008. How to improve performance of robotic totally endoscopic coronary artery bypass grafting. *American Journal of Surgery 195*, 5, 711–716.

BOOCH, G., RUMBAUGH, J., AND JACOBSON, I. 2005. *The unified modeling language user guide*. Addison-Wesley Longman, Amsterdam, Reading, MA.

BOUARFA, L., JONKER, P., AND DANKELMAN, J. 2010. Discovery of high-level tasks in the operating room. *Journal of Biomedical Informatics 44*, 3, 455–462.

BRAUMANN, C., JACOBI, C., MENENAKOS, C., ISMAIL, M., RUECKERT, J., AND MUELLER, J. 2008. Robotic-assisted laparoscopic and thoracoscopic surgery with the da Vinci system: a 4-year experience in a single institution. *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques 18*, 3, 260–266.

BRAY, T., PAOLI, J., SPERBERG-MCQUEEN, C., MALER, E., AND YERGAU, F. 2008. Extensible markup language (XML) 1.0.

BREITENSTEIN, S., NOCITO, A., PUHAN, M., HELD, U., WEBER, M., AND CLAVIEN, P. 2008. Robotic-assisted versus laparoscopic cholecystectomy: outcome and cost analyses of a case-matched control study. *Annals of Surgery 247*, 6, 987–993.

BURGERT, O., NEUMUTH, T., GESSAT, M., JACOBS, S., AND LEMKE, H. 2007. Deriving DICOM surgical extensions from surgical workflows. In: S. Horii and K. Andriole, eds., *Medical Imaging 2007: PACS and Imaging Informatics*. SPIE, Bellingham, WA, CID65150B.

CAO, A., CHINTAMANI, K., PANDYA, A., AND ELLIS, R. 2009. NASA TLX: software for assessing subjective mental workload. *Behavior Research Methods 41*, 1, 113–117.

CAO, C., MACKENZIE, C., AND PAYANDEH, S. 1996. Task and Motion analysis in endoscopic surgery. In: *ASME Dynamic Systems and Controls Division (5th Annual Symposium on Haptic Interfaces for Virtual Environments and Teleoperation Systems)*. ACM Press, Atlanta, Georgia, 583–590.

CARLSON, L. 1981. Aspect and quantification. In: P. Tedeschi and A. Zaenen, eds., *Syntax and Semantics*. Academic Press, New York, 31–64.

CARVALHO, L., STEFANI, M., ROMÃO, A., ET AL. 2002. Videokeratoscopes for dioptric power measurement during surgery. *Journal of Cataract and Refractive Surgery 28*, 11, 2006–2016.

CASALETTO, J. AND RAJARATNAM, V. 2004. Surgical process re-engineering: carpal tunnel decompression - a model. *Hand Surgery 9*, 1, 19–27.

CASATI, R. AND VARZI, A. 2008. Events. In: E. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Stanford University, Stanford.

CASTELLANO, J., PEREA, A., ALDAY, L., AND MENDO, A. 2008. The measuring and observation tool in sports. *Behavior Research Methods 40*, 3, 898–905.

CEGLOWSKI, A., CHURILOV, L., AND WASSERTHEIL, J. 2005. Knowledge discovery through mining emergency department data. In: *Hawaii International Conference on System Sciences*. IEEE Computer Society, Los Alamitos, CA, USA, 142c.

CHEUNG, C., LI, H., LAMOUREUX, E., ET AL. 2011. Validity of a new computer-aided diagnosis imaging program to quantify nuclear cataract from slit-lamp photographs. *Investigative Ophthalmology & Visual Science 52*, 3, 1314–1319.

CHITWOOD JR., W., KYPSON, A., AND NIFONG, L. 2003. Robotic mitral valve surgery: a technologic and economic revolution for heart centers. *The American Heart Hospital Journal 1*, 1, 30–39.

CHOI, K., SOO, S., AND CHUNG, F. 2009. A virtual training simulator for learning cataract surgery with phacoemulsification. *Computers in Biology and Medicine 39*, 11, 1020–1031.

CICEKLI, N. AND CICEKLI, I. 2006. Formalizing the specification and execution of workflows using the event calculus. *Information Sciences 176*, 15, 2227–2267.

CINQUIN, P. AND TROCCAZ, J. 2003. Model driven therapy - the instance of computer assisted medical interventions. *Methods of Information in Medicine 42*, 2, 169–176.

CLEARY, K. AND KINSELLA, A. 2005. OR 2020: the operating room of the future. *Journal of Laparoendoscopic & Advanced Surgical Techniques 15*, 5, 495, 497–573.

CLEARY, K., KINSELLA, A., AND MUN, S. 2005. OR2020 Workshop report: operating room of the future. In: H. Lemke, K. Inamura, K. Doi, M. Vannier and A. Farman, eds., 832–838.

DE CLERCQ, P., KAISER, K., AND HASMAN, A. 2008. Computer-interpretable guideline formalisms. *Studies in Health Technology and Informatics 139*, 22–43.

COMBI, C., GOZZI, M., OLIBONI, B., JUAREZ, J., AND MARIN, R. 2009. Temporal similarity measures for querying clinical workflows. *Artificial Intelligence in Medicine 46*, 1, 37–54.

COMRIE, B. 1976. *Aspect: an introduction to the study of verbal aspect and related problems*. Cambridge University Press, Cambridge.

COOK, J. AND WOLF, A. 1995. Automating process discovery through event-data analysis. In: New York, 73–82.

CREMERS, S., CIOLINO, J., FERRUFINO-PONCE, Z., AND HENDERSON, B. 2005. Objective assessment of skills in intraocular surgery (OASIS). *Ophthalmology 112*, 7, 1236–1241.

CRONBACH, L. AND MEEHL, P. 1955. Construct validity in psychological tests. *Psychological Bulletin 52*, 281–302.

CUADROS, J. AND BRESNICK, G. 2009. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *Journal of Diabetes Science and Technology 3*, 3, 509–516.

CUSCHIERI, A., FERREIRA, E., GOH, P., ET AL. 1997. Guidelines for conducting economic outcomes studies for endoscopic procedures. *Surgical Endoscopy 11*, 3, 308–314.

DARZI, A. AND MACKAY, S. 2002. Skills assessment of surgeons. *Surgery 131*, 2, 121–124.

DAVULCU, H., KIFER, M., RAMAKRISHNAN, C., AND RAMAKRISHNAN, I. 1998. Logic based modeling and analysis of workflows. In: ACM, Seattle, Washington, United States, 25–33.

DEINHARDT, M. 2003. Manipulators and integrated OR systems - requirements and solutions. *Minimally Invasive Therapy & Allied Technologies 12*, 6, 284–292.

DEUTSCHE KRANKENHAUS VERLAGSGESELLSCHAFT GMBH. 2009. Zahlen, Daten, Fakten 2009: Statistikmaterial zum deutschen und internationalen Krankenhauswesen. .

DEUTSCHE OPHTHALOMOLIGSCHE GESELLSCHAFT E.V. DOG. 2009. Veröffentlichungen der Deutschen Ophthalmologischen Gesellschaft e.V. *Deutsche Ophthalomoligsche Gesellschaft e.V.* http://www.dog.org/publikationen/index.html.

DEVI, S., RAO, K., AND SANGEETHA, S. 2012. Prediction of surgery times and scheduling of operation theaters in Optholmology department. *Journal of Medical Systems 36*, 2, 415–430.

DEXTER, F., ABOULEISH, A., EPSTEIN, R., WHITTEN, C., AND LUBARSKY, D. 2003a. Use of operating room information system data to predict the impact of reducing turnover times on staffing costs. *Anesthesia & Analgesia 97*, 4, 1119 –1126.

DEXTER, F., DEXTER, E., AND LEDOLTER, J. 2010. Influence of procedure classification on process variability and parameter uncertainty of surgical case durations. *Anesthesia and Analgesia 110*, 4, 1155–1163.

DEXTER, F., TRAUB, R., AND MACARIO, A. 2003b. How to release allocated operating room time to increase efficiency: predicting which surgical service will have the most underutilized operating room time. *Anesthesia & Analgesia 96*, 2, 507 –512.

DICKHAUS, C., BURGHART, C., TEMPANY, C., ET AL. 2004. Workflow modeling and analysis of computer guided prostate brachytherapy under MR imaging control. *Studies in Health Technology and Informatics 98*, 72–74.

DOCUMET, J., LE, A., LIU, B., CHIU, J., AND HUANG, H. 2010. A multimedia electronic patient record (ePR) system for image-assisted minimally invasive spinal surgery. *International Journal of Computer Assisted Radiology and Surgery 5*, 3, 195–209.

VAN DONGEN, B., DIJKMAN, R., AND MENDLING, J. 2008. Measuring similarity between business process models. In: J. Roddick and A. Hinze, eds., *Advanced Information Systems Engineering*. 450–464.

DOWTY, D. 1979. *Word meaning and Montague grammar*. Kluwer Academic Publishers, Dordrecht.

DOYLE, L., GAUTHIER, N., RAMANATHAN, S., AND OKAMURA, A. 2008. A simulator to explore the role of haptic feedback in cataract surgery training. *Studies in Health Technology and Informatics 132*, 106–111.

DURRANT-WHYTE, H. 1988. Sensor models and multisensor integration. *The International Journal of Robotics Research 7*, 6, 97 –113.

EUROPEAN COMMITTEE FOR STANDARDIZATION (CEN). 2002. Health informatics - Categorial structure for classifications and coding systems of surgical procedures (EN 1828).

EVERMANN, J. 2009. A UML and OWL description of Bunge's upper-level ontology model. *Software and Systems Modeling 8*, 2, 235–249.

EVERMANN, J. AND WAND, Y. 2001. Towards ontologically based semantics for UML constructs. In: H. Kunii, S. Jajodia and A. Solvberg, eds., *Conceptual Modeling - ER 2001, 20th International Conference on Conceptual Modeling*. Springer, Berlin, Heidelberg, 354–367.

EZRA, D., CHANDRA, A., OKHRAVI, N., SULLIVAN, P., MCDONNELL, P., AND LEE, J. 2010. Higher surgical training in ophthalmology: trends in cumulative surgical experience 1993-2008. *Nature Eye 24*, 1466–1473.

FANNING, J., FENTON, B., AND PUROHIT, M. 2008. Robotic radical hysterectomy. *American Journal of Obstetrics and Gynecology 198*, 6, 649.e1–4.

FERNANDEZ-LLATAS, C., MENEU, T., BENEDI, J., AND TRAVER, V. 2010. Activity-based process mining for clinical pathways computer aided design. In: *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*. 6178–6181.

FIGUEIRA, E., WANG, L., BROWN, T., ET AL. 2008. The grape: an appropriate model for continuous curvilinear capsulorhexis. *Journal of Cataract and Refractive Surgery 34*, 9, 1610–1611.

FISCHER, M., BURGERT, O., NEUMUTH, T., DIETZ, A., LEMKE, H., AND STRAUß, G. 2006. Limitations in endolaryngeal surgery - analysis of development potential by surgical workflows. *International Journal of Computer Assisted Radiology and Surgery 1*, sup. 1, 440–442.

FISHER, J., BINENBAUM, G., TAPINO, P., AND VOLPE, N. 2006. Development and face and content validity of an eye surgical skills assessment test for ophthalmology residents. *Ophthalmology 113*, 12, 2364–2370.

FISSELL, K. 2007. Workflow-based approaches to neuroimaging analysis. *Methods in Molecular Biology 401*, 235–266.

FOLLIGUET, T., VANHUYSE, F., CONSTANTINO, X., REALLI, M., AND LABORDE, F. 2006. Mitral valve repair robotic versus sternotomy. *European Journal of Cardio-Thoracic Surgery 29*, 3, 362–366.

FRIEDMAN, D., SOKAL, S., CHANG, Y., AND BERGER, D. 2006. Increasing operating room efficiency through parallel processing. *Annals of Surgery 243*, 1, 10–14.

FRIGG, R. AND HARTMANN, S. 2006. Models in science. In: E. Zalta, ed., *Stanford encyclopedia of philosophy*. Stanford University, Stanford.

GADATSCH, A. 2002. *Management von Geschäftsprozessen. Methoden und Werkzeuge für die IT-Praxis: Eine Einführung für Studenten und Praktiker*. Vieweg Verlag, Wiesbaden.

GÄRDENFORS, P. 2000. *Conceptual spaces: The geometry of thought*. MIT Press, Cambridge.

GEBHARD, E., HARTWIG, E., ISENMANN, R., ET AL. 2003. OR-manager: surgeon or anesthesiologist? An interdisciplinary study. *Der Anaesthesist 52*, 11, 1062–1067.

GELDNER, G., EBERHART, L., TRUNK, S., ET AL. 2002. Efficient OP management. Suggestions for optimisation of organisation and administration as a basis for establishing statutes for operating theatres. *Der Anaesthesist 51*, 9, 760–767.

GERMAN FEDERAL STATISTICAL OFFICE, ED. 2008a. International Classification of Diseases 10: National health statistics for 2007. .

GERMAN FEDERAL STATISTICAL OFFICE, ED. 2008b. International Classification of Diseases 10: National health statistics for 2007. .

GERMAN STATISTICAL FEDERAL OFFICE. 2007. International Classification of Diseases 10: National health statistics for 2006. .

GREEN, P. AND ROSEMANN, M. 2000. Integrated process modeling: an ontological evaluation. *Information Systems 25*, 2, 73–87.

GREINER, U., MUELLER, R., RAHM, E., RAMSCH, J., HELLER, B., AND LOEFFLER, M. 2005. AdaptFlow: protocol-based medical treatment using adaptive workflows.

Methods of Information in Medicine 44, 1, 80–88.

GRUBER, T. 1993. A translation approach to portable ontologies. *Knowledge Acquisition 5*, 2, 199–220.

HALL, D. AND LLINAS, J. 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE 85*, 1, 6–23.

HALSTED, M. AND FROEHLE, C. 2008. Design, implementation, and assessment of a radiology workflow management system. *American Journal of Roentgenology 191*, 2, 321–327.

HÄNDEL, A., MARTUS, P., KÜCHLE, M., AND SCHÖNHERR, U. 2002. Microsurgical quality assurance exemplified by cataract surgery. *Der Ophthalmologe 99*, 5, 352–357.

HÄNNINEN, L. AND PASTELL, M. 2009. CowLog: open-source software for coding behaviors from digital video. *Behavior Research Methods 41*, 2, 472–476.

HANSEN, T. AND BARDRAM, J. 2007. Applying mobile and pervasive computer technology to enhance coordination of work in a surgical ward. *Studies in Health Technology and Informatics 129*, 1, 107–111.

VAN HARMELEN, F. AND ANTONIOU, G. 2003. Web ontology language: OWL. In: S. Staab and R. Studer, eds., *Handbook on Ontologies*. Springer, Berlin, 76–92.

HART, S. AND STAVELAND, L. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P. Hancock and N. Meshkati, eds., *Human Mental Workload*. Elsevier Science Publishing Company, North-Holland, 53–71.

HARTMANN, J., JACOBI, C., MENENAKOS, C., ISMAIL, M., AND BRAUMANN, C. 2008. Surgical treatment of gastroesophageal reflux disease and upside-down stomach using the Da Vinci robotic system. A prospective study. *Journal of Gastrointestinal Surgery 12*, 3, 504–509.

HASHIMOTO, C., KUROSAKA, D., AND UETSUKI, Y. 2001. Teaching continuous curvilinear capsulorhexis using a postmortem pig eye with simulated cataract. *Journal of Cataract & Refractive Surgery 27*, 6, 814–816.

HAUX, R., SEGGEWIES, C., BALDAUF-SOBEZ, W., ET AL. 2003. Soarian--workflow management applied for health care. *Methods of Information in Medicine 42*, 1, 25–36.

HELD, J. AND MANSER, T. 2005. A PDA-based system for online recording and analysis of concurrent events in complex behavioral processes. *Behavior Research Methods 37*, 1, 155–164.

HENDERSON, B., KIM, J., GOLNIK, K., ET AL. 2010a. Evaluation of the virtual mentor cataract training program. *Ophthalmology 117*, 2, 253–258.

HENDERSON, B., KIM, J., GOLNIK, K., ET AL. 2010b. Evaluation of the virtual mentor cataract training program. *Ophthalmology 117*, 2, 253–258.

HENDLER, K. AND NAU, D. 1994. UMCP: A sound and complete procedure for hierarchical task network planning. In: *In Proceedings of the International Conference on AI Planning & Scheduling (AIPS)*. 249–254.

HERRE, H. 2009. General Formal Ontology: A foundational ontology for conceptual modeling. In: *Theory and Application of Ontology*. Springer, Berlin.

HERRE, H. AND HELLER, B. 2006. Semantic foundations of medical information systems based on top-level ontologies. *Knowledge-Based Systems 19*, 2, 107–115.

HERRE, H., HELLER, B., BUREK, P., HOEHNDORF, R., LOEBE, F., AND MICHALEK, H. 2007. General Formal Ontology (GFO): A foundational ontology integrating objects and processes. Part I: Basic principles. .

HERRE, H. AND LOEBE, F. 2005. A meta-ontological architecture for foundational ontologies. In: R. Meersmann and Z. Tari, eds., *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE*. Springer, Berlin, 1398–1415.

HORROCKS, I., PATEL-SCHNEIDER, P., AND VAN HARMELEN, F. 2003. From SHIQ and RDF to OWL: The making of a web ontology language. *Journal of Web Semantics 1*, 1, 7–26.

HU, J., NELSON, R., WILSON, T., ET AL. 2006. Perioperative complications of laparoscopic and robotic assisted laparoscopic radical prostatectomy. *The Journal of Urology 175*, 2, 541–546.

HUBERT, J. AND SIEMER, S. 2008. Nephrectomy, donor nephrectomy, and partial kidney resection: indications for robot-assisted renal surgery. *Der Urologe 47*, 4, 425–430.

IBBOTSON, J., MACKENZIE, C., CAO, C., AND LOMAX, A. 1999. Gaze patterns in laparoscopic surgery. *Studies in Health Technology and Informatics 62*, 154–60.

JABLONSKI, S. AND BUSSLER, C. 1996. *Workflow-management: Modelling concepts, architecture and implementation*. International Thomson Computer Press, London.

JAKISCH, R. 2007. *Controlling strategischer Medizintechnik-Investitionen: ein Leitfaden für Effizienz und Effektivität in Planung und Betrieb*. Diplomica Verlag, Hamburg.

JAMES, A., VIEIRA, D., LO, B., DARZI, A., AND YANG, G. 2007. Eye-gaze driven surgical workflow segmentation. In: N. Ayache, S. Ourselin and A. Maeder, eds., *Medical Image Computing and Computer-Assisted Intervention: (MICCAI 2007)*. Springer, Berlin, 110–117.

JANK, E., ROSE, A., HUTH, S., ET AL. 2006. A new fluoroscopy based navigation system for milling procedures in spine surgery. *International Journal of Computer*

Assisted Radiology and Surgery 1, sup. 1, 196–198.

JANNIN, P., GROVA, C., AND MAURER, C. 2008. Model for defining and reporting reference-based validation protocols in medical image processing. *International Journal of Computer Assisted Radiology and Surgery 1*, 2, 63–73.

JANNIN, P., KRUPINSKI, E., AND WARFIELD, S. 2006. Validation in medical image processing. *25*, 11, 1405–1409.

JANNIN, P. AND MORANDI, X. 2007. Surgical models for computer-assisted neurosurgery. *Neuroimage 37*, 3, 783–91.

JANNIN, P., RAIMBAULT, M., MORANDI, X., RIFFAUD, L., AND GIBAUD, B. 2003. Model of surgical procedures for multimodal image-guided neurosurgery. *Computer Aided Surgery 8*, 2, 98–106.

JASMAN, A., SHAHARUDDIN, B., NOOR, R., ISMAIL, S., GHANI, Z., AND EMBONG, Z. 2010. Prediction error and accuracy of intraocular lens power calculation in pediatric patient comparing SRK II and Pediatric IOL Calculator. *BMC Ophthalmology 10*, 20.

JOHN, H., SCHMID, D., AND FEHR, J. 2007. Extraperitoneal radical prostatectomy Da Vinci. *Actas Urologicas Españolas 31*, 6, 580–586.

JOLESZ, F. AND SHTERN, F. 1992. The operating room of the future. Report of the National Cancer Institute workshop, "Imaging-guided stereotactic tumor diagnosis and treatment."*Investigative Radiology 27*, 4, 326–328.

JUNG, W., YOUN, C., KIM, H., KIM, D., HAZEMI, F., AND SHIM, E. 2009. Policy-based hybrid workflow management system for heart disease identification. *Proceedings of the 15th Asia-Pacific conference on Communications*, IEEE Press, 826–829.

KAUL, S. AND MENON, M. 2007. Da Vinci assisted cystoprostatectomy and urinary diversion: a paradigm shift in surgical management of bladder cancer. *Minerva Urologica E Nefrologica 59*, 2, 149–157.

KENNY, A. 2003. *Action, emotion and will*. Routledge, London.

KHALIFA, Y., BOGORAD, D., GIBSON, V., PEIFER, J., AND NUSSBAUM, J. 2006. Virtual reality in ophthalmology training. *Survey of Ophthalmology 51*, 3, 259–273.

KOKAR, M., TOMASIK, J., AND WEYMAN, J. 2004. Formalizing classes of information fusion systems. *Information Fusion 5*, 3, 189–202.

KOWALSKI, R. AND SERGOT, M. 1986. A logic-based calculus of events. *New Generation Computing 4*, 1, 67–95.

KRAUSS, A., MÜNSTERER, O., NEUMUTH, T., ET AL. 2009. Workflow analysis of laparoscopic Nissen fundoplication in infant pigs – a model for surgical feedback and training. *Journal of Laparoendoscopic & Advanced Surgical Techniques 19*, Sup.1, 117–122.

KREFTING, D., VOSSBERG, M., HOHEISEL, A., AND TOLXDORFF, T. 2010. Simplified implementation of medical image processing algorithms into a grid using a workflow management system. *Future Generation Computer Systems 26*, 4, 681–684.

KREUTZER, T., AL SAEIDI, R., KAMPIK, A., AND GRUETERICH, M. 2010. Real-time intraocular pressure measurement in standard and microcoaxial phacoemulsification. *Journal of Cataract and Refractive Surgery 36*, 1, 53–57.

KREYHER, V. 2001. *Handbuch Gesundheits- und Medizinmarketing: Chancen, Strategien und Erfolgsfaktoren*. Hüthig Jehle Rehm, Heidelberg.

KROMREY, H. 2006. *Empirische Sozialforschung*. Lucius & Lucius, Stuttgart.

KUBITZ, J., EPPLE, J., LÜTZELBERGER, U., SCHMIDT, H., MOTSCH, J., AND BACH, A. 2001. Computer simulation and pharmacoeconomics. Computer simulation as an aid for the analysis of operating room efficiency: an example. *Der Anaesthesist 50*, 2, 122–127.

KUEHME, T. 1993. Adaptive action prompting: A complementary aid to support task-oriented interaction in explorative user interfaces. http://smartech.gatech.edu/handle/1853/3629.

KYRIACOU, E., FAKAS, G., AND PAVLAKI, V. 2006. A completely decentralized workflow management system for the support of emergency telemedicine and patient monitoring. In: *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*. IEEE, 5663–5666.

LALYS, F., RIFFAUD, L., BOUGET, D., AND JANNIN, P. 2011. An application-dependent framework for the recognition of high-level surgical tasks in the OR. *Med Image Comput Comput Assist Interv*.

LALYS, F., RIFFAUD, L., MORANDI, X., AND JANNIN, P. 2010. Automatic phases recognition in pituitary surgeries by microscope images classification. *Proceedings of the First international conference on Information processing in computer-assisted interventions*, Springer-Verlag, 34–44.

LANG, M., BÜRKLE, T., LAUMANN, S., AND PROKOSCH, H. 2008. Process mining for clinical workflows: challenges and current limitations. *Studies in Health Technology and Informatics 136*, 229–234.

LATOSZEK-BERENDSEN, A., TANGE, H., VAN DEN HERIK, H.J., AND HASMAN, A. 2010. From clinical practice guidelines to computer-interpretable guidelines. A literature overview. *Methods of Information in Medicine 49*, 6, 550–570.

LE, C. AND GETTMAN, M. 2006. Laparoscopic and robotic radical prostatectomy. *Expert Review of Anticancer Therapy 6*, 7, 1003–1011.

LEMKE, H. 2007. Summary of the white paper of DICOM WG24: DICOM in surgery. 651603.

LEMKE, H. AND VANNIER, M. 2006. The operating room and the need for an IT infrastructure and standards. *International Journal of Computer Assisted Radiology and Surgery 1*, 3, 117–122.

LENG, T., LUJAN, B., YOO, S., AND WANG, J. 2008. Three-dimensional spectral domain optical coherence tomography of a clear corneal cataract incision. *Ophthalmic Surgery, Lasers & Imaging 39*, 4, S132–134.

LEONG, J., NICOLAOU, M., ATALLAH, L., MYLONAS, G., DARZI, A., AND YANG, G. 2007. HMM assessment of quality of movement trajectory in laparoscopic surgery. *Computer Aided Surgery 12*, 6, 335–346.

LIGABUE, E. AND GIORDANO, C. 2007. Interpretation of aberrometry measurements in cataract surgery. *Journal of Refractive Surgery 23*, 9, 996–1004.

LIN, H. AND HAGER, G. 2009. User-independent models of manipulation using video contextual cues. *Workshop on modeling and monitoring of computer-assisted intervetntions*.

LIN, H., SHAFRAN, I., YUH, D., AND HAGER, G. 2006. Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. *Computer Aided Surgery 11*, 5, 220–230.

LINK, R., BHAYANI, S., AND KAVOUSSI, L. 2006. A prospective comparison of robotic and laparoscopic pyeloplasty. *Annals of Surgery 243*, 4, 486–491.

LLINAS, J., BOWMAN, C., ROGOVA, G., STEINBERG, A., WALTZ, E., AND WHITE, F. 2004. Revisiting the JDL data fusion model II. In: P. Svensson and J. Schubert, eds., *Proceedings of the 7th International Conference on Information Fusion*. International Society of Information Fusion, 1218–1230.

LOVE, B., JONES, M., TOMLINSON, M., AND HOWE, M. 2008. Predicting information needs: Adaptive display in dynamic environments. Proceedings of cognitive science society.

LUO, R., YIH, C., AND SU, K. 2002. Multisensor fusion and integration: approaches, applications, and future research directions. *IEEE Sensors Journal 2*, 2, 107–119.

MA, J. 2007. Ontological considerations of time, meta-predicates and temporal propositions. *Applied Ontology 2*, 1, 37–66.

MACKENZIE, C., IBBOTSON, J., CAO, C., AND LOMAX, A. 2001. Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. *Minimally Invasive Therapy & Allied Technologies 10*, 3, 121–127.

MACLIN, O. AND MACLIN, M. 2005. Coding observational data: a software solution. *Behavior Research Methods 37*, 2, 224–231.

MAEDA, N. 2009. Clinical applications of wavefront aberrometry - a review. *Clinical & Experimental Ophthalmology 37*, 1, 118–129.

MALIK, R., WHITE, P., AND MACEWEN, C. 2003. Using human reliability analysis to detect surgical error in endoscopic DCR surgery. *Clinical Otolaryngology and Allied Sciences 28*, 5, 456–460.

MANS, R., VAN DER AALST, W., RUSSELL, N., BAKKER, P., AND MOLEMAN, A. 2010a. Process-aware information system development for the healthcare domain - consistency, reliability, and effectiveness. In: *Business Process Management Workshops 2009*. 635–646.

MANS, R., SCHONENBERG, H., LEONARDI, G., ET AL. 2008. Process mining techniques: an application to stroke care. *Studies in Health Technology and Informatics 136*, 573–578.

MANS, R., SCHONENBERG, M., SONG, M., VAN DER AALST, W., AND BAKKER, P. 2009. Application of process mining in healthcare – A case study in a Dutch hospital. In: A. Fred, J. Filipe and H. Gamboa, eds., *Biomedical Engineering Systems and Technologies*. Springer, Berlin, Heidelberg, 425–438.

MANS, R.S., RUSSELL, N.C., VAN DER AALST, W., BAKKER, P.J.M., AND MOLEMAN, A.J. 2010b. Simulation to analyze the impact of a schedule-aware workflow management system. *SIMULATION 86*, 8-9, 519–541.

MARUSTER, L., VAN DER AALST, W., WEIJTERS, T., VAN DEN BOSCH, A., AND DAELEMANS, W. 2001. Automated discovery of workflow models from hospital data. In: B. Kröse, M. Rijke, G. Schreiber and M. Someren, eds., *Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2001)*. 183–190.

MASOLO, C., BORGO, S., GANGEMI, A., GUARINO, N., AND OLTRAMARI, A. 2003. WonderWeb deliverable D18: ontology library (final). http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf.

MATHWORKS. 2009. *MATLAB*. Natick, MA.

MAURO, C., LEIMEISTER, M., AND KRCMAR, H. 2011. Serviceorientierte Integration medizinischer Geräte - ganzheitliche IT-Unterstützung klinischer Prozesse. *Informatik-Spektrum 34*, 3, 276–285.

MCPHERSON, K. AND BUNKER, J. 2007. Costs, risks and benefits of surgery: a milestone in the development of health services research. *Journal of the Royal Society of Medicine 100*, 8, 387–390.

DE MEDEIROS, A., VAN DER AALST, W., AND WEJITERS, A. 2008. Quantifying process equivalence based on observed behavior. *Data & Knowledge Engineering 64*, 1, 55–74.

DE MEDEIROS, A., WEIJTERS, A., AND VAN DER AALST, W. 2005a. Genetic process mining: a basic approach and its challenges. In: *Workshop on Business Process*

Intelligence (BPI). Nancy.

DE MEDEIROS, A.K.A., AND WEIJTERS, A.J.M.M., AND AND VAN DER AALST, W.M.P. 2005b. Genetic process mining: a basic approach and its challenges. In: *Workshop on Business Process Intelligence (BPI)*. Nancy.

MEGALI, G., SIGNIGAGLIA, S., TONET, O., AND DARIO, P. 2006. Modelling and evaluation of surgical performance using hidden markov models. *IEEE Transactions on Biomedical Engineering 53*, 10, 1911–1919.

MEHTA, N., HALUCK, R., FRECKER, M., AND SNYDER, A. 2002. Sequence and task analysis of instrument use in common laparoscopic procedures. *Surgical Endoscopy 16*, 2, 280–285.

MENG, F., D'AVOLIO, L., CHEN, A., TAIRA, R., AND KANGARLOO, H. 2005. Generating models of surgical procedures using UMLS concepts and multiple sequence alignment. *AMIA Annual Symposium Proceedings*, 520–524.

MINEKUS, J., ROZING, P., VALSTAR, E., AND DANKELMAN, J. 2003. Evaluation of humeral head replacements using time-action analysis. *Journal of Shoulder and Elbow Surgery 12*, 2, 152–157.

MISHRA, Y., WASIR, H., SHARMA, K., MEHTA, Y., AND TREHAN, N. 2006. Totally endoscopic coronary artery bypass surgery. *Asian Cardiovascular & Thoracic Annals 14*, 6, 447–451.

MOENS, M. AND STEEDMAN, M. 1988. Temporal ontology and temporal reference. *Computational Linguistics 14*, 2, 15–28.

MOHR, C., NADZAM, G., ALAMI, R., SANCHEZ, B., AND CURET, M. 2006. Totally robotic laparoscopic Roux-en-Y Gastric bypass: results from 75 patients. *Obesity Surgery 16*, 6, 690–696.

MORINO, M., BENINCÀ, G., GIRAUDO, G., DEL GENIO, G., REBECCHI, F., AND GARRONE, C. 2004. Robot-assisted vs laparoscopic adrenalectomy: a prospective randomized controlled trial. *Surgical Endoscopy 18*, 12, 1742–1746.

MUDUNURI, R., NEUMUTH, T., STRAUß, G., DIETZ, A., MEIXENSBERGER, J., AND BURGERT, O. 2007. SOCAS-Surgical ontologies for computer assisted surgery. *International Journal of Computer Assisted Radiology and Surgery 2*, sup. 1, 481.

MULYAR, N., VAN DER AALST, W., AND PELEG, M. 2007. A pattern-based analysis of clinical computer-interpretable guideline modeling languages. *Journal of the American Medical Informatics Association 14*, 6, 781.

MÜNCHENBERG, J., BRIEF, J., RACZKOWSKY, J., WÖRN, H., HASSFELD, S., AND MÜHLING, J. 2001a. Operation Planning of robot supported surgical Interventions. In: *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE Computer Society, 547–552.

MÜNCHENBERG, J., BRIEF, J., RACZKOWSKY, J., WÖRN, H., HASSFELD, S., AND MÜHLING, J. 2001b. Operation Planning of robot supported surgical Interventions. In: *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE Computer Society, 547–552.

MYER, E. AND WAGNER, J. 2007. Robotic assisted laparoscopic bladder diverticulectomy. *The Journal of Urology 178*, 6, 2406–2410.

NARA, A., IZUMI, K., ISEKI, H., SUZUKI, T., NAMBU, K., AND SAKURAI, Y. 2010. Surgical workflow analysis based on staff's trajectory patterns. *Workshop on modeling and monitoring of computer-assisted interventions.*

NAZEMI, T., GALICH, A., STERRETT, S., KLINGLER, D., SMITH, L., AND BALAJI, K. 2006. Radical nephrectomy performed by open, laparoscopy with or without hand-assistance or robotic methods by the same surgeon produces comparable perioperative results. *International Brazilian Journal of Urology 32*, 1, 15–22.

NEUMUTH, D., LOEBE, F., HERRE, H., AND NEUMUTH, T. 2011a. Modeling surgical processes: a four-level translational approach. *Artificial Intelligence in Medicine 51*, 3, 147–161.

NEUMUTH, T., CZYGAN, M., STRAUSS, G., MEIXENSBERGER, J., AND BURGERT, O. 2009a. Computer assisted acquisition of surgical process models with a sensor-driven ontology. *Workshop on modeling and monitoring of computer-assisted interventions.*.

NEUMUTH, T., DURSTEWITZ, N., FISCHER, M., ET AL. 2006a. Structured recording of intraoperative surgical workflows. CID 61450A.

NEUMUTH, T., JANNIN, P., SCHLOMBERG, J., MEIXENSBERGER, J., WIEDEMANN, P., AND BURGERT, O. 2011b. Analysis of surgical intervention populations using generic surgical process models. *International Journal of Computer Assisted Radiology and Surgery 6*, 1, 59–71.

NEUMUTH, T., JANNIN, P., STRAUSS, G., MEIXENSBERGER, J., AND BURGERT, O. 2009b. Validation of knowledge acquisition for surgical process models. *Journal of the American Medical Informatics Association 16*, 1, 72–80.

NEUMUTH, T., KASCHEK, B., GOLDSTEIN, D., ET AL. 2010. An observation support system with an adaptive ontology-driven user interface for the modeling of complex behaviors during surgical interventions. *Behavior Research Methods 42*, 4, 1049–1058.

NEUMUTH, T., KRAUSS, A., MEIXENSBERGER, J., AND MUENSTERER, O. 2011d. Impact quantification of the DaVinci telemanipulator system on the surgical workflow using resource impact profiles. *International Journal of Medical Robotics 7*, 2, 156–64.

NEUMUTH, T., LOEBE, F., AND JANNIN, P. 2012. Similarity metrics fo surgical process models. *Artificial Intelligence in Medicine 54*, 1, 15–27.

NEUMUTH, T. AND MEIßNER, C. 2012. Online recognition of surgical instruments by information fusion. *International Journal of Computer Assisted Radiology and Surgery 7*, 2, 297–304.

NEUMUTH, T., MUDUNURI, R., JANNIN, P., MEIXENSBERGER, J., AND BURGERT, O. 2007a. SWAN-Suite: The tool landscape for surgical workflow analysis. In: J. Troccaz and P. Merloz, eds., *Computer Assisted Medical and Surgical Interventions (SURGETICA)*. Sauchamps Medical, Paris, 199–204.

NEUMUTH, T., STRAUSS, G., MEIXENSBERGER, J., LEMKE, H., AND BURGERT, O. 2006b. Acquisition of process descriptions from surgical interventions. Database and Expert Systems Application 2006, Springer, 602–611.

NEUMUTH, T., TRANTAKIS, C., ECKHARDT, F., DENGL, M., MEIXENSBERGER, J., AND BURGERT, O. 2007b. Supporting the analysis of intervention courses with surgical process models on the example of fourteen microsurgical lumbar discectomies. 436–438.

NEUMUTH, T., TRANTAKIS, C., RIFFAUD, L., STRAUSS, G., MEIXENSBERGER, J., AND BURGERT, O. 2009c. Assessment of technical needs for surgical equipment by surgical process models. *Minimally Invasive Therapy & Allied Technologies 18*, 6, 341–349.

NEZHAT, C., LAVIE, O., LEMYRE, M., UNAL, E., NEZHAT, C., AND NEZHAT, F. 2009. Robot-assisted laparoscopic surgery in gynecology: scientific dream or reality? *Fertility and Sterility 91*, 6, 2620–2622.

NISSEN, R. 1956. A simple operation for control of reflux esophagitis. *Schweizerische Medizinische Wochenschrift 86*, 20, 590–592.

NISSEN, R. 1961. Gastropexy and “fundoplication” in surgical treatment of hiatal hernia. *The American Journal of Digestive Diseases 6*, 954–961.

OETTING, T. 2009. Surgical competency in residents. *Current Opinion in Ophthalmology 20*, 1, 56–60.

OFFERMANNS, M. 2002. *Krankenhaus Barometer*. Deutsches Krankenhausinstitut, Düsseldorf.

PADOY, N., BLUM, T., AHMADI, S., FEUSSNER, H., BERGER, M., AND NAVAB, N. 2012. Statistical modeling and recognition of surgical workflow. *Medical Image Analysis 16*, 3, 632–641.

PADOY, N., BLUM, T., ESSA, I., FEUSSNER, H., BERGER, M., AND NAVAB, N. 2007. A boosted segmentation method for surgical workflow analysis. In: N. Ayache, S. Ourselin and A. Maeder, eds., *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2007)*. Springer, Berlin, 102–109.

PADOY, N., HORN, M., FEUSSNER, H., BERGER, M., AND NAVAB, N. Recovery of surgical workflow: a model-based approach. *International Journal of Computer*

Assisted Radiology and Surgery 1, sup. 1.

PANZARASA, S., QUAGLINI, S., CAVALLINI, A., ET AL. 2006. Workflow technology to enrich a computerized clinical chart with decision support facilities. *AMIA Annual Symposium Proceedings 2006*, 619–623.

PARK, S., JEONG, W., CHOI, Y., CHUNG, B., HONG, S., AND RHA, K. 2008. Yonsei experience in robotic urologic surgery-application in various urological procedures. *Yonsei Medical Journal 49*, 6, 897–900.

PARSONS, T. 1994. *Events in the semantics of English: a study in subatomic semantics*. The MIT Press, Boston.

PATEL, V., THALY, R., AND SHAH, K. 2007. Robotic radical prostatectomy: outcomes of 500 cases. *BJU International 99*, 5, 1109–1112.

PATKIN, M. 2003. What surgeons want in operating rooms. *Minimally Invasive Therapy & Allied Technologies 12*, 6, 256–262.

PAYNE, P., MENDONCA, E., JOHNSON, S., AND STARREN, J. 2007. Conceptual konwledge acquisition in biomedicine: A methodological review. *Journal of Biomedical Informatics 40*, 5, 582–602.

PIETROSKI, P. 2005. *Events and semantic architecture*. Oxford University Press, Oxford.

PINON, C. 1997. Achievements in an event semantics. CLC Publications, 273–296.

PITTER, M., ANDERSON, P., BLISSETT, A., AND PEMBERTON, N. 2008. Robotic-assisted gynaecological surgery-establishing training criteria; minimizing operative time and blood loss. *The International Journal of Medical Robotics + Computer Assisted Surgery 4*, 2, 114–120.

PLOUSSARD, G., XYLINAS, E., PAUL, A., ET AL. 2009. Is robot assistance affecting operating room time compared with pure retroperitoneal laparoscopic radical prostatectomy? *Journal of Endourology 23*, 6, 939–943.

POSTGRESQL GLOBAL DEVELOPMENT GROUP. 2009. *PostgreSQL: The world's most advanced open source database*. .

POSTGRESQL GLOBAL DEVELOPMENT GROUP. 2010. PostgreSQL: The world's most advanced open source database. http://www.postgresql.org/.

POULYMENOPOULOU, M., MALAMATENIOU, F., AND VASSILACOPOULOS, G. 2003. Emergency healthcare process automation using workflow technology and web services. *Medical Informatics and the Internet in Medicine 28*, 3, 195–207.

PRAKASH, G., JHANJI, V., SHARMA, N., GUPTA, K., TITIYAL, J., AND VAJPAYEE, R. 2009. Assessment of perceived difficulties by residents in performing routine steps in phacoemulsification surgery and in managing complications. *Canadian Journal of*

Ophthalmology 44, 3, 284–287.

PRINYAPOL, N., LAU, S., AND FAN, J. 2009. A dynamic nursing workflow management system: a Thailand hospital scenario. In: X. Huang, S. Ao and O. Castillo, eds., *Intelligent Automation and Computer Engineering*. Springer Netherlands, Dordrecht, 489–501.

PRIVETT, B., GREENLEE, E., ROGERS, G., AND OETTING, T. 2010. Construct validity of a surgical simulator as a valid model for capsulorhexis training. *Journal of Cataract and Refractive Surgery 36*, 11, 1835–1838.

QI, J., JIANG, Z., ZHANG, G., MIAO, R., AND SU, Q. 2006. A surgical management information system driven by workflow. 1014–1018.

QUAGLINI, S., STEFANELLI, M., CAVALLINI, A., MICIELI, G., FASSINO, C., AND MOSSA, C. 2000. Guideline-based careflow systems. *Artificial Intelligence in Medicine 20*, 1, 5–22.

QUAGLINI, S., STEFANELLI, M., LANZOLA, G., CAPORUSSO, V., AND PANZARASA, S. 2001. Flexible guideline-based patient careflow systems. *Artificial Intelligence in Medicine 22*, 1, 65–80.

R FOUNDATION. 2010. *The R project for statistical computing*. .

RACER SYSTEMS. 2009. *RacerPro - Renamed abox and concept expression reasoner*.

RAIMBAULT, M., MORANDI, X., AND JANNIN, P. 2005. Towards models of surgical procedures: analyzing a database of neurosurgical cases. In: O. Ratib and S. Horii, eds., *Medical Imaging 2005: PACS and imaging informatics*. SPIE, Bellingham, WA, 97–104.

REICHERT, M., DADAM, P., MANGOLD, R., AND KREIENBERG, R. 2000. Computer support of workflow in the hospital: concepts, technology and application. *Zentralblatt Für Gynäkologie 122*, 1, 53–67.

REITER, R. 1991. The frame problem in situation the calculus: a simple solution (sometimes) and a completeness result for goal regression. In: *Artificial intelligence and mathematical theory of computation: papers in honor of John McCarthy*. Academic Press Professional, Inc., 359–380.

RENEMAN, M., FOKKENS, A., DIJKSTRA, P., GEERTZEN, J., AND GROOTHOFF, J. 2005. Testing lifting capacity: validity of determining effort level by means of observation. *Spine 30*, 2, E40–46.

RICHARDS, C., ROSEN, J., HANNAFORD, B., PELLEGRINI, C., AND SINANAN, M. 2000. Skills evaluation in minimally invasive surgery using force/torque sigantures. *Surgical Endoscopy 14*, 9, 791–798.

RIDDER, L. 2002. *Mereologie. Ein Beitrag zur Ontologie und Erkenntnistheorie*. Klostermann, Frankfurt am Main.

RIEDL, S. 2003. Modern operating room management in the workflow of surgery. Spectrum of tasks and challenges of the future. *Der Anaesthesist 52*, 10, 957–963.

RIFFAUD, L., NEUMUTH, T., MORANDI, X., ET AL. 2010. Recording of surgical processes: a study comparing senior and junior neurosurgeons during lumbar disc herniation surgery. *Neurosurgery 67*, 2, 325–332.

RODRÍGUEZ, E., KYPSON, A., MOTEN, S., NIFONG, L., AND CHITWOOD, W. 2006. Robotic mitral surgery at East Carolina University: a 6 year experience. *The International Journal of Medical Robotics + Computer Assisted Surgery 2*, 3, 211–215.

ROGERS, G., OETTING, T., LEE, A., ET AL. 2009. Impact of a structured surgical curriculum on ophthalmic resident cataract surgery complication rates. *Journal of Cataract and Refractive Surgery 35*, 11, 1956–1960.

ROSEN, J., BROWN, J., CHANG, L., AND HANNAFORD, B. 2006. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. *IEEE Transactions on Biomedical Engineering 53*, 3, 399–413.

ROSEN, J., HANNAFORD, B., RICHARDS, C., AND SINANAN, M. 2001. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *IEEE Transactions on Biomedical Engineering 48*, 5, 579–591.

ROSENBLOOM, S., MILLER, R., JOHNSON, K., ELKIN, P., AND BROWN, S. 2008. A model for evaluating interface terminologies. *Journal of the American Medical Informatics Association 15*, 1, 65–76.

ROSENBLOOM, S.T., MILLER, R.A., JOHNSON, K.B., ELKIN, P.L., AND BROWN, S.H. 2006. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *Journal of the American Medical Informatics Association 13*, 3, 277–288.

ROSENKRANZ, F. 2005. *Geschäftsprozesse: Modell- und computergestützte Planung*. Springer, Heidelberg.

ROSSE, C. AND MEJINO, J. 2003. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics 36*, 6, 478–500.

ROTHSTEIN, S. 1999. Fine-grained structure in the eventuality domain: the semantics of predicative adjective phrases and be. *Natural Language Semantics 7*, 4, 347–420.

ROZINAT, A. AND VAN DER AALST, W. 2008. Conformance checking of processes based on monitoring real behavior. *Information Systems 33*, 1, 64–95.

ROZINAT, A., DE MEDEIROS, A., GUNTHER, C., WEIJTERS, A., AND VAN DER AALST, W. 2007. Towards an evaluation framework for process mining algorithms. In: *BPM'07 Proceedings of the 2007 international conference on Business process management*. Springer, Berlin, Heidelberg.

RUMBAUGH, J., JACOBSON, I., AND BOOCH, G. 2005. *The unified modeling language reference manual*. Addison-Wesley Longman, Amsterdam, Reading, MA.

RYLE, G. 1984. *The concept of mind*. University Of Chicago Press, Chicago.

SACKETT, D., ROSENBERG, W., GRAY, J., HAYNES, R., AND RICHARDSON, W. 1996. Evidence based medicine: what it is and what it isn't. *British Medical Journal 312*, 7023, 71 –72.

SALEH, G., GAUBA, V., MITRA, A., LITWIN, A., CHUNG, A., AND BENJAMIN, L. 2007. Objective structured assessment of cataract surgical skill. *Archives of Ophthalmology 125*, 3, 363–366.

SANDBERG, W., GANOUS, T., AND STEINER, C. 2003. Setting a research agenda for perioperative systems design. *Seminars in Laparoscopic Surgery 10*, 2, 57–70.

SANGHA, O., SCHNEEWEIS, S., WILDNER, M., ET AL. 2002. Metric properties of the appropriateness evaluation protocol and predictors of inappropriate hospital use in Germany: an approach using longitudinal patient data. *International Journal for Quality in Health Care 14*, 6, 483–492.

SANTERRE, N., BLONDEL, F., RACOUSSOT, F., ET AL. 2007. A teaching medical simulator: phacoemulsification in virtual reality. *Journal Français D'ophtalmologie 30*, 6, 621–626.

SARKAR, A., DUTTA, A., DHINGRA, U., ET AL. 2006. Development and use of behavior and social interaction software installed on Palm handheld for observation of a child's social interactions with the environment. *Behavior Research Methods 38*, 3, 407–415.

SCHALLHORN, J., TANG, M., LI, Y., SONG, J., AND HUANG, D. 2008. Optical coherence tomography of clear corneal incisions for cataract surgery. *Journal of Cataract and Refractive Surgery 34*, 9, 1561–1565.

SCHEER, A. 1997. *Wirtschaftsinformatik: Referenzmodelle für industrielle Geschäftsprozesse*. Springer, Heidelberg.

SCHEER, A. 1999. *ARIS - vom Geschäftsprozess zum Anwendungssystem*. Springer, Heidelberg.

SCHIMM, G. 2004. Mining exact models of concurrent workflows. *Computers in Industry 53*, 3, 265–281.

SCHLENOFF, C., GRUNINGER, M., TISSOT, F., VALOIS, J., LUBELL, J., AND LEE, J. 1999. The process specification language (PSL) overview and version 1.0 specification. .

SCHUHMANN, T. 2008. Hospital financial performance: trends to watch: financial indicators derived from Medicare cost report data are reliable tools for assessing the effectiveness of a hospital's operations. *Healthcare Financial Performance 62*, 7, 59–66.

SCHUSTER, M., WICHA, L., FIEGE, M., AND GOETZ, A. 2007. Utilization rates and turnover times as indicators of OR workflow efficiency. *Der Anaesthesist 56*, 10, 1060–1066.

SEEBURGER, J., LEONTJEV, S., NEUMUTH, T., ET AL. 2012. Trans-apical beating-heart implantation of neo-chordae to mitral valve leaflets: results of an acute animal study. *European Journal of Cardio-Thoracic Surgery 41*, 1, 173–176.

SEIBT, J. 2007. *Process theories: crossdisciplinary studies in dynamic categories*. Springer Netherlands, Amsterdam.

SHAH, J. AND DARZI, A. 2003. The impact of inherent and environmental factors on surgical performance in laparoscopy: a review. *Minimally Invasive Therapy & Allied Technologies 12*, 1-2, 69–75.

SIM, H., YIP, S., LAU, W., TAN, Y., WONG, M., AND CHENG, C. 2006. Team-based approach reduces learning curve in robot-assisted laparoscopic radical prostatectomy. *International Journal of Urology 13*, 5, 560–564.

SJOERDSMA, W., MEIJER, D., JANSEN, A., DEN BOER, K., AND GRIMBERGEN, C. 2000. Comparison of efficiencies of three techniques for colon surgery. *Journal of Laparoendoscopic & Advanced Surgical Techniques 10*, 1, 47–53.

SONG, J., KANG, W., OH, S., HYUNG, W., CHOI, S., AND NOH, S. 2009. Role of robotic gastrectomy using da Vinci system compared with laparoscopic gastrectomy: initial experience of 20 consecutive cases. *Surgical Endoscopy 23*, 6, 1204–1211.

SPSS INC. 2008. *SPSS 15.0*. Chicago.

SPSS INC. 2010. SPSS, Data Mining, Statistical Analysis Software, Predictive Analysis, Predictive Analytics, Decision Support Systems. http://www.spss.com/de/.

STAHL, J., EGAN, M., GOLDMAN, J., ET AL. 2005. Introducing new technology into the operating room: measuring the impact on job performance and satisfaction. *Surgery 137*, 5, 518–526.

STANFORD UNIVERSITY. 2009. *The Protégé ontology editor and knowledge acquisition system*. .

STITZENBERG, K. AND SHELDON, G. 2005. Progressive specialization within general surgery: adding to the complexity of workforce planning. *Journal of the American College of Surgeons 201*, 6, 925–932.

STRAUß, G., FISCHER, M., MEIXENSBERGER, J., ET AL. 2006a. Workflow analysis to assess the efficiency of intraoperative technology using the example of functional endoscopic sinus surgery. *HNO 54*, 7, 528–535.

STRAUß, G., KOULECHOV, K., RICHTER, R., ET AL. 2005a. Navigated control: a new concept in computer assisted ENT-surgery. *Laryngorhinootologie 84*, 8, 567–576.

STRAUß, G., KOULECHOV, K., RICHTER, R., DIETZ, A., TRANTAKIS, C., AND LÜTH, T. 2005b. Navigated control in functional endoscopic sinus surgery. *The International Journal of Medical Robotics + Computer Assisted Surgery 1*, 3, 31–41.

STRAUß, G., KOULECHOV, K., RÖTTGER, S., ET AL. 2006b. Ist der Vorteil eines Navigationssystems in der HNO-Chirurgie messbar? *HNO 54*, 12, 947–957.

SUDRA, G., BECKER, A., BRAUN, M., SPEIDEL, S., MUELLER-STICH, B., AND DILLMANN, R. 2009. Estimating similarity of surgical situations with case-retrieval-nets. *Studies in Health Technology and Informatics 142*, 358–363.

SUDRA, G., SPEIDEL, S., MÜLLER-STICH, B., BECKER, A., OTT, M., AND DILLMANN, R. 2007. Situation modelling and situation recognition for a context-aware augmented reality system. *International Journal of Computer Assisted Radiology and Surgery 2*, sup. 1, 434–436.

SUNYAEV, A., LEIMEISTER, J., SCHWEIGER, A., AND KRCMAR, H. 2006. Integrationsarchitekturen für das Krankenhaus. *IMC-Information Management & Consulting 21*, 1, 28–35.

SUTHERLAND, J. AND VAN DEN HEUVEL, W. 2006. Towards an intelligent hospital environment: adaptive workflow in the OR of the future. In: *Proceedings of the 39th Annual Hawaii International Conference on System Siences 2006*. IEEE, 100b–100b.

SUTHERLAND, J., VAN DEN HEUVEL, W., GANOUS, T., BURTON, M., AND KUMAR, A. 2005. Towards an intelligent hospital environment: OR of the future. *Studies in Health Technology and Informatics 118*, 278–312.

TANABE, N., GO, K., SAKURADA, Y., ET AL. 2011. A remote operating slit lamp microscope system. Development and its utility in ophthalmologic examinations. *Methods of Information in Medicine 50*, 5, 427–434.

TAYLOR, J., BINENBAUM, G., TAPINO, P., AND VOLPE, N. 2007. Microsurgical lab testing is a reliable method for assessing ophthalmology residents' surgical skills. *The British Journal of Ophthalmology 91*, 12, 1691–1694.

TEN TEIJE, A., MIKSCH, S., AND LUCAS, P. 2008. *Computer-based medical guidelines and protocols*. IOS Press, Amsterdam.

TERENZIANI, P., GERMAN, E., AND SHAHAR, Y. 2008. The temporal aspects of clinical guidelines. *Studies in Health Technology and Informatics 139*, 81–100.

TERENZIANI, P. AND TORASSO, P. 1995. Time, action-types, and causation: an integrated analysis. *Computational Intelligence 11*, 3, 529–552.

THIEL, D. AND WINFIELD, H. 2008. Robotics in urology: past, present, and future. *Journal of Endourology / Endourological Society 22*, 4, 825–830.

THIELSCHER, M. 2010. A unifying action calculus. *Artificial Intelligence 175*, 1, 120–

141.

THÜSING, C. 2005. Qualitätsmanagement im Krankenhaus. *Medizinische Klinik 100*, 3, 149–153.

VAN DER TOGT, R., VAN LIESHOUT, E., HENSBROEK, R., BEINAT, E., BINNEKADE, J., AND BAKKER, P. 2008. Electromagnetic interference from radio frequency identification inducing potentially hazardous incidents in critical care medical equipment. *Journal of the American Medical Association 299*, 24, 2884 –2890.

TRABULSI, E., LINDEN, R., GOMELLA, L., MCGINNIS, D., STRUP, S., AND LALLAS, C. 2008. The addition of robotic surgery to an established laparoscopic radical prostatectomy program: effect on positive surgical margins. *The Canadian Journal of Urology 15*, 2, 3994–3999.

VARADARAJAN, B., REILEY, C., LIN, H., KHUDANPUR, S., AND HAGER, G. 2009. Data-derived models for segmentation with application to surgical assessment and training. In: G. Yang, D. Hawkes, D. Rueckert, J. Noble and C. Taylor, eds., *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2009)*. Springer, Berlin, 426–434.

VAWDREY, D., GARDNER, R., EVANS, R., ET AL. 2007. Assessing data quality in manual entry of ventilator settings. *Journal of the American Medical Informatics Association 14*, 3, 295–303.

VENDLER, Z. 1967. *Linguistics in philosophy*. Cornell University Press, Ithaca.

VILLAVICENCIO MAVRICH, H., ESQUENA FERNÁNDEZ, S., SALVADOR BAYARRI, J., ET AL. 2009. Robotic radical prostatectomy: first 100 patients in Fundació Puigvert. *Actas Urologicas Españolas 33*, 1, 24–29.

WANG, G., BAROCAS, D., RAMAN, J., AND SCHERR, D. 2008. Robotic vs open radical cystectomy: prospective comparison of perioperative outcomes and pathological measures of early oncological efficacy. *BJU International 101*, 1, 89–93.

WAQAR, S., PARK, J., KERSEY, T., MODI, N., ONG, C., AND SLEEP, T. 2011. Assessment of fatigue in intraocular surgery: analysis using a virtual reality simulator. *Graefe's Archive for Clinical and Experimental Ophthalmology 249*, 1, 77–81.

WEBSTER, R., SASSANI, J., SHENK, R., ET AL. 2005. Simulating the continuous curvilinear capsulorhexis procedure during cataract surgery on the EYESI system. *Studies in Health Technology and Informatics 111*, 592–595.

WEILER, T., HOFFMANN, R., AND STREHLAU-SCHWOLL, H. 2003. Qualitätsmanagement und Zertifizierung. *Der Unfallchirurg 106*, 8, 692–697.

WEINSTEIN, G., O'MALLEY, B., DESAI, S., AND QUON, H. 2009. Transoral robotic surgery: does the ends justify the means? *Current Opinion in Otolaryngology & Head and Neck Surgery 17*, 2, 126–131.

WENTE, M., SEILER, C., UHL, W., AND BÜCHLER, M. 2003. Perspectives of evidence-based surgery. *Digestive Surgery 20*, 4, 263–269.

WEST, M., SULLIVAN, J., AND TEIJGELER, H. 2003. ISO/FDIS 15926-2: Lifecycle integration of process plant data including oil and gas production facilities. http://www.tc184-sc4.org/wg3ndocs/wg3n1328/lifecycle_integration_schema.html.

WESTBROOK, J. AND AMPT, A. 2009. Design, application and testing of the work observation method by activity timing (WOMBAT) to measure clinicians' patterns of work and communication. *International Journal of Medical Informatics 78*, sup. 1, S25–33.

WHITE, S. 2004. Introduction to BPM. .

WHITE, S. AND MIERS, D. 2008. *BPMN modeling and reference guide: understanding and using BPMN*. Future Strategies Inc., Lighthouse Point, FL.

WORKFLOW MANAGEMENT COALITION. 1999a. Terminology & Glossary - Document Number WFMC-TC-1011, Document Status - Issue 3.0. .

WORKFLOW MANAGEMENT COALITION. 1999b. *Terminology & Glossary - Document Number WFMC-TC-1011, Document Status - Issue 3.0*. Winchester, UK.

WORKFLOW MANAGEMENT COALITION (WFMC). 1995. The workflow reference model. .

WORLD HEALTH ORGANIZATION. 2008. International classification of dieseases. *International classification of dieseases*. http://www.who.int/classifications/icd/en.

XIAO, Y., HU, P., HU, H., ET AL. 2005. An algorithm for processing vital sign monitoring data to remotely identify operating room occupancy in real-time. *Anesthesia and Analgesia 101*, 3, 823–829.

XIONG, N. AND SVENSSON, P. 2002. Multi-sensor management for information fusion: issues and approaches. *Information Fusion 3*, 2, 163–186.

YEE, D., SHANBERG, A., DUEL, B., RODRIGUEZ, E., EICHEL, L., AND RAJPOOT, D. 2006. Initial comparison of robotic-assisted laparoscopic versus open pyeloplasty in children. *Urology 67*, 3, 599–602.

ZAI, A., GRANT, R., ESTEY, G., ET AL. 2008. Lessons from implementing a combined workflow-informatics system for diabetes management. *Journal of the American Medical Informatics Association 15*, 4, 524–533.

ZHANG, J., LU, X., NIE, H., HUANG, Z., AND VAN DER AALST, W. 2009. Radiology information system: a workflow-based approach. *International Journal of Computer Assisted Radiology and Surgery 4*, 5, 509–516.

ZHOU, W. AND PIRAMUTHU, S. 2010. Healthcare process mining with RFID. In: S. Rinderle-Ma, S. Sadiq and F. Leymann, eds., *Proceedings of Business Process*

Management Workshops 2009. Springer Berlin Heidelberg, Berlin, Heidelberg, 405–411.

ZIARKO, W. AND SHAN, N. 1996. A method for computing all maximally general rules in attribute-value systems. *Computational Intelligence 12*, 2, 223–234.

8 Glossary

Term	Comment
bottom-up modeling	Modeling by generalization of specific entities.
DaVinci telemanipulator	Surgical telemanipulator system of the vendor Intuitive Surgical Inc, USA.
generic surgical process model (gSPM)	A type of a surgical process model. The gSPM is a statistical mean model of several individual surgical process models.
individual surgical process model (iSPM)	A type of a surgical process model. The iSPM is a model of one surgical case resp. one surgical process.
information and communication technology (ICT)	General term to name systems and strategies concerning information management and communication by the help of computers.
ontology	A specification of an conceptualization (39).
radio frequency identification (RFID)	Radio-wave based technology for wireless information transmission.
resource impact profile (RIP)	A gSPM that was generated from iSPMs that were selected according to the same resource or material usage.
surgeon-individual treatment profile (SiTP)	A gSPM that was generated from iSPMs that were selected according to the same performing surgeon.
surgical assist system (SAS)	A technical system that supports the surgeon in his/her work.
surgical process (SP)	Term used for a surgical intervention/a surgical case.
surgical process model (SPM)	A simplified pattern of a surgical process that reflects a predefined subset of interest of the SP in a formal or semiformal representation.
surgical workflow management system (SWFMS)	A workflow management system to support a surgical intervention.
top-down modeling	Modeling by refinement.
workflow management	The management of a surgical process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules (adapted from (40)).
workflow management system (WFMS)	A system that defines, creates and manages the execution of workflows through the use of software, running on one or more workflow engines, which is able to interpret the process definition, interact with workflow participants and, where required, invoke the use of IT tools and applications (40).
workflow schema	A model-like description to specify the behavior of a workflow management system.

ISBN 978-3-00-038630-5