Methods for Automatic Statistical Modeling of Surgical Workflow

Tobias Blum
blum@in.tum.de

Nassir Navab
navab@in.tum.de

Technische Universität München
Computer Aided Medical Procedures,
Boltzmannstr. 3, 85748 Garching, Germany

Hubertus Feußner
Technische Universität München
Department of Surgery, Klinikum Rechts der Isar,
Ismaninger Str. 22, 81675 München, Germany
feussner@chir.med.tu-muenchen.de

ABSTRACT
An increasing amount of data can be recorded automatically during surgeries. This has led to a recent interest in methods that allow statistical modeling of surgical workflow from this data. One main challenge of these methods is how to automatically generate a statistical model from a set of recorded surgeries. In this paper we will review different methods and show advantages and drawbacks.

Author Keywords
Surgical workflow, statistical models of time series.

INTRODUCTION
Trough the introduction of new technology into the operating room it is now possible to record an increasing amount of signals during a surgery. Video images can be obtained, especially for laparoscopic surgeries, position and orientation data is available from tracking systems or robotic systems. Motion sensors, like accelerometers, are becoming cheap and small and technologies like RFID might be used in the future to identify instruments and estimate their presence. Also training systems that use e.g. haptic feedback devices are an important source of data.

As this kind of information becomes available, there is recent interest in methods to make use of this data. Possible applications are analysis of behavior and workflows, assessment of surgical skill and context-sensitive systems that can detect the current work step of a surgery. All these applications require some sort of statistical model of the workflow. Automatically building such models is an important topic in this domain. We will review and compare different methods for this task.

METHODS
The most common method for modeling surgical workflow is to use Hidden Markov Models (HMM). HMMs model time series by a directed graph. Each node or state in the graph has transition probabilities representing the probability of advancing to other states. Each state also has an observation symbol distribution that represents the probability of observing a certain observation symbol in that state. For most applications in the field of surgical workflow modeling one tries to build a model from a set of training data such that the model represents the training data well. We will discuss several methods to automatically build HMMs. We use data that represents the use of instruments during ten laparoscopic cholecystectomies. Each surgery was split into 14 phases. The models are trained independently for each phase. For each method, the capabilities to automatically detect the current phase of a surgery and to generate models that are understandable by humans are discussed. Results are summarized in Table 1.

The most common method to train a HMM is to initialize it with random transition and observation probabilities as for example done in [1]. Next expectation maximization (EM) is used to adapt the HMM to the training data. This procedure is repeated multiple times until a good HMM is found. Many extensions of this method have been proposed e.g. to estimate the number of states that are required to model a certain procedure. These methods work, are easy to implement and do not require long computation time. Furthermore they have shown good results for automatically detecting the phase of a workflow. However the resulting model is usually hardly understandable to humans, which makes them practically useless for analysis of human behavior.

In the case of discrete observations, like the use of instruments, Markov Chains are an alternative. They are similar to HMMs and have been used for workflow modeling in [2]. While states of an HMM can represent the use of different instruments, a Markov Chain state represents only one instrument. Markov Chains can be built simply by counting transitions from one observation symbol to another. They give a relative good impression of

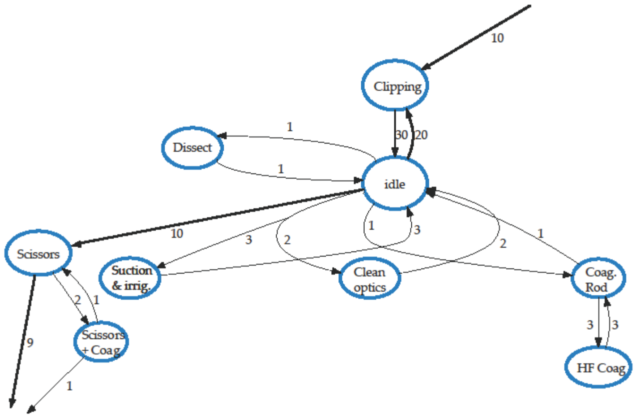

Figure 1. Markov chain representing one phase of a laparoscopic cholecystectomies.

the frequency of instrument use and are much easier to understand for humans then randomly initialized HMMs. Also automatic recognition rates are good. However they do not model the temporal order of actions within one phase are do therefore not model repetitions well. An example can be seen in Figure 1.

The methods that showed best results for generating human-understandable models are data-derived HMMs. One example is successive state splitting (SSS) [3]. This method starts with a model consisting only of one state and splits states until a good model is found. So e.g. if we have one state that represents that first one instrument is used and next another, this method will do a temporal split so that two successive states are used instead. If one state represents alternating use of two instruments it will do a contextual split into two states. While this method delivers models that are intuitive to understand, as can be seen in Figure 2, for our data it failed to deliver models that statistically well represent the workflow and are therefore not suited for automatic recognition systems.

Another data-derived method that works the other way round is model merging [4], where an initial very big model is built that represents every single action that occurred in any of the training surgeries. Iteratively, a smaller model is generated by merging similar nodes. As SSS, model merging well represents the temporal order of actions and is also well suited to detect workflow steps. The main drawback of this method is that it takes several hours to train it, even for a relative small training set consisting only of nine surgeries.

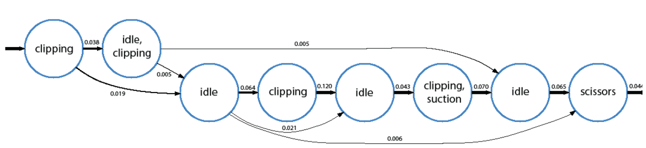

Figure 2. HMM that has been generated by model merging.

Model	Detection	Speed	Understandable
Random	6,7%	fast	bad
Markov Chain	5,3%	very fast	medium
SSS	12,5%	medium	good
Merging	6,4%	slow	good

Table 1. Results from different methods.

RESULTS

We have compared different aspects of the models. One interesting application is to build context-aware system. In order to estimate whether the methods are suitable for this application, we tested whether it is possible to detect the current phase of a running surgery. This measure also indicates if the model is a good statistical representation of the workflow. All methods but SSS have shown to deliver reasonable results here. Another important metric is whether a certain method can create models that reflect the workflow in a human understandable way. This is true for the data-derived methods. However, from our experience it has shown that these methods are harder to implement and unlike for standard methods, there are no toolkits or libraries available where these methods are already implemented.

CONCLUSION

Automatically building statistical methods is important, especially as a huge amount of data is becoming available, which makes manual modeling difficult. Using the right training methods, HMMs can be built that are suitable for automatic recognition systems and for human understandable models. Challenges in the future will be to extend these models such that they work on huge datasets and to incorporate different kinds of sensors at the same time. Future models should also take into account patient or surgeon related information.

REFERENCES

1. Leong , J., Nicolaou, M., Atallah, L., Mylonas, G., Darzi, A., Yang, G-Z. HMM Assessment of Quality of Movement Trajectory in Laparoscopic Surgery. ***MICCAI*** (2006), 335-346.
2. Mackel, T.R., Rosen, J., Pugh, C. M. Markov Model Assessment of Subjects' Clinical Skill Using the E-Pelvis Physical Simulator. ***IEEE Trans Biomed Eng, 5***, 12 (2007), 2133-2141.
3. Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., Hager, G. Data-Derived Models for Segmentation with Application to Surgical Assessment and Training, ***MICCAI*** (2009), 426-434.
4. Blum, T., Padoy, N., Feußner, H., Navab, N. Modeling and Online Recognition of Surgical Phases using Hidden Markov Models. ***MICCAI*** (2008), 379-386.