Towards unified dataset for Modeling and Monitoring of Computer Assisted Medical Interventions

Fabien Despinoy, Sandrine Voros, Nabil Zemiti, Nicolas Padoy, Pierre Jannin

Towards unified dataset for Modeling and Monitoring of Computer Assisted Medical Interventions

Fabien DESPINOY [1], Sandrine VOROS [2], Nabil ZEMITI [3], Nicolas PADOY [4] and Pierre JANNIN [1]

1 : LTSI, UMR 1099, University of Rennes 1, Inserm, F-35043, France
2 : TIMC-IMAG, UMR 5525, University of Grenoble-Alpes, CNRS, F-38706, France
3 : LIRMM, UMR 5506, University of Montpellier, Inserm, F-34095, France
4 : ICube, UMR 7357, University of Strasbourg, CNRS, F-67000, France
Contact: fabien.despinoy@univ-rennes1.fr

M2CAMI is a working group dedicated to Modeling and Monitoring of Computer Assisted Medical Interventions within the CAMI Labex. It aims at unifying data acquired from different surgical trainers and procedures for collaborative research. In this paper, we propose a generic structure for multimodal dataset that allows faster development and easier processing. With such formalization, our objective is to go beyond the state of the art by sharing various types of data between international institutions and merge methodological approaches for better detection and understanding of surgical workflows.

1 Introduction

For the past decade, the medical community has shown a growing interest in biomedical international competitions, named challenges. Since 2007, more than 120 challenges have been proposed on the grand-challenge platform[1] where various research fields are addressed such as image segmentation and registration, anatomical structure localization and tracking or surgical workflow detection and understanding. Through such platform, multiple types of data are shared going from point cloud depth image acquisition to CT or MRI segmented images. These challenges share the same goal: use pre-clinical and clinical data in order to validate and evaluate methodologies that are developed for biomedical analysis purposes.

[1] `https://grand-challenge.org/`

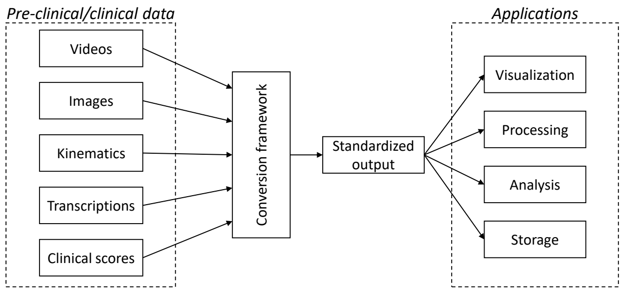
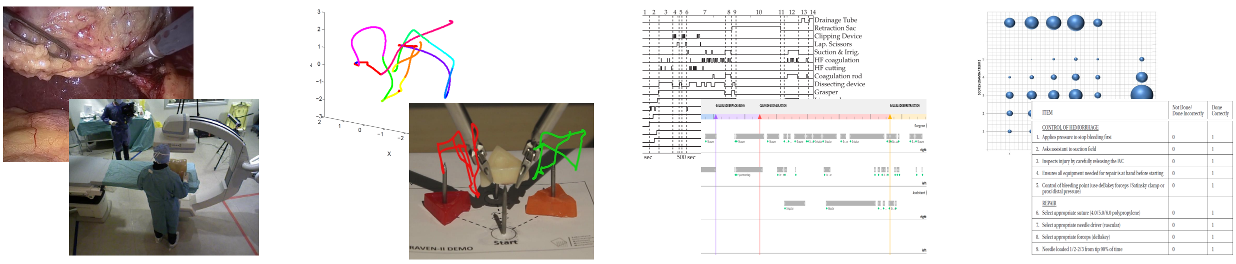


Figure 1: *Pipeline for data conversion into a standardized output for M2CAMI applications.*

Focusing on surgical workflow detection and understanding, various datasets have been released over the last years, where multiple types of data are accessible. For each dataset, data representation and storage are unique because of the acquisition setup and the personal requirements. However, this sharing model is not suitable for collaborative research due to the heterogeneous representation of the same data type. Moreover, such approach does not promote the use of multimodal data for multiscale processing because of the complex integration of various data types and formats.
In this paper, we address the need of a standard formalism for multimodal data representation and storage. By taking into account the available datasets in the M2CAMI community, we propose a generic structure and a first implementation for data conversion to a generic format that is readable and usable in the different applications developed into the M2CAMI group (fig. 1).



(a) *Video data.* **(b)** *Kinematic data.* **(c)** *Transcription data.* **(d)** *Evaluation data.*

Figure 2: *Multimodal data of surgical workflow available into the M2CAMI collaborative research group.*

2 Methods

In the M2CAMI group, data acquisition and analysis are oriented towards surgical workflow detection and understanding. In order to build a unified framework to convert data into a generic representation, we proposed a two-step approach. The first step was the inventory of the different available data into the working group. Then, in the second step, we described a common formalism for data representation and storage. From the M2CAMI community, two datasets are publicly available: NeuroSurgicalTools [1] and Cholec80 [2] where the latter has been already used for a previous scientific challenge[2]. To generalize our work, we used also three other available datasets: JIGSAWS [3] as well as acquisitions performed during previous CAMI works [4][5].

2.1 Data inventory

These datasets contain various data types that can be sorted into four distinct categories. Videos are (fig. 2a) automatically captured from the surgical field by recording the endoscopic point of view, or from the entire operating room with additional cameras. They can also be expressed as a set of images. Kinematic data (fig. 2b) is automatically captured from robotic systems or additional tracking devices to describe the surgical instrument motions over the time. Transcription data (fig. 2c), or annotations, is most of the time manually generated by human observers and describes the different actions, instruments or events taking place during the procedure. Evaluation data (fig. 2d) quantifies performance of the surgeon to achieve a clinical task or procedure. It can be automatically calculated from surgical trainers or manually generated by human experts and is necessary for surgical workflow referenced-based evaluations. As an extension to these four categories, unusual clinical data types are emerging and combine annotations with kinematics such as human body pose annotations for medical staff analysis [6]. To cover the full spectrum from training to interventional clinical data available into the M2CAMI group, we developed an inventory form based on a previous formalism [7] to make a list of the different data and structures available.

[2] `http://camma.u-strasbg.fr/m2cai2016/`

2.2 Generic data representation

As previously mentioned, a same data type can be represented in different ways. For instance, video is a set of images that is embedded into a streaming container. However, depending on the format, the encoder and the output container it can be difficult to use it for processing purpose. Thus, our objective is to define a standard representation that rely on open source protocols (e.g. x265 for video codec, Boost.PropertyTree for configuration file) to ease broadcast of our developments. This standard representation can be developed at two different scales. At the global level, we provide guidelines and tools to uniformly structure dataset in order to ease sharing, processing and hard data storage. Additionally, standardized header files are integrated into datasets to regroup data properties for conversion management. At a second level, we propose a C++ language-based container that stores data synchronously into memory for processing (e.g. as an extended map for data series). The main advantage of the presented structure is to be easily extended to add new types of data. The only assumption we rely on is the data synchronization where for each timestamp in the timeline there is at least one sample from each data type available in the dataset.

3 Conclusion and Perspectives

At the current stage, data conversion seems feasible to provide larger datasets for each partner of the M2CAMI group. Relying on open source protocols, the current implementation is shareable between members and improvements can be easily undertaken. The added value of the current work will be emphasized by the connection of various developments coming from different teams, including surgical context detection [2] that will provide entry frame for instrument tracking [8]. The output information from the tool tracking will be used for gesture recognition [9] in order to, at the end, obtain a complete comprehension of the surgical context. These are the first steps towards generation of new decision-making systems where multimodal data allows to capture and understand surgical events for training, teaching and surgical assistance purposes.

References

[1] D. Bouget, R. Benenson, M. Omran, L. Riffaud, B. Schiele and P. Jannin (2015). Detecting Surgical Tools by Modelling Local Appearance and Global Shape. *IEEE Transactions on Medical Imaging*, 34(12):2603–2617.

[2] A.P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. de Mathelin and N. Padoy (2016). EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Video. *IEEE Transactions on Medical Imaging*, 36(1):86–97.

[3] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B.Béjar, D. D. Yuh, C. C. G. Chen, R. Vidal, S. Khudanpur and G. D. Hager (2014). The JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling. *Modeling and Monitoring of Computer Assisted Interventions (M2CAI) - MICCAI Workshop.*

[4] R. Wolf (2014). Quantification de la qualité d'un geste chirurgical à partir de connaissances a priori. *Université Grenoble Alpes.*

[5] F. Despinoy (2016). Analyse, reconnaissance et réalisation des gestes pour l'entraînement en chirurgie laparoscopique robotisée. *Université de Montpellier.*

[6] A. Kadkhodamohammadi, A. Gangi, M. De Mathelin and N. Padoy (2017). Articulated clinician detection using 3D pictorial structures on RGB-D data. *Medical Image Analysis* 35:215–224.

[7] P. Jannin, C. Grova and C. R. Maurer Jr. (2006). 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.

[8] A. Agustinos and S. Voros (2015). 2D/3D Real-Time Tracking of Surgical Instruments Based on Endoscopic Image Processing. *Computer-Assisted and Robotic Endoscopy*, 9515:90–100.

[9] F. Despinoy, D. Bouget, G. Forestier, C. Penet, N. Zemiti, P. Poignet and P. Jannin (2016). Unsupervised Trajectory Segmentation for Surgical Gesture Recognition in Robotic Training. *IEEE Transactions on Biomedical Engineering*, 63(6):1280–1291.