Surgical Phases Detection from Microscope Videos by Combining SVM and HMM

Florent Lalys[1,2,3], Laurent Riffaud[4], Xavier Morandi[1,2,3,4], and Pierre Jannin[1,2,3]

[1] INSERM, U746, Faculté de Médecine CS 34317, F-35043 Rennes Cedex, France
[2] INRIA, VisAGeS Unité/Projet, F-35042 Rennes, France
[3] University of Rennes I, CNRS, UMR 6074, IRISA, F-35042 Rennes, France
[4] Department of Neurosurgery, Pontchaillou University Hospital, F-35043 Rennes, France

Abstract. In order to better understand and describe surgical procedures by surgical process models, the field of workflow segmentation has recently emerged. It aims to recognize high-level surgical tasks in the Operating Room, with the help of sensors or human-based systems. Our approach focused on the automatic recognition of surgical phases by microscope images analysis. We used a hybrid method that combined Support Vector Machine and discrete Hidden Markov Model. We first performed features extraction and selection on surgical microscope frames to create an image database. SVMs were trained to extract surgical scene information, and then outputs were used as observations for training a discrete HMM. Our framework was tested on pituitary surgery, where six phases were identified by neurosurgeons. Cross-validation studies permitted to find a percentage of detected phases of 93% that will allow the use of the system in clinical applications such as post-operative videos indexation.

Keywords: Surgical phase, digital microscope, neurosurgery, SVM, HMM.

1 Introduction

In recent years, due to the progress of medicine and computers, there has been an increased use of technologies and in the Operating Room (OR). To develop computer assisted systems that better handle and integrate this new OR [1], a more detailed comprehension of the surgical workflow is needed. From this area, the challenge of surgical workflow recovery has emerged. Clinical applications are the evaluation of surgeons, OR management optimization or the creation of context-sensitive user interfaces. As mentioned in [2], the modelling must address behavioural, anatomical, pathological aspects and instruments. The concepts of Surgical Process (SP) and SP models (SPM) have been introduced for such purposes [2,3].

Related data extraction techniques can be classified according to the addressed level of granularity where the surgery is studied. These approaches yield to various methods used for data acquisition: patient specific procedures description [2,3], interview of surgeons [4], sensor-based methods [5-14], using fixed protocols created by expert surgeons [15], or combination between them [16]. Within sensor-based methods, an approach for finer classification is to differentiate materials used: such as a robot-simulator in virtual environments [5], using existing or additional materials.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 54–62, 2011.

Most of studies used sensors additionally installed. Padoy et al. [6] positioned sensors on instruments, and the workflow was segmented using Dynamic Time Warping (DTW) and Hidden Markov Model (HMM). Similarly, mechanisms for dataset preprocessing using Bayesian network before HMM training were presented in [7]. In both works, data acquisition was performed manually. Accelerometers placed on the operator were used in [8] to identify alphabets of activity. James et al. [9] installed an eye-gaze tracking system on surgeons combined with visual features to detect one important phase. Nara et al. [10] introduced an ultrasonic location aware system that tracks 3-D positions of the staff for the identification of surgical events.

From existing sensors within the OR, videos are a rich source of information, as demonstrated on laparoscopy [11]. A situation recognition process was created based on augmented reality and computer vision techniques. Helpful information such as 3D map were also extracted from laparoscopic videos in [12]. Bhatia et al. [13] analyzed global view videos for better OR management. Finally, Xiao et al. [14] implemented a system that record patient vital signs to situate the intervention process.

Our project focused on the extraction of information from microscope videos for high-level tasks recognition. Compare to other techniques, it permits not only to avoid the installation of materials, but also to have a source of information that has not to be controlled by human. Even if the approach and the application differ, we followed a methodology similar to the one described in [13] along with results of our previous work [17], where we presented in detail the image feature extraction process and performed studies on machine learning algorithms. Here the goal was to add a temporal reasoning for better detection. That's why we first took advantage of the ability of SVMs as binary classifiers to extract scene information from frames. Then outputs of the classification were treated as observations to train a HMM. This combination permitted to take into account the sequential nature of high-level tasks for accurate recognition. We focused in this paper on the detection of surgical phases and validated our methodology with a specific type of neurosurgical interventions: the pituitary surgeries.

2 Materials and Methods

The process for automatic recognition is introduced here: frames were first extracted from microscope videos, and reduced with spatio-temporal downsampling to perform feature extraction. Image signatures were composed of 185 features, in which discriminant ones were chosen with a specific feature selection. SVMs were then used to classify relevant surgical scene information. These results alone were not enough informative to correctly classify surgical phases, so a HMM has then been trained, taking as observations the outputs of the SVMs and as hidden states the surgical phases. The Viterbi decoder finally permitted to find the optimal path for a given observation sequence. We assessed this process with cross-validation studies.

2.1 Data-Set

We evaluated our algorithm on pituitary surgeries [18], presented in [17]. Pituitary adenomas are tumors that occur in the pituitary gland, where neurosurgeons use a

trans-nasal approach with an incision in the back wall of the nose. The anonymous data set was composed of 16 pituitary surgeries (mean time: 40min), all performed in the neurosurgical department of Rennes University Hospital by three expert surgeons. Videos focused on the operative field of view and were recorded using the surgical microscope OPMI Pentero (Carl Zeiss) (Videos: 768 x 576 pixels at 33 frames per second). The labeling of surgical phases was manually performed by surgeons (Fig. 1.).

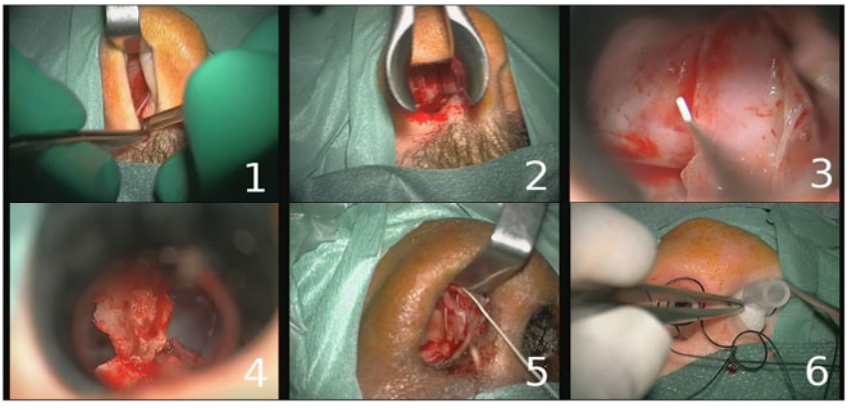

Fig. 1. Example of typical digital microscope images for the six phases: 1) nasal incision, 2) nose retractors installation, 3) access to the tumor along with tumor removal, 4) column of nose replacement, 5) suturing, 6) nose compress installation

Original frames were first spatially downsampled by a factor of 8 with a 5-by-5 Gaussian kernel (internal studies have shown that it had no impact on accuracy) and then downsampled to 1 frame every 2s (0.5Hz). We performed a statistical intensity normalization of images, where data closely followed a normal distribution.

2.2 Feature Extraction and Selection

From these videos, we randomly extracted 500 frames which were supposed to correctly represent the six phases of a common pituitary surgery. We defined for each frame a feature vector, representing a signature. Signatures are composed of three main information that usually describe an image: the color, the texture and the form.

The color has been extracted with two complementary spaces [19]: RGB space (3 x 16 bins) along with Hue (32 bins) and Saturation (32 bins) from HSV space. The texture has been extracted with the co-occurrence matrix along with Haralick descriptors [20]. The form was represented with spatial moments [21], and we also computed the Discrete Cosine Transform (DCT) [22] coefficients. Each signature was finally composed of 185 complementary features.

The main goal of feature selection is to remove redundancy information and to keep essential ones. Feature selection methods can be divided into two groups [23]: the filter and the wrapper approach. We fused them using the method described by Mak and Kung [24], where they argued that both methods are complementary to each other. Algorithms were first independently applied to find two feature subsets. They

were then merged by selecting one feature at a time from the two subsets, starting from the highest rank. The Recursive Feature Elimination (RFE) SVM [25] was chosen for wrapper method. The principle is to generate the ranking using backward feature elimination. The mutual information (MI) [26] was chosen for the filter method, where a feature is more important if the MI between the target and the feature distributions is larger. In order to a have a good compromise between computation time and accuracy, we kept the 40 first features.

2.3 Supervised Classification

Based on our last work [17], multiclass SVMs [27] have been found to be effective for microscope images classification. We decided to use binary SVMs for scene information extraction. SVMs are supervised learning algorithms used for classification and regression. Mathematically, given training data $\{x_1...(x_n)\}$where $x \in \Re^d$ and their labels $\{y_1...(y_n)\}$where $y \in \{-1,(1)\}$. The goal is to find the optimal hyperplane $w.x+b=0$ that separates the data into two categories. The idea is to maximize the margin between the positive and negative examples. The parameter pair $(w;b)$ is finally the solution to the optimization problem:

$$\varphi(w)=\frac{1}{2}|w|^2=\frac{1}{2}(w.w) \tag{1}$$

following constraints:

$$y_i(x_i.w+b)-1\geq 0, \forall i \tag{2}$$

Four discriminant scene information were defined: global-zoom views, presence-absence of nose retractors, of the column of nose and of compress. Combinations of these 4 binary outputs resulted in 16 possible observations for the HMM.

2.4 HMM

HMMs [28] are statistical models used for modeling of non-stationary vector times-series. An HMM is formally defined by a five-tuple (S,O,Π,A,B), where $S=\{s_1...(s_N)\}$ is a finite set of N states, $O=\{o_1...(o_M)\}$ is a set of M symbols in a vocabulary, $\Pi=\{\pi(i)\}$ are the initial state probabilities, $A=\{a(ij)\}$ the state transition probabilities and $B=\{b_i(o(k))\}$the output probabilities. Here, outputs of SVMs were treated as observations for the HMM. States were represented by the surgical phases that generated a left-right HMM (Fig. 2.). The transition probabilities were low because of the sampling rate of frames (0.5Hz). We set two probabilities for the transition from one state to its consecutive state: $\alpha=0.005$ for state n°3, $\beta=0.01$for the others. The probability of remaining in the same state is then: $1-\alpha$ or $1-\beta$. The outputs probabilities were obtained from SVMs results. They were computed as the probability of having an observation in a specific state. Videos were applied to SVMs and observation probabilities were manually computed. All these probability

computations were part of the training process of our framework, and were therefore obtained only from the training sample of the cross-validation. Furthermore, the Baum-Welch algorithm has not been used for the training because of the limited size of the training sample.

Lastly, given observations and the HMM structure, the Viterbi algorithm [29] find the most likely sequence of states.

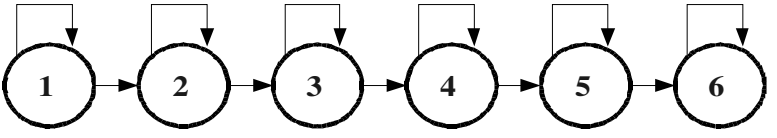

Fig. 2. Left-right HMM, where each state corresponds to one surgical phase

2.5 Cross-Validation

SVM classifiers and HMM were both evaluated with a random 10-fold cross-validation study. The image database and videos were divided into 10 random subsets. Nine were used for training while the prediction was made on the 10^{th} subset. This procedure was repeated 10 times and results were averaged. We computed the correct classification rate for SVMs evaluation and the percentage of phases misclassified, namely the Frequency Error Rate (FER), for HMM assessment. In addition, the confusion matrix was extracted, showing exactly where states were misclassified.

3 Results

Statistical results of the cross-validation study for the SVMs (Tab. 1.) showed that very good detections (~88%) along with low standard deviations (max=2.5%) were obtained, getting a maximum accuracy of 94.6% for the column of nose detection.

Table 1. Mean accuracy and standard deviation (Std) of the 4 scene information recognition

	Global-zoom view	Presence-absence of nose retractors	Presence-absence of column of nose	Presence-absence of compress
Accuracy (%)	87.6	88.0	94.6	87.4
Std (%)	2.4	2.2	1.6	2.5

The HMM study showed a **mean FER of 7.1 +/- 5.3%**, with a **min of 2.5%** and a **max of 15.2%.** This error rate is low, but values are very scattered (resulting in a high standard deviation). A recognized sequence compare to the true one is shown on Fig. 3. On this particular example, each state is correctly classified with a maximum delay of 40s.

From Tab. 2., we see that state n°3 contains the bigger number of frames, and all confusions are always between neighbouring states. The most significant error is for state n°5, where the detection is around 75%. The highest accuracy (excluding the first and the last state) is for state n°4, where the detection reaches 95%.

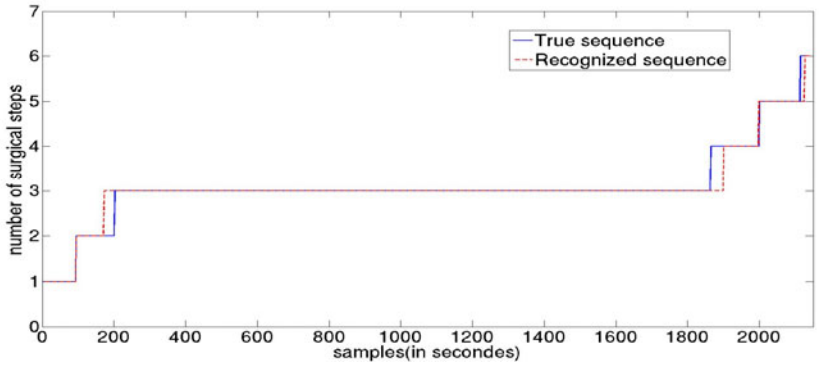

Fig. 3. Phases recognition of one video made by the HMM compared to the ground truth

Table 2. Confusion matrix for the surgical phases detection with the Viterbi decoder. Rows indicate the recognized surgical steps and columns the ground truth.

	1	2	3	4	5	6
1	5.68	0.97	0	0	0	0
2	0	4.68	4.09	0	0	0
3	0	0	72.99	0.12	0	0
4	0	0	0.45	3.12	0.07	0
5	0	0	0	0.04	3.31	0
6	0	0	0	0	0.99	3.49

4 Discussion

In this paper, we proposed a framework that automatically recognizes surgical phases from microscope videos. The combination of SVMs and HMM showed a total accuracy of 93% of detected phases.

4.1 Microscope Video Data

As mentioned in [13], the information extracted from the OR must be discriminant, invariant to task distortion, compact in size and easy to monitor. Microscope video data meet all of these constraints. Image features are first very discriminant for scene information extraction, as the SVMs validation indicated. Secondly, within a same surgical environment, procedures are reproducible and image features are thus invariant to task distortion. This constraint addresses the issue of the adaptability of the system. Due to the different equipments in each department, the system could be not flexible. The solution would be to train dedicated image databases for each department which would be adapted to the corresponding surgical environment and microscope scene layout. The idea would also be to have several models for each type of procedure, adapting the scene information to be extracted and optimising the HMM. The third crucial parameter is the sample size which must be compact. Image signatures are composed of 40 features and are thus reduced. Finally, the real added value is the use of microscope videos. This device is not only already installed in the OR, but it has also not to be monitored by the staff.

4.2 Accuracy of the Detection

We decided to use frames in a static way (without motion), and deliberately remained at a high level of granularity with the detection of phases. The recognition of lower level information, such as gestures, is difficult with microscope videos only. Spatio-temporal features will have to be inserted for the segmentation of such information.

The possible phases “access to the tumor” and “tumor removal” were fused because the transition between both was not clear due to similar tools and zooms. The confusion matrix showed that there was no main confusion and that the HMM was helpful for separating phases with same image features (like phase n°1 and n°5).

The high recognition rates of binary SVMs, associated with their small standard deviations, indicates that they are very robust for images classification. Then graphical probabilistic models allow an efficient representation of the problem by modelling time varying data. This association permitted to obtain good detection accuracy.

4.3 Clinical Applications

Workflow recovery might be helpful for various applications. Purposes are generally to bring a plus-value to the surgery or to the OR management. This work could be integrated in an architecture that would extract microscope frames and transform it into information helping the decision making process. For now, even with a low computation time (feature extraction/selection + classification < 0.5s), accuracy must definitively be higher than our results before establishing on-line surgical phase detection.

However, the system could be introduced for video indexation. Surgical videos are very useful for learning and teaching purposes, but surgeons often don't use them because of the huge amount of surgical videos. The data-base would contain relevant surgical phases of each procedure for easy browsing. Moreover, we could imagine the creation of pre-filled reports that will have to be completed by surgeons. For such applications, even with few errors, the automatic indexation would be helpful, as there is no need of perfect detection and it has no impact on the surgery.

5 Conclusion

Using the proposed framework, we are now able to recognize the major surgical phases of every new procedure, by computing frames signatures, classifying scene information, and decoding SVMs outputs with the Viterbi algorithm. Thanks to this combination, we obtained high detection accuracy. We have validated this framework with pituitary surgeries. Six phases were defined by an expert, and we found a global accuracy of 93% of detected phases. This recognition process is a first step toward the construction of context-aware surgical systems. Currently, it could be used for post-operative video indexation or reports generation. In future works, spatial image features will have to be mixed with other information (such as spatio-temporal features) to generate a more robust system.

Acknowledgments. The authors would like to acknowledge the financial support of Carl Zeiss Surgical GmbH.

References

1. Cleary, K., Chung, H.Y., Mun, S.K.: OR 2020: The operating room of the future. Laparoendoscopic and Advanced Surgical Techniques 15(5), 495–500 (2005)
2. Jannin, P., Morandi, X.: Surgical models for computer-assisted neurosurgery. Neuroimage 37(3), 783–791 (2007)
3. Neumuth, T., Jannin, P., Strauss, G., Meixensberger, J., Burgert, O.: Validation of Knowledge Acquisition for Surgical Process Models. J. Am. Med. Inform. Assoc. 16(1), 72–82 (2008)
4. Morineau, T., Morandi, X., Le Moëllic, N., Diabira, S., Haegelen, C., Hénaux, P.L., Jannin, P.: Decision making during preoperative surgical planning. Human Factors 51(1), 66–77 (2009)
5. Darzi, A., Mackay, S.: Skills assessment of surgeons. Surg. 131(2), 121–124 (2002)
6. Padoy, N., Blum, T., Feuner, H., Berger, M.O., Navab, N.: On-line recognition of surgical activity for monitoring in the operating room. In: Proc. of IAAI (2008)
7. Bouarfa, L., Jonker, P.P., Dankelman, J.: Discovery of high-level tasks in the operating room. Journal of Biomedical Informatics (in Press, 2010)
8. Ahmadi, S.A., Padoy, N., Rybachuk, K., Feussner, H., Heinin, S.M., Navab, N.: Motif discovery in OR sensor data with application to surgical workflow analysis and activity detection. In: M2CAI Workshop, MICCAI, London (2009)
9. James, A., Vieira, D., Lo, B.P.L., Darzi, A., Yang, G.-Z.: Eye-gaze driven surgical workflow segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 110–117. Springer, Heidelberg (2007)
10. Nara, A., Izumi, K., Iseki, H., Suzuki, T., Nambu, K., Sakurai, Y.: Surgical workflow analysis based on staff's trajectory patterns. In: M2CAI Workshop, MICCAI, London (2009)
11. Speidel, S., Sudra, G., Senemaud, J., Drentschew, M., Müller-stich, B.P., Gun, C., Dillmann, R.: Situation modeling and situation recognition for a context-aware augmented reality system. Progression in Biomedical Optics and Imaging 9(1), 35 (2008)
12. Sánchez-González, P., Gayá, F., Cano, A.M., Gómez, E.J.: Segmentation and 3D reconstruction approaches for the design of laparoscopic augmented reality environments. In: Bello, F., Edwards, E. (eds.) ISBMS 2008. LNCS, vol. 5104, pp. 127–134. Springer, Heidelberg (2008)
13. Bhatia, B., Oates, T., Xiao, Y., Hu, P.: Real-time identification of operating room state from video. In: AAAI, pp. 1761–1766 (2007)
14. Xiao, Y., Hu, P., Hu, H., Ho, D., Dexter, F., Mackenzie, C.F., Seagull, F.J.: An algorithm for processing vital sign monitoring data to remotely identify operating room occupancy in real-time. Anesth Analg. 101(3), 823–832 (2005)
15. MacKenzie, C.L., Ibbotson, A.J., Cao, C.G.L., Lomax, A.: Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. Min. Invas. Ther. All Technol. 10(3), 121–128 (2001)
16. Neumuth, T., Czygan, M., Goldstein, D., Strauss, G., Meixensberger, J., Burgert, O. Computer assisted acquisition of surgical process models with a sensors-driven ontology. In: M2CAI Workshop, MICCAI, London (2009)
17. Lalys, F., Riffaud, L., Morandi, X., Jannin, P.: Automatic phases recognition in pituitary surgeries by microscope images classification. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 34–44. Springer, Heidelberg (2010)

18. Ezzat, S., Asa, S.L., Couldwell, W.T., Barr, C.E., Dodge, W.E., Vance, M.L., McCutcheon, I.E.: The prevalence of pituitary adenomas: a systematic review. Cancer 101(3), 613–622 (2004)
19. Smeulders, A., Worrin, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. 22(12), 1349–1380 (2000)
20. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. on Systems, Man, and Cybernetics 3(6), 610–621 (1973)
21. Hu, M.: Visual pattern recognition by moment invariants. Trans. Inf. Theory 8(2), 79–87 (1962)
22. Ahmed, N., Natarajan, T., Rao, R.: Discrete Cosine Transform. IEEE Trans. Comp., 90–93 (1974)
23. Duda, R.O., Hart, P.E.: Pattern classification and scene analysis. Wiley, New York (1973)
24. Mak, M.W., Kung, S.Y.: Fusion of feature selection methods for pairwise scoring SVM. Neurocomputing 71, 3104–3113 (2008)
25. Guyon, I., Weston, J., Barhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machine. Machine Learning 46, 389–422 (2002)
26. Hamming, R.W.: Coding and Information Theory. Prentice-Hall Inc., Englewood Cliffs (1980)
27. Crammer, K., Singer, Y.: On the Algorithm implementation of multiclass SVMs. JMLR (2001)
28. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proc of IEEE 77(2) (1989)
29. Viterbi, A.: Errors bounds for convolutional codes. IEEE TIT 13(2), 260–269 (1967)