

Understanding expertise in surgical gesture by means of Hidden Markov Models *

G. Megali, S. Sinigaglia, O. Tonet, F. Cavallo, P. Dario
CRIM Lab, Scuola Superiore Sant'Anna
viale R. Piaggio 34, 56025 Pontedera, Italy
EndoCAS, University of Pisa,
Ospedale di Cisanello, via Paradisa 2, 56124 Pisa, Italy
{g.megali, s.sinigaglia, o.tonet, p.dario}@crim.sssup.it

Abstract—Minimally invasive surgery (MIS) has become very widespread in the last ten years. Due to the difficulties encountered by the surgeons to learn and manage this technique, a huge importance has the improvement of training procedures, the improvement of surgical instrumentation and the robotic automation of surgical gesture. All these purposes require the analysis of surgical performance with the aim to understand it and to define what is *expertise* in surgical gesture. In this paper for the first time the Hidden Markov Models (HMMs) are used as a tool for the understanding of surgical performance and of the human factors that characterize it. In our experiments we used position data concerning the tools movements during exercises performed on a surgical simulator. Using Hidden Markov theory, we create a model of the *expert surgeon performance* able to evaluate surgical capability and to distinguish between expert and non-expert surgeons. By analyzing the trained model of the *expert surgeon performance* we show that it is possible to deduce information about features characterizing the *surgical expertise*.

Index Terms—Endoscopic surgery, Surgical Training, Hidden Markov Model, Evaluation of surgical performance.

I. INTRODUCTION

Minimally invasive surgery (MIS) has assumed, in the medical scenario, a dominant role as a consequence to the remarkable social (and economic) improvement that it involves. Laparoscopic surgery is a procedure that uses ports varying in size from 5 mm to 12 mm, throughout the abdomen for accessing the internal organs. The abdomen is then distended with gas (CO₂). A small telescope is introduced into the abdominal area for visual examination. Additional ports are required to place special instruments to operate. While on one hand MIS procedures ensure many advantages to patients, on the other hand they require surgeons to undergo a long and difficult training in order to manage and master these techniques. Mainly, surgeons encounter perceptual limitations (such as lack of stereoscopic view, limited field of view, and reduced force and tactile sensing), and motor limitations (such as reversed motion, movement scaling and limited degrees-of-freedom). The skills required for MIS are based on unique perceptual-motor relationships that make them very difficult to master.

*This work has been supported in part by the FIRB-2001 Project *ApprEndo* (no. RBNE013TYM) and by *EndoCAS*, the Centre of Excellence for Computer-Assisted Surgery (COFINLAB-2001), both funded by MIUR, the Italian Ministry of Education, University and Research.

In this context, designing more efficient training exercises and defining objective methods to evaluate surgical performance, so as the ergonomic and functional design of new surgical instruments and the robotic automation of surgical procedures [1] are the major challenges for the surgery of the future. To these purposes the analysis of surgical performance and the understanding of human factors that characterize it, is a crucial step along this path. Analyzing and understanding the surgical gesture, consists in quantitatively defining what optimal surgical performance means, and what characterizes the expert gesture.

Concerning this problem a lot of works, such as [2], [3], [4], focused on the definition of metrics for an objective evaluation of surgical performance, are principally based on uncorrelated measurements of simple data (such as time, number of targets missed, trajectories, velocity, acceleration, smoothness, etc.).

In this paper we create a model of the *expert surgical performance* and show that, by analyzing this model, it is possible to deduce information that can help to identify features of the *surgical expertise*. The work is based on the processing of data concerning the timed trajectories of the movements performed by experienced surgeons during exercises on a surgical simulator. The data could be collected also with other different training methods or directly during surgical intervention, the only requirements is the tracking of the position of surgical tools.

With respect to the above cited works, we follow a different approach based on machine learning methods. The machine learning method we used to model the performances of expert surgeon is hidden Markov theory.

The strength of Hidden Markov Models (HMMs) is that they do not require an a priori definition of what surgical expertise is. Similar studies based on the same approach [5], [6] use Markov model theory to evaluate surgical performance during training exercises, choosing manually the states of the model (Grasping state, Idle state ecc.). Other studies [7], always with the objective to develop an evaluation metric for surgical performance, use HMMs to automatically recognize 10 simple user defined tasks. All these works implicitly suppose to exactly know a priori what is the surgical expertise, in fact they define manually a set of simple physical states, and then

they describe performances by means of a sequence of transitions between these states.

Our study differs from the above cited works because we use the full power of HMMs. In fact we only suppose the existence of hidden physical/mental states, and leave the system to train this hidden layer on the basis of observable physical measurements. The training phase makes the model converge to the definition of states and to the definition of the state transition rule that better describes the surgical expertise. Once the model is trained, paraphrasing the study made by Cave and Neuwirth [8], we analyze the structures that define it, i.e. the probability transition matrix and the observation matrix. This analysis can help us to determine the relationship between the hidden states and the observations (i.e. the physical measurements).

In the paper first we shortly introduce the mathematics of HMMs. After we provide a description of the data acquisition procedure and of the experimental protocol. We acquired data in a simple exercise that allows subjects to train moving and positioning of laparoscopic instruments in the surgical scenario. In section IV we describe the approach followed for the modelling of a generic *expert* performance, and we provide the results of the validation of the model. These results show that the model recognizes expert performances. In section V we provide the analysis of the trained model, and we conclude the paper with a discussion of the results. Analyzing our expert model we noticed that the hidden states converge to real physical states, easy to recognize and describe. In fact, the results of our experiments reveal that in simple exercises the expertise of a surgeon is coded in his bi-manual skill.

II. HMM, SHORT DESCRIPTION

Consider a process which can be described at any time, as being in one of the N distinct states $S_1, S_2, S_3, \dots, S_N$ (see Fig. 1).

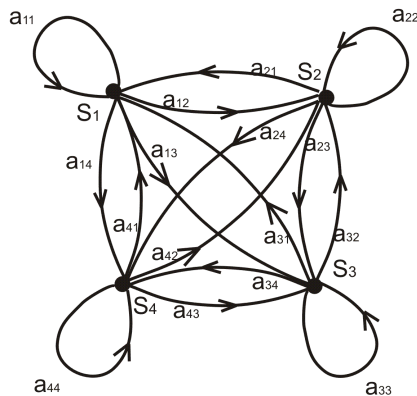


Fig. 1. A Markov chain with 4 states fully connected.

Denote the time instant with $t = 1, 2, \dots$, and the state at time t with q_t . At each instant t the transition probability between two states i and k , with $i, k \in \{1, \dots, N\}$, follows a statistical distribution.

In a Markov process the probability of q_t depends uniquely from the state assumed at the previous instant $t - 1$:

$$\begin{aligned} P[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots] \\ = P[q_t = S_j | q_{t-1} = S_i] \end{aligned}$$

With this assumption the process can be statistically described by a matrix A whose elements are the state transition probabilities (Fig. 1):

$$\begin{aligned} a_{ij} &= P[q_t = S_j | q_{t-1} = S_i] \quad 1 \leq i, j \leq N \\ a_{ij} &\geq 0 \quad \sum_{j=1}^N a_{ij} = 1 \end{aligned}$$

The HMM theory is an extension of the Markov model concept used to describe physical phenomena where states are not directly observable but observations are statistically dependent by them. Suppose to have N discrete states and M possible observables ν_k , $1 \leq k \leq M$. Like for the definition of the state transition probabilities matrix A , it is possible to define the $N \times M$ observation probabilities distribution matrix B , having elements $b_j(k)$ defined as follows:

$$b_j(k) = P[\nu_k \text{ at time } t | q_t = S_j]$$

with $1 \leq j \leq N$ and $1 \leq k \leq M$.

The process is modelled if matrices A and B are known, and if the probability distribution of the initial state

$$\pi_i = P[q_1 = S_i] \quad 1 \leq i \leq N$$

is known. In compact notation a HMM is indicated as $\lambda = (A, B, \pi)$.

HMMs have three basic problems that must be solved to completely model physical phenomena:

- Pr. 1 Given a phenomenon described by the model $\lambda = (A, B, \pi)$, and the observation sequence $O = o_1 o_2 \dots o_T$, compute $P(O|\lambda)$, consisting in the probability to obtain the sequence O from the phenomenon modelled with λ .
- Pr. 2 Given the model $\lambda = (A, B, \pi)$, find the most probable state sequence $Q = q_1 q_2 \dots q_T$ related to the observation sequence $O = o_1 o_2 \dots o_T$.
- Pr. 3 Given the observation sequence $O = o_1 o_2 \dots o_T$ determine the parameters A, B, π of the model $\lambda = (A, B, \pi)$ so that the observation probability $P(O|\lambda)$ is maximal.

III. DATA ACQUISITION

We used the commercial laparoscopic simulator LapSim Basic Skills 2.2 (Surgical Science AB, Göteborg, Sweden) which allowed to perform exercises in a virtual environment and to easily acquire data concerning instrument positions (Fig. 2).

The use of a surgical simulator, avoiding the needs of an operating room and of human cadavers or animals for testing, will allow us to consider a high number of exercises.



Fig. 2. The LapSim system.

The LapSim simulator provides the endoscopic view of virtual environments, with which the surgeon can interact using the Virtual Laparoscopic Interface (VLI, produced by Immersion Corporation, San Jose, USA). The VLI interface consists in two mock-ups of laparoscopic instrument handles fixed to a sensorized mechanical support that gives a 3D position value of each tool tip (Fig. 3) and two additional values about its orientation and grasper angle.

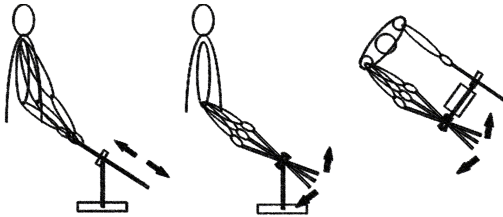


Fig. 3. Graphic representation of the 3 degree of freedom related to tip position.

Moreover the LapSim software allows the user to edit and customize exercises and to reproduce them identically multiple times. This feature gives different subjects the possibility to execute identical exercises, and gives the same subject the possibility to perform the same exercise at different times, allowing the monitoring of his/her learning curve. So we created several exercises to study the various characteristics of the gesture and the cognitive ability of the surgeon during specific surgical actions and collected data concerning the performances of different surgeons at different levels of expertise. In particular we created some basic exercises to evaluate the surgeon's ability in the management of surgical instruments and in depth perception using laparoscopic view. In these exercises the surgeon has to reach, alternatively with the left and right instrument, spheres that appear in the virtual scenario.

We choose to study tip positions measured by the hardware interface, ignoring grasper and tool orientation values that are not influent in this exercise, acquiring data at 60 Hz frequency, high enough to describe human gesture [9].

Components of both tips, right and left, was stored in PC memory and off-line processed, using MATLAB.

IV. EXPERIMENTAL PROTOCOL

A group of 4 novices and a group of 2 expert surgeons were asked to complete an experimental session of 4 consecutive trials. The exercises consisted in reaching, alternatively with the right and left tool, 10 spheres which appeared one at a time in the virtual scenario. Once a sphere was touched with the tip of the instrument, it disappeared.

V. MODELING EXPERT SURGICAL GESTURE

The first part of the work consists in creating a HMM that describes the surgical performance made by expert surgeons. To this purpose we trained a $\lambda = (A, B, \pi)$ to give high value of $P(O|\lambda)$ for expert observation series and low value of $P(O|\lambda)$ for non-expert observation series. We used the Baum-Welch algorithm [10] to solve the training problem (Pr. 3). The algorithm, given a finite number of observation sequences O_i as training data, iteratively adjusts the model parameters (A, B, π) maximizing $P(O_i|\lambda)$, the probability of observation sequences O_i given the model λ .

A. Data Processing

For creating a HMM of the process we need to represent the input data in a form that maintains the *interesting characteristics*. In a previous work [11] we noticed that the spectrum analysis of accelerations highlights the differences in the gesture of expert and inexperienced surgeons. Applying the Short Time Fourier Transform (STFT) we obtain a frequencies spectrum function of the gesture in the time domain. The STFT has been computed using the Fourier transform of the signal $x(t)$ in a rectangular time window $\gamma^*(t-t')$ centered around t and of width W [12]. For each time window we defined the gesture spectrum as:

$$STFT_x^\gamma(t, f) = \int_{t'} [x(t')\gamma^*(t-t')]e^{-j2\pi ft'} dt'$$

and we used it to characterize the process at time t .

So, at instant t (i.e. for each time window) we have three power spectra for each instrument (corresponding to the tool tip acceleration given in spherical polar coordinates). To apply the HMM (discrete and one-dimensional) theory we need to map them in a discrete space using a clusterization algorithm. The clusters are defined iteratively optimizing a function of the distances between the points of the cluster and its centroid [13]. Once a clusterization map has been defined, for each time window, we define as *observation* the number of the cluster in which the corresponding set of power spectrum falls. In this way the sequence of the cluster numbers represents a one-dimensional observation sequence that we can use for the HMM training. We performed the clusterization process using the *K-means* algorithm [14].

B. Topology of the expert model

To model expert gesture, we acquired a set of data concerning repetitions of the same exercise performed by two experienced surgeons (*ES1* and *ES2*). Our objective,

at this stage, is to define a unique HMM able to recognize as *expert* the gesture executed by both expert surgeons. We used ergodic models (where all state transitions are allowed) since we have no reasons to suppose limited state transitions. The topology of the HMM is determined by the following parameters: the number N of hidden states, the number obs of possible observations (number of clusters of the vector quantization), and the sampling time interval Wl (the duration time of the STFT windows). We choose the parameters N , obs and Wl such that the two models, λ_{ES1} trained on data collected from the first expert surgeon, and λ_{ES2} trained on data collected from the second, recognize as their own the performances carried out by both expert surgeons [15].

C. Expert model validation

Once we fixed the topology parameters $N = 3$, $obs = 20$ and $Wl = 0.5$ s and trained the expert model λ_{ES} using data relative to expert performances of both expert surgeons, we defined an evaluation metric S_m based on the value of the log-likelihood of a generic performance compared to the expert model [15]. The log-likelihood value, defined as $\log P(O|\lambda_{ES})$, consists in a measure of the probability that a generic performance (described by means of the observation sequence O) is generated by the expert model λ_{ES} . The expert model created recognizes expert surgeon performances, in fact observation sequences relative to performances executed by expert surgeons will produce values of S_m smaller than the values produced using the data relative to performances of unexperienced surgeons as shown in the Fig. 4.

VI. ANALYZING THE TRAINED EXPERT MODEL

As described before blindness is the real power of the training algorithm, which only requires kinematic data of the tool tips without a priori assumptions on what expert gesture is characterized by. At this stage the model appears like a black box able to distinguish between expert and non-expert performances, but if we analyze the matrices A and B of the model we can obtain information concerning the physical phenomenon modeled, and hence infer knowledge about it and its characteristics.

Once the model is trained by analyzing the structures that define it, i.e. the probability state transition matrix and the observation matrix, we can obtain informations that help us to determine the relationship between the hidden states of the trained model and the physical observations of the phenomenon [8].

The analysis of matrix B of the trained model let us see if the hidden states of the trained model have converged to real physical states, easy to recognize and describe. In other words we investigate if a state describes a particular phase of the performance, and hence attracts observables whose *physical similarity* is easy to identify. The analysis of matrix A reveals the nature and the state sequence of the physical phenomenon described by the trained model, i.e. it can help us to provide a graphical representation of the most probable state sequences describing phenomenon.

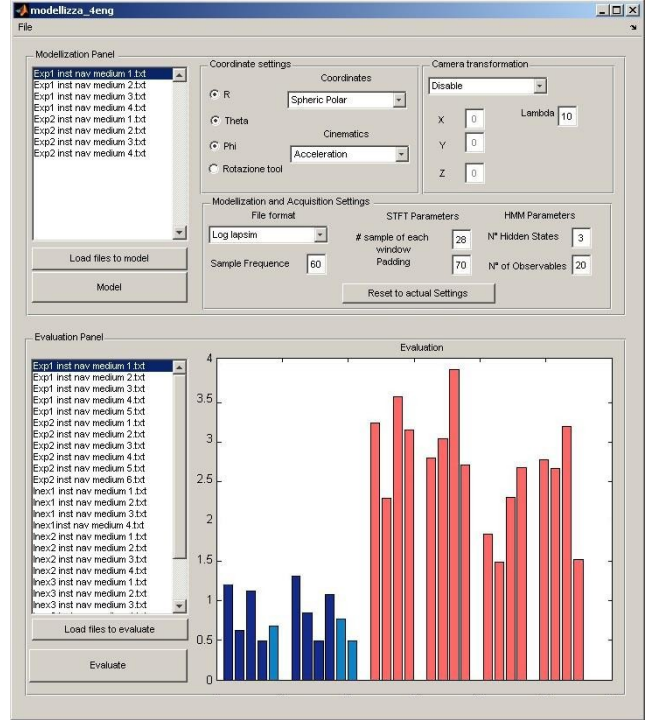


Fig. 4. Evaluation of surgical performance: the first two groups of bars are the evaluation of the expert surgeons performances (darker bar indicate exercises used to train the model). The remaining 4 groups are the evaluation of the 4 inexpert surgeons performances.

A. Analysis of the relationship between states and observables

By studying the observation matrix B , for each state, we selected the three observables that have the maximum probability to appear (as reported in Table I).

State number	Most probable observable		
1	18	12	7
2	13	4	17
3	3	16	9

TABLE I
STATE LIST AND MOST PROBABLE OBSERVABLES

Given the trained model $\lambda = (A, B, \pi)$, for each exercise (described by means of an observation sequence $O = o_1 o_2 \dots o_T$) we found the most probable states sequence $Q = q_1 q_2 \dots q_T$ relative to O (i.e. we solved the Pr. 2 of the HMM [10]). Then we compared the states q_i with the corresponding observations o_i , $0 \leq i \leq T$ and we selected those pairs (o_i, q_i) that match the state-observable combinations reported in Table I (i.e. those pairs for which o_i and q_i stay on the same row in Table I). Analyzing accelerations related to the selected pairs (o_i, q_i) , results that the power spectrum of gesture has very different characteristics for observation recognized being in different states.

State 1 and state 2 (respectively Fig. 6 and Fig. 5) are characterized by a dominance of right and left instrument respectively, indeed they clearly has very different power

(power spectrum integral) of the two tools, specially in the radial coordinate. State 3 (Fig. 7) shows a similar power of the two instruments, highlighting a bi-manual movement.

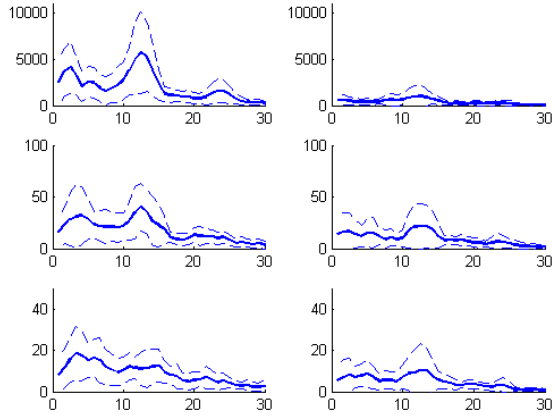


Fig. 5. Power spectrum of polar acceleration for the observations recognized as State 1. Continuous bold line represents the mean μ , while the dashed lines represent $\mu \pm \sigma$ (σ = standard deviation). In the left and right column are depicted radial, azimuthal and transversal accelerations spectra respectively of left and right instruments.

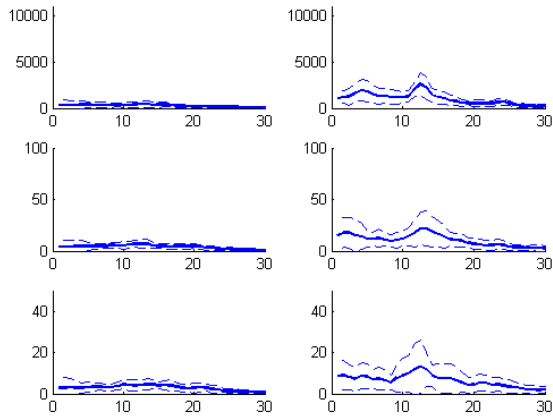


Fig. 6. Power spectrum of polar acceleration for the observations recognized as State 2. Continuous bold line represents the mean μ , while the dashed lines represent $\mu \pm \sigma$ (σ = standard deviation). In the left and right column are depicted radial, azimuthal and transversal accelerations spectra respectively of left and right instruments.

By analyzing these graphs it is possible to assert that the 3-states model tends to associate at each state a real and easy to recognize physical phase: indeed state 1 represents the phase *Moving Right Hand*, state 2 represents the phase *Moving Left Hand* and state 3 represents the phase *Moving Both Hands*. This is a very interesting result that focuses in the instrument dominance and bi-manual coordination an important characteristic for the study of the expertise.

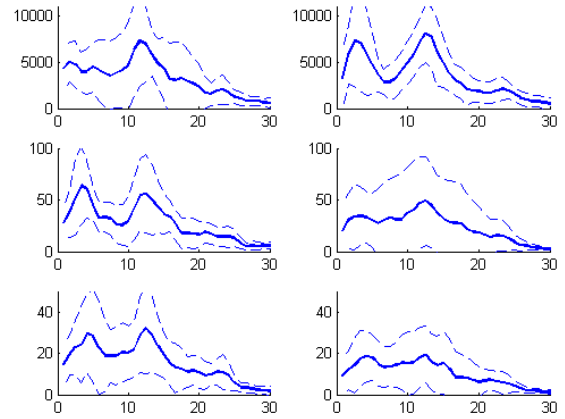


Fig. 7. Power spectrum of polar acceleration for the observations recognized as State 3. Continuous bold line represents the mean μ , while the dashed lines represent $\mu \pm \sigma$ (σ = standard deviation). In the left and right column are depicted radial, azimuthal and transversal accelerations spectra respectively of left and right instruments.

B. Analysis of Model Topology

Just by studying the state transition matrix A :

$$\mathbf{A} = \begin{pmatrix} 0.84 & 0.12 & 0.04 \\ 0.04 & 0.81 & 0.15 \\ 0.24 & 0.13 & 0.63 \end{pmatrix}$$

is it possible to note that the model tends to converge to a *cyclic left-right* topology where transitions are allowed only to the same or next state (Fig. 8).

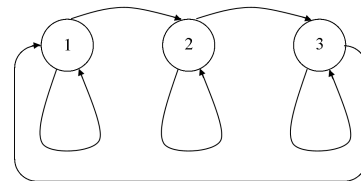


Fig. 8. Representation of the ideal topology of the 3-states model: state 1 corresponds to *Moving Right Hand*, state 2 corresponds to *Moving Left Hand* and state 3 corresponds to *Moving Both Hand*.

According to the exercise performed, that consists in an alternate reach of balls with left and right instrument, it suggests that after a left reach the two surgeons put the instrument in the idle position and then they start to use the right instrument. On the contrary, after a right reaching, the left instrument starts to work while right instrument is still reaching the idle position. A possible interpretation of this phenomena, according to the fact that both surgeon are right handed, is that they manage to handle the right instrument for simple tasks also when their attention is not focused on it, whereas the use of the left instrument requires the attention focused on it.

C. Consideration about a 4-states model

Using the same procedure we trained and validated a 4-states *expert model*. This model, as the 3-state model, has

3 states related to left, right or bi-manual behavior and has a new *Moving Fine Left Hand* state (Fig. 9). This state is characterized by a very low power for right instrument and left radial coordinate.

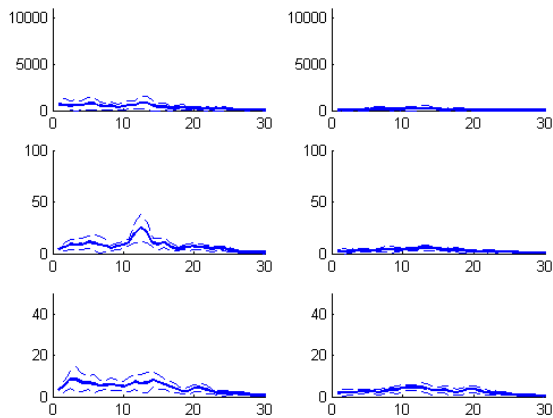


Fig. 9. Power spectrum of the polar acceleration for the observations recognized as State 4. Continuous bold line represents the mean μ , while the dashed lines represent $\mu \pm \sigma$ (σ = standard deviation). In the left and right column are depicted radial, azimuthal and transversal accelerations spectra respectively of left and right instruments.

A study of the transition matrix A shows that essentially transition probabilities between the first 3 states remain the same as in the 3-states model, and the *Moving Fine Left Hand* state is reachable only from the 1st state and to this state it returns:

$$\mathbf{A} = \begin{pmatrix} 0.60 & 0.14 & 0.03 & 0.23 \\ 0.03 & 0.79 & 0.18 & 0.00 \\ 0.25 & 0.11 & 0.62 & 0.03 \\ 0.22 & 0.03 & 0.01 & 0.74 \end{pmatrix}$$

A possible interpretation of this state is that using the left

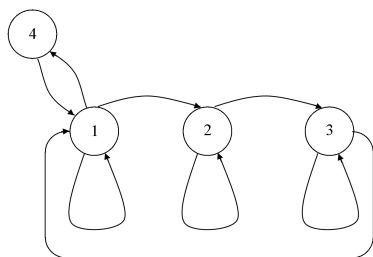


Fig. 10. Representation of the ideal topology of the 3-states model: state 1 corresponds to *Moving Right Hand*, state 2 corresponds to *Moving Left Hand*, state 3 corresponds to *Moving Both Hands* and state 4 corresponds *Moving Fine Left Hand*.

instrument there is a phase of fine movements in the last phase of the ball reaching.

VII. CONCLUSIONS

In this paper we used Hidden Markov Models (HMMs) as a tool for the understanding and characterization of surgical gesture. We created an *expert model* without a

priori assumptions about the different phases of a surgical performance, training it automatically from the data. By analyzing the *expert model*, we show that it is possible to deduce information about the features that characterize the *surgical expertise* required in the examined performances. We made our tests on the basis of data collected during exercises that allow to train the surgeon's ability in the bi-manual movement of surgical instruments and in depth perception using laparoscopic view. In this particular case the model converged to physical and readable states that underline relations between left and right hands movements, according to the characteristic of the exercise. Furthermore, an important results is that the model gives also information concerning the bi-manual coordination, highlighting the different use of left and right tools.

The ongoing work is focused on the statistical validation of the proposed method by training the expert model on data collected from a larger number of expert surgeons, and on data concerning more complex exercises.

REFERENCES

- [1] H. Mayer, I. Nagy, and A. Knoll, "Skill transfer and learning by demonstration in a realistic scenario of laparoscopic surgery," in *IEEE International Conference on Humanoids*, October, 2003.
- [2] S. Cotin, N. Stylopoulos, M. Ottensmeyer, P. Neumann, D. Rattner, and S. Dawson, "Metrics for laparoscopic skills trainers: the weakest link!," in *LNCS 2488* (T. Dohi and R. Kikinis, eds.), (Berlin Heidelberg), pp. 35–43, MICCAI 2002, Springer Verlag, 2002.
- [3] R. Katz, A. Hoznek, L. Salomon, P. Antiphon, A. de la Taille, and C. Abbou, "Skill assessment of urologic laparoscopic surgeons: can criterion levels of surgical performance be determined using the pelvic box trainer?," *European Urology*, vol. 47, pp. 482–487, 2005.
- [4] S. Payandeh, A. Lomax, J. Dill, C. Mackenzie, and C. Cao, "On defining metrics for assessing laparoscopic surgical skills in a virtual training environment," in *Medicine Meets Virtual Reality*, 2002.
- [5] J. Rosen, M. Solazzo, B. Hannaford, and M. Sinanan, "Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using Hidden Markov Model," *Computer Aided Surgery*, vol. 7, pp. 49–61, 2002.
- [6] J. Rosen, J. Brown, M. Barreca, L. Chang, B. Hannaford, and M. Sinanan, "The blue dragon - a system for monitoring the kinematics and the dynamics of endoscopic tools in minimally invasive surgery for objective laparoscopic skill assessment," in *Medicine Meets Virtual Reality*, vol. 85, pp. 412–418, 2002.
- [7] T. E. Murphy, C. M. Vignes, D. D. Yuh, and A. M. Okamura, "Automatic motion recognition and skill evaluation for dynamic tasks.," in *Eurohaptics 2003*, pp. 363–373, 2003.
- [8] R. L. Cave and L. P. Neuwirth, *Hidden Markov models for Speech*, ch. Hidden Markov models for English. Princeton, NJ: IDA-CRD, October 1980.
- [9] G. Burdea, *Force and Touch Feedback for Virtual Reality*. Wiley Interscience Publication, July 1996.
- [10] L. R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition proceedings," *Proceedings of the IEEE*, vol. 77, pp. 257–286, February 1989.
- [11] S. Sinigaglia, G. Megali, O. Tonet, A. Pietrabissa, and P. Dario, "Defining metrics for objective evaluation of surgical performances in laparoscopic training," in *Computer Assisted Radiology and Surgery*, pp. 509–5014, 2005.
- [12] F. Hlawatsch and D. Bordeaux-bartles, "Linear and quadratic time-frequency signal representation," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.
- [13] J. Yang, Y. Xu, and C. S. Chen, "Hidden Markov Model approach to skill learning and its application to telerobotics," *IEEE Transactions on Robotics and Automation*, vol. 10, pp. 621–631, October 1994.
- [14] R. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol. 1, pp. 4–29, April 1984.
- [15] G. Megali, S. Sinigaglia, O. Tonet, and P. Dario, "Modelling and evaluation of surgical performance using hidden markov models," *IEEE Trans Biomed Eng*, 2005, submitted.