

Gesture recognition and classification for surgical skill assessment

G. Saggio, G. L. Santosuosso, P. Cavallo,

C. A. Pinto, M. Petrella, F. Giannini

Dept. of Electronic Engineering, Univ. of "Tor Vergata"
Via del Politecnico, 1 - Rome, Italy
saggio@uniroma2.it

N. Di Lorenzo, A. Lazzaro, A. Corona,

F. D'Auria L. Iezzi, A. L. Gaspari

Department of Surgical Sciences Univ. of Tor Vergata
Via, Montpellier 1- Rome, Italy
alessandralazz@gmail.com

Abstract—The understanding of surgical gesture, by means of measuring apparatus, can play a key role for a possible evaluation of the surgical performance and the human factors characterizing it. To this aim a neural network classification algorithm can be helpful, since combines good generalization performances along with a parsimonious architecture when dealing with high dimensional classification problems. So, here it is presented and proposed the development of an innovation in surgical training system, as a fundamental objective support for training of novice surgeons.

Keywords—gesture recognition, data glove, classifiers, surgical training

I. INTRODUCTION

Despite the enormous improvements in effectiveness of surgical treatments of the latest years, the criteria of evaluating the surgical skills of the trainees still remains mainly based on subjective criteria. The assessment relies on expert examiners who judge the skills of the learners on behalf of observation and experience. In addition by now the learning model in apprenticeship is mainly based on observation, imitation and instructions. In such a way the mentors transfer their expertise to trainee primarily with the traditional approach of implicit mode, because of the clear reason that a didactic setting does not provide truly helpful tools.

Within this frame, any new tool which is capable to provide objective parameters, as a measure of the surgical skill capabilities of trainees, is desirable and welcome. For this reason, this paper intends to propose an innovative method to measure, quantify but, primarily, store, analyze and classify the clinical gesture in surgery.

The importance of measuring the human gesture is evident in itself and underlined by the fact that many applications deal with motion and gesture signal obtained by *ad hoc* electronic transducers. Examples can be character animation, biomechanics, musical gesture interfaces and virtual surgery too. Stands this importance, even some file formats have been defined as the BVA/BVH (BioVision Hierarchical data), the AMC (Acclaim Motion Capture data) [1], the ASK (Alias Skeleton), the HTR (Hierarchical Translation Rotation), .. now recognized as standards. Closely related to the medical field, there are examples of gesture measurement especially devoted

to rehabilitation purposes [2], and a interesting application was proposed as a video-based hand gesture capture and recognition system, utilized to manipulate magnetic resonance images within a graphical user interface [3]. According to our knowledge, some efforts have been paid to utilize the measure of the gestures of trainees to quantitatively evaluate their skill, or to furnish them an objective method to correct wrong postures during surgical sessions. One of the few examples comes from the skills assessment tool based on a Body Sensor Network by means of a fiber optic utilized to measure just the index finger movements for laparoscopic procedures [4].

With our system we realize a pervasive acquisition, storage, analysis and classification of the gestures (stands the kinematic, dynamic and physiologic data of the surgeon's hand postures) during a simulated or real operation, making possible to obtain a quantitative knowledge about the surgeon "action", and to define the evaluation metrics of the surgical performance. This is to "quantify" the operator's surgical skills and ergonomics, and this is our goal.

II. HISTORICAL BACKGROUND

In the new modalities of surgical education, basic surgical skills are learned and practiced on models and simulators, with the aim of better preparing trainees for the operating room experience.[5, 6]. The understanding of surgical gesture in open surgery represents a big step forward towards the improvement of surgical training procedures and of surgical instruments. The wide acceptance of the objective structured clinical examination (OSCE) led a group in Toronto to develop a similar concept for the assessment of technical skills. OSATS (objective structured assessment of technical skills) is a method used to standardized the evaluation process outside the operating room. It consists of six stations where residents and trainees perform procedures on live animal or bench models in fixed time periods [7]. The observers mark the performance of the task using two validated marking systems, a checklist and a global scoring sheet. The checklist comprises a series of yes/no items which have been developed by analysis of the task and also of the specific tuition that has been provided in skills training sessions [8]. The global scoring sheet includes multiple items which were selected to be general markers of technical skill, not procedure-specific steps; each of these scored with a

value from 1 to 5 to help in the assessment of performance [7]. They are useful in assessing technical skills in terms of knowledge and dexterity aspects but they do not offer the scope to assess judgment as the tasks are highly standardized. Other methods of assessment include the MISTELS (Mc Gill Inanimate System for Training and Evaluation of Laparoscopic Skills) as a program to assess a resident's performance of basic Laparoscopic Skills, and the ICSAD (Imperial College Surgical Assessment Device), in both laparoscopic and open procedures, that uses motion analysis to determine how many movements a subject uses to perform a standardized surgical task using a sensors placed on the trainee's hands during the performance [9, 10]. Objective assessment is essential because deficiencies in training and performance are difficult to correct without objective feedback.

III. MATERIALS AND METHODS

In order to preliminarily delimit the working area we operated a bounding box decomposition to surgeon's hand movements analysis and gesture recognition during training of invasive surgery. This feature was applied to analyze trajectories of surgeon's wrist and finger postures, so to recognize different hand gestures.

A. Dataset of surgical gestures

5 master surgeons, 5 resident surgeons and 5 attending surgeons volunteered to participate in the study. The tasks studied in this assessment have included:

- Skin pad incision
- Tissues dissection
- Interrupted stitch
- Running suture
- Knot Tying exercise

B. Measure of gestures

In order to measure the hand gestures of expert surgeons and trainees, we developed a data glove on the basis of acquired experiences [5].



Figure 1. The so called "Hiteg" glove since our acronym "Health Involved Technical Engineering Group"

This glove is provided with sensors capable to measure the movements of the Distal Interphalangeal (DIP), Proximal Interphalangeal (PIP), MetacarpoPhalangeal (MCP) finger joints, the wrist and the elbow postures, and to sense the pinch thanks to pressure sensors placed in correspondence of the fingertips of the thumbs. The accuracy and repeatability of the data acquired with this data glove are similar to those of other ones commercially available and reported in literature [6], but the number of degree of freedom which we measure are the largest ever reported. The fig. 2 shows a hand's posture measured by the glove and replayed on a pc screen by an ad-hoc realized avatar. On the picture's sides are visible two hardware's board designed for the data acquisition.

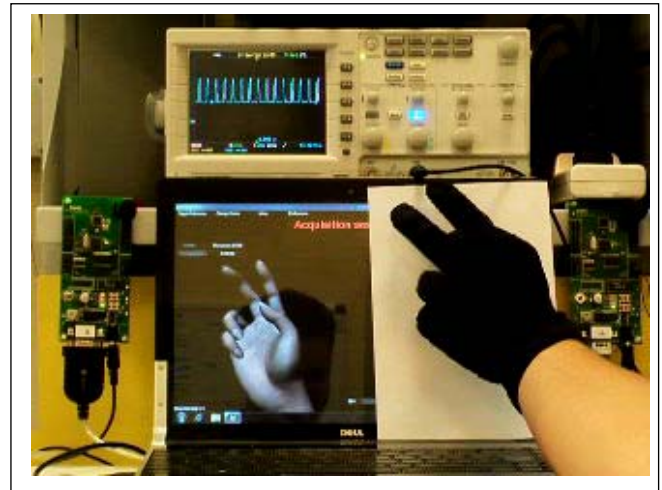


Figure 2. Data are acquired from the glove and the measured movement are replayed by an avatar on a pc screen

At this stage, in order to simplify our preliminary work, we decided to do not take into account the movements of the forearms, which we believe not meaningfully affect the results in any case. The following fig. 3 illustrates a typical surgical gesture, i.e. a running suture made for practice and measured by our system of a couple of sensorized gloves. There can be noticed added sensors on the surgeon's arms to measure the postures of the elbows.

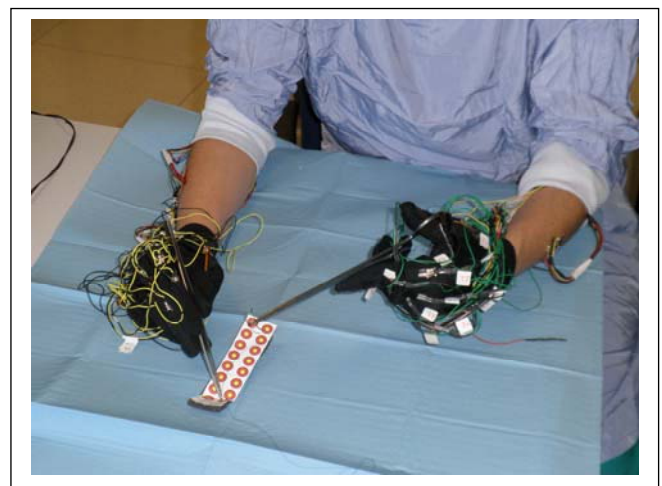


Figure 3. A typical surgical gesture measured by our system

Despite the complexity of the connection wires, according to different operators, the gloves seem to do not substantially limit the natural gestures.

Two twins hardware architectures are used to record the data from each of the gloves, and the synchronization between them is assured by a common timer.

C. Gesture classification

The next step in our work is concerned with the synthesis of an algorithm that automatically assigns each gesture to a predefined class. The most popular strategies include: support vector machines (SVM), classifiers based on mixture density models (CMM), fuzzy classifiers (FCL), radial basis function neural networks (RBF) and multilayer perceptron neural networks (MPN) (see [8] for a tutorial on this issue).

SVM has been shown in [11] to be an effective tool to classify hand gestures recorded by using data gloves to generate variable-length data streams. In the context of performance evaluation, SVM is a *discriminative* classifier based on the principle of structural risk minimization, which is expected to have an optimal classification performance on new data (generalization), at the cost of computationally complex learning algorithms. To be specific, for binary classifications of linearly separable patterns, SVM aims at separating two classes in the input space via a hyperplane 'optimally' defined. A hyperplane is called 'optimal' if it has a maximal margin. The term 'margin' refers to the distance between the hyperplane and the closest input samples of the two classes, the so-called support vectors situated on two hyperplanes which are parallel to the separating hyperplane. When dealing with multidimensional non separable patterns, the samples are mapped into a (typically higher-dimensional) so-called feature space in which a linear separation is expected to be possible. This is done by means of kernel functions. Often, Gaussian or polynomial kernels are used. A new cost function is introduced, and the optimization problem is solved by specific algorithms. For large problems they rely on heuristics for breaking the problem down into smaller ones, like the sequential minimal optimization (SMO) algorithm, which solves two-dimensional sub-problems analytically..

The choice of CMM, on the other side, can be regarded as a *generative* classifier, i.e. additionally aims at modeling the processes from which the observed data originate. The idea behind CMM is to estimate the posterior probabilities density functions of the different classes based on the available samples. CMM achieves good generative properties, but it fails for sparse data or data that are very far from uniform distribution.

FCL derive from the theory of fuzzy sets; their elements are characterised by membership degrees. Fuzzy sets allow gradual membership determination for each element in a set via the 'membership functions' valued in the interval between 0 and 1. The most frequently used type of fuzzy systems are based on triangular membership functions and max–min inference, or Mamdani-type fuzzy system with Gaussians and singletons as membership functions, sum-prod inference, and height method for defuzzification. FCL are *comprehensible*

classifiers, i.e. are also intended to be parameterized and understood by human domain experts, but at the cost of lower classification rates. Furthermore, the task of parameter tuning by the designer is rather complex when dealing with higher-dimensional input spaces.

Neural networks classifiers such as radial basis function neural networks (RBF) and multilayer perceptron neural networks (MPN) (see [14] for recent advances on his topic) are shown to be efficient tools for related classification tasks, as in [13], [15]. They are both *parsimonious* in their architectures for input space with many dimensions and admit 'simple' learning algorithms.

Radial basis functions (RBF) form artificial neural networks having three layers of neurons: an input layer, a hidden layer and an output layer. For a classification problem, each output neuron is associated to a class. Input and hidden layers are feedforward connected as well as hidden and output layers. The activation of a hidden neuron is computed by a Gaussian function using the Euclidean distance between a weight vector of parameters and the external input vector of the network. Each output neuron computes its activation as a weighted sum of the Gaussian functions, so that when the trained RBF is tested to classify new data, the final decision on class membership is made with a winner-takes-all approach. A central issue about RBF concerns the training algorithms employed to determine the suitable values for the network parameters i.e. centers, radii along with output weights of an RBF network, given a set of training patterns. An accurate learning strategy addresses the problem by selecting the centers of the basis functions by means of unsupervised clustering algorithms, as k-means. The radii are then chosen via a nearest neighbour algorithm or the empirical variances of the clusters. The weights in the output layer are computed solving a classical linear least-squares problem. In a simpler approach for training RBF, the network parameters are determined applying an iterative, stochastic training procedure, as the scaled conjugate gradients algorithm.

Multilayer perceptron neural networks (MPN) are characterized by an input layer, an arbitrary number of hidden layers and an output layer. Again, for a classification problem, each output neuron performs the linear weighted sum of its inputs, via a set of tunable parameters associated with the neuron itself. Each hidden layer neuron output is obtained by 'squashing' the weighted sum, plus a bias, into a nonlinear sigmoidal function, i.e. a bounded monotone increasing function, which in turn feeds all the neurons of the next layer. Notice that, by the approximation theorems, a MPN with three layers can virtually approximate any nonlinear function, including Gaussian functions of RBF networks, so that MPN can be regarded as architectures generalising the ones of RBF. Training algorithms for MPN include traditional iterative, stochastic training procedures, for example a gradient-based (i.e., first-order) technique such as backpropagation or resilient propagation.

Notice that the mathematical formulation of the classification problem associated with hand gestures involves an input space with many dimensions and relatively few

output classes, so that 'curse of dimensionality' problems must be avoided. Therefore, the classifier structure has to be a *parsimonious* nonlinear architecture (see [16]), i.e. presenting a polynomial complexity increase with respect to the input space dimension increase as well as having a simple training algorithm.

In conclusion, to comply with the requirement above, we implement RBF and MPN for an efficient and parsimonious hand gesture classification on the basis of the data collected in our experiments.

D. Operator's training

Currently mentors transfer their expertise to trainee via practical demonstrations and oral instructions. But with recorded data of measures of expert surgeon gestures, it is possible to reproduce such movements via avatar representation on a pc screen. It gets the important aspect that the same gesture can be represented several times always in the same manner, and that it is possible to look at the gesture from all possible points of view, just rotating, translating, zooming the avatar.

The avatar representation can be not only for the gestures of the mentors, but also for the movements performed by the trainees. So, we developed a graphical interface capable to superimpose a "ghost" avatar of the learner upon the "guide" avatar of the expert (see Fig. 4). In this manner the trainee is capable to easily auto-evaluate his/her performance, with instinctive ability.



Figure 4. Training session. A ghost view of the trainer hand is superimposed on the avatar hand of the expert surgeon

IV. DISCUSSION

Surgical skills are required by a wide number of health care professionals. It is demonstrated in a broad range of tasks from simple wound closure to highly complex diagnostic and therapeutic procedures. Technical expertise, although essential, is only one component of a complex picture. The measure of hand movement in 3D space is important for several applications in surgical training. Recent new modalities in

surgical technique are inseparably linked to teaching and learning and the pace of this change is constantly accelerating. It must be stressed that reconstruction of surgical tasks through electronic modalities has been dedicated to laparoscopic procedures in almost all cases. This can be easily explained with the similarity of true laparoscopy, with a screen-mediated bidimensional view and a trocar-mediated access to the abdominal cavity to electronic visual and tactile simulation. Open surgery, on the other hand, acts in a tridimensional open space, with the interaction with the surrounding environment. Most aspects of medicine have historically been learnt in an apprenticeship model by means of observation, imitation, and instruction. In such a setting, much of the expertise transferred from mentor to trainee is implicit, and cannot be transferred easily to a didactic setting. Subjective assessment relies on expert examiners to judge the skills of the trainees based on observation.

Dexterity is an important component in motor skills' acquisition [17]. Manual dexterity has a number of components such as aiming ability, arm-wrist speed and fine finger dexterity. As a general principle, repetitive performance of a specific task results in a much better performance of that task than does general training in a variety of skills [18].

We focus on the training in open surgery tracking and recording movements of surgeons [19]. Surgical procedures, in fact, can be viewed as a series of gestures performed in a sequence with an end objective [20]. By recognizing these gestures, it is possible to determine how the task is performed and the intermediate steps leading to the task.

Existing research has shown that different surgeons can have different approaches to performing a given task, and their maneuvers are characteristic of the basic skills attained. For instance, an expert surgeon may require much less movements than a novice in performing a similar task, therefore identifying the intrinsic pattern of the hand movement can provide important information on basic surgical skills.

V. CONCLUSION

This work would be an innovative, accurate and non-invasive method to measure and evaluate surgical gestures. It will be useful to accelerate the attending surgeon's learning curve who can compare the basic level of his expertise with master surgeon's level and verify step by step his improvement.

REFERENCES

- [1] J. Lander "Working with motion capture fileformats" Game developer, Jan 1988
- [2] G. Saggio, M. De Sanctis, E. Cianca, G. Latessa, F. De Santis, F. Giannini "Long Term Measurement of Human Joint Movements for Health Care and Rehabilitation Purposes" Wireless Vitae09 - Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, Aalborg (Denmark), 17-20 May, 2009 - pp. 674-678
- [3] J. P. Wachs, H. I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, M. Smith "A gesture-based tool for sterile browsing of radiology images" J Am Med Inform Assoc. 2008 May-Jun; 15(3): 321-323

- [4] R. King, Benny P.L. Lo, A. Darzi, G-Z. Yang "Hand gesture recognition with body sensor networks" in BSN '05, International Workshop on Wearable and Implantable Body Sensor Networks, April 12-13, 2005
- [5] Giovanni Saggio, Stefano Bocchetti, Carlo Alberto Pinto, Giancarlo Orengo "Wireless Data Glove System developed for HMI" ISABEL2010, 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies, Roma, Italy, November 7-10, 2010
- [6] L. Dipietro, A. M. Sabatini, P. Dario "A Survey of Glove-Based Systems and Their Applications" IEEE Transactions on Systems, Man, and Cybernetics—part C: Applications and Reviews, vol. 38, no. 4, July 2008.
- [7] Krishna Moorthy, Yaron Munz, Sudip K Sarker, Ara Darzi; Objective assessment of technical skills in surgery- Clinical Review; BMJ Volume 327 1 November 2003
- [8] A Darzi, S Mackay Assessment of surgical competence; Quality in Health Care 2001;10(Suppl II):ii64–ii69.
- [9] Fitts PM, Posner MI, Human performance. Belmont, CA: Brooks/Cole, 1967
- [10] Kopta JA. The development of motor skills in orthopaedic education. Clin Orthop 1971; 75:80-5
- [11] C. Li · P.R. Kulkarni · B. Prabhakaran, "Segmentation and recognition of motion capture data", Multimed Tools Appl vol. 35, pp. 55–70, 2007.
- [12] D. Fisch, B. Kühbeck, B. Sick, S. J. Ovaska "So near and yet so far: New insight into properties of some well-known classifier paradigms" Information Sciences, vol. 180, pp.3381–340,1 2010.
- [13] W. Pawel, T. Amaral, O. Dias A. Wolczowski M.Kurzynski, "Hand movement recognition based on biosignal analysis" Engineering Applications of Artificial Intelligence Vol. 22, Issue: 4-5, pp. 608-615, June, 2009.
- [14] I. Mrazova, D. Wangb, "Improved generalization of neural classifiers with enforced internal representation", Neurocomputing, Vol. 70, pp. 2940–2952, 2007.
- [15] B. Guillaume, B. Stéphane, L. Minotti, M. Antonakios, R. Guillemaud, P. Kahane "Classification of epileptic motor manifestations using inertial and magnetic sensors", Journal: Computers in Biology and Medicine, Vol. 41, pp. 46-55, Jan, 2011.
- [16] G. Dreyfus, "Neural Networks - Methodology and Applications", Springer,2005.
- [17] Dankelman J, N. Di Lorenzo "Surgical Training and Simulation" Minim Invasive Ther Allied Technol.;14 (4):211-3. 2005.
- [18] N. Di Lorenzo, G. Coscarella."Il Paziente Virtuale come nuova modalità di Apprendimento". Monografia "Tecnologia e sicurezza in sala operatoria". SIC pag. 60-63 – 2003
- [19] G. Coscarella, N. Di Lorenzo, P. Gentileschi, L. Faraci, F. Liroso, A. De Majo, F. Russo, M. Pietrantuono, A. Gaspari "I Sistemi Robotici sono utili nella formazione del Giovane Chirurgo" Tumori, Suppl. to vol 89, n 3, pg 44 May-Jun 2003.
- [20] V.N. Paltern, T.P. Grantcharov "Virtual reality in surgical skills training" Surg Clin North Am 2010 Jun; 90 (3): 605-617.