Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Recognition of the Surgeon's Motions During Endoscopic Operation by Statistics Based Algorithm and Neural Networks Based ANARX Models ⋆

Sven Nõmm * **Eduard Petlenkov** ** **Jüri Vain** * **Juri Belikov** *
Fujio Miyawaki *** **Kitaro Yoshimitsu** ***

* *Institute of Cybernetics at TUT, Akadeemia tee 21, 12618, Tallinn, Estonia, (Tel: +372-6204196; e-mail: sven@cc.ioc.ee).*
** *Department of Computer Control, Tallinn University of Technology, Ehitajate tee 5, Tallinn, Estonia (e-mail: eduard.petlenkov@dcc.ttu.ee)*
*** *Graduate School of Advanced Science and Technology, Tokyo Denki University, Ishizaka, Hatoyama-machi, Hiki-gun, Saitama, 350-0394, Japan miyawaki@b.dendai.ac.jp, kitaro@atl.n.dendai.ac.jp,*

Abstract:
The problem of recognition and short time prediction of the surgeon's hand motions during surgical endoscopic operation are approached in the present contribution using neural network based nonlinear modeling techniques and statistics based segmentation of the operating room. It is shown that proposed technique provide precise recognition of surgeon's motions.

Keywords: Human adaptive mechatronics, neural networks, Nonlinear systems

1. INTRODUCTION

Accurate recognition of the human motions have a significant importance for successful cooperation between humans and machines. Once robot or computer recognizes human motion it can respond by providing necessary assistance. Since for each particular situation motion or action can be defined only in the context defined by this situation, general framework for motion understanding seems quite a difficult task. At the same time there are many results achieved in the area of human motions recognition and detection Moeslund et al. (2006). In Yam et al. (2004) approach to person recognition by analysis of walking and running was proposed. Parameterized modeling is applied to recognition of human atomic activities in Yacoob and Black (1999). Analysis and classification of trajectories in gesture recognition with Bezier were studied in Shin et al. (2004). Differential geometric approach is used in representations of human actions in A.Yilmaz and M.Shah (2007).

In many cases where necessity to recognize human actions has risen, the robot or machine is meant to replace human which is usually assists another human. Present contribution is devoted to the problem of recognizing medical surgeon's motions during surgical endoscopic operation. This research is a part of large project which final goal is to create a robot which will be able to replace a human scrub nurse (specially trained nurse who directly assists a surgeon during an operation) during endoscopic surgery Ohnuma et al. (2006), Miyawaki et al. (2005). In the framework of present contribution main attention is payed to the classification and definition of surgeon's motions during the endoscopic operation and to the problem of motion recognition by scrub-nurse robot. Following notations proposed in Nõmm et al. (2007) and taking in to account dynamical and statistical points of view "adjusted" system of basic motions is proposed. Unlike Nõmm et al. (2007) where motion recognition was purely neural networks based, present contribution uses NN-based model of surgeon hand just to make short time prediction of the surgeon's wrist coordinates (allows to compensate delay caused by video processing), and motion recognition algorithm is probability based.

The paper is organized as follows. Section 2 describes general scheme of the data flow in Scrub Nurse Robot (SNR) and provides necessary information about robot itself. Mathematical tools employed, are presented in Section 3. Prediction of the surgeon's hand movements described in Section 4. Actions of the surgeon during endoscopic medical operation are explained in Section 5, which also provide definition of motion in the context of present research and classification of surgeon's motions, both form medical and mathematical point of views. Practical implementation illustrated by examples presented in Section 6. Conclusions are drawn in the last section.

2. GENERAL SCHEME OF THE ROBOT

Mechanical part of the robot is shown in Figure 1 . It can be seen that this robot is oriented to perform the tasks of the scrub nurse during medical operation. The robot can move inside the surgical room, take and give instruments

⋆ This work was supported in part by the Estonian Science Foundation ETF grants 6884, 6837, 7667.

Fig. 1. Scrub nurse robot

to the surgeon, and observe surgeon's actions. General scheme of the data flow, depicted in Figure 2, explains work of the SNR.

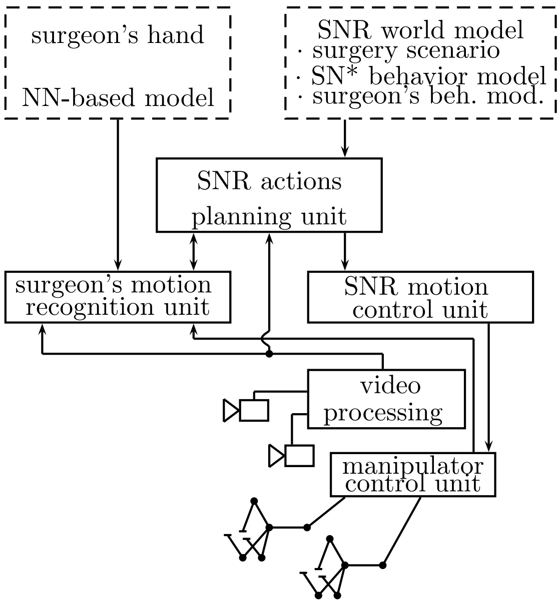

Fig. 2. Dataflow of the scrub nurse robot

Robot observes the surgeon's motions and by simple video processing determines the coordinates of three key points: chest, elbow and wrist. This information together with the data from the manipulator sensors is passed to the motion recognition unit, which recognizes current motion or switching from one motion to another and passes this information to the unit responsible for decision making strategy planning (SNR actions planning unit). Based on the surgery scenario, behavioral models of the surgeon and scrub nurse together with data from the motion recognition unit, SNR action planning unit generates orders for the low level manipulator controllers. In spite of multicore processors and large amounts of memory available above mentioned operations still take some time which leads the necessity to make short time coordinate prediction of the points representing surgeon wrist, elbow and chest.

3. MATHEMATICAL MODELS

3.1 Mathematical tools

The choice of the model was defined by the following three factors:

- The visual sensor of the robot provides the sampled data to the robot "brains"
- Model should allow fast and easy adaptation from one surgeon to another
- advantages of ANARX class model over the ordinary NARX

Since only sampled data is used, discrete-time model is preferable. There are large variety of results available on adaptive control for NN-based models. The advantages of the ANARX structure form the point of view of system analysis and NN-training makes it better choice compare to general NARX structure. Based on the above listed arguments NN-ANARX discrete-time model was chosen to model the surgeon's hand.

3.2 The ANARX structure

In order to make this paper self sufficient the ANARX structure Kotta et al. (2006) and its main properties will be described below. The ANARX is the subclass of the more general Nonlinear Autoregressive Exogenous (NARX) models class

$$\begin{aligned} y(t+n) = f(y(t) + \ldots \\ \ldots + y(t+n-1), u(t), \ldots, u(t+n-1)) \end{aligned} \quad (1)$$

where u is the scalar input, y is the scalar output and $f : \mathbb{R}^n \to \mathbb{R}$ is the analytic function (of its arguments). The main difference which distinguishes ANARX from NARX is that ANARX does not allow arbitrary coupling of different time instances.

$$y(t+n) = f_1(y(t), u(t)) + \ldots + f_n(y(t+n-1), u(t+n-1)) \quad (2)$$

The NN-based ANARX structure is described by the following equation

$$y(t+n) = \sum_{i=1}^{n} C_i \phi_i(W_i z(t+i-1)), \quad (3)$$

where $z(t) = [y(t), u(t)]^T$, C_i and W_i are $1 \times l$ and $l \times 2$ dimensional matrices. Schematic diagram of the ANARX structure represented by a neural network is depicted in Figure 3.

Two main advantages of the ANARX over NARX are: ANARX type system is always realizable in the classical state-space form and it is linearizable by dynamic output feedback. While neither of those advantages are essential on the present stage of the project both can be important

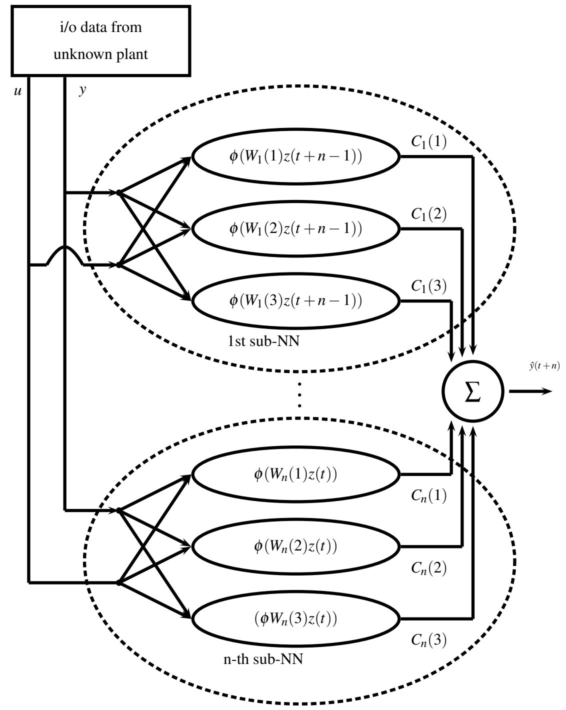

Fig. 3. Additive NARX model represented by a neural network

later, especially form the analysis point of view. Additionally restrictions imposed by ANARAX on neural network topology require less weights to be calculated which result in shorter training time or less computation power needed. Those advantages will become essential when adaptation will be added. In Kotta et al. (2001) and Petlenkov et al. (2006) it was shown that ANARX is suitable to represent NN-based models. Since at each time human hand position is described by coordinates of its wrist chest and elbow generalization of ANARX to MIMO systems type case is used Petlenkov.

4. COORDINATES PREDICTION

4.1 Experimental data acquisition

In order to obtain experimental data from the real surgical operation the operating room was equipped with four cameras. Introduce coordinates as shown in Figure 4. Axis x is directed from the scrub nurse to the surgeon in parallel to the table. Axis y is directed outside of the table orthogonally to x and z is directed vertically. Positions of medical staff during the surgical operation and placement of the cameras depicted in figure 4. In order to simplify monitoring process colored markers were placed on the surgeons chest, elbow and wrist. The cameras were set to film surgical operation at 60 frames per second. Later

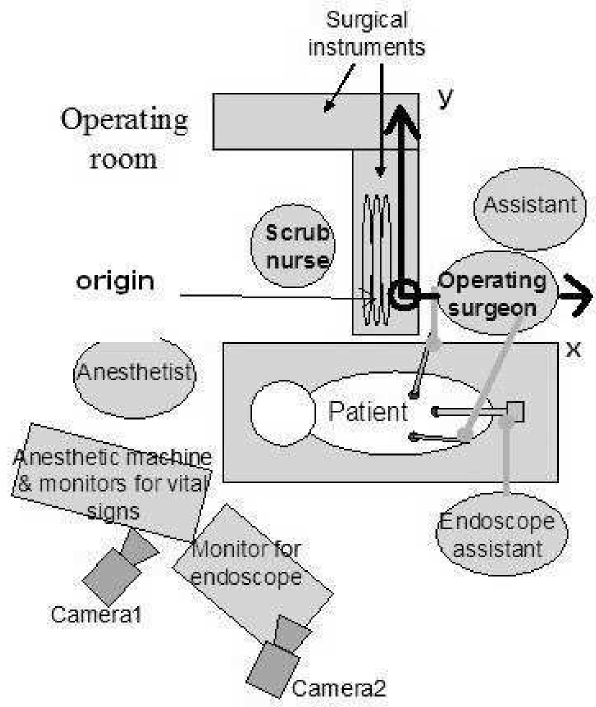

Fig. 4. Schematic diagram of the operating room

on the basis of recorded operation information about the coordinates of chest, elbow and wrist was extracted.

4.2 Prediction of The Wrist Coordinates

NN-ANARX structure based models of chest, elbow and wrist were trained. These models were successfully used for prediction of chest, elbow and wrist coordinates. In this section only prediction of wrist coordinates are shown as an example because this point is the most important to detect current motion of the surgeon. Results obtained for the surgeon chest and elbow are very similar to those shown in the section and they are available from the authors upon request.

For identification of the wrist model we used real measurements of three coordinates (X, Y and Z) of the point in time with sample time 0.017 sec. Data set consisting of 1227 measurements of each coordinate was used as training set. We trained NN-ANARX model with three sub-layers of the hidden layer (3) corresponding to the third order of the model.

$$\Big[X(t), Y(t), Z(t)\Big]^T = \sum_{i=1}^{3} C_i f_i \left(W_i \Big[X(t-i), Y(t-i), Z(t-i)\Big]^T \right) \quad (4)$$

where $X(t), Y(t)$ and $Z(t)$ are the coordinates of chest, elbow and wrist at time-step t. Levenberg-Marquardt (LM) algorithm was chosen to perform training of the network, since it is much more efficient compared to other techniques when the network contains no more than a few hundred neurons Declercq and De Keyser (1996). Also the training speed of LM algorithm is much higher and the feed-forward network trained with it can better model nonlinearity Billings and Fadzil (1985). The structure of trained network representing ANARX structure is shown in figure 1. After training the network different data sets were used for validation of the model. Figure 5 represents results of prediction simulation for 5 time steps.

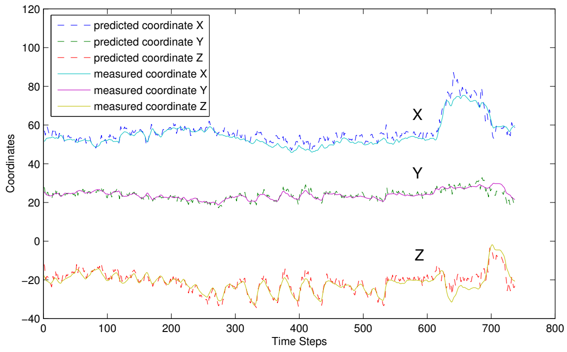

Fig. 5. Prediction of surgeon wrist coordinates for 5 time-steps

It can be seen from figures that the model is capable of predicting wrist coordinates up to 5 time-steps with high degree of accuracy. These predicted coordinates can be used for detecting current motion of the surgeon described in the next section.

5. SURGEON'S ACTIONS, MOVEMENTS AND MOTIONS

Well trained human scrub-nurse is planning his/her actions based on knowledge of surgery scenario and observing surgeon's motions. From the human scrub nurse point of view the surgeon's movements during endoscopic surgery can be segmented in a following way Nõmm et al. (2007). Receiving the instrument, inserting it in to the pipe, working, extracting, passing the instrument and finally waiting for the new instrument. Those stages were considered as basic, in some sense primitive parts, of the surgery and named as *motions* in Nõmm et al. (2007). Such segmentation is also suitable for SNR actions planning unit realized by timed-automata. While all above mentioned motions can be detected Nõmm et al. (2007), the accuracy of such detection suffer from the fact that initial segmentation was made purely from human view point. Analysis performed in Nõmm et al. (2007) clearly shows that detection of motions which take place during shorter periods of time was less accurate compare to longer lasting motions. This leads the necessity to adjust motion's segmentation in such a way that on one side different motions will be distinguishable by their dynamical or statistical characteristics and on the other side will be suitable for the design of SNR actions planning unit.

5.1 Motions

We divided intraoperative surgeon's motions into the following five motions. Unlike Nõmm et al. (2007) where surgeon's movements were divided into six motions, present contribution does not distinguish between 'extracting' and 'passing' due to their similarity. This poses one more open problem for the future research.

(1) 'inserting' -defined as the motion observed from the moment that instrument is received by surgeon's hand up to the moment when the surgeon inserted it into the abdominal cavity through the cannula (cylindrical tube)
(2) 'working' defined as the motion observed while the surgeon was conducting some kind of surgical procedure with a surgical instrument
(3) 'returning' defined as the motion which started from the moment that the surgeon began to draw the surgical instrument out of the the abdominal cavity (operative field) and which lasted until the surgeon released the instrument by passing it to the scrub nurse or returning it to a surgical tray by himself.
(4) 'waiting' - defined as the motion continuing from the moment that the surgeon released the instrument to the moment when he received the next instrument from the scrub nurse.
(5) 'get' -defined as the motion observed while surgeon takes the instrument from the nurse. (compared to other motions *get* takes place during extremely short period of time, in spite of it, "get" will be called motion in the sense that it is important stage of the surgery).

One can spot that differently to Nõmm et al. (2007) one motion disappears and remaining motions slightly shifted compared to the initial segmentation. Motion 'inserting' starts immediately after getting the instrument and lasts until the instrument reaches the abdominal cavity. While 'working' remains the same 'extracting' and 'passing' are joint into one motion 'return' which is symmetrical, in some sense, to 'inserting'. On one side 'extracting' and 'passing' are nearly indistinguishable and on the other side distinguishing does not give any advantage to the SNR actions planning unit. As it was already mentioned 'get' remains in the list of motions only because of its role in the surgery scenario and its importance for the actions planning. 'Get' would not be detected on the basis of video data but by the force sensors of the SNR hand manipulator.

Observing 110 experiments one can calculate an 'average switching point' for each pair of neighboring motions simply by calculating average for each of coordinates x y z. Lilliefors test shows, that all switching points are normally distributed. This allows to calculate probability that next motion has started (switching probability) as a function of the distance between actual point of the surgeon wrist and average switching point for current motion.

$$p = \int\int_D\int \frac{1}{(2\pi)^{\frac{3}{2}} \mid \Sigma \mid^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)} \mathrm{d}x\mathrm{d}y\mathrm{d}z \quad (5)$$

where Σ is the covariation matrix, $X = (x, y, z)$ coordinates of current point and $\mu = (\mu_x, \mu_y, \mu_z)$ vector of coordinates of average switching point.

Based on the same 110 experiments, liminal values of switching probabilities for each pair of consequent motions are calculated. Denote liminal value of switching probability as $P_{lim}n$ where n is a number of motion. 5 This leads

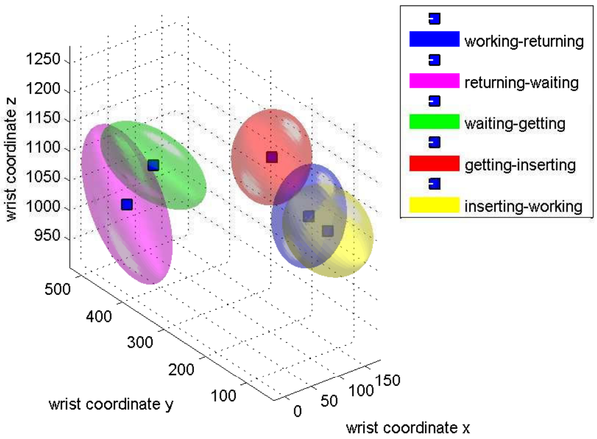

Fig. 6. Working space segmentation by ellipsoids

to the segmentation of working space by the ellipsoids with a center at corresponding average switching point and radiuses defined by liminal switching probability. Such segmentation is depicted in Figure 6

5.2 Motion detection algorithm

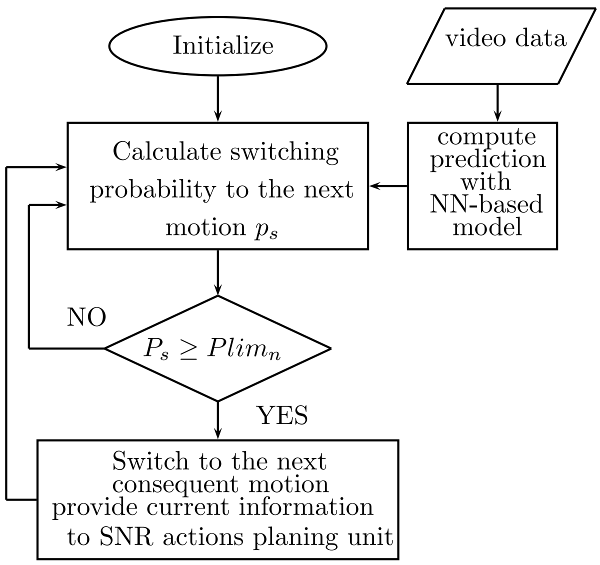

Fig. 7. Surgeon motions detection algorithm

Algorithm depicted in Figure 7 requires to be initialized. In real life it would require surgeon to put hands in a certain position. On the present stage initializations simply define starting motion. On the basis of 5 time-steps prediction, computer controlling the scrub nurse robot computes the probability of switching to the next motion and compares it to liminal probability of switching. If current switching probability is greater or equal of liminal then algorithm updates value of current motion and sends a signal to the SNR actions planning unit about motion switching time and coordinates of the surgeon wrist. Comparison of the motion detection results (on the data which was not used for segmentation) obtained by algorithm and motions observed by human depicted in Figure 8

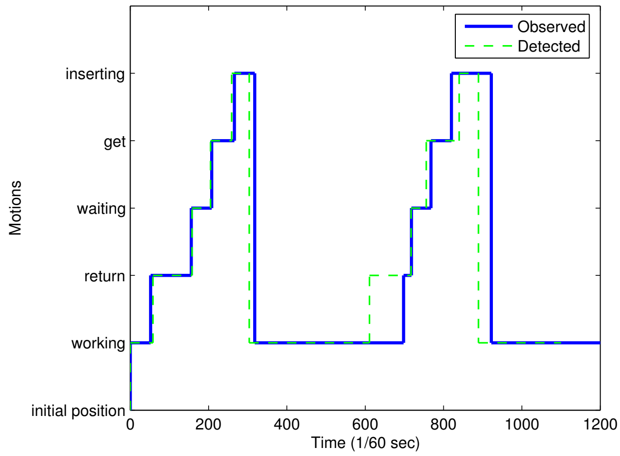

Fig. 8. Working space segmentation by σ ellipsoids

6. DISCUSSION

Compared to the method described in Nõmm et al. (2007), where detection is made on the basis of 7 NN-based models, present algorithm provides more accurate results. In average detection delay is less than 1 sec. Also the algorithm depicted in Figure 7 does not require running in parallel 7 models and consequently less computational power is needed. In real life SNR should be able to work with different surgeons. The algorithm proposed in Nõmm et al. (2007) is more adaptation friendly in the sense of training due to the fact that it is purely NN-based. It can be adapted to the different surgeons during initialization by measuring distances between key points. While conducting normality tests for switching points, high significance levels were observed. While it does not necessary influence the accuracy of detection there can exist another distribution which can be more suitable to model switching points. Another weakness of present technique is that it is yet unable to handle unexpected switchings and exceptional movements.

7. CONCLUSIONS

Movements of surgeon's hand during the endoscopic operation have been segmented by a system of five "primitive" (basic) parts called motions. System of motions is an important part of the operation scenario which is used by scrub nurse robot actions planning unit. Statistic based algorithm is proposed to detect switching from one motion to another by observing current coordinates of surgeon wrist. NN-based ANARX model of the surgeon hand was used to "filter" coordinate data and provide five time-step prediction. MATLAB simulation is used to illustrated the results of the paper. Future work will be pointed on developing more accurate and robust detection algorithm with ability to adapt to a different surgeon's.

REFERENCES

A.Yilmaz and M.Shah. Differential geometric approach to representing human actions. *Computer Vision and Image Understanding*, 2007.

S. A. Billings and M. B. Fadzil. The practical identification of systems with nonlinearities. *Proc. of 7th*

IFAC/IFORS Symp. Identification Syst. Parameter Estimation, pages 155–160, 1985.

F. Declercq and R. De Keyser. Comparative study of neural predictors in model based predictive control. In *Proc. of Neural Networks for Identification, Control, Robotics, and Signal/Image Processing International Workshop on*, pages 20–28. Bratislava, 1996.

Ü. Kotta, S. Nõmm, and F.N. Chowdhury. On a new typr of neural-network-based input-output model: The anarma structure. *Prock of the 5th IFAC Symposium on nonlinear control systems NOLCOS, St. Petersbourg, Russia*, July 2001.

Ü. Kotta, F. Chowdhury, and S. Nõmm. On realizability of neural networks-based input-output models in the classical state space form. *Automatica*, 42(6):1211–1216, 2006.

F. Miyawaki, K. Masamune, S. Suzuki, K. Yoshimitsu, and J. Vain. Scrub nurse robot system - intraoperative motion analysis of a scrub nurse and timed-automata-based model for surgery. *IEEE Industrial Electronics Transaction on*, 5(52):1227–1235, 2005.

T. B. Moeslund, A. Hilton, and V. Krüger. Asurvey of advances in vision based human motion capture and analysis. *Computer Vision and Image Understanding*, (104):90–126, 2006.

S. Nõmm, E. Petlenkov, J. Vain, K. Yoshimitsu, K. Ohnuma, T. Sadahiro, and F. Miyawaki. Nn-based anarx model of the surgeon's hand for the motion recognition. In *Proceedings of the 4th COE Workshop on Human Adaptive Mechatronics (HAM)*, pages 19–24. Tokyo Denki University, Tokyo, Japan, March 2007.

K. Ohnuma, K. Masamune, K. Yoshimitsu, J. Vain, Y. Fukui, and F. Miyawaki. Analysis and recognition of a surgeon's motions in laparoscopic cholecystectomy giving a scrub nurse robot suitable timings for instrument exchange. In *Proc. of the 3rd COE HAM Workshop*, pages 1–5. Tokyo Denki University, Japan, March 2006.

E. Petlenkov. Nn-anarx structure based dynamic output feedback linearization for control of nonlinear mimo systems. In *Proc. Of the 15th Mediterranean Conference on Control and Automation (MED'07)*. June .

E. Petlenkov, S. Nõmm, and Ü. Kotta. Neural networks based anarx structure for identification and model based control. In *Proceedings of the 9th IEEE International Conference on Control, Automation, Robotics and Vision*, pages 2284–2288. Singapore, December 2006.

M.S. Shin, L.V. Tsap, and D.B. Goldgof. Gesture recognition using bezier curves for visualization navigation from registrated 3-d data. *Pattern Recognition*, (37): 1011–1024, 2004.

Y. Yacoob and M.J. Black. Parametrized modeling and recognition of activities. *Computer Vision and Image Understanding*, 73(2):232–247, 1999.

C.Y. Yam, M.S. Nixon, and J.N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, (37):1057–1072, 2004.