

Received September 11, 2021, accepted October 17, 2021, date of publication October 29, 2021, date of current version November 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3124304

Sensorless Real-Time Force Estimation in Microsurgery Robots Using a Time Series Convolutional Neural Network

JUUYUN XIA^{ID}, (Student Member, IEEE), AND KAZUO KIGUCHI^{ID}, (Senior Member, IEEE)

Department of Mechanical Engineering, Kyushu University, Fukuoka 819-0395, Japan

Corresponding author: Kazuo Kiguchi (kiguchi@ieee.org)

This work was supported in part by the Okawa Foundation for Information and Telecommunications under Grant 20–18.

ABSTRACT Robotic-assisted microsurgeries provide several benefits to both patients and surgeons. Nevertheless, there are still some limitations and challenges associated with their outcome, one of which is a lack of force feedback. Without force information, the risk of delicate tissue damage from the excessive force applied by surgeons would be increased. Since it is difficult to install force sensors on microsurgical tools, a novel approach for estimating a force vector from the deformation of the surgical tool is proposed in this paper. In the proposed approach, a surgical instrument that deforms according to the magnitude of the tool-to-tissue force is designed, and a time series convolution neural network is used to make the nonlinear relationship between the visual information of the deformation of the surgical tool and the applied forces in such a way that the tool-to-tissue force can be estimated according to the deformation of the surgical instrument in a real-time manner. The experimental results prove that the applied force can be successfully estimated with high accuracy in three dimensions using the proposed method.

INDEX TERMS Force estimation, surgical tool, visual information, machine learning, microsurgery.

I. INTRODUCTION

Microsurgery is an operation for suturing vessels or nerves under a microscope and is regarded as one of the most technically demanding surgical disciplines because it requires precise motion to manipulate delicate tissue in a small and constrained workspace [1], [2]. Microsurgeries have also been widely used in the areas of ophthalmology, otolaryngology, and neurosurgery. However, the accuracy of the surgeries is limited by the surgeons' capabilities. The robotic-assisted surgery system with a master-slave configuration brings several potential advantages in the field of microsurgery, such as enhancing the surgeons' dexterity and allowing precise manipulation by providing motion scaling and tremor filtering [3]–[5].

Although microsurgical robotic systems have several benefits to both surgeons and patients, there are still some limitations and challenges associated with their outcome, one of which is a lack of force feedback. Force feedback is a major feature that can improve the microsurgical performance since

it enables surgeons to control the interaction forces [6], and thus, it helps in the proper execution of surgical procedures. In addition, force feedback can also enable surgeons to feel extremely small forces; as a result, surgeons can avoid exertion of excessive force and reduce the risk of damage to delicate tissues. Without force information, it is difficult for surgeons to feel how much force is applied to delicate tissues, and excessive force might result in irreversible damage [7]–[9]. To address this problem, a method to measure the interaction force between a surgical instrument and a tissue on the slave side and a method to feedback the force to the surgeon on the master side should appropriately be considered in robotic-assisted microsurgery. This paper proposes a method to estimate the interaction force vector on the slave side.

To date, many researchers have studied force feedback for robotic-assisted minimally invasive surgery (widely used in endoscopic surgery such as laparoscopic surgery). Existing force estimation methods in the surgery robot can be roughly divided into two methods: direct force sensing and indirect force estimation [10], [11]. In the direct force sensing method, the tool-to-tissue force is measured by integrated sensors that are installed on or close to the surgical tool. For example,

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine^{ID}.

several sensorized tools have been developed with force sensors or strain gauges mounted on the shaft or the base of the surgical tool [12], [13]. Other researchers [14], [15] have measured the force using force sensors installed on the tip of the surgical tool. Furthermore, a surgical forceps tool with the capability of force sensing, including pull and grasp forces has been developed using strain gauges attached to the complaint hinge [16]. The main advantage of this method is that the forces can be measured directly from the tool-to-tissue interaction. Nonetheless, this approach still suffers from many constraints, such as biocompatibility, sterilization, size, and cost [17], [18]. For these reasons, indirect force estimation has been proposed as an alternative solution to estimate the interaction forces. In contrast to the direct force sensing method, the indirect force estimation method involves estimating forces through driven system information, such as the current, torque and displacement, or vision, instead of direct force measuring. For example, force sensing was conducted based on changes in driving cable tension for a 3-DOF manipulator [19]. In [20], a dynamic model-based estimator was proposed to estimate cable pretension in cable-driven robots. However, the use of the driven system information for force detection could affect its accuracy due to its nonlinear characteristics.

The vision-based approach, which is called vision-based force measurement (VBFM) [21], was proposed to estimate the interaction forces from the observable displacements of the deformable object. To prove the feasibility and effectiveness of the VBFM, researchers have made a large amount of effort. In the study in [22], a method based on image processing techniques and a continuum mechanics model was proposed to estimate the force, which can be used in both micro- and macroscale environments. A virtual-template-based approach that estimates the tool-organ force interaction from monocular camera images was presented in [23]. The authors in [24] built a biomechanical model from the organ shape using 3D reconstruction and meshing techniques to estimate the contact forces. These research showed that the vision-based approach is stable, robust, and effective in estimating the applied forces according to the object deformation, whether in macro or microsystems. Thus, although the accuracy provided by the VBFM method is limited, it is still a potential solution to estimate the force.

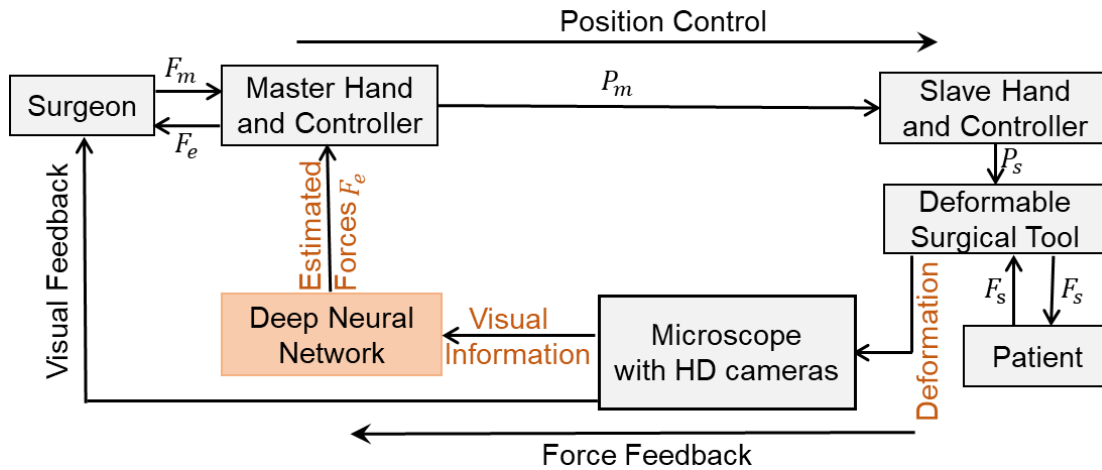
In addition, some researchers have used artificial neural networks to improve the accuracy of vision-based force estimation. In [25], two different neuro-fuzzy inference systems were proposed to identify the tool-tissue force and the maximum local stress in laparoscopic surgery. A recurrent neural network (RNN) approach for 3D vision-based force estimation was proposed in [26]. In their work, the force is estimated through the RNN, which uses kinematic variables and deformation mapping as the input. To increase the accuracy of the learning system, a modification method was presented by the same authors in [27]. The solution extracts the geometry of motion of the tissue's surface first, and then, a deep

network based on an LSTM-RNN architecture is used to find the accurate mapping between the extracted visual-geometric information and the applied force. This vision-based neural network method mainly relies on tissue deformation observed by monocular or stereo vision systems as the input of the deep learning models.

As described in the aforementioned literature, most existing methods have been proposed for laparoscopic surgical robots. Since microsurgery involves suturing blood vessels or nerves under an optical microscope, the size of the surgical instruments is smaller than that in laparoscopic surgery, which makes installing force sensors on microsurgical tools difficult. Moreover, the view of the region of interest in microsurgery is limited, which makes it difficult to observe the deformation of delicate tissues. Therefore, to date, there have been few studies on force feedback in robotic-assisted microsurgery. A 3-DOF force-sensing micro-forceps for robot-assisted vitreoretinal surgery has been developed based on fiber Bragg grating (FBG) strain sensors [28]–[30]. However, this direct detection method also has some shortcomings, such as temperature sensitivity, sterilization, and high cost.

Due to the reasons mentioned above, a novel vision-based sensorless force estimation approach for robotic-assisted microsurgery with the capability of multiaxis force sensing is proposed in this paper. In the proposed approach, a surgical instrument that easily deforms according to the magnitude of the applied force is designed, and a time series convolution neural network (CNN) model is proposed to determine the relationship between the deformation of the surgical tool and the applied forces in such a way that the tool-to-tissue forces can be estimated according to the deformation of the surgical instrument in real time.

Deep neural networks, such as CNNs, have achieved dramatic progress on image recognition tasks, and a number of extensions to process video have also been proposed [31]–[33]. Some of the most popular CNN architectures used today are AlexNet [34], VGG16 [35], ResNet [36] and GoogLeNet [37]. Among them, VGG16 is currently a widely used choice for extracting features from images and has excellent performance. For example, transfer learning using CNNs has been proven to be an effective method that has the benefit of decreasing the training time and can result in lower generalization error [38]. In transfer learning, a neural network model is first trained on a base dataset and task and then repurposes the learned features or transfers them to a second target network to be trained on a target dataset and task [39]. Shin *et al.* [40] accomplished two specific computer-aided detection problems in medical images by fine-tuning CNN models pretrained from a natural image dataset. Esteva *et al.* [41] demonstrated that a pretrained CNN model in the ImageNet dataset can be used in the classification of images that concern skin cancer lesions. Therefore, in the current work, the VGG16 model is used as an example of the pretrained model, and the transfer learning technique is exploited.



P_m : position of master hand P_s : position of slave hand F_m : operation force F_s : tool-to-tissue interaction force F_e : estimated force

FIGURE 1. Flowchart of the senseless force estimation approach in a microsurgical robotic system. We first designed a surgical tool that deforms according to the magnitude of the applied force, and then, visual information from the deformation of the surgical tool is used in a deep learning system that can estimate the applied force.

The remainder of this paper is organized as follows: Section II details the proposed approaches. Sections III and IV present experimental evaluations to demonstrate the feasibility and effectiveness of the proposed approach. Section V presents the conclusions of this paper.

II. VISION-BASED SENSORLESS FORCE ESTIMATION APPROACH WITH A TIME SERIES CNN MODEL

In the proposed approach, deep neural networks are applied together with time series visual information from a deformable surgical tool to estimate the applied force. The force estimation approach is a part of the robot-assisted microsurgical system based on a master-slave configuration, as shown in Fig. 1. The surgeon watches the motion of the surgical tool in the slave hand on a display and controls the master hand to perform microsurgery. The operation of the master hand can be scaled down with a motion-scaling ratio by the controller and transmitted to the slave hand; then, the slave hand directly acts on the patient with the deformable surgical tool. In this way, the slave hand can replicate the surgeon’s motion with greater precision. A high-definition (HD) microscope is used to provide stereoscopic visualization of the surgical area, allowing the surgeon to perform operations with three-dimensional (3D) perception. To provide the surgeon with a sense of touch, the system integrates a force feedback function using the proposed approach, i.e., the interaction forces between the surgical tool and the tissue on the slave hand can be fed back to the surgeon.

In the force estimation approach, visual information on the displacements of the deformable surgical tool is provided to the proposed deep neural networks. In the proposed deep neural networks, a time series CNN model is trained to analyze the visual information and generate an accurate force estimation. Then, the estimated forces are transmitted to the surgeon’s hand by the master hand.

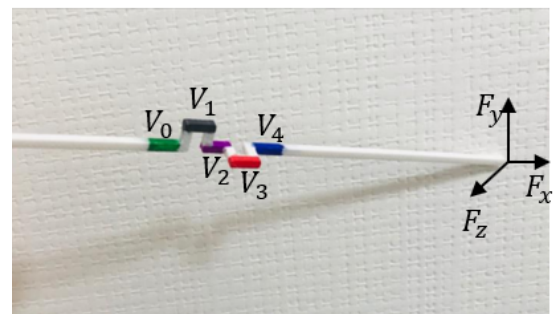


FIGURE 2. 3D model of the deformable surgical tool. It deforms according to the applied forces on the tip of the surgical tool.

A. THE DEFORMABLE SURGICAL TOOL

The proposed method in this paper enables the system to estimate the tool-to-tissue forces based on the deformation of the surgical tool. The first part of our approach is to design the deformable surgical tool. Figure 2 shows a 3D model of the surgical tool that is deformable according to applied forces.

The five parts ($V_0 - V_4$) in the middle of the surgical tool are marked with different colors, and each adjacent part is designed to be perpendicular to each other part to make it clearer to observe its deformation. During the operation, interaction forces (F_x , F_y and F_z) are generated between the tip of the surgical tool and the delicate tissue; as a result, the position of the five parts will change according to the deformation. Then, the relationship between the applied forces on the tip of the surgical tool and the deformation of the surgical instrument is established through the time series CNN model, which is detailed later in Section B. This method makes it possible to estimate the tool-to-tissue forces from the deformation of the surgical tool while in actual use.

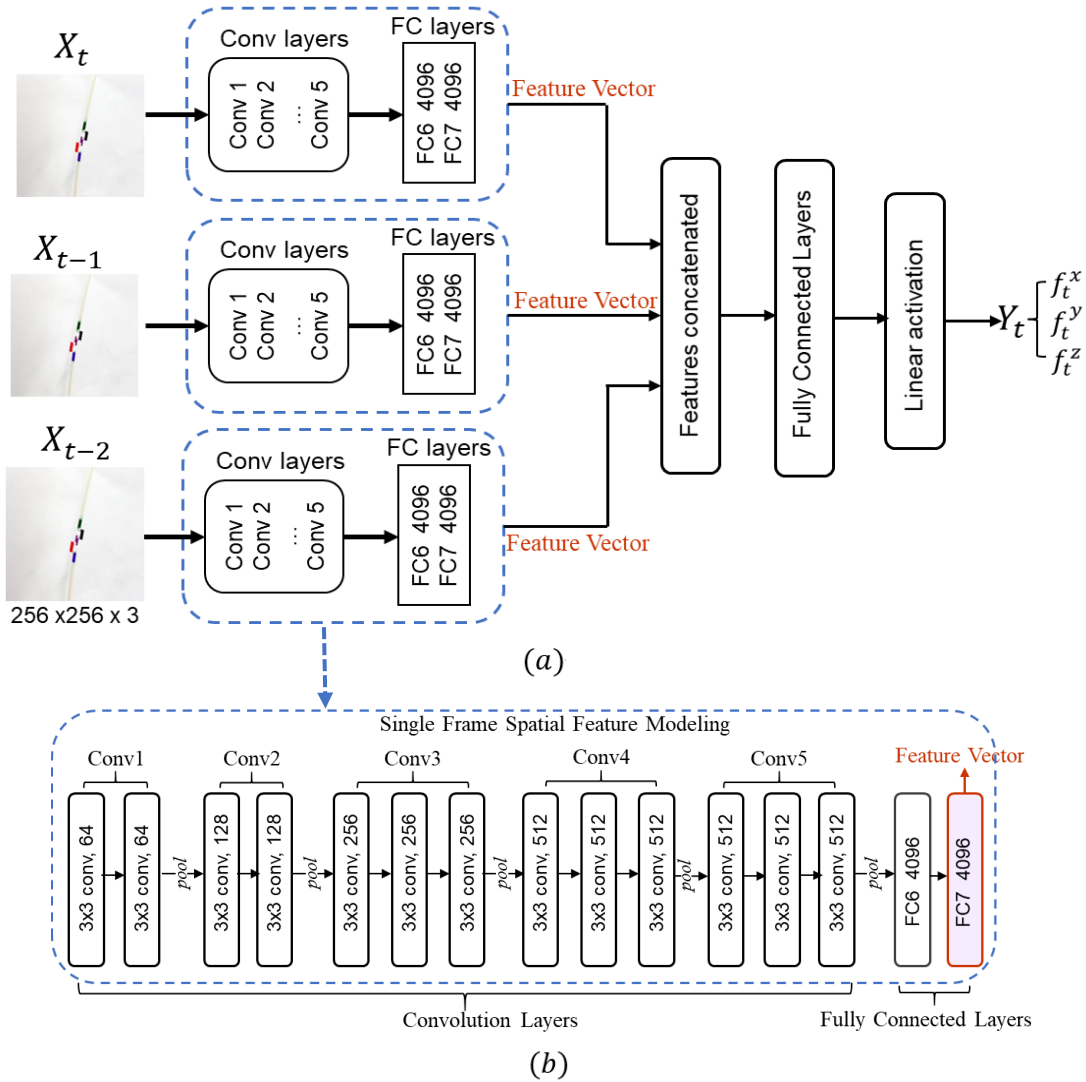


FIGURE 3. (a) A novel time series CNN model consists of three VGG16 architectures. First, a sequence of the current and previous 2 timestep frames (X_t, X_{t-1} and X_{t-2}) is used as the input data of the neural network model. Then, each timestep of the sequence is trained by the fine-tuned VGG16 network model, and the feature vector is extracted from the penultimate layer. Afterward, the extracted feature vectors of 3 timestep frames are concatenated as a new feature vector. In addition, finally, the new feature vector is followed by a fully connected layer with linear activation to estimate the force vector (Y_t) at the current timestep. (b) The fine-tuned VGG16 network model. After the fine-tuning process, the feature vectors can be extracted from the penultimate fully connected layer FC7.

B. THE TIMESERIES CNN MODEL

The approach presented in this paper is used to predict interaction forces using the image information of the deformable tool. Because of the long duration of the surgical image information, the neural network model for force estimation should be established as a function that maps an input image sequence to an output force sequence. Moreover, due to the dynamic characteristics of the sequence, the speed and acceleration of the surgical tool deformation must be accounted for. Therefore, with the aim of performing the force estimation task, a time series CNN model is proposed in this paper, i.e., a sequence of 3 time series frames is used as input to the model, rather than utilizing only a single frame. It can be

represented by the following equation:

$$Y_t = f((X_t, X_{t-1}, X_{t-2}), W) \tag{1}$$

where X_t, X_{t-1} and $X_{t-2} \in \mathbf{R}^{h \times d \times c}$ (h, d represent the height and width of the pixel images, and c is the number of RGB channels) are the sequence of the current and previous 2 timestep frames. The nonlinear model $f(\cdot)$ with weight vector parameters W maps the sequence into estimated force vector $Y_t = [f_t^x \ f_t^y \ f_t^z]$ in the current timestep. The proposed time series CNN model is depicted in Fig. 3(a). First, a sequence of the current and previous 2 timestep frames (X_t, X_{t-1} and X_{t-2}) is used as the input data of the time series CNN model. Subsequently, each frame of the sequence is

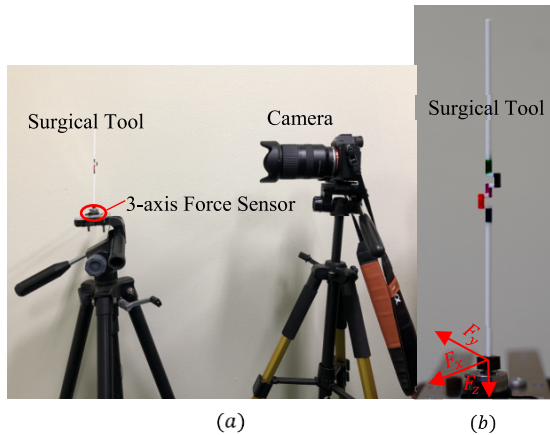


FIGURE 4. (a) Experimental setup, used to acquire the dataset. (b) The 3-axis force sensor is used to obtain the actual force values.

trained by a fine-tuned VGG16 network model. Afterward, features of size 4096 are extracted from the penultimate layer (FC7 shown in Fig. 3(b)) of the VGG16 architecture. Finally, the extracted feature vector of 3 timestep frames is concatenated as new feature vectors, which are followed by a fully connected layer with linear activation to estimate the force vector (Y_t) at the current timestep.

The VGG16 network is fine-tuned on our task to extract feature vectors from image information that can be used in the time series model (3 timesteps). To match the force estimation task, after fine-tuning, the feature vectors can be extracted from the penultimate fully connected layer FC7. After the features are concatenated, a nonlinear activation function *Tanh* is used in the FC layers. By applying tanh, the output results are compressed between -1 and $+1$. A linear activation is used in the output layer for networks that estimate interaction forces.

The loss function plays an important role in machine learning to evaluate how well a learning system can predict the expected results. In this study, the mean square error (MSE) loss function is utilized for the regression task that predicts force vectors. The MSE loss function is given as:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

where y_i represents the actual force vector, \hat{y}_i stands for the predicted force vector, and the index i ($1, 2, \dots, n$) represents the samples in the dataset.

III. EXPERIMENTAL EVALUATION

In this section, experiments to verify the accuracy of the proposed approach are described in detail. In the proposed solution, the established time series CNN model must first be trained using the image information and real forces. Once the model is trained, it can be used to estimate the applied interaction forces. Then, the estimated forces are validated against the real forces to verify the feasibility of the proposed approach.

TABLE 1. Hyperparameters used for the timeseries CNN model.

Hyperparameter	Value
Batch Size	16
Epochs	30
Learning Rate	1.0×10^{-4}
Learning Rate Decay	1.0×10^{-6}
Dropout (FC Layers)	30%

A. DATASET ACQUISITION

The experimental setup is composed of a camera, a 3D printed model of the surgical tool and a 3-axis force sensor (NITTA PD-3-32-05), as shown in Fig. 4. This setup was designed to acquire a dataset of image information of the proposed deformable surgical tool and the actual applied interaction forces. The dataset samples are used as the training and test sets of the model. The video sequence of the deformation of the surgical tool is obtained with a resolution of 1920×1080 pixels at 120 frames per second in RGB color space. At the same time, the applied force on the tip of the surgical tool was measured with the 3-axis force sensor to prepare the training dataset and validation dataset. The surgical tool deforms according to the magnitude of the applied forces, as shown in Fig. 5.

The raw video frames were preprocessed to effectively train and validate the proposed deep learning network. In pre-processing operations, the video is converted into an image frame by frame, and then, a resize method is used to extract a region of interest with a size of 256×256 pixels from 1920×1080 pixels. Exactly from this environment, a dataset with 5400 images and the corresponding force information labels was established, and this dataset was used to train and evaluate our proposed time series CNN model network.

B. TRAINING APPROACH

To fit and evaluate the learning model, the stages of machine learning are divided into two phases: a training phase and a test phase. Here, 67% of the dataset was used for training, and 33% was used for testing, which is a common practice in data science. In the training phase, the 3 timestep CNN learning model is trained using the training dataset to obtain an accurate mapping function of the given information and applied forces. In the test phase, the trained model uses the image information in the test dataset to estimate the applied force and then compares the estimated force with the actual force in the test dataset to validate the accuracy of the network model.

Hyperparameters are the variables that determine how the network is trained, and a proper choice of hyperparameters can enable neural networks to learn faster and achieve higher performance. The hyperparameters used for the training of the proposed time series CNN model are listed in Table 1, such as the batch size, epochs and learning rate. The number of epochs in this paper is defined as 30 to maintain the generalization capacity of the neural network. During the training stage, the error backpropagation method is used to

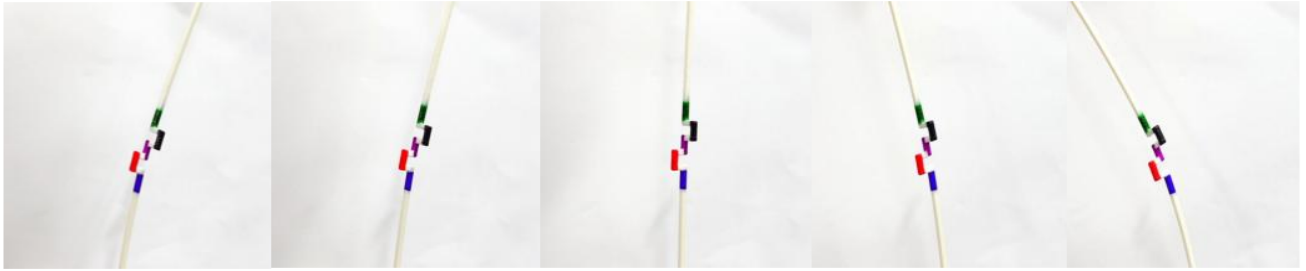


FIGURE 5. Deformation of surgical that result from applying a force at different time instants.

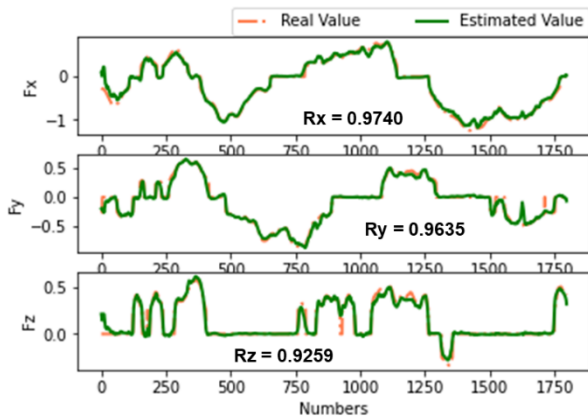


FIGURE 6. The figures show the real force measures in the X, Y, and Z directions, and those estimated by our approach.

update the weights of the model. In addition, the root mean squared error propagation (RMSProp) optimizer [42] is used as the learning rate optimization method. After the training process, the test dataset is used to assess the performance of the proposed model by evaluating the difference between the estimated forces and the real forces.

IV. RESULTS AND DISCUSSION

In this section, the experimental results and discussion are presented. To demonstrate the benefits of the proposed approach, the learning results of the proposed time series CNN network were compared with those of the other two comparison networks (described later in subsection B).

A. RESULTS OF THE PROPOSED METHOD

The results of the experiment are shown in Fig. 6. The accuracy of the proposed approach is verified by the plots shown in the figure in which the estimated force against the real force is compared in the X, Y, and Z directions. The figure shows that the estimated values in the three directions (X, Y and Z) are very close to the real values.

The R^2 score accuracy function is used to evaluate the consistency between the estimated value and the actual value in the proposed model. The accuracy function is described as

follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

$$SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5)$$

where y_i represents the actual force vector, \hat{y}_i stands for the predicted force vector, \bar{y} is the mean value of the actual force vector, SS_{res} is the sum of squares of the residual errors, and SS_{tot} is the total sum of the errors. The value of R^2 varies between 0 and 1. A higher R^2 value indicates that the effect of the model is better.

The R^2 value results of three directions (X, Y and Z) are also shown in Fig. 6. R_x^2 is 0.9740, R_y^2 is 0.9635 and R_z^2 is 0.9259. During the test, the proposed networks estimated the interaction force accurately at every 30ms. The calculation time of estimated forces can be shortened by using a faster processor. The maximum absolute error between the estimated forces and the real forces are 0.1435N, 0.1531N and 0.2284N in the X, Y, and Z directions, respectively. The experimental results prove that the proposed method can successfully estimate the applied force with high accuracy in three dimensions. The root mean square errors (RMSE) of the proposed force estimation method were less than 0.066N in all directions whereas that of another indirect force detection method was 0.07N [23]. Consequently, the precision of the proposed force estimation is acceptable level.

B. COMPARISON RESULTS

To verify the effectiveness of the proposed approach, the results of the proposed approach were compared with those of the other two models, as shown in Fig. 7. Comparison Model 1 is based on the CNN architecture and uses only a single image as input to predict the force vectors, while comparison Model 2 is composed of CNN and LSTM networks and estimates the force vectors through the sequence image information. The comparison models were trained using the same dataset. Figure 8 shows the comparison graphs between the actual forces and the estimated forces with each method, and Table 2 shows the results of the comparison in terms of R^2 values.

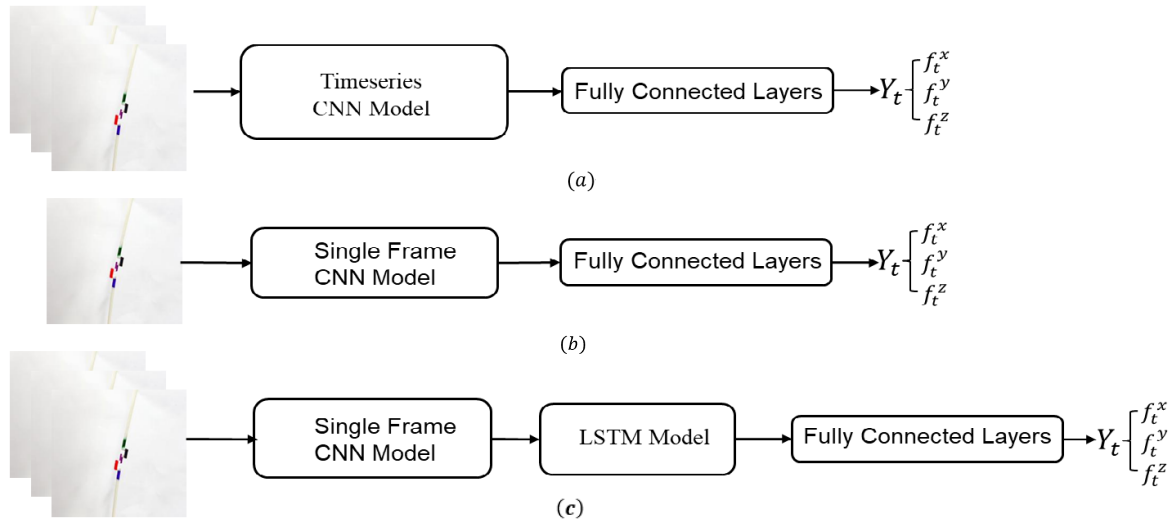


FIGURE 7. Comparison architecture: (a) our proposed model, (b) comparison Model 1 with single frame CNN model, (c) comparison Model 2 with a combined of CNN and LSTM network.

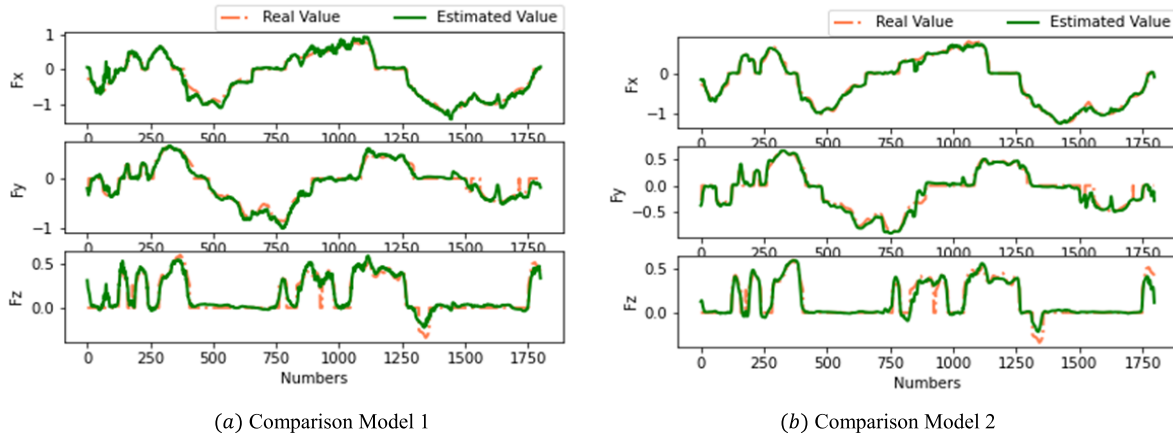


FIGURE 8. Comparison graphs between the actual forces and the estimated forces of comparison Model 1 and comparison Model 2.

TABLE 2. The comparison results of the three models.

Model	R_x^2	R_y^2	R_z^2
Timeseries CNN Model	0.9740	0.9635	0.9259
Single Frame CNN Model	0.9510	0.9143	0.8659
CNN+LSTM Model	0.9709	0.9601	0.9189

Comparing the three architectures, the proposed time series CNN model showed the best result. Although the R^2 value of comparison Model 2 is close to that of the proposed method, the model structure of the proposed method is simpler. Furthermore, using the same training data, the same learning completion condition (as detailed in Section III) and the same computer for training, the training time of the proposed method is approximately 30% less than that of comparison Model 2. This outcome occurs because the tool dynamics are effectively trained with the time series input in the proposed

CNN, while LSTMs are trained by looping through the input time sequences. These experimental results show the effectiveness of the proposed force estimation method.

The movement of the surgical tool during the operation is dynamic and continuous, and thus, a major challenge in force estimation is to understand the dynamic information of the surgical tool. The time series CNN network proposed in this work considers the dynamic characteristics of real surgical scenarios, using a sequence of 3 timestep images as data input instead of using a single frame. Among them, a conventional CNN processes spatial information in a single image, while time series CNNs can handle dynamic information such as velocity and acceleration using 3 time series images. By applying this proposed model, a dynamic mapping of a nonlinear system can be effectively constructed, and the output can be calculated through the past and current frames to provide better force estimation performance. Other studies have also proven the effectiveness of neural networks for the learning of uncertain dynamics using dynamic

information. Time series input data were also applied with damping neurons in the adaptive neuron network controller for force control of robotic manipulators in unknown environments to consider the velocity and acceleration terms and effectively compensate for the unknown dynamics of the environment [43]. In addition, a time-delay feature matrix is used to provide inputs for neural networks and support vector machine-based classifiers, which collect the dynamic characteristics of EEG signals for motion intention prediction in [44]. Furthermore, the authors in [45] confirmed that the convolutive architecture with dynamic information as input is more accurate and robust than recurrent networks in longitudinal vehicle dynamics modeling. In summary, the time series CNN model in this study is effective for practical dynamic systems in constructing dynamic mapping of non-linear systems.

V. CONCLUSION

In robotic-assisted microsurgeries, the surgeons perform microsurgery through a master-slave configuration instead of directly contacting the patients. Therefore, in the absence of force feedback, this approach can affect the accuracy of the operation procedure and cause irreversible tissue damage. Therefore, this paper considers a method of estimating the interaction force vector between a surgical instrument and tissue on the slave side. Because the size of the surgical tool in microsurgery is smaller than that in laparoscopic surgery, which makes it difficult to install force sensors on microsurgical tools, a novel vision-based sensorless force estimation approach for robotic-assisted microsurgery with the capability of multi-axis force sensing is proposed.

In this paper, a method to estimate the amount of force (using a time series CNN model) from the deformation of the surgical tool, which is designed to easily deform according to the applied force, was proposed. The approach presented in this work offers a feasible alternative that overcomes the limitations of integrating sensors into surgical tools. The accuracy of the approach is validated by the experimental results. In particular, the evaluation of the other two comparison models further proves the effectiveness of this method. Future work could extend our approach to estimating the force vectors in real surgical scenarios.

REFERENCES

- [1] N. S. Clarke, J. Price, T. Boyd, S. Salizzoni, K. J. Zehr, A. Nieponice, and P. Bajona, "Robotic-assisted microvascular surgery: Skill acquisition in a rat model," *J. Robot. Surg.*, vol. 12, no. 2, pp. 331–336, Jun. 2018.
- [2] X. He, M. Balicki, P. Gehlbach, J. Handa, R. Taylor, and I. Iordachita, "A novel dual force sensing instrument with cooperative robotic assistant for vitreoretinal surgery," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2013, pp. 213–218.
- [3] J. Selber, "Robotic surgery," *J. Reconstructive Microsurg.*, vol. 28, no. 7, pp. 433–434, Sep. 2012, doi: [10.1055/s-0032-1324408](https://doi.org/10.1055/s-0032-1324408).
- [4] T. J. M. Mulken, R. M. Schols, S. S. Qiu, K. Brouwers, L. T. Hoekstra, D. I. Booi, R. Cau, F. Schoenmakers, A. M. J. Scharnga, and R. W. J. Hulst, "Robotic (super) microsurgery: Feasibility of a new master-slave platform in an *in vivo* animal model and future directions," *J. Surgical Oncol.*, vol. 118, no. 5, pp. 826–831, Oct. 2018.
- [5] M. Shahinpoor and S. Gheshmi, "Introduction to surgical robots' general configurations," in *Robotic Surgery—Smart Materials, Robotic Structures, and Artificial Muscles*. New York, NY, USA: Taylor & Francis, 2015.
- [6] L. Mattos, D. Caldwell, G. Peretti, F. Mora, L. Guastini, and R. Cingolani, "Microsurgery robots: Addressing the needs of high-precision surgical interventions," *Swiss Med. Weekly*, vol. 146, no. 4344, pp. 1–14, 2016.
- [7] O. A. J. van D. Meijden and M. P. Schijven, "The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: A current review," *Surgical Endoscopy*, vol. 23, no. 6, pp. 1180–1190, Jun. 2009.
- [8] C. Pacchierotti, L. Meli, F. Chinello, M. Malvezzi, and D. Prattichizzo, "Cutaneous haptic feedback to ensure the stability of robotic teleoperation systems," *Int. J. Robot. Res.*, vol. 34, no. 14, pp. 1773–1787, Dec. 2015.
- [9] P. Puangmalai, K. Althoefer, L. D. Seneviratne, D. Murphy, and P. Dasgupta, "State-of-the-art in force and tactile sensing for minimally invasive surgery," *IEEE Sensors J.*, vol. 8, no. 4, pp. 371–381, Apr. 2008, doi: [10.1109/JSEN.2008.917481](https://doi.org/10.1109/JSEN.2008.917481).
- [10] Y. Fu, K. Li, B. Pan, J. C. Zhan, and S. G. Wang, "A survey of force sensing and force feedback technology for robot-assisted minimally invasive surgical system," *Robot.*, vol. 36, no. 1, pp. 117–128, 2014.
- [11] A. A. Nazari, F. Janabi-Sharifi, and K. Zareinia, "Image-based force estimation in medical applications: A review," *IEEE Sensors J.*, vol. 21, no. 7, pp. 8805–8830, Apr. 2021.
- [12] Y. Hu, H. Jin, L. Zhang, P. Zhang, and J. Zhang, "State recognition of pedicle drilling with force sensing in a robotic spinal surgical system," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 1, pp. 357–365, Feb. 2014.
- [13] J. Rosen, B. Hannaford, C. G. Richards, and M. N. Sinanan, "Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 5, pp. 579–591, May 2001.
- [14] J. B. Gafford, S. B. Kesner, R. J. Wood, and C. J. Walsh, "Force sensing surgical grasper enabled by pop-up book MEMS," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Tokyo, Japan, Nov. 2013, pp. 2552–2558.
- [15] A. Menciassi, A. Eisinger, M. C. Carrozza, and P. Dario, "Force sensing micro-instrument for measuring tissue properties and pulse in microsurgery," *IEEE/ASME Trans. Mechatronics*, vol. 8, no. 1, pp. 10–17, Mar. 2003.
- [16] M. B. Hong and Y.-H. Jo, "Design and evaluation of 2-DOF compliant forceps with force-sensing capability for minimally invasive robot surgery," *IEEE Trans. Robot.*, vol. 28, no. 4, pp. 932–941, Aug. 2012.
- [17] A. M. Okamura, "Haptic feedback in robot-assisted minimally invasive surgery," *Current Opinion Urol.*, vol. 19, no. 1, pp. 102–107, Jan. 2009.
- [18] S. Sokhanvar, J. Dargahi, S. Najarian, and S. Arbatani, "Clinical and regulatory challenges for medical devices tactile sensing and displays," in *Haptic Feedback for Minimally Invasive Surgery and Robotics*. Hoboken, NJ, USA: Wiley, 2012.
- [19] L. Yu, W. Wang, and F. Zhang, "External force sensing based on cable tension changes in minimally invasive surgical micromanipulators," *IEEE Access*, vol. 6, pp. 5362–5373, 2018.
- [20] M. Haghpanah, M. Miyasaka, and B. Hannaford, "Utilizing elasticity of cable-driven surgical robot to estimate cable tension and external force," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1593–1600, Jul. 2017, doi: [10.1109/LRA.2017.2676347](https://doi.org/10.1109/LRA.2017.2676347).
- [21] A. I. Aviles, S. Alsaleh, P. Sobrevilla, and A. Casals, "Sensorless force estimation using a neuro-vision-based approach for robotic-assisted surgery," in *Proc. 7th Int. IEEE/EMBS Conf. Neural Eng.*, Apr. 2015, pp. 86–89, doi: [10.1109/NER.2015.7146566](https://doi.org/10.1109/NER.2015.7146566).
- [22] J. Kim, F. Janabi-Sharifi, and J. Kim, "A haptic interaction method using visual information and physically based modeling," *IEEE/ASME Trans. Mechatronics*, vol. 15, no. 4, pp. 636–645, Aug. 2010, doi: [10.1109/TMECH.2009.2031240](https://doi.org/10.1109/TMECH.2009.2031240).
- [23] E. Noohi, S. Parastegari, and M. Zefran, "Using monocular images to estimate interaction forces during minimally invasive surgery," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 4297–4302, doi: [10.1109/IROS.2014.6943169](https://doi.org/10.1109/IROS.2014.6943169).
- [24] N. Haouchine, W. Kuang, S. Cotin, and M. Yip, "Vision-based force feedback estimation for robot-assisted surgery using instrument-constrained biomechanical three-dimensional maps," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 2160–2165, Jul. 2018, doi: [10.1109/LRA.2018.2810948](https://doi.org/10.1109/LRA.2018.2810948).
- [25] A. Mozaffari, S. Behzadipour, and M. Kohani, "Identifying the tool-tissue force in robotic laparoscopic surgery using neuro-evolutionary fuzzy systems and a synchronous self-learning hyper level supervisor," *Appl. Soft Comput.*, vol. 14, pp. 12–30, Jan. 2014.

- [26] A. I. Aviles, A. Marban, P. Sobrevilla, J. Fernandez, and A. Casals, "A recurrent neural network approach for 3D vision-based force estimation," in *Proc. 4th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Oct. 2014, pp. 1–6, doi: [10.1109/IPTA.2014.7001941](https://doi.org/10.1109/IPTA.2014.7001941).
- [27] A. I. Aviles, S. M. Alsaleh, J. K. Hahn, and A. Casals, "Towards retrieving force feedback in robotic-assisted surgery: A supervised neuro-recurrent-vision approach," *IEEE Trans. Haptics*, vol. 10, no. 3, pp. 431–443, Jul. 2017.
- [28] X. He, J. Handa, P. Gehlbach, R. Taylor, and I. Iordachita, "A sub-millimetric 3-DOF force sensing instrument with integrated fiber Bragg grating for retinal microsurgery," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 522–534, Feb. 2014.
- [29] H. Xie, H. Liu, Y. Noh, J. Li, S. Wang, and K. Althoefer, "A fiber-optics-based body contact sensor for a flexible manipulator," *IEEE Sensors J.*, vol. 15, no. 6, pp. 3543–3550, Jun. 2015.
- [30] B. Gonenc, A. Chamani, J. Handa, P. Gehlbach, R. H. Taylor, and I. Iordachita, "3-DOF force-sensing motorized micro-forceps for robot-assisted vitreoretinal surgery," *IEEE Sensors J.*, vol. 17, no. 11, pp. 3526–3541, Jun. 2017.
- [31] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 677–691, Apr. 2017, doi: [10.1109/TPAMI.2016.2599174](https://doi.org/10.1109/TPAMI.2016.2599174).
- [32] L. Zhang, G. Gui, A. M. Khattak, M. Wang, W. Gao, and J. Jia, "Multi-task cascaded convolutional networks based intelligent fruit detection for designing automated robot," *IEEE Access*, vol. 7, pp. 56028–56038, 2019, doi: [10.1109/ACCESS.2019.2899940](https://doi.org/10.1109/ACCESS.2019.2899940).
- [33] C. Wang, K. Huang, W. Ren, J. Zhang, and S. Maybank, "Large-scale weakly supervised object localization via latent category learning," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1371–1385, Apr. 2015, doi: [10.1109/TIP.2015.2396361](https://doi.org/10.1109/TIP.2015.2396361).
- [34] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015, pp. 1–14.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2015, pp. 1–9.
- [38] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sens.*, vol. 9, no. 9, p. 907, 2017.
- [39] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 1–14.
- [40] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: [10.1109/TMI.2016.2528162](https://doi.org/10.1109/TMI.2016.2528162).
- [41] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [42] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
- [43] K. Kiguchi and T. Fukuda, "Neural network controllers for robot manipulators application of damping neurons," *Adv. Robot.*, vol. 12, no. 3, pp. 191–208, Jan. 1997.
- [44] D. Bandara, J. Arata, and K. Kiguchi, "A noninvasive brain-computer interface approach for predicting motion intention of activities of daily living tasks for an upper-limb wearable robot," *Int. J. Adv. Robot. Syst.*, vol. 15, no. 2, Mar. 2018, Art. no. 172988141876731.
- [45] M. D. Lio, D. Bortoluzzi, and G. P. R. Papini, "Modelling longitudinal vehicle dynamics with neural networks," *Veh. Syst. Dyn.*, vol. 58, no. 11, pp. 1675–1693, 2020.



JIUYUN XIA (Student Member, IEEE) received the B.E. degree from the Hebei University of Science and Technology, Hebei, China, in 2013, and the M.E. degree in instrument and meter engineering from Beihang University, Beijing, China, in 2016. She is currently pursuing the D.E. degree from Kyushu University, Fukuoka, Japan.

Her research interests include robotics and machine learning.



KAZUO KIGUCHI (Senior Member, IEEE) received the B.E. degree from Niigata University, Niigata, Japan, in 1986, the M.A.Sc. degree from the University of Ottawa, Ottawa, Canada, in 1993, and the D.E. degree from Nagoya University, Nagoya, Japan, in 1997.

He worked with Mazda Motor Corporation, from 1986 to 1989, and MHI Aerospace Systems Corporation, from 1989 to 1991. From 1994 and 1999, he was with the Niigata College of Technology, Niigata. He was with the Graduate School of Science and Engineering, Saga University, Japan, from 1999 and 2012. He is currently a Professor with the Department of Mechanical Engineering, Faculty of Engineering, Kyushu University, Japan. His research interests include biorobotics, human-assist robots, rehabilitation robots, and application of robotics in medicine.

Dr. Kiguchi is a member of the Robotics Society of Japan, the Japan Society of Computer-Aided Surgery, the Society of Biomechanisms Japan, and the Japanese Society for Medical and Biological Engineering. He is a Senior Member of the IEEE Robotics and Automation Society, the IEEE Systems, Man, and Cybernetics Society, the IEEE Computational Intelligence Society, and the IEEE Engineering in Medicine and Biology Society. He is a fellow of the Japan Society of Mechanical Engineers (JSME) and the Society of Instrument and Control Engineers (SICE). He received the JSME Funai Award, the JSME Medal for Outstanding Paper, the JSME Medal for Distinguished Engineers, and the Lifetime Achievement Award at WAC2014.

• • •