2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Macau, China, November 4-8, 2019

Hysteresis Compensator with Learning-based Pose Estimation for a Flexible Endoscopic Surgery Robot

Donghoon Baek, Ju-Hwan Seo, Joonhwan Kim, and Dong-Soo Kwon, *Member, IEEE*

***Abstract*—The use of the tendon-sheath mechanism (TSM) is common in flexible surgery robots, because it can flexibly work in limited spaces and provides efficient power transmission. However, hysteresis from nonlinearities such as friction and backlash poses a challenge in controlling precise motion in the surgical instrument. Moreover, this hysteresis is also affected by changes in the various configurations of sheath which limits traditional model-based compensation approaches. Recently, feedback approach using an endoscopic camera is presented, but they use markers which are not appropriate for applying to a real surgical instruments. In this paper, a novel hysteresis compensator with learning-based pose estimation is proposed. Unlike previous studies, the proposed compensator can reduce hysteresis of the surgical instrument in various sheath configurations without using markers. In order to estimate an actual angle of the surgical instrument's joint, we employ the learning-based pose estimation using a siamese convolutional neural network (SCNN). The proposed compensator reduces hysteresis by partially controlling the position command, similar to the instinctive adjustments that physicians make with their visual feedback. To validate the proposed method, a testbed was constructed considering several requirements of flexible surgery robots. As a result, the results show the proposed method reduces hysteresis to less than 10°, for various configurations of sheath. In addition, we confirmed that the learning-based pose estimation is sufficient to apply to the proposed compensator for reducing hysteresis in real-time.**

I. Introduction

Conventional flexible endoscopic systems have been used frequently in non-invasive surgical interventions such as intraluminal surgery and natural orifice transluminal endoscopic surgery (NOTES). With these flexible endoscopes, physicians can perform surgeries without making external incisions in a patient's body. However, it has several limitations due to its low degree of freedom of surgical tip and non-intuitive manipulation. In order to overcome these limitations, a number of flexible surgery robots with higher degrees of freedom and intuitive operation using a master device emerged in subsequent research [1, 2].

Tendon-sheath mechanism (TSM) was widely adopted in these flexible surgery robots due to its flexibility, lightweight, and ability to provide efficient power transmission to a surgical instrument from an external actuator. However, hysteresis caused by friction, backlash, and the time-varying configuration of the sheath during its operations brings many challenges in controlling the precise motion of surgical instruments [3]. Although physicians can compensate for hysteresis through visual feedback, this increases the fatigue in performing the task.

A lot of studies of hysteresis reduction have been devoted. Mathematical hysteresis models based on the system dynamics have been suggested [4, 5] and applied to model-based feed-forward control approach [6, 7]. However, this hysteresis model will be changed if the configuration of sheath changes. So, it was only valid for one particular configuration [7]. Because it is difficult to obtain a complete model applicable to all situations, the feedback approaches using electromagnetic sensors have been suggested [8, 9]. However, because of the size of sensors and the requirements for sterilization of instrument, the traditional sensors are difficult to integrate into surgery robots. To overcome some of these challenges, pose estimation algorithms for obtaining an actual pose of the surgical instruments are also being actively studied [10 – 14]. Although some of them estimate the pose of the surgical instrument in the clinical image, there is insufficient research showing that the application of this pose estimation method reduces hysteresis of the robotic system. In practical applications, image-based hysteresis reduction method as feedback approaches using an endoscopic camera and markers has been proposed [15]. However, the markers are easily obscured by surgical debris, and it is likely to reduce the performance of hysteresis compensation.

In this paper, a hysteresis compensator with learning-based pose estimation was proposed. Unlike previous image-based hysteresis compensator, this novel approach can reduce hysteresis without using any markers by using the learning-based method. The feedback approach of the proposed compensator facilitates hysteresis compensation in various sheath configurations as well. The proposed pose estimation calculates the actual joint angle of surgical instrument using a siamese convolutional neural network (SCNN) [16] in real-time. The concept of this method is the image-retrieval [17] which does not require precise geometric and kinematic models. Moreover, the speed of the estimation process is fast and suitable for application to real-time systems.

*Research supported by a grant of the Korea Evaluation Institute of Industrial Technology (KEIT), funded by the Ministry of Trade, Industry and Energy, Republic of Korea(Grant#:10060158), and International Joint Technology Development Project funded by the Korean Ministry of Trade, Industry and Energy (Grant #: P0006718)

Donghoon Baek is with the robotics program, Korea Advanced Institute of Science and Technology, Daejeon, South Korea (e-mail: romansabaek@kaist.ac.kr).

Ju-Hwan Seo and Joonhwan Kim are with the Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea e-mail: (seojh1989@gmail.com, joonhwan@kaist.ac.kr).

Dong-Soo Kwon (corresponding author) is with the Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea (Tel.: +82 42 350 3042; e-mail: kwonds@kaist.ac.kr). He is a CEO of the EasyEndo Surgical Inc.

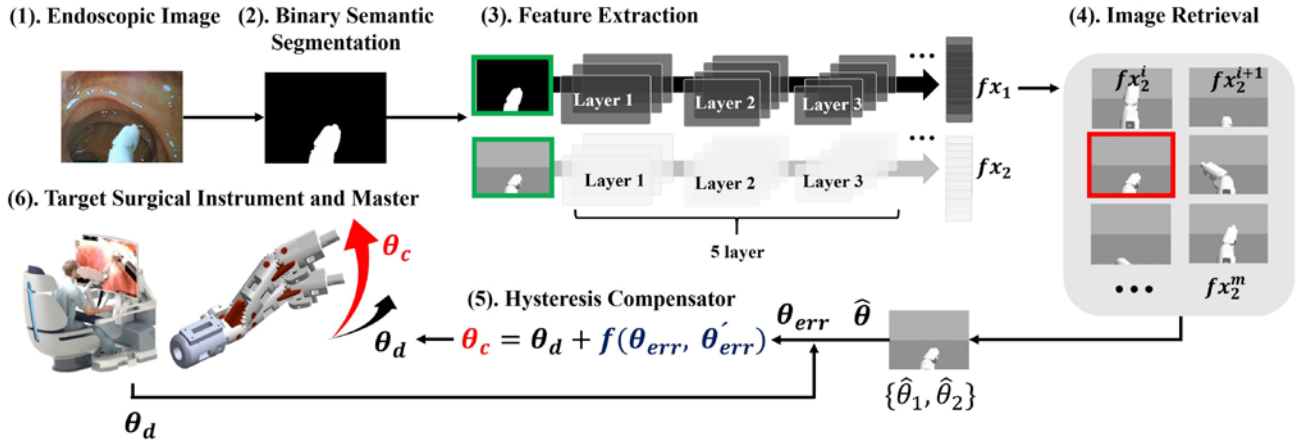

Fig. 1. Steps of the proposed hysteresis compensation process. (1) Obtaining an image from an endoscopic camera. (2) Binary semantic segmentation for background subtraction. (3) Feature extraction using a SCNN. (4) Image retrieval for searching the most similar image. (5) Hysteresis compensator. (6) Target surgical instrument.

II. Materials and Method

Our objective is to propose a hysteresis compensator that reduces the hysteresis in the various configurations of sheath, by relying on an embedded endoscopic camera without using markers. This method includes an image segmentation to distinguish between the surgical instrument and a background in a clinical image. Subsequently, a SCNN and k-d tree are used for feature extraction and efficient searching respectively. When the searching is finished, the surgical instrument joint angle is estimated. In the last step, this estimated angle is applied to the hysteresis compensator. The overall process of the proposed method is presented in Fig. 1.

A. Testbed of flexible surgical robotic system

a) Design requirements for testbed:

In order to construct the testbed similar to the surgical instrument used in real flexible surgery robots, the testbed was designed according to several requirements for flexible surgical robot as shown in Fig. 2(b). Considerations are as follows: (1) The coupled tendon-sheath mechanism applied to most surgical robots for a bending motion is utilized as shown in Fig. 2(a). (2) An endoscopic camera, a micro USB-cam used in existing surgical robotic systems, is utilized. (3) The surgical instrument has two degree of freedom bending motions (pitch θ_1, yaw θ_2) that are most affected by hysteresis. (4) The insertion tube with a sheath and a wire which can represent the various configurations of sheath is used.

b) Kinematics:

The proposed method is to add a compensator to a general kinematic-based control approach. An inverse kinematics model is needed for a kinematic-based control in the coupled TSM. The wire length and a total variation of the wires across a joint with angle θ can be expressed as follows:

$$L_a(\theta) = 2R - 2dsin\left(\frac{\theta}{2}\right) - 2(R - B)\cos\left(\frac{\theta}{2}\right) \quad (1)$$

$$L_b(\theta) = L_a(-\theta) \quad (2)$$

where $L_a(\theta)$ and $L_b(\theta)$ represent the vertical length from the wire hole to the hole as shown in Fig. 2(a), and each corresponds to the pulling and releasing side. The variation of each wire, pulling and releasing respectively, is described as:

$$\Delta L_a(\theta) = L_a(\theta) - 2B \quad (3)$$

$$\Delta L_b(\theta) = L_b(\theta) - 2B \quad (4)$$

where B is defined as the wire length when a single joint angle is 0°. The rotation angles q_a and q_b for each motor are calculated, taking into account the initial position q_{a0} and q_{b0} of each motors:

$$q_a = q_{a0} - \Delta L_a(\theta) * \mathrm{K} \quad (5)$$

$$q_b = q_{b0} - \Delta L_b(\theta) * \mathrm{K} \quad (6)$$

where K indicates the ratio between the change in wire length ΔL and the motor rotation Δq. Finally, two motors pull and release each wire.

B. hysteresis compensator

The proposed hysteresis compensator, which controls the position command as through adding a supplemental movement similar to the concept of visual feedback, is designed as follows:

$$\theta_c = \theta_d + f(\theta_{err}, \dot{\theta_{err}}) \quad (7)$$

$$f(\theta_{err}, \dot{\theta}_{err}) = K_p\theta_{err} + K_d\dot{\theta}_{err} + K_i\int\theta_{err}dt \quad (8)$$

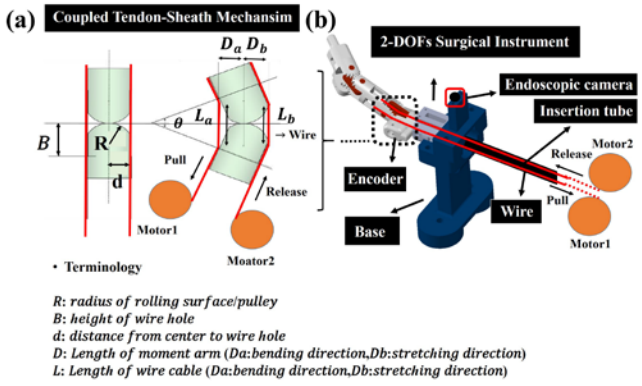

Fig. 2. (a) Coupled tendon-sheath mechanism. (b) Design configuration.

where θ_c is the compensated position input, θ_d is the desired angle from the master, $\theta_{err} = \theta_d - \hat{\theta}$, $\hat{\theta}$ is the estimated angle from the learning-based pose estimation, $\acute{\theta_{err}}$ is the differential of θ_{err}, and K_p, K_d, K_i are the hyperparameters for the proposed compensator.

This approach is based on a visual-servo system [18] that uses a vision sensor to provide feedback information. The compensated position input θ_c is updated with a sampling rate of 30Hz, which is determined by the camera sampling time. In this case, the camera sampling time $T_c(33\text{ms})$ is much longer than the control sampling time $T_i(1\text{ms})$and it has a negative effect on the control performance. The multi-rate control approach [19], which is widely adopted to alleviate the problem caused by the delay time between T_c and T_i [20], is used in the proposed compensator. This can be expressed as:

$$\theta_c[j] = \begin{cases} \theta_d[j] + f\left(\theta_{err}[j,i], \acute{\theta_{err}}[j,i]\right) & if \;\; i = 33j, \;\; j \in N \\ \theta_d[j] + f\left(\theta_{err}[j,i-1], \acute{\theta_{err}}[j,i-1]\right) & otherwise \end{cases} \quad (9)$$

where $\theta_{err}[j,i] = \theta_d[j] - \hat{\theta}[i]$, $\theta_{err}[j,i-1] = \theta_d[j] - \hat{\theta}[i-1]$, i is the vision sampling time, and j is the control sampling time. The value of $f\left(\theta_{err}, \acute{\theta}_{err}\right)$ is limited for system safety.

C. Learning-based pose estimation

The basic idea is an image-retrieval approach, searching for the image candidate that is the most similar to the image obtained from an endoscopic camera in large database. When obtaining the most similar image, the estimated angle $\hat{\theta}_1$ and $\hat{\theta}_2$ are determined by the label of the image. This method consists of three stages: Image segmentation, feature extraction, and searching.

a) Image Segmentation for background subtraction:

In order to eliminate the effect of the background for the purpose of reducing the variance of the input in a clinical image, image segmentation is adopted as shown in Fig. 3. Considering the real-time performance, image segmentation using distance transform and watershed algorithm [21] is applied to a small resized endoscopic image. A Laplacian filter which is a robust to the effects of noise is applied to obtain the edge of the surgical instrument. Then a threshold is used in order to filter out main edges not corresponding to the outline of the surgical instrument. A distance transform and normalizing are then applied to the binary image, and a skeleton image is obtained for detecting the foreground image. Finally, a binary segmented image is taken by watershed algorithm using two images Fig. 3(b) and Fig. 3(e).

b) Database construction:

In practice, it is challenging to get the trustful ground-truth image describing the actual joint angle of the surgical instrument due to the influence of friction, noise, illumination,

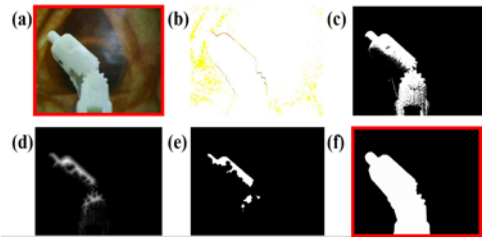

Fig. 3. The process of the image segmentation. (a) Small resized endoscopic image. (b) Edge of the surgical instrument. (c) Binary image of (b). (d) Apply the distance transform and normalizing. (e) Binary image of (d). (f) Binary segmented image.

and etc. However, in the simulator, it is straightforward to obtain the trustful data due to the absence of these influences.

In order to get the ground-truth data, we developed a simulator consisting of the 3D-CAD model of the surgical instrument and a virtual camera with the same interior calibration parameters of the endoscopic camera. The database consists of two parts: the binary segmented image dataset $\mathbb{X}^b$ obtained by applying image segmentation to the images obtained from the endoscopic camera and the rendered-image dataset $\mathbb{X}^r$ obtained from the virtual camera in the simulator. Each sample of data is labeled with the actual joint angles $\theta_1$and θ_2 are defined as follows:

$$X_i^b = \{x_i^b, \theta_{i1}^b, \theta_{i2}^b\}, X_i^b \in \mathbb{X}^b, i \in [1,2,\dots,N] \quad (10)$$

$$X_i^r = \{x_i^r, \theta_{i1}^r, \theta_{i2}^r\}, X_i^r \in \mathbb{X}^r, i \in [1,2,\dots,N] \quad (11)$$

where x_i^b and x_i^r are a binary segmented image and a rendered-image respectively. θ_{i1} and θ_{i2} refer to the actual joint angles at the surgical instrument. In this case, the range of each θ is determined by considering commonly used ranges in surgeries [22] and the field of view (FOV) of the endoscopic camera. The range of each θ is as follows: $-5^\circ < \theta_{i1} < 50^\circ, -50^\circ < \theta_{i2} < 50^\circ$. The entire dataset consists of three column vectors $P_j = \{x_j^b, x_j^r, S_j\}$, where label S_j that represents true or false is defined as below.

$$S_j = \begin{cases} 1 & if \;\; \theta_{j1}^b = \theta_{j1}^r, \; \theta_{j2}^b = \theta_{j2}^r \\ 0 & otherwise \end{cases} \quad (12)$$

The pose of the endoscopic camera is fixed as $(x_0, y_0, z_0, \alpha, \beta, \gamma)$ since it has no independent degrees of freedom.

The database construction process is shown in Fig. 4. In this figure, (a) - (c) show the process of collecting $\mathbb{X}^r$ and (d) - (f) show the process of collecting $\mathbb{X}^b$.

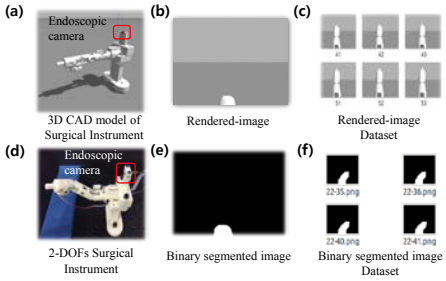

Fig. 4. The process of the database construction. (a) – (c) rendered-image dataset construction using the simulator. (d) – (f) binary segmented image dataset construction using the testbed.

c) Feature extraction using SCNN:

The SCNN, which is effective in extracting the feature vectors f_{x1} and f_{x2} that determine a "similarity" which means whether the paired images x_i^b and x_i^r are in the same category or not, is used. This network consists of two branches sharing the same architecture and weight values [16]. In the first step, the paired images x_i^b and x_i^r are separated and fed into the each branch. Because of these reasons, two input images are mapped into the same feature space and it is possible that one of the image can be converted into a vector form in advance. This means that computational efficiency can be improved in the process of comparing images. A contrastive loss function is used for the purposed of minimizing the Euclidean distance of images in the same category and maximizing the Euclidean distance of images in the different category. The contrastive loss function is defined as follows:

$$L(x_1, x_2, y) = \frac{1}{2N}\sum_{n=1}^{N} [y \times d(x_1 \circ x_2)^2 + (1-y)max(1 - d(x_1 \circ x_2), 0)^2] \quad (13)$$

where the Euclidean distance is given by

$$d(x_1 \circ x_2)^2 = \sum_J \left| f_{x_1}^j - f_{x_2}^j \right|^2, \ j \in \mathbb{N} \quad (14)$$

To prevent overfitting problems, regularization is applied as shown in Eq. (15).

$$L(x_1, x_2, y)_{reg} = L(x_1, x_2, y) + \alpha \sum_i W_i^2 \quad (15)$$

where α is the regularization parameter and W_i is the weight values of network. x_1 and x_2 represent the sample of paired images and two feature vectors f_{x1} and f_{x2} are used to compute the Euclidean distance. y represents the same thing as S_j from Eq. (12), and j is the dimension of the feature vector extracted by each branch.

The network was trained by the Adam-optimizer [23] with a batch size of *200*, and the size of the input image is *100*×*100*. Learning parameters are $1e^{-5}$ for initial learning rate, *0.96* for learning rate decay per epoch. The weights of the networks were initialized by using the Xavier initialization algorithm [24]. The number of layers is five and each has convolution layer, rectified linear unit, and pooling layer as illustrated in Fig. 5.

d) Data distributed learning strategy:

In general, for an efficient learning, the true and false data should be balanced in the database [25]. According to the principle, the database size becomes very large, because x_i^b has 5655 $(56 \times 101 - 1)$ different images in our method. Therefore, it results in an increase of the variance of input. To solve this problem, θ_1 and θ_2 are separately estimated by using four networks $\{N_{\theta_1}, N_{\theta_2}^1, N_{\theta_2}^2, N_{\theta_2}^3\}$. When the binary segmented image is taken, θ_1 is estimated by using N_{θ_1}. Subsequently, θ_2 is estimated by using one of the other networks $\{N_{\theta_2}^1, N_{\theta_2}^2, N_{\theta_2}^3\}$ depending on the result of $\hat{\theta}_1$. We divided the datasets into four cases $\{D_{\theta_1}, D_{\theta_2}^1, D_{\theta_2}^2, D_{\theta_2}^3\}$ and each is handled by $\{N_{\theta_1}, N_{\theta_2}^1, N_{\theta_2}^2, N_{\theta_2}^3\}$ separately. Each dataset is organized as shown in table I.

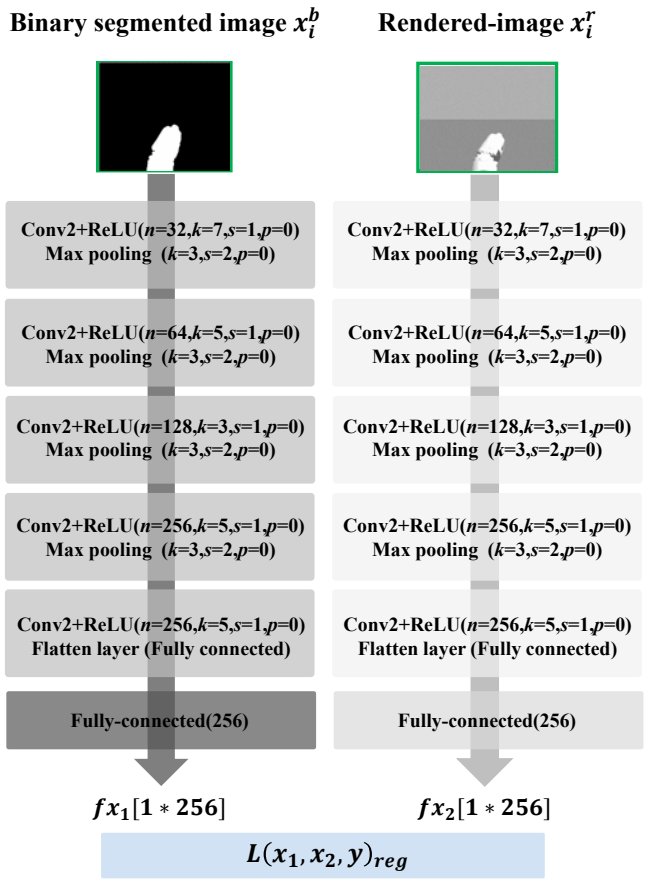

Fig. 5. The structure of SCNN. *n*: the number of filters, *k*: the size of filters, *s*: stride, *p*: padding.

TABLE I. CONFIGURATION OF EACH DATASET

	X^b		X^r	
	Range of θ_1^b	*Range of* θ_2^b	*Range of* θ_1^r	*Range of* θ_2^r
D_{θ_1}	$-5° \le \theta_1^b \le 50°$	$-50° \le \theta_2^b \le 50°$	$-5° \le \theta_1^r \le 50°$	$\theta_2^r = 0°$
$D_{\theta_2}^1$	$-5° \le \theta_1^b \le 10°$	$-50° \le \theta_2^b \le 50°$	$\theta_1^r = \theta_1^b$	$-50° \le \theta_2^r \le 50°$
$D_{\theta_2}^2$	$11° \le \theta_1^b \le 30°$	$-50° \le \theta_2^b \le 50°$	$\theta_1^r = \theta_1^b$	$-50° \le \theta_2^r \le 50°$
$D_{\theta_2}^3$	$31° \le \theta_1^b \le 50°$	$-50° \le \theta_2^b \le 50°$	$\theta_1^r = \theta_1^b$	$-50° \le \theta_2^r \le 50°$

e) Searching the best candidate using k-d tree:

To determine the best candidate efficiently that is the most similar to the binary segmented image x_i^b obtained from an endoscopic camera in the rendered-image dataset $\mathbb{X}^r$, k-d tree algorithm is used. This is an efficient searching method that creates a binary spatial partition at each level of the tree by recursively partitioning the space region. The rendered images in $\mathbb{X}^r$ labeled with the actual joint angles θ_1 and θ_2 of the surgical instrument are pre-transformed into feature vector f_{x2} using the learned network model. The equation used in searching for the best candidate is given by the following, Eq. (17).

$$C(x_1, \mathbb{X}^r) = argmin_c d(x_1 \circ x_2)^2, \ x_2 \in \mathbb{X}^r, \ x_1 \in \mathbb{X}^b \quad (17)$$

where $d(x_1 \circ x_2)^2$ represents same as equation (14). Finally, the estimated angles $\hat{\theta}_1$ and $\hat{\theta}_2$ are determined as the label of the selected image x_2.

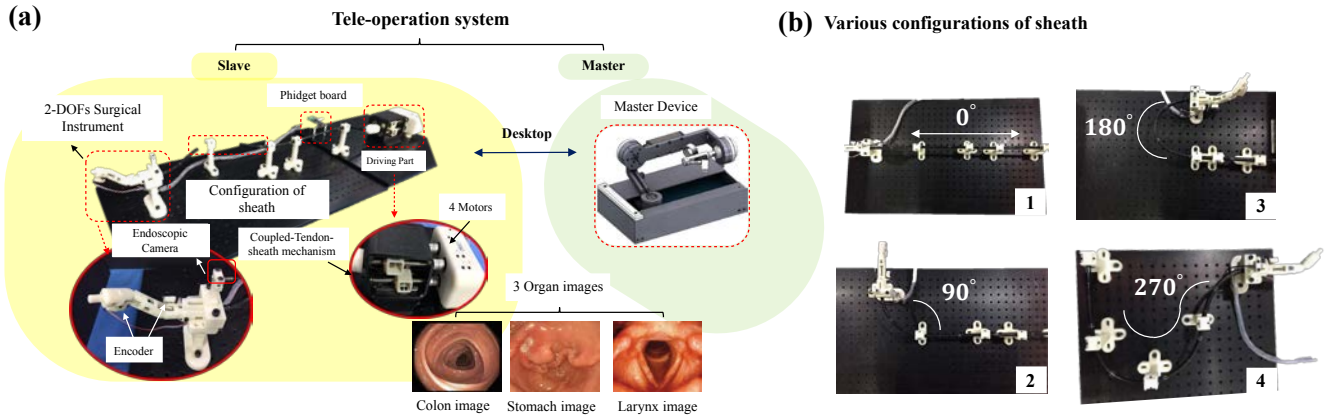

Fig. 6. The experiment setup. (a) Tele-operation system. (b) Various configurations of sheath used in the experiment. 1) 0°. 2) 90°. 3) 180°. 4) 270°.

III. Experiments and Results

The testbed of the flexible endoscopic surgery robot system and master device for tele-operation were used for evaluation as shown in Fig. 6(a). Experiments were carried out in the range of θ_1 and θ_2 as defined in a training dataset. The actual joint angle of the surgical instrument is measured by an encoder attached to each joint for reliable measurements. Three different printed images of organ {Colon, Stomach, Larynx} were used as the background.

A. Testbed setup

A 2-DOFs flexible surgical robot testbed driven by a tendon-sheath mechanism using a SUS spring and stainless steel wire was used to evaluate the performance of the proposed method. The main driving part consisted of four motors (Dynamixel xm430, ROBOTIS, Korea) and was connected via ball screws and linear guides together. The micro USB endoscopic camera (MD-V20904L-76, MISUMI, Korea) with 640×480 resolution was fixed on the base to take images. A mini-rotary encoder (7S-400-2MC-50-00E, NEMICON, Japan) with 400° resolution was attached to each joint to measure the actual angle. All system software was constructed on Linux (ubuntu16.04LTS) and ROS (kinetic). This process was conducted by using a NVIDIA RTX-2080ti GPU in a desktop.

B. Evaluation of Learning-based pose estimation

To evaluate the proposed pose estimation method, the training dataset as detailed in Sec. II and the testbed were used. Periodic sinusoidal signal and non-periodic signal from the master were given. As a measurement of performance, the RMS deviation between the estimated angle and the actual angle from the encoder attached at the joint was used.

The result is presented in Fig. 7. In the test using the training datasets, the RMSE was 0.22° for θ_1 and 0.06° for θ_2 as shown in Fig. 7(a) and (b). However, in the physical environment, using the image from an endoscopic camera, the RMSE was 2.11° for θ_1 and 5.09° for θ_2 when the sinusoidal signal is given as shown in Fig. 7(c) and (d). In the case of the non-periodic signal from a tele-operation system, the RMSE was 2.18° for θ_1 and 6.51° for θ_2. The time taken to estimate the angle is within 10ms, which is less than the camera sampling time $T_c(33\text{ms})$.

The performance result in the real world is relatively inferior to those evaluated using the training data. This is expected to be an overfitting problem. In practice field, the physical influence such as noise and illumination affect the performance of the image segmentation, which creates more variance of the input image that enters the network. If the trained network does not learn all variances of input image, its performance will degrade in different test environments. However, despite some overfitting problems, the performance of the pose estimation was acceptable in reducing hysteresis as shown in following experiment. To further reduce the impact of overfitting problem, dropout and batch normalization techniques will be applied in our method.

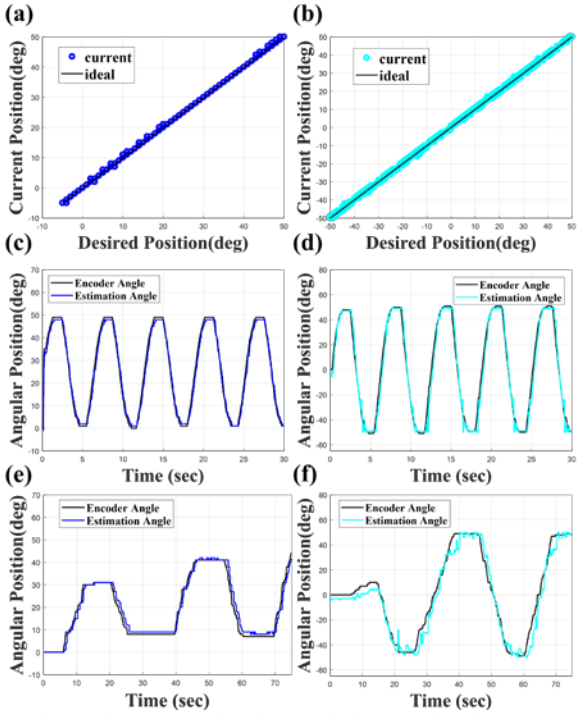

Fig. 7. The result of estimated angle from the learning-based pose estimation. (a) – (b) Estimated angle in the test using training datasets (left: θ_1, right: θ_2). (c) – (d) Estimated angle with respect to a periodic sinusoidal signal (left: θ_1, right: θ_2). (e) – (f) Estimated angle with respect to a non-periodic signal (left: θ_1, right: θ_2).

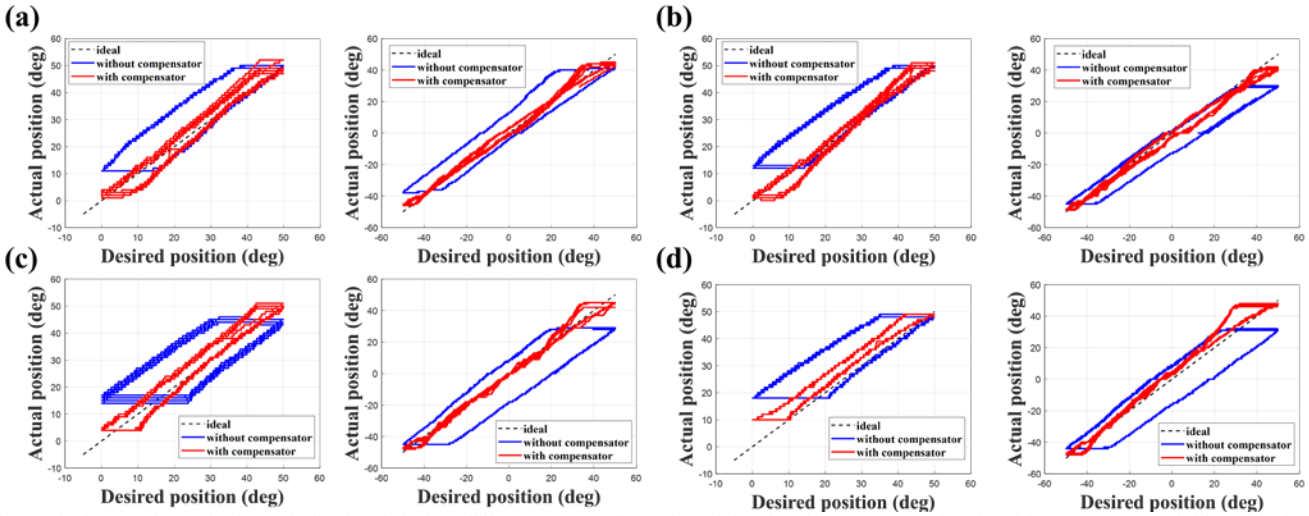

Fig. 8. The hysteresis results for periodic signals in four different case configuration of sheath (with compensated (red), without compensated (blue), ideal case (black dotted line)). (a) 0°. (b) 90°. (c) 180°. (d) 270°.

TABLE II. RESULT OF HYSTERESIS REDUCTION

		Case 1: $\mathbf{0^{\circ}}$		*Case 2:* $\mathbf{90^{\circ}}$		*Case 3:* $\mathbf{180^{\circ}}$		*Case 4:* $\mathbf{270^{\circ}}$	
		$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$
Hysteresis size	***Without compensator***	17°	23°	18°	17°	30°	28°	19°	30°
	With compensator	**8°**	**9°**	**7°**	**7°**	**9°**	**8°**	**8°**	**12°**
RMSE	***Without compensator***	8.03°	7.97°	8.20°	11.64°	9.57°	13.99°	11.18°	12.51°
	With compensator	**2.99°**	**3.37°**	**2.96°**	**4.38°**	**3.51°**	**3.28°**	**5.46°**	**4.94°**

Additionally, we will change the architecture of network and weight initialization method for improving our proposed pose estimation. The proposed pose estimation method may have some limitations in real surgical environment where optical illumination, dynamic background and surgical smoke can affect the image segmentation outcome. The effect of these factors in pose estimation will be further evaluated. Moreover, a more robust image segmentation method using deep learning will be tried in our further works.

C. Evaluation of Hysteresis reduction

To validate the proposed compensator to reduce the hysteresis in various sheath configurations, four different sheath configurations were used as shown in Fig. 6(b). Seven trials periodic sinusoidal signals were given. As a measurement of evaluation, the hysteresis size defined as the peak-to-peak error between the desired trajectory and measured output was used. In addition, the RMS deviation between the desired angle and the actual angle measured by an encoder attached at the joint was also utilized. The hyperparameters K_p, K_i, and K_d were determined by manual tuning.

The results are summarized in table II and illustrated in Fig. 8. In all four cases, the size of hysteresis was reduced to less than 10° by averaging the results at θ_1 and θ_2. Each value shown in table II corresponds to the maximum hysteresis size. The average reduction rate of RMSE in the four cases was $60.24 \pm 0.37\%$ at $\theta_1$and $65.15 \pm 0.65\%$ at θ_2.

Previous studies have reduced the hysteresis by 4° which is stronger result, but this was achieved using encoders that cannot be employed in surgical applications [6, 7]. Compared to the image-based approach [15], the proposed method shows better performance. The proposed method reduces hysteresis without degrading performance in various configurations of sheath. It is expected that the accuracy of the surgical instrument can be improved even in the stomach and colon which have multiple bending sections due to the ability to reduce hysteresis from various angles. Some graphs in Fig. 8 show a relatively large hysteresis at both ends, where the position direction changes. This is the caused by the significant influence of backlash due to TSM at each ends. To reduce this gap, the proposed compensator should converge quickly to the target point. This can be implemented by increasing the hyperparameters K_p and K_d, but this is limited due to a slower sampling time T_c. The performance of the compensator may be improved by increasing the feedback sampling time using a filter such as a Kalman filter and this will be conducted in the future works. When considering the expansion of degree of freedom, the rotational and translational motions are not driven by TSM. This corresponds that the additional motions are less affected by hysteresis. Therefore, it is not necessary to estimate the value of additional motions. However, as the degree of freedom increases, the variance of input into the network also increases. This enables to reduce the performance of the proposed pose estimation, but it may be improved by reflecting the effect of increased degree of freedom on training dataset.

IV. CONCLUSION AND FURTHER WORKS

In this paper, a hysteresis compensator with learning-based pose estimation was proposed. This compensator reduces the hysteresis of the surgical instrument by controlling the position input in various configurations of sheath. The

proposed pose estimation, based on an image retrieval approach with a SCNN, estimates the actual joint angle without using any markers. A flexible endoscopic surgery robot testbed was used to validate the proposed method. The testbed was designed to take into account several considerations required of an actual flexible surgical robot. As a result, the observed hysteresis was significantly reduced to less than 10° by the proposed compensator, in various sheath configurations. Moreover, the proposed pose estimation estimated the actual joint angle in real time without using any markers in clinical background image as well.

In future works, we will apply a more robust image segmentation algorithm to reduce the variance of input in the network and to consider the disturbances caused by factors in the actual surgical environment. Then, we will apply the proposed method to the flexible endoscopic surgery robot K-FLEX [1] which is developed by the authors group.

ACKNOWLEDGMENT

This research was supported by a grant of the Korea Evaluation Institute of Industrial Technology (KEIT), funded by the Ministry of Trade, Industry and Energy, Republic of Korea(Grant#:10060158), and International Joint Technology Development Project funded by the Korean Ministry of Trade, Industry and Energy (Grant #: P0006718)

REFERENCES

[1] M. Hwang, D. S. Kwon ″Surgical Joint with Enhanced Payload Capability with its application to Flexible Endoscopic Surgery Robot", Ph.D. Dissertation, Korea Advanced Institute of Science and Technology, August 2017.

[2] De Donno A, Zorn L, Zanne P, et al. Introducing STRAS: a new flexible robotic system for minimally invasive surgery. 2013 IEEE International Conference on Robotics and Automation (ICRA). 2013 1213-1220. http://doi.org/10.1109/ICRA.2013.6630726

[3] Camarillo DB, Milne CF, Carlson CR, et al. Mechanics modeling of Tendon - driven continuum manipulators. IEEE Trans Robot. 2008;24(6):1262-1273.

[4] Agrawal V, Peine WJ, Yao B. Modeling of transmission characteristics across a cable-conduit system. IEEE Trans Robot. 2010;26(5):914-924. http://doi.org/10.1109/TRO.2010.2064014

[5] Ikhouane F, Rodellar J. Systems with hysteresis: analysis: identification and control using the Bouc–Wen model. Hoboken: John Wiley & Sons, Ltd.; 2007.

[6] TN Do, T Tjahjowidodo, MWS Lau, T Yamamoto, and SJ Phee. Hysteresis modeling and position control of tendon-sheath mechanism in flexible endoscopic systems. Mechatronics, 24(1):12–22, 2014.

[7] T.N.Do, T.Tjahjowidodo , M.W.S.Lau S.J.Phee. Adaptive control for enhancing tracking performances of flexible tendonsheath mechanism in natural orifice transluminal endoscopic surgery (notes). Mechatronics (2015), 16(2):161–188, 2015.

[8] Bardou B, Zanne P, Nageotte F, de Mathelin M. Control of a multiple sections flexible endoscopic system. In Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. 2010: 2345–2350.

[9] Reichl T, Gardiazabal J, Navab N. Electromagnetic servoing - a new tracking paradigm. IEEE Trans on Medical Imaging. 2013; 32(8):1526-1535.

[10] M. Allan, S. Ourselin, D. J. Hawkes, J. D. Kelly, and D. Stoyanov, "3-d pose estimation of articulated instruments in robotic minimally invasive surgery," in TMI, 2017.

[11] R. Reilink et al., "Three-dimensional pose reconstruction of flexible instruments from endoscopic images", Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst., pp. 2076-2082, 2011.

[12] Cabras P, Nageotte F, Zanne P, Doignon C. An adaptive and fully automatic method for estimating the 3D position of bendable instruments using endoscopic images. Int J Med Robotics Comput Assist Surg. 2017;13:e1812. https://doi.org/10.1002/rcs.1812

[13] L. Zhang, M. Ye, P.L. Chan et al., "Real-time surgical tool tracking and pose estimation using a hybrid cylindrical marker", Int. J. Comput. Assist. Radiol. Surg., vol. 12, no. 6, pp. 921-930, 2017.

[14] P. Cabras et al., "Comparison of methods for estimating the position of actuated instruments in flexible endoscopic surgery", *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 3522-3528, 2014.

[15] Reilink R, Stramigioli S, Misra S. Image-based hysteresis reduction for the control of flexible endoscopic instruments. Mechatronics 2013; 23:652–8.

[16] Bromley, Jane, Bentz, James W, Bottou, Leon, Guyon, ´ Isabelle, LeCun, Yann, Moore, Cliff, Sackinger, Ed- ¨ uard, and Shah, Roopak. Signature verification using a siamese time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 7 (04):669–688, 1993

[17] https://en.wikipedia.org/wiki/Image_retrieval

[18] S. Hutchinson, G.D. Hager, P. Corke, "A Tutorial Introduction to Visual Servo Control", IEEE Trans. Robotics and Automation, vol. 12, no. 5, pp. 651-670, 1996.

[19] H. M. AI-Rahmani and G. F. Franklin, "Multirate control: A new approach," Automatica, vol. 28, pp. 35-44, 1992.

[20] H. Fujimoto, Y. Hori, and A. Kawamura, "Perfect Tracking Control Based on Multirate Feedforward Control With Generalized Sampling Periods," IEEE Trans. on Industrial Electronics, vol. 48, no. 3, pp.636-644, June 2001.

[21] P. P. Acharjya, A. Sinha, S.Sarkar, S. Dey,and S.Ghosh, "A new Approach Of Watershed Algorithm Using Distance Transform Applied To Image Segmentation",International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, 2013,Issue 2.ISSN (Print) :pp. 2320 –9798.

[22] Kinematic and Dynamic Analysis of A Surgical Tool Manipulator Towards Robotic Surgery Sanjeev Soni, Kuldeep Singh, Sanjeev Verma, Dinesh Pankaj, Amod Kumar.

[23] D. Kingma, J. Ba, "Adam: A method for stochastic optimization", ICLR, 2015.

[24] Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 2010.

[25] Xufeng Han, Thomas Leung, Yangqing Jia, Rahul Sukthankar, and Alexander C Berg. MatchNet: Unifying feature and metric learning for patch-based matching. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.