A03-KB102

Autonomous Control of a Surgical Assistant Robot with a Stereoscopic Endoscope That Can Obtain Depth Information in Real Time

- Progress Overview FY2017 -

Atsushi Nishikawa [*1], Noriyasu Iwamoto [*2], Toshikazu Kawai [#3], Hisashi Suzuki [$4], Hitoshi Katai [%5]

[*] *Department of Mechanical Engineering and Robotics, Faculty of Textile Science and Technology, Shinshu University*
3-15-1 Tokida, Ueda 386-8567 Japan

[1] nishikawa@shinshu-u.ac.jp

[2] iwamoto@shinshu-u.ac.jp

[#] *Department of Robotics, Faculty of Robotics and Design, Osaka Institute of Technology*
1-45 Chayamachi, Kita-ku, Osaka 530-8568 Japan

[3] toshikazu.kawai@oit.ac.jp

[$] *Department of Information and System Engineering, Faculty of Science and Engineering, Chuo University*
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551 Japan

[4] suzuki@ise.chuo-u.ac.jp

[%] *Department of Gastric Surgery, National Cancer Center Hospital*
5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045 Japan

[5] hkatai@ncc.go.jp

Abstract—**This research presents the use of a stereoscopic endoscope that obtains depth information in real time for assistant-free robotic laparoscopic surgery. First, the image plane-based motion controller was implemented on a robot manipulator holding the stereo endoscope. Next, an algorithm for real-time visual tracking of surgical instruments was developed. The proposed stereo-matching engine was then applied to a stereo-image pair during laparoscopic surgery. Our next plans include real-time estimation of the goodness of the field of view, called the *activity*, during the surgery. The activity will be estimated using both the surgeon's joint angle information obtained from ergonomic sensors and the depth information of the instrument tip tracked in the endoscopic images. Based on multidisciplinary computational anatomy information and a biologically inspired mathematical model called the *attractor selection model*, we finally establish a new methodology for autonomous "stereoscopic" endoscope manipulation.**

I. INTRODUCTION

This study aims to develop an autonomous laparoscope manipulating robot that can replace human camera assistants (Fig.1). In 1997, Wei et al. [1] pioneered an autonomous "stereoscopic" endoscope guidance system for laparoscopic surgery and evaluated its performance in clinical tests [2]. Since then, several automatic camera positioning systems have been developed over the past 20 years (e.g., [3],[4]). These systems assume that the distal end of the surgical instrument in the laparoscopic image is the surgeon's region of interest. Immediately before surgery, the desired position of the instrument tip in the image is obtained by the "teach by showing" approach [1]-[3], and is provided in advance or updated during the surgery by simple heuristic approaches [4]. Although these approaches remove the camera-control burden from the surgeon, the presented view may be sub-optimal because the most appropriate field of view (FOV) varies dynamically and adaptively throughout the surgery. The best FOV depends on both the surgical procedure/phase and the habits/preferences of the operating surgeon. The FOV problem can be overcome by a biologically inspired mathematical model called the *attractor selection model*. Proposed by Kashiwagi et al. [5] and described in the next section, this model formalizes the dynamical and self-adaptive behavior of gene regulatory networks found in biological cells, which has been generalized as a stochastic differential equation [6]. We will develop a new autonomous stereo-laparoscope manipulating robot based on the generalized model of attractor selection.

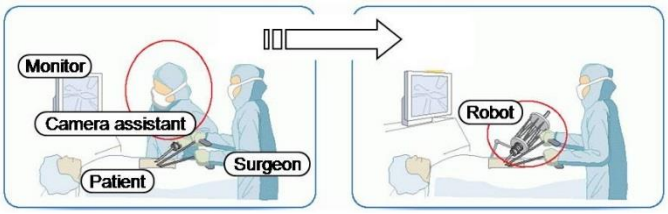


Fig. 1. Laparoscopic surgery. (left) Conventional case in which the laparoscope is operated by a human camera assistant. (right) Robot-assisted case in which the laparoscope is operated by a robot manipulator.

II. Concept Overview

A. Problem Setting

Suppose that a stereo-endoscopic manipulator observes a surgical instrument during laparoscopic surgery. Let $\boldsymbol{x_L^*} = (x_L^*, y_L^*)$ and $\boldsymbol{x_R^*} = (x_R^*, y_R^*)$ be the current positions of the surgical instrument tip in the left and right images, respectively, and let $\boldsymbol{x_L} = (x_L, y_L)$ and $\boldsymbol{x_R} = (x_R, y_R)$ be the corresponding desired positions. If the current tip positions $\boldsymbol{x_L^*}$ and $\boldsymbol{x_R^*}$ are estimated robustly though real-time visual tracking, and the desired tip positions $\boldsymbol{x_L}$ and $\boldsymbol{x_R}$ are explored adaptively as the surgical task proceeds, the direction and speed of the manipulator motion can be determined from the following (dx, dy, dz) parameters [1]:

$$dx = \frac{(x_L^* + x_R^*)}{2} - \frac{(x_L + x_R)}{2} \quad (1)$$

$$dy = \frac{(y_L^* + y_R^*)}{2} - \frac{(y_L + y_R)}{2} \quad (2)$$

$$dz = \|\boldsymbol{x_L^*} - \boldsymbol{x_R^*}\| - \|\boldsymbol{x_L} - \boldsymbol{x_R}\| \quad (3)$$

Without loss of generality, we can assume a standard (parallel) stereo endoscope realized through image rectification. In this case, we have $y_L^* = y_R^*$ and $y_L = y_R$ (epipolar line constraint). Denoting the current and desired disparities of the surgical instrument tip by $d^* = x_L^* - x_R^*$ and $d = x_L - x_R$ respectively, we can rewrite Equations (1)–(3) as follows:

$$dx = (x_L^* - x_L) - \frac{(d^* - d)}{2} \quad (4)$$

$$dy = y_L^* - y_L \quad (5)$$

$$dz = d^* - d \quad (6)$$

Thus, to generate the manipulator control command, we must solve the following two problems:

Problem I (real-time visual tracking & stereo matching):

Estimate $\boldsymbol{x}^* = (x_L^*, y_L^*, d^*)$

Problem II (attractor selection & FOV evaluation):

Explore $\boldsymbol{x} = (x_L, y_L, d)$

In FY2017, we focused on the first problem. The result (progress overview in FY2017) is partially given in Section III. The remainder of this section outlines the attractor selection model for solving the second problem, which will appear in FY2018.

B. Attractor selection model

A possible solution for $\boldsymbol{x}$ is referred to as an *attractor*. Suppose $U(\boldsymbol{x})$ is a potential function with multiple attractors. Let t ($\in [0, t_{max}]$) be the "virtual" time for exploring solutions of $\boldsymbol{x}$. Attractor selection can be generalized by the following stochastic differential equation [6]:

$$\frac{d}{dt}\boldsymbol{x}(t) = \boldsymbol{f}(\boldsymbol{x}(t)) \cdot activity + \boldsymbol{\eta}(t) \quad (7)$$

Suppose that the function $\boldsymbol{f}$ depends on the state $\boldsymbol{x}(t)$ at time t and can be represented as $\boldsymbol{f}(\boldsymbol{x}) = -\partial U(\boldsymbol{x})/\partial \boldsymbol{x}$. The scalar value $activity$ ($\in [0,1]$) (which is constant during the virtual exploration time) indicates the fitness of the current FOV to the surgical task. The fluctuation term $\boldsymbol{\eta}(t)$ is assumed as zero-mean white Gaussian noise with a given standard deviation. The value $activity$ should be designed to be large (small) when the current FOV is suited (not suited) to the surgical task, and should be acquired/updated in real time during the surgery. Notice that as $activity$ increases, the term $\boldsymbol{f}(\boldsymbol{x}) \cdot activity$ becomes more dominant in Eq. (7), and the state transition becomes more deterministic. Consequently, state $\boldsymbol{x}$ tends to be entrained into a suitable attractor, where it remains despite the persistent noise terms. On the other hand, decreasing $activity$ increases the dominance of the noise $\boldsymbol{\eta}$, thereby flattening the potential landscape. In this scenario, the state transition becomes more probabilistic, like a random walk, and $\boldsymbol{x}$ is driven away from the current attractor.

In our current plans, the potential function $U(\boldsymbol{x})$ will be derived from various surgical video databases or from multidisciplinary computational anatomy information. Moreover, the $activity$ value will be based on the surgeon's joint angle information from ergonomic sensors and the depth (disparity) information detected in endoscopic images.

III. Progress Overview (FY2017)

In this section, we briefly introduce our research progress presented in FY2017.

A. Robot Controller: (dx, dy, dz) based motion control

Fig. 2 shows the robot manipulator AESOP 3000 (Computer Motion Inc., USA) and the stereo endoscope (Kaneko Manufacturing Co. Ltd, Japan) used in our evaluations. We implemented eight computer-controlled commands via the middleware ORiN (Open Resource Interface for the Network, by the Japan Robot Association) [8]:
MOVE_LEFT, MOVE_RIGHT, MOVE_UP, MOVE_DOWN, MOVE_FORWARD, MOVE_BACK,
STOP_MANIPULATION, SET_LIMIT.

Fig.2. The robot manipulator AESOP3000 and our stereo endoscope

Fig. 3. Example of estimating the tip positions of surgical instruments

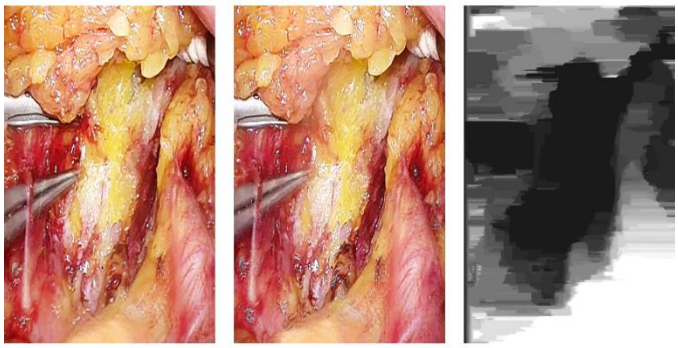


Fig. 4. Example of disparity estimation in a stereo image during surgery

The commands MOVE_LEFT, MOVE_RIGHT, MOVE_UP, and MOVE_DOWN directly control the motion in the image plane. The commands MOVE_FORWARD and MOVE_BACK provide motions orthogonal to the image plane, that is, insertion and retraction of the laparoscope. During the surgery, the robot controller assesses the sign and magnitude of (dx, dy, dz), and autonomously selects one of these six commands. STOP_MANIPULATION and SET_LIMIT indicate the robot motion stop command and the robot movable-limit setting command, respectively.

B. Instrument Tracking: tip position (x_L^*, y_L^*) estimation

The proposed instrument-tracking method binarizes the images based on their color features, including their R, G, and B values in RGB color space and their S (Saturation) value in HSV color space. Binarization is followed by three dilation iterations, three erosion iterations, and labeling processes. Fig. 3 shows the real-time tracking result of four surgical instruments using the proposed method. Note that the positions of the surgical instrument tips are correctly estimated.

C. Stereo Matching: disparity d^* estimation

Previously, we developed a stereo-matching engine [9] that controls the trade-off between the correct-match percentage and the processing speed. This engine achieved correct-matching rates of 94.4% and 92.1% for typical stereo images in the Middlebury Stereo Datasets [7] at processing speeds of 21.0 MDE/s and 198.3 MDE/s, respectively (where the unit MDE/s denotes one million disparity estimations per second). Fig. 4 shows the disparity detection result of the proposed system for a stereo-image pair during laparoscopic surgery. Furthermore, Kaneko Manufacturing Co., Ltd. (Saitama, Japan) has implemented a 3D reconstruction unit that inputs stereo images of 24-bit color FHD (1920×1080 pixels) in side-by-side format. From an RGB left image and an RGB right image, this unit calculates depth images at 33 fps, and outputs stereo images of 24-bit color FHD in side-by-side format. The output images comprise an RGB left image and a gray-scaled image denoting the depth distribution [9].

IV. Conclusion

This paper briefly reported on the research progress (FY2017) in the Recruited Group A03-KB102: "Autonomous Control of a Surgical Assistant Robot with a Stereoscopic Endoscope That Can Obtain Depth Information in Real Time."

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP17H05287. The endoscopic robot controllers have been developed in collaboration with Professor Ken Masamune of the Project Group A03-3. The stereo endoscope and the 3D reconstruction unit have been developed in collaboration with Kaneko Manufacturing Co., Ltd. We have no international collaborative research partners.

References

[1] G. Q. Wei, K. Arbter, G. Hirzinger, "Automatic tracking of laparoscopic instruments by color-coding," in *Proc. First Int. Joint Conf. CVRMed-MRCAS'97*, Springer LNCS 1205, pp.357-366, 1997.

[2] K. Omote, et al., "Self-guided robotic camera control for laparoscopic surgery compared with human camera control," *American Journal of Surgery*, vol. 177, no. 4, pp. 321-324, Apr. 1999.

[3] A. Nishikawa, et al., "Automatic positioning of a laparoscope by preoperative workspace planning and intraoperative 3D instrument tracking," in *MICCAI2006 Workshop Proc. on Medical Robotics*, pp. 82-91, Copenhagen, Denmark, Oct. 2006.

[4] A. Nishikawa, et al., "How does the camera assistant decide the zooming ratio of laparoscopic images? Analysis and implementation," in *Proc. MICCAI2008*, Springer LNCS 5242, pp. 611-618, 2008.

[5] A. Kashiwagi, I. Urabe, K. Kaneko, and T. Yomo, "Adaptive response of a gene network to environmental changes by fitness-induced attractor selection," *PLoS ONE*, vol. 1, no. 1, e49, Dec. 2006.

[6] K. Leibnitz and M. Koda, "Analysis of noise sensitivity of attractor selection," *RIMS Kôkyûroku*, vol. 1620, pp. 150-161, Jan. 2009.

[7] (2018) The Middlebury Computer Vision Pages, Middlebury Stereo Datasets [Online]. Available: http://vision.middlebury.edu/stereo/data/

List of Accepted and Published Papers since 1st April 2017

[8] N. Iwamoto, A. Nishikawa, T. Kawai, Y. Horise, K. Masamune, "MRLink: An ORiN-Based Medical Robot Architecture for Connecting Freely Selected Master/Slave Robotic Devices," in *Proc. IROS2017 Workshop on Shared Platforms for Medical Robotics Research*, Vancouver, BC, Canada, Sep. 24, 2017.

[9] J. Z. Lim, H. Suzuki, S. Utsugi, H. Katai, "Experimental Development of a Multi-View Stereo Endoscope System," in *Proc. 2nd Russian-Pacific Conference on Computer Technology and Applications (RPC 2017)*, Vladivostok, Russia, Sep. 25-29, 2017.

[10] A. Nishikawa, "Human-Robot Interaction in Minimally Invasive Surgery and Therapy," *JSPS Alumni Club in Sweden (SAC) Activity Seminar: Smart Textiles - Technology for Medicine and Healthcare*, University of Borås, Sweden, Oct. 19, 2017.

A03-KB102 Autonomous Control of a Surgical Assistant Robot with a Stereoscopic Endoscope That Can Obtain Depth Information in Real Time

Atsushi Nishikawa (*Shinshu University*)
Noriyasu Iwamoto (*Shinshu University*), *Toshikazu Kawai* (*Osaka Institute of Technology*),
Hisashi Suzuki (*Chuo University*), *Hitoshi Katai* (*National Cancer Center Hospital*)

☑ SPACE
☑ TIME
☐ FUNC.
☐ PATHOL.

1. Introduction (summary)

This research presents the use of a stereoscopic endoscope that obtains depth information in real time for assistant-free robotic laparoscopic surgery.

First, the image plane-based motion controller was implemented on a robot manipulator holding the stereo endoscope. Next, an algorithm for real-time visual tracking of surgical instruments was developed. The proposed stereo-matching engine was then applied to a stereo-image pair during laparoscopic surgery.

Our next plans include real-time estimation of the goodness of the field of view (FOV), called the $activity$, during the surgery. Based on a biologically inspired mathematical model called the attractor selection model, we finally establish a new methodology for autonomous "stereoscopic" endoscope manipulation.

2. Concept overview

◆ Problem setting

Suppose that a stereo-endoscopic manipulator observes a surgical instrument during laparoscopic surgery. Let $\boldsymbol{x_L^*} = (x_L^*, y_L^*)$ and $\boldsymbol{x_R^*} = (x_R^*, y_R^*)$ be the current positions of the surgical instrument tip in the left and right images, respectively, and let $\boldsymbol{x_L} = (x_L, y_L)$ and $\boldsymbol{x_R} = (x_R, y_R)$ be the corresponding desired positions.

Without loss of generality, we can assume a standard (parallel) stereo endoscope realized through image rectification. In this case, we have $y_L^* = y_R^*$ and $y_L = y_R$ (epipolar line constraint). Denoting the current and desired disparities of the surgical instrument tip by $d^* = x_L^* - x_R^*$ and $d = x_L - x_R$ respectively. In order to generate the manipulator control command, we must solve the following two problems:

☑ *Problem I*
Real-time instrument tracking & stereo matching:
Estimate $\boldsymbol{x}^* = (x_L^*, y_L^*, d^*)$

☐ *Problem II*
Real-time attractor selection & FOV evaluation:
Explore $\boldsymbol{x} = (x_L, y_L, d)$

◆ Attractor selection model

A possible solution for $\boldsymbol{x}$ is referred to as an *attractor*. Suppose $U(\boldsymbol{x})$ is a potential function with multiple attractors. Let t ($\in [0, t_{max}]$) be the "virtual" time for exploring solutions of $\boldsymbol{x}$. Attractor selection can be generalized by the following stochastic differential equation:

$$\frac{d}{dt}\boldsymbol{x}(t) = \boldsymbol{f}(\boldsymbol{x}(t)) \cdot activity + \boldsymbol{\eta}(t) \qquad (1)$$

Suppose that the function $\boldsymbol{f}$ depends on the state $\boldsymbol{x}(t)$ at time t and can be represented as $\boldsymbol{f}(\boldsymbol{x}) = -\partial U(\boldsymbol{x})/\partial \boldsymbol{x}$. The scalar value $activity$ ($\in [0,1]$) (which is constant during the virtual exploration time) indicates the fitness of the current FOV to the surgical task. The fluctuation term $\boldsymbol{\eta}(t)$ is assumed as zero-mean white Gaussian noise with a given standard deviation. The value $activity$ should be designed to be large (small) when the current FOV is suited (not suited) to the surgical task, and should be acquired/updated in real time during the surgery. Notice that as $activity$ increases, the term $\boldsymbol{f}(\boldsymbol{x}) \cdot activity$ becomes more dominant in Eq. (1), and the state transition becomes more deterministic. Consequently, state $\boldsymbol{x}$ tends to be entrained into a suitable attractor, where it remains despite the persistent noise terms. On the other hand, decreasing $activity$ increases the dominance of the noise $\boldsymbol{\eta}$, thereby flattening the potential landscape. In this scenario, the state transition becomes more probabilistic, like a random walk, and $\boldsymbol{x}$ is driven away from the current attractor.

3. Progress overview

◆ Robot controller: Image plane-based motion control

Fig. 1 shows the robot manipulator AESOP 3000 (Computer Motion Inc., USA) and the stereo endoscope (Kaneko Manufacturing Co. Ltd, Japan) used in our evaluations. We implemented eight computer-controlled commands. The commands MOVE_LEFT, MOVE_RIGHT, MOVE_UP, and MOVE_DOWN directly control the motion in the image plane. The commands MOVE_FORWARD and MOVE_BACK provide motions orthogonal to the image plane, that is, insertion and retraction of the laparoscope. During the surgery, the robot controller assesses the sign/magnitude of (dx, dy, dz): $dx = x_L^* - x_L,\ \ dy = y_L^* - y_L,\ \ dz = d^* - d$ and autonomously selects one of these six commands. STOP_MANIPULATION and SET_LIMIT indicate the robot motion stop command and the robot movable-limit setting command, respectively.

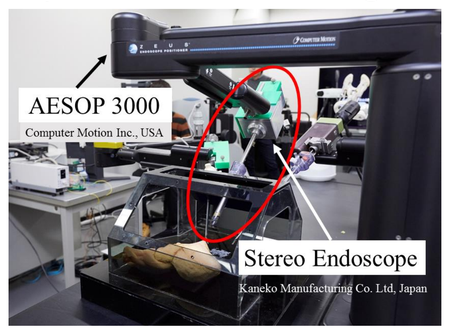


Fig. 1 Our stereo-endoscope manipulator

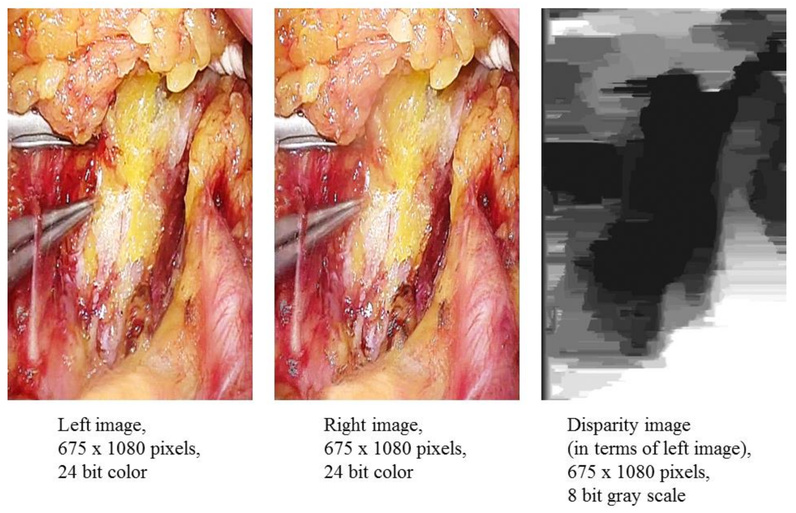
◆ Instrument tracking: Tip position (x_L^*, y_L^*) estimation

The proposed instrument-tracking method binarizes the images based on their color features, including their R, G, and B values in RGB color space and their S (Saturation) value in HSV color space. Binarization is followed by three dilation iterations, three erosion iterations, and labeling processes. Fig. 2 shows the real-time tracking result of four surgical instruments using the proposed method. Note that the positions of the surgical instrument tips are correctly estimated.

Fig.2 Example of estimating the tip positions of surgical instruments

◆ Stereo matching: Disparity d^* estimation

We developed a stereo-matching engine that controls the trade-off between the correct-match percentage and the processing speed. This engine achieved correct-matching rates of 94.4% and 92.1% for typical stereo images in the Middlebury Stereo Datasets at processing speeds of 21.0 MDE/s and 198.3 MDE/s, respectively (where the unit MDE/s denotes one million disparity estimations per second). Fig. 3 shows the disparity detection result of the proposed system for a stereo-image pair during laparoscopic surgery.



Left image, 675 x 1080 pixels, 24 bit color
Right image, 675 x 1080 pixels, 24 bit color
Disparity image (in terms of left image), 675 x 1080 pixels, 8 bit gray scale

Fig. 3 Example of disparity estimation in a stereo image during surgery

4. Conclusion and next plans

This poster briefly reported on the research progress (FY2017) in the Recruited Group A03-KB102. In our current plans, the potential function $U(\boldsymbol{x})$ will be derived from various surgical video databases preoperatively. Moreover, the $activity$ value will be based on the surgeon's joint angle information from ergonomic sensors and the depth (disparity) information detected intraoperatively in endoscopic images.

This work was supported by JSPS KAKENHI Grant Number JP17H05287.
The stereo endoscope and the 3D reconstruction unit have been developed in collaboration with Kaneko Manufacturing Co., Ltd. (Saitama, Japan).

多元計算解剖学
Multidisciplinary Computational Anatomy