2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

A Hybrid Vision/Force Control Strategy for Handheld Robotic Devices Enhancing Probe-Based Confocal Laser Endomicroscopy

Ingu Choi, Eunchan Kim, and Sungwook, *Member, IEEE,*

***Abstract*— We introduce a novel hybrid vision/force control strategy for robotic devices designed to obtain clear and consistent images using probe-based confocal laser endomicroscopy (pCLE). Due to the variable nature of tissue characteristics encountered during pCLE imaging, the conventional approach of pre-setting forces or focus metrics for either force or vision control is often impractical and inadequate. To address this, our strategy employs a blur metric called the CR score to assess the level of blur in pCLE images, enabling the attainment of clear and focused images. At the onset of a pCLE scan, the system autonomously determines the target CR score for vision control, in tandem with a real-time peak detection algorithm. Concurrently, force control is applied judiciously to prevent excessive force on the tissue, adjusting to ensure minimal force is applied, thus preserving image focus. This innovative approach facilitates seamless transitions between vision and force control, depending on the imaging conditions, thereby ensuring the acquisition of consistent pCLE images with minimal force. Our method marks a notable improvement over conventional PID force control techniques. By dynamically adjusting target forces and minimizing force application during operation, we not only enhance the precision and quality of pCLE imaging but also eliminate the dependency on manual pre-settings.**

I. Introduction

Optical imaging systems equipped with probe-based confocal laser endomicroscopy (pCLE) enable the diagnosis and treatment of biological tissues at the cellular level, both *in situ* and *in vivo* [1]. pCLE facilitates the real-time acquisition of optical biopsies with its flexibility and compactness, making it widely applicable in clinical settings, particularly in the gastrointestinal, urological, and respiratory tracts [2].

However, the field of view (FoV) of pCLE is typically limited to micro- to sub-millimeter ranges, which restricts the diagnostic coverage to a narrow area [3]. To address this limitation, mosaicking techniques have been employed to create expansive maps by stitching images from consecutive frames [4]. This approach allows for real-time observation of broader areas, effectively overcoming the inherent narrow FoV of pCLE. However, image quality is compromised by various disturbance parameters, including deformation, dragging, respiration, and hand tremor, which hinder seamless mosaicking. Specifically, the method is susceptible to acquiring blurred images that are out of focus due to the shallow depth of field during handheld scanning of the imaging probe.

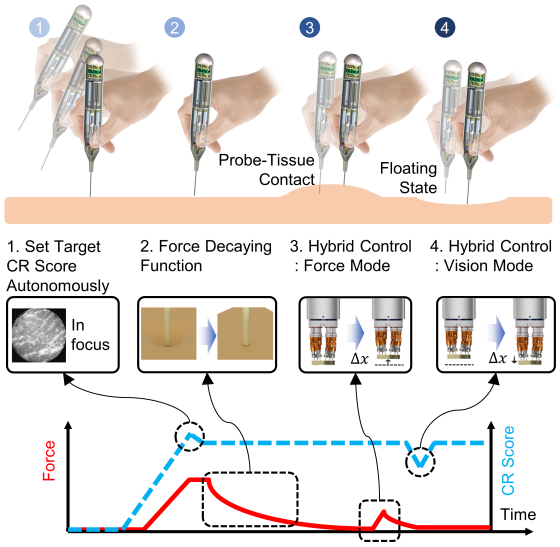

Fig. 1. Proposed semi-autonomous hybrid control strategy: It dynamically adjusts to maintain high-quality pCLE imaging with minimal force, responding to changes such as tissue contact and respiratory motion by alternating between force and vision modes.

Force-based control has been a key focus in various robotic approaches introduced to address the limitations arising from manual operation, particularly when interacting with biological tissues [5]–[9]. For example, strategies maintaining the distance between the probe and the tissue by controlling the force value have been proposed. Latt *et al.* present the results of an early study, in which pCLE imaging was performed using a voice coil actuator in ex-vivo experiments with a force magnitude of 100 mN using a simple PID force control [5]. By extending this study, a device that can be used in minimally invasive surgery has also been studied [6]. Wisanuvej *et al.* showed the results of improved mosaicking by controlling the orientation and force simultaneously [7]. Choi *et al.* proposed a handheld robotic system with three-axis force control to interact with biological tissues [8]. A critical factor influencing the quality of pCLE images is the distance between the probe and tissue.

Vision-based control has also been employed in research on robotic approaches for acquiring the high quality of pCLE images. Giataganas *et al.* have also conducted a study of mosaicking in a wide area by applying visual servoing

This work was supported partially by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. 2019M3A9E2061784) and also by by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00422069)

I. Choi is with the SW Campus, HL Mando, Seongnam, 13488, South Korea. cik970418@naver.com

E. Kim is with the LG Electronics, Seoul, 04763, Korea.

S. Yang are with the Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology, Seoul, 02792, South Korea. swyang@kist.re.kr

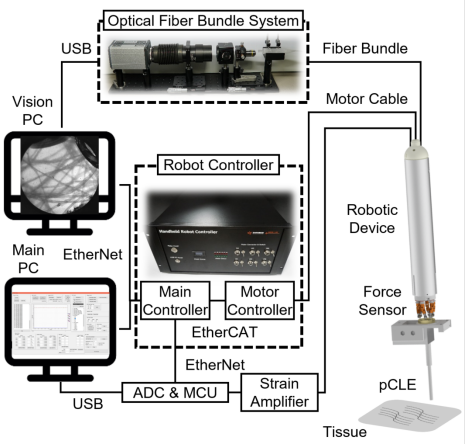

Fig. 2. Overview of the system comprising a handheld robot, optical fiber bundle imaging system, and robot controller.

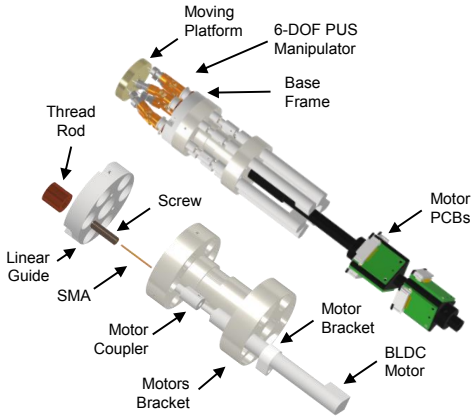

Fig. 3. Mechanical design of the handheld micromanipulator.

technology to a handheld device with added degrees of freedom [10]. Triantafyllou *et al.* obtained consistent images by controlling the position of the probe without a force sensor despite the tissue moving with in vivo respiration motion [11]. They implemented vision-based feedback control by evaluating the image blur level using the Crété-Roffet (CR) metric, a no-reference perceptual blur metric introduced in [12]. Li *et al.* conducted a study to acquire pCLE images in the eye in a non-contact type by applying image-based position control to the cooperative robot using the CR metric [13].

In light of the challenges posed by the variability of tissue properties during pCLE imaging, previous methodologies have faced limitations due to their reliance on predetermined target forces or CR scores. Recognizing the impracticality of these static approaches in the dynamic and complex real-world environment of pCLE imaging, we introduce a semiautonomous hybrid vision and force control strategy as illustrated in Fig. 1. This approach dynamically adjusts applied forces in real-time, ensuring the acquisition of high-quality pCLE images with minimal contact force. Our method is rigorously validated through comprehensive benchtop experiments and further assessed by executing precise imaging tasks on a dynamic phantom model.

II. Materials and Methods

A. Overall System

Our system incorporates a handheld robot integrated with the imaging probe of a custom-built pCLE system, as shown in Fig. 2. This handheld robot, outfitted with a GS0-500 load cell (Transducer Technique, USA), enables precise control of the contact force between the imaging probe and tissue. When operated by a user, the robot adeptly manipulates the imaging probe, facilitating both vision and force control during pCLE imaging sessions. For instance, it adjusts the probe to either maintain a specific focus metric—referred to here as the CR score—or to adhere to a predetermined target force, thereby ensuring optimal vision and force control.

B. Handheld Robot

The handheld robot integrates a six degrees-of-freedom (6-DoF) miniature micromanipulator for hybrid vision/force control, as depicted in Fig. 3. This micromanipulator, leveraging a six-prismatic-universal-spherical (6-PUS) architecture, underwent geometrical optimization to endure external forces up to 3.0 N, as detailed in [14]. This design enables precise 6-DoF manipulation of the tool tip, achieving both position and orientation control within a workspace defined by a cylindrical volume of 3 mm in diameter and 6 mm in height. Since the physiological hand tremor is known to have an RMS (root-mean-square) amplitude on an order of 50-200 μm at a frequency commonly in a 6-14 Hz band [15], the manipulator has sufficient workspace to compensate for both physiological tremors and respiratory motion of the tissue. Given the focus of this study on the hybrid vision and force control strategy, linear motion along the z-axis is used only for control.

C. Probe-based Confocal Endomicroscopy System (pCLE)

The pCLE system features an optical fiber bundle (FIGH-30-650S, Fujikura Ltd., Japan) serving as the imaging probe, enabling cellular-level tissue imaging. Images are captured using a sCMOS camera (Zyla 4.2 PLUS, Oxford Instruments, U.K.) through custom-designed optics. The fiber bundle, comprising 30,000 cores, achieves a resolution of 2–3 μm. Images, with a resolution of 1024 x 1024 pixels from the camera, are streamed to the vision PC at a rate of 100 fps. A CNN-based honeycomb artifact removal algorithm is used to obtain an artifact-free image, which is inevitably yielded by optical path blocking of cladding layers in fiber bundle imaging [16]. This preprocessing step is crucial, as the presence of artifacts can lead to inaccurate CR score calculations due to the pronounced honeycomb patterns. This processing is conducted on the vision PC for each image frame, and the resulting CR score is transmitted to the main controller via TCP/IP communication in real time, facilitating precise vision control.

III. Hybrid Vision/Force Control

The objective of the hybrid vision/force control strategy is to obtain clear and consistent images from pCLE, while ensuring the contact force on the tissue is minimized. Given the nature of the probe-based confocal imaging system, such that the focal plane is established at the end of the fiber bundle, it would be ideal to position the CLE probe directly above the tissue surface, exerting little to no force. However, achieving zero contact force during scanning to observe the lesion area is nearly impossible. Variations in the distance between the probe and the tissue, due to movements of the tissue or its structural differences, may lead to the probe's separation from the tissue, surpassing the depth of field and causing image blur. Conversely, accidental collisions can result in excessive force. This section describes the methodology for autonomously determining the target CR score at the onset of hybrid control and delves into the workings of the proposed control algorithm, aligning with the challenges and objectives described.

A. Setting Autonomous Target CR Score

We present an algorithm designed to autonomously ascertain the desired CR score, denoted as q_d, at the initiation of hybrid control. Fig. 4 (a) outlines the algorithm devised to achieve this objective. Among various metrics for assessing image focus or blur, CR is preferred for robustly evaluating the clarity and clinical diagnostic value of pCLE images [13]. The CR score quantifies image sharpness or blur by evaluating the contrast differences between neighboring pixels, with its calculation process detailed in [13].

Given the likelihood of the optimal image focus aligning with the maximal CR score, this peak value is determined at the imaging session's outset by examining the associated slope, as depicted in Fig. 4(b). Thus, when initially attempting to establish contact with the tissue for imaging, identifying a moment characterized by a pronounced increase and subsequent gradual decrease in the slope Sk becomes essential.

The CR score q_k is derived from the kth image I_k, stored in a FIFO (First In, First Out) buffer $q_{[k\ \ end,k]}$, whose capacity is user-defined, adopting a window of 1,000 corresponding to 1 s. The slope Sk is calculated as the difference between the initial value q_k and the end value $q_{k\ \ end}$ within the FIFO buffer. A sharp rise in the CR score accompanied by a mild decrease in the slope, indicating $MAX_{S_k} > SLOPE\_THRESHOLD$ and $S_k \leq 0$, respectively, identifies the peak CR score. Upon meeting this criterion, the peak CR score is established as the target CR score for hybrid control. Furthermore, if the CR score exceeds a specific threshold (around 0.4), suggesting comparable image quality, a safety factor f_{safety} (e. g., 95%) is applied to the peak CR score to set the desired CR score q_d. Given that this calibration involves probe-tissue contact, there's a potential risk of tissue damage. To minimize this risk, a force limitation strategy is integrated by defining a threshold $FORCE\_LIMIT$ to avert excessive force application. Figure 4(b) illustrates that the force exerted does not surpass 50 mN.

(a)

Algorithm 1: Set Target CR Score with Force Limit Control

Input: Image (I_k) at the current time k
Output: Target CR Score (q_d)

while targe CR score calibration mode **do**
$q_k = CR(I_k)$
$q_{[k-end,\,k]} = Window(q_k)$
$S_k = Slope\left(q_{[k-end,\,k]}\right)$
$MAX\_S_k = MAX(S_k)$
if ($F_c \geq FORCE\_LIMIT$)
  $F_d = 0$
  P_d = FORCE_PID(F_d, F_c)
end
If ($MAX\_S_k > SLOPE\_THRESHOLD$) && ($S_k \leq 0$) **then**
  return $q_d = q_k * f_{safety}$
end
end while

(b)

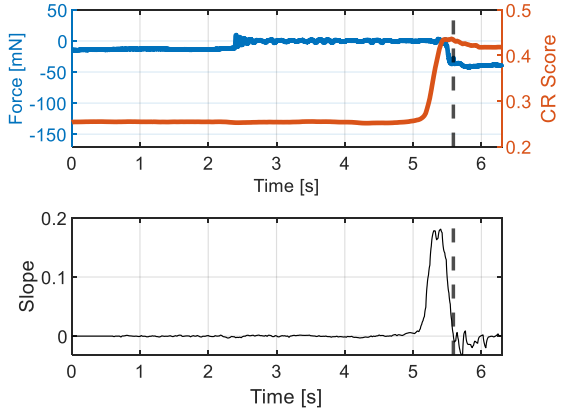

Fig. 4. Setting autonomous target CR score. (a) Autonomous target CR score setting algorithm with force limitation feature. (b) Implementation of Algorithm 1, with the dotted line indicating the moment the target CR score is determined.

B. Hybrid Vision/Force Control Algorithm

Figure 5 (a) presents the three main operational modes of this algorithm: Calibration (C), Vision Control (VC), and Force Control (FC). In Calibration mode, the algorithm calculates the target CR score as detailed in Section III. A. Once established, this target CR score becomes the reference for vision control. Within vision feedback control, two scenarios are possible: First, a minimal contact force approach may result in the probe floating in air, identified by the condition $q_c < q_d$, determining whether the probe is in contact or floating, where q_c is the current CR score. Second, excessive pressing of the tissue with the probe, when the CR score exceeds a certain threshold, leads to image degradation. To avoid this, a second condition, $F_c < F_{limit}$, ensures the probe does not exert too much force, where F_c is the current contact force. Outside these conditions, a switch from vision to force control is warranted. Force control introduces a scenario where the CR score's maintenance or lack thereof guides action. Maintaining the CR score suggests image quality is

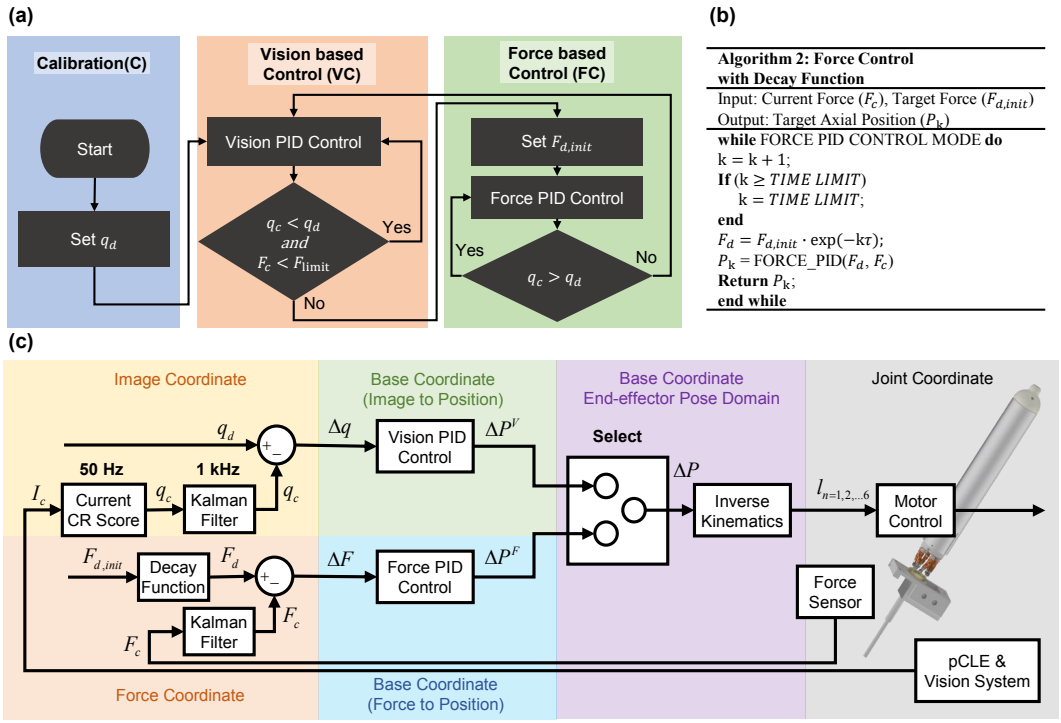

Fig. 5. Hybrid vision/force control algorithm: (a) switching algorithm for hybrid control, (b) force control algorithm with decay function, (c) control block diagram of hybrid vision/force control.

upheld with minimal force, while failing to do so indicates a floating state. Hence, the criterion for persisting in force control is $q_c > q_d$. Consequently, toggling between VC and FC creates a comprehensive closed-loop control system.

C. Implementation

This hybrid control scheme is implemented through active manipulation of the imaging probe's position via the handheld robot, which ultimately adjusts the length of the robot's prismatic joint as needed. As previously mentioned, the objective of the vision feedback control is to enable the probe to approach the tissue while acquiring clear, in-focus images via the CR score. To facilitate real-time calculation of the CR score, the size of the pCLE image is reduced to 256 x 256 pixels; however, this resizing process has minimal impact on the CR score itself. The vision feedback loop is established by computing the error Δq between the target CR score q_d and the current CR score q_c.

$$\Delta q = q_d - q_c \tag{1}$$

The vision feedback control utilizes PID control for rapid convergence. The PID gains for the vision control are denoted as K_p^q, K_i^q, and K_d^q, respectively.

$$\Delta P^V = K_p^q \Delta q + K_i^q \int \Delta q dt + K_d^q \frac{d}{dt} \Delta q \tag{2}$$

The goal of the force feedback control is to minimize the error ΔF between the target force F_d and the current force F_c.

$$\Delta F = F_d - F_c \tag{3}$$

To attenuate high-frequency noise from the force sensor, a linear Kalman filter that uses a constant velocity motion model with minimal phase delay is employed. Previous studies have demonstrated the efficacy of PID closed-loop force control for handheld robots used in pCLE imaging [8]. Therefore, this study adopts a closed-loop PID control scheme for force feedback control. The PID gains for the force control are denoted as K_p^F, K_i^F, and K_d^F, respectively.

$$\Delta P^F = K_p^F \Delta F + K_i^F \int \Delta F dt + K_d^F \frac{d}{dt} \Delta F \tag{4}$$

The hybrid control scheme also features a decaying function designed to ensure consistent imaging with minimal contact force by progressively lowering the initial target force value towards a minimal level. Unlike conventional force control schemes, this algorithm is employed to gradually reduce the contact force by decaying the target force, as presented in Fig. 5(b). The desired force F_d decreases over time k following an exponential function $\exp(-k\tau)$, where the decay rate is modulated by adjusting the τ parameter. This exponential decay function plays a pivotal role in minimizing the duration of tissue contact. The initially set target force $Fd, init$ experiences a gradual reduction over time k, enabling the robot's end effector to apply minimal force on the tissue persistently. Such an approach is vital for achieving consistent tissue imaging while avoiding tissue damage and deformation caused by excessive force.

The output of each control loop indicates the axial position change of the tool tip, ΔP^V for vision control and ΔP^F for

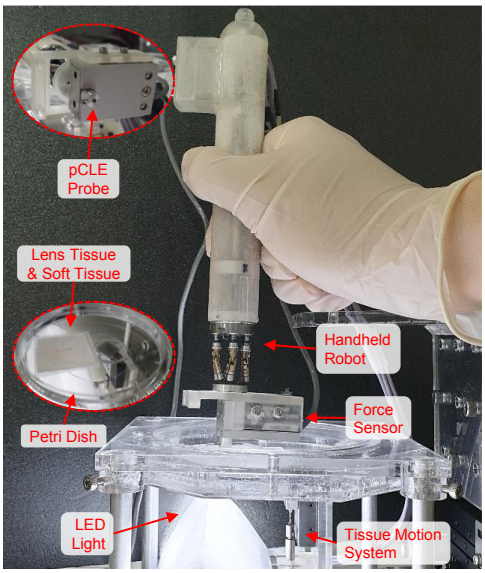

Fig. 6. Experimental setup with a tissue motion generation system mimicking vertical respiratory movements. An Eco-Flex phantom, augmented with lens tissue, serves as the biological tissue model.

force control, respectively. As shown in Fig. 5(a), the system determines whether to adopt the position change value from either vision feedback, ΔP^V, or force feedback, ΔP^F, based on the mode selected by the algorithm in (5), as described in Section III.B. The final position is determined by summing the current value with the preceding one, as in (6), with the P_k value being inputted into the inverse kinematics of the robot to calculate the lengths of the prismatic links.

$$\Delta P = Select(\Delta P^V, \Delta P^F) \tag{5}$$

$$P_k = P_{k\ 1} + \Delta P \tag{6}$$

IV. Experiments and Results

To assess the effectiveness of our proposed hybrid vision/force control strategy, this study embarked on a comprehensive series of experiments. The goal was to validate the strategy's performance in a controlled benchtop environment, offering a direct comparison with established baselines; both manual techniques and other conventional control methods previously documented in the field. To replicate the real-world conditions encountered during pCLE imaging sessions, we designed our experimental setup to emulate the dynamic nature of tissue movements. This setup, detailed in Fig. 6, leveraged linear actuators to simulate the vertical respiratory movements of biological tissues, adopting a sinusoidal motion pattern with a 3-mm amplitude and a 5-s period, in alignment with parameters established in the existing literature [17]. To further enhance the realism of our simulations, an eco-flex phantom augmented with lens tissue served as our biological tissue model, providing a robust platform for testing the precision and adaptability of our hybrid control strategy under conditions that closely mimic those encountered in clinical pCLE imaging scenarios.

A. Implementation

To verify the feasibility of our hybrid vision/force control strategy, an initial experiment was designed where the handheld robotic device was affixed to a benchtop setup, designed to emulate the conditions of pCLE within a simulated respiratory environment. Over the course of 25 s, the system was subjected to the synthetic respiratory movements of a phantom tissue model, with the hybrid control being engaged 8 s into the experiment. This setup allowed for the detailed observation of the lens tissue through our pCLE system, capturing critical data points such as the exerted forces and CR scores throughout the experiment. Fig. 7 shows the outcomes of this trial, presenting a comprehensive view of the pCLE imaging process, CR score fluctuations, and the forces applied during the entirety of the experiment.

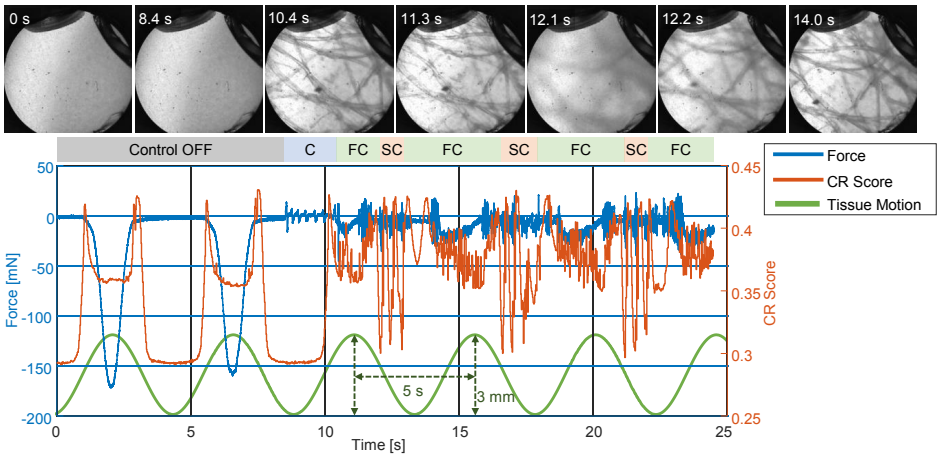

Fig. 7. Force and CR score results from the proposed hybrid vision/force control validation with respiratory motion: C = Calibration, FC = Force Control, SC = Switching Control between force and vision control. The handheld robot is fixed on a benchtop.

TABLE I

CONTACT FORCE AND CR SCORE WITH MEAN AND STANDARD DEVIATIONS DURING HOLD-STILL

Hold-Still		
Control Mode	with Respiration (Force, mN/CR)	without Respiration (Force, mN/CR)
Unaided	-208.0±24.4 /0.40±0.00	-102.7±91.3 /0.41±0.05
Force Control 10 mN	17.3±10.2 /0.43±0.01	-12.9±6.9 /0.38±0.04
Force Control 100 mN	-103.8±10.1 /0.45±0.02	-99.39±16.39 /0.46±0.00
Vision Control	**-7.3±2.8** /0.39±0.04	-12.7±8.4 /0.40±0.02
Hybrid Control	-10.7±7.5 /0.46±0.01	**-9.1±6.4** /0.46±0.04

This step was crucial in demonstrating the hybrid control strategy's capability to adapt and perform under conditions that closely mimic real-world imaging scenarios, validating its potential for enhancing the precision and reliability of pCLE imaging in clinical applications.

At the beginning of the experiment, without the proposed control, the force profile fluctuated with the periodic tissue motion until 8 s, reaching a contact force of 150 mN. Additionally, the CR score periodically dropped when contact with the tissue phantom was lost, then increased upon re-establishing contact, followed by a slight drop due to excessive compression of the tissue by the pCLE probe. The hybrid control was initiated at approximately 8 s, with the image initially appearing out of focus. During this phase, the algorithm transitions into calibration mode upon calculation of the desired CR score value. Subsequently, the initial target force value is determined, activating force control.

It is observed that the force decreases by approximately 1 mN within 0.1 s due to the decaying function of the algorithm 2. Despite the declining force from 10.4 to 11.3 s, the image remains clear. However, due to respiratory motion tissue descent causes the image to become unclear around 12 s. During this period, the SC mode, which alternates between FC and VC, strives to approach the previously set target CR score. Control reverts to FC mode once the target CR score value is attained, thus demonstrating the feasibility of obtaining a clear image with minimal force.

B. Control Evaluation with Baseline

To evaluate the performance of the hybrid vision/force control strategy, it was compared against conventional PID force control methods [5] and vision control methods [11] serving as the baseline. Additionally, experiments were conducted by varying the target force values to -10 mN (FC-10) and -100 mN (FC-100) to assess the quality of the pCLE image according to changes in force. Both vision control and hybrid control were implemented, and contact force and CR scores were compared across all experiments.

Subjects were assigned two primary tasks: maintaining a stationary position (hold still) and performing a scanning motion along a 1 mm straight line on the phantom (scan motion). Tissue motion was continuously enabled and disabled for each task. Subjects were instructed to observe a monitor displaying real-time pCLE images and manipulate the handheld robot to improve image quality and minimize contact force.

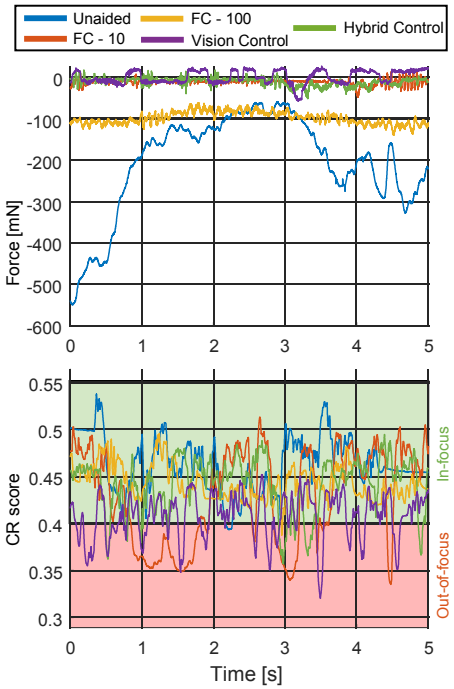

Fig. 8. Force and CR score resulted from control validation with respiratory motion. Baseline control method: simple PID force control. In the CR score graph, the red area and the green area mean the in-focus and out-of-focus of the pCLE image, respectively.

Fig. 8 illustrates the results of tissue scanning with respiratory motion. In the unaided scenario, the CR score exhibits an overall high value; however, subjects without robot assistance struggled to control force precisely due to tissue respiratory motion. In the FC-10 scenario, force is controlled to near -10 mN, yet the CR score briefly falls below 0.4 at 2-3 s and 6-7 s, indicating delayed response to tissue movement. In contrast, the FC-100 scenario demonstrates a consistent force level near -100 mN, with the CR score mostly maintained at 0.4 or higher across most regions. However, the force of 100 mN may induce tissue deformation and dragging, potentially leading to tissue damage. In the vision control scenario, the control loop was configured based on the CR score rather than the force. We observed that the probe failed to respond to the respiratory motion of the tissue at around 3.1 s and measured a contact force of about -50 mN. In the proposed hybrid control scenario, force is regulated near -10 mN. Although the CR score occasionally drops below 0.4, it promptly rebounds above this threshold.

Tables I and II present the statistical analysis results regarding force and CR scores during the hold-still and scan-motion tasks, respectively. Two main notable points emerge here. First, the CR score is calculated to be approximately 0.4

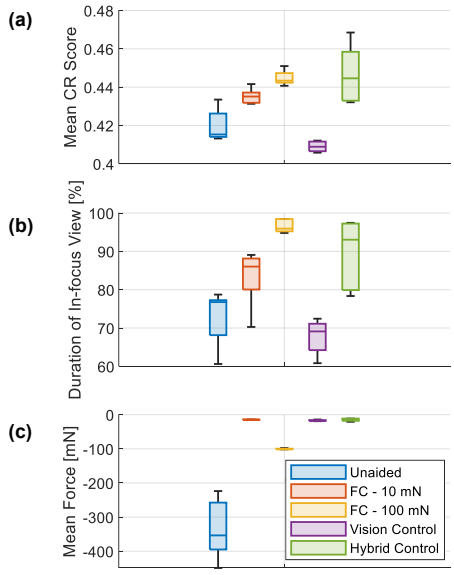

Fig. 9. Statistical analysis results for (a) mean CR score, (b) duration of in-focus view, and (c) mean force. The experiment evaluated performance under conditions where subjects performed scan-motion tasks, repeated five times.

or higher on average in most control modes except for FC-10 and VC mode. This indicates that clear pCLE images were consistently acquired under most control modes. Second, the proposed hybrid control algorithm showed minimal contact force among the results where CR scores were calculated above 0.4. In summary, the findings demonstrate that the hybrid control mode introduced in this study enables the acquisition of clear pCLE images with minimal force.

C. Result Analysis

To validate the statistical analysis results of the proposed algorithm, an experiment was conducted to mimic real surgical conditions. This experiment aimed to gather statistical data by repeatedly scanning tissue under simulated respiratory motion conditions. Subjects were instructed to scan a 1 mm straight line drawn on the tissue. Statistical analysis was conducted based on 5 repeated trials. Fig. 9 illustrates the statistical analysis results, including the mean CR score, duration of in-focus view, and mean force exerted during the scans.

The mean metric of the CR score indicates the consistency of the pCLE image quality. On average, the image showed a CR score of 0.4 or more, which was clearly visible in all control modes. Notably, in the proposed hybrid control mode, the average CR score was the highest, demonstrating that the proposed algorithm yields the clearest image result on average.

The duration of the in-focus view indicates the proportion occupied by the in-focus area for the full time. Under unaided conditions, this ratio was below 80%. For force control, the ratio ranged from 80% to 90% for FC-10 and exceeded 95% for FC-100. In contrast, for vision control, the ratio

TABLE II
CONTACT FORCE AND CR SCORE WITH MEAN AND STANDARD DEVIATIONS DURING SCAN-MOTION

Scan-Motion		
Control Mode	with Respiration (Force, mN/CR)	without Respiration (Force, mN/CR)
Unaided	-204.1±120.2 /0.46±0.03	-176.4±33.2 /0.44±0.02
Force Control 10mN	-11.3±6.0 /0.41±0.05	-14.5±9.1 /0.45±0.04
Force Control 100 mN	-98.1±14.7 /0.44±0.02	-103.1±10.4 /0.46±0.02
Vision Control	-11.5±5.7 /0.40±0.02	-12.2±3.9 /0.40±0.03
Hybrid Control	**-11.0±8.6** /0.44±0.02	**-7.7±5.6** /0.44±0.03

was approximately 70%. Regarding the proposed method, the ratio varied from 80% to 97%. While the proposed method did not achieve the highest ratio, it demonstrated similar performance to FC-10 and FC-100.

The mean force metric indicates the extent of tissue damage. Under unaided conditions, this metric measured approximately 350 mN, indicating a lack of response to respiratory motion and the potential for fatal tissue damage. In contrast, force control methods FC-10 and FC-100 achieved force levels of 10 mN and 100 mN, respectively. Generally, force control methods showed precise responses to respiratory motion, although FC-100 still risked tissue dragging and deformation. Both vision control and hybrid control demonstrated force levels comparable to FC-10. Based on a combination of the three metrics above, the proposed method achieved pCLE image quality similar to FC-100 with minimal force comparable to FC-10, as evidenced statistically.

V. DISCUSSION

Our findings highlight the potential for developing semi-autonomous robotic systems, such as handheld devices and cooperative robots used for pCLE imaging. In this study, we introduced a novel hybrid vision/force control strategy for pCLE, demonstrating its utility in enhancing the acquisition of consistent, high-quality pCLE images with minimal force application. This strategy integrates a semiautonomous surgical robot that significantly reduces manual adjustments by surgeons, automatically adapting target values for both vision and force controls. This approach diverges from traditional feedback control methods, which often depend on static target settings for force or image contrast, and instead focuses on dynamically balancing force and visual feedback to maintain minimal probe-tissue interaction. This is crucial for preventing tissue deformation and avoiding the blurring effects often seen with manual or less adaptive systems. Our experimental results confirm that our hybrid control strategy successfully mitigates these challenges, suggesting a promising direction for the development of more intelligent and sensitive robotic systems in medical imaging.

A limitation of this proposed method is the discontinuity of switching control. Given that switching control

involves different physical quantities, instability may arise when transitioning between modes. As observed in Fig. 9, the hybrid control exhibits higher standard deviation values, such as mean CR score, duration of in-focus view, and mean force, compared to conventional force control methods. Therefore, future work will focus on implementing optimal control strategies to minimize contact force and maximize the CR score. Additionally, in terms of application, we aim to address the narrow FoV limitation of pCLE by incorporating image stitching or mosaicking functions [18].

REFERENCES

[1] S. Zuo and G.-Z. Yang, "Endomicroscopy for computer and robot assisted intervention," *IEEE Rev. Biomed. Eng*., vol. 10, pp. 12–25, 2017.

[2] A. Perperidis, K. Dhaliwal, S. McLaughlin, and T. Vercauteren, "Image computing for fibre-bundle endomicroscopy: A review," *Med. image analysis*, vol. 62, p. 101620, 2020.

[3] N. D. Pilonis, W. Januszewicz, and M. di Pietro, "Confocal laser endomicroscopy in gastro-intestinal endoscopy: technical aspects and clinical applications," *Transl. Gastroenterol. Hepatol*., vol. 7, 2022.

[4] T. Vercauteren, A. Perchant, G. Malandain, X. Pennec, and N. Ayache, "Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy," *Med. Image Anal*., vol. 10, no. 5, pp. 673–692, 2006.

[5] W. T. Latt, R. C. Newton, M. Visentini-Scarzanella, C. J. Payne, D. P. Noonan, J. Shang, and G.-Z. Yang, "A hand-held instrument to maintain steady tissue contact during probe-based confocal laser endomicroscopy," *IEEE Trans. Biomed. Eng*., vol. 58, no. 9, pp. 2694–2703, 2011.

[6] W. T. Latt, T. P. Chang, A. Di Marco, P. Pratt, K.-W. Kwok, J. Clark, and G.-Z. Yang, "A hand-held instrument for in vivo probe-based confocal laser endomicroscopy during minimally invasive surgery," in *Proc. 2012 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, IEEE, 2012, pp. 1982–1987.

[7] P. Wisanuvej, P. Giataganas, K. Leibrandt, J. Liu, M. Hughes, and G.-Z. Yang, "Three-dimensional robotic-assisted endomicroscopy with a force adaptive robotic arm," in *Proc. 2017 IEEE Int. Conf. Robot. Autom. (ICRA)*, IEEE, 2017, pp. 2379–2384.

[8] I. Choi, E. Kim, M.-T. Lim, and S. Yang, "Contact force control during soft tissue interaction using handheld robot," in *Proc. 19th Int. Conf. Ubiquitous Robot*, 2022, pp. 149–152.

[9] R. Newton, D. Noonan, C. Payne, J. Andreyev, A. Di Marco, M. V. Scarzanella, A. Darzi, and G. Yang, "Probe tip contact force and bowel distension affect crypt morphology during confocal endomicroscopy," *Gut*, vol. 60, no. Suppl 1, pp. A12–A13, 2011.

[10] P. Giataganas, M. Hughes, C. J. Payne, P. Wisanuvej, B. Temelkuran, and G.-Z. Yang, "Intraoperative robotic-assisted large-area high-speed microscopic imaging and intervention," *IEEE Trans. Biomed. Eng*., vol. 66, no. 1, pp. 208–216, 2018.

[11] P. Triantafyllou, P. Wisanuvej, S. Giannarou, J. Liu, and G.-Z. Yang, "A framework for sensorless tissue motion tracking in robotic endomicroscopy scanning," in *Proc. 2018 IEEE Int. Conf. Robot. Autom.*, 2018, pp. 2694–2699.

[12] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in *Proc. Hum. Vis. Electron. Imaging XII*, vol. 6492, SPIE, 2007, pp. 196–206.

[13] Z. Li, M. Shahbazi, N. Patel, E. O'Sullivan, H. Zhang, K. Vyas, P. Chalasani, A. Deguet, P. L. Gehlbach, I. Iordachita, *et al.*, "Hybrid robot-assisted frameworks for endomicroscopy scanning in retinal surgeries," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 2, pp. 176–187, 2020.

[14] E. Kim, I. Choi, and S. Yang, "Design and control of fully handheld microsurgical robot for active tremor cancellation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2021, pp. 12289–12295.

[15] S. Tatinati, K. C. Veluvolu, and W. T. Ang, "Multistep prediction of physiological tremor based on machine learning for robotics assisted microsurgery," *IEEE transactions cybernetics*, vol. 45, no. 2, pp. 328–339, 2014.

[16] E. Kim, S. Kim, M. Choi, T. Seo, and S. Yang, "Honeycomb artifact removal using convolutional neural network for fiber bundle imaging," *Sensors*, vol. 23, no. 1, p. 333, 2022.

[17] H. Wang, D. Kong, and S. Zuo, "A miniature robotic-assisted tool for large area endomicroscopy scanning," *J. Mech. Robot*., vol. 12, no. 3, p. 031012, 2020.

[18] J. Kim, H. Lee, S.-R. Oh, and S. Yang, "Real-time endomicroscopic image mosaicking with an ekf-based sensor fusion approach," in *Proc. 2023 45th Annu. Int. Conf. IEEE Eng. Med. & Biol. Soc. (EMBC)*, IEEE, 2023, pp. 1–5.