IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Real-Time SLAM-Guided Closed-Loop Photodynamic Therapy with Pixel-Accurate Light-Dose Control

Hyesung Lee and Sungwook Yang, *Member, IEEE*

***Abstract*— Precise light-dose delivery is essential for photodynamic therapy (PDT), yet current handheld systems remain operator-dependent and lose accuracy under motion. We present a SLAM-guided, closed-loop control framework that enables co-temporal and co-spatial photodynamic diagnosis (PDD) and PDT with a single handheld endomicroscopic probe, while enforcing pixel-level dose control. The probe integrates a fiber bundle that shares a common optical path for both PDD and PDT and is paired with a digital micromirror device (DMD) for μm-scale pattern projection. An extended Kalman filter fuses optical-tracking measurements with texture-limited endomicroscopic images at 30 Hz, providing robust six-degree-of-freedom pose estimates that expand the probe's effective field of view and drive real-time pattern updates. A dose-map SLAM algorithm accumulates light dose over the reconstructed lesion surface during handheld scanning, while pixel-level dose control is enforced by referencing previously accumulated light at each location. Quantitative evaluation shows a spatial registration error between diagnostic and therapeutic systems within 5.2 μm. Experiments on fluorescence phantoms achieved sub-millimeter localization accuracy (0.3 mm RMSE), significantly outperforming vision-only and tracker-only baselines. Finally, tests on targets with quadrant-specific dose limits confirmed SLAM-based dose control, achieving dose uniformity within ± 0.186 mJ/cm^2 across millimeter-scale regions.**

I. INTRODUCTION

Photodynamic therapy (PDT) is a light-activated treatment in which photosensitizers (PSs) are excited by specific wavelengths of light to produce cytotoxic reactions that selectively destroy tumor tissue. Its therapeutic effectiveness and tumor selectivity are well documented [1], [2]; however, the limited specificity of current PSs can also damage surrounding healthy tissue [3], posing critical challenges in applications such as brain-tumor treatment, where preserving normal tissue is essential.

Photodynamic diagnosis (PDD) mitigates this risk by exciting PSs with a secondary wavelength, allowing surgeons to visualize tumor margins in real time [4]–[6]. Yet today's handheld PDT tools remain operator dependent: accurate light delivery over a moving, irregular surface demands simultaneous high-resolution imaging and dose feedback—capabilities rarely integrated in a single closed-loop pipelineT, let alone in a form that executes PDD and PDT simultaneously on the same tissue region.

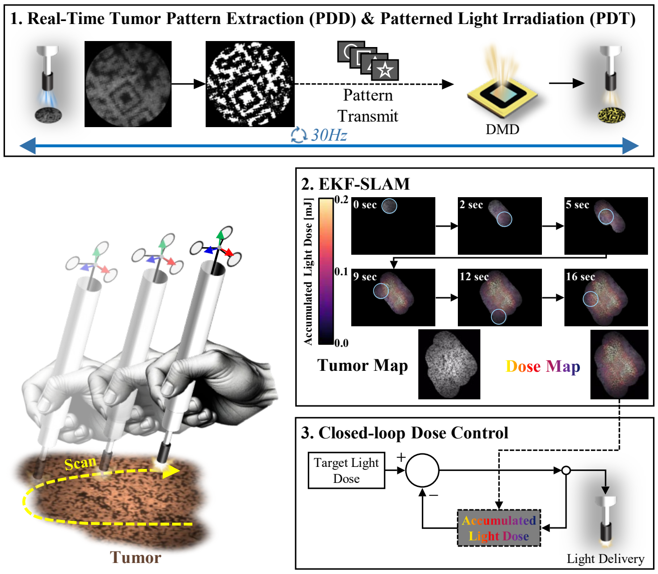

Fig. 1. Proposed integrated system for closed-loop photodynamic therapy.

To satisfy these concurrent requirements, we turn to the digital micromirror device (DMD). This integration enables dynamic spatial modulation based on real-time feedback, allowing the system to project adaptive light patterns that conform to the lesion's morphology to spare healthy tissue. Although DMDs can project μm-scale, arbitrarily shaped light patterns and have improved accuracy in several *open-loop* biomedical-optics demonstrations [7]–[9], they have not yet been embedded in a sensor-guided *closed-loop* pipeline for PDT. When a DMD is coupled to a high-magnification fiber endoscope, two additional bottlenecks emerge: 1) the diagnostic and therapeutic channels must share a common coordinate frame; however, existing systems typically rely on two independent probes or manual mode switching, making them susceptible to patient or hand motion [10], [11]; 2) light-dose uniformity must be verified over the entire lesion, yet most studies monitor dose at only a single point, providing no spatial feedback [12].

Achieving such uniformity is critical, as non-uniformity may compromise therapeutic efficacy or overheat healthy tissue [13], [14]. Recent advances have addressed this challenge on a macro-scale: systems utilizing 3D cameras with spatially-addressable LED arrays dynamically plan light delivery for complex surfaces [15], [16], or employ optimized source profiles to create uniform fields in non-spherical internal organs [17]. While these approaches effectively solve

---

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (RS-2024-00422069, 2019M3A9E2061784), and also by Korea Insitute of Science and Technology (KIST) Institutional Program.

H. Lee and S. Yang are with the Center for Humanoid Research, Korea Institute of Science and Technology (KIST), Seoul 02792, South Korea. H. Lee is also with Korea University, Seoul 02841, South Korea `hs981002@kist.re.kr, swyang@kist.re.kr`

area-level uniformity, their macro-scale control is limited given that PDT is a cellular-level response; furthermore, they do not address the dynamic challenges of high-speed endomicroscopic scanning where pixel-level precision is critical. To widen the treatment field, handheld probes were introduced for freehand scanning (e.g., for photoacoustic-guided theranostics) [18], [19]. However, these platforms still operate in an open-loop manner and lack a global map of deposited light dose. Other studies enhanced tip localization using ultrasound, electromagnetic, or multicore optical trackers fused with Kalman filters [20]–[22]; however, their focus was on pose accuracy rather than coupling localization with pixel-level dose control. Likewise, mosaicking and stitching methods have been applied to fluorescence images to visualize extended tumor areas [23]–[26], but—being offline and open-loop—they do not provide real-time guidance for dose delivery.

To address these gaps, we present the first *single-probe platform that enables simultaneous and co-spatial PDD–PDT* under closed-loop dose control. A 30 Hz EKF-SLAM [27], [28] loop fuses optical tracking and endomicroscopic data to steer the DMD in real time, while a dose-map module accumulates deposited lightdose and enforces pixel-level uniformity over millimeter-scale regions. Unlike prior open-loop or offline methods, the proposed framework is fully closed-loop and hardware-agnostic. An overview of the proposed system is illustrated in Fig. 1.

The main contributions are as follows: 1) a probe-centered, trigger-synchronized architecture that unifies imaging, localization, and DMD projection; 2) a sensor-fusion EKF-SLAM method that achieves sub-millimeter six-DoF tracking in texture-limited fluorescence scenes; and 3) a pixel-accurate closed-loop dose-control algorithm.

The rest of this paper is organized as follows. Section II details the overall system design and SLAM-based control framework; Section III reports experimental validation and quantitative results; and Section IV discusses limitations and future directions.

II. METHODOLOGY

A. *Overall System Configuration*

The proposed system is illustrated in Fig. 2. The handheld endomicroscopic probe contains a fiber bundle (FIGH-30-650S, Fujikura, Japan) and a 3 mm-diameter aspheric doublet lens (#49270, Edmund Optics, USA) that share a common optical path for both photodiagnosis and phototherapy.

Therapeutic or diagnostic light exits the probe tip, while tumor fluorescence returns through the same tip, propagates along the fiber bundle, and is magnified by a 10× objective lens (RMS10X, 0.25 NA, 10.6 mm WD, Olympus, Japan). The image is then projected onto an sCMOS camera (Zyla 4.2 PLUS, Oxford Instruments, UK) at 2048 × 2048 resolution and streamed to the processing PC at 30 Hz via USB 3.0.

For illumination, an LED driver (LEDD1B, Thorlabs, USA) provides trigger control to a 1 W, 470 nm LED (M470L5) for diagnosis and a 1 W, 595 nm LED (M595L4) for therapy. A six-degrees-of-freedom (6-DoF) optical tracking system (fusionTrack 250, Atracsys, Switzerland) reports the pose of a passive fiducial marker (Navex™ Shield Geometry 1) at 120 Hz over TCP/IP, enabling precise probe localization.

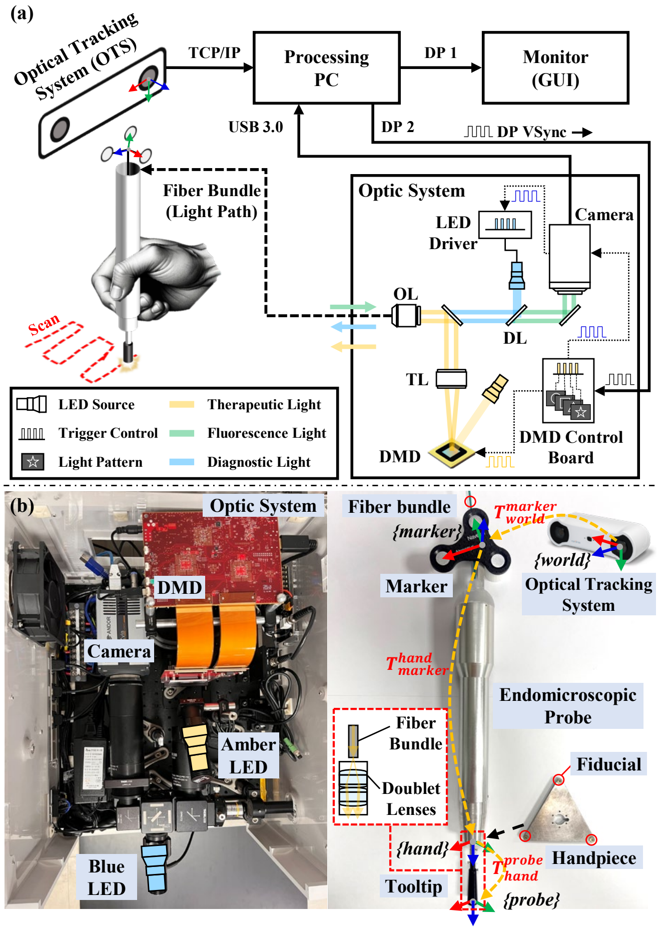

Fig. 2. Hardware configuration of the proposed system. (a) Trigger-synchronized system diagram showing camera, LEDs, DMD, and optical tracker. (b) Optical layout and handheld endomicroscopic probe with their respective coordinate frames.

Patterned light is generated by a digital micromirror device (DMD, DLPLCR50XEVM, Texas Instruments, USA) driven by a control board (DLPCRC900DEVM). The board appears to the PC as a secondary monitor over DisplayPort (DP); pattern frames sent through this link are reflected by the DMD, relayed through a tube lens (TTL200-A, Thorlabs), and delivered back to the probe tip via the fiber bundle.

The PC synchronizes the camera, LEDs, DMD, and tracker via hardware triggers, processes incoming data, and presents a GUI for real-time visualization and control. Details of the trigger sequence and timing are discussed in Section II-B.

B. *Real-Time Framework for Simultaneous Photodynamic Diagnosis and Therapy (PDD/PDT)*

The real-time control framework for simultaneous PDD and PDT is illustrated in Fig. 3. Fig. 3(a) shows a timing di-

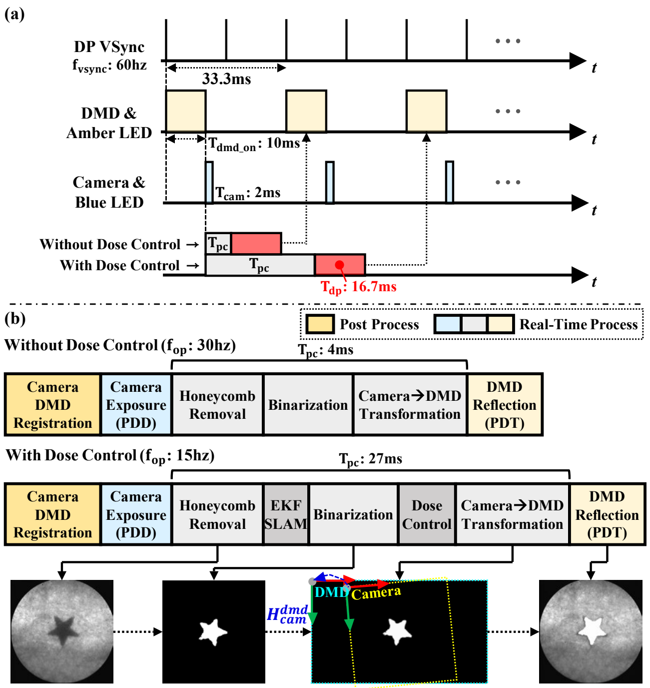

Fig. 3. Real-time control framework for simultaneous PDD and PDT. (a) Timing diagram showing LED, DMD, and camera synchronization via the DisplayPort (DP) VSync signal. (b) The open-loop pipeline without dose control operates at 30 Hz (top), while the closed-loop pipeline with EKF-SLAM and dose feedback runs at 15 Hz (bottom).

agram in which the LED sources, DMD, and camera are synchronized via the DisplayPort (DP) vertical-synchronization (VSync) signal.

A single PDD–PDT cycle proceeds as follows: (i) the camera acquires a fluorescence frame; (ii) the tumor pattern is extracted on the PC and sent to the DMD over the DP link; and (iii) the DMD projects the corresponding patterned therapeutic light through the probe. The timing variables are defined as T_{cam} (camera exposure), $T_{\text{dmd\_on}}$ (DMD illumination), T_{pc} (PC processing), and T_{dp} (pattern transfer), yielding a cycle time $T_{\text{proc}} = T_{\text{cam}} + T_{\text{pc}} + T_{\text{dp}} + T_{\text{dmd\_on}}$.

Fig. 3(b) details the processing pipeline. First, honeycomb artifacts caused by the fiber bundle are removed with HAR-CNN [29] (T_{harcnn}). The denoised image is then binarized to isolate the tumor pattern (T_{bin}) and transformed from camera to DMD coordinates (T_{trans}). Thus $T_{\text{pc}} = T_{\text{harcnn}} + T_{\text{bin}} + T_{\text{trans}}$.

The devices are driven by a VSync-aligned trigger sequence. Each trigger window begins with the first VSync edge and ends just before the third; during this window the DMD displays the pattern computed in the previous cycle while the camera acquires a new frame. Immediately after VSync, the amber LED (therapy, 595 nm) is enabled for $T_{\text{dmd\_on}}$, after which the blue LED (diagnosis, 470 nm) is flashed and the camera is triggered. The newly computed pattern is then uploaded via DP and becomes active at the next VSync pulse.

With $f_{\text{vsync}} = 60$ Hz, $T_{\text{dp}} = 16.7$ ms, $T_{\text{dmd\_on}} = 10$ ms, $T_{\text{cam}} = 2$ ms, and $T_{\text{pc}} = 4$ ms, the baseline cycle time is $T_{\text{proc}} = 32.7$ ms, corresponding to an operating frequency $f_{\text{op}} = 30$ Hz.

When the SLAM-based dose-control module is enabled, T_{pc} additionally includes T_{slam} and T_{dose}, giving $T_{\text{pc}} = 27$ ms and $T_{\text{proc}} = 55.7$ ms. Because the pattern arrives one VSync late, the effective operating frequency halves to $f_{\text{op}} = 15$ Hz.

C. Camera–DMD Registration

Accurate patterned-light therapy requires the therapeutic (DMD) and diagnostic (camera) coordinate frames to be aligned, so that projected pixels coincide with the intended treatment sites. To achieve this, we calibrate the two channels by using the DMD itself as the pattern generator. The checkerboard patterns described below are generated digitally on the DMD pixel array and then projected by reflecting the system's illumination source onto a target plane, where they are captured by the camera. We therefore adopt a fast coarse-to-fine calibration strategy.

Coarse stage. A 4×4 checkerboard, projected by the DMD, is swept across the entire DMD field of view—left-to-right and bottom-to-top—while the camera records each position. The scan locates an approximate overlap region and provides an initial homography that maps DMD pixels to camera pixels.

Fine stage. At the estimated overlap, the DMD projects a denser 8×8 checkerboard. Corner points are detected with sub-pixel accuracy in the camera image, and a refined planar homography is computed by least-squares fitting. This step reduces the residual reprojection error to within a few micrometers on the sample plane, ensuring that therapeutic patterns are delivered precisely to the target.

The two-step procedure shortens calibration time while achieving the pixel-level alignment required for subsequent closed-loop dose control.

D. EKF-Based Sensor-Fusion SLAM

Accurate tracking of the accumulated light dose at every tissue location demands sub-millimetre knowledge of the probe tip's pose. We therefore fuse two complementary sensing streams in an extended Kalman filter (EKF): (i) an optical-tracking system (OTS) that delivers an absolute six-DoF pose at 120 Hz but can be occluded, and (ii) image-based visual odometry derived from consecutive endomicroscopic frames, which is immune to external occlusion yet drifts on texture-poor tissue. By coupling the global stability of the OTS with the local, high-frequency image cues, the EKF provides a robust, drift-free pose estimate that underpins pixel-accurate, closed-loop dose control.

Coordinate initialization. Fig. 2(b) shows the frames. Pivot calibration gives the marker-to-handpiece transform $T^{\text{marker}}_{\text{hand}}$; the handpiece-to-probe transform $T^{\text{hand}}_{\text{probe}}$ is fixed by design. With the OTS pose $T^{\text{marker}}_{\text{world}}$, the probe pose is

$$T^{\text{probe}}_{\text{world}} = T^{\text{probe}}_{\text{hand}}\, T^{\text{hand}}_{\text{marker}}\, T^{\text{marker}}_{\text{world}}. \tag{1}$$

State vector. The EKF state comprises position $\mathbf{p}_t \in \mathbb{R}^3$, linear velocity $\mathbf{v}_t \in \mathbb{R}^3$, orientation quaternion $\mathbf{q}_t \in \mathbb{R}^4$, and angular velocity $\boldsymbol{\omega}_t \in \mathbb{R}^3$:

$$\mathbf{x}_t = \left[\mathbf{p}_t^\top \quad \mathbf{v}_t^\top \quad \mathbf{q}_t^\top \quad \boldsymbol{\omega}_t^\top\right]^\top \in \mathbb{R}^{13}. \tag{2}$$

Prediction. Assuming constant body-frame velocity over Δt,

$$\hat{\mathbf{p}}_{t+1|t} = \mathbf{p}_t + \mathbf{R}_t \mathbf{v}_t \Delta t, \tag{3}$$

$$\hat{\mathbf{q}}_{t+1|t} = \mathbf{q}_t \oplus \tfrac{1}{2}\, \mathbf{q}_t \otimes [0, \boldsymbol{\omega}_t]^\top \Delta t, \tag{4}$$

$[0, \boldsymbol{\omega}_t]$ denotes the pure (zero-scalar) quaternion built from the body-frame angular velocity $\boldsymbol{\omega}_t \in \mathbb{R}^3$; this embedding is required by quaternion kinematics, i.e., $\dot{\mathbf{q}} = \frac{1}{2}\, \mathbf{q} \otimes [0, \boldsymbol{\omega}_t]$.

where $\mathbf{R}_t$ is the rotation matrix of $\mathbf{q}_t$, $\oplus$ denotes quaternion integration, and $\otimes$ is quaternion multiplication. Collecting the terms,

$$\hat{\mathbf{x}}_{t+1|t} = f(\mathbf{x}_{t|t}) = \begin{bmatrix} \mathbf{p}_t + \mathbf{R}_t \mathbf{v}_t \Delta t \\ \mathbf{v}_t \\ \mathbf{q}_t \oplus \frac{1}{2}\, \mathbf{q}_t \otimes [0, \boldsymbol{\omega}_t]^\top \Delta t \\ \boldsymbol{\omega}_t \end{bmatrix}. \tag{5}$$

Measurement fusion and update. Consecutive-frame Scale-invariant feature transform (SIFT) matches yield an in-plane displacement and rotation, which we convert—via pixel size and frame rate—into linear and angular velocities at the probe tip. Under the 2-D scanning assumption, the image-derived motion is modeled as $\mathbf{v}_t = [v_x, v_y, 0]^\top$ and $\boldsymbol{\omega}_t = [0, 0, \omega_z]^\top$, i.e., $v_z = 0$ and $\omega_x = \omega_y = 0$. These terms are fused together with the OTS pose $(\mathbf{p}_t, \mathbf{q}_t)$ in a single stacked measurement and expressed in the probe-tip frame to ensure consistency between modalities. We retain adaptive covariance inflation to down-weight image terms when the inlier count is low and to down-weight OTS pose during partial/complete occlusions. A standard EKF innovation, Kalman gain, and state update complete the cycle.

Noise and adaptive covariance. The process noise $\mathbf{Q}$ and the nominal measurement noise $\mathbf{R}$ are initialized from variance measured during stationary hold. To handle sensing dropouts, we apply adaptive covariance inflation: when the number of inlier SIFT matches falls below a threshold, we increase the variances of $\mathbf{v}_t$ and $\boldsymbol{\omega}_t$; when the OTS is partially/fully occluded, we increase the pose variances accordingly. This makes the filter automatically down-weight unreliable channels while maintaining consistency.

Mapping and localization. Each registered frame is down-scaled to 30% and mosaicked at its world pose, producing a tumor map suitable for real-time dose feedback. This fusion keeps the RMS pose error below 0.3 mm even on texture-poor tissue and supports drift-free 30 Hz mosaics, satisfying the accuracy required for closed-loop, pixel-level dose control.

E. SLAM-Based Closed-Loop Light-Dose Control

Accurate PDT requires uniform light dosing across the lesion, even during probe motion. Fig. 4(a) summarizes our closed-loop algorithm, which achieves this by fusing the EKF pose with live camera frames.

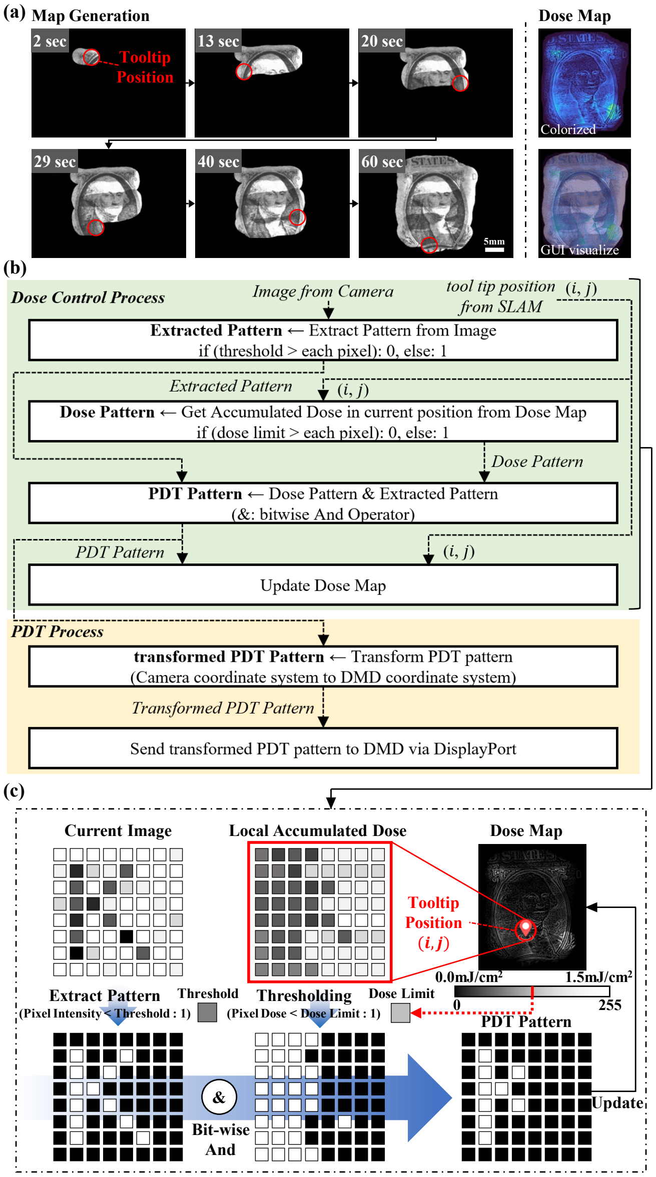

Fig. 4. SLAM-based closed-loop light-dose control. (a) Growth of the tumor map and dose map during a freehand scan; the phantom is a scaled-down one-dollar bill. (b) Block diagram of the pixel-level dose-control pipeline. (c) The detailed process for generating PDT patterns for pixel-level dose control.

Our dosimetric model relies on three key assumptions: 1) the probe tip maintains direct contact (or fixed distance) with the tissue, 2) scanning occurs on a 2D planar surface, and 3) the projected light intensity is spatially uniform. These conditions simplify distance-dependent attenuation to a fixed constant, allowing the dose to be calculated as time-integrated irradiance. We calibrated the total optical power at the probe tip P_{tip} to be 0.06 mW. Given the effective area ($A \approx 0.130\,\mathrm{cm}^2$) of the 4.072 mm diameter tip, the constant irradiance I is derived as:

$$I = \frac{P_{tip}}{A} \approx 0.461\,\mathrm{mW/cm^2}. \tag{6}$$

The algorithm, illustrated in Fig. 4(b) and (c), updates a

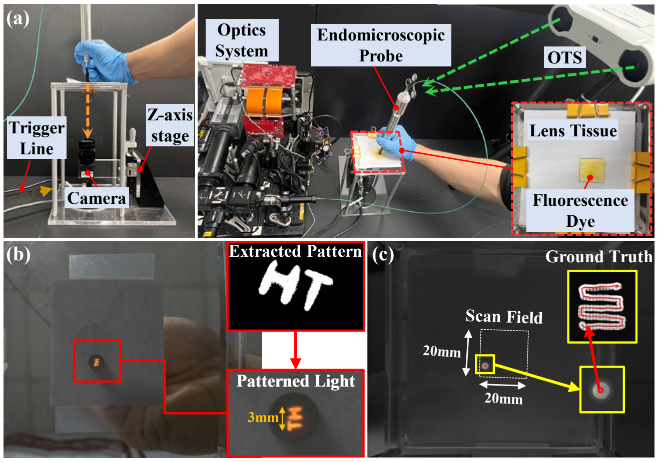

Fig. 5. Validation of timing and pattern fidelity. (a) Experimental apparatus used to synchronize camera exposure with the therapeutic-light pulse. (b) Image of the "HT" test pattern acquired under the synchronized illumination. (c) Ground-truth verification of the projected pattern on the scan field.

global dose map (D) initialized to zero. At each cycle, the camera image is binarized to isolate the fluorescent tumor mask. Using the current EKF pose, the system extracts a local patch from the global map corresponding to the probe's FoV.

This patch is thresholded against a user-defined limit, D_{limit}, identifying pixels eligible for irradiation. A pixel-wise AND operation between the eligible-dose mask and the tumor mask yields the active pattern P_{pdt}:

$$P_{\text{pdt}} = \{(i_1, j_1), (i_2, j_2), \ldots, (i_p, j_p)\}, \tag{7}$$

defining the pixels to be activated. For every pixel (x, y) in the local patch, the accumulated light dose is updated based on the DMD binary state $\delta(x, y)$:

$$D_{new}(x, y) = D_{old}(x, y) + \delta(x, y) \cdot (I \times T_{\text{dmd}}), \tag{8}$$

where T_{dmd} is the frame duration. To visualize dose progress, the map stores a normalized intensity $V(x, y)$:

$$V(x, y) = \frac{D_{new}(x, y)}{D_{limit}} \times 255. \tag{9}$$

The map grows monotonically from 0 to 255, providing a direct visual readout. Pixels in P_{pdt} are mapped to DMD coordinates and transmitted via DisplayPort at the next VSync edge, ensuring real-time, uniform dosing over the target.

III. EXPERIMENTS AND RESULTS

To evaluate the system's overall performance, we developed the experimental setup shown in Fig. 5(a). This system synchronizes the camera exposure with the DMD trigger to ensure that therapeutic illumination is captured only during the intended period. This allows for accurate verification that the projected pattern, such as a planned 'HT' test image, agrees with the intended light dose and establishes the ground truth for the user's scan trajectory in practical scenarios.

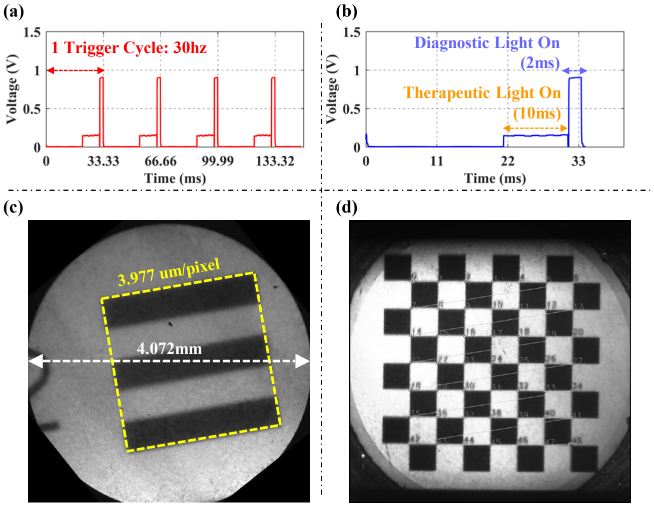

Fig. 6. (a) Photometer and DAQ logging of raw voltage values over time (b) Single system cycle validating synchronization of diagnostic and therapeutic light sources. (c) Resolution measurement of the endomicroscopic probe using USAF 1951 Group 0 Element 2. (d) Corners detection for fine registration.

A. Validation of Temporal Concurrency of Diagnosis and Therapy

To validate the real-time diagnosis and therapy performance of the system, a photometer (S132C, Thorlabs Inc., USA) and a DAQ (Data Acquisition System) were used to measure the raw light power values over time. Before starting the experiment, the power of the therapeutic light was set lower than that of the diagnostic light to facilitate easy differentiation. Fig. 6(a) shows the logging results over 4 trigger cycles, indicating that the system can perform single diagnosis and therapy cycle at a frequency of 30 Hz. Fig. 6(b) illustrates a single cycle of the system. The waveforms corresponding to the therapeutic and diagnostic lights were in the high state for 10 ms and 2 ms, respectively. This confirms that the diagnostic and therapeutic lights are accurately synchronized with each triggering control device.

B. Evaluation of Spatial Registration Error between Camera and DMD

To assess the resolution of the endomicroscopic probe, we employed the 1951 standard resolution chart (see Fig. 6(c)). The probe demonstrated a resolution of 3.977 µm per pixel for a 1024 × 1024 image, with a FoV of 4.072 mm. To further evaluate the spatial alignment between the diagnostic and therapeutic coordinate system, we repeated the registration process between the camera and the DMD 20 times. During each alignment, as illustrated in Fig. 6(d), a fine checkerboard pattern was projected onto the camera via the DMD to calculate the fiducial registration error (FRE) at the checkerboard corners, defined as follows:

$$FRE = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\left\| H_{\text{cam}}^{dmd} \cdot p_i^{\text{cam}} - q_i^{dmd} \right\|^2} \tag{10}$$

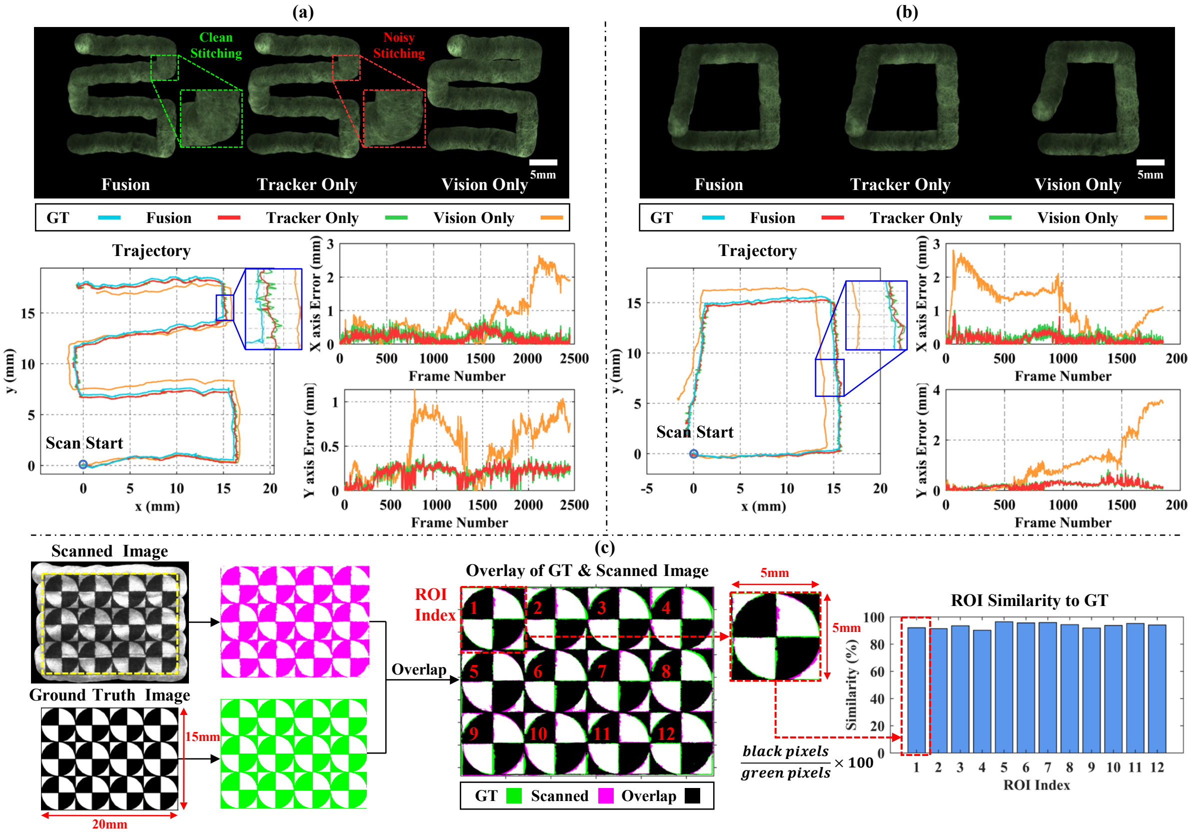

Fig. 7. (a) Localization accuracy for a curved path using Sensor Fusion, Tracker Only, and Vision Only methods, with corresponding errors in the x and y axes. (b) Localization accuracy for a rectangular path using the same methods. (c) Map construction performance evaluated by comparing similarity with the ground truth image.

n, H_{cam}^{dmd}, p_i^{cam}, q_i^{dmd} represent the number of checkerboard corners, the transformation from the camera to the DMD, the coordinates of the corners recognized by the camera, and the ground truth coordinates of the checkerboard corners on the DMD, respectively.

The average FRE across 20 trials was found to be $5.222 \pm 2.215\,\mu$m, scaled from pixels to millimeters using the previously determined resolution. Considering that the average human cell size is approximately $10\,\mu$m, these results indicate that the proposed system offers spatial precision adequate for cellular-level diagnosis and treatment. This high level of precision enables more targeted and effective tumor treatment, aligning diagnostic and therapeutic coordinate system with exceptional accuracy.

C. EKF-based SLAM performance Evaluation

In this system, localization and mapping performance significantly influence dose control accuracy. To evaluate the performance of EKF-based sensor fusion localization, we compared it with vision-only and tracker-only methods using two trajectory patterns: a curved path and a rectangular path. These two trajectories were chosen to probe complementary aspects of the SLAM performance. The curved path contains continuous changes in the velocity direction and is used to assess whether the sensor-fusion scheme can maintain accurate localization under realistic, non-linear handheld scanning motions. In contrast, the squared path forms a closed loop and is therefore suited to evaluating long-term drift and loop-closure consistency. We scanned the lens tissue coated with fluorescence dye to enable fluorescence imaging, as shown in Fig. 5(a). Fig. 7(a) and (b) present the mosaicked maps, trajectory comparisons with the ground truth, and the x and y axis errors for each method. We calculated the Euclidean distance of the root-mean-square errors (RMSEs) relative to the ground truth to quantitatively assess each approach. The sensor fusion method achieved RMSEs of 0.306 mm and 0.300 mm for the curved and rectangular paths, respectively, outperforming the other methods in Table I. These results confirm that sensor fusion can accurately track the tooltip even in feature-poor fluorescence images. In contrast, the vision-only method struggles due to insufficient features, and the tracker-only method, while similar in performance, exhibits increased noise due to the inherent limitations of

TABLE I

COMPARISON OF RMSE VALUES ACROSS FUSION, TRACKER, AND VISION METHODS.

RMSE [mm]	**Fusion**	Tracker Only	Vision Only
Curved Path	**0.306**	0.334	1.203
Rectangular Path	**0.300**	0.320	1.981

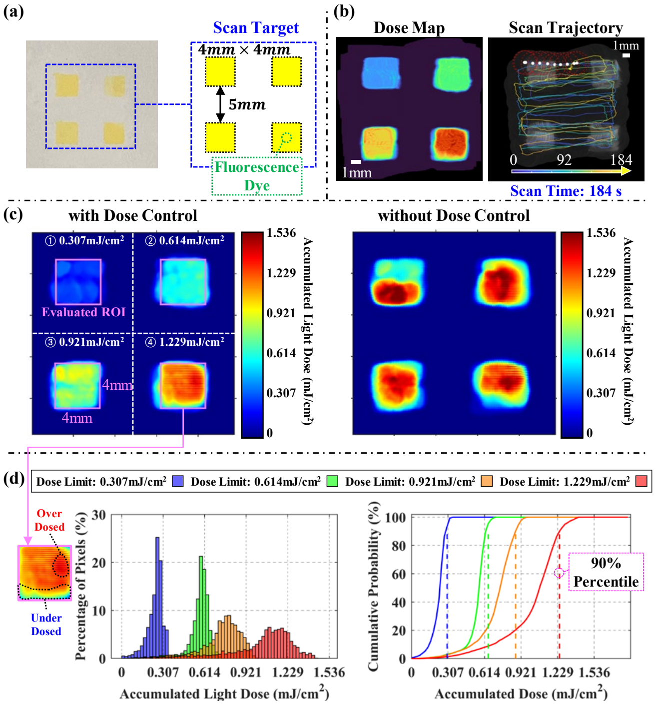

Fig. 8. (a) Target printed with fluorescence dye. (b) Scan Result: Dose map and corresponding scan trajectory. (c) Accumulated light dose visualization with colormap (d) Quantitative evaluation of dose control using histograms

OTS.

To evaluate the performance of map construction, we scanned a 3×4 binary pattern consisting of rectangular and circular shapes, as illustrated in Fig. 7(c). The scanned image was binarized and compared with the ground truth by overlapping it to assess similarity across each region of interest (ROI). The similarity was measured by calculating the overlap percentage of black pixels between the binarized image and the ground truth. The results, shown on the right side of Fig. 7(c), indicate that the system achieved an average similarity rate of 93.72%, demonstrating a high level of precision in map construction.

D. Dose Control Performance Evaluation

We assessed the effectiveness of SLAM-based dose control using a scan target with four 4 mm × 4 mm fluorescent dye squares, as shown in Fig. 8(a). The experiment involved two groups: with dose control and without dose control. In the dose control group, the target was divided into four quadrants, each assigned specific dose limits of 0.307 mJ/cm^2, 0.614 mJ/cm^2, 0.921 mJ/cm^2, and 1.229 mJ/cm^2, with scans conducted over 184 s (Fig. 8(b)). A tooltip irradiance of 0.461 mW/cm^2 required 3.3 s to deliver a 1.536 mJ/cm^2 dose, facilitating evaluation of dose distribution.

We analyzed the accumulated dose within the 4 mm magenta boundary shown on the left side of Fig. 8(c). In the dose control group, the color map corresponding to the set limits demonstrated a controlled dose distribution (Fig. 8(c)). In contrast, the without dose control group showed frequent exceedances of the dose limits across the target, indicating uncontrolled dose delivery.

TABLE II

QUADRANT-WISE RESULTS OF SLAM-BASED LIGHT DOSE CONTROL

Dose Limit $[mJ/cm^2]$	0.307	0.614	0.921	1.229
With SLAM-based Dose Control				
90% Percentile $[mJ/cm^2]$	0.300	0.645	0.114	1.244
Mean Dose $[mJ/cm^2]$	0.249	0.574	0.737	1.043
Overdosed Rate [%]	6.48	23.50	2.21	13.49
Without SLAM-based Dose Control				
90% Percentile $[mJ/cm^2]$	1.498	1.452	1.359	1.359
Mean Dose $[mJ/cm^2]$	0.944	1.080	1.019	0.975
Overdosed Rate [%]	98.61	90.05	63.07	26.21

Table II contains quantitative evaluation metrics for each quadrant of the dose-controlled group. The percentage histogram of the accumulated dose in Fig. 8(d) showed peaks near the set limits, confirming the effectiveness of the dose control. Within the 1.229 mJ/cm^2 dose limit boundary, the dose control group exhibited both underdosed and overdosed pixels: underdosed pixels were located near the edges, while overdosed pixels were found in the interior. A cumulative histogram highlighted deviations from the ideal limits, with the 90th percentile values of 0.300 mJ/cm^2, 0.645 mJ/cm^2, 0.876 mJ/cm^2, and 1.244 mJ/cm^2, indicating that most pixels remained within the limits. The mean values recorded were 0.249 mJ/cm^2, 0.575 mJ/cm^2, 0.737 mJ/cm^2, and 1.043 mJ/cm^2, with deviations of 0.058 mJ/cm^2, 0.039 mJ/cm^2, 0.184 mJ/cm^2, and 0.186 mJ/cm^2, respectively. The rates of overdosed pixels were 6.48%, 23.50%, 2.21%, and 13.49% per quadrant, demonstrating the system's capability for precise dose control.

IV. CONCLUSION

This study presents a novel photo diagnosis and therapy (PDD/PDT) system that employs a real-time framework and a common-path optical configuration, enabling high-magnification lesion observation and adaptive patterned light projection. By integrating SLAM technology within a closed-loop control system, precise light dose delivery across lesions is achieved. However, the system has certain limitations that offer avenues for further enhancement.

Experimental results in Fig. 8(b) revealed instances of overdosed pixels within the designated boundaries, along with shifts in the cumulative histogram compared to ideal distributions. These deviations stem from two primary issues: (1) localization errors due to the limited precision of SLAM, and (2) inconsistencies in thresholding during overlapping scans near boundaries.

A core limitation of the current EKF-based SLAM approach is its planar lesion model, which does not account for tissue 3D structure and deformation during scanning.

Deformation-aware 3D SLAM could improve spatial registration and dose-mapping accuracy, enhancing PDT.

Moreover, detecting the target shape in fluorescence imaging is challenging due to blurred edges and variable intensity, making simple binarization inadequate. Advanced shape extraction algorithms are necessary to enhance morphology extraction and dose selectivity [30], [31]. Additionally, while downscaling maps can reduce computational load, it may lead to a loss of critical detail. Therefore, efficient algorithms that maintain resolution without imposing excessive resource demands are needed.

In summary, this research demonstrates a technological advancement in PDT through the integration of an optical system and a real-time framework for patterned photodynamic diagnosis and therapy, alongside SLAM-based dose control. Nevertheless, challenges related to localization accuracy, segmentation fidelity, and computational efficiency must be addressed. Continued research and optimization are essential to develop more precise and effective light-based therapies.

REFERENCES

[1] D. Bartusik-Aebisher, A. Żołyniak, E. Barnaś, A. Machorowska-Pieniążek, P. Oleś, A. Kawczyk-Krupka, and D. Aebisher, "The use of photodynamic therapy in the treatment of brain tumors—a review of the literature," *Molecules*, vol. 27, no. 20, p. 6847, 2022.

[2] G. Gunaydin, M. E. Gedik, and S. Ayan, "Photodynamic therapy—current limitations and novel approaches," *Front. Chem.*, vol. 9, p. 691697, 2021.

[3] T. C. Pham, V.-N. Nguyen, Y. Choi, S. Lee, and J. Yoon, "Recent strategies to develop innovative photosensitizers for enhanced photodynamic therapy," *Chem. reviews*, vol. 121, no. 21, pp. 13454–13619, 2021.

[4] T. Yamamoto, E. Ishikawa, S. Miki, N. Sakamoto, A. Zaboronok, M. Matsuda, H. Akutsu, K. Nakai, W. Tsuruta, and A. Matsumura, "Photodynamic diagnosis using 5-aminolevulinic acid in 41 biopsies for primary central nervous system lymphoma," *Photochem. photobiology*, vol. 91, no. 6, pp. 1452–1457, 2015.

[5] J. He, L. Yang, W. Yi, W. Fan, Y. Wen, X. Miao, and L. Xiong, "Combination of fluorescence-guided surgery with photodynamic therapy for the treatment of cancer," *Mol. imaging*, vol. 16, p. 1536012117722911, 2017.

[6] R. Rabbani, H. Najafiaghdam, M. Roschelle, E. P. Papageorgiou, B. R. Zhao, M. M. Ghanbari, R. Muller, V. Stojanovic, and M. Anwar, "Towards a wireless image sensor for real-time fluorescence microscopy in cancer therapy," *IEEE Trans. Biomed. Circuits Syst.*, 2024.

[7] C. Gong and D. Mehrl, "Characterization of the digital micromirror devices," *IEEE Trans. Electron Devices*, vol. 61, no. 12, pp. 4210–4215, 2014.

[8] T. Yoon, C.-S. Kim, K. Kim, and J.-r. Choi, "Emerging applications of digital micromirror devices in biophotonic fields," *Opt. & Laser Technol.*, vol. 104, pp. 17–25, 2018.

[9] M. Praeger, J. Scott, J. A. Grant-Jacob, J. Barnsley, M. Niranjan, R. W. Eason, E. Healy, and B. Mills, "Machine learning for automated, targeted, phototherapy," *Opt. Continuum*, vol. 3, no. 5, pp. 738–750, 2024.

[10] B. Pettiford and R. J. Landreneau, "Endobronchial stents and bronchial sparing surgery in the management of lung cancer," *Revista del Instituto Nac. de Enfermedades Respir.*, vol. 20, no. 1, pp. 33–41, 2007.

[11] Y. Hu and K. Masamune, "Flexible laser endoscope for minimally invasive photodynamic diagnosis (pdd) and therapy (pdt) toward efficient tumor removal," *Opt. Express*, vol. 25, no. 14, pp. 16795–16812, 2017.

[12] M. M. Kim and A. Darafsheh, "Light sources and dosimetry techniques for photodynamic therapy," *Photochem. photobiology*, vol. 96, no. 2, pp. 280–294, 2020.

[13] J. F. Algorri, M. Ochoa, P. Roldan-Varona, L. Rodriguez-Cobo, and J. M. López-Higuera, "Light technology for efficient and effective photodynamic therapy: A critical review," *Cancers*, vol. 13, no. 14, p. 3484, 2021.

[14] Y. Wei, J. Song, and Q. Chen, "In vivo detection of chemiluminescence to monitor photodynamic threshold dose for tumor treatment," *Photochem. & Photobiol. Sci.*, vol. 10, no. 6, pp. 1066–1071, 2011.

[15] X. Wang, W.-R. Kang, X.-M. Hu, and Q. Li, "Irradiance uniformity optimization for a photodynamic therapy treatment device with 3d scanner," *J. Biomed. Opt.*, vol. 26, no. 7, pp. 078001–078001, 2021.

[16] X. Wang, T. Jin, J. Xiong, H. Zhao, X. Hu, Q. Li, J. Ren, and Y. Zhao, "Three-dimensional image-guided topical photodynamic therapy system with light dosimetry dynamic planning and monitoring," *Biomed. Opt. Express*, vol. 14, no. 1, pp. 453–466, 2022.

[17] Y. Tan, S. Sun, D. Chen, H. Qiu, J. Zeng, Y. Wang, H. Zhao, and Y. Gu, "Light delivery device modelling for homogenous irradiation distribution in photodynamic therapy of non-spherical hollow organs," *Photodiagnosis Photodyn. Ther.*, vol. 34, p. 102320, 2021.

[18] M. Basij, S. John, D. Bustamante, L. Kabbani, W. Maskoun, and M. Mehrmohammadi, "Integrated ultrasound and photoacoustic-guided laser ablation theranostic endoscopic system," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 1, pp. 67–75, 2022.

[19] S. Liu, X. Feng, H. Jin, R. Zhang, Y. Luo, Z. Zheng, F. Gao, and Y. Zheng, "Handheld photoacoustic imager for theranostics in 3d," *IEEE transactions medical imaging*, vol. 38, no. 9, pp. 2037–2046, 2019.

[20] Y. Cai, R. Li, A. Davoodi, M. Ourak, J. Deprest, and E. Vander Poorten, "Autonomous robotic ultrasound approach for fetoscope tracking by fusing optical and 2d ultrasound data," *IEEE Robot. Autom. Lett.*, 2024.

[21] X. T. Ha, D. Wu, M. Ourak, G. Borghesan, A. Menciassi, and E. Vander Poorten, "Sensor fusion for shape reconstruction using electromagnetic tracking sensors and multi-core optical fiber," *IEEE Robot. Autom. Lett.*, vol. 8, no. 7, pp. 4076–4083, 2023.

[22] Z. Xia, H. Wang, Y. Men, H. Ma, Z. Cao, W. Wang, and Z. Du, "Kalman filter-based em-optical sensor fusion for bone needle position tracking," *IEEE Sensors J.*, 2024.

[23] B. Rosa, B. Dahroug, B. Tamadazte, K. Rabenorosoa, P. Rougeot, N. Andreff, and P. Renaud, "Online robust endomicroscopy video mosaicking using robot prior," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4163–4170, 2018.

[24] L. Zhang, M. Ye, P. Giataganas, M. Hughes, and G.-Z. Yang, "Autonomous scanning for endomicroscopic mosaicing and 3d fusion," in *Proc. 2017 IEEE Int. Conf. Robot. Autom. (ICRA)*, IEEE, 2017, pp. 3587–3593.

[25] S. Zuo, M. Hughes, and G.-Z. Yang, "Flexible robotic scanning device for intraoperative endomicroscopy in mis," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 4, pp. 1728–1735, 2017.

[26] X. Xu, R. Tang, L. Gong, B. Chen, and S. Zuo, "Two dimensional position-based visual servoing for soft tissue endomicroscopy," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 5728–5735, 2021.

[27] R. C. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *The international journal Robot. Res.*, vol. 5, no. 4, pp. 56–68, 1986.

[28] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (slam): Part ii," *IEEE robotics & automation magazine*, vol. 13, no. 3, pp. 108–117, 2006.

[29] 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.

[30] S. G. K. Patro, N. Govil, S. Saxena, B. K. Mishra, A. T. Zamani, A. B. Miled, N. Parveen, H. Elshafie, and M. Hamdan, "Brain tumor classification using an ensemble of deep learning techniques," *IEEE Access*, 2024.

[31] G. Florimbi, H. Fabelo, E. Torti, S. Ortega, M. Marrero-Martin, G. M. Callico, G. Danese, and F. Leporati, "Towards real-time computing of intraoperative hyperspectral imaging for brain cancer detection using multi-gpu platforms," *IEEE Access*, vol. 8, pp. 8485–8501, 2020.