

UC Davis

UC Davis Previously Published Works

Title

FLImBrush: dynamic visualization of intraoperative free-hand fiber-based fluorescence lifetime imaging.

Permalink

<https://escholarship.org/uc/item/5j4233pn>

Journal

Biomedical Optics Express, 11(9)

ISSN

2156-7085

Authors

Marsden, Mark
Fukazawa, Takanori
Deng, Yu-Cheng
[et al.](#)

Publication Date

2020-09-01

DOI

10.1364/boe.398357

Peer reviewed



FLImBrush: dynamic visualization of intraoperative free-hand fiber-based fluorescence lifetime imaging

MARK MARSDEN,^{1,5}  TAKANORI FUKAZAWA,^{1,2,5} YU-CHENG DENG,¹ BRENT W. WEYERS,¹ JULIEN BEC,¹  D. GREGORY FARWELL,^{3,4,6} AND LAURA MARCU^{1,4,7}

¹Department of Biomedical Engineering, University of California, Davis, CA 95616, USA

²Sony Imaging Products and Solutions Inc., Japan

³Department of Otolaryngology, University of California, Davis, CA 95817, USA

⁴Corresponding authors

⁵Equal Contribution

⁶dgfarwell@ucdavis.edu

⁷lmarcu@ucdavis.edu

Abstract: A free-hand scanning approach to medical imaging allows for flexible, lightweight probes to image intricate anatomies for modalities such as fluorescence lifetime imaging (FLIm), optical coherence tomography (OCT) and ultrasound. While very promising, this approach faces several key challenges including tissue motion during imaging, varying lighting conditions in the surgical field, and sparse sampling of the tissue surface. These challenges limit the coregistration accuracy and interpretability of the acquired imaging data. Here we report FLImBrush as a robust method for the localization and visualization of intraoperative free-hand fiber optic fluorescence lifetime imaging (FLIm). FLImBrush builds upon an existing method while employing deep learning-based image segmentation, block-matching based motion correction, and interpolation-based visualization to address the aforementioned challenges. Current results demonstrate that FLImBrush can provide accurate localization of FLIm point-measurements while producing interpretable and complete visualizations of FLIm data acquired from a tissue surface. Each of the main processing steps was shown to be capable of real-time processing (> 30 frames per second), highlighting the feasibility of FLImBrush for intraoperative imaging and surgical guidance. Current findings show the feasibility of integrating FLImBrush into a range of surgical applications including cancer margins assessment during head and neck surgery.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Free-hand scanning approaches to medical imaging enable physicians to have greater control over the imaged region through the use of hand-held imaging probes. This approach has been investigated for a range of imaging modalities including optical coherence tomography (OCT) [1,2], ultrasound [3,4], high resolution microendoscope (HRME) [5] and fluorescence lifetime imaging (FLIm) [6–8]. A free-hand approach has several advantages over whole-field imaging, including the potential to employ small, lightweight probes for the intraoperative imaging of intricate anatomies as well as the flexibility to image a desired region of interest with non-uniform surface topology from a variety of angles. Several challenges are also observed for free-hand scanning implementations of various modalities [1,3,6] including localization and coregistration errors due to tissue motion and sparse sampling resulting in incomplete representations of the tissue region of interest. Overcoming these challenges is key to the development of a clinically feasible method for intraoperative free-hand scanning. This work aims to address these challenges for fiber-based FLIm and to develop an accurate and rapid localization and visualization method for

FLIm data acquired intraoperatively. This study addresses the limitations of a previously reported fiber-based FLIm data augmentation method [9] that employed a continuous wave aiming beam emitted from the fiber optic to register FLIm point-measurement positions before augmenting a white-light image of the surgical field-of-view (FOV) with a transparent overlay representing the acquired spectroscopic data. Label-free, time-resolved autofluorescence techniques such as FLIm have been shown to detect variation in the endogenous molecular composition of tissue (e.g. metabolic co-factors, matrix proteins, porphyrins) [10]. Point-scanning FLIm has shown potential to intraoperatively discriminate between tissue types and pathologies (e.g. cancer) during neurosurgery [7] and head and neck cancer resection procedures [6]. Fiber-based FLIm has also been integrated into transoral robotic surgery (TORS) using the da Vinci Surgical System [8].

While the current method for free-hand FLIm shows great promise for surgical guidance, it faces several key challenges common to free-hand modalities. These challenges can be defined in the context of FLIm as follows: (i) localization errors due to tissue motion over time (e.g. caused by respiration, pulsation from blood flow and camera motion) as well as aiming beam localization errors caused by reflections within the FOV from surgical instruments and smooth tissue surfaces, (ii) sparse sampling of the tissue surface due to various constraints encountered (e.g. limited time for imaging, fast scanning of the tissue surface, difficulties when positioning the FLIm probe due to tortuous anatomies) which lead to an incomplete and difficult to interpret visualization of the targeted region, and (iii) the need for real-time visualization and feedback during intraoperative imaging while addressing the aforementioned challenges.

The overall objective of this study is to develop a refined FLIm localization and visualization method, referred to as FLImBrush, capable of providing dynamic, accurate and interpretable visualization of intraoperative fiber-based FLIm data. To achieve this, the following specific goals were defined: (i) Develop an image analysis method capable of the accurate localization of FLIm aiming beam positions which accounts for tissue motion, changing lighting conditions and the presence of surgical instruments. (ii) Demonstrate an interpolation-based visualization method which produces interpretable, coherent visualizations of sparse FLIm point-measurement data. (iii) Demonstrate the capability of these methods to display FLIm data in real-time (> 30 frames-per-second (FPS)) and thus allow for intraoperative visualization.

2. Method

2.1. FLIm instrument

A custom-built, fiber-based fluorescence lifetime imaging (FLIm) system (described previously in [11]) was used for intraoperative free-hand imaging. Tissue autofluorescence was excited with a 355 nm (< 600 ps FWHM) pulsed laser (micro Q-switched laser, 120 Hz repetition rate, Teem Photonics, France) and delivered through a 365 μm core diameter multimode optical fiber (Thorlabs Inc, numerical aperture 0.22). The same fiber was used to collect tissue autofluorescence. The lateral resolution of the system is determined by the illumination spot size and collection geometry, which is improved as the distance between the probe and tissue decreases. The proximal end of the fiber was coupled to a wavelength selection module (WSM) which features a set of four dichroic mirrors and bandpass filters to separate 355 nm excitation from collected autofluorescence, spectrally resolving the autofluorescence signal (CH1: 390 ± 20 nm; CH2: 470 ± 14 nm; CH3: 542 ± 25 nm; and CH4: 629 ± 26.5 nm). Each spectral band targets the reported autofluorescence emission maxima of the following endogenous fluorophores (CH1: collagen, CH2: nicotinamide adenine dinucleotide (NADH), CH3: flavin adenine dinucleotide (FAD), CH4: porphyrins) [10]. The autofluorescence signal from each spectral band was time-multiplexed onto a single microchannel plate photomultiplier tube (MCP-PMT, R3809U-50, 45ps FWHM, Hamamatsu, Japan), amplified (AM-1607-3000, Miteq Inc., USA), and time-resolved by a high sampling frequency digitizer (12.5 GS/s, 3GHz, 8-bit, 512 Mbytes,

PXIe-5185, National Instruments, Austin, TX, USA). Signal-to-noise ratio (SNR) was calculated separately for each spectral channel for each individual point-measurement acquired during FLIm scanning. Background subtraction was performed using a probe background signal acquired at the start of each imaging session. A Laguerre expansion based deconvolution [12] was performed on the raw decay signal acquired from each channel using the system impulse response function (IRF) before a set of FLIm signal parameters (average fluorescence lifetime, spectral intensity ratio) were calculated for each spectral band.

The FLIm system has been integrated into various surgical workflows for intraoperative imaging [6,7]. The main differences between workflows are the source of white-light video feed observing the surgical FOV and the method used to integrate the fiber probe. The data used for this study was acquired during head and neck cancer resection surgeries, under two distinct experimental situations: (i) the da Vinci Surgical System (both Si and SP models) (Intuitive Surgical Inc.). Both da Vinci models are equipped with an integrated camera for white-light video capture. For the Si model the fiber probe was inserted into a fiber introducer instrument (described in [6]). For the SP model, a 3D printed stainless steel grasper tab was placed on the fiber optic distal end to facilitate integration with the SP models grasping instrument (Maryland or Fenestrated bipolar forceps) for direct manipulation of the fiber for FLIm imaging and (ii) a non-robotic approach which combines a hand-held fiber probe (OmniGuide Laser Handpiece) for imaging and endoscopic camera (Stryker) for video capture. 720p video (1280 x 720) was captured at 30 frames per second from the da Vinci robot and Stryker endoscope using a frame-grabber (Epiphan Systems Inc., Ottawa, ON, Canada). Examples of these experimental setups are presented in Figure 1.

2.2. Review of existing augmentation method

The FLImBrush method was designed to address key limitations of a previously developed FLIm localization and visualization method [8] described in brief below.

2.2.1. Aiming beam localization via color space thresholding

The earlier method performed FLIm point-measurement localization through color space thresholding. Each point-measurement was localized via a 445 nm continuous wave aiming beam (TECBL-50G-440-USB, World Star Tech, Canada). This aiming beam was integrated into the optical path of the systems WSM and delivered to the tissue through the same optical path used to induce tissue autofluorescence [9]. The RF amplifier of the instrument was AC coupled with a cutoff frequency of 10 kHz in order to filter out any signal contribution from the aiming beam and operating room lights as described previously [9]. The aiming beam was then localized within a 2-D white light image of the surgical field captured via an integrated camera. The aiming beam was first segmented by transforming the image to the hue saturation value (HSV) color space (H (0-360), S (0-255), V (0-255)) before thresholding the hue channel between 85 and 255 and the saturation and value channels between 55 and 255. This thresholding step targets saturated blue regions (i.e. the 445 nm aiming beam) within the surgical field, producing a segmentation mask. Morphological erosion and dilation were then performed to remove isolated noise. The centroid was then calculated for all contours in the segmentation mask, with the contour centroid closest to the beam centroid for the previous frame (in Euclidean distance) selected in order to ensure accurate tracking of the beam. Selected contours must have a minimum size > 5 pixels and must be within 50 pixels of the previous frames beam location. If these criteria are not met, then the previous frames beam location was used for the current frame. The FLImBrush method employs the same instrumentation as the method, only altering the image analysis and contour selection steps.

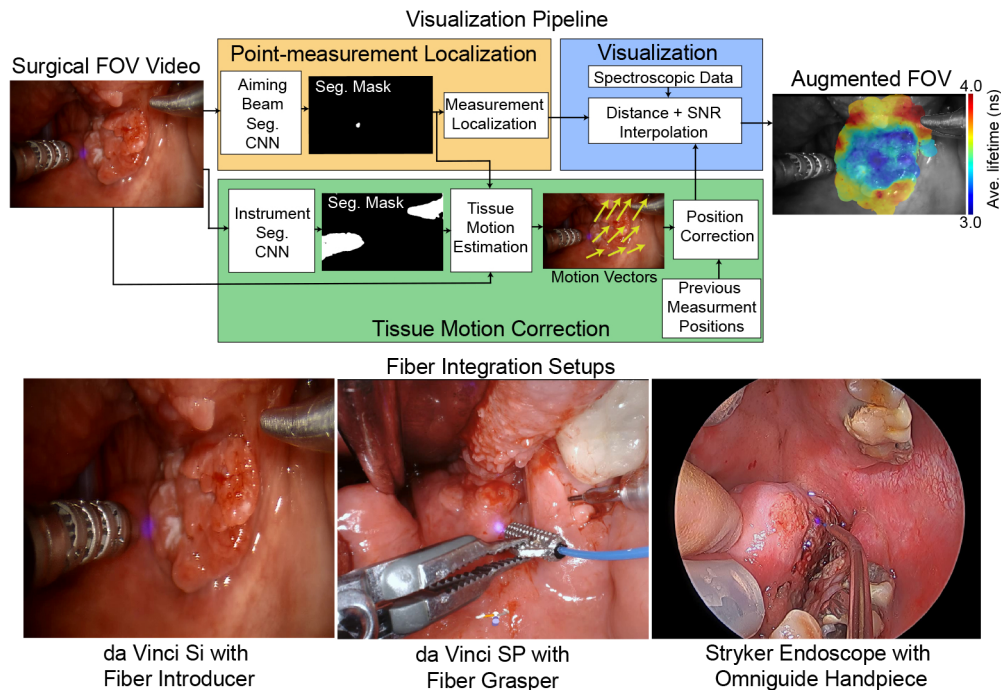


Fig. 1. Top: full processing pipeline of the FLImBrush method for the localization and visualization of free-hand FLIm point measurements. White light images of the surgical field-of-view (FOV) were taken from an integrated camera. Point-measurement localization (orange) was performed by segmenting the aiming beam emitted from the fiber using a U-Net CNN before calculating the centroid. For frames in which the beam was not successfully segmented, a linear interpolation was performed using estimated positions within a 10-frame window. Tissue motion correction (green) was performed by segmenting the surgical instruments using a U-Net CNN in order to exclude these regions from tracking. Motion estimation via an optimized block-matching approach was then applied. Point-measurement position correction was performed on a frame-by-frame basis for previous measurement locations using the estimated motion vectors. Data visualization (blue) was then performed by generating an augmented overlay using measured spectroscopic data (e.g. average fluorescence lifetime for a specific spectral band) for all point-measurements acquired. An optimized distance and SNR-based interpolation method was applied on a frame-by-frame basis to generate the visualization. Bottom: sample frames taken from the distinct fiber integration setups used in this study.

2.2.2. Overlapping discs visualization

Visualization was performed by fitting an ellipse to the segmented beam contour for each point-measurement, applying a chosen color map (e.g. jet) using the measured value for a chosen FLIm parameter (e.g. avg. lifetime) before overlapping and averaging rendered ellipses to produce an overall FLIm map for the scanned tissue region. This FLIm map was then augmented onto the surgical field as a transparent overlay.

2.3. FLImBrush augmentation method

The FLImBrush method builds upon the earlier method and consists of three interdependent processing steps: (i) Convolutional neural network (CNN) based aiming beam localization, (ii) block-matching based tissue motion correction, and (iii) interpolation-based visualization of

point-measurements. Figure 1 presents an overview of the full processing pipeline and how these distinct processing steps are integrated together.

2.3.1. CNN-based aiming beam localization

The U-Net convolutional neural network (CNN) architecture [13] was used to train a segmentation model for improved aiming beam localization. This fully convolutional architecture is highly symmetric in nature, with a high number of feature channels in the latter layers. A series of skip connections from the earlier layers allow for image context to be propagated forward. These improvements allow for precise segmentation to be performed when compared to more conventional architectures [13]. A sigmoid activation function was applied to network output to produce a probability overlay for the occurrence of the aiming beam. The use of a CNN approach allows for local shape, color, and texture features to be utilized. Images were down-sampled to an image size of 640×384 to allow for accelerated beam segmentation, before the output was restored to the original image size for the subsequent processing steps. This image size was chosen as U-Net input images must have a width and height divisible by 64. Scaling and padding was performed to ensure the original image aspect ratio was maintained. A combination of Jaccard index and binary cross entropy loss, given in Eq. (1) and (2), was minimized during model training as described by Shvets et al. [14]. This joint loss function enforces accurate segmentation shape and correct pixel labels during training. H refers to pixel-wise binary cross entropy loss for the n pixels in the image, while y_i and \hat{y}_i refer to the ground truth and predicted value for pixel i .

$$J = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i \hat{y}_i}{y_i + \hat{y}_i - y_i \hat{y}_i} \right) \quad (1)$$

$$L = H - \log J \quad (2)$$

Model optimization was performed for 5000 training iterations using the adagrad optimizer [15] with an L_2 weight decay of $1e-4$ and a batch size of 1. This configuration allowed for gradual model convergence with the employed training data without overfitting. Random horizontal and vertical flips were performed with a probability of 50% during training to increase the diversity of training data. A threshold of 0.5 was applied to the generated probability overlay to produce binary segmentation masks, allowing for weaker candidate locations to also be considered. The CNN segmentation approach was developed using the PyTorch numerical library [16] and the OpenCV image processing library [17].

Following beam segmentation, a similar contour selection approach was used as for the earlier method with some refinements. In the event of an unsuccessful localization, a linear interpolation approach was used within a 10-frame window to retrospectively estimate beam centroid positions for these missing frames. This interpolation was only performed if 5 successful localizations were performed in this 10-frame window, otherwise this point-measurement was not included in the generated visualization. Employing a wider window for linear interpolation can result in the erratic interpolation of aiming beam positions.

2.3.2. Block matching-based tissue motion correction

Tissue motion estimation and aiming beam position correction was performed on a frame-by-frame basis during FLIm acquisition. The target application of FLImBrush is intraoperative scanning of a tissue region prior to excision, resulting in locally correlated motion patterns observed across the tissue surface with gradual motion between frames. Informed by this observation, motion estimation was performed using Adaptive Rood Pattern Search (ARPS) block-matching [18], which allows for efficient and accurate estimation of local motion vectors. Block-matching based motion estimation was selected over feature matching methods [19,20] due to the presence of visually homogeneous tissue regions lacking in strong landmarks. These homogeneous regions

can be more accurately tracked by comparing image patches (i.e. block-matching) rather than individual interest points. One of the major challenges of tissue motion correction that must be accounted for is occlusions due to surgical instruments and the fiber probe itself. An example of this is shown in Figure 2.

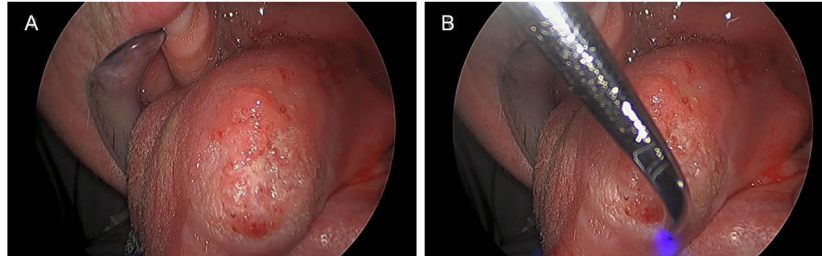


Fig. 2. Example frames taken from a head and neck surgery video in which (a) the FOV is clear (b) the tissue is occluded by a surgical instrument, in this case the fiber probe itself.

An overview of the motion estimation and correction workflow is presented in Figure 3. Each input frame was divided into non-overlapping 32×32 pixels macroblocks (MBs), based on the scale typically observed in intraoperative FLIm scans. Prior to motion estimation for a given frame, two initialization steps must be performed. First, a region-of-interest for motion correction was selected in terms of MBs based on prior aiming beam segmentation locations. This region of interest selection allows for accelerated processing. Second, occluded tissue regions (in terms of MBs) were detected for a given frame using a previously developed surgical instrument segmentation CNN [14] before being excluded from direct motion estimation. This CNN model also employs the U-Net architecture and uses a pre-trained set of model weights [14]. This step prevents motion from surgical instruments affecting motion estimation for the underlying tissue. ARPS motion estimation was then performed following a conversion to grayscale (color data was not used for motion tracking with this method). Mean Absolute Difference (MAD) between MBs was minimized during the matching stage with a search range of 7 pixels. The searching range allows for accelerated processing and was selected based on the observation that $> 99\%$ of inter-frame motion was within 7 pixels for a typical intraoperative FLIm scan. A set of motion vectors were estimated for each frame, allowing for previously estimated aiming beam positions to be refined through a position correction step using these motion vectors.

In the event that a given MB was occluded, motion vectors were interpolated from the 3 nearest non-occluded MBs through a weighted average inversely proportional to the distance between blocks. All 3 MBs used for vector interpolation must fall within a 5-MB radius and not be marked as occluded. If this condition is not met then no motion vector interpolation is performed for a given occluded MB. The rationale for applying motion vector interpolation is that tissue motion within a local region is observed to be largely homogeneous due to tissue stretching and pulsation. With an image width of 1280 pixels employed the 32×32 macroblock size corresponds to 2.5% of the overall image width. When performing motion estimation at this scale it is observed that adjacent macroblocks will have similar motion vectors. The 3 MB constraint ensures highly occluded regions are not interpolated incorrectly, preventing vector interpolation from contributing to overall coregistration error. Calculating local motion vectors across a full video sequence also allows for estimated measurement positions to be translated to any desired reference frame through a vector summation step which can be employed for point-measurement labelling (i.e. cancer vs. healthy) through the coregistration of histopathology with white light images. The developed algorithm was implemented in Matrix Laboratory (MATLAB). Image downsampling was not performed for motion correction to ensure low magnitude motions were accurately estimated.

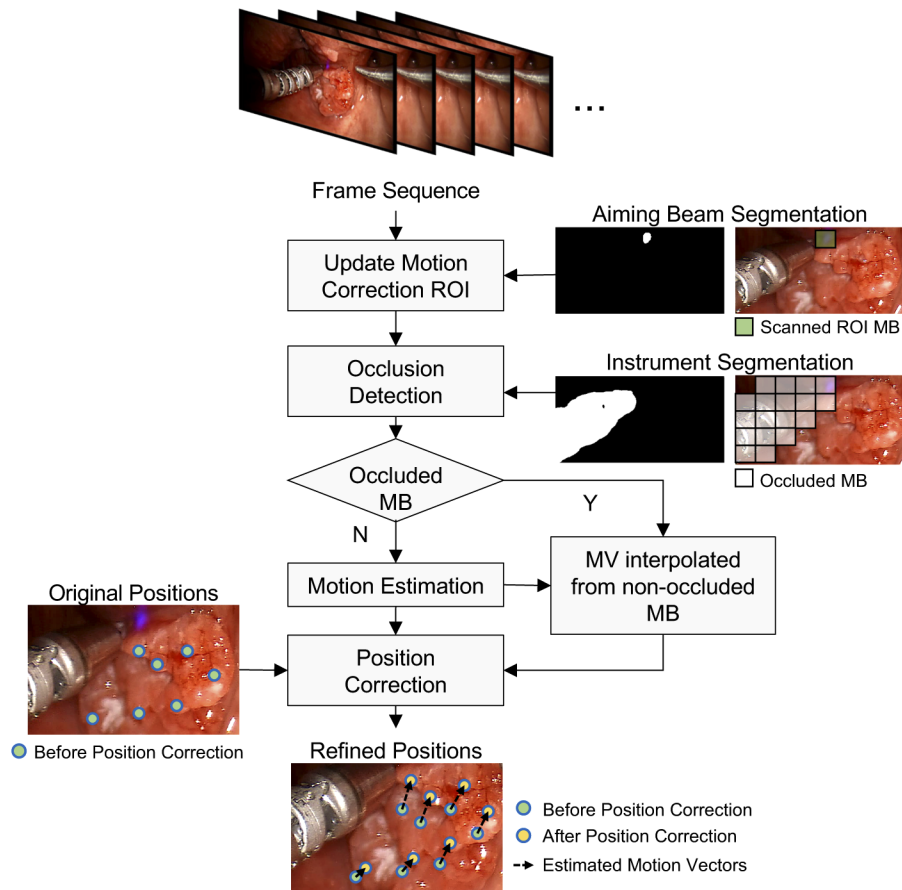


Fig. 3. Overview of the developed tissue motion correction workflow. First, the motion correction ROI was updated based on the scanned tissue region to remove redundant processing. The size of this ROI increases over the course of a given FLIm run as more tissue surface is scanned. In practice, only a small portion of the surgical FOV is scanned and included in this ROI, limiting the computational demands. Then, tissue occlusion was detected using a surgical instrument segmentation CNN to prevent errors due to instrument motion. Adaptive rood pattern search (ARPS) block-matching was then performed to estimation local motion vectors. In the event that a given macro-block is occluded, motion vectors were interpolated from neighboring macroblocks included in the ROI. With an estimated set of motion vectors for a given frame, position correction was performed for prior FLIm point-measurements to ensure an accurate visualization is subsequently generated.

2.3.3. Interpolation-based visualization of point-measurements

Interpolation-based visualization was performed for each frame to allow for real-time visualization. Processing was performed at a reduced image size (640×384 , consistent with CNN downsampling) with point measurement locations adjusted to this new scale. An inverse distance [21] and SNR based interpolation method was applied to visualize data acquired from FLIm point-measurements. Aggregation and interpolation of point-measurements allows for isolated point-measurement errors to be overcome. The rendered FLIm parameter, referred to as the f value, can be a specific spectroscopic parameter (e.g. average lifetime) or the output of a trained classifier. Any pixel location in the surgical FOV within a 15-pixel radius (approximately 0.55 mm at reduced image size) of at least 5 point-measurement centers was included in the

generated overlay for a given frame. This allows for efficient computation and ensures that only data in close proximity to a given pixel influences the final output. For each included pixel x_i , the N point-measurements centered within this 15-pixel radius were aggregated together using Eq. (3)-(5).

$$F(x_i) = \frac{1}{2} \sum_{i=1}^N a_i f_i + b_i f_i \quad (3)$$

$$a_i = \frac{d_i^{-P}}{\sum_{j=1}^N d_j^{-P}} \quad (4)$$

$$b_i = \frac{\min(s_i)^P}{\sum_{j=1}^N \min(s_j)^P} \quad (5)$$

N is the number of data-points within the radius of a given pixel location x_i , d is the Euclidean distance between the influencing data-point i and the pixel of interest x_i . s_i is the set of channel level SNR values recorded for the influencing data-point i . P is the weighting exponent. Greater values of P assign a greater influence on the data-points closest to the pixel of interest (x, y) as well as data-points with higher SNR. If SNR is not available in a given context, interpolation can be performed solely using distance values. After performing this aggregation for an entire scan, the generated overlay was colorized using an RGB colormap (e.g. jet, hot). The overlay was then overlaid onto the white light image of the tissue as a transparent overlay. Only points with a measured SNR greater than 30 dB for all channels were included in a given visualization to mitigate the impact of noise. Overlay pixels were kept in memory and only recalculated if necessary (e.g. tissue motion changing the measurement location, scanning of a nearby location) to reduce computational redundancy. Once a given frame was fully generated it was upsampled back to the original image using bilinear interpolation. The developed visualization method was implemented and evaluated in MATLAB.

2.4. Evaluation dataset

FLImBrush was evaluated using video and FLIm data acquired intraoperatively during 30 head and neck cancer resection surgeries. All procedures took place at the University of California Davis Medical Center once patient eligibility was determined. Research was performed under institutional review board (IRB) approval and with patient's informed consent. A comprehensive set of FLIm point-measurements (1 per video frame) was acquired prior to tumor excision to evaluate the potential of FLIm for assisting in cancer margin assessment. The FLIm scan performed for each patient was acquired following preoperative planning and was less than 90 seconds in duration. This time-efficient scan allows for a comprehensive mapping of the tissue region of interest.

From this 30 patient dataset, 300 images (10 per patient) were annotated to train and validate the aiming beam segmentation model. Only images for which the aiming beam was visible were selected. For each image the outline of the illuminated aiming beam region was annotated to produce a binary ground truth mask for segmentation. As the level of illumination is observed to decrease towards the edges of the beam, the outline of the region was deemed to be the point where the blue aiming beam light was no longer clearly visible. This was performed consistently for all images and inspected by another researcher. In order to then evaluate localization accuracy for an entire FLIm scan, aiming beam center positions for all frames in all 30 videos were manually annotated to serve as ground truth. Error accumulation during tissue motion correction was validated on a set of 4 videos (two using the da Vinci Si and two with a Stryker Endoscope). Interpolation-based visualization was evaluated using an intraoperative scan as well a synthetic two-lifetime image in which noise properties could be controlled. This synthetic two-lifetime

image was generated by first performing a binary threshold (128) to a grayscale image (0-255) of some objects on a desk (hands, cell phone). The two regions were then populated with lifetime values of 3.3 ns and 3.7 ns respectively. Gaussian noise was applied to each region reflecting the statistical properties of cancer and healthy point-measurements for a given head and neck surgery case (standard deviation of 0.35 ns and 0.45 ns respectively).

2.5. Evaluation approach

Each of the three processing steps were evaluated separately to assess the overall performance of the FLImBrush method. Evaluation was performed using an Intel Core i9-9900K @ 3.6 GHz CPU and a Nvidia RTX 2080 Ti GPU.

CNN-based aiming beam segmentation was compared with the previously developed color space thresholding method using two evaluation approaches. First, a 7-fold cross validation was used to evaluate the CNN-based beam segmentation model (33-48 images per fold), with images from a given patient limited to a single fold. Intersection over union (IOU) and Dice coefficient were used as the evaluation metrics. Second, aiming beam localization accuracy across a full video was evaluated using both methods. Euclidean distance from the ground truth position for each frame served as the error metric, with Root Mean Squared Error (RMSE) and max error calculated for each of the 30 videos. The CNN used when evaluating this step for a given video was trained using images from the other 6 cross-validation folds previously described. Processing time was calculated for CNN based aiming beam segmentation in terms of mean Frames-Per-Second (FPS).

Error accumulation during motion correction was investigated for each of the 4 surgical videos by selecting 9 tissue positions as regions of interest (ROIs) in the first frame and comparing the estimated position in the final frame with a manually labelled set of ground truth positions. These manual ground truth positions were reviewed by two researchers prior to analysis. Each of the 9 ROIs were selected from the observed tissue region based on the standard deviation (SD) of grayscale pixel intensity (0-255) in the containing MB. Three ROIs were selected for each of the following groups (High: $SD > 10$, Middle: $5 < SD \leq 10$, Low: $SD \leq 5$). This is referred to as the feature level. Landmark rich ROIs will have a higher feature level and vice-versa. Mean absolute error in terms of Euclidean distance (both in pixels and mm) was calculated for each video to estimate error accumulation. An approximate scale in mm was calculated for each video using the known dimensions of surgical instruments as a reference. This experiment allows any specific sources of error accumulation to be identified. Comparisons were also made with unaltered point-measurement positions in which no motion correction was performed. Processing time was also calculated in terms of mean FPS. This mean FPS was also calculated for the final 30 frames to present processing speed once the majority of the FLIm scan has been performed and a high number of points are being corrected.

Interpolation-based visualization was evaluated for various weighting exponent values (i.e. P value) on both clinical and synthetic data. Qualitative comparisons are made on real and synthetic data in terms of how interpretable the produced visualization is. Direct comparisons are also made with the Overlapping Discs method for a given clinical case. Quantitative comparisons are made on a synthetic two-lifetime FLIm map for which Gaussian noise was applied reflecting the statistical properties of cancer and healthy point-measurements for a given head and neck surgery case. 95% of the pixels were removed from the synthetic pattern to evaluate the interpolation method. Accuracy of the interpolation relative to the original pattern for various P values was calculated using the Structural Similarity Index (SSIM). For synthetic data a purely distance-based interpolation was performed. Reduced image size processing was evaluated on clinical data in terms of processing speed (FPS) and the image deterioration relative to full image size processing (SSIM).

3. Results

3.1. Aiming beam position estimation

Table 1 compares the aiming beam pixel-level segmentation performance and full video beam localization performance of the CNN-based method and color space thresholding. In each case the CNN method is shown to produce far more robust localization of the aiming beam. While the observed IOU and Dice Coefficient scores are not very high, the ground truth occupies a very small number of pixels in this task, resulting in small errors being heavily penalized with these metrics. Ultimately, the accuracy of the calculated beam center is the most important aspect of this evaluation, with an RMSE of 16 ± 10 pixels (1.1 ± 0.7 mm) observed across 30 surgical videos, a 49% decrease from color space thresholding. Maximum error was also observed to be noticeably higher for color space thresholding. Figure 4 presents an example of the difference in output for the two methods, with the CNN method far more robust to specular reflections in the surgical FOV. The trained CNN was observed to process at 37.5 FPS at the chosen image size (640×384).

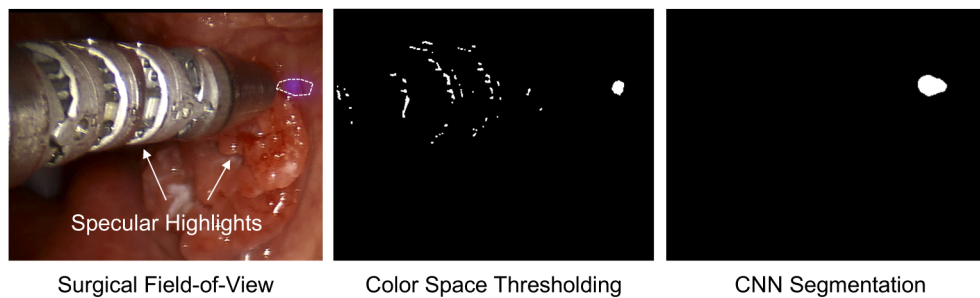


Fig. 4. Comparison of aiming beam segmentation using HSV color space thresholding and the developed CNN method. The CNN method is observed to be more robust to specular highlights on the instrument and tissue, while also segmenting the more transparent edges of the aiming beam. The outline of the aiming beam region used for model training and evaluation is shown.

Table 1. Comparison of the aiming beam localization methods

Beam Localization Method	Pixel-level Segmentation		Full Video Localization	
	IOU ($\pm \sigma$)	Dice Coef. ($\pm \sigma$)	RMSE (px $\pm \sigma$ / mm $\pm \sigma$)	Max Error (px $\pm \sigma$ / mm $\pm \sigma$)
U-Net CNN	0.41 ± 0.03	0.54 ± 0.03	$16 \pm 10 / 1.1 \pm 0.7$	$93 \pm 51 / 6.8 \pm 3.7$
Color Space Thresholding	0.23 ± 0.01	0.39 ± 0.3	$31 \pm 21 / 2.2 \pm 1.5$	$217 \pm 141 / 15.9 \pm 10.4$

3.2. Motion correction

Table 2 presents error accumulation results and computational speed for 4 head and neck surgery videos (all with 1280×720 pixel image size). Figure 5 highlights the error accumulation for 9 individual ROIs from a given video. The application of motion correction results in a noticeable reduction in position error (35%) relative to not performing any correction. In all cases the mean error accumulation with correction is less than 1.23 mm, well within the acceptable margin of error for imaging of this kind. An overall mean error across the four patients of 0.95 ± 0.2 mm was observed. The key source of error is shown to be visually homogeneous tissue regions with a low feature level which lack discerning landmarks to track (e.g. surface of the tongue). The smaller the feature level, the larger the position error tends to be. Processing speed in all instances is greater than 30 FPS, even for the final 30 frames in which the majority of the point-measurement

locations are being tracked. Video examples of motion correction are provided as supplementary material (see [Visualization 1](#) and [Visualization 2](#)).

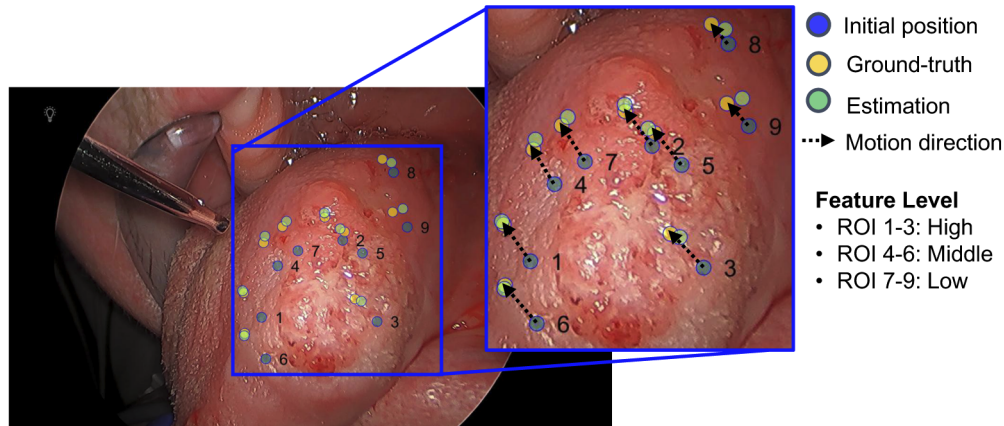


Fig. 5. Comparison of error accumulation for the 9 ROIs for a given surgical video. Greater error is observed for homogenous tissue regions such as ROI 8 and 9 which have less variation in the surrounding pixel intensity values and less clear landmarks to track.

Table 2. Motion correction error accumulation and computational performance

Video	Approach	MC OFF				MC ON	
		Error (px ± σ)	Error (mm ± σ)	Error (px ± σ)	Error (mm ± σ)	Overall FPS (± σ)	FPS Final 30 Frames
1	Endoscope	58.05 ± 10.15	5.46 ± 0.95	13.05 ± 5.84	1.23 ± 0.55	43.59 ± 0.12	39.40 ± 2.96
2	Endoscope	70.43 ± 7.17	3.03 ± 0.31	19.48 ± 8.25	0.84 ± 0.35	38.56 ± 0.21	37.04 ± 2.55
3	da Vinci	87.58 ± 7.47	4.29 ± 0.37	21.62 ± 8.73	1.06 ± 0.43	44.23 ± 0.16	49.65 ± 8.12
4	da Vinci	36.03 ± 3.36	1.98 ± 0.18	12.56 ± 4.59	0.69 ± 0.25	55.98 ± 0.48	49.47 ± 2.23

3.3. Interpolation-based visualization

Figure 6 presents the output of various visualization configurations on a head and neck surgery video as well a synthetic sample. Average fluorescence lifetime data from CH2 (linked with the NADH emission maxima) was rendered for the surgical data. The aiming beam was successfully localized for 1886/2244 video frames of the surgical scan. The cause of unsuccessful localization in this case was the aiming beam occasionally being obstructed by the fiber probe itself. For the surgically acquired data, interpolation-based visualization with a weighting exponent of 1.0 leads to a coherent and complete FLIm map that is qualitatively deemed to be the most interpretable when compared to the Overlapping Discs method as well as interpolation with higher P values. A higher P value leads to small, local variations in average lifetime being highlighted and producing a noisier final visualization. This local variation in clinical lifetime values results in the need for smoothing to be applied without obscuring the delineation of distinct lifetime regions. For the synthetic sample, a P value of 1.0 leads to the most coherent visualization with the highest SSIM relative to the original synthetic image. While an SSIM of 0.56 is not very high, it is clearly superior to higher exponent configurations and acceptable given the high number of pixels removed and the level of Gaussian noise applied.

A mean processing speed of 31.5 FPS was recorded for interpolation-based visualization on the surgical video shown in Figure 6 (2244 frames/point-measurements in total). When comparing the rendered visualization of the final frame, reduced image size processing was observed to have

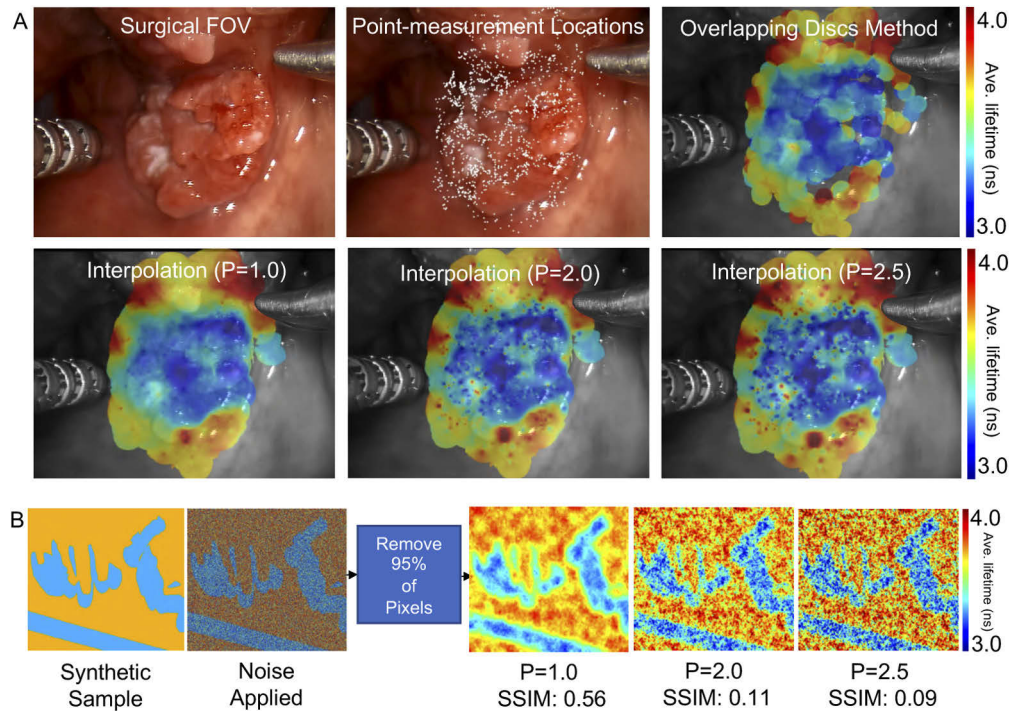


Fig. 6. Comparison of various FLIm visualization methods for a clinical head and neck surgery case as well as a synthetic sample: (a) sparse point-measurements taken from a clinical head and neck sample are visualized using the Overlapping Discs method as well as the developed interpolation-based method with various P values (i.e. weighting exponent). Interpolation-based visualization with a P value of 1.0 leads to the most complete and interpretable overlay of the scanned tissue. No SNR filtering was applied to the clinically acquired data-points acquired for the purpose of this experiment, (b) a synthetic two lifetime FLIm sample was generated with noise properties representative of cancer and healthy measurements acquired clinically. 95% of the pixels were removed before interpolation-based visualizations were generated with various P values. A P value of 1.0 leads to the highest SSIM relative to the original synthetic sample while also being the easiest to visually interpret.

an SSIM of 0.968 relative to full image size processing. This highlights the minimal impact of reduced image size processing on the final output. Full image size processing was observed to visualize at 14 FPS.

4. Discussion

In this study, FLImBrush, a robust method for the localization and visualization of free-hand FLIm point-measurements was developed. We showed that this approach overcomes several key challenges associated with free-hand medical imaging including tissue motion, reflections in the surgical field and sparse sampling to allow for the dynamic, accurate and interpretable visualization of intraoperative FLIm data. Each of the key processing steps of FLImBrush have been demonstrated to be capable of real-time processing, highlighting the feasibility for intraoperative use. The developed method is suitable for applications such as cancer margin assessment, where an overall representation of the transition between cancer and healthy tissue needs to be clearly delineated prior to tumor excision. In a transoral robotic surgery (TORS)

procedure surgeons operate with a fixed level of control and precision and include a margin of care to prevent residual tumor. A surgeon (DGF) involved in our research team indicated an acceptable margin of error of 1-2 mm when dealing with muscle or mucosa. The goal of this method is to provide guidance that can match the level of precision capable with the surgical tools in use. A mean aiming beam localization error of 1.1 ± 0.7 mm and mean motion correction error of 0.95 ± 0.2 mm were observed, suggesting it is feasible to achieve the required level of precision with this method.

The main outcomes of this study include the following: (i) CNN-based aiming beam segmentation results in noticeably superior localization to the previously developed color space thresholding method. This CNN-based segmentation model can be extended to other surgical contexts (e.g. neurosurgery, breast cancer) with the generation of a more extensive training set. Occasional unsuccessful aiming beam localizations were encountered due to the aiming beam being obstructed by the fiber probe in certain da Vinci cases using the Si system. This issue has been addressed through the development of a smaller and more flexible fiber probe used for the da Vinci SP System cases which prevents such obstructions. This issue is not as prevalent for endoscopic cases due to the thin probe employed. (ii) Tissue motion correction is an essential step for any intraoperative point-measurement based imaging approach, improving the accuracy of the rendered visualization as well as any subsequent coregistration with histopathology. Motion correction error was observed to be at an acceptable level (< 1.23 mm, mean of 0.95 ± 0.2 mm) across 4 surgical videos from two distinct setups (da Vinci surgical system, Strkyer Endoscope). These scans were performed by a surgeon and deemed to be representative of a typical scenario for pre-resection imaging both in terms of scanning time and the size of the tissue region of interest. Interestingly, mean error accumulation was comparable using both the da Vinci system and the more conventional endoscope, despite greater camera motion observed in the endoscopic surgery videos. (iii) Interpolation-based visualization with a weighting exponent of 1.0 leads to accurate and easy to interpret visualizations of FLIm data in which minor local variations are attenuated. The developed interpolation method is capable of highlighting the distinct lifetime regions within a scanned tissue surface, even with relatively few point-measurements acquired and the presence of signal noise.

While FLImBrush shows great promise, it currently has some potential limitations. First, aiming beam segmentation via the integrated camera is not possible if the beam is temporarily obscured by the surgical instruments or an uneven tissue surface. This can result in a loss of FLIm data during an intraoperative scan. A possible solution is to employ an alternative method to track the position of the fiber probe which does not exclusively rely on white light image data (e.g. kinematics). Second, higher error accumulation was encountered when performing motion correction for smooth tissue surfaces which lack discerning landmarks. This can be addressed by approximating the motion vectors in low feature level regions using the estimated motion vectors in neighboring high features level regions. Since this process would be performed in terms of macroblocks, the impact on the processing time would be negligible. Ultimately, the presence of difficult to track tissue surfaces will vary with anatomy. Third, the developed method applies a 2-D FLIm overlay onto a 3-D tissue surface with non-uniform topology. This limits how accurately an intricate anatomy can be imaged and prevents highly detailed coregistration. This issue can potentially be resolved through integration with augmented reality technologies such as the Microsoft HoloLens [22] and/or the stereoscopic 3-D imaging capability of the da Vinci Surgical System. Such integration would allow surface reconstruction of the surgical field to be performed and a 3-D visualization rendered. Scale can also vary between anatomies. The dynamic adjustment of the interpolated overlay based on image scale and magnification is another potential improvement that will be investigated in a future study.

FLImBrush has the potential to provide intraoperative diagnostic contrast in applications such as cancer margin assessment and image guided biopsy in a range of anatomies including

head and neck, brain and breast. The output of a trained classification model such as a random forest can potentially be visualized to distinguish specific tissue conditions in the surgical field (e.g. cancer, necrosis, dysplasia). FLImBrush also has the potential to be integrated with other imaging setups such as high resolution microendoscope (HRME) [5] or Raman spectroscopy [23,24], and is compatible with fiber-based methodologies which use exogenous probes such as Panitumumab/Cetuximab-IRDye800CW [25]. Future studies will look to carry out a more extensive assessment of FLImBrush in an intraoperative setting with the development of an optimized software implementation and a qualitative assessment of the rendered visualizations from a team of physicians. The development and extensive validation of a tissue type classifier for head and neck cancer margin assessment will be carried out in a future study. These next steps will allow for key refinements to be made to the software and the visualization approach, resulting in a fiber-based FLIm method that can provide actionable, real-time visualizations for surgical guidance in a range of applications.

Funding

Intuitive Surgical; National Institutes of Health (R01 CA187427).

Acknowledgments

We thank our clinical coordinator Angela Beliveau for her efforts in recruiting head and neck cancer patients for this work. We are grateful to Athena Tam and Tianchen Sun for their various contributions to the project, which include assisting with the acquisition and preparation of clinical data. We would like to thank Jonathan Sorger for his assistance in the design of the updated fiber probe used with the da Vinci SP robot.

Disclosures

The authors declare no conflicts of interest

References

1. X. Liu, Y. Huang, and J. U. Kang, "Distortion-free freehand-scanning OCT implemented with real-time scanning speed variance correction," *Opt. Express* **20**(15), 16567–16583 (2012).
2. B. Y. Yeo, R. A. McLaughlin, R. W. Kirk, and D. D. Sampson, "Enabling freehand lateral scanning of optical coherence tomography needle probes with a magnetic tracking system," *Biomed. Opt. Express* **3**(7), 1565–1578 (2012).
3. F. Cenni, D. Monari, K. Desloovere, E. Aertbeliën, S.-H. Schless, and H. Bruyninckx, "The reliability and validity of a clinical 3D freehand ultrasound system," *Comput. Methods Programs Biomedicine* **136**, 179–187 (2016).
4. M. H. Mozaffari and W.-S. Lee, "Freehand 3-D ultrasound imaging: a systematic review," *Ultrasound in Med. & Biol.* **43**(10), 2099–2124 (2017).
5. Y. Tang, A. Kortum, S. G. Parra, I. Vohra, A. Milbourne, P. Ramalingam, P. A. Toscano, K. M. Schmeler, and R. R. Richards-Kortum, "In vivo imaging of cervical precancer using a low-cost and easy-to-use confocal microendoscope," *Biomed. Opt. Express* **11**(1), 269–280 (2020).
6. B. W. Weyers, M. Marsden, T. Sun, J. Bec, A. F. Bewley, R. F. Gandour-Edwards, M. G. Moore, D. G. Farwell, and L. Marcu, "Fluorescence Lifetime Imaging (FLIm) for Intraoperative Cancer Delineation in Transoral Robotic Surgery (TORS)," *Transl. Biophotonics* **1**(1-2), e201900017 (2019).
7. A. Alfonso-Garcia, J. Bec, S. Sridharan Weaver, B. Hartl, J. Unger, M. Bobinski, M. Lechpammer, F. Girgis, J. Boggan, and L. Marcu, "Real-time augmented reality for delineation of surgical margins during neurosurgery using autofluorescence lifetime contrast," *J. Biophotonics* **13**(1), e201900108 (2020).
8. D. Gorpas, J. Phipps, J. Bec, D. Ma, S. Dochow, D. Yankelevich, J. Sorger, J. Popp, A. Bewley, and R. Gandour-Edwards, Others, "Autofluorescence lifetime augmented reality as a means for real-time robotic surgery guidance in human patients," *Sci. Rep.* **9**(1), 1187 (2019).
9. D. Gorpas, D. Ma, J. Bec, D. R. Yankelevich, and L. Marcu, "Real-time visualization of tissue surface biochemical features derived from fluorescence lifetime measurements," *IEEE Trans. Med. Imaging* **35**(8), 1802–1811 (2016).
10. L. Marcu, P. M. W. French, and D. S. Elson, *Fluorescence lifetime spectroscopy and imaging: principles and applications in biomedical diagnostics - Chapter 3: Tissue fluorophores and their spectroscopic characteristics* (CRC Press, 2014).

11. D. R. Yankelevich, D. Ma, J. Liu, Y. Sun, Y. Sun, J. Bec, D. S. Elson, and L. Marcu, "Design and evaluation of a device for fast multispectral time-resolved fluorescence spectroscopy and imaging," *Rev. Sci. Instruments* **85**(3), 034303 (2014).
12. J. Liu, Y. Sun, J. Qi, and L. Marcu, "A novel method for fast and robust estimation of fluorescence decay dynamics using constrained least-squares deconvolution with Laguerre expansion," *Phys. Med. Biol.* **57**(4), 843–865 (2012).
13. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer, 2015), pp. 234–241.
14. A. A. Shvets, A. Rakhlin, A. A. Kalinin, and V. I. Iglovikov, "Automatic Instrument Segmentation in Robot-Assisted Surgery using Deep Learning,".
15. J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Machine Learning Res.* **12**, 2121–2159 (2011).
16. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds. (Curran Associates, Inc., 2019), pp. 8024–8035.
17. G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools* (2000).
18. Y. Nie and K.-K. Ma, "Adaptive rood pattern search for fast block-matching motion estimation," *IEEE Trans. on Image Process.* **11**(12), 1442–1449 (2002).
19. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).
20. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understanding* **110**(3), 346–359 (2008).
21. D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM national conference*, (ACM, 1968), pp. 517–524.
22. M. G. Hanna, I. Ahmed, J. Nine, S. Prajapati, and L. Pantanowitz, "Augmented reality technology using Microsoft HoloLens in anatomic pathology," *Arch. Pathol. & Lab. Med.* **142**(5), 638–644 (2018).
23. K. Guze, H. C. Pawluk, M. Short, H. Zeng, J. Lorch, C. Norris, and S. Sonis, "Pilot study: Raman spectroscopy in differentiating premalignant and malignant oral lesions from normal mucosa and benign lesions in humans," *Head Neck* **37**, 511–517 (2015).
24. S. Dochow, D. Ma, I. Latka, T. Bocklitz, B. Hartl, J. Bec, H. Fatakdawala, E. Marple, K. Urmey, and S. Wachsmann-Hogiu, Others, "Combined fiber probe for fluorescence lifetime and Raman spectroscopy," *Anal. Bioanal. Chem.* **407**(27), 8291–8301 (2015).
25. R. W. Gao, N. T. Teraphongphom, N. S. van den Berg, B. A. Martin, N. J. Oberhelman, V. Divi, M. J. Kaplan, S. S. Hong, G. Lu, and R. Ertsey, Others, "Determination of tumor margins with surgical specimen mapping using near-infrared fluorescence," *Cancer Res.* **78**(17), 5144–5154 (2018).