

Received February 3, 2021, accepted March 4, 2021, date of publication March 10, 2021, date of current version March 18, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3065104

The Future of Endoscopic Navigation: A Review of Advanced Endoscopic Vision Technology

ZUOMING FU¹, ZIYI JIN¹, CHONGAN ZHANG¹, ZHONGYU HE¹, ZHENZHOU ZHA¹, CHUNYONG HU¹, TIANYUAN GAN¹, QINGLAI YAN², PENG WANG¹, AND XUESONG YE¹

¹Biosensor National Special Laboratory, Key Laboratory of Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou 310027, China

²Hangzhou Center for Medical Device Quality Supervision and Testing, CFDA, Hangzhou 310000, China

Corresponding authors: Peng Wang (pengwangoptimus@zju.edu.cn) and Xuesong Ye (yexuesong@zju.edu.cn)


This work was supported in part by the National Key Research and Development Project under Grant 2019YFC0117901, Grant 2019YFC0117900, Grant 2017YFC0109603, and Grant 2017YFC0110802; in part by the National Major Scientific Research Instrument Development Project under Grant 81827804, in part by the Robotics Institute of Zhejiang University under Grant K11806 and Grant K11807, and in part by the Key Research and Development Plan of Zhejiang Province under Grant 2017C03036 and Grant 2018C03064.

ABSTRACT Minimally invasive medicine has become mainstream because of its crucial clinical significance in providing a low risk of postoperative complications, limited blood loss, short postoperative recovery time, and small sizes of associated physiological tissue wounds. Endoscopic navigation systems comprise a research hot spot in medical science and technology and are an essential means to achieve precision medicine and improve surgical operation safety. As a core component in endoscopic navigation during minimally invasive surgery, endoscopes play a critical role in disease diagnosis and treatment. The development of endoscopic vision technologies has resulted in a renewed drive to further develop endoscopic navigation systems. Multiple endoscopic optical imaging modalities provide data sources for endoscopic vision technology, including white-light endoscopy, contrast-enhanced imaging and technologies involving magnified observation. Endoscopic vision is a specific application of computer vision involving the use of endoscopes that include instrument tracking, endoscopic view expansion, and suspicious lesion tracking in the application of endoscopic navigation. These techniques help surgeons or surgical robots locate instruments and lesions and expand the field of view of the endoscope. Although these technologies have been applied to various clinical and pre-clinical diagnoses and treatments, the use and combination of these advanced technologies in endoscopic navigation system for specific clinical requirements remains challenging. This review performs a broad survey of advanced endoscopic vision technologies and their application in endoscopic navigation systems. Finally, we discuss the challenges and future directions in implementing and developing endoscopic navigation systems.

INDEX TERMS Endoscopic navigation, endoscopic vision, endoscopic view expansion, instrument tracking, suspicious lesion tracking.

ACRONYMS

Acronyms	Definition
AFI	Autofluorescence Imaging
AR	Augmented Reality
BE	Binocular Endoscopy
BLI	Blue-light Imaging
CADe	Computer-aided Detection
CADx	Computer-aided Diagnosis
CBCT	Cone-beam Computed Tomography Systems

The associate editor coordinating the review of this manuscript and approving it for publication was Taous Meriem Laleg-Kirati .

CCD	Charged-coupled Device
CF	Close Focus
CLE	Confocal Laser Endomicroscopy
CMOS	Complementary Metal-oxide Semiconductor
CNN	Convolutional Neural Network
CT	Computed Tomography
CV	Computer Vision
EC	Endocytoscopy
EM	Electromagnetic
ESCC	Oesophageal Squamous Cell Carcinoma

ESCCs	Early Oesophageal Squamous Cell Carcinomas
EVE	Endoscopic View Expansion
FICE	Fujinon Intelligent Chromoendoscopy
GAN	Generative Adversarial Networks
GC	Gastric Cancer
GI	Gastrointestinal
GPU	Graphics Processing Unit
HOG	Histogram of Oriented Gradient
HPs	Hyperplastic Polyps
HRME	High-resolution Microendoscopy
IC	Indigo Carmine
ICG	Indocyanine Green
iMRI	Intraoperative Magnetic Resonance Imaging
IOUS	Intraoperative Ultrasonography
I-scan OE	I-Scan Optical Enhancement
IT	Instrument Tracking
LCI	Linked Colour Imaging
ME	Microendoscope
ETS	Endoscopy with Tracking Sensor
MIS	Minimally Invasive Surgery
MOET	Multiple Optical Endoscopy Technologies
MRI	Magnetic Resonance Imaging
NA	Not Applicable
NBI	Narrow-band Imaging
NIR	Near-infrared
NPs	Neoplastic Polyps
ORB	Oriented Fast and Rotated Brief
RCNNs	Recurrent Convolutional Neural Networks
Refs	References
RGB	Red, Green, Blue
ROBUST-MIS	Robust Medical Instrument Segmentation
SFM	Structure from Motion
SIFT	Scale-invariant Feature Transformation
SLAM	Simultaneous Localization and Mapping
SLT	Suspicious Lesions Tracking
SME	Standard Monocular Endoscopy
SPIES	Storz Professional Image Enhancement System
SSD	Single-shot Multibox Detector
US	Ultrasound
VCE	Virtual Chromoendoscopy
VO	Visual Odometry
WLI	White-light Image

I. INTRODUCTION

In the 21st century, minimally invasive medicine has become a mainstream medical process, overtaking traditional surgical operations, which can cause severe trauma to patients during diagnosis and treatment. Minimally invasive medical technology uses less traumatic or even non-invasive methods and

has many advantages, such as less pain for the patients, quick recovery after surgery, a significantly shortened hospital stay, and a reduction in medical resource uses [1]. With its rapid progression, it has become an emerging medical discipline alongside surgery, internal medicine, paediatrics, etc.

However, as Fuchs [2] described, most surgeons have a precipitous learning curve with minimally invasive surgery (MIS) that leads to a longer operating time than with open surgery. Due to the limitations of human vision and touch, it can be challenging to accurately locate surgical instruments and lesions during the MIS procedure. This affects the two fundamental issues in surgical navigation systems: “where to go” and “how to get there” [3]. In recent years, a number of commercialized surgical navigation systems have been implemented in different hospital departments. In orthopaedics and neurosurgery, the current mainstream surgical navigation systems are primarily obtained from Medtronic, Stryker, General Electric, BrainLAB, etc. [4]. The superDimensionTM system and the VeranTM system are two primary electromagnetic (EM) navigational bronchoscopy platforms [5]. EDDA Technology has developed a puncture surgery navigation system (IQQA[®]-BodyImaging; EDDA Tech., Princeton Junction, NJ, USA) to support various types of MISs for chest and abdomen tumours.

The endoscope is one of the key pieces of equipment in MIS, serving as the “eye” of the surgical navigation system. Since 1806, medical endoscopy has been widely used in numerous parts of the body and has evolved over many generations [6]. The traditional endoscope system generally consists of a light source, a camera, an image controller, a display, and the body of the endoscope. Endoscopes are designed in different forms for different clinical applications. Furthermore, they implement different signal transmission methods (analogue or digital signal transmission), mechanical structures for the inserted section (flexible or rigid), and optical sensors (resolution and spectrum). In recent years, the latest generation of endoscopes have been built based on a variety of disciplines, such as optics, electronics, mechanics, informatics, and graphics. Advanced endoscopic technology introduces new surgical approaches to surgeons and serves as the basis for minimally invasive surgical navigation systems.

Multiple optical endoscopy technologies are able to obtain more information than ever before [7]. For example, the binocular stereoscopic endoscope uses two calibrated endoscope lenses to obtain stereo data. Then, depth information can be extracted by the stereo matching method after epipolar correction. Additionally, multispectral imaging modalities help doctors observe the morphology of tissues in non-white-light conditions, such as narrow-band imaging (NBI, Olympus) and Fujinon intelligent chromoendoscopy (FICE, Fujinon). Surgeons can also use technologies for magnification observation to see more details regarding tissue textures and even cellular objects, then call on their own experience or use algorithms to perform pathological detection. In this way, multimodal optical endoscopes provide a diverse

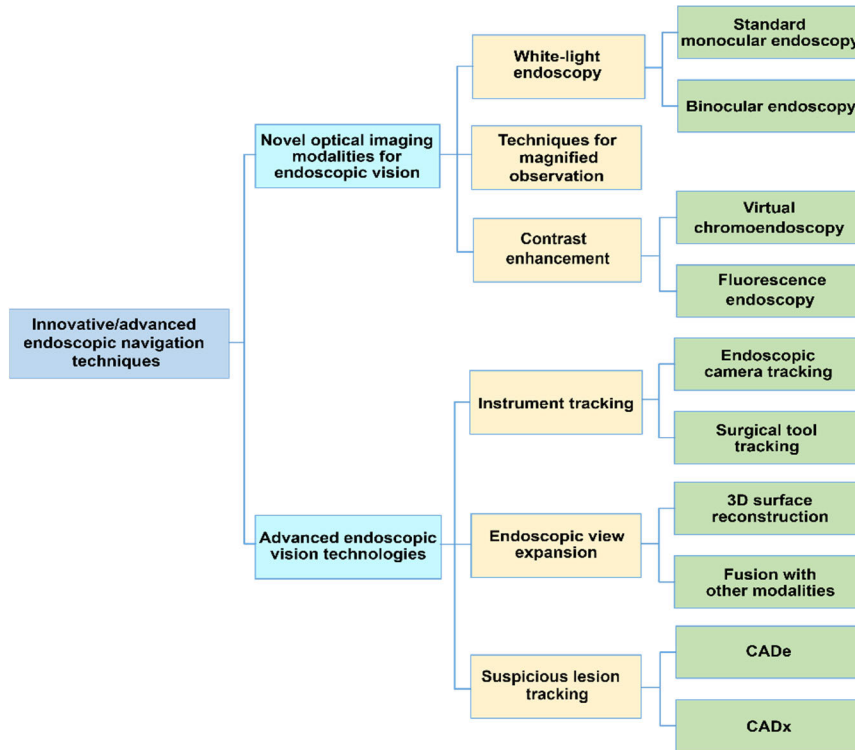


FIGURE 1. Overview of innovative/advanced endoscopic navigation techniques.

range of data forms for endoscopic navigation systems, which also leads to new challenges in the field of computer vision (CV).

CV is a type of science that investigates ways to enable a computer to understand its environment from visual information [8]. Early researchers in CV developed mathematical techniques for recovering the 3D shape and appearance of objects in imagery in an attempt to make computers understand and interpret images the way humans do, which is very difficult [9]. Recently, the development of deep learning has led to great strides in a variety of CV problems, such as object detection, motion tracking, and semantic segmentation [10]. Currently, CV based on deep learning is one of the most popular research fields; as a result, a growing spectrum of CV applications can be observed in minimally invasive surgery, which are mainly based on endoscopes. Endoscopic vision was developed to provide assistance in endoscopic surgery. For example, with the help of endoscopic vision, laparoscopic instruments can be located precisely inside the abdominal cavity through visual feedback [11]. In short, endoscopic vision is a specific application of computer vision involving the use of endoscopes.

As a result, endoscopic vision technology has gradually become the focus of research efforts by laboratories and endoscope manufacturers [12]–[17]. Extracting key information from the electronic signals collected by the various sensors in the endoscope will assist physicians in diagnosis and treatment at multiple stages. Three-dimensional surface reconstruction technology extracts depth information from

the organ surface, which can extend the doctor's field of vision and provide a source of information for the precise positioning and navigation of the surgical robot. Device tracking technology can evaluate the doctor's surgical process, and lesion identification allows doctors to achieve faster and more accurate detection and diagnosis. These developments have evolved significantly and can be applied in endoscopic navigation systems.

For this review, a PubMed and Springer Link literature search was systematically performed for studies published from 1990 to 2020 on novel endoscopic technologies for the implementation of endoscopic navigation systems (Fig. 1) by using the search terms “Endoscopy,” “Endoscopic navigation,” “Endoscopic tracking,” “Endomicroscopy,” “Endoscopic medical imaging,” “Endoscopic detection,” “Endoscopic vision,” “Computer vision,” “Instrument tracking,” “SLAM,” “3D surface reconstruction,” and “Depth estimation.” Additional articles were obtained through a review of the quoted references from the selected reference articles. Only full manuscripts and case reports published in English were collected. We included studies that first proposed novel endoscopic vision techniques that have previously been applied or have the potential to be applied in endoscopic navigation systems. In addition, we placed special focus on technologies that have a variety of medical applications. However, for new or emerging technology where related pieces of literature are scarce, articles with other study designs were also included. As the endoscope continues to evolve in the next generation, endoscopic

navigation will continue to be redefined. Therefore, we firmly believe that this article will provide a broad technical vision for the development of minimally invasive surgical navigation systems and provide an important reference to surgeons and manufacturers of minimally invasive surgical instruments.

The rest of the paper is organized as follows. In Chapter II, multiple optical endoscopy modalities are reviewed, including white-light endoscopy, contrast-enhancement techniques, and technologies that implement magnified observation. Chapter III introduces endoscopic vision technologies, including instrument tracking, endoscopic view expansion, and suspicious lesion tracking. Chapter IV further introduces endoscopic navigation system applications that combine advanced vision technologies for different parts of the body. Chapter V discusses significant research directions for future works. Finally, a summary is provided in Chapter VI.

II. OPTICAL ENDOSCOPY MODALITIES FOR ENDOSCOPIC VISION

Optical endoscopy technologies have great potential to implement endoscopic vision algorithms to extract multidimensional information across several scales from the region of interest. These modalities can also enhance the reliability of the whole navigation system.

A. WHITE-LIGHT ENDOSCOPY

1) STANDARD MONOCULAR ENDOSCOPY

A standard monocular endoscope is a tube with a light and lens that captures the scene in question with charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) sensors. The endoscope can be inserted into the body through a natural orifice or a small incision during surgery. A flexible endoscope can bend easily around corners of the body. With the aid of endoscopes, traditional operations requiring large incisions can be performed with only minor cuts. Endoscopy is also the most effective and least invasive way to screen the precancerous lesions of gastric and colorectal cancer.

2) BINOCULAR ENDOSCOPY

Binocular endoscopes, also known as stereo laparoscopes, are the core surgical instruments of modern “precision” minimally invasive surgery and surgical robots due to their ability to obtain clear images, a wide field of view with flexible distal portions, stereo measurements, precision positioning, and high diagnostic efficiency. The 3D field-of-view real-time reconstruction algorithms used in binocular endoscopy, based on the stereo imaging method and the parallax principle, restore spatial depth information from image pixels. We can reconstruct the 3D view through binocular endoscopy, which enhances the perception of stereovision and field of view of the endoscopic examination scene [18]. The real-time reconstruction algorithm for the 3D field of view can assist the doctor in locating, tracking, and navigating the lesion.

Binocular endoscopy offers a new type of intraoperative navigation technology for endoscopic surgery [19].

B. CONTRAST ENHANCEMENT

Endoscopists can diagnose lesions by assessing the following characteristics: (i) mucosal morphology (ulcer, erosion, protuberance, etc.), (ii) mucosal colour (suspicious red or white spots), and (iii) vascular information (thickness, distribution, blood concentration, etc.). Additionally, according to the images obtained of the mucosal capillaries microstructures, endoscopists can differentiate between cancer and normal tissue due to the significant differences in neoangiogenesis between the two. Contrast-enhanced endoscopic imaging techniques (as shown in Fig. 2) are capable of highlighting these targets. In this way, they can facilitate endoscopists in reducing the miss rates of lesion detection and increase the accuracy in characterizing the lesions [20].

1) FLUORESCENCE ENDOSCOPY

Fluorescence can provide comprehensive and detailed detection of the structure and dynamics of the targeted tissue. Different fluorophores have different fluorescent properties, and their fluorescence spectra are often used in diagnostics [21] because they provide detailed information on fluorescence molecules, such as conformation, binding sites, and interaction within cells and tissues [22]. Fluorophores can be divided into endogenous fluorophores and exogenous fluorophores [23]. With customized optical filters blocking false excitation light to a sufficiently low level, the sensors implemented in fluorescence imaging optical platforms capture part of the emitted fluorescence from the tissue. This part of the fluorescence will contribute to the formation of the final images [24]. Existing clinically available fluorescent endoscopic systems include autofluorescence imaging (AFI) [25]–[27] and near-infrared imaging with indocyanine green (NIR/ICG) [28]–[30]. AFI uses violet light to excite endogenous fluorophores, while NIR/ICG uses a modified range of light to excite exogenous fluorochromes applied to the region of interest.

2) VIRTUAL CHROMOENDOSCOPY

Virtual chromoendoscopy (VCE) augments and detects the spatial variance in light absorption and the scattering properties of tissue and organs in diagnosis and therapy, improving the contrast between abnormal and healthy tissue. According to the illumination light and image processing methods used, VCE can be categorized into three groups: (i) pre-processing VCE (e.g., NBI [31] and blue-light imaging (BLI) [32]), which uses a modified spectrum of light that coincides with the central peak of the absorption characteristics of haemoglobin in the blood vessels, enhancing the contrast between capillaries and adjacent tissue from the mucosa and submucosa; (ii) post-processing VCE (e.g., FICE, i-scan and the Storz professional image enhancement system (SPIES) [33]), which processes digital images to achieve a similar effect as pre-processing VCE via spectral

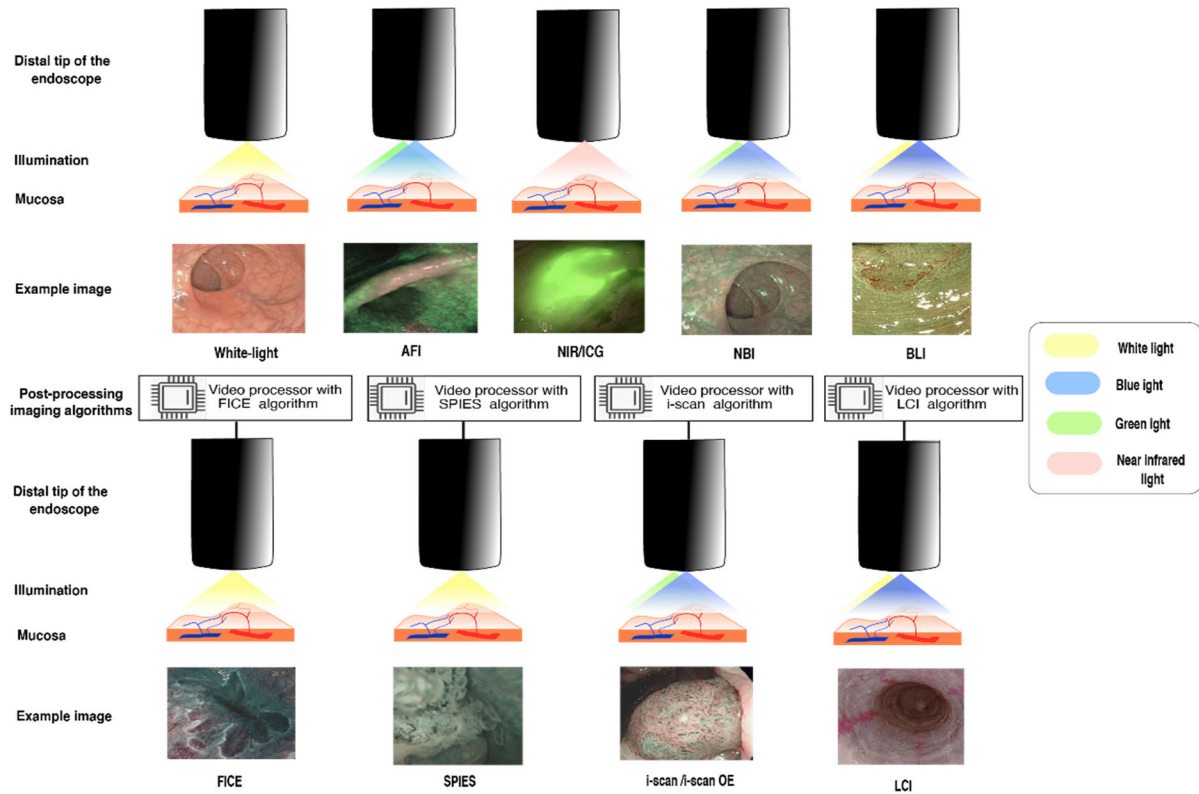


FIGURE 2. Schematic diagram of contrast-enhanced endoscopic imaging techniques, illustrating their basic working principles and corresponding example image results.

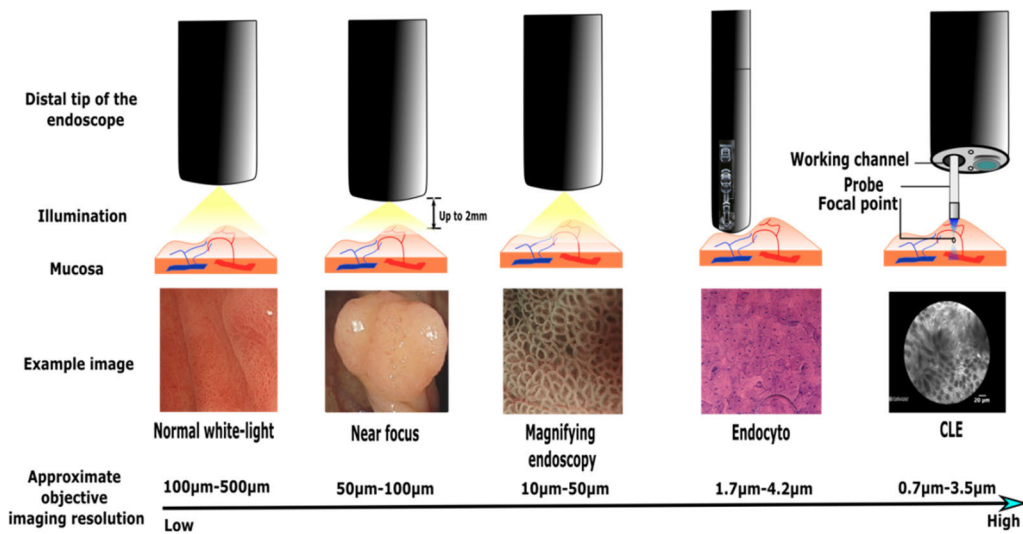


FIGURE 3. Schematic diagram of technologies used in magnified observation. The real objective imaging resolution gradually increases from left to right. EC and CLE are in direct contact with the corresponding tissues during imaging.

reconstruction algorithms; and (iii) linked colour imaging (LCI) [34] and i-scan optical enhancement (i-scan OE [35]), which incorporate both pre- and post-processing methods.

C. MAGNIFIED OBSERVATION

Images with a higher objective resolution than the human eye can perceive can be used to provide more effective diagnostic information to endoscopists. Fig. 3 illustrates several

technologies that have already been implemented for clinical use. Close focus (CF) [36] and optical and electronic magnification [37] provide a clearer vision to the physician while permitting a low miss rate for small lesions. Additionally, surface details, including pit patterns and vascular structures, can be further enhanced with zoom lenses, which can magnify images by up to 150 times [38]. Moreover, with endocytoscopy (EC) [39] or confocal laser endomicroscopy

(CLE) [40], which enables real-time micron-level imaging, endoscopists can characterize suspicious lesions by visualizing multicellular structures such as capillaries and villiform structures; cellular morphologies (for example, crypt, goblet, or epithelial cells); and subcellular organelles (nuclei and cytoplasm). Thus, EC and CLE can facilitate in vivo “optical biopsy” and are promising techniques for replacing ex vivo histology, which is the current gold standard for endoscopic diagnosis. Furthermore, these techniques can be integrated into the distal end of conventional white-light endoscopy, allowing endoscopists to visualize the mucosa both macro- and microscopically [20].

III. ADVANCED ENDOSCOPIC VISION TECHNOLOGY

There are diverse goals and perspectives in the domain of endoscopic vision. This review intends to provide a broad overview of related research in endoscopic navigation and to combine research points from a variety of fields in solving practical problems. In our extensive literature research, articles that met the inclusion criteria were divided into three categories: (i) instrument tracking, (ii) endoscopy view expansion, and (iii) suspicious lesion tracking.

A. INSTRUMENT TRACKING

Minimally invasive surgical instruments include endoscopic cameras and surgical tools such as endoscopic ultrasound sensors, biopsy forceps and minimally invasive surgical robots. It can be difficult for the surgeons to visually track these instruments due to the limited field of view. Therefore, instrument tracking technology forms a very important component of endoscopic navigation. This technology can be divided into endoscopic camera tracking and surgical tool tracking depending on the tracking methods implemented.

1) ENDOSCOPIC CAMERA TRACKING

a: VIDEO-BASED TRACKING

Video-based tracking is used to locate the endoscopic camera by using video and photos only. This tracking technology is also known as visual odometry (VO) and is mainly based on simultaneous localization and mapping (SLAM) algorithms [41], which are commonly used in the robotic and autonomous driving fields. The movement of the camera can be predicted by detecting the salient features in the video and analysing the differences in the locations of those features in different frames.

Fig. 4 shows two frames (I_1, I_2) of camera movement. P is a point in 3D space, which can be written as $[X, Y, Z]^T$. P_1 and P_2 are the projections of point P on the two imaging planes, called feature points. The pixel coordinates of these two projection points can be expressed as in (1):

$$p_1 = KP, p_2 = K(RP + t) \quad (1)$$

R and t are the homogeneous rotation and translation matrices for the camera pose, respectively, and K is the intrinsic camera matrix. Letting $x_1 = K^{-1}p_1, x_2 = K^{-1}p_2$, the

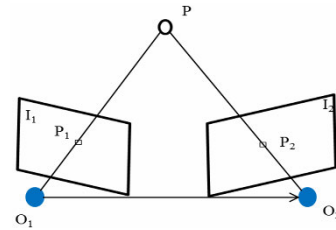


FIGURE 4. Two frames of camera movement in visual odometry.

formula can be written as in (2):

$$t^\wedge x_2 = t^\wedge R x_1 \quad (2)$$

t^\wedge is the skew-symmetric matrix of t . If we then left multiply by x_2^T , we obtain (3):

$$x_2^T t^\wedge R x_1 = 0 \quad (3)$$

We define $E = t^\wedge R$, where E is called the essential matrix, which can be solved with the pixel coordinates of feature points, allowing us to then determine the translations of R and t .

In the field of endoscopic tracking, Grasa *et al.* [12] proposed a visual SLAM module to provide an up-to-scale 3D map of the observed cavity and endoscope trajectory, which was validated with synthetic data and human in vivo image sequences corresponding to 15 laparoscopic hernioplasties. Lin *et al.* [42] focused on the preoperative and intraoperative data, adopted a parallel tracking and mapping framework, and extended it for use in stereoscopy [42]. Mahmoud *et al.* [43] applied ORB-SLAM to estimate the location of the endoscope and the 3D structure of the surgical scene. Later, he proposed a monocular, quasi-dense reconstruction algorithm through a depth propagation algorithm that used keyframe images [44, 45]. This method is robust to severe illumination changes, poor textures, and small deformations in endoscopic surgery. Prendergast *et al.* [46] used the ORB-SLAM2 method in capsule endoscopy to diagnose gastrointestinal (GI) abnormalities. Their method was demonstrated with a robotic endoscope and in an intestine model. Turan *et al.* [47] used convolutional neural network (CNN)-based vision odometry and depth learning methods in capsule endoscopy and obtained good results. The performance of the method was demonstrated by using ex vivo porcine stomach models. Wang *et al.* [48] proposed a visual SLAM method for bronchoscope tracking that obtained a large improvement over ORB-SLAM.

No additional sensors are required for video-based tracking methods. Therefore, the diameter of the endoscope can be designed to be smaller. However, owing to the poor imaging environment in the human body, pure video-based tracking methods still have much room for improvement in accuracy and application.

b: EXTERNAL SENSOR-BASED TRACKING

An external sensor is often used in endoscope tracking. During the surgery, the pose of the endoscope will be calculated

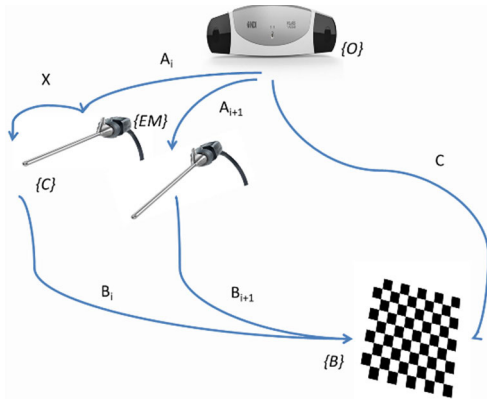


FIGURE 5. Example of hand-eye calibration.

by the sensor tracker and the “hand-eye” calibration matrix, which indicates the relative pose relationship between the sensor tracker and the endoscope [49]–[52]. When using this method, the endoscope and tracker must be rigidly connected. The endoscope pose can be calculated as in (4):

$$\begin{bmatrix} R_e & T_e \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R_s & T_s \\ 0 & 1 \end{bmatrix} \cdot X \quad (4)$$

where R_e and R_s are the rotation matrix of the endoscope and the positioner relative to the world coordinate system, respectively, T_e and T_s are the respective translation vectors, and X is the hand-eye calibration matrix. It is easy to obtain the pose of the sensor tracker from R_s and T_s with high precision and accuracy. Therefore, obtaining an accurate hand-eye calibration matrix becomes the primary problem.

The key to calculating the hand-eye calibration matrix is to solve the formula $AX = XB$. In Fig. 5, $\{O\}$, $\{EM\}$, $\{C\}$ and $\{B\}$ represent the tracking system, endoscope-attached marker, endoscope camera and calibration board, respectively. A_i and B_i represent, respectively, the transformations from $\{O\}$ to $\{EM\}$ and from $\{C\}$ to $\{B\}$ in the i th movement. The endoscope is moved to obtain different images of the calibration board, which is kept still. It is then easy to obtain the following equation:

$$A_i X B_i = A_{i+1} X B_{i+1} \quad (5)$$

Let $A'_i = A_{i+1}^{-1} A_i$, $B'_i = B_{i+1} B_i^{-1}$. Then, the equation can be rewritten as in (6):

$$A'_i X = X B'_i \quad (6)$$

In the field of endoscopic tracking, Lee *et al.* [50] used a checkerboard and an optical tracking system to complete calibration and then analysed the error. He found that the factors contributing to the error included the distance between the sensor and the camera lens and the order of solving the different matrices. Thompson *et al.* [49] chose to use an invariant point and compared the errors of optical and EM tracking systems. He found that even if the error of the optical tracker sensor was smaller than that of the EM tracker due to the closer distance between the installation position of the electromagnetic tracker and the lens, the total error was smaller.

Ha *et al.* [53] used a checkerboard composed of triangles rather than quadrilaterals to improve the accuracy of corner image extraction, maintaining stability during video shooting with a robot arm. Lai *et al.* [54] designed a 3D calibration object and compared it with a 2D calibration plate. They also sought to define the minimum number of poses required to obtain a good calibration result. The EM waved-based localization method was used with a wireless capsule endoscope [55]. Other methods have also been developed to perform “hand-eye” calibration for endoscopes [56]–[58], with principles similar to those for solving the formula $AX = XB$.

Due to the high accuracy of the additional sensor, the total accuracy of external sensor-based tracking is better than that of the video-based method. However, the current inability to reduce the size of the sensor is an essential factor restricting its development.

c: HYBRID TRACKING

Some methods combine external tracking sensors and visual information. Turan *et al.* [59] proposed a non-rigid and deformable RGB depth fusion method, which combines magnetic positioning and RGB images to provide excellent, precise tracking results for endoscopic capsule robots. Charreyron *et al.* [60] integrated magnetic kinematic information into a bundle adjustment procedure as a regularization term to penalize deviations between camera poses. Their method yielded a correctly scaled and lower-drift result compared with visual-only methods. Yang *et al.* [61] introduced a vision-based endoscope tracking method by combining 3D ultrasonography with endoscopic vision, which solved the problem of scale ambiguity and interest-point inadequacy.

2) SURGICAL TOOL TRACKING

Surgical tool tracking is an important concern in endoscopic navigation and is usually a prerequisite for computer- and robot-assisted intervention [62]. Currently, surgical tool detection and tracking data are provided through image-based (also called vision-based) and sensor-based technologies [63]. Sensor-based methods are now widely used in clinical practice. After calibration and coordinate system transformation, the position of the tip of the surgical tool can be stably acquired in real time. In bronchoscopic navigation surgery, biopsy forceps with EM sensors can be rendered in the virtual endoscope in real time after an initial calibration [64], [65]. Puncture needles with an optical sensor can help the surgeon achieve precise ablation in liver navigation systems [49], [66].

In early studies, some image processing methods were proposed for the segmentation of colour markers on surgical tools to track them [67]–[69]. Although the above methods could be efficiently run on a computer, they also had obvious limitations. For example, the selected materials need to have good biocompatibility, most of the existing endoscopic impact data do not meet the requirements, and they are easily affected by light and shadow. [70]. Subsequently, many studies have proposed a series of feature extraction

methods to enhance the robustness of surgical tool tracking. Colour is the most widespread natural feature, and almost all existing surgical tracking methods use colour information as input. Lee *et al.* [71] first proposed using RGB colour space as part of the framework in a minimally invasive surgical context. Although colour features were more convenient to use, they did not perform well in places with shadows and highlights. Gradients are another very popular feature. Typically, the Hough transform [72], [73] and histogram of oriented gradients (HOG) descriptors [74] are widely used to extract the edges of surgical tools. These gradient features can extract the edges and corners of surgical tools well, but noise easily interferes with them. Furthermore, texture feature extraction methods such as scale-invariant feature transformation (SIFT) [75] and Colour-SIFT [76] have been proposed to improve the robustness of the tracking algorithm.

In recent years, with the continuous development of CNNs, an increasing number of studies have begun to apply them in surgical tool tracking. CNNs have been demonstrated to have high feature extraction and expression capabilities. Initially, CNN methods were used to replace specific steps in surgical tool tracking. Zhang *et al.* [77] proposed a method that uses a line segment detector to detect the positions of selected lines and a CNN to find the tip of the surgical tool. Wang *et al.* [78] combined VGGNet [79] and GoogleNet [80] to detect surgical tools and then used an average ensemble learning method to avoid overfitting two deep neural networks. Recently, end-to-end CNN approaches have become a popular research direction in surgical tool tracking. Redmon *et al.* [81] proposed a real-time detection method based on the ‘You Only Look Once’ (YOLO) algorithm [82]. Methods based on U-Net [83] have been widely applied in surgical tool tracking [84]. Colleoni *et al.* [13] proposed a U-Net-based structure combined with 3D fully convolutional neural networks (3D FCNNs) for detecting the locations of surgical tools. In the 2019 Robust Medical Instrument Segmentation (ROBUST-MIS) Challenge [62], the Haoyun team used the DeepLabV3+ [85] architecture to focus on high-level information; a pre-trained ResNet101 [86] was used as an encoder and the focal loss was combined with the Dice similarity coefficient to train the network [87], leading the group to become the most successful participating team in the challenge.

Sensor-based methods for tracking surgical instruments require expensive equipment and hardware. Methods based on endoscopic vision can directly infer the position of the tool in the video frame through endoscopic image vision without modifying the tool or the surgical procedure [88], causing them to become states-of-the-art in surgical tool tracking tasks in endoscopic navigation.

B. ENDOSCOPIC VIEW EXPANSION

Due to the limits of the field of view, doctors have very limited sense of the environment in endoscopic surgery. Endoscopes operate via visible-light optical camera imaging. Thus, internal information more than 10 mm below the surface of the

organ cannot be observed. Endoscopic view expansion can effectively solve this problem. Three-dimensional surface reconstruction can help to expand the field of view, and fusion with other modalities, such as ultrasound, CT, and MRI, can help expand the field of view to visualize deeper parts of the tissue.

1) THREE-DIMENSIONAL SURFACE RECONSTRUCTION

Conventional endoscopy, which lacks 3D vision and is unable to provide proper depth perception, severely limit the effect of diagnostic examinations and the administration of therapy. Via 3D surface reconstruction technology, doctors can better enhance the perception of endoscopic displays on augmented reality (AR) systems and avoid the surgical risks caused by poor visibility and insufficient experience. The pipeline of 3D reconstruction consists of camera calibration, depth estimation, point cloud reconstruction, and point cloud registration, as shown in Fig. 6. In this section, we will focus on camera calibration and depth estimation, which have the greatest effect on accuracy.

a: CAMERA CALIBRATION

Reconstruction of a 3D surface using a computer vision technique requires prior information on the relationship between the pixel location of the camera image and the location of the corresponding spot in the actual scene, which can be obtained from camera calibration. Distortion calibration can eliminate the effect of geometric deformation arising from optical imaging systems. For binocular endoscopy, external parameter calibration is needed for epipolar line rectification. Accurate camera calibration is a prerequisite for synthesizing accurate spatial information from relevant images. For ordinary cameras, the pinhole model is a good approximation of its projection relation. Equation (7) expresses the relationship between point p in pixel coordinates and point P in real-world coordinates. $[R|t]$ is an external parameter matrix composed of a rotation and a translation. Distortion is defined in the normalized image coordinate. There are different kinds of distortion models, but the most commonly used is radial distortion. Sometimes, tangent distortion is also added. Radial distortion is derived from light bent away from the centre of the lens, which is expressed by the first term of equation (8). Tangent distortion is caused by assembly errors in which the lens is not ideally parallel to the sensor [89], which is represented in the last two terms of (8).

$$p = K [R|t] P \quad (7)$$

$$\begin{cases} x_{corrected} = x (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ \quad + 2p_1 xy + p_2 (r^2 + 2x^2) \\ y_{corrected} = y (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ \quad + p_1 (r^2 + 2y^2) + 2p_2 xy \end{cases} \quad (8)$$

where $(x_{corrected}, y_{corrected})$ are the corrected, undistorted image coordinates, (x, y) are the original, distorted image coordinates, and r is $\sqrt{x^2 + y^2}$. k_1, k_2 and k_3 is the radial distortion factor, p_1 and p_2 is the tangential distortion factor.

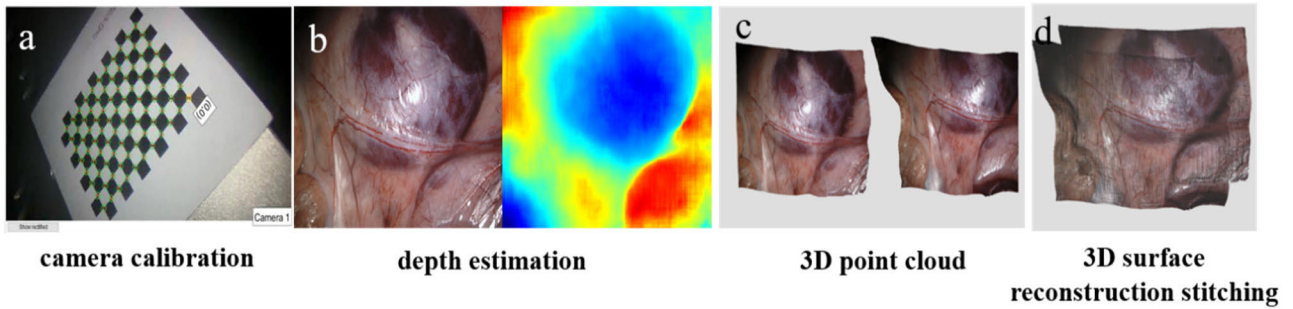


FIGURE 6. Process of 3D surface reconstruction of the abdominal cavity.

Calibration methods can be classified into self-calibration and object-based calibration. Self-calibration is performed by moving the camera in a static scene to obtain sufficient images for parameter estimation. Object-based calibration, as its name suggests, requires a standard target object and is suitable for endoscopic scenes because of its high precision. Consequently, we will focus on object-based calibration next.

Tsai *et al.* [90] pioneered using a 2D calibration target with a two-step method that involves a closed-form solution and nonlinear minimization. Weng *et al.* [91] developed a two-step method to perform optimization that takes all the parameters into account. Zhang *et al.* [92] proposed a brief and effective calibration process that takes advantage of the homographic transformation and orthogonality of the rotation matrix. The work of Zhang *et al.* has been most widely used, and many subsequent works have made improvements within their framework [93].

More-recognizable planar patterns and more effective control point estimation strategies are critical for accurate calibration. Therefore, they are the focus of many works [53], [94]–[97]. Heikkilä and Silven [94] first stressed that perspective projection was generally not a shape-preserving transformation, and therefore the centre of a circle in an image is usually not the projection of the circle centre on the plane. Thus, they derived formulations [94], [95] to calculate the real projection of the circle centre. Rufli *et al.* [98] proposed a method to detect chessboards in blurred and highly distorted images. Ouellet *et al.* [96] summarized existing circle centre estimation methods, explained their applicable scopes and proposed an improvement. Geiger *et al.* [97] developed an effective method to estimate sub-pixel corners of a chessboard pattern, which was adopted in MATLAB and OpenCV. Ha *et al.* [53] used a deltile grid pattern with a polynomial fitting-based method to detect monkey saddle points and obtained excellent calibration accuracy. Additionally, phase-based patterns [99]–[101] have been proposed that are robust to out-of-focus images. The processes include the use of the phase-shift method to encode stripe or circle patterns and unwrapped phases in images taken by the camera to obtain the corresponding points.

Recently, Zhu *et al.* [102] used a camera equipped with a polarizer and obtained an image at an optimal polarization

angle in different spatial positions to eliminate the impact of highlighted regions. Chen *et al.* [103] used random speckle patterns as a calibration target and the digital image correlation method to extract the corresponding points. This method provides a large number of control points (up to 10e3) to perform the calibration, improving the calibration accuracy. Ramírez *et al.* [104] used a triangulation method to calculate 3D coordinates from the corresponding 2D points, which performed an optimization and obtained a formulation to compensate for the errors caused by image digitalization and lens distortion when other camera parameters were omitted.

In the field of endoscopy, the accuracy of 3D reconstruction is important, even at the appropriate expense of operational convenience and computational efficiency. Advanced optimization algorithms [105] and new projection or distortion models are also desirable [106], [107].

b: DEPTH ESTIMATION

Monocular Depth Estimation: Reconstruction of the tissue surface with a monocular endoscope does not have a simple or direct solution, but it is attractive because 3D structural information can significantly improve diagnostic and surgical performance. Such reconstruction was initially implemented with the traditional CV method, structure from motion (SfM), and SLAM technology.

The SLAM method [41], [108] can be expressed by (9):

$$\begin{cases} x_k = f(x_{k-1}, u_k, w_k) \\ z_{k,j} = h(y_j, x_k, v_{k,j}) \end{cases} \quad (9)$$

where f is the motion function and h is the observation function. In the case of endoscopy, x_k denotes the location of the endoscope; u_k denotes the pose transformation between x_{k-1} and x_k ; y_j is the 3D location of key points on the tissue, $z_{k,j}$ is a photo of the tissue snapped by the endoscope at x_k , w_k and $v_{k,j}$ is noise. We can build a series of functions f and h as the endoscope is moved in the space of interest and obtain the 3D tissue structure and the motion of the endoscope by solving these functions.

Early works used general, well-performing SLAM algorithms [43] or those designed to be suitable for endoscopic conditions [12]. Mahmoud *et al.* [44] and Qiu and Ren [109] used methods based on ORB-SLAM.

For deformation, Turan *et al.* [110] used Surfel-based scene reconstruction to cope with non-rigid deformations throughout the frame sequence. Mahmoud *et al.* [45] segmented video frames into clusters during live tracking, outperforming pure stereo methods.

If we have only the equations of motion in (7)—in other words, given photos without order—this becomes an SFM problem, which is suitable for a specific space. Recently, Sun *et al.* [111], Widya *et al.* [112] first used chromoendoscopy video after spreading indigo carmine (IC) dye on the stomach surface to reconstruct the whole 3D shape of the stomach with details of the mucosal surface by the SFM method and found that the red channel data provided the most complete and densest result.

A learning-based method used for monocular depth estimation was first proposed by Eigen *et al.* [113]. Since an accurate depth value is difficult to obtain in large-scale training, an unsupervised frame was developed that could use spatial (between left-right pairs) or temporal (forward-backward pairs) photometric warp error [114]–[116] or both [117]. The latest research [118] sought to obtain clear boundaries in estimating depth by resampling pixels around occlusion boundaries. One obstacle in training is the patient-specific nature of the texture of the tissue when first used in depth reconstruction in colonoscopy [119]. Early works taught their models on a per-patient basis [120], [121], while some later works used a synthetic training set [122], [123] or a generative adversarial network (GAN) [124] with a generator and discriminator pre-trained on real video frames. Unsupervised attempts have also been made [47], [125]–[127]. CT or video-CT registration data are often used as ground truths to test trained networks.

Recently, SFM/ SLAM and learning methods, especially GAN, were combined to obtain acceptable results [128], [129]. Rau *et al.* [124] proposed an unsupervised learning-based approach for estimating depth and motion with monocular endoscopy videos. This work jointly trained an EndoMotionNet and an EndoDepthNet, and the depth information of the previous frame can be used as a priori information for obtaining the depth information of the next frame. Widya *et al.* [128], Chen *et al.* [129] expanded their prior work by training a network for image-to-image style translation from a no-IC image to an IC-sprayed image. Therefore, whole stomach 3D reconstruction can be performed without the need for IC dye. In poorer environments, such as during colonoscopy, the paucity of distinguishing features and tissue homogeneity can lead to insufficient features for tracking or dense reconstruction. To address this, Widya *et al.* [128], Chen *et al.* [129] used an adversarial CNN to predict depth from monocular endoscopy images and proposed a framework for performing SLAM endoscopy by fusing monocular RGB images with corresponding depth predictions.

Although monocular methods are most convenient when they lack other attachments, the images obtained for the motion of the monocular camera, zooming trail, and scene are

the same for the same multiplier (since the epipolar constraint is equal to 0 [130]). Therefore, we cannot obtain the scale of the object with monocular methods.

Stereo Matching: For binocular endoscopy, we can obtain the 3D structure of the tissue surface along with the scale by the stereo matching method with a pre-determined baseline distance. Traditional stereo matching methods [131]–[133] have been proposed to match corresponding points between two images, usually with a four-step pipeline—cost calculation, cost aggregation, parallax calculation, and post-processing—in order to minimize function (10):

$$E(d_x) = C(x, y, d_x) + \sum_{x, y \in N_{x, y}} E_s(C(x, y, d_x)) \quad (10)$$

where C is the cost of pixel $I(x, y)$ having a disparity of d_x , E_s is a cost aggregation term calculated according to a set of pixels $N_{x, y}$ around $I(x, y)$ that is used to impose constraints of smoothness, consistency between two images, etc. The task of stereo matching is to find d_x for all pixels $I(x, y)$.

The stereo matching method was first used in endoscopic research to localize two coronary arteries on a beating heart [134]. Some early works designed a matching method strategy applicable to the minimally invasive surgical environment [135], [136] as well as a GPU-enhanced strategy [137]. The attractive results of the application of learning methods in stereo matching led to their natural transfer to binocular endoscopy [35]–[38]. Due to the lack of labelled data, a self-supervised network based on ideas from the non-endoscope area was used. Ye *et al.* [138] proposed a network that optimized the combination of reconstruction loss and disparity consistency loss. Armin *et al.* [139] took the motion of two image points and visible areas into consideration. Luo *et al.* [51] used a simple architecture with an encoder-decoder network to predict two disparity maps and then estimated the original images. Wang *et al.* [14] built a precise GI environment via software to provide training data for their network.

Recently, Chen *et al.* [140], Zhou *et al.* [141] provided a comprehensive approach that includes enlarging the radius of the constraint of low texture regions, screening inliers from feature matching results with a histogram voting-based method, and tracking camera motion with a novel random sample consensus-based (RANSAC-based) perspective-n-point (PnP) algorithm. Luo *et al.* [51] presented an AR-assisted navigation system for liver resection based on stereoscopy. In their study, an unsupervised convolutional network was used to estimate the depth and generate an intraoperative 3D liver surface. A V-Net architecture was also used to segment volumetric image data from preoperative CT images. Finally, the preoperative model and intraoperative model were registered to the same space by a globally optimal iterative closest-point algorithm, providing detailed information during surgery. Due to the lack of binocular endoscopic images with ground truths, Wang *et al.* [14] built a precise GI environment with software to provide training data for their network. They proposed a 23-layer deep CNN

to generate depth from stereo images and introduced a novel scale-invariant loss function to increase precision.

Because of the limited diameter of the endoscope, the binocular design is only used laparoscopically. Both monocular and binocular endoscopy encounter similar obstacles that limit the performance of depth estimation: (i) a limited field of view and low signal-to-noise ratio, (ii) complicated reflections and weak textures, and (iii) deformation of soft tissue.

2) FUSION WITH OTHER MODALITIES

Although endoscopy can provide clear images of the internal organs and tissues of patients, some limitations remain. First, it is difficult to accurately locate and differentiate the target tissue (for example, distinguishing cancerous from normal tissues or assessing the distribution of the subcutaneous blood vessel network) with endoscopic images alone. Second, doctors typically take much time to predict the real-time position of the endoscope with video, which makes it difficult to shift the endoscope to the target location. Fortunately, many detection options exist, such as CT, MRI, and ICG. By fusing other modalities with endoscopic images, the insufficiency of the endoscope can be improved, leading to better performance in the clinic.

a: MULTIMODAL DATA

Multimodal data mainly come from preoperative and intraoperative medical images. The former are mainly obtained from CT and MRI, while the latter involves CBCT, ICG, ultrasound, iMRI, etc.

CT and MRI are considered morphological imaging modalities, that is, imaging of the appearance of the tissue structure. CT is a type of high-density imaging that involves penetrating the human body with X-rays and obtains images according to the difference in the absorption rate of radiation by different tissues of the human body and tomographic images. MRI is a type of tomography that obtains human body structure information by magnetic resonance phenomena. Compared with CT, MRI can achieve better resolution of soft tissues. Through the segmentation and rendering of CT and MR images, we can obtain a 3D model of the target tissue and its anatomical structure, which is the primary method to obtain surgical navigation data. This process is called anatomical reconstruction.

Compared with preoperative data, intraoperative data can provide real-time anatomical information and observations that can enable physicians to make an accurate diagnosis and administer interventional treatments.

Cone-beam computed tomography (CBCT) systems can be used intraoperatively by obtaining data using a cone-shaped X-ray beam rotated around the patient. The resulting signals can then be used to reconstruct the 3D structure of the patient's anatomy.

Indocyanine green (ICG) is a type of clinically fluorescent dye that is injected into veins for NIR fluorescence imaging of human tissues.

Endoscopic ultrasound refers to the use of a miniature ultrasound probe installed on the top of the endoscope to directly detect the mucosal surface of the digestive tract and its lesions. This technique obtains the histological features of all levels of the digestive tract wall, intracavitary lesions, and surrounding tissue with the help of ultrasound scanning methods. Mainly used for disease diagnosis, endoscopic ultrasound can significantly improve upon the resolution of ultrasound and display lesions at the end of the common bile duct and the head of the pancreas in the deep part of the abdominal cavity, resulting in increased image resolution and accuracy of endoscopic and ultrasonic diagnosis.

Intraoperative magnetic resonance imaging (iMRI) has demonstrated a significant benefit in surgery. Additionally, fast MRI techniques have been developed for clinical real-time requirements and have great application potential for fusion with endoscopic images.

b: FUSION WITH PREOPERATIVE DATA

Compared with open surgery, endoscopic-based MIS has great advantages. However, surgeons often have to operate within the narrow field of view of the endoscope during MIS. Preoperative data, such as CT and MR images, contain anatomical information that could enhance the stereo perception and inspection of the visual field of the endoscopic examination. Many studies have emerged that mainly focused on accurate matching between preoperative models and the intraoperative scene. Dey *et al.* [142] fused endoscopic video images obtained from a tracked neuro-endoscope to polygonal surfaces extracted from preoperative CT images of a standard brain phantom to assist in surgical planning, training, and guidance. Thoranaghate *et al.* [143] developed an AR endoscope system for paranasal and transnasal surgeries based on landmarks defined by the surgeon in preoperative CT volumes, which were augmented onto the endoscopic video stream and helped in orienting and navigating the surgical field. The deformation correction of the middle registration model effectively reduced the model registration error under large-scale deformation by 30 mm [144].

On the one hand, preoperative CT/MR images have high resolution and quality. On the other hand, due to the inability to intraoperatively image tissue deformation, real-time imaging of the tissue interior cannot be performed, which increases the complexity and risk of navigating during endoscopic surgery.

c: FUSION WITH INTRAOPERATIVE DATA

Because of the deformation of non-rigid tissues such as the liver, some scholars have studied the registration of real-time intraoperative images and endoscopic images to guide operations in the corresponding parts of the body. Methods for the fusion and registration of intraoperative CBCT images with endoscopic video have been established. Prisman *et al.* [145] developed custom software to integrate intraoperative CBCT images with endoscopic video for surgical navigation and guidance and tested it on a cadaveric head. Lai *et al.* [15]

presented a method based on a surgical navigation system (Philips, Eindhoven, NL) for intraoperative CBCT image fusion on endoscopic views for endonasal skull-base surgery, allowing submillimetre accuracy. Clinically approved fluorescent dyes, such as fluorescein and ICG, are also useful in intraoperative guidance. Harada *et al.* [146] developed a method for visualising the placental vascular network in detail in the presence of turbid amniotic fluid by a rigid-type fluorescence endoscope coupled with intravenous administration of ICG. Intraoperative ultrasound images are also helpful. For example, the navigation system under the da Vinci surgical robot from Intuitive Surgical uses a binocular stereoscopic laparoscope to track the movement of an abdominal ultrasound probe in real time. After positioning and registering the separate laparoscope and ultrasound probes, the resulting 2D ultrasound image and optical image are fused and presented. Fast MRI techniques have also been developed that can achieve acquisition times down to 139 ms per slice [147]. As a result, the fusion of iMRI data and endoscopic images in real time will be a meaningful research direction in the future.

The fusion of CBCT images with endoscopic video can be of great benefit for intraoperative registration accuracy. However, fusion with endoscopic images still cannot be performed in real time. With the help of fluorescent dyes, such as ICG, fluorescence endoscopy can visualize the vascular network with high resolution. However, the detection depth of this method can only reach 10 mm under the skin. In contrast, endoscopic ultrasound can obtain deep structure information beneath the surface of human tissue, but it has poor image resolution.

In summary, the fusion of images from different modalities with endoscopic images has different effects resulting from their relative advantages and disadvantages. Fusion with preoperative data has a substantial guiding effect on MIS navigation, but it cannot solve the problem of intraoperative tissue deformation. Intraoperative fusion also has great potential value, but there remains much room for improvement before it can be implemented in real time. Multimodal data fusion is expected to be one of the main directions for the future development of endoscopic navigation systems.

C. SUSPICIOUS LESION TRACKING

Endoscopes are standard tools for detecting and locating lesions. In minimally invasive surgery, quickly and accurately locating lesions has become a key research issue for current computer-aided diagnosis (CAD) systems. Suspicious lesion tracking technology has gradually become one of the core components of endoscopic navigation systems [148], especially for the GI system. Based on white-light imaging (WLI), contrast-enhancement techniques can be used to enhance the detectability of lesions. Magnified observation techniques can improve the resolution of the object image and allow visualization of microscopic structures. These techniques provide a basis for accurately locating lesions in endoscopic navigation systems. Within the field of endoscopic

suspicious lesion tracking, CAD algorithms can be generally divided into two classes: computer-aided detection (CADE) and computer-aided diagnosis (CADx) [149].

1) COMPUTER-AIDED DETECTION

CADE is a subfield of CAD in which the aim is the automatic detection of lesions missed by the endoscopist [149]. To date, different CADE systems for lesions in different parts of the GI tract have been developed. For the oesophagus, Horie *et al.* [16] designed a CNN-based CADE algorithm for conventional WLI gastroscopy. They used 8428 endoscopic images of oesophageal cancer from 384 patients as a training set, obtaining a model that provided 98% sensitivity in detecting oesophageal squamous cell carcinoma (ESCC). Guo *et al.* [150] developed a model based on SegNet for the automated detection of precancerous lesions and early oesophageal squamous cell carcinomas (ESCCs) in real time. They used a total of 6473 NB images, including those of precancerous lesions, early ESCCs, and noncancerous lesions, as a training set, and tested their algorithm using a dataset containing 1480 malignant NBI images from 59 consecutive cancerous cases and 5191 noncancerous NB images from 2004 cases. The authors ultimately achieved a sensitivity of 98.04% and a specificity of 95.03% with their algorithm. They also used video datasets of precancerous lesions or early ESCCs, including 27 non-magnified videos and 20 magnified videos, to evaluate their model. The model was capable of processing at least 25 frames per second with a clinically meaningful latency of less than 100 ms in real-time video analysis. For the stomach, Hirasawa *et al.* [151] developed an SSD-based system for gastroscopy capable of utilizing a WLI, NBI, or CE modality. They used a dataset of 51558 images to train a CNN model and tested it on another dataset of 2296 images, achieving a sensitivity of 92.2% in the detection of gastric cancer (GC). Itoh *et al.* [152] developed an algorithm based on GoogleNet for detecting *Helicobacter pylori* infection. They used a total of 179 upper gastrointestinal endoscopy images obtained from 139 patients to train and test their model, achieving a sensitivity and specificity of 86.7% and 86.7%, respectively, and an area under the curve of 0.956. In colonoscopy, Wang *et al.* [149] developed a CNN model based on SegNet for conventional WLI colonoscopy. They used 3634 static images with polyps and 1911 static images without polyps as a training set and tested with 3 testing datasets: set A (5541 static images with polyps and 21,572 static images without polyps), set B (29 public videos with polyps), and set C (54 videos without polyps). Across all testing datasets, the model achieved a sensitivity of over 90% in the detection of polyps.

Although these CADE algorithms achieved good results, many problems remain. First, the current indicators for evaluating algorithm performance are technically simplistic. However, to reflect the actual effect of a CADE algorithm, some clinical measurement metrics need to be created, such as the adenoma detection rate (ADR) and polyp detection rate, in colonoscopy. In addition, most current research on CADE

algorithms is retrospective, and their effect on real-time video has not been well verified. We also note that several prospective CADE studies have been performed recently in the field of colon polyp detection [153], [154]. However, similar studies for other lesions need to be verified.

2) COMPUTER-AIDED DIAGNOSIS

CADx is a subfield of CAD also referred to as an optical biopsy, focusing on the prediction of histology results without the need for tissue biopsy. CADx is performed based on CADE; if a lesion is missed by CADE, then it would be irrelevant to a CADx system [149]. Unlike CADE, CADx always uses endoscopy images with advanced optical modalities such as NBI, ME, CE to train CNNs to perform histological diagnosis. In the oesophagus, Quang *et al.* [17] constructed a CADx system based on a 2-class linear discriminant algorithm for high-resolution microendoscopy (HRME). Using 104 biopsy-sampled sites from 54 patients as a training set and 104 biopsy-sampled sites from 45 patients as a test set, this CADx algorithm achieved a sensitivity and specificity of 95% and 91%, respectively, in evaluating the malignancy of oesophageal squamous cell neoplasia. In the stomach, Zhu *et al.* [155] developed a CADx algorithm based on ResNet for conventional WLI gastroscopy. A dataset of 790 conventional endoscopy images was used for training the network, and an independent dataset of 203 images was used for testing its performance. The results showed 95.56% specificity in the discrimination of SM2 status or worse from M/SM1, relative to the invasion depth of the gastric cancer and significantly higher than that of expert endoscopists. Some CADx systems have also been developed for colonoscopy inspection. Chen *et al.* [156] used a pre-trained CNN based on a natural image dataset and fine-tuned the classifier layer with a dataset composed of 1476 NB images with neoplastic polyps (NPs) and 681 NB images with hyperplastic polyps (HPs) to build a CADx system for differentiating the two kinds of polyp. Then, they tested the model using another colonoscopy NB image dataset, in which 188 images were NPs and 96 images were HPs. The system achieved 96.3% sensitivity and 78.1% specificity in differentiating NPs from HPs. Takenaka *et al.* [157] developed a model for evaluating the status of ulcerative colitis (UC) based on Inception v3. They used a training dataset of 40578 images to train their model and tested it using a dataset of 4187 images. The algorithm identified patients with endoscopic and histologic remission with 90.1% and 92.9% accuracy, respectively, and required less than 0.2 seconds to obtain a result from the input of an endoscopic image using a GTX 1080 Ti GPU.

In recent studies, researchers combined CADE algorithms with CADx algorithms to develop clinically meaningful integrated systems that perform sequential detection and diagnosis [158]. Moreover, some researchers are working on developing a computer-aided monitor (CADm) system, which can perform quality control during endoscopy [159]. For example, one CADm system can monitor the time of caecal incubation and withdrawal speed during colonoscopy.

This would allow physicians to perform more detailed examinations and minimize the number of blind spots [160]. Although CADx systems are capable of producing satisfactory results on many types of lesions, some issues remain. The performance of CADx systems is overly reliant on advanced endoscopic imaging modalities such as NBI, CE, and HRME; under the most common endoscopy modality, WLI, the performance is not as acceptable.

IV. APPLICATION

According to a study published by The Lancet in 2020 on global diseases from 1980 to 2017, cardiovascular disease and cancer were listed as the main causes of death from non-communicable diseases [161]. For example, in 2017, cardiovascular diseases accounted for 43.3% of deaths caused by non-communicable diseases, and cancer accounted for 23.3%. Among the causes of cancer deaths, lung cancer, gastrointestinal cancer, and liver cancer accounted for more than half; brain tumours and urinary system cancers also account for high proportions of deaths. Endoscopy is widely used in departments focused on the abovementioned diseases. Combined with advanced endoscopic navigation technology, endoscopy can play a vital role in clinical and pre-clinical applications (in Table 1).

A. CARDIOVASCULAR

Endoscopy plays a very important role in the field of cardiovascular disease treatment and diagnosis. Angioscopy enables the macroscopic pathological diagnosis of cardiovascular diseases from the arterial wall [162]. Multimodal scanning fibre endoscopy has been reported to identify thrombotic lesions and define the risk of plaque rupture in the diagnosis of atherosclerosis [163], [164]. The method is combined with fluorescence emissions to classify atherosclerotic plaques in different stages of development. Intravascular photoacoustics is considered a promising imaging modality for atherosclerosis, as it could provide chemical-specific optical information on the arterial wall [165]. The cardio-endoscopic technique is a useful adjunct in the resection of left ventricular tumours and thrombi [166]. A cardiovascular virtual endoscopy system was applied to support a diagnosis of congenital heart disease [167]. This system uses preoperative CT and MR images to reconstruct a 3D model of blood vessels and simulates the entry of an endoscope for surgical path planning.

B. RESPIRATORY TRACT

There are many types of endoscopes used in the diagnosis and treatment of diseases in the respiratory tract, such as sinusscopes, laryngoscopes, and bronchoscopes. In sinuscopic navigation, self-supervised learning for dense depth estimation with a monocular endoscope has achieved good results in the absence of CT images [125], [168]. For laryngoscopic navigation, in combination with a tracking sensor, Wu *et al.* [169] utilized deformation modelling to improve accuracy and reliability in laryngoscopic surgery. Yamamoto *et al.* [170]

TABLE 1. Clinical and pre-clinical endoscopic navigation system using those above technologies in the different regions of the human body.

Device	Refs.	Object	MOET	IT	EVE	SLT	Purpose
Cadiovascular							
Angioscope	Luis E. Savastano et al., [164]	22 fresh carotid arteries and 1 vertebral artery	AFI; CLE	NA	Fusion with multimodal autofluorescence images	NA	Diagnosis, prognosis, and image-guided therapy for atherosclerosis
	Li-Ping Yao et al., [167]	40 children with congenital heart diseases	NA	NA	Fusion with CT	CADe	Automatically planned path for diagnosis of congenital heart diseases
Respiratory tract							
Sinuscope	Ayushi Sinha et al., 2018 [168]	Statistical shape models from 53 head CTs; Videos of the nasal cavity from patients	SME	Video-based camera tracking	Monocular depth estimation	NA	Navigation during clinical endoscopic exploration in the absence of CT
	Xiongbiao Luo et al., 2019 [199]	Preoperative images, endoscopic video sequences, and ultrasound images	SME	Video-based camera tracking; Image-based Ultrasonic vibrator tracking	Fusion with CT and US	NA	Simultaneously navigate a flexible endoscope and US without using external tracking devices
	Sajad Majid Qazi et al., 2020 [200]	20 patients with the specific inclusion and exclusion criteria	ETS	Optical-based camera and surgical tools tracking	Fusion with CT	NA	The first experience of using image-guided endoscopic sinus surgery in Kashmir valley
Laryngoscope	Xiaotian Wu et al., 2019 [169]	Preoperative and intraoperative images from 15 patients	ETS	EM-based camera tracking	Fusion with iCT and deformation model	NA	Utilizing intraoperative imaging and deformation modeling to improve surgeon accuracy and confidence
	Clyde Matava et al., 2020 [201]	775 video laryngoscopy and bronchoscopy videos	SME	NA	NA	CADe	Classification of vocal cords and tracheal airway anatomy in real-time
Bronchoscope	Belanger, Adam R et al., 2019 [202]	108 patients for the diagnosis of peripheral lung lesions and/or the placement of FMs for stereotactic body radiotherapy	ETS	EM-based surgical tools tracking	Fusion with CT	NA	Peripheral lung nodule diagnosis and fiducial marker placement with Veran™ system
	Xinqi Liu et al., 2020 [203]	Videos of Bronchoscope	ETS	EM-based camera tracking	Monocular depth estimation	NA	Utilizing CNN-generated depth images to localize the bronchoscope
	D Kyle Hogarth et al., 2018 [172]	113 patients from the LungVision Clinical Trial	ETS	Fluoroscopic tracking	Fusion with CT	NA	Utilizing augmented fluoroscopy to peripheral lung nodules in real-time
	Qi Chang et al., 2020 [173]	39,899 video frames constituting four lung-cancer patients	AFI	NA	NA	CADx	Early frame detection of bronchial lesions
Gastrointestinal tract							
Esophagoscope	Shi-Lun Cai, MD et al., 2019 [204]	2428 (1332 abnormal, 1096 normal) esophagosopic images from 746 patients	SME	NA	NA	CADe	Localization and Identification of early ESCC
	Yuan-Yuan Zhao et al., 2019 [180]	1350 images from 282 patients with a suspicious esophageal condition	EC	NA	NA	CADx	Automated classification of ESCC
	Joost van der Putten et al., 2019 [178]	Images of 40 subtle early neoplastic BE lesions and 20 non-dysplastic BE patients	BLI	NA	NA	CADx	Detection of early neoplasia in Barrett's esophagus (BE)
	Youichi Kumagai et al., 2019 [181]	4715 endocytoscopic system images of the esophagus	SME	NA	NA	CADx	Supporting endoscopists for the replacement of biopsy-based histology
Gastroscope	Jiquan Liu et al., 2015 [174]	Phantom model produced by 3B Scientific Company	ETS	EM-sensor based camera tracking	Fusion with CT	NA	Developing an image panoramic system in the diagnosis of gastric disease
	Aji Resindra Widya et al., 2020 [128]	Gastrosopy videos	VCE	Video-based camera tracking	Monocular depth estimation	NA	Tracking the lesion in the global view of the stomach
	Xu Zhang et al., 2019 [205]	354 images from 215 patients	SME	NA	NA	CADe	Detection of gastric polyp in real-time
	Rie Miyaki et al., 2013 [182]	Images from 46 patients	ME-FICE	NA	NA	CADx	Differentiation of cancerous areas from non-cancerous areas
	Hiroya Ueyama et al., 2020 [183]	5574 images for training and validation, 2230 images for testing	ME-NBI	NA	NA	CADx	Diagnosis of early gastric cancer
Hirotaaka Nakashima et al., 2018 [206]	Image from 222 patients currently infected or uninfected with H.pylori	BLI; LCI	NA	NA	CADx	Prediction of H. pylori infection	
WCE	Mehmet Turan et al., 2018 [59]	Ex-vivo porcine stomach models	ETS	Hybrid camera tracking	Monocular depth estimation	NA	Real-time 3D reconstruction and localization
	Kutsev Bengisu Ozyoruk et al., 2020 [175]	42,700 frames from 8 ex-vivo porcine GI-tract organs	ETS	Video-based camera tracking	Monocular depth estimation	NA	Endoscopic VO and depth estimation and available datasets
	Mohammad Wajih Alam et al., 2020 [207]	Porcine intestine sprayed by various concentrations of fluorescein solution	AFI	Sensor-based camera tracking	NA	NA	Early detection of colorectal cancer
Colonoscope	Jianjun Yang et al., 2020 [208]	500 normal images including 500 polyps	SME	Sensor-based camera tracking	NA	CADe	Detection of polyp
	Gwangbin Bae et al., 2020 [176]	Colonoscopy videos	SME	Video-based camera tracking	Monocular depth estimation	NA	3D reconstruction from monocular endoscopic images in real-time
	Yuchen Luo et al., 2020 [209]	112,199 images with 69,716 polyps in 150 patients	SME	NA	NA	CADe	Detection of colorectal polyp

TABLE 1. (Continued.) Clinical and pre-clinical endoscopic navigation system using those above technologies in the different regions of the human body.

	Mamoru Tokunaga <i>et al.</i> , 2020 [184]	3442 images from 1035 consecutive colorectal lesions	SME	NA	NA	CADx	Prediction of invasion depth in colorectal cancer
	Saad Nadeem and Arie Kaufman 2016 [119]	Optical colonoscopy images	SME	NA	Monocular depth estimation	CADe	Detection of colorectal polyp
	Eun Mi Song <i>et al.</i> , 2020 [185]	1352 images of 1169 polyps in 951 patients	ME-NBI	NA	NA	CADx	Prediction of colorectal polyp histology
	Hideka Horiuchi <i>et al.</i> , 2019 [186]	429 polyps of 95 patients from AFI scope and 258 polyps of 95 patients from macroscopic type	AFI and ME-AFI	NA	NA	CADx	Diagnosis of diminutive rectosigmoid polyp using AFI in real-time
	Yasuharu Meada <i>et al.</i> , 2019 [210]	22835 images of 187 patients	EC	NA	NA	CADe and CADx	Prediction of histologic inflammation in real-time
	Min Min <i>et al.</i> , 2019 [211]	139 images of adenomatous polyps and 69 images of non-adenomatous polyps	LCI	NA	NA	CADx	Prediction of histological results of polyps
	Daniela Ștefănescu <i>et al.</i> , 2016 [212]	1035 artifact-free endomicroscopy images	CLE	NA	NA	CADe and CADx	Automatic diagnosis of colorectal cancer
Chest and Abdomen							
Thoracoscope	Takashi Anayama <i>et al.</i> , 2015 [187]	A resected porcine lung	ICG; ETS	EM-based camera tracking	Fusion with ICG	NA	Localization of Pulmonary Nodules
	Chang Young Lee <i>et al.</i> , 2018 [188]	A realistic tumor phantom	ETS	Optical-based surgical tools tracking	Fusion with CBCT	NA	Real-time guidance for chest wall tumors
Laparoscope	Takeshi Aoki <i>et al.</i> , 2020 [190]	27 patients underwent laparoscopic partial liver resection	SME	EM-based surgical tools tracking	Fusion with IOUS and CT	NA	Real-time anatomical feedback during hepatic resections
	Pascale Tinguely <i>et al.</i> , 2017 [189]	54 patients with 346 lesions	SME	Optical-based surgical tools tracking	Fusion with CT	NA	Microwave ablation for malignant liver tumors
	Gian Andrea Prevost <i>et al.</i> , 2019 [191]	10 patients with 18 lesions	BE	Optical-based camera and surgical tools tracking	Fusion with CT	NA	To investigate efficiency, accuracy, and clinical benefit of AR-3D navigations system
	Peng Zhang <i>et al.</i> , 2019 [213]	64 patients underwent laparoscopic hepatectomy	ICG	Video-based camera tracking	Fusion with CT and ICG images	NA	Achieve real-time surgical navigation
Brain							
Neuroendoscope	Arthur Wang <i>et al.</i> , 2019 [171]	50 ommaya reservoir patient	SME	EM-based surgical tools tracking	NA	NA	Placing Ommaya reservoirs with the zero-error precision protocol
	P Tabakow <i>et al.</i> , 2019 [214]	21 patients with symptomatic arachnoid cysts	SME	NA	Fusion with iMRI	NA	Demonstrating for the usefulness of intraoperative cysternography with iMRI
	T. Finger <i>et al.</i> , 2017 [215]	A 3D-printed head model with a right parietal lesion	ETS	Optical -based camera and surgical tools tracking	Fusion with CT	NA	AR in intraventricular neuroendoscopy
	Simon Drouin <i>et al.</i> , 2017 [216]	NA	ETS	Optical -based camera tracking	Fusion with IOUS	NA	An open-source platform for image-guided neurosurgery
Cardiovascular							
Angioscope	Luis E. Savastano <i>et al.</i> , [164]	22 fresh carotid arteries and 1 vertebral artery	AFI; CLE	NA	Fusion with multimodal autofluorescence images	NA	Diagnosis, prognosis, and image-guided therapy for atherosclerosis
	Li-Ping Yao <i>et al.</i> , [167]	40 children with congenital heart diseases	NA	NA	Fusion with CT	CADe	Automatically planned path for diagnosis of congenital heart diseases
Urinary tract							
Ureteroscope	Kenji Yoshida <i>et al.</i> , 2019 [217]	3D pyelocaliceal system model	SME	EM-based surgical tools tracking	Fusion with CT	NA	An experimental ureteroscopic navigation system
	Wesley W. Ludwig <i>et al.</i> , 2018 [192]	30 kidney stone fragments	SME	Image-based surgical tools tracking	NA	NA	Measurement of kidney stones fragments
	Ali H Aldoukhi <i>et al.</i> , 2019 [193]	128 images with 63 kidney stones	SME	NA	NA	CADe	Detection of kidney stone
Others: Hysteroscope, Cystoscope, Nephroscope, Arthroscope, and Otoscope							

developed a method for the transoral removal of schwannoma using a laryngoscope with the help of nerve integrity monitoring (NIM) and NBI. Endoscopic navigation systems have also been widely used in bronchoscopic surgery [171]. Many commercial platforms, such as superDimension™ and Veran™, can help guide surgery by registering CT images reconstructed before surgery with intraoperative instruments, and a large number of experiments have been conducted with these systems in the clinic. Recently, early clinical trials have been conducted utilizing fluoroscopy with novel algorithms in the LUNGVISION™ system [172]. Chang *et al.* [173] utilized a CADx system with 39899 AFI video frames to detect bronchial lesions.

C. GASTROINTESTINAL TRACT

In GI navigation, some systems similar to the superDimension™ bronchial navigation system have been tested in the early stages of development [174]–[176]. However, due to the large deformation of the stomach and its relatively simple shape, this type of system is used for approximated tracking. However, CAD technology has been widely used in the GI tract [177] and has been combined with many advanced imaging technologies in clinical trials. In the upper GI tract, CAD has been used in the detection of early neoplasia in Barrett’s oesophagus (BE). Putten *et al.* [178], [179] proposed a patch-based deep learning algorithm combining the WLI and BLI modalities for early neoplasia in BE.

In their research, CAD was used for the multi-frame analysis of volumetric laser endomicroscopy (VLE) images from oesophageal wall layers, improving BE neoplasia detection. In other studies, a microendoscope with narrow-band imaging (ME-NBI) and EC were used in the classification of ESCC [180], [181]. FICE and ME-NBI were used in the diagnosis of GC [182], [183]. BLI and LCI were employed in the prediction of *H. pylori* infections. In the lower GI tract, CAD systems have been widely used in the detection and characterization of polyps. In reference [184], 112199 images of 69716 polyps from 150 patients were used in the detection of colorectal polyps. ME-NBI, AFI, ME-AFI, EC, LCI, and CLE have also been employed in the diagnosis of polyps [185], [186].

D. CHEST AND ABDOMEN

Laparoscopic surgery is the most popular MIS, and surgical navigation technology for the thoracic and abdominal cavities has been continuously improved in recent years. However, due to the dynamic, unstructured nature of the abdominal cavity environment, the abdominal soft tissue experiences large deformations and displacements during surgery, which remains the major challenge in accurate navigation. In thoracoscopy, Anayama *et al.* [187] combined ICG injections, intraoperative NIR fluorescence thoracoscopic detection, and EM tracking sensors and achieved acceptable results in locating pulmonary nodules in a resected porcine lung. Lee *et al.* [188] developed a thoracoscopic surgical navigation system utilizing CBCT and optical tracking sensors to guide the search for chest wall tumours in real time, significantly improving the localization accuracy for tumour margins. In laparoscopy, optical tracking sensors combined with CT were used to guide the microwave ablation of liver tumours for 54 patients, which resulted in low complication rates and thus a feasible option for patients with hepatic disease [189]. In laparoscopic hepatectomy, intraoperative ultrasonography (IOUS) and ICG were employed and fused with CT to achieve real-time surgical navigation during hepatic resections on 27 and 64 patients, respectively, achieving high efficiency and accuracy [190]. In addition, binocular endoscopy was employed in an AR-3D navigation system [191] that merged preoperative cross-sectional imaging with laparoscopic images in liver resection, verifying the feasibility of AR-3D laparoscopic liver surgery with landmark-based registration.

E. URINARY TRACT

In ureteroscopy, endoscopic navigation is usually applied to detect, measure, and track kidney stones. At present, it is difficult to accurately detect kidney stones and measure their size during flexible ureteroscopy. Consequently, Ludwig *et al.* [192] developed novel intraoperative software to measure the size of stone fragments based on an external fixed-length tool, the accuracy and precision of which reached 0.17 and 0.15 mm, respectively. Aldoukhi *et al.* [193] applied a deep learning CV algorithm

for automatically detecting the composition of kidney stones from their images with high accuracy.

F. OTHERS

Endoscopic navigation technology is also being used in other departments. For instance, video-based navigation technology in knee arthroscopy can be used for anterior cruciate ligament reconstruction [194]. Moreover, hysteroscopy, cystoscopy, nephroscopy, and otoscopy systems have also been used in virtual reality simulators [195]–[198].

V. DISCUSSION

In recent years, endoscopic navigation systems have gradually become one of the core instruments in MIS. After a period of development, endoscopic navigation systems have effectively solved many issues encountered in the clinic. However, different clinical applications have diverse goals and encounter different challenges:

- 1) In cardiovascular endoscopy, advanced imaging technology has reached the stage of clinical trials in lesion identification. Instrument tracking technology has been widely used in vascular interventional surgery, which mostly uses radiography technology (such as ultrasound and MRI) and magnetic tracking technology. Future research should focus on ways to use endoscopic vision technology to guide and accurately track cardiovascular endoscopes.
- 2) Some clinical applications and commercial products have been developed in the application of endoscopic navigation to the respiratory tract. The branched structure of the respiratory tract is complex and narrow, and therefore the main concern in the endoscopic navigation of the respiratory tract is the guidance of the endoscope through the complex branches to reach the target, such as a biopsy location at the distal end of the bronchus. Currently, the most commonly used navigation system for the respiratory tract is based on electromagnetic tracking technology, which has high accuracy but is expensive. Navigation methods based on endoscopic vision technology are attractive, low-cost solutions. As developments in deep learning technology continue, this method will increasingly become the focus of research.
- 3) The main goal of digestive endoscopy navigation systems at present is to help doctors comprehensively and accurately detect and characterize lesions in the gastrointestinal tract. This raises three key questions: how can the endoscope be made to traverse the entire gastrointestinal system completely, how can advanced sensor technology be used to obtain more information about the lesion, and how can the lesion be accurately detected, characterized and located? Endoscopic vision technology has shown great potential in attempting to solve these problems. Deep learning-based methods can effectively increase the detection rate of BE, ESCC,

GC, and other lesions, and can reduce the number of blind spots in gastrointestinal endoscopy.

- 4) In laparoscopic navigation, methods based on the fusion of preoperative data (such as CT and MRI) cannot cope with the large-scale deformation of intraoperative organs. Intraoperative modalities, such as IOUS, may provide real-time structural details on deep tissues, but the image resolution is low, and therefore the doctor's experience and spatial visualizing ability are required. ICG fluorescence endoscopes can perform high-resolution angiography, but the detection depth is only approximately 10 mm below the tissue. Navigation technologies focused on the multimodal image fusion of preoperative and intraoperative data could potentially solve the problem of lesion navigation under conditions of large-scale changes. The primary challenge is the high-precision registration between the multiple modalities.
- 5) Endoscopic navigation systems have been widely used in clinical treatment during neurosurgery. Because of the strict safety requirements, the accuracy of the registration between preoperative data and the endoscopic visualization is a key indicator for evaluating the endoscopic navigation system. During navigation, changes in the anatomical structures during the operation could still affect the registration accuracy; however, the stiffness of the brain tissue is relatively high. As an evolving technology, the fusion of intraoperative data, such as CBCT and iMRI, with endoscopic images can potentially reduce the impact of intraoperative anatomical structure changes on registration accuracy, which can also be improved by the real-time identification of critical anatomical structures with endoscopic images.
- 6) Ureteroscopic navigation is currently in the research and development stage. The main problem lies in determining methods to completely traverse the endoscope through the renal pelvis structure to and accurately identify, locate and measure the lesion or stone. Future research should identify ways to improve the flexibility of the ureteroscope and ways to locate it in an environment with limited textures and feature points.

Below, we summarize future issues for common technologies in endoscopic navigation:

Multiple optical endoscopy technologies provide more information than ever before. Nevertheless, these technologies are currently limited to some departments. For example, binocular endoscopes are generally only used for laparoscopy, while EC is generally only used in GI applications. This is mainly limited by the sizes of the tips of different endoscopes. Miniaturization of these modules and wider application in other departments will significantly improve disease diagnosis.

In endoscopic navigation, many sensor modules, including the EM sensor and CLE probe, must pass through the instrument channel to function, which can affect the surgeon's surgical procedures. For example, in urinary navigation surgery,

these sensors and laser fibres have to be exchanged frequently through the instrument channel. Therefore, integrating these sensors into existing endoscopes may lead to greater convenience during the surgical procedure.

In 3D reconstruction, monocular endoscopes cannot measure scale without a fixed-size reference object. Therefore, it is necessary to use binocular cameras or external positioning sensors to obtain scale information. However, this can lead to errors from just the initial calibration. High-precision camera and hand-eye calibration for endoscopes will effectively minimize these errors.

Deep learning has become an important technology in navigation systems. Today, it is extensively used in a variety of fields and has achieved feasible performance. Constructing a standard dataset with a large number of labelled endoscopic images is important for the performance of endoscopic artificial intelligence-assisted diagnosis algorithms.

Finally, in MIS, movement interference and the large-scale deformation of soft tissue remain vital issues. Many previous studies were based on registration algorithms using preoperative data. With the improvement in endoscopic imaging and fast MRI technology, combining intraoperative data to track and visualize lesions in real time is expected to be the primary research direction in the future.

VI. CONCLUSION

Endoscopic navigation provides surgeons with more accurate, effective, and reliable diagnoses and treatments in MIS. Combined with advanced endoscopic vision technology, the next generation of endoscopic navigation systems are already under development. In this paper, we summarized multiple optical endoscopy modalities, including white-light imaging, contrast-enhancement and advanced endoscopic vision technologies that are currently or will potentially be used in endoscopic navigation systems. Endoscopic optical imaging modalities include white-light imaging, contrast-enhancement techniques, and technologies of magnifying observation. Meanwhile, endoscopic vision technologies consist of instrument tracking, endoscopic view expansion, and suspicious lesion tracking. All these technologies will bring changes to future surgical navigation systems. Combining diverse and complex clinical needs, it is necessary to integrate these advanced endoscopic technologies to achieve more precise navigation.

ACKNOWLEDGMENT

The authors thank Dr. Xiao Liang at Sir Run-Run Shaw Hospital, Hangzhou, China, for helping inspire the development of experiments and this article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] S. Nicolau, L. Soler, D. Mutter, and J. Marescaux, "Augmented reality in laparoscopic surgical oncology," *Surgical Oncol.*, vol. 20, no. 3, pp. 189–201, Sep. 2011.
- [2] K. J. E. Fuchs, "Minimally invasive surgery," *Endoscopy*, vol. 34, no. 2, pp. 154–159, 2002.

- [3] X. Luo, K. Mori, and T. M. Peters, "Advanced endoscopic navigation: Surgical big data, methodology, and applications," *Annu. Rev. Biomed. Eng.*, vol. 20, no. 1, pp. 221–251, Jun. 2018.
- [4] M. Wang, D. Li, X. Shang, and J. Wang, "A review of computer-assisted orthopaedic surgery systems," *Int. J. Med. Robot. Comput. Assist. Surgery*, vol. 16, no. 5, pp. 1–28, Oct. 2020.
- [5] J. Cicienia, S. K. Avasarala, and T. R. Gildea, "Navigational bronchoscopy: A guide through history, current use, and developing technology," *J. Thoracic Disease*, vol. 12, no. 6, p. 3263, 2020.
- [6] J. Marlow, "History of laparoscopy, optics, fiberoptics, and instrumentation," *Clin. Obstetrics Gynecology*, vol. 19, no. 2, pp. 261–275, Jun. 1976.
- [7] Z. He, P. Wang, and X. Ye, "Novel endoscopic optical diagnostic technologies in medical trial research: Recent advancements and future prospects," *Biomed. Eng. OnLine*, vol. 20, no. 1, pp. 1–38, Dec. 2021.
- [8] Y. Shirai, *Three-Dimensional Computer Vision*. Springer, 2012.
- [9] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2010.
- [10] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, Feb. 2018, Art. no. 7068349.
- [11] C. Doignon, F. Nageotte, and M. De Mathelin, "Segmentation and guidance of multiple rigid objects for intra-operative endoscopic vision," in *Dynamical Vision*. Springer, 2006, pp. 314–327.
- [12] O. G. Grasa, E. Bernal, S. Casado, I. Gil, and J. M. M. Montiel, "Visual SLAM for handheld monocular endoscope," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 135–146, Jan. 2014.
- [13] E. Colleoni, S. Moccia, X. Du, E. De Momi, and D. Stoyanov, "Deep learning based robotic tool detection and articulation estimation with spatio-temporal layers," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2714–2721, Jul. 2019.
- [14] X.-Z. Wang, Y. Nie, S.-P. Lu, and J. Zhang, "Deep convolutional network for stereo depth mapping in binocular endoscopy," *IEEE Access*, vol. 8, pp. 73241–73249, 2020.
- [15] M. Lai, C. Shan, D. Babic, R. Homan, A. E. Terander, E. Edstrom, O. Persson, G. Burstrom, and P. de With, "Image fusion on the endoscopic view for endo-nasal skull-base surgery," *Proc. SPIE*, vol. 10951, Mar. 2019, Art. no. 109511D.
- [16] Y. Horie, T. Yoshio, K. Aoyama, S. Yoshimizu, Y. Horiuchi, A. Ishiyama, T. Hirasawa, T. Tsuchida, T. Ozawa, S. Ishihara, Y. Kumagai, M. Fujishiro, I. Maetani, J. Fujisaki, and T. Tada, "Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks," *Gastrointestinal Endoscopy*, vol. 89, no. 1, pp. 25–32, Jan. 2019.
- [17] T. Quang, R. A. Schwarz, S. M. Dawsey, M. C. Tan, K. Patel, X. Yu, G. Wang, F. Zhang, H. Xu, S. Anandasabapathy, and R. Richards-Kortum, "A tablet-interfaced high-resolution microendoscope with automated image interpretation for real-time evaluation of esophageal squamous cell neoplasia," *Gastrointestinal Endoscopy*, vol. 84, no. 5, pp. 834–841, Nov. 2016.
- [18] M. Kim, C. Lee, N. Hong, Y. J. Kim, and S. Kim, "Development of stereo endoscope system with its innovative master interface for continuous surgical operation," *Biomed. Eng. OnLine*, vol. 16, no. 1, p. 81, Dec. 2017.
- [19] T. Zhang, Y. Gu, X. Huang, E. Tu, and J. Yang, "Stereo endoscopic image super-resolution using disparity-constrained parallel attention," Tech. Rep., 2020.
- [20] Z. He, P. Wang, Y. Liang, Z. Fu, and X. Ye, "Clinically available optical imaging technologies in endoscopic lesion detection: Current status and future perspective," *J. Healthcare Eng.*, vol. 2021, Feb. 2021, Art. no. 7594513.
- [21] R. Liang, "Optical design for biomedical imaging," in *Book Optical Design for Biomedical Imaging*. Bellingham, WA, USA: SPIE, 2011.
- [22] B. C. Wilson and S. L. Jacques, "Optical reflectance and transmittance of tissues: Principles and applications," *IEEE J. Quantum Electron.*, vol. 26, no. 12, pp. 2186–2199, Dec. 1990.
- [23] M. Monici, "Cell and tissue autofluorescence research and diagnostic applications," *Biotechnol. Annu. Rev.*, vol. 11, pp. 227–256, Jan. 2005.
- [24] R. Liang, *Optical Design for Biomedical Imaging*. Bellingham, WA, USA: SPIE Press, 2010.
- [25] J. Borovicka, J. Fischer, J. Neuweiler, P. Netzer, J. Gschossmann, T. Ehmann, P. Bauerfeind, G. Dorta, U. Zürcher, J. Binek, and C. Meyenberger, "Autofluorescence endoscopy in surveillance of Barrett's esophagus: A multicenter randomized trial on diagnostic efficacy," *Endoscopy*, vol. 38, no. 9, pp. 867–872, Sep. 2006.
- [26] M. A. Kara, F. P. Peters, F. J. W. ten Kate, S. J. van Deventer, P. Fockens, and J. J. G. H. M. Bergman, "Endoscopic video autofluorescence imaging may improve the detection of early neoplasia in patients with Barrett's esophagus," *Gastrointestinal Endoscopy*, vol. 61, no. 6, pp. 679–685, May 2005.
- [27] G. W. Falk, "Autofluorescence endoscopy," *Gastrointestinal Endoscopy Clinics North Amer.*, vol. 19, no. 2, pp. 209–220, Apr. 2009.
- [28] W. Polom, M. Markuszewski, Y. Soo Rho, and M. Matuszewski, "Usage of invisible near infrared light (NIR) fluorescence with indocyanine green (ICG) and methylene blue (MB) in urological oncology. Part 1," *Central Eur. J. Urol.*, vol. 67, no. 2, p. 142, 2014.
- [29] E. J. Aslim, F. J. Lee, and V. H. L. Gan, "The utility of intraoperative near infrared fluorescence (NIR) imaging with indocyanine green (ICG) for the assessment of kidney allograft perfusion," *J. Transplantation*, vol. 2018, pp. 1–4, Aug. 2018.
- [30] F. Ris, R. Hompes, C. Cunningham, I. Lindsey, R. Guy, O. Jones, B. George, R. A. Cahill, and N. J. Mortensen, "Near-infrared (NIR) perfusion angiography in minimally invasive colorectal surgery," *Surgical Endoscopy*, vol. 28, no. 7, pp. 2221–2226, Jul. 2014.
- [31] K. Kuznetsov, R. Lambert, and J.-F. Rey, "Narrow-band imaging: Potential and limitations," *Endoscopy*, vol. 38, no. 01, pp. 76–81, Jan. 2006.
- [32] H. Neumann, H. N. Sen, M. Vieth, R. Bisschops, F. Thieringer, K. F. Rahman, T. Gamstätter, G. E. Tontini, and P. R. Galle, "Leaving colorectal polyps in place can be achieved with high accuracy using blue light imaging (BLI)," *United Eur. Gastroenterology J.*, vol. 6, no. 7, pp. 1099–1105, Aug. 2018.
- [33] E. Emiliani, M. Talso, M. Baghdadi, A. Barreiro, A. Orosa, P. Serviàn, P. Gavrilov, S. Proietti, and O. Traxer, "Evaluation of the Spies modalities image quality," *Int. Braz. J. Urol.*, vol. 43, no. 3, pp. 476–480, Jun. 2017.
- [34] H. Kanzaki, R. Takenaka, Y. Kawahara, D. Kawai, Y. Obayashi, Y. Baba, H. Sakae, T. Gotoda, Y. Kono, K. Miura, M. Iwamura, S. Kawano, T. Tanaka, and H. Okada, "Linked color imaging (LCI), a novel image-enhanced endoscopy technology, emphasizes the color of early gastric cancer," *Endoscopy Int. Open*, vol. 05, no. 10, pp. E1005–E1013, Oct. 2017.
- [35] H. Neumann, M. Fujishiro, C. M. Wilcox, and K. Mönkemüller, "Present and future perspectives of virtual chromoendoscopy with i-scan and optical enhancement technology," *Digestive Endoscopy*, vol. 26, pp. 43–51, Jan. 2014.
- [36] J. B. Levine, M. Grupka, and M. Parente, "Narrow band imaging-close focus colonoscopy detection of colon aberrant crypt foci levine, J. grupka, M. parente, M. Colon cancer prevention program, neag cancer center, university of connecticut health center, farmington, CT," *Gastrointestinal Endoscopy*, vol. 63, no. 5, p. AB243, Apr. 2006.
- [37] M. B. Wallace and R. Kiesslich, "Advances in endoscopic imaging of colorectal neoplasia," *Gastroenterology*, vol. 138, no. 6, pp. 2140–2150, May 2010.
- [38] Y. M. Bhat, B. K. A. Dayyeh, S. S. Chauhan, K. T. Gottlieb, J. H. Hwang, S. Komanduri, V. Konda, S. K. Lo, M. A. Manfredi, J. T. Maple, F. M. Murad, U. D. Siddiqui, S. Banerjee, and M. B. Wallace, "High-definition and high-magnification endoscopes," *Gastrointestinal Endoscopy*, vol. 80, no. 6, pp. 919–927, Dec. 2014.
- [39] R. S. Kwon, L. M. W. K. Song, D. G. Adler, J. D. Conway, D. L. Diehl, F. A. Farraye, S. V. Kantsevov, V. Kaul, S. R. Kethu, and P. Mamula, "Endocytoscopy," *Gastrointestinal Endoscopy*, vol. 70, no. 4, pp. 610–613, 2009.
- [40] R. Kiesslich, M. Goetz, M. Vieth, P. R. Galle, and M. F. Neurath, "Confocal laser endomicroscopy," *Gastrointestinal Endoscopy*, vol. 2012, no. 4, p. 715, 2014.
- [41] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.
- [42] B. Lin, A. Johnson, X. Qian, J. Sanchez, and Y. Sun, "Simultaneous tracking, 3D reconstruction and deforming point detection for stereoscope guided surgery," in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*. Springer, 2013, pp. 35–44.
- [43] N. Mahmoud, I. Cirauqui, A. Hostettler, C. Doignon, L. Soler, J. Marescaux, and J. Montiel, "ORB-SLAM-based endoscope tracking and 3D reconstruction," in *Book ORB-SLAM-Based Endoscope Tracking and 3D Reconstruction*. Springer, 2016, pp. 72–83.
- [44] N. Mahmoud, A. Hostettler, T. Collins, L. Soler, C. Doignon, and J. M. M. Montiel, "SLAM based quasi dense reconstruction for minimally invasive surgery scenes," Tech. Rep., 2017.

- [45] N. Mahmoud, T. Collins, A. Hostettler, L. Soler, C. Doignon, and J. M. M. Montiel, "Live tracking and dense reconstruction for hand-held monocular endoscopy," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 79–89, Jan. 2019.
- [46] J. M. Prendergast, G. A. Formosa, C. R. Heckman, and M. E. Rentschler, "Autonomous localization, navigation and haustral fold detection for robotic endoscopy," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2018, pp. 783–790.
- [47] M. Turan, E. P. Ornek, N. Ibrahimli, C. Giracoglu, Y. Almalioglu, M. F. Yanik, and M. Sitti, "Unsupervised odometry and depth learning for endoscopic capsule robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2018, pp. 1801–1807.
- [48] C. Wang, M. Oda, Y. Hayashi, T. Kitasaka, H. Honma, H. Takabatake, M. Mori, H. Natori, and K. Mori, "Visual SLAM for bronchoscope tracking and bronchus reconstruction in bronchoscopic navigation," *Proc. SPIE*, vol. 10951, Mar. 2019, Art. no. 109510A.
- [49] S. Thompson, D. Stoyanov, C. Schneider, K. Gurusamy, S. Ourselin, B. Davidson, D. Hawkes, and M. J. Clarkson, "Hand-eye calibration for rigid laparoscopes using an invariant point," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 6, pp. 1071–1080, Jun. 2016.
- [50] S. Lee, H. Lee, H. Choi, S. Jeon, H. Ha, and J. Hong, "Comparative study of hand-eye calibration methods for augmented reality using an endoscope," *J. Electron. Imag.*, vol. 27, no. 4, p. 1, Jul. 2018.
- [51] H. Luo, D. Yin, S. Zhang, D. Xiao, B. He, F. Meng, Y. Zhang, W. Cai, S. He, W. Zhang, Q. Hu, H. Guo, S. Liang, S. Zhou, S. Liu, L. Sun, X. Guo, C. Fang, L. Liu, and F. Jia, "Augmented reality navigation for liver resection with a stereoscopic laparoscope," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 105099.
- [52] A. Malti and J. P. Barreto, "Robust hand-eye calibration for computer aided medical endoscopy," in *Proc. IEEE Int. Conf. Robotics Automat.*, May 2010, pp. 5543–5549.
- [53] H. Ha, M. Perdoch, H. Alismail, I. So Kweon, and Y. Sheikh, "Deltille grids for geometric camera calibration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 5344–5352.
- [54] M. Lai, C. Shan, and P. H. de Wit, "Hand-eye camera calibration with an optical tracking system," in *Proc. 12th Int. Conf. Distrib. Smart Cameras*, 2018, pp. 1–6.
- [55] H. Mateen, R. Basar, A. U. Ahmed, and M. Y. Ahmad, "Localization of wireless capsule endoscope: A systematic review," *IEEE Sensors J.*, vol. 17, no. 5, pp. 1197–1206, Mar. 2017.
- [56] W. He, K. Kang, Y. Li, W. Shi, Y. Miao, F. He, F. Yan, H. Yang, H. Zhang, and K. Mori, "An improved robust hand-eye calibration for endoscopy navigation system," *Proc. SPIE*, vol. 9786, Mar. 2016, Art. no. 97861P.
- [57] Z. Wang, Z. Liu, Q. Ma, A. Cheng, Y.-H. Liu, S. Kim, A. Deguet, A. Reiter, P. Kazanzides, and R. H. Taylor, "Vision-based calibration of dual RCM-based robot arms in human-robot collaborative minimally invasive surgery," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 672–679, Apr. 2018.
- [58] K. Pachtrachai, F. Vasconcelos, F. Chadebecq, M. Allan, S. Hailes, V. Pawar, and D. Stoyanov, "Adjoint transformation algorithm for hand-eye calibration with applications in robotic assisted surgery," *Ann. Biomed. Eng.*, vol. 46, no. 10, pp. 1606–1620, Oct. 2018.
- [59] M. Turan, Y. Almalioglu, E. P. Ornek, H. Araujo, M. F. Yanik, and M. Sitti, "Magnetic-visual sensor fusion-based dense 3D reconstruction and localization for endoscopic capsule robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1283–1289.
- [60] S. L. Charreyron, Q. Boehler, A. J. Millane, and B. J. Nelson, "Visual-kinematic monocular SLAM using a magnetic endoscope," in *Proc. Hamlyn Symp. Med. Robot.*, London, U.K.: The Hamlyn Centre, Imperial College London, 2018, pp. 81–82.
- [61] L. Yang, J. Wang, T. Ando, A. Kubota, H. Yamashita, I. Sakuma, T. Chiba, and E. Kobayashi, "Vision-based endoscope tracking for 3D ultrasound image-guided surgical navigation," *Computerized Med. Imag. Graph.*, vol. 40, pp. 205–216, Mar. 2015.
- [62] T. Ross, A. Reinke, P. M. Full, M. Wagner, H. Kenngott, M. Apitz, H. Hempte, D. M. Filimon, P. Scholz, T. N. Tran, and P. Bruno, "Robust medical instrument segmentation challenge 2019," Tech. Rep., 2020.
- [63] C. Yang, Z. Zhao, and S. Hu, "Image-based laparoscopic tool detection and tracking using convolutional neural networks: A review of the literature," *Comput. Assist. Surg.*, vol. 25, no. 1, pp. 15–28, Jan. 2020.
- [64] N. Shinagawa, K. Yamazaki, Y. Onodera, K. Miyasaka, E. Kikuchi, H. Dosaka-Akita, and M. Nishimura, "CT-guided transbronchial biopsy using an ultrathin bronchoscope with virtual bronchoscopic navigation," *Chest*, vol. 125, no. 3, pp. 1138–1143, Mar. 2004.
- [65] L. Ma, Y. Fang, T. Zhang, P. Xue, L. Bo, W. Liu, and E. Fu, "Comparison in efficacy and safety of forceps biopsy for peripheral lung lesions guided by endobronchial ultrasound-guided sheath (EBUS-GS) and electromagnetic navigation bronchoscopy combined with EBUS (ENB-EBUS)," *Amer. J. Transl. Res.*, vol. 12, no. 8, p. 4604, 2020.
- [66] J. Hou, R. Yang, Q. Lin, Z. Zhang, Y. Xie, and M. Huang, "Automatic robotic puncture system for accurate liver cancer ablation based on optical surgical navigation," *Zhongguo Yi Liao Qi Xie Za Zhi= Chin. J. Med. Instrumentatio*, vol. 42, no. 1, pp. 27–30, 2018.
- [67] L. Bouarfa, O. Akman, A. Schneider, P. P. Jonker, P. P. Jonker, and J. Dankelman, "In-vivo real-time tracking of surgical instruments in endoscopic video," *Minimally Invasive Therapy Allied Technol.*, vol. 21, no. 3, pp. 129–134, 2012.
- [68] G.-Q. Wei, K. Arbter, and G. Hirzinger, "Real-time visual servoing for laparoscopic surgery. Controlling robot motion with color image segmentation," *IEEE Eng. Med. Biol. Mag.*, vol. 16, no. 1, pp. 40–45, Jan. 1997.
- [69] A. Krupa, J. Gangloff, C. Doignon, M. F. de Mathelin, G. Morel, J. Leroy, L. Soler, and J. Marescaux, "Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing," *IEEE Trans. Robot. Autom.*, vol. 19, no. 5, pp. 842–853, Oct. 2003.
- [70] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, "Vision-based and marker-less surgical tool detection and tracking: A review of the literature," *Med. Image Anal.*, vol. 35, pp. 633–654, Jan. 2017.
- [71] C. Lee, Y. F. Wang, D. R. Uecker, and Y. Wang, "Image analysis for automated tracking in robot-assisted endoscopic surgery," in *Proc. 12th Int. Conf. Pattern Recognit.*, Oct. 1994, pp. 88–92.
- [72] J. Climent and R. A. Hexsel, "Particle filtering in the Hough space for instrument tracking," *Comput. Biol. Med.*, vol. 42, no. 5, pp. 614–623, May 2012.
- [73] S. Haase, J. Wasza, T. Kilgus, and J. Hornegger, "Laparoscopic instrument localization using a 3-D time-of-flight/RGB endoscope," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Oct. 2013, pp. 449–454.
- [74] S. Kumar, M. S. Narayanan, P. Singhal, J. J. Corso, and V. Krovci, "Product of tracking experts for visual tracking of surgical tools," in *Proc. IEEE Int. Conf. Automat. Sci. Eng. (CASE)*, Aug. 2013, pp. 480–485.
- [75] M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly, and D. Stoyanov, "Toward detection and localization of instruments in minimally invasive surgery," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1050–1058, Apr. 2013.
- [76] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2006, pp. 1978–1983.
- [77] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M. H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Computer Vision—ECCV*. Springer, 2014, pp. 127–141.
- [78] S. Wang, A. Raju, and J. Huang, "Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 620–623.
- [79] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Tech. Rep., 2014.
- [80] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [81] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.
- [82] K. Jo, Y. Choi, J. Choi, and J. W. Chung, "Robust real-time detection of laparoscopic instruments in robot surgery using convolutional neural networks with motion vector prediction," *Appl. Sci.*, vol. 9, no. 14, p. 2865, Jul. 2019.
- [83] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Springer, 2015, pp. 234–241.
- [84] T. Kurmann, P. M. Neila, X. Du, P. Fua, D. Stoyanov, S. Wolf, and R. Sznitman, "Simultaneous recognition and pose estimation of instruments in minimally invasive surgery," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Springer, 2017, pp. 505–513.
- [85] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 801–818.
- [86] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

- [87] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [88] Z. Zhao, S. Voros, Z. Chen, and X. Cheng, "Surgical tool tracking based on two CNNs: From coarse to fine," *J. Eng.*, vol. 2019, no. 14, pp. 467–472, Feb. 2019.
- [89] D. C. Brown, "Decentering distortion of lenses," Tech. Rep., 1966.
- [90] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Autom.*, vol. 3, no. 4, pp. 323–344, Aug. 1987.
- [91] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 10, pp. 965–980, Oct. 1992.
- [92] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
- [93] A. Datta, J. S. Kim, and T. Kanade, "Accurate camera calibration using iterative refinement of control points," in *Proc. 12th Int. Conf. Comput. Vis. Workshops, ICCV Workshops*, Sep. 2009, pp. 1201–1208.
- [94] J. Heikkilä and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 1106–1112.
- [95] J. Heikkilä, "Geometric camera calibration using circular control points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1066–1077, Oct. 2000.
- [96] J. N. Ouellet and P. Hébert, "Precise ellipse estimation without contour point extraction," *Mach. Vis. Appl.*, vol. 21, no. 1, p. 59, 2009.
- [97] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2012, pp. 3936–3943.
- [98] M. Ruffli, D. Scaramuzza, and R. Siegwart, "Automatic detection of checkerboards on blurred and distorted images," in *Proc. IEEE/RSS Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 3121–3126.
- [99] F. Forster, "Camera calibration: Active versus passive targets," *Opt. Eng.*, vol. 50, no. 11, Nov. 2011, Art. no. 113601.
- [100] L. Huang, Q. Zhang, and A. J. O. L. Asundi, "Camera calibration with active phase target: Improvement on feature detection and optimization," *Opt. Lett.*, vol. 38, no. 9, pp. 1446–1448, 2013.
- [101] B. Cai, Y. Wang, K. Wang, M. Ma, and X. Chen, "Camera calibration robust to defocus using phase-shifting patterns," *Sensors*, vol. 17, no. 10, p. 2361, Oct. 2017.
- [102] Z. Zhu, X. Wang, Q. Liu, and F. Zhang, "Camera calibration method based on optimal polarization angle," *Opt. Lasers Eng.*, vol. 112, pp. 128–135, Jan. 2019.
- [103] B. Chen and B. Pan, "Camera calibration using synthetic random speckle pattern and digital image correlation," *Opt. Lasers Eng.*, vol. 126, Mar. 2020, Art. no. 105919.
- [104] L. R. Ramírez-Hernández, J. C. Rodríguez-Quinones, M. J. Castro-Toscano, D. Hernández-Balbuena, W. Flores-Fuentes, R. Rascón-Carmona, L. Lindner, and O. Sergiyenko, "Improve three-dimensional point localization accuracy in stereo vision systems using a novel camera calibration method," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 1, Jan. 2020, Art. no. 172988141989671.
- [105] M. A. Gunen, E. Besdok, P. Civioglu, and U. H. Atasever, "Camera calibration by using weighted differential evolution algorithm: A comparative study with ABC, PSO, COBIDE, DE, CS, GWO, TLBO, MVMO, FOA, LSHADE, ZHANG and BOUGUET," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17681–17701, 2020.
- [106] D. Jin and Y. Yang, "Using distortion correction to improve the precision of camera calibration," *Opt. Rev.*, vol. 26, no. 2, pp. 269–277, Apr. 2019.
- [107] S. Ramalingam and P. Sturm, "A unifying model for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1309–1319, Jul. 2017.
- [108] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
- [109] L. Qiu and H. Ren, "Endoscope navigation and 3D reconstruction of oral cavity by visual SLAM with mitigated data scarcity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 2197–2204.
- [110] M. Turan, Y. Almalioglu, H. Araujo, E. Konukoglu, and M. Sitti, "A non-rigid map fusion-based direct SLAM method for endoscopic capsule robots," *Int. J. Intell. Robot. Appl.*, vol. 1, no. 4, pp. 399–409, Dec. 2017.
- [111] D. Sun, J. Liu, C. A. Linte, H. Duan, and R. A. Robb, "Surface reconstruction from tracked endoscopic video using the structure from motion approach," in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*. Springer, 2013, pp. 127–135.
- [112] A. R. Widya, Y. Monno, K. Imahori, M. Okutomi, S. Suzuki, T. Gotoda, and K. Miki, "3D reconstruction of whole stomach from endoscope video using structure-from-motion," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 3900–3904.
- [113] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Proc. 28th Annu. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2366–2374.
- [114] R. Garg, V. K. Bg, G. Carneiro, and I. Reid, "Unsupervised CNN for single view depth estimation: Geometry to the rescue," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 740–756.
- [115] C. Godard, O. M. Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 270–279.
- [116] T. Zhou, M. Brown, N. Snavely, and D. G. Lowe, "Unsupervised learning of depth and ego-motion from video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1851–1858.
- [117] H. Zhan, R. Garg, C. S. Weerasekera, K. Li, H. Agarwal, and I. Reid, "Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 340–349.
- [118] M. Ramamonjisoa, Y. Du, and V. Lepetit, "Predicting sharp and accurate occlusion boundaries in monocular depth estimation using displacement fields," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 14648–14657.
- [119] S. Nadeem and A. Kaufman, "Depth reconstruction and computer-aided polyp detection in optical colonoscopy video frames," Tech. Rep., 2016.
- [120] A. Reiter, S. Léonard, A. Sinha, M. Ishii, R. H. Taylor, and G. D. Hager, "Endoscopic-CT: Learning-based photometric reconstruction for endoscopic sinus surgery," *Proc. SPIE*, vol. 9784, Mar. 2016, Art. no. 978418.
- [121] M. Visentini-Scarzanella, T. Sugiura, T. Kaneko, and S. Koto, "Deep monocular 3D reconstruction for assisted navigation in bronchoscopy," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 7, pp. 1089–1099, Jul. 2017.
- [122] F. Mahmood and N. J. Durr, "Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy," *Med. Image Anal.*, vol. 48, pp. 230–243, Aug. 2018.
- [123] F. Mahmood and N. J. Durr, "Deep learning-based depth estimation from a synthetic endoscopy image training set," in *Proc. SPIE*, vol. 10574, Mar. 2018, Art. no. 1057421.
- [124] A. Rau, P. J. E. Edwards, O. F. Ahmad, P. Riordan, M. Janatka, L. B. Lovat, and D. Stoyanov, "Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 7, pp. 1167–1176, Jul. 2019.
- [125] X. Liu, A. Sinha, M. Unberath, M. Ishii, G. D. Hager, R. H. Taylor, and A. Reiter, "Self-supervised learning for dense depth estimation in monocular endoscopy," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 128–138.
- [126] X. Liu, A. Sinha, M. Ishii, G. D. Hager, A. Reiter, R. H. Taylor, and M. Unberath, "Dense depth estimation in monocular endoscopy with self-supervised learning methods," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1438–1447, May 2020.
- [127] L. Li, X. Li, S. Yang, S. Ding, A. Jolfaei, and X. Zheng, "Unsupervised learning-based continuous depth and motion estimation with monocular endoscopy for virtual reality minimally invasive surgery," Tech. Rep., 2020.
- [128] A. R. Widya, Y. Monno, M. Okutomi, S. Suzuki, T. Gotoda, and K. Miki, "Stomach 3D reconstruction based on virtual chromoendoscopic image generation," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 1848–1852.
- [129] R. J. Chen, T. L. Bobrow, T. Athey, F. Mahmood, and N. Durr, "Slam endoscopy enhanced by adversarial depth prediction," Tech. Rep., 2019.
- [130] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [131] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

- [132] K. Zhang, J. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 1073–1079, Jul. 2009.
- [133] M. Bleyer, C. Rhemann, and C. Rother, "PatchMatch stereo-stereo matching with slanted support windows," in *Proc. BMVC*, 2011, pp. 1–11.
- [134] F. Devernay, F. Mourgues, and E. Coste-Manière, "Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery," in *Proc. Int. Workshop Med. Imag. Augmented Reality*, Jun. 2001, pp. 16–20.
- [135] D. Stoyanov, M. V. Scarzanella, P. Pratt, and G. Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Springer, 2010, pp. 275–282.
- [136] S. Bernhardt, J. Abi-Nahed, and R. Abugharbieh, "Robust dense endoscopic stereo reconstruction for minimally invasive surgery," in *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*. Springer, 2012, pp. 254–262.
- [137] S. Röhl, S. Bodenstedt, S. Suwelack, H. Kenngott, B. P. Müller-Stich, R. Dillmann, and S. Speidel, "DenseGPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration," *Med. Phys.*, vol. 39, no. 3, pp. 1632–1645, Mar. 2012.
- [138] M. Ye, E. Johns, A. Handa, L. Zhang, P. Pratt, and G. Yang, "Self-supervised siamese learning on stereo image pairs for depth estimation in robotic surgery," in *Proc. 10th Hamlyn Symp. Med. Robot.*, Jun. 2017, pp. 1–2.
- [139] M. A. Armin, N. Barnes, S. Khan, M. Liu, F. Grimpen, and O. Salvado, "Unsupervised learning of endoscopy video frames' correspondences from global and local transformation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 108–117.
- [140] L. Chen, W. Tang, and N. W. John, "Real-time geometry-aware augmented reality in minimally invasive surgery," *Healthcare Technol. Lett.*, vol. 4, no. 5, pp. 163–167, Oct. 2017.
- [141] H. Zhou and J. Jagadeesan, "Real-time dense reconstruction of tissue surface from stereo optical video," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 400–412, Feb. 2020.
- [142] D. Dey, D. G. Gobbi, P. J. Slomka, K. J. M. Surry, and T. M. Peters, "Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: Creating stereoscopic panoramas," *IEEE Trans. Med. Imag.*, vol. 21, no. 1, pp. 23–30, Aug. 2002.
- [143] R. Thoranaghatte, J. Garcia, M. Caversaccio, D. Widmer, M. A. Gonzalez Ballester, L.-P. Nolte, and G. Zheng, "Landmark-based augmented reality system for paranasal and transnasal endoscopic surgeries," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 5, no. 4, pp. 415–422, Dec. 2009.
- [144] N. Haouchine, J. Dequidt, I. Peterlik, E. Kerrien, M.-O. Berger, and S. Cotin, "Towards an accurate tracking of liver tumors for augmented reality in robotic assisted surgery," in *Proc. IEEE Int. Conf. Robotics Automat. (ICRA)*, Jun. 2014, pp. 4121–4126.
- [145] E. Prisman, M. J. Daly, H. Chan, J. H. Siewerdsen, A. Vescan, and J. C. Irish, "Real-time tracking and virtual endoscopy in cone-beam CT-guided surgery of the sinuses and skull base in a cadaver model," in *International Forum of Allergy Rhinology*. Hoboken, NJ, USA: Wiley, 2011, pp. 70–77.
- [146] K. Harada, M. Miwa, T. Fukuyo, S. Watanabe, S. Enosawa, and T. Chiba, "ICG fluorescence endoscope for visualization of the placental vascular network," *Minimally Invasive Therapy Allied Technol.*, vol. 18, no. 1, pp. 3–7, Jan. 2009.
- [147] A. E. Bourque, S. Bedwani, J.-F. Carrier, C. Ménard, P. Borman, C. Bos, B. W. Raaymakers, N. Mickevicius, E. Paulson, and R. H. N. Tijssen, "Particle Filter–Based target tracking algorithm for magnetic resonance-guided respiratory compensation: Robustness and accuracy assessment," *Int. J. Radiat. Oncol. *Biol. *Phys.*, vol. 100, no. 2, pp. 325–334, Feb. 2018.
- [148] M. Fujii and H. Isomoto, "Next generation of endoscopy: Harmony with artificial intelligence and robotic-assisted devices," *Digestive Endoscopy*, vol. 32, no. 4, pp. 526–528, May 2020.
- [149] P. Wang, X. Xiao, J. R. Glissen Brown, T. M. Berzin, M. Tu, F. Xiong, X. Hu, P. Liu, Y. Song, D. Zhang, X. Yang, L. Li, J. He, X. Yi, J. Liu, and X. Liu, "Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy," *Nature Biomed. Eng.*, vol. 2, no. 10, pp. 741–748, Oct. 2018.
- [150] L. Guo, X. Xiao, C. Wu, X. Zeng, Y. Zhang, J. Du, S. Bai, J. Xie, Z. Zhang, Y. Li, X. Wang, O. Cheung, M. Sharma, J. Liu, and B. Hu, "Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos)," *Gastrointestinal Endoscopy*, vol. 91, no. 1, pp. 41–51, Jan. 2020.
- [151] T. Hirasawa, K. Aoyama, T. Tanimoto, S. Ishihara, S. Shichijo, T. Ozawa, T. Ohnishi, M. Fujishiro, K. Matsuo, J. Fujisaki, and T. Tada, "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," *Gastric Cancer*, vol. 21, no. 4, pp. 653–660, Jul. 2018.
- [152] T. Itoh, H. Kawahira, H. Nakashima, and N. Yata, "Deep learning analyzes helicobacter pylori infection by upper gastrointestinal endoscopy images," *Endoscopy Int. Open*, vol. 6, no. 2, pp. E139–E144, Feb. 2018.
- [153] P. Wang, T. M. Berzin, J. R. Glissen Brown, S. Bharadwaj, A. Becq, X. Xiao, P. Liu, L. Li, Y. Song, D. Zhang, Y. Li, G. Xu, M. Tu, and X. Liu, "Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study," *Gut*, vol. 68, no. 10, pp. 1813–1819, Oct. 2019.
- [154] Y. Mori, S. E. Kudo, M. Misawa, Y. Saito, H. Ikematsu, K. Hotta, K. Ohtsuka, F. Urushibara, S. Kataoka, Y. Ogawa, and Y. Maeda, "Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: A prospective study," *Ann. Internal Med.*, vol. 169, no. 6, p. 357, Sep. 2018.
- [155] Y. Zhu, Q.-C. Wang, M.-D. Xu, Z. Zhang, J. Cheng, Y.-S. Zhong, Y.-Q. Zhang, W.-F. Chen, L.-Q. Yao, P.-H. Zhou, and Q.-L. Li, "Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy," *Gastrointestinal Endoscopy*, vol. 89, no. 4, p. 806, 2019.
- [156] P.-J. Chen, M.-C. Lin, M.-J. Lai, J.-C. Lin, H. H.-S. Lu, and V. S. Tseng, "Accurate classification of diminutive colorectal polyps using computer-aided analysis," *Gastroenterology*, vol. 154, no. 3, pp. 568–575, Feb. 2018.
- [157] K. Takenaka, K. Ohtsuka, T. Fujii, M. Negi, K. Suzuki, H. Shimizu, S. Oshima, S. Akiyama, M. Motobayashi, M. Nagahori, E. Saito, K. Matsuoka, and M. Watanabe, "Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis," *Gastroenterology*, vol. 158, no. 8, pp. 2150–2157, Jun. 2020.
- [158] Y. Mori, S.-E. Kudo, M. Misawa, and K. Mori, "Simultaneous detection and characterization of diminutive polyps with the use of artificial intelligence during colonoscopy," *VideoGIE*, vol. 4, no. 1, pp. 7–10, Jan. 2019.
- [159] M. Misawa, S. Kudo, Y. Mori, Y. Maeda, Y. Ogawa, K. Ichimasa, T. Kudo, K. Wakamura, T. Hayashi, H. Miyachi, T. Baba, F. Ishida, H. Itoh, M. Oda, and K. Mori, "Current status and future perspective on artificial intelligence for lower endoscopy," *Digestive Endoscopy*, vol. 33, no. 2, pp. 273–284, Jan. 2021.
- [160] J. R. Su, Z. Li, X. J. Shao, C. R. Ji, R. J. Ji, R. C. Zhou, G. C. Li, G. Q. Liu, Y. S. He, X. L. Zuo, and Y. Q. Li, "Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: A prospective randomized controlled study," *Gastrointestinal Endoscopy*, vol. 91, no. 2, p. 415, 2020.
- [161] G. A. Roth, D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, and A. J. T. L. Abdelalim, "Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: A systematic analysis for the global burden of disease study 2017," *Lancet*, vol. 392, no. 10159, pp. 1736–1788, 2018.
- [162] Y. Uchida, "Recent advances in coronary angiography," *J. Cardiol.*, vol. 57, no. 1, pp. 18–30, Jan. 2011.
- [163] M. Shibuya, K. Fujii, H. Hao, T. Imanaka, M. Fukunaga, K. Miki, H. Tamaru, T. Nakata, H. Sawada, Y. Naito, S. Hirota, and T. Masuyama, "Ex vivo comparison of angiography and histopathology for the evaluation of coronary plaque characteristics," *Int. J. Cardiovascular Imag.*, vol. 32, no. 6, pp. 863–869, Jun. 2016.
- [164] L. E. Savastano, Q. Zhou, A. Smith, K. Vega, C. Murga-Zamalloa, D. Gordon, J. McHugh, L. Zhao, M. M. Wang, A. Pandey, B. G. Thompson, J. Xu, J. Zhang, Y. E. Chen, E. J. Seibel, and T. D. Wang, "Multimodal laser-based angiography for structural, chemical and biological imaging of atherosclerosis," *Nature Biomed. Eng.*, vol. 1, no. 2, pp. 1–15, Feb. 2017.
- [165] S. S. S. Choi and A. Mandelis, "Review of the state of the art in cardiovascular endoscopy imaging of atherosclerosis using photoacoustic techniques with pulsed and continuous-wave optical excitations," *J. Biomed. Opt.*, vol. 24, no. 8, Aug. 2019, Art. no. 080902.

- [166] E. Soylu, E. Kidher, H. Ashrafian, G. Stavridis, L. Harling, and T. Athanasiou, "A systematic review of left ventricular cardio-endoscopic surgery," *J. Cardiothoracic Surg.*, vol. 12, no. 1, p. 41, Dec. 2017.
- [167] L.-P. Yao, J. Mei, F.-B. Ding, L. Zhang, H.-M. Li, M. Ding, X. Yang, X.-M. Li, and K. Sun, "Application of cardiovascular virtual endoscopy: A pilot study on roaming path planning for diagnosis of congenital heart diseases in children," *Sci. Rep.*, vol. 8, no. 1, pp. 1–7, Dec. 2018.
- [168] A. Sinha, X. Liu, A. Reiter, M. Ishii, G. D. Hager, and R. H. Taylor, "Endoscopic navigation in the absence of CT imaging," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist.*, Springer, 2018, pp. 64–71.
- [169] X. Wu, C. A. Sánchez, P. W. Kahng, C. A. Rees, A. S. Ponukumati, E. A. Eisen, D. A. Pastel, H. Borgard, J. E. Lloyd, and S. Fels, "Multi-modal framework for image-guided transoral surgery with intraoperative imaging and deformation modeling," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2019, pp. 6975–6978.
- [170] K. Yamamoto, M. Kurose, R. Yadamura, R. Yajima, T. Okuni, and K. Takano, "Endoscopy-assisted transoral resection of a parapharyngeal space schwannoma without mandibular dissection," Tech. Rep., 2020.
- [171] A. Wang, M. S. Tenner, M. H. Schmidt, and C. Bowers, "Placement of ommaya reservoirs using electromagnetic neuronavigation and neuroendoscopy: A retrospective study with cost-benefit analysis," *World Neurosurg.*, vol. 122, pp. e723–e728, Feb. 2019.
- [172] D. K. Hogarth, "Use of augmented fluoroscopic imaging during diagnostic bronchoscopy," *Future Oncol.*, vol. 14, no. 22, pp. 2247–2252, Sep. 2018.
- [173] Q. Chang, R. Bascom, J. Toth, D. Ahmad, and W. E. Higgins, "Autofluorescence bronchoscopy video analysis for lesion frame detection," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 1556–1559.
- [174] J. Liu, B. Wang, W. Hu, P. Sun, J. Li, H. Duan, and J. Si, "Global and local panoramic views for gastroscopy: An assisted method of gastroscopic lesion surveillance," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 9, pp. 2296–2307, Apr. 2015.
- [175] K. B. Ozyoruk, G. I. Gokceler, G. Coskun, K. Incetan, Y. Almalioglu, F. Mahmood, E. Curto, L. Perdigoto, M. Oliveira, H. Sahin, H. Araujo, H. Alexandrino, N. J. Durr, H. B. Gilbert, and M. Turan, "EndoSLAM dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos: Endo-SfMLearner," 2020, *arXiv:2006.16670*. [Online]. Available: <http://arxiv.org/abs/2006.16670>
- [176] G. Bae, I. Budvytis, C.-K. Yeung, and R. Cipolla, "Deep multi-view stereo for dense 3D reconstruction from monocular endoscopic video," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Springer, 2020, pp. 774–783.
- [177] S. Gulati, M. Patel, A. Emmanuel, A. Haji, B. Hayee, and H. Neumann, "The future of endoscopy: Advances in endoscopic image innovations," *Digestive Endoscopy*, vol. 32, no. 4, pp. 512–522, May 2020.
- [178] J. van der Putten, R. Wildeboer, J. de Groof, R. van Sloun, M. Struyvenberg, F. van der Sommen, S. Zinger, W. Curvers, E. Schoon, and J. Bergman, "Deep learning biopsy marking of early neoplasia in Barrett's esophagus by combining WLE and BLI modalities," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Aug. 2019, pp. 1127–1131.
- [179] M. R. Struyvenberg, F. van der Sommen, A. F. Swager, A. J. de Groof, A. Rikos, E. J. Schoon, J. J. Bergman, P. H. N. de With, and W. L. Curvers, "Improved Barrett's neoplasia detection using computer-assisted multi-frame analysis of volumetric laser endomicroscopy," *Diseases Esophagus*, vol. 33, no. 2, 2020, Art. no. doz065.
- [180] Y.-Y. Zhao, D.-X. Xue, Y.-L. Wang, R. Zhang, B. Sun, Y.-P. Cai, H. Feng, Y. Cai, and J.-M. Xu, "Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy," *Endoscopy*, vol. 51, no. 04, pp. 333–341, Apr. 2019.
- [181] Y. Kumagai, K. Takubo, K. Kawada, K. Aoyama, Y. Endo, T. Ozawa, T. Hirasawa, T. Yoshio, S. Ishihara, M. Fujishiro, J.-I. Tamaru, E. Mochiki, H. Ishida, and T. Tada, "Diagnosis using deep-learning artificial intelligence based on the endocystoscopic observation of the esophagus," *Esophagus*, vol. 16, no. 2, pp. 180–187, Apr. 2019.
- [182] R. Miyaki, S. Yoshida, S. Tanaka, Y. Kominami, Y. Sanomura, T. Matsuo, S. Oka, B. Raytchev, T. Tamaki, T. Koide, K. Kaneda, M. Yoshihara, and K. Chayama, "Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement," *J. Gastroenterology Hepatology*, vol. 28, no. 5, pp. 841–847, May 2013.
- [183] H. Ueyama, Y. Kato, Y. Akazawa, N. Yatagai, H. Komori, T. Takeda, K. Matsumoto, K. Ueda, K. Matsumoto, M. Hojo, and T. Hepatology, "Application of artificial intelligence using a convolutional neural network for diagnosis of early gastric cancer based on magnifying endoscopy with narrow-band imaging," Tech. Rep., 2020.
- [184] M. Tokunaga, T. Matsumura, R. Nankinzan, T. Suzuki, H. Oura, T. Kaneko, M. Fujie, S. Hirai, R. Saiki, and N. Akizue, "A computer-aided diagnosis system using only white-light endoscopy for the prediction of invasion depth in colorectal cancer," Tech. Rep., 2020.
- [185] E. M. Song, B. Park, C.-A. Ha, S. W. Hwang, S. H. Park, D.-H. Yang, B. D. Ye, S.-J. Myung, S.-K. Yang, N. Kim, and J.-S. Byeon, "Endoscopic diagnosis and treatment planning for colorectal polyps using a deep-learning model," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, Dec. 2020.
- [186] H. Horiuchi, N. Tamai, S. Kamba, H. Inomata, T. R. Ohya, and K. Sumiyama, "Real-time computer-aided diagnosis of diminutive rectosigmoid polyps using an auto-fluorescence imaging system and novel color intensity analysis software," *Scandin. J. Gastroenterology*, vol. 54, no. 6, pp. 800–805, Jun. 2019.
- [187] T. Anayama, J. Qiu, H. Chan, T. Nakajima, R. Weersink, M. Daly, J. McConnell, T. Waddell, S. Keshavjee, D. Jaffray, J. C. Irish, K. Hirohashi, H. Wada, K. Orihashi, and K. Yasufuku, "Localization of pulmonary nodules using navigation bronchoscope and a near-infrared fluorescence thoracoscope," *Ann. Thoracic Surg.*, vol. 99, no. 1, pp. 224–230, Jan. 2015.
- [188] C. Y. Lee, H. Chan, H. Ujiie, K. Fujino, T. Kinoshita, J. C. Irish, and K. Yasufuku, "Novel thoracoscopic navigation system with augmented real-time image guidance for chest wall tumors," *Ann. Thoracic Surg.*, vol. 106, no. 5, pp. 1468–1475, Nov. 2018.
- [189] P. Tinguely, M. Fusaglia, J. Freedman, V. Banz, S. Weber, D. Candinas, and H. Nilsson, "Laparoscopic image-based navigation for microwave ablation of liver tumors—A multi-center study," *Surgical Endoscopy*, vol. 31, no. 10, pp. 4315–4324, Oct. 2017.
- [190] T. Aoki, D. A. Mansour, T. Koizumi, Y. Wada, Y. Enami, A. Fujimori, T. Kusano, K. Matsuda, K. Nogaki, and Y. Tashiro, "Laparoscopic liver surgery guided by virtual real-time CT-guided volume navigation," Tech. Rep., 2020.
- [191] G. A. Prevost, B. Eigl, I. R. T. Paolucci, M. Peterhans, S. Weber, G. Beldi, D. Candinas, and A. Lachenmayer, "Efficiency, accuracy and clinical applicability of a new image-guided surgery system in 3D laparoscopic liver surgery," *J. Gastrointestinal Surg.*, to be published.
- [192] W. W. Ludwig, S. Lim, D. Stoianovici, and B. R. Matlaga, "Endoscopic stone measurement during ureteroscopy," *J. Endourology*, vol. 32, no. 1, pp. 34–39, Jan. 2018.
- [193] A. H. Aldoukhi, H. Law, K. M. Black, W. W. Roberts, J. Deng, and K. R. Ghani, "PD04-06 Deep learning computer vision algorithm for detecting kidney stone composition: Towards an automated future," *J. Urol.*, vol. 201, no. 4, pp. e75–e76, Apr. 2019.
- [194] C. Raposo, J. P. Barreto, C. Sousa, L. Ribeiro, R. Melo, J. P. Oliveira, P. Marques, F. Fonseca, and D. Barrett, "Video-based computer navigation in knee arthroscopy for patient-specific ACL reconstruction," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 9, pp. 1529–1539, Sep. 2019.
- [195] S. G. Vitale, S. Caruso, A. Vitagliano, G. Vilos, L. M. Di Gregorio, B. Zizolfi, J. Tesarik, and A. Cianci, "The value of virtual reality simulators in hysteroscopy and training capacity: A systematic review," *Minimally Invasive Therapy Allied Technol.*, vol. 29, no. 4, pp. 185–193, Jul. 2020.
- [196] G. Niazi and W. M. Hetta, "Role of multidetector CT virtual cystoscopy compared to conventional cystoscopy in the diagnosis of urinary bladder neoplasms," *Egyptian J. Radiol. Nucl. Med.*, vol. 50, no. 1, pp. 1–10, Dec. 2019.
- [197] D. Jiao, Z. Zhang, Z. Sun, Y. Wang, X. J. D. Han, and I. Radiology, "Percutaneous nephrolithotripsy: C-Arm CT with 3D virtual navigation in non-dilated renal collecting systems," *Diagnostic Interventional Radiol.*, vol. 24, no. 1, p. 17, 2018.
- [198] B. Wickens, J. Lewis, D. P. Morris, M. Husein, H. M. Ladak, S. Agrawal, and N. Surgery, "Face and content validity of a novel, Web-based otoscopy simulator for medical education," *J. Otolaryngology-Head Neck Surg.*, vol. 44, no. 1, pp. 1–8, 2015.
- [199] P. Tabakow, A. Weiser, K. Chmielak, P. Blauciak, J. Bładowska, and M. Cyz, "Navigated neuroendoscopy combined with intraoperative magnetic resonance cysternography for treatment of arachnoid cysts," *Neurosurgical Rev.*, to be published.

- [200] X. Luo, H. Q. Zeng, Y.-P. Du, and X. Cheng, "A novel endoscopic navigation system: Simultaneous endoscope and radial ultrasound probe tracking without external trackers," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Springer, 2019, pp. 47–55.
- [201] S. M. Qazi, A. A. Bhat, and S. A. Patigaroo, "Image guided endoscopic sinus surgery: First experience from kashmir valley," *Indian J. Otolaryngology Head Neck Surg.*, pp. 1–10, Apr. 2020.
- [202] C. Matava, E. Pankiv, S. Raisbeck, M. Caldeira, and F. Alam, "A convolutional neural network for real time classification, identification, and labelling of vocal cord and tracheal using laryngoscopy and bronchoscopy video," *J. Med. Syst.*, vol. 44, no. 2, pp. 1–10, Feb. 2020.
- [203] A. R. Belanger, A. C. Burks, D. M. Chambers, S. Ghosh, C. R. MacRosty, A. J. Conterato, M. P. Rivera, and J. A. Akulian, "Peripheral lung nodule diagnosis and fiducial marker placement using a novel tip-tracked electromagnetic navigation bronchoscopy system," *J. Bronchology Interventional Pulmonology*, vol. 26, no. 1, pp. 41–48, 2019.
- [204] X. Liu, J. Berg, F. King, and N. Hata, "Computer vision-guided bronchoscopic navigation using dual CNN-generated depth images and ICP registration," in *Proc. SPIE*, vol. 11315, Mar. 2020, Art. no. 113152C.
- [205] S.-L. Cai, B. Li, W.-M. Tan, X.-J. Niu, H.-H. Yu, L.-Q. Yao, P.-H. Zhou, B. Yan, and Y.-S. Zhong, "Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video)," *Gastrointestinal Endoscopy*, vol. 90, no. 5, pp. 745–753, 2019.
- [206] X. Zhang, F. Chen, T. Yu, J. An, Z. Huang, J. Liu, W. Hu, L. Wang, H. Duan, and J. Si, "Real-time gastric polyp detection using convolutional neural networks," *PLoS ONE*, vol. 14, no. 3, Mar. 2019, Art. no. e0214133.
- [207] H. Nakashima, "Artificial intelligence diagnosis of helicobacter pylori infection using blue laser imaging-bright and linked color imaging: A single-center prospective study," *Ann. Gastroenterology*, vol. 31, no. 4, p. 462, 2018.
- [208] M. W. Alam, S. S. Vedaei, and K. A. Wahid, "A fluorescence-based wireless capsule endoscopy system for detecting colorectal cancer," *Cancers*, vol. 12, no. 4, p. 890, Apr. 2020.
- [209] J. Yang, L. Chang, S. Li, X. He, and T. Zhu, "WCE polyp detection based on novel feature descriptor with normalized variance locality-constrained linear coding," *Tech. Rep.*, 2020.
- [210] Y. Luo, Y. Zhang, M. Liu, Y. Lai, P. Liu, Z. Wang, T. Xing, Y. Huang, Y. Li, A. Li, Y. Wang, X. Luo, S. Liu, and Z. Han, "Artificial intelligence-assisted colonoscopy for detection of colon polyps: A prospective, randomized cohort study," *J. Gastrointestinal Surg.*, pp. 1–8, Sep. 2020.
- [211] Y. Maeda, S.-E. Kudo, Y. Mori, M. Misawa, N. Ogata, S. Sasanuma, K. Wakamura, M. Oda, K. Mori, and K. Ohtsuka, "Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video)," *Gastrointestinal Endoscopy*, vol. 89, no. 2, pp. 408–415, Feb. 2019.
- [212] M. Min, S. Su, W. He, Y. Bi, Z. Ma, and Y. Liu, "Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Dec. 2019.
- [213] D. Ștefanescu, C. Streba, E. T. Cârțnă, A. Săftoiu, G. Gruionu, and L. G. Gruionu, "Computer aided diagnosis for confocal laser endomicroscopy in advanced colorectal adenocarcinoma," *PLoS ONE*, vol. 11, no. 5, May 2016, Art. no. e0154863.
- [214] P. Zhang, H. Luo, W. Zhu, J. Yang, N. Zeng, Y. Fan, S. Wen, N. Xiang, F. Jia, and C. Fang, "Real-time navigation for laparoscopic hepatectomy using image fusion of preoperative 3D surgical plan and intraoperative indocyanine green fluorescence imaging," *Surgical Endoscopy*, vol. 34, no. 8, pp. 3449–3459, 2019.
- [215] T. Finger, A. Schaumann, M. Schulz, and U.-W. Thomale, "Augmented reality in intraventricular neuroendoscopy," *Acta Neurochirurgica*, vol. 159, no. 6, pp. 1033–1041, Jun. 2017.
- [216] S. Drouin, A. Kochanowska, M. Kersten-Oertel, I. J. Gerard, R. Zelmann, D. De Nigris, S. Bériault, T. Arbel, D. Sirhan, A. F. Sadikot, J. A. Hall, D. S. Sinclair, K. Petrecca, R. F. DelMaestro, and D. L. Collins, "IBIS: An OR ready open-source platform for image-guided neurosurgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 3, pp. 363–378, Mar. 2017.
- [217] K. Yoshida, S. Naito, and T. Matsuda, *Navigation in Endourology, Ureterscopy*: 'Endourology Progress. Springer, 2019, pp. 289–295.



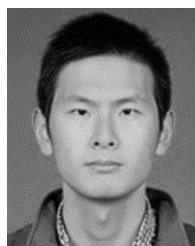
ZUOMING FU received the B.S. degree in biomedical engineering from the South China University of Technology, Guangzhou, China, in 2013. She is currently pursuing the Ph.D. degree in biomedical engineering with Zhejiang University, Hangzhou, Zhejiang, China.

Since 2013, she continues to work in medical endoscope related research with the Biosensor National Special Laboratory, Key Laboratory of Biomedical Engineering of Ministry of Education, Zhejiang University. Her research interests include endoscopic imaging and real-time medical video process techniques and visual-based navigation algorithm including multi-systems calibration, endoscope tracking, multi-view reconstruction, and three-dimension visualization.



ZIYI JIN received the B.S. degree in biomedical engineering from Northeastern University, Shenyang, China, in 2016. He is currently pursuing the Ph.D. degree in biomedical engineering with Zhejiang University, Hangzhou, Zhejiang, China.

Since 2016, he continues to work in medical endoscope related research with the Biosensor National Special Laboratory, Key Laboratory of Biomedical Engineering of Ministry of Education, Zhejiang University. His research interests include the endoscopic imaging and 3D surface reconstruction techniques and endoscopic computer-aided diagnosis algorithm-based on CNN.



CHONGAN ZHANG received the B.S. degree in measurement and control technology and instrument from the Harbin Institute of Technology, Harbin, China, in 2014. He is currently pursuing the Ph.D. degree in biomedical engineering with Zhejiang University, Hangzhou, Zhejiang, China.

Since 2014, he continues to work in medical endoscope related research with the Biosensor National Special Laboratory, Key Laboratory of Biomedical Engineering of Ministry of Education, Zhejiang University. His research interests include the endoscopic imaging and real-time medical video process techniques and electromagnetic navigation system and shape sensing technique for continuum robot.



ZHONGYU HE was born in Shanxi, Mainland China. He received the B.S. degree in optoelectronic information science and engineering from Shandong University, Jinan, Mainland China, in 2018. He is currently pursuing the Ph.D. degree in biomedical engineering with Zhejiang University, Hangzhou, Mainland China.

His research interests include the development of biomedical optics, medical imaging, spectroscopy, and advanced endoscopic optical medical detection technologies.



ZHENZHOU ZHA was born in Zhejiang, Mainland China. He received the B.S. degree from the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, Mainland China, in 2019. He is currently pursuing the master's degree in biomedical engineering with Zhejiang University, Hangzhou, Mainland China.

His research interests include advanced computer vision of biomedical images, endoscopic surgery navigation technologies, and endoscopic structure design.



CHUNYONG HU was born in Liaoning, Mainland China. He received the B.S. degree in mechatronic engineering from Zhejiang University, Hangzhou, Mainland China, in 2020, where he is currently pursuing the master's degree in biomedical engineering.

His research interests include sensors, mechanical design, mechatronic, and advanced endoscopic surgery navigation technologies.



TIANYUAN GAN was born in Zhejiang, Mainland China. He received the B.S. degree in biomedical engineering from Zhejiang University, Hangzhou, Mainland China, in 2020, where he is currently pursuing the Ph.D. degree.

His research interests include the designing of computer-aided diagnosis system in endoscope, the development of deep learning algorithms, and real-time medical video process techniques.



QINGLAI YAN was born in Wenling, Zhejiang, China, in 1982. He received the B.S. and M.S. degrees from Zhejiang University, China.

Since 2013, he has been a Senior Engineer with the Hangzhou Center for Medical Device Quality Supervision and Testing, CFDA. He focused on the standardization of evaluating and testing the safety and effectiveness of optical medical devices. His main achievement is to lead or participate in drafting more than 20 Chinese national standards and industrial standards of optical medical devices. He is the Project Leader of two ISO standards WD draft with respect to the evaluation of endoscope optical characteristics. His research interests include establishing evaluation methods of the medical endoscope system's characteristics, the study of endoscope optical path design, and the fundamental study of endoscope photoelectric imaging.



PENG WANG was born in Penglai, China, in 1982. He received the B.S., M.S. and Ph.D. degrees in biomedical engineering from Zhejiang University, Hangzhou, China, in 2005 and 2012, respectively.

From 2012 to 2015, he was a Research Associate with the Electrical Engineering and Computer Science Department, Case Western Reserve University. Since 2016, he has been a Research Associate with the Biosensor National Special Laboratory, Key Laboratory of Biomedical Engineering of Ministry of Education, Zhejiang University. He is the author of one book chapter, more than ten articles, and seven Chinese patents. His research interests include minimal invasive instrument and implantable telemetry technology, including endoscopic navigation, ultra-low power integrated circuit, biomedical telemetry, wireless powering, energy harvesting, wireless sensor networks, and embedded system design.



XUESONG YE received the Ph.D. degree in biomedical engineering from Zhejiang University, Hangzhou, in 1997.

He is currently a Full Professor with the College of Biomedical Engineering and Instrument Science, Zhejiang University. He is also the Director of the Zhejiang University Medical and Health Information Engineering Technology Research Institute, an Executive Director of the Zhejiang Intelligent Medical Equipment Manufacturing Innovation Center, and the Deputy Director of the Zhejiang Provincial Key Laboratory of Endoscope Technology Research. He has long been engaged in sensor detection technology and scientific instrument research and development. His research interests include micro-nano manufacturing of biomedical sensors, low power interface design of CMOS integrated circuit, miniaturized package of detection system, and related intelligent detection technology at the level of molecules, cells, tissues, and systems, closely combined with basic and clinical medical demands.

...