



HAL
open science

Optical biopsy mapping on endoscopic image mosaics with a marker-free probe

Omar Zenteno, Dinh-Hoan Trinh, Sylvie Treuillet, Yves Lucas, Thomas Bazin,
Dominique Lamarque, Christian Daul

► To cite this version:

Omar Zenteno, Dinh-Hoan Trinh, Sylvie Treuillet, Yves Lucas, Thomas Bazin, et al.. Optical biopsy mapping on endoscopic image mosaics with a marker-free probe. *Computers in Biology and Medicine*, 2022, 143, pp.105234. 10.1016/j.combiomed.2022.105234 . hal-03657261

HAL Id: hal-03657261

<https://hal.science/hal-03657261v1>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Optical biopsy mapping on endoscopic image mosaics with a marker-free probe

Omar Zenteno^a, Dinh-Hoan Trinh^c, Sylvie Treuillet^a, Yves Lucas^a, Thomas Bazin^b, Dominique Lamarque^b and Christian Daul^{c,*}

^aLaboratoire PRISME, Université d'Orléans, Orléans, France

^bService d'Hépatogastroentérologie et oncologie digestive, Hôpital Ambroise Paré, Boulogne-Billancourt, France

^cCRAN, UMR 7039 CNRS and Université de Lorraine, Vandœuvre-lès-Nancy, France

ARTICLE INFO

Keywords:

Image mosaicing
Multimodal registration
Hyperspectral Imaging
Gastroendoscopy

ABSTRACT

Gastric cancer is the second leading cause of cancer-related deaths worldwide. Early diagnosis significantly increases the chances of survival; therefore, improved assisted exploration and screening techniques are necessary. Previously we made use of an augmented multi-spectral endoscope by inserting an optical probe into the instrumentation channel. However, the limited field of view and the lack of markings left by optical biopsies on the tissue complicate the navigation and revisit of the suspect areas probed in-vivo. In this contribution two innovative tools are introduced to significantly increase the traceability and monitoring of patients in clinical practice: (i) video mosaicing to build a more comprehensive and panoramic view of large gastric areas; (ii) optical biopsy targeting and registration with the endoscopic images. The proposed optical flow-based mosaicing technique selects images that minimize texture discontinuities and is robust despite the lack of texture and illumination variations. The optical biopsy targeting is based on automatic tracking of a free-marker probe in the endoscopic view using deep learning for estimating dynamically its pose during exploration. The accuracy of pose estimation is sufficient to ensure a precise overlapping of the standard white-light color image and the hyperspectral probe image, assuming that the small target area of the organ is almost flat. This allows the mapping of all spatio-temporally tracked biopsy sites onto the panoramic mosaic. Experimental validations are carried out from videos acquired on patients in hospital. The proposed technique is purely software-based and therefore easily integrable into clinical practices. It is also generic and compatible to any imaging modalities connected to a cylindrical fibroscope.

1. Introduction

Gastric cancer is one of the world's primary causes of morbidity and mortality from malignant diseases. If diagnosed and treated at an early stage, the 5-year survival rate is as high as 70 percent [1, 2]. The standard medical procedure for diagnosing pathologies of the upper gastrointestinal tract (esophagus, stomach and duodenum) is Endoscopy. The visual examination in white light or with color enhancement of vascular structures such as narrow-band imaging (NBI from Olympus), is the main tool for screening and is usually complemented by incisional biopsies to take tissue samples and perform histological analysis of the mucosa. However, biopsy is invasive and the histological sections analysis increases processing time. Furthermore, early inflammatory precancerous lesions may present very subtle alterations of the mucosa almost invisible under white light video, leading the gastroenterologist to perform multiple random biopsies. This technique is operator-dependent and causes delays in diagnosis. Moreover, the complications for navigation and diagnosis are further increased by the observation of a very limited field of view (FoV). Indeed, even when the endoscopic camera has a approximate FoV of 140 degrees, the

distal tip must be moved close to the internal epithelial wall to provide high-resolution images, which effectively reduces it to a few square centimeters. The limited field of view through the endoscope can however be expanded by using image stitching and surface reconstruction methods [3].

Innovative techniques, known as optical biopsy, have been proposed to perform additional targeted in-vivo tissue assessment using light properties [4, 5, 6]. They make use of miniaturized photonic probes that have been designed to be inserted into the operating channel of a standard endoscope. Optical biopsy has several advantages over conventional biopsy: the probe provides the surgeon with an on-line tissue assessment and larger areas can be scanned. However, since optical biopsy does not leave any mark on the tissue, it is necessary to assist the surgeon in locating the scanned area and retargeting the biopsies for patient follow-up.

This contribution aims to overcome the aforementioned limitations by providing a wider FoV through a panoramic image constructed from the endoscopic video where optical biopsies can be localized. The technique exploits an hyperspectral-augmented endoscopic prototype which comprises a multispectral snapshot camera sensitive to visible and near-infrared light, and connected to a fibroscope inserted through the instrument channel of a commercial endoscope [7, 8]. The inserted fibroscope works as a localized hyperspectral probe for detection of precancerous lesions in stomach (see Fig. 1).

* This document is the result of the EMMIE research project funded by the french Agence Nationale de la Recherche under ID ANR-15-CE17-0015.

*Corresponding author

✉ christian.daul@univ-lorraine.fr (C. Daul)

ORCID(s):

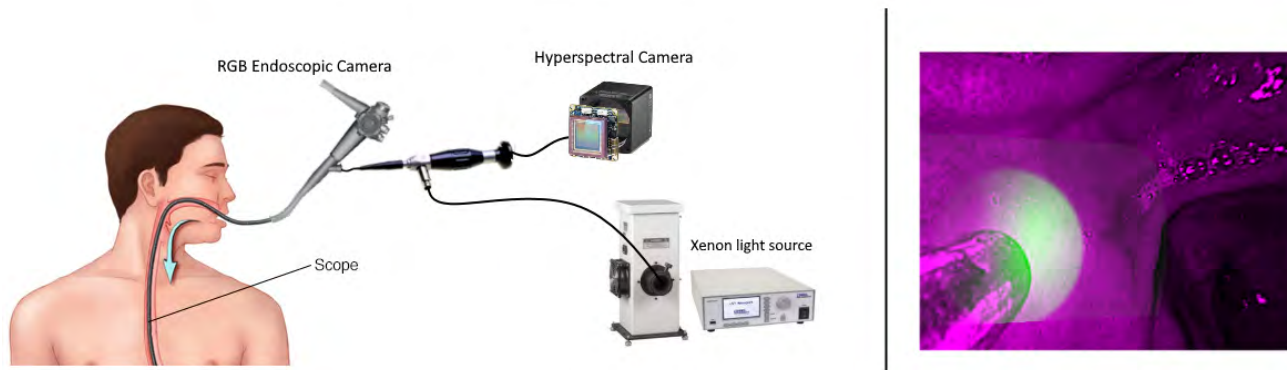


Figure 1: The augmented endoscope. Left : Experimental prototype. Right : Overlay of the zone targeted by the optical probe (in green) on the endoscopic view (in purple)

The recording of the two video streams (endo- and fiberoptic) provided by the prototype is performed simultaneously, but the multi-modal registration is challenging for three reasons. First, the two imaging sensors have different intrinsic parameter values (focal lengths, distortions, resolutions). Second, the sensors have different viewpoints with a correspondence that can vary over time because the clinician can move them independently to each other during exploration. Third, the data is from bi-modal cameras (standard visible spectrum on one side and near infrared on the other) and conventional registration methods using image content generally fail because of the different nature of the information and the lack of texture of the gastric wall. The registration of the multi-modal images can be achieved by the on-line estimation of the relative pose of the optical probe observed in the endoscopic view. Nevertheless, off-line calibration of each sensor is needed for distortion rectification of each modality separately.

This paper presents a novel method to combine video mosaicing based on optical flow with markerless probe tracking in the endoscopic video to locate the biopsy on the panoramic view. Video mosaicing is the process of stitching endoscopic frames together to form a comprehensive view of the scene. Registration of annotated biopsies on mosaics is useful for patient follow-up and traceability of examinations (having a history of the areas examined). The proposed method can be generalized to any endoscopic system in which an optical probe is inserted into the instrument channel, regardless of the type of imaging sensor that it is connected to.

The remainder of this document is organized as follows: Section II makes a review of the related work. Section III details a mosaicing algorithm for complex scenes, as well as an optical biopsy localization and re-targeting approach. Section IV presents an experimental validation of the hyperspectral biopsy localization on image mosaics from endoscopic videos acquired on patients. Finally, Section V concludes the manuscript and opens up some perspectives for the medical practice.

2. Related Work

This section gives an overview of the state-of-the-art of the two issues addressed in this contribution, namely endoscopic image mosaicing of hollow organs and optical biopsy localization and re-targeting.

2.1. Mosaicing of hollow organ image sequences

Image mosaicing is an effective mean to construct a single seamless panoramic image by aligning multiple partially overlapped images. Image mosaicing of the inner wall of hollow organs is a challenging issue, especially in the stomach in which no textures and structures are available: the uncontrolled endoscope displacements generate motion blur or defocusing, the viewpoint changes lead to illumination changes between the images and specular reflections affect the images.

In the last decade, most of the works dealing with 2D image mosaicing of hollow organs focused on the construction of panoramic views of the bladder epithelium using cystoscopic video-sequences. Based on the assumption that the endoscope's distal tip is close to the tissue (i.e., the surfaces are almost planar due to the small field of view), consecutive images are re-aligned by computing homographies that superimpose their homologous pixels, then all images are stitched together in a global mosaic coordinate system.

In a first approach applied on fluorescence bladder images [9, 10], robust feature extractors, such as SIFT [11] or SURF [12] and a RANSAC outlier rejection algorithm [13] are used to find the best parameters of the homographies. Such an approach is possible for textured images of the bladder, but is less effective in the white-light cystoscopic modality [14, 15] and fails in gastroscopy due to lack of texture.

Other approaches exploit directly the colour information to determine the homographies without any feature extraction [16, 17, 18]. These three methods were developed for the bladder and are based on the mutual information optimization [19], on graph-cuts [20], and on optical flow [21], respectively.

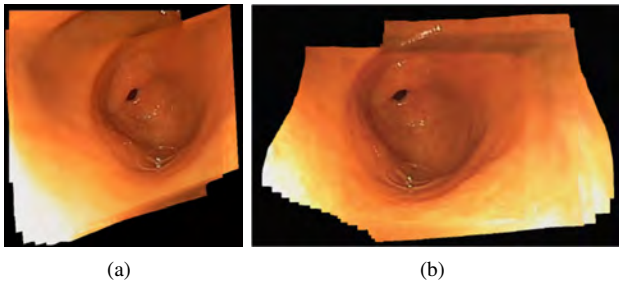


Figure 2: Two mosaic examples of the pyloric antrum built from same images: (a) with the method in [18] which only uses homographies without an image selection strategy, (b) with the method in [24] which directly uses flow fields.

However, determining homographies between consecutive images along the video-sequence lead to coherent mosaic with a limited extend. Indeed, due to small accumulating registration errors along the video sequence, texture discontinuities and geometrical distortions (the scene is not planar in the complete mosaic) affect the mosaic which cannot be always corrected, even by elaborated global bundle adjustment methods [22, 23].

The authors in [24] proposed a new approach for building 2D mosaics in the stomach, based on a robust algorithm to determine the optical flow between consecutive texture-free images affected by strong illumination changes [25]. The optical flow information was directly used to place the pixels into the global mosaic coordinate system.

As seen in Fig. 2, there is a significant difference depending on the method used to reconstruct the mosaic. In the work [24], the optical flux vectors are added iteratively. However, even with accurate optical flow determination, errors accumulate and visual inconsistencies become visible as the mosaic grows. In addition, placing pixels from many images on the mosaic plane using only the flow fields is a complex task when strategies must be implemented to minimize mapping errors.

2.2. Optical biopsy localization and re-targeting

Optical biopsy refers to all methods that use the properties of light to enable on-line tissue assessment during endoscopy. New biophotonic probes that can be inserted into the channel of a standard white light endoscope have been proposed, in particular for optical coherence tomography [4], miniaturized confocal laser [5] or spectral band imaging [6]. It helps to improve lesion targeting and reduce the number of required incisional biopsies. Our prototype exploits hyperspectral imaging. Wavelengths $415nm$ and $540nm$ are commonly used to highlight vascular structures (as NBI). Recent publications have demonstrated that using broader spectral information could better detect early lesions such as inflammatory gastritis [26, 27, 28, 29, 30]. Since optical biopsy leaves no marks on the tissue, it is necessary to provide gastroenterologists with on-line localization and re-targeting of biopsies. Two approaches have been explored in the literature : (i) by region matching from image content

[31, 32, 33] or (ii) using epipolar geometry in endoscopic video [34, 35, 36, 37].

Atasoy et al.[31] presented an affine-covariant region-matching approach in narrow-band endoscopy. An off-line machine learning process was subsequently added [32]. More recently, a learning-based tracker [38] was applied for on-line biopsy re-targeting on in-vivo gastrointestinal endoscopic videos [33]. Region matching is only efficient if both cameras have the same modality, and cannot be used in our case to register the hyperspectral data on the endoscopic white light image. Another way is to estimate the 3D relative pose of the dual cameras for reprojecting the biopsy site using epipolar geometry. Allain et al. [36] used intersecting epipolar lines assuming a local affine transformations in single modality images. A visual simultaneous localization and mapping (vSLAM) algorithm was used by [34] with a fluorescence augmented prototype. By tracking the pose of the camera and the location of salient points on the organ wall, a 3D map of the organ being explored is generated. Then the biopsy site is targeted by 3D-to-2D reprojection assuming the camera is relatively static when the biopsy is taken and the organ is a rigid environment. A variant with a compensated rhythmic movement was subsequently proposed [35], and validated on a two minutes long stereo laparoscopic video with a silicon phantom. However, the use of these techniques remains rather unreliable in the gastrointestinal tract due to the difficulty of tracking salient features since endoscopic images are very poorly textured. Additionally, it is well known that vSLAM has critical limitations due to cumulative error along the trajectory (3D reconstruction drift) leading to inconsistent 3D-maps. This technique is also too computationally and memory-intensive for an on-line use in an operating room. More recently, a work has jointly exploited chromoendoscopy and color channels to increase the number of salient points for better 3D reconstruction [37].

Due to the lack of texture in the white light endoscopic images, it was decided to reproject the targeted biopsy onto the 2D mosaic from the relative pose of the instrument, without creating a 3D model of the organ. This method is based on a reliable 3D pose tracking of the free-moving optical probe. Markerless approaches for tracking surgical tools [39, 40, 41, 42] have historically been introduced as an interesting alternative to electromagnetic or optical markers [43, 44], because they can be fully implemented in software, do not interfere with the surgical workflow and do not require modifications to the tracked instrument. The latest advances in deep neural network segmentation make them more effective for detecting and tracking surgical tools in in-vivo scenarii [45, 46].

This work makes use of our previously presented markerless tracking algorithm [46], which makes it possible to achieve high precision in the localization of biopsies, even with low frame rate acquisition, and remains applicable to any cylindrical optical probe inserted through the operating channel of a standard endoscope. The following section develops our method.

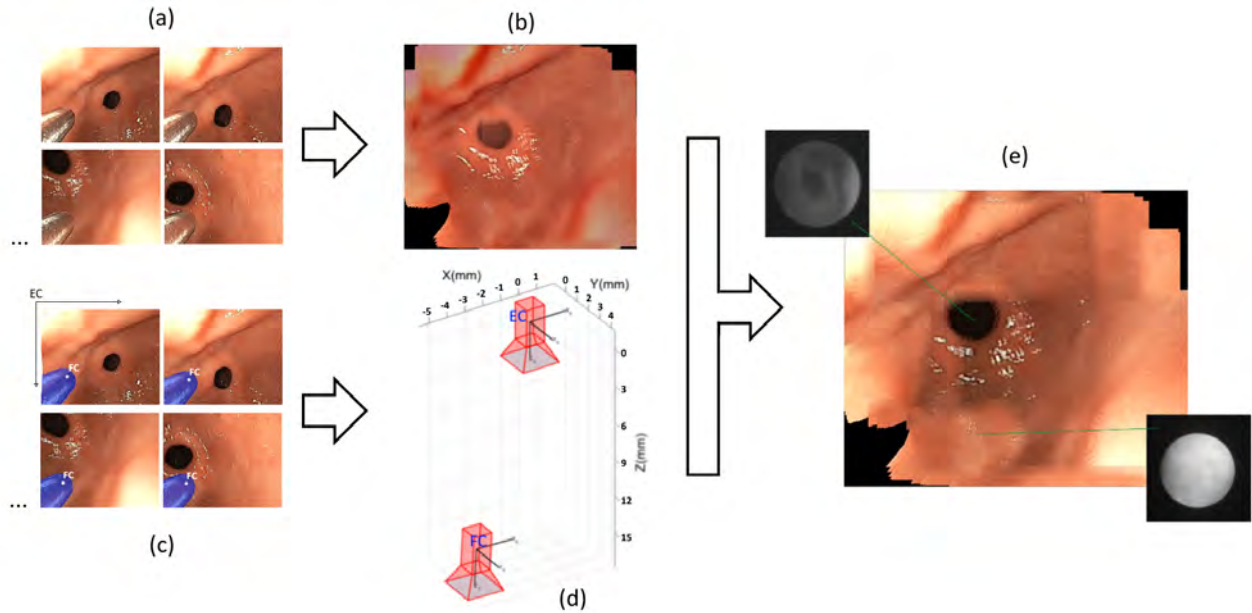


Figure 3: Overview of the processing flow: (a) white-light or NBI endoscopic video, (b) mosaic generation, (c) fiberscope segmentation by convolutional neural network, (d) 3D pose estimation of the optical probe relative to the endoscopic camera, (e) localization of the HSI biopsy in mosaic.

3. Biopsy localization on image mosaics

The overall process for localizing the spectral biopsies on the panoramic mosaic is described in Fig. 3. The endoscopic RGB video is processed to detect the visible tip of the probe and to construct the panoramic mosaic. The 3D pose of the probe estimated from the segmentation allows the biopsies to be located in the mosaic. The hyper-spectral information (HSI) corresponding to each point biopsy can be viewed on-line by the physician or revisited on demand for follow-up.

The proposed method is divided in two main tasks. The first corresponds to the panoramic mosaic generation, while the second is related to the location of the biopsy onto the mosaic based on the instrument pose estimation in individual frames.

3.1. Video mosaicing

This section presents a novel mosaicing scheme specially designed for video-sequences of complex gastroscopic scenes (e.g., the pyloric antrum region of the stomach). A sequence S consists of (time) consecutive images $\{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_N\}$ acquired from different viewpoints. The aim of the mosaicing algorithm is to construct a panoramic image including all scene parts seen in images \mathbf{I}_i . The mosaicing algorithm takes the following scene and acquisition conditions into account:

1. The endoscope's distal tip (with the CCD-sensor matrix) is close to the epithelial surface so that gastroenterologists can observe details in high resolution images. The pyloric antrum region is globally far to be planar. However, due to the short acquisition distances, the surfaces are locally smooth enough to consider that the images include planar surface parts

and that the homologous pixels of two “neighbour images” are geometrically linked by a field of almost parallel vectors.

2. During a gastroscopy, numerous parts of the pyloric antrum surface are usually observed several times. Thus, there exists a large number of overlapping images in a sequence.
3. The image content and quality varies along the sequences. Some images are without textures, while in other images the texture amount, shape and contrast are not uniform.
4. The illumination conditions are changing strongly from one image to another, notably due to camera viewpoint changes and/or varying surface orientations.

The proposed mosaicing algorithm consists of four steps. In the first step one estimates the trajectory of the image centers in a 2D plane whose coordinate system is defined by the first image of sequence S . Then, this trajectory is exploited both to select a subset of images in S which allows to cover the largest tissue area and to find the reference image which minimizes pixel misalignment (This image also defines the coordinate system of the mosaic). The aim of the next step is to determine the shortest paths from the selected images to the reference image to limit the texture discontinuities during the mosaicing process. In the final step, the pixels of the selected images are placed onto the mosaicing plane using a method which minimizes the color discontinuities.

3.1.1. Determination of the image center trajectory

The coordinate system of the image center trajectory plane is that of image \mathbf{I}_1 . As justified in point 1) the displacements of homologous pixels of consecutive images can be approximated by parallel translations.

The displacement vector $\mathbf{v}_{n,n+1}$ linking the centers of two consecutive images \mathbf{I}_n and \mathbf{I}_{n+1} is computed as follows. The pyramidal Lucas-Kanade feature tracker algorithm [47] is used to compute the (sparse) optical flow of corner features detected by the Shi-Tomasi detector [48]. This flow field between \mathbf{I}_n and \mathbf{I}_{n+1} is determined in a square of 50×50 pixels placed on the center of image \mathbf{I}_n . If it is possible to track at least 10 points, then vector $\mathbf{v}_{n,n+1}$ is given by the mean of the vectors of the sparse flow field. On the contrary, if the Lucas-Kanade method led to less than 10 correspondences, then the algorithm detailed in [49] is used to determine a dense flow field between \mathbf{I}_n and \mathbf{I}_{n+1} for a square of 200×200 pixels centered in image \mathbf{I}_n . Vector $\mathbf{v}_{n,n+1}$ corresponds also to the average flow field vector. Thus, when textures are available, the Lucas-kanade method allows for a fast image motion estimation, whereas for images with very poor information, the illumination invariant optical flow method conceived for textureless scenes [49] is used to robustly find the image motion. The image center coordinates P_{n+1} of image \mathbf{I}_{n+1} in the trajectory plane are defined by:

$$P_1 = (0, 0) \text{ and } P_{n+1} = P_n + \mathbf{v}_{n,n+1}. \quad (1)$$

3.1.2. Image selection

This step has two objectives: on the one hand one have to select the images which lead to a mosaic covering the largest possible tissue area, and on the other hand, it is important to select the reference image defining the mosaic coordinate system and which minimizes the texture discontinuities in the mosaic. The convex hull of the trajectory points $\mathcal{P} = \{P_1, P_2, \dots, P_N\}$ facilitates the selection of images covering a large tissue area. Indeed, the vertices of this convex hull correspond to the images which are on the periphery of the scanned organ area. Moreover, by considering the whole pyloric antrum area and the tissue areas seen in the images, the images corresponding to the vertices alone cover a significant part of the surface to be mosaicked.

A convex hull is defined by the set of vertices $\mathcal{P}_c = \{P_1^c, P_2^c, \dots, P_K^c\}$ which correspond to images $\mathcal{S}_c = \{\mathbf{I}_1^c, \mathbf{I}_2^c, \dots, \mathbf{I}_K^c\}$. In the proposed method, the reference image \mathbf{I}_{ref} is determined using the set \mathcal{S}_c of vertices:

$$P_{ref}^c P_0 = \max\{P_k^c P_0\}_{k=1}^K, \quad (2)$$

where $P_0 = \frac{1}{N} \sum_{i=1}^N P_i$ is the centroid of \mathcal{P} and P_{ref}^c is the image center associated to \mathbf{I}_{ref} .

The set of the shortest paths from the reference vertex P_{ref}^c to all other vertices allows to cover at most the scanned area while minimizing the image superimposition (i.e., the texture discontinuities) into the mosaic plane.

3.1.3. Determination of the geometrical link between the selected images and the reference image

During the mosaicing process, the pixels of the images in $\mathcal{S}_c \setminus \{\mathbf{I}_{ref}\}$ are placed in the coordinate system of \mathbf{I}_{ref} using homographies. For the specific shape of the pyloric antrum regions two situations can occur when searching for the geometrical relationship between the coordinate system of a vertex image $\mathbf{I}_k^c \in \mathcal{S}_c \setminus \{\mathbf{I}_{ref}\}$ and the reference vertex image \mathbf{I}_{ref} : either the two vertices are close enough so that the corresponding images can directly be superimposed by homography $\mathbf{H}_{k \rightarrow ref}$, or the two vertices are distant and an intermediate (non-vertex) image \mathbf{I}_m located close to the centroid P_0 acts as a gateway between the images \mathbf{I}_k^c and \mathbf{I}_{ref} . For this last situation, the homography linking the two vertices is given by $\mathbf{H}_{k \rightarrow ref} = \mathbf{H}_{m \rightarrow ref} \mathbf{H}_{k \rightarrow m}$.

Following condition is used to check whether an overlap area $\tau_k = \mathbf{I}_{ref} \cap \mathbf{I}_k^c$ (in pixels) is large enough or not to link two images \mathbf{I}_{ref} and \mathbf{I}_k^c with a homography :

$$\begin{cases} -W < v_k^1 < W \\ -H < v_k^2 < H \\ \tau_k = (W - |v_k^1|)(H - |v_k^2|) > \frac{WH}{2} \end{cases} \quad (3)$$

where (v_k^1, v_k^2) are the components of vector $\overline{P_{ref}^c P_k^c}$, W stands for width of the images and H corresponds to the height of the images. The two first equations in (3) ensure that $\mathbf{I}_{ref} \cap \mathbf{I}_k^c \neq \emptyset$. If condition (3) is satisfied, then homography $\mathbf{H}_{k \rightarrow ref}$ is directly computed with the homologous points between \mathbf{I}_{ref} and \mathbf{I}_k^c . Otherwise, $\mathbf{H}_{k \rightarrow ref}$ is indirectly computed through the intermediate image \mathbf{I}_m . The latter is the image in \mathcal{S} such that the distance $P_m P_0$ is the smallest. Since \mathbf{I}_m lies in a central position on the mosaic plane, it is most often overlapped both by \mathbf{I}_{ref} and \mathbf{I}_k^c . If both image pairs $(\mathbf{I}_{ref}, \mathbf{I}_m)$ and $(\mathbf{I}_m, \mathbf{I}_k^c)$ fulfill (3), then $\mathbf{H}_{k \rightarrow ref} = \mathbf{H}_{m \rightarrow ref} \mathbf{H}_{k \rightarrow m}$. In the very rare cases where one of these overlap conditions is not verified, the pixels of vertex image \mathbf{I}_k^c are not placed in the mosaic coordinate system. As shown by the results, this case is infrequent enough not to create gaps in the mosaics.

In the proposed algorithm, the homography between two overlapping images is determined using the dense point correspondence provided by the optical flow field estimated with the method described in [49].

3.1.4. Mosaic construction

In this final step, the selected images $\mathbf{I}_k^c \in \mathcal{S}_c \setminus \{\mathbf{I}_{ref}\}$ are placed in the coordinate system of reference image \mathbf{I}_{ref} and optimal seams are determined to minimize colour discontinuities in the mosaic. $\mathbf{I}_k^{c,warped}$ refers to image \mathbf{I}_k^c which is warped by homography $\mathbf{H}_{k \rightarrow ref}$, and placed in the mosaic plane \mathbf{I}^{mosaic} whose coordinate system is defined by image \mathbf{I}_{ref} .

Initially, \mathbf{I}^{mosaic} corresponds to \mathbf{I}_{ref} . The first image used to extend \mathbf{I}^{mosaic} is the one belonging to \mathcal{P}_c and being the farthest from \mathbf{I}_{ref} . Farthest image \mathbf{I}_t^c is given by the greatest distance $P_{ref}^c P_t^c$ determined as follows:

$$P_{ref}^c P_t^c = \max\{P_k^c P_{ref}^c \mid P_k^c \in \mathcal{P}_c\}. \quad (4)$$

Then, the mosaic grows by placing pixels of the warped images into the panoramic image:

$$\mathbf{I}^{mosaic} \leftarrow \mathbf{I}^{mosaic} \oplus \hat{\mathbf{I}}_t^{warped}, \quad (5)$$

where \oplus is a blending operation, and $\hat{\mathbf{I}}_t^{warped}$ is defined by

$$\hat{\mathbf{I}}_t^{warped} = \mathbf{I}_t^{c,warped} \cap D(\mathbf{I}_t^{c,warped} \setminus \mathbf{I}^{mosaic}, se). \quad (6)$$

In (6), D is a morphological dilation operator which uses the structuring element se . $\hat{\mathbf{I}}_t^{c,warped}$ includes the pixels of $\mathbf{I}_t^{c,warped}$ which are not overlapped with \mathbf{I}^{mosaic} , as well as the pixels of a small common area of \mathbf{I}^{mosaic} and $\mathbf{I}_t^{c,warped}$ used for the blending operation.

The remaining images in $\Omega = S \setminus \{\mathbf{I}_{ref}, \mathbf{I}_t^c\}$ are iteratively added to the mosaic. At each iteration i , the image \mathbf{I}_i having the smallest intersection with \mathbf{I}^{mosaic} is selected, and some of its pixels are added to \mathbf{I}^{mosaic} using (5) (i.e., $\mathbf{I}^{mosaic} \leftarrow \mathbf{I}^{mosaic} \oplus \hat{\mathbf{I}}_i^{warped}$).

In this way, the visual coherence of the mosaic is maximized since the discontinuities caused by the seams are minimized.

3.2. Biopsy localization

The location of the biopsy on the mosaic is based on the markerless tracking algorithm of the instrument previously presented in [46]. This tracking capability allows for a reliable 3D pose estimation of the free-moving optical probe relative to the endoscopic camera. The probe tip is detected in the endoscopic view by a deep convolutional neural network segmentation. Then, the 3D pose is estimated frame-by-frame by an Expectation-Maximization algorithm from the segmented silhouette assuming a cylindrical shape with a known radius. A prior alignment of the cylinder axis based on Plücker coordinates makes the convergence faster and more reliable. Our approach demonstrated a high precision on a simulated trajectory mimicking endoscopic exploration with a deviation of 0.33 ± 0.68 mm in probe position and 0.32 ± 0.11 degree in the axis orientation. More details on the tracking algorithm can be found in the original paper [50].

From the estimated 3D pose, multi-modal registration is performed by applying a homographic transformation from the fibroscopic image to the endoscopic image, assuming that the small target area of the organ is almost flat. This process is illustrated in Fig. 4. A realistic configuration of the two cameras observing the wall is simulated in Fig. 4(a). The rectangular dot pattern on the wall represents a 9×7 mm² area. The reference point EC_{REF} is the intersection of the optical axis of the endoscopic camera with the wall. The projection of the rectangular dot pattern can be observed in the fibroscope image plane Fig. 4(b) and in the endoscope image plane Fig. 4(c), as well. The registration is established by the

homography linking the two views from the knowledge of the relative 3D pose of the two cameras, assuming an average distance to the wall. A result of multimodal registration of real endoscopic images is shown in Fig. 5.

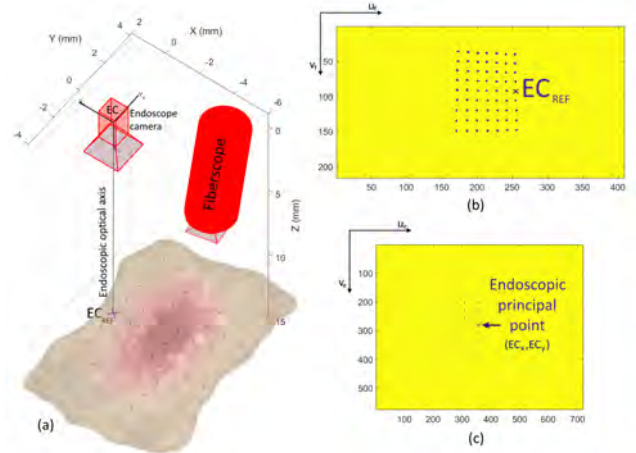


Figure 4: Illustration of multimodal registration: (a) simulation of a realistic configuration of the two cameras observing a rectangular dot pattern lying on the wall, (b) projection of the points in the plane of the fibroscopic image, and (c) same points projected in the plane of the endoscopic image.

The accuracy of this registration was quantitatively assessed by the acquisition of real images of a chessboard pattern. Fig. 6 shows six examples with wide variations in the configuration of the two cameras obtained by moving and sliding the fibroscope in the operating channel; some configurations, such as (e), go far beyond what the surgeon can do in standard endoscopic exploration.

The true pose can be obtained from the chessboard pattern and compared to the one estimated from the segmentation of the fibroscope according to the method detailed in [46]. The pose estimation error is smaller than 3 pixels, i.e. less than 0.33% of the diagonal of the endoscopic image, which is sufficient to ensure a precise overlapping of the RGB and HS images (See Fig. 5).

Multi-modal image registration allows the surgeon to visualise the spectral bands of the tissue being examined as presented in Fig. 7. The average spectrum is calculated over all correctly exposed pixels of the examined region (saturated pixels are automatically discarded by a level-set approach). A prior spectral calibration was performed to transform the raw data captured by the sensor into reflectance values [7]. This data has the potential to be used for the discrimination of healthy and pathological tissues [8, 51]. In these pilot studies in patients with gastritis, changes in visible and near-infrared reflectance spectra were observed that may be associated with inflammatory gastric conditions.

The frame-by-frame registration process can be applied to all images selected for the panoramic mosaic to locate the biopsies. The spectral biopsies corresponding to the selected images \mathbf{I}_k^c are warped by homographies $\mathbf{H}_{k \rightarrow ref}$ and

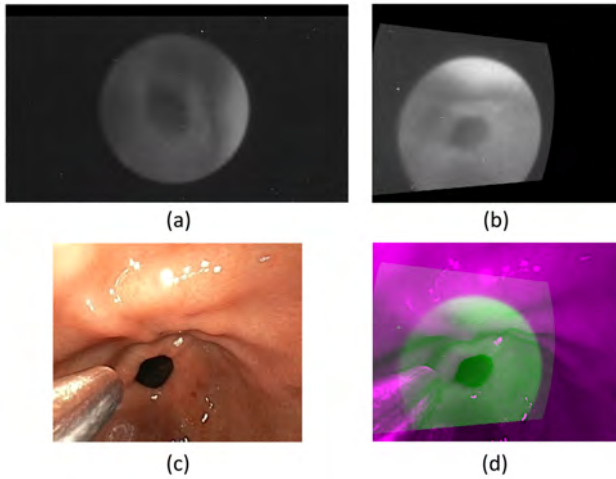


Figure 5: Multi-modal image registration : (a) Original HS image, (b) HS image warped in the coordinate system of the RGB image, (c) Original endoscopic RGB image and (d) Superimposed multi-modal data.

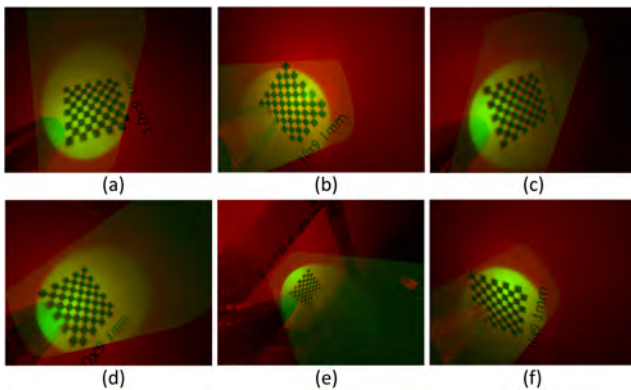


Figure 6: Six different cases showing the correct registration of the images with the chessboard pattern: the HS image (green) is superimposed on the RGB image (red). (e) represents an extreme pose configurations beyond reality.

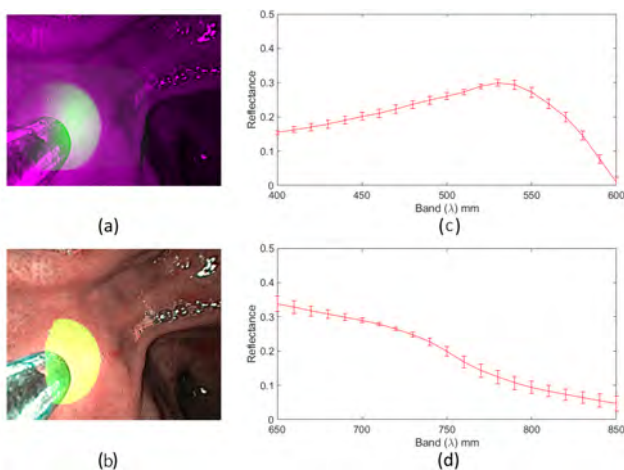


Figure 7: HS-RGB overlay and corresponding reflectance spectral data:(a-b) visible spectrum, (c-d) near-infrared spectrum.

placed in the mosaic plane \mathbf{I}^{mosaic} whose coordinate system is defined by image \mathbf{I}_{ref} .

4. Experimental validation

This section presents the experimental results to validate the proposed method on patient endoscopic sequences.

4.1. Data acquisition

The data acquisition was performed at Ambroise-Paré Hospital (AP-HP) located in Boulogne-Billancourt (France), using a commercial endoscopic unit (Olympus EVIS EX-ERA III) with gastrointestinal videoscope, augmented with a microFlex m25-2500 fiberscope from IT Concepts GmbH connected to a hyperspectral snapshot camera of XIMEA (See fig. 1).

The prototype is designed for the simultaneous acquisition of two video streams of white light and spectral images. The RGB videos were captured with a resolution of 720×576 pixels. The HS camera connected to the fiberscope provides video sequences of resolution 409×216 pixels for each of the 41 spectral bands in the range of 470 to 975 nm.

Automatic segmentation of the fiber-optic instrument in endoscopic videos is processed by the deep convolutional neural network Deeplabv3+. This is the third version of a popular deep convolutional neural architecture for semantic segmentation. Compared to previous versions, modifications are introduced to handle the problem of multiple scale segmentation by using Atrous convolution and encode the global context from image-level features for improving performance. More details on the CNN's architecture can be found in its original paper [52].

The initial model pre-trained on the PASCAL VOC dataset was fine-tuned on a database comprising a total of 10 218 endoscopic frames of 8 different patients. The videos include typical challenges of endoscopic vision such as non-rigid movements and deformations of the organ, bleeding, mucus, etc. The fiberscope used has a very bright and reflective metal head, which represents a serious difficulty for the segmentation task. The only visible instrument is a cylindrical fiberscope (superposition of instruments is a situation that does not occur in this context of optical biopsy).

Our deeplabv3 based segmentation model has been evaluated in our previous paper [46]. It provides a robust detection of the probe tip in various illumination conditions, as it produces a Jaccard similarity index with ground truth manually outlined of more than 0.98 in mean on a total of more 5000 images of the test base.

The chosen stomach parts to show the reliability of the biopsy location on panoramic mosaic were the pyloric antrum and cardia which are critical sites for precancerous lesions. The results are extracted from different videos with various trajectories, number of frames and duration, and large illumination and tissue colors changes according to the patient.

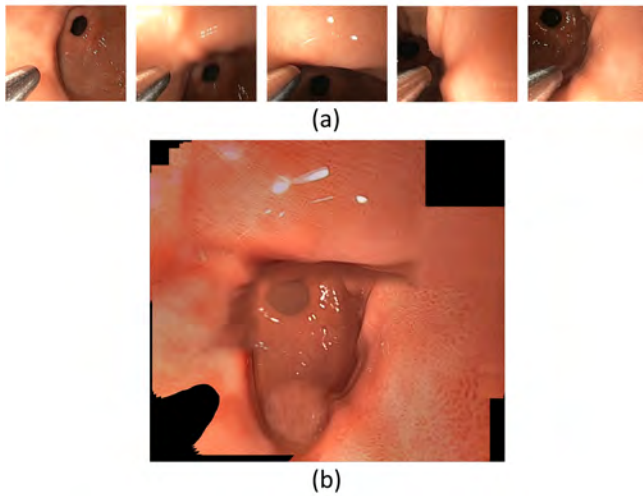


Figure 8: Illustration of the mosaicing method: (a) five images from the white-light video sequence, (b) obtained panoramic view

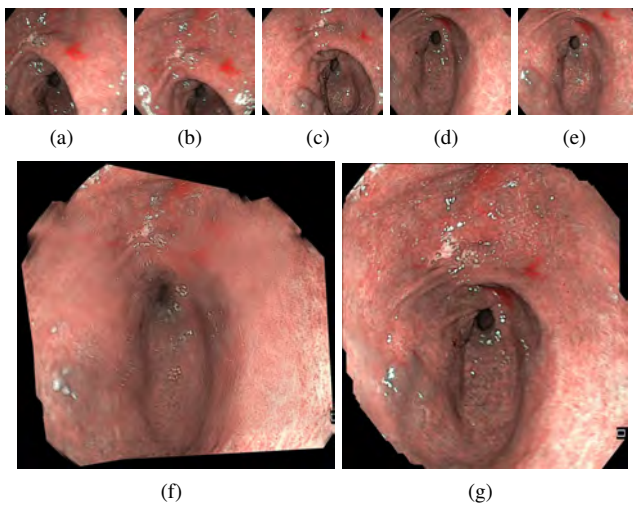


Figure 9: Comparison of reconstructed mosaics of the pyloric antrum in NBI: (a)-(e) five images from the 90 images in the sequence. (f) with Brown and Lowe's method [53] and (g) with the proposed method

4.2. Mosaic construction

4.2.1. Image mosaicing in the context of optical biopsy

The panoramic image in Fig 8(b) shows that the distal tip of the fiberscope which is visible in Fig 8(a) can be segmented and removed, and that the remaining pixels are sufficient to build a mosaic. This 2D map was obtained from almost textureless data and is without colour discontinuities, despite the changing white light illumination conditions and specular reflections.

4.2.2. Independence from the image modality

The panoramic image given in Fig. 9(g) was determined with 90 images of the pyloric antrum seen in the narrow band imaging modality (five small field of view images of the sequence are shown in Fig. 9(a)-(e)). The white light

Table 1

Characteristics of data used in the mosaics presented in the Fig. 10

Fig.10	Number of frames	Number of selected images	Index of the reference image	Computation time (in seconds)
(a)	139	32	103	18
(b)	135	28	210	11
(c)	154	31	213	12
(d)	131	31	150	15

mosaic in Fig. 8(b) and the NBI mosaic in Fig. 9(g) were both obtained with the mosaicing method described in Section 3.1 and with the same algorithm parameter values. Figs. 8 and 9 illustrate the ability of the proposed method to provide contrasted and visually exploitable mosaics for different texture amount and data qualities (with almost no textures in the white light modality, and with different textures in NBI). This modality independence is confirmed by additional results given in Section 4.2.4. It is also noticeable in Fig. 9(f) that, in comparison to standard SIFT based mosaicing approaches [53], the proposed method leads to significantly less blurred or smoothed mosaics which are more appropriate for a diagnosis. The comparison of the proposed mosaicing approach with a reference method is detailed in Section 4.2.4.

4.2.3. Robustness, constancy and computation time of the mosaicing algorithm

The four mosaics given in Fig.10 were constructed using the same video sequence of the pyloric antrum region. It comprises 652 images acquired in the white light modality. The mosaics are built with different tracked region of interest (50x50 pixel region whose position is marked by a gastroenterologist or randomly chosen, see Section 3.1.1), and varying starting and end images in the video-sequence. The blue rectangular boxes correspond to the regions to be tracked in the first sequences of images (upper left images in Fig. 10(a)-(d)). The green rectangular boxes outline the final position of the tracked region. The yellow curves represent the trajectory of the center of the tracked region in the coordinate system of the first image.

Table 1 gives information relating to the mosaics presented in Fig.10: the total number of frames on the sub-sequences extracted from the complete sequence of 652 images; the number of images automatically selected for the mosaic construction, the index of the automatically chosen reference image, as well as the computation time (in seconds). The image selection step described in Section 3.1.2 typically reduces by a factor of 4 or 5 the amount of images effectively used for the mosaicing. For instance, in figure 10.(a), 32 images were selected among a sub-sequence of 139 images, which corresponds to an image reduction by a factor of 4.34. After the tracking phase, a 2D mosaic is available after some seconds (from 10 up to 20 seconds). Even if this delay is appropriate for a clinical use, is it noticeable that the program running on a laptop equipped with an Intel(R) Core(TM) i7 2.80 GHz processor and a 16 GByte RAM written with the C language was not optimized. Besides

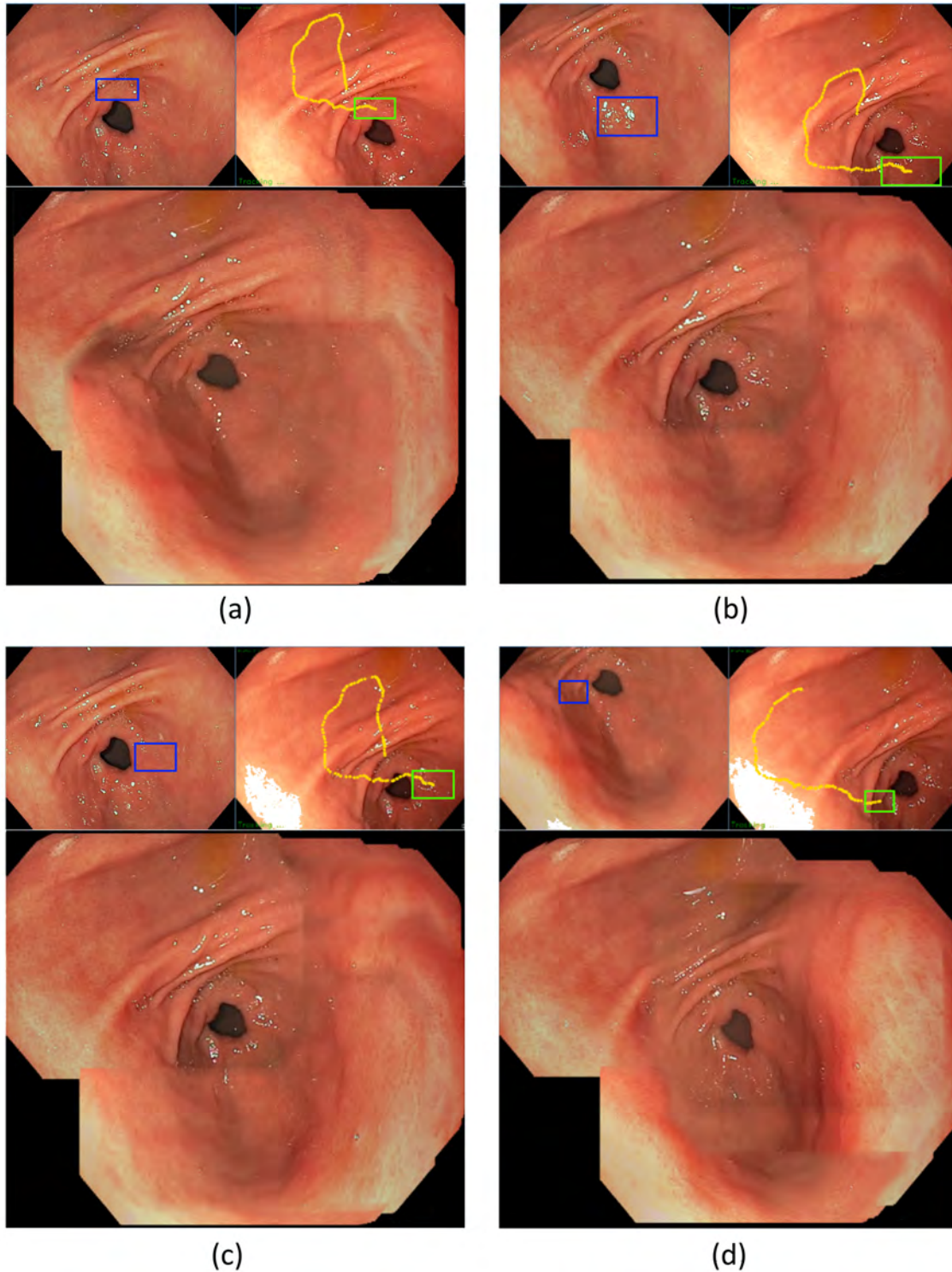


Figure 10: Illustration of the algorithm constancy. The mosaics given in the lower parts of subfigures (a), (b), (c) and (d) were all computed with the same video-sequence. The images with the blue (upper left images) and green (upper right images) rectangular boxes represent the first and last frame of the sub-sequence used for the mosaicing respectively, while the yellow curves (upper right images) show the endoscope trajectory in the coordinate system of the first frame during the image tracking phase (see section 3.1.1). Tables 1 gives information like the automatically chosen reference image index, the number of selected vertex images (see section 3.1.2), and the mosaicing time (time taken by of the two steps described in Sections 3.1.3 and 3.1.4). It is noticeable that for all four subsequences, the mosaics are visually similar and well contrasted, making them exploitable for a diagnosis.

algorithmic and numerical improvements, the algorithms can be parallelized using CUDA (no GPU card equipped the laptop used to assess the method). The authors believe that after the tracking, a mosaic can be displayed in some few seconds. In the four examples shown in Fig. 10, it is visible that the tracked boxes are located in four different places around the opening on the duodenum (dark “circular shape” in the image centres). Moreover, these regions include very few textures and one of them (see the blue box of subfigure 10(b)) is even affected by numerous specular reflections. Even in this situation, the real-time tracking was successful and contributes to the overall robustness of the mosaicing. Moreover, when comparing the images with the blue and green boxes, it is also noticeable that there are significant illumination (intensity) changes between the images. The dense optical flow method [41] used in the homography determination step in Section 3.1.3 contributes greatly to the algorithm robustness. The mosaics given in Fig. 10 and Table 1 show that the mosaicing method provides panoramic views which are visually very close, even if different reference and vertex images were selected with the method detailed in Section 3.1.2. It is also noticeable that for different tracking trajectories into the mosaicing plane (see the shapes of the yellow trajectory curves in the subsequence end image with the green rectangle) the mosaics remain visually similar and contrasted. These observations indicate that the proposed method ensures constancy in terms of visual content of the mosaics for a given patient.

4.2.4. Comparison with a reference method

It is impossible to quantitatively compare two mosaicing methods since for gastroscopic data no ground truth is available (the ideal mosaic to obtain for patient data cannot be known). However, a visual comparison can be done with the results obtained by a SIFT based reference method on which numerous mosaicing methods mentioned in Section 2.1 (related work) are based [16]. It can be seen in Fig. 11 (see sub-figures (a), (d) and (g)) that the mosaics obtained with Brown and Lowe’s approach [16] are blurred. This is mainly due to the fact that all pixels of all images are used to build the mosaics. Such an approach inevitably leads to a significant loss of contrast into the mosaics. This loss of contrast is very detrimental to the visualization of inflammations. It is also noticeable that when using all images, the pixel placement errors into the mosaic reference coordinate system accumulate themselves and become high. In the mosaics of Figs 11.(c), (f) and (i) these accumulation errors were minimized, which leads to larger mosaic surfaces. The fact that the proposed algorithm both searches for the best reference image for the mosaicing (this image determines the mosaic coordinate system) and determines the smallest set of images to cover a large surface effectively optimizes the contrast into the mosaics. It is also noticeable that the methods based on [16] often fail to build mosaics of the pyloric antrum region in the white light modality. Figure 11 confirms that the proposed approach is efficient both for

white light sequences and narrow band imaging videos. Contrast preservation is essential to allow for the visualization of inflammations, to anticipate more serious lesions such as tumours, and to report optical biopsy information on it.

4.3. Optical biopsy mapping

Fig. 12 shows the optical biopsy locations in a white light mosaic in the pyloric antrum region of the stomach. This part of the stomach is a frequent site of chronic inflammation that can lead to the development of cancerous lesions. The collection and localization of hyper-spectral optical biopsies can improve early detection of inflammation, thus anticipating the risk of cancer. The mosaic in Fig. 12(a) was built from a set of homographies $\mathbf{H}_{k \rightarrow ref}$ linking the pixels of the selected vertex images \mathbf{I}_k^c to the reference image \mathbf{I}_{ref} , as detailed in Section III.A. Then, the HS images can be registered on their homologous white light or NBI images applying the registration process described in Section III.B. The green squares in Fig. 12(b) represent the centre of the HS images in the mosaic, where optical biopsies are available.

The white light mosaic shown in Fig. 13 is the same as the one in Fig. 12(a), but the location of some of the optical biopsies are given together with their corresponding HS images for one spectral band. The white light modality is the most popular representation for gastroenterologists since these images represent the stomach in the most natural way. The proposed data visualization allows the clinician to explore all spectral bands in localized HS images, even off-line, which may allow detection of inflammation of the gastric mucosa at an earlier stage.

The Fig. 14 shows another sample in NBI mode (green-blue light). Even if the NBI image modality is not systematically used by all gastroenterologists, it is a useful image enhancement of blood vessels and mucosal structures. With numerous gastro-endoscopes it is possible to switch between the white and green-blue light (or other virtual chromo-endoscopy solutions, as blue light imaging of the Fuji endoscopes). The two mosaics of Fig. 13 and 14 were built with the same mosaicing algorithm and the same technique is applied for the localization of the optical biopsies in both cases, in particular the segmentation step. The solution proposed in this contribution can therefore be used to make the link between the three image modalities.

5. Conclusion and perspectives

This contribution is centered on the localization of optical biopsies performed during endoscopic exploration in a panoramic view. It combines a new 2D mosaic technique with a markerless probe tracking to provide the surgeon with a multimodal augmented data visualization for intra operative guidance and biopsy site re-targeting.

The proposed mosaicing algorithm allows to build and record a panoramic view of a particular area explored, like the pyloric antrum region of the stomach, known to be particularly sensitive to chronic inflammation. The mosaic is built by switching, according to the image quality, between a

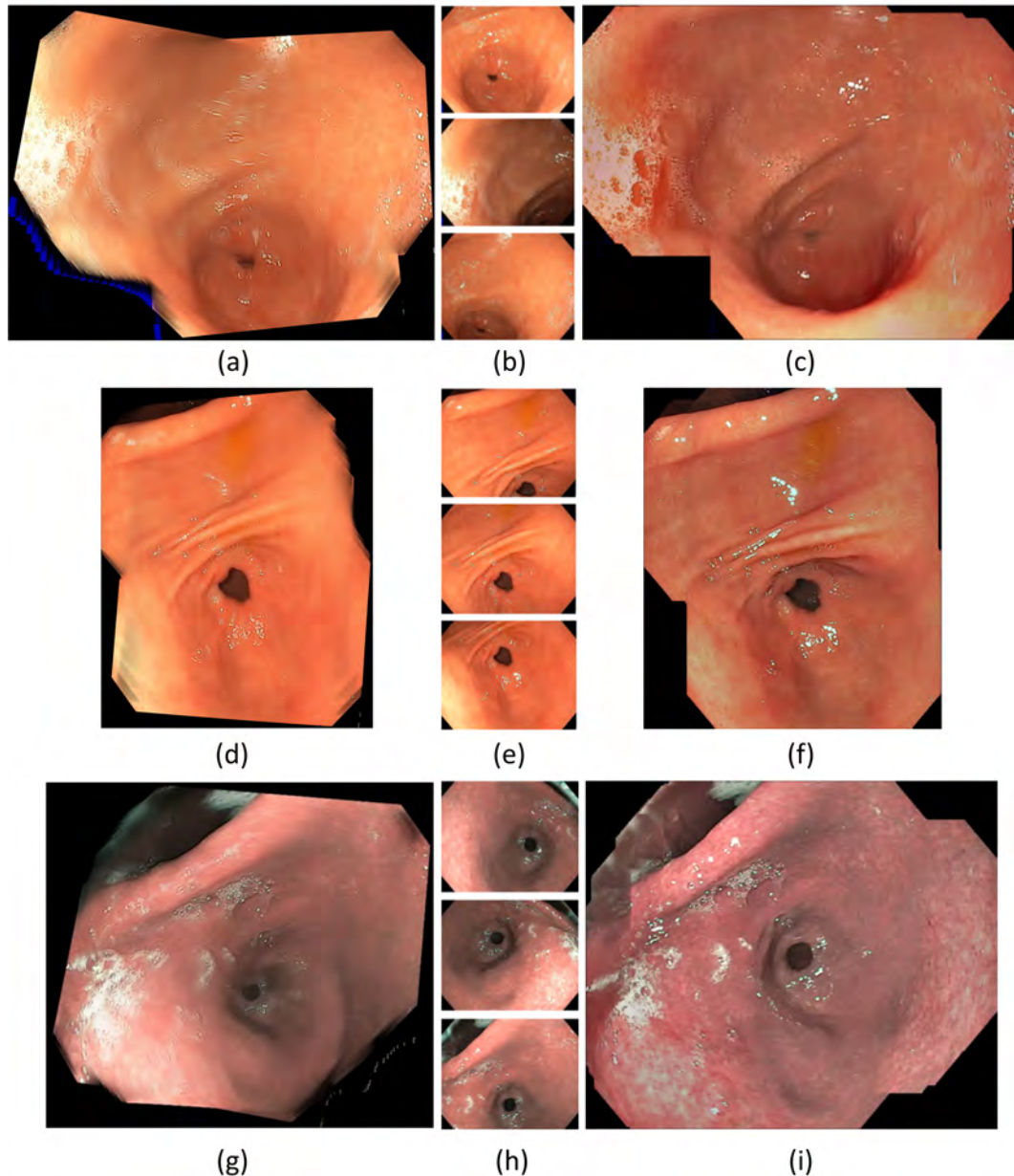


Figure 11: Comparison of the proposed mosaicing method with Brown and Lowe's one [16]. The two first lines give results for two white light sequences of 94 and 79 images respectively, while the third line shows a result for a NBI acquisition consisting of 77 images. Subfigures (b), (e) and (h) give respectively the first, middle and last images of the sequence, and the mosaics in (a), (d) and (g) can be compared to those in (c), (f) and (i) computed with the proposed method. The latter are more contrasted and they represent the organ on larger mosaics, making the diagnosis more easy

sparse or dense optical flow method computation during image tracking step, while an image selection step and a dense optical flow method are used to build a map minimizing colour and texture discontinuities. The resulting panoramic view presents a high visual coherence despite strong lighting variations and specular reflections.

Moreover, the augmented prototype allows the acquisition of optical biopsies by means of a fiberscope slid into the operating channel. The registration of the area targeted by the optical probe in the endoscopic image is made from the frame-by-frame estimation of the 3D pose of the fiberscope

with respect to the endoscopic camera. This tracking is estimated without any marking from an automatic segmentation of the cylindrical silhouette of the probe in the endoscopic image, based on deep learning neural network DeepLabV3+ and works equally well in white light or NBI. We designed a specific data base because the existing databases do not offer the same acquisition configuration with an optical probe inserted in the operating channel and simultaneous acquisition of multimodal data.

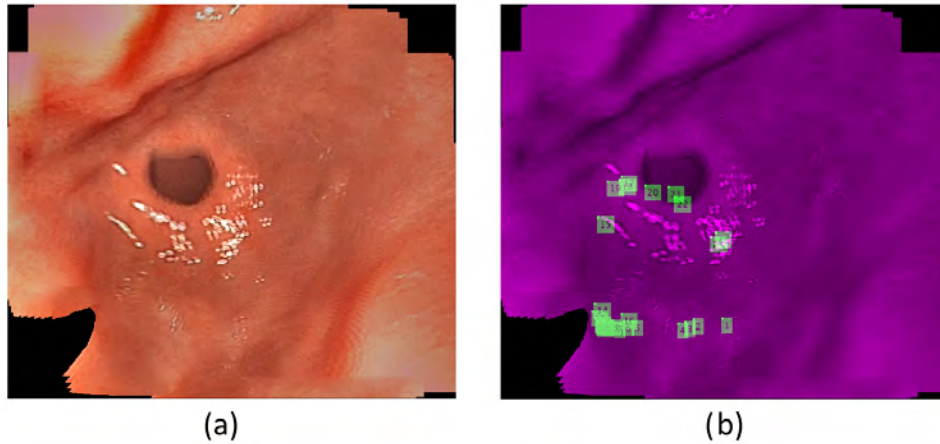


Figure 12: Illustration of the optical biopsy location on the white-light mosaic. (a) White light mosaic. (b) Green squares point to the registered HS biopsy sites.

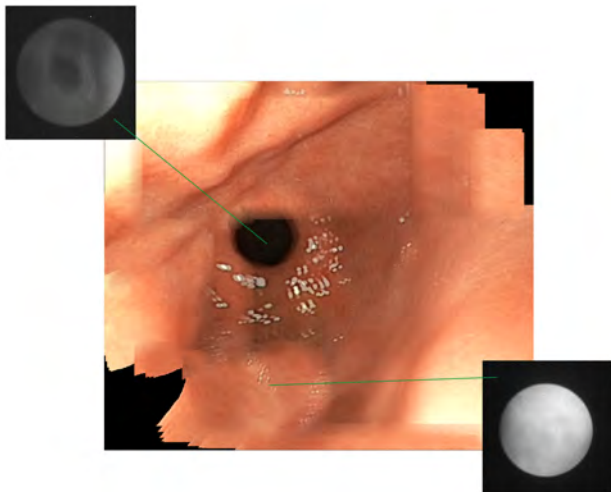


Figure 13: Observation of selected spectral data at biopsy points in the white-light mosaic (grayscale image corresponding to one spectral band).

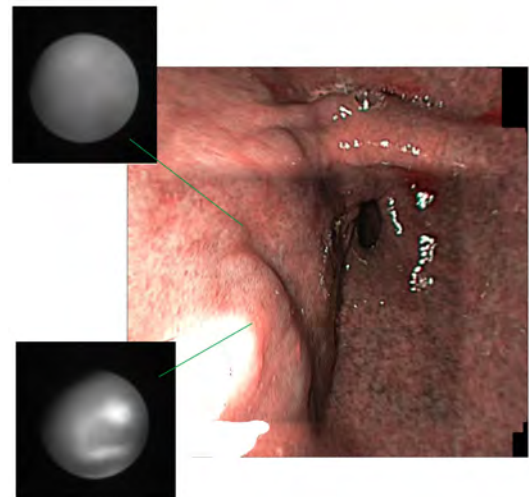


Figure 14: Observation of selected spectral data at biopsy points in the NBI mosaic (grayscale image corresponding to one spectral band).

From the perspective of clinical use, the proposed method is purely software based and does not require any hardware modification. It can thus be adapted to any type of endoscope. Improved intraoperative visualization could help clinicians provide a more rapid and accurate diagnosis, or at least significantly limit the number of incisional biopsies and make sampling less random, by highlighting areas of tissue more likely to have pre-cancerous lesions or inflammation. The recording of the biopsy sites on a panoramic view also facilitates the follow-up of the same patient from one examination to another, as well as sharing information with other colleagues to compare opinions. On the other hand to have an accurate record of the examination may be required by regulatory authorities for forensic purposes. Regarding the optical biopsy method, current data show a correlation between hyperspectral data and diffuse histological features,

but the precision of the analysis and its application to patch lesions is limited by the fact that it is very difficult to find the precise location of the mucosa that was analyzed. Our method of accurately locating the probe on a panoramic view will allow this lock to be lifted and advance in the automated recognition of non-diffuse lesions.

The feasibility of this technique has been demonstrated with a HSI-augmented endoscope prototype. However, this technique can be extended to any other modality camera that could also be connected to a cylindrical optical probe or a fibroscopic instrument, slid into the operative channel. Future works will include an additional functionality for computer-aided diagnosis using hyperspectral data to automatically discriminate healthy and pathological tissues from the optical biopsy and a more extensive clinical study to better quantify the benefit of our technique. 3D image

mosaics can also be built using only 2D endoscopic images and superimposed by markerless biopsy data. However, due to higher computation times, a 3D map [54] can be useful for an accurate off-line examination documentation (and a second diagnosis), while 2D mosaics can be obtained in realtime and facilitate a diagnosis during gastroscopy itself.

Acknowledgments

This study was funded by French National Research Agency (grant ANR-15-CE17-0015).

CRedit authorship contribution statement

Omar Zenteno: Methodology, Software, Investigation and Writing - Original Draft. **Dinh-Hoan Trinh:** Methodology, Software, Investigation and Validation. **Sylvie Treuillet:** Conceptualization and Writing - Review & Editing. **Yves Lucas:** Conceptualization. **Thomas Bazin:** Data curation, Investigation and Validation. **Dominique Lamarque:** Data curation, Resources, Investigation. **Christian Daul:** Conceptualization and Writing - Review & Editing.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] I. A. for Research on Cancer, "Globocan: Cancer today." <https://gco.iarc.fr/>, 2018.
- [3] T. Bergen and T. Wittenberg, "Stitching and surface reconstruction from endoscopic image sequences: A review of applications and methods," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 304–321, 2016.
- [4] R. Wessels, D. de Bruin, D. Faber, T. van Leeuwen, M. Beurden, and T. Ruers, "Optical biopsy of epithelial cancers by optical coherence tomography (oct)," *Lasers in medical science*, vol. 29, 03 2013.
- [5] B. Glover, J. Teare, and N. Patel, "The status of advanced imaging techniques for optical biopsy of colonic polyps," *Clinical and translational gastroenterology*, vol. 11, no. 3, 2020.
- [6] B. Jansen-Winkeln, M. Barberio, C. Chalopin, K. Schierle, M. Diana, H. Köhler, I. Gockel, and M. Maktabi, "Feedforward artificial neural network-based colorectal cancer detection using hyperspectral imaging: A step towards automatic optical biopsy," *Cancers*, vol. 13, no. 5, 2021.
- [7] O. Zenteno, A. Krebs, S. Treuillet, Y. Lucas, Y. Benezeth, and F. Marzani, "Spatial and spectral calibration of a multispectral-augmented endoscopic prototype," vol. 997, pp. 262–280, 2019.
- [8] A. Krebs, Y. Benezeth, T. Bazin, F. Marzani, and D. Lamarque, "Pre-cancerous stomach lesion detections with multispectral-augmented endoscopic prototype," *Applied Sciences*, vol. 10, no. 3, p. 795, 2020.
- [9] A. Behrens, T. Stehle, S. Grossa, and T. Aach, "Local and global panoramic imaging for fluorescence bladder endoscopy," in *31st International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6690–6693, 2009.
- [10] A. Behrens, M. Bommers, T. Stehle, S. Gross, S. Leonhardt, and T. Aach, "Real-time image composition of bladder mosaics in fluorescence endoscopy," *Computer Science - RD*, vol. 26, pp. 51–64, 2011.
- [11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [12] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, pp. 346–359, 2008.
- [13] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [14] N. Shevchenko, J. Fallert, H. K. A. Stepp, H. and Sahli, and T. Lueth, "A high resolution bladder wall map: Feasibility study," in *IEEE International Conference on Engineering in Medicine and Biology Society (EMBC)*, p. 5761–5764, 2012.
- [15] T. Bergen, T. Wittenberg, C. Munzenmayer, C. Chi, and G. Hager, "A graph-based approach for local and global panorama imaging in cystoscopy," in *SPIE Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*, (Florida, United States), pp. 1–7, 2013.
- [16] R. Miranda-Luna, C. Daul, W. Blondel, Y. Hernandez-Mier, D. Wolf, and F. Guillemin, "Mosaicing of bladder endoscopic image sequences: distortion calibration and registration algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 541–553, 2008.
- [17] T. Weibel, C. Daul, R. Röscher, and D. Wolf, "Endoscopic bladder image registration using sparse graph cuts," in *IEEE Int. Conf. on Image Processing, ICIP*, (Hong-Kong, China), pp. 157–160, 2011.
- [18] S. Ali, C. Daul, E. Galbrun, and W. Blondel, "Illumination invariant optical flow using neighborhood descriptors," *Computer Vision and Image Understanding*, vol. 145, pp. 95–110, 2016.
- [19] P. Viola and W.M.Wells, "Alignment by maximization of mutual information," *International journal of computer vision*, vol. 24, pp. 137–154, 1997.
- [20] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/maxflow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, pp. 1124–1137, 2004.
- [21] S. N. M. Drulea, "Motion estimation using the correlation transform," *IEEE Trans. Image Process.*, vol. 22, p. 3260–3270, 2013.
- [22] F. A. Marzotto, R. and V. Murino, "High resolution video mosaicing with global alignment," in *IEEE Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, p. 692–698, 2004.
- [23] T. Weibel, C. Daul, D. Wolf, R. Röscher, and F. Guillemin, "Graph based construction of textured large field of view mosaics for bladder cancer diagnosis," *Pattern Recognition*, vol. 45, pp. 4138–4150, 2012.
- [24] D.-H. Trinh, C. Daul, W. Blondel, and D. Lamarque, "Mosaicing of images with few textures and strong illumination changes: Application to gastroscopic scenes," in *IEEE Int. Conf. on Image Processing*, pp. 1263–1267, 2018.
- [25] D.-H. Trinh, W. Blondel, and C. Daul, "A general form of illumination-invariant descriptors in variational optical flow estimation," in *IEEE Int. Conf. on Image Processing, ICIP*, (Beijing, China), pp. 2533–2537, 2017.
- [26] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of biomedical optics*, vol. 19, no. 1, p. 010901, 2014.
- [27] D. Nouri, Y. Lucas, and S. Treuillet, "Hyperspectral interventional imaging for enhanced tissue visualization and discrimination combining band selection methods," *International journal of computer assisted radiology and surgery*, vol. 11, no. 12, pp. 2185–2197, 2016.
- [28] Z. Han, A. Zhang, X. Wang, Z. Sun, M. D. Wang, and T. Xie, "In vivo use of hyperspectral imaging to develop a noncontact endoscopic diagnosis support system for malignant colorectal tumors," *Journal of biomedical optics*, vol. 21, no. 1, p. 016001, 2016.
- [29] J. Yoon, J. Joseph, D. J. Waterhouse, A. S. Luthman, G. S. Gordon, M. Di Pietro, W. Januszewicz, R. C. Fitzgerald, and S. E. Bohndiek, "A clinically translatable hyperspectral endoscopy (hyse) system for imaging the gastrointestinal tract," *Nature communications*, vol. 10, no. 1, pp. 1–13, 2019.
- [30] N. T. Clancy, G. Jones, L. Maier-Hein, D. S. Elson, and D. Stoyanov, "Surgical spectral imaging," *Medical image analysis*, vol. 63, p. 101699, 2020.

- [31] S. Atasoy, B. Glocker, S. Giannarou, D. Mateus, A. Meining, G.-Z. Yang, and N. Navab, "Probabilistic region matching in narrow-band endoscopy for targeted optical biopsy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 499–506, Springer, 2009.
- [32] S. Atasoy, D. Mateus, A. Meining, G. Yang, and N. Navab, "Endoscopic video manifolds for targeted optical biopsy," *IEEE Trans. Med. Imag.*, vol. 31, pp. 637–653, 2012.
- [33] M. Ye, S. Giannarou, A. Meining, and G.-Z. Yang, "Online tracking and retargeting with applications to optical biopsy in gastrointestinal endoscopic examinations," *Medical Image Analysis*, vol. 30, pp. 144–157, 2016.
- [34] P. Mountney, S. Giannarou, D. Elson, and G.-Z. Yang, "Optical biopsy mapping for minimally invasive cancer screening," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 483–490, Springer, 2009.
- [35] P. Mountney and G.-Z. Yang, "Motion compensated slam for image guided surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 496–504, Springer, 2010.
- [36] B. Allain, M. Hu, L. B. Lovat, R. J. Cook, T. Vercauteren, S. Ourselin, and D. J. Hawkes, "Re-localisation of a biopsy site in endoscopic images and characterisation of its uncertainty," *Medical Image Analysis*, vol. 16, no. 2, pp. 482–496, 2012.
- [37] A. R. Widya, Y. Monno, M. Okutomi, S. Suzuki, T. Gotoda, and K. Miki, "Whole stomach 3d reconstruction and frame localization from monocular endoscope video," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019.
- [38] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, pp. 1409–1422, 2012.
- [39] 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 Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1050–1058, 2012.
- [40] S. Haase, J. Wasza, T. Kilgus, and J. Hornegger, "Laparoscopic instrument localization using a 3-d time-of-flight/rgb endoscope," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 449–454, IEEE, 2013.
- [41] R. Sznitman, C. Becker, and P. Fua, "Fast part-based classification for instrument detection in minimally invasive surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 692–699, Springer, 2014.
- [42] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, "Vision-based and marker-less surgical tool detection and tracking: a review of the literature," *Medical image analysis*, vol. 35, pp. 633–654, 2017.
- [43] M. K. Chmarra, C. Grimbergen, and J. Dankelman, "Systems for tracking minimally invasive surgical instruments," *Minimally Invasive Therapy & Allied Technologies*, vol. 16, no. 6, pp. 328–340, 2007.
- [44] S. Speidel, G. Sudra, J. Senemaud, M. Dreuschew, B. P. Müller-Stich, C. Gutt, and R. Dillmann, "Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling," in *Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling*, vol. 6918, p. 69180X, International Society for Optics and Photonics, 2008.
- [45] X. Du, T. Kurmann, P. Chang, M. Allan, S. Ourselin, R. Sznitman, J. D. Kelly, and D. Stoyanov, "Articulated multi-instrument 2-d pose estimation using fully convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 37, no. 5, pp. 1276–1287, 2018.
- [46] O. Zenteno, P. VanTrung, S. Treuillet, and Y. Lucas, "Markerless tracking of micro-endoscope for optical biopsy in stomach," in *IEEE Int. Conference in Medicine and Biology (EMBC)*, pp. 4419–4422, 2019.
- [47] J.-Y. Bouguet, "Pyramidal implementation of the Lucas Kanade feature tracker," tech. rep., Intel Corporation, Microprocessor Research Labs, 2000.
- [48] J. Shi and C. Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition*, pp. 593–600, 1994.
- [49] D.-H. Trinh and C. Daul, "On illumination-invariant variational optical flow for weakly textured scenes," *Computer Vision and Image Understanding*, vol. 179, pp. 1–18, 2019.
- [50] O. Zenteno, S. Treuillet, and Y. Lucas, "Pose estimation of a markerless fiber bundle for endoscopic optical biopsy," in *J. Med. Imag.*, vol. 8, p. 25001, 2021.
- [51] T. Bazin, A. Krebs, A. Jobart-Malfait, V. Camilo, V. Michel, Y. Benezeth, F. Marzani, E. Touati, and D. Lamarque, "Multimodal imaging as optical biopsy system for gastritis diagnosis in humans, and input of the mouse model," *EBioMedicine*, vol. 69, p. 103462, 2021.
- [52] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *CoRR*, vol. abs/1802.02611, 2018.
- [53] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.
- [54] T.-B. Phan, D.-H. Trinh, D. Wolf, and C. Daul, "Optical flow-based structure-from-motion for the reconstruction of epithelial surfaces," *Pattern Recognition*, vol. 105, p. 107391, 2020.

Graphical Abstract

Optical biopsy mapping on endoscopic image mosaics with a marker-free probe

Omar Zenteno, Dinh-Hoan Trinh, Sylvie Treuillet, Yves Lucas, Thomas Bazin, Dominique Lamarque, Christian Daul

