How Big is this Neoplasia? Live Colonoscopic Size Measurement using the Infocus-Breakpoint

F. Chadebecq[a,b] C. Tilmant[b] A. Bartoli[a]

[a] *ISIT UMR 6284 CNRS / Université d'Auvergne*
Bâtiment 3C, Faculté de Médecine
28 place Henri Dunant, BP 38
63001 Clermont-Ferrand, France

[b] *Institut Pascal UMR 6602 CNRS / Université Blaise Pascal / IFMA*
Complexe Universitaire des Cézeaux
24 avenue des Landais, BP 80026
63171 Aubière cedex, France

Abstract

Colonoscopy is the reference medical examination for early diagnosis and treatment of colonic diseases. This minimally invasive technique allows endoscopists to explore the colon cavity and remove neoplasias - abnormal growths of tissue - which may develop into malignant tumors. The size, shape and appearance of a neoplasia are essential cues for diagnostic. However, the size is difficult to estimate because the absolute scale of the observed tissue is not directly conveyed in the 2D colonoscopic images. An erroneous size estimate may lead to inappropriate treatment. There currently exist no solutions to reproducible neoplasia size measurement adapted to colonoscopy.

We propose a colonoscopic size measurement system for neoplasias. By using a simple planar geometry, the key technical problem is reduced to resolving scale. Our core contribution is introducing the *Infocus-Breakpoint* (IB) that allows us to resolve scale from a regular colonoscopic video. We define the IB as the lower limit of the colonoscope's depth of field. The IB corresponds to a precise colonoscope to tissue distance, called the *reference depth*, which we calibrate preoperatively. We detect the IB intraoperatively thanks to two novel modules: deformable *Blur-Estimating Tracking* (BET) and *Blur-Model Fitting* (BMF). With our system, the endoscopist may in-

teractively measure the length and area of a neoplasia in a 2D colonoscopic image directly. Our system needs no hardware modification to standard monocular colonoscopes, yet reaching a size measurement accuracy of the order of a millimeter, as shown on several phantom and patient datasets.

Keywords: colonoscopy, neoplasia, size, Thin-Plate Spline, registration, Depth-from-Focus.

---

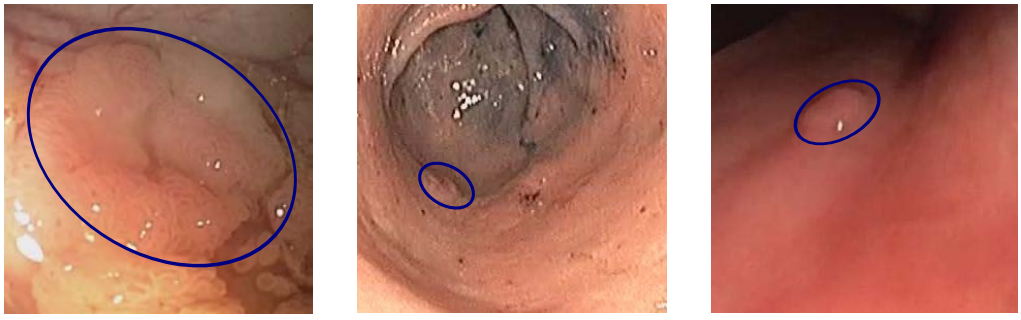
1. Introduction

Clinical problem. Colorectal cancer is the fourth leading cause of death by cancer according to the World Health Organization (World Health Organization, 2008). Regular screening is necessary for early diagnosis and surveillance of abnormal growths of tissue, neoplasias, which may spread and transform into malignant tumors (see figure 1). Despite the development of non-invasive screening techniques such as CT colonography and the Pill Cam®, colonoscopy is currently the reference medical examination (Kim et al., 2007). This procedure consists in inserting a flexible endoscope called a colonoscope through the rectum to explore the colon cavity. The main advantage of this minimally invasive technique is that it allows the endoscopist to excise and biopsy neoplasias during the examination. The visual diagnosis relies on criteria involving the size, shape and appearance of neoplasias. By appearance, we mean the pit pattern of the colonic mucosa as described in the *pit pattern classification* (Kudo et al., 1994). The influence of shape and size is defined in a gold standard, the *Paris classification of superficial neoplastic lesions* (Endoscopic Classification Review Group, 2005). Roughly speaking a neoplasia whose size is larger than 1 cm has a high rate of malignancy, and a neoplasia whose size lies between 5 mm and 1 cm should be reassessed (Talbot, 1996). This size is also essential to determine the surveillance interval time. However, a small neoplasia imaged close to the colonoscope looks like a larger neoplasia imaged at a further depth: this is the scale ambiguity.

In practice, the endoscopist may place a surgical tool next to the neoplasia in order to visually gauge its size. While they could use an endoscopic ruler or a marked biopsy forceps as a measurement reference (Hyun et al., 2011), this technique suffers from unreproducibility and requires a fine manipulation of the surgical instrument (which could deform the colonic wall). It is thus inconvenient and time consuming particularly for interventions such as polypectomy which could require the endoscopist to use different instruments

(for instance only a snare or an electric probe) while the biopsy channel is already occupied by an endoscopic ruler.

It has been shown that an endoscopist's visual estimation of a neoplasia's size is generally biased (Margulies et al., 1994; Schoen et al., 1997; Moug et al., 2010). A recent study (Chaptini et al., 2014) evaluated the visual estimation accuracy among endoscopists and the impact on the surveillance intervals after polypectomy (removal of a neoplasia). This study reveals that erroneous visual estimation[1] leads to 10% of surveillance intervals to be inappropriate. Moreover, the lack of intra and inter-observer reproducibility compromises the cancer risk prediction for neoplasias removed in colonoscopy (Rubio et al., 2010). The lack of a convenient and accurate size measurement facility, and the potential clinical impact it may have form the core motivations for our proposed system. Finally, a neoplasia measurement system may also be used for training purpose (Chang et al., 2010).



(a) Laterally spreading tumor (b) Sessile antral neoplasia (c) Sessile polyp (Oblong)

Figure 1: Images of various neoplasias. (a) is a suspicious neoplasia which may become or already be malignant. (b) and (c) are begnin neoplasias. These three examples show that colonoscopic images are difficult for computer vision. The lack of textural features (or 'natural landmarks') both on the colon cavity and the neoplasias defeats feature-based methods for image registration and 3D reconstruction.

We argue that neoplasia, and more generally colonoscopic, size measurement can be accurately solved from colonoscopic images only using computer vision. This involves resolving scale, a difficult and open technical problem.

[1]The study was realized on 2812 cases. Erroneous visual estimation greater than 20% occured in 52% of the visual estimates.

Active techniques and electromagnetic (EM) tracking. Active 3D sensing techniques such as structured lighting usually resolve scale but require heavy modifications of the colonoscope's hardware (Mirota et al., 2011). They may enlarge the distal end or permanently occupy a biopsy channel. Scale may also be resolved from the magnitude of the colonoscope's distal end motion. In this respect, EM tracking may resolve scale but would, similarly to active 3D sensing, involve substantial hardware changes (Reichl et al., 2013).

Passive techniques. Passive 3D sensing techniques such as Structure-from-Motion (SfM, Hartley and Zisserman (2004)) and Shape-from-Shading (SfS, Trucco and Verri (1998)) use regular images but do not resolve scale and thus do not facilitate absolute size measurements. Resolving scale would require the endoscopist to use a biopsy forceps printed with visual landmarks (as it is difficult to accurately detect visual landmarks on the tip of the tool). The forceps would have to be placed in contact with the neoplasia. This would be inconvenient. Alternatively, the forceps would have to remain fix in the colonoscopic scene in at least two views which is not possible in practice for the forceps moves with the colonoscope. Depth-from-Focus (DfF) uses optical blur as a visual cue to provide absolute measurements (Pentland, 1987). However it has two major limitations which prevent its use in colonoscopy: *(i)* it requires one to control the camera motion and *(ii)* it infers an absolute depth for every pixel. *(i)* is a limitation because accurately servoing the colonoscope's distal end intraoperatively is not possible. *(ii)* is a limitation because it means that noise is not reduced by redundancy and colonoscopic images may be quite noisy due to moisture and blur.

Proposed technical solution. We propose to use optical blur to resolve scale by releasing the two limitations of DfF. Monocular colonoscopes generally house a prime lens (a fixed optical sytem) which forms a focused image for a typical range of distances (generally between 2 and 10 cm), or a dual-focus system which allows the endoscopist to choose between normal and near focus zoom (a two-states mechanical system). The focus of such colonoscopes is thus easily fixed. Our method is based on what we called the *Infocus-Breakpoint* (IB). The IB is the sharp/blur breakpoint which arises while the colonoscope moves toward the tissue. It occurs at the lower bound d_0 of the focusing distance range below which the image of a neoplasia becomes blurred. We call d_0 the *reference depth*.

Our system relies on two novel technical modules: *Blur-Estimating Tracking* (BET) and *Blur-Model Fitting* (BMF). Together, they robustly find the

IB in a video for a specific Region Of Interest (ROI) which contains the neoplasia to be measured. Because the reference depth is constant, it can be calibrated; finding the IB thus resolves scale and facilitates size measurement. In practice, we use our BET module once preoperatively to calibrate the reference depth using a calibration object with known structure. We assume the neoplasia's boundaries to be planar and frontoparallel to the colonoscope's distal end. It is obvious that tracking the neoplasia overcomes DfF's limitation *(i)*. Accuracy stems from the fact that all pixels in the ROI contribute to find the IB, releasing DfF's limitation *(ii)*.

The usage of our interactive neoplasia measurement system is illustrated in figure 2. It relies on simple preoperative and intraoperative protocols which, from our preliminary trials, meet the operating room's constraints. The cleaning and sterilization procedures are not affected as the colonoscope's hardware is not modified. The biopsy channel is left free and thus can be used for the biopsy tools whenever required. Our system's usage time is extremely mild. Preoperative calibration takes less than a minute. Once the neoplasia is selected, computation time is a few dozens of seconds (on a standard computer) before measurements can be made. The accuracy of visual size estimation by endoscopists is of the order of 3 to 4 mm (Summers, 2010); the accuracy of our system is generally within 1 mm.

Contributions. We propose the first passive measurement system applicable to colonoscopy with no hardware modifications. Technically, we propose to extract the IB from a video stream obtained by moving the colonoscope toward the neoplasia. To this end, we introduce a complete optical and geometric model of colonoscopic images. This model is implemented in our BET and BMF modules which, thanks to robust estimation, successfully extract the IB even in the presence of unmodeled artifacts causing local erroneous blur estimation.

This paper markedly extends our previous conference publications (Chadebecq et al., 2012, 2013). First, our previous work relied on a simple non-robust affine BET while our novel BET handles motion blur, colon deformations and image artifacts such as those caused by moisture. Our previous method was more sensitive to noise, causing inaccuracies in optical blur estimation and detection ambiguities in the IB which could sometimes not be resolved. These do not happen with our novel BMF module. Our current system reaches less than 1 mm measurement accuracy while our previous system reached an accuracy of the order of 3 mm.

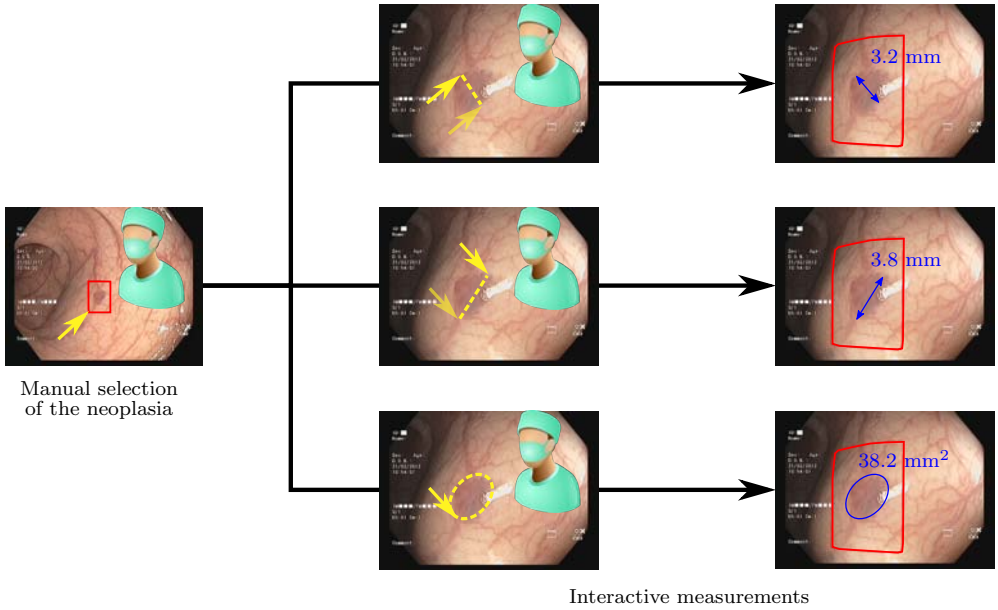


Figure 2: Neoplasia measurement during colonoscopic examination with our system. The endoscopist first selects a neoplasia. They may then make measurements of lengths and areas interactively.

Paper organization. Section 2 reviews state-of-the-art. Section 3 gives an overview of our system. Sections 4 gives our optical model and formally introduces the notion of IB. Section 5 gives our geometric model. Section 6 presents BET and BMF, and their combination for robust IB estimation. Finally, section 7 reports experimental results and section 8 discusses our contributions and future research.

2. State-of-the-Art

2.1. Neoplasia Size Measurement

Resolving scale is a challenging task in 3D computer vision. This is made even more difficult in colonoscopic conditions where the colonoscope cannot be accurately controlled. Some of the active (Hensley et al., 2009; Reichl et al., 2013) and most passive 3D sensing methods (Alcantarilla et al., 2013; Chen et al., 2010; Koppel et al., 2007; Parot et al., 2013) proposed

for colonoscopy focused on the topography of colonic lesions, which may be obtained by reconstructing the shape, without resolving scale.

Active methods. Some of the active methods such as structured light projection resolve scale but require modifications to the colonoscope hardware. A review of active methods can be found in (Mirota et al., 2011). Structured light probes have mainly been designed for rigid endoscopes (Schmalz et al., 2012). This is because of the difficulty to convey a structured light source inside the flexible colonoscope.

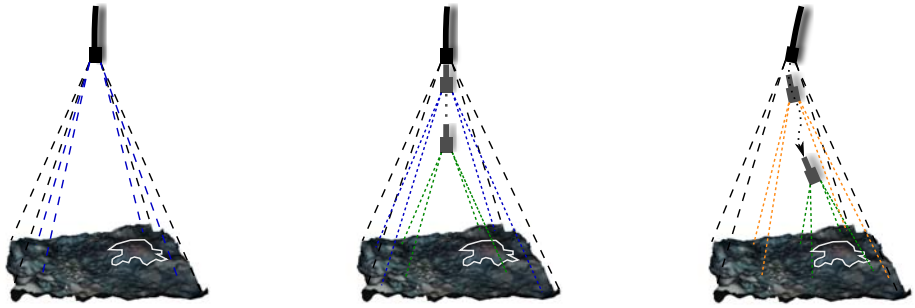
Scale from EM tracking. EM tracking (Reichl et al., 2013; Lee et al., 2010) was also proposed to track the colonoscope's distal end inside the colon cavity. While EM reaches an accuracy of around 2 mm, it occupies a biopsy channel to host the EM 'needle' and an external tracking equipment inside the operating room. Olympus® has recently introduced the ScopeGuide which features EM tracking. It is mainly used for training purposes and due to cost may not replace existing colonoscopes in the near future (Lee et al., 2010).

Passive methods. Passive 3D sensing techniques do not generally resolve scale in monocular optical systems. For instance, a photometric stereo colonoscopic system has been proposed in (Parot et al., 2013). This approach reconstructs the topography of the colonic environment but does not resolve scale. Note that scale is easily resolved in stereo systems. However, stereo has mainly been developed for rigid endoscopes (Field et al., 2009). Depth-from-Optical-Blur techniques apply to monocular systems and resolve scale by servoing some camera parameters; they are reviewed directly below.

2.2. Depth-from-Optical-Blur

Optical blur has been used to measure absolute depth from multiple images. There exist two main lines of work, Depth-from-Defocus (DfD) and Depth-from-Focus (DfF), which were both introduced in (Pentland, 1987). These techniques require a calibration process to relate depth with the amount of optical blur measured in an image. This relationship depends on the camera's internal and external parameters. DfD (Namboodiri and Chaudhuri, 2007; Favaro et al., 2008; Zhuo and Sim, 2009; Zhou et al., 2010) generally requires at least two images taken from the same camera pose with varying but known internal camera parameters (figure 3-a). DfF (Takeshita et al., 2009; Shim and Choi, 2010) requires a set of images created by accurately servoing the camera's displacement (figure 3-b) but keeping fixed

its known internal parameters. For instance DfF has been recently used to reconstruct surface topography in microscopy (Mannan and Choi, 2011).



(a) Requirements for DfD:
- still camera
- controlled optical changes

(b) Requirements for DfF:
- controlled camera motion
- fixed optical setting

(c) Colonoscopic conditions:
- loosely controlled camera
- fixed optical setting

Figure 3: Requirements for (a) Depth-from Defocus (DfD) and (b) Depth-from Focus (DfF), and (c) practical colonoscopy conditions. DfD and DfF cannot be used in colonoscopy because the internal (optical) and external (motion) colonoscope's parameters are respectively fixed and uncontrolled.

Current off-the-shelf DfD and DfF methods are not applicable to colonoscopy (figure 3-c). Directly adapting DfF would require one to modify the colonoscope's hardware to track its distal end and incorporating some image unwarping mechanism to align corresponding pixels. DfD cannot be used because most of the colonoscopes' optical properties are fixed. DfD has been adapted to monofocal cameras (Wei et al., 2009; Wohler et al., 2009) or to a single image (Zhuo and Sim, 2009). However the price to pay is that scale is then unresolved, and the influence of noise increases dramatically (Kaufman and Wang, 2008). The former adaptation would also require one to extract and match image keypoints. This is not adapted to colonoscopic images which generally contain very few reliable keypoints, as illustrated in figure 1 and discussed in (Alcantarilla et al., 2013).

2.3. Proposed Approach

Unlike state-of-the-art methods, our system is adapted to colonoscopic conditions in that it does not require one to accurately control the internal or external camera parameters and does not either require modifications

to the colonoscope's hardware. Our system uses a short colonoscopic video corresponding to an approximate constant speed motion of the colonoscope toward a neoplasia (figure 4). The endoscopists involved in our experiments easily performed this motion (see section 7.3). First, our area-based BET module automatically tracks the neoplasia and estimates coherent and redundant blur change along the video. Second, our BMF module reliably extracts the sharp/blur breakpoint, giving the IB. Our method does not require to estimate absolute optical blur accurately, neither does it require a complex preoperative calibration procedure. Finally, our experiments show the feasibility of our method to handle regular colonoscopic artifacts such as motion blur or noise which are not handled by most state-of-the-art passive methods.

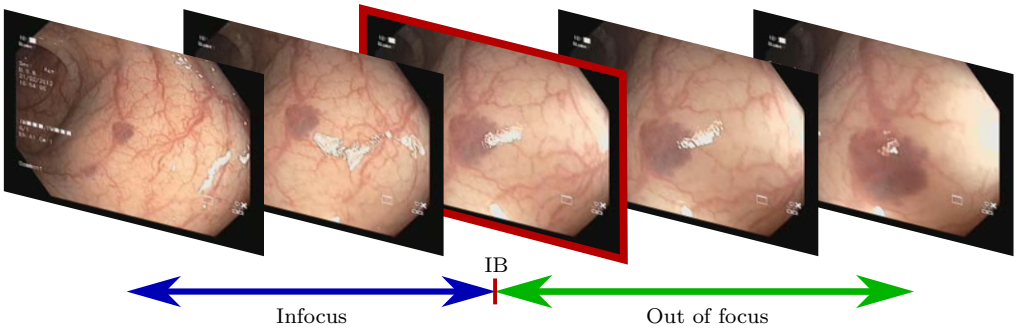


Figure 4: Colonoscopic video corresponding to an approximate approach motion of the colonoscope toward a neoplasia. The three leftmost images of the neoplasia extracted from the video stream are infocus while the two rightmost images are out of focus. The central image thus corresponds to the sharp/blur breakpoint, namely the *Infocus-Breakpoint*.

3. System Overview

Our system has two main steps, illustrated in figure 5: preoperative calibration, done just before colonoscopy, and intraoperative measurement, where actual neoplasia size measurements are performed.

Preoperative calibration. Calibration is done once, preoperatively. Geometric calibration is required to compute the internal camera parameters, including the lens distortion function. We use the pinhole camera model (Hartley and

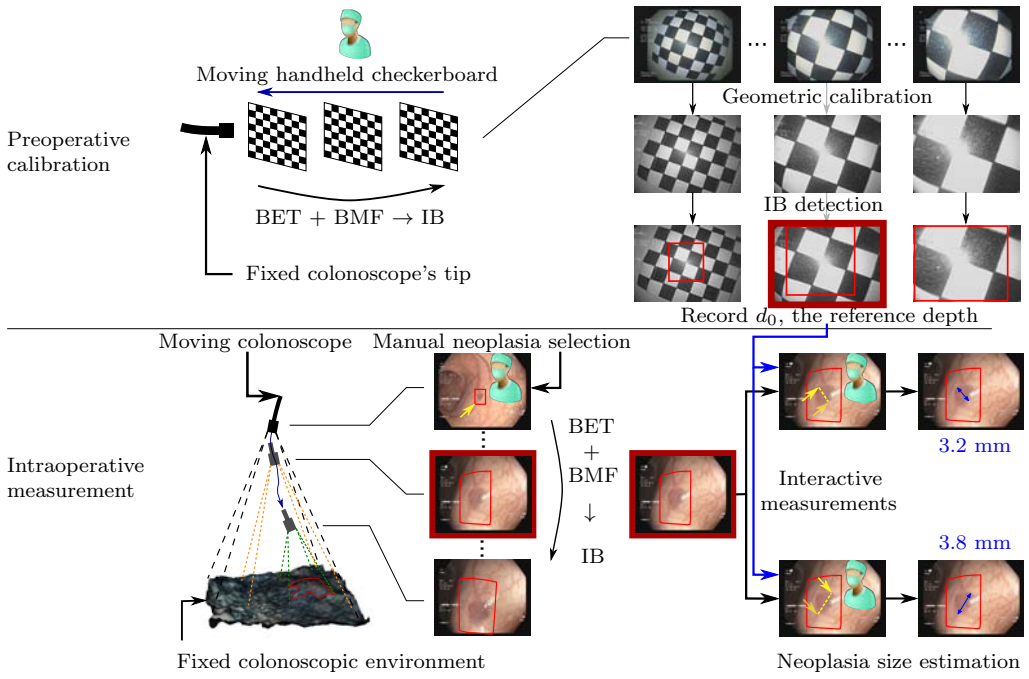


Figure 5: Proposed system overview. Top row: preoperative calibration protocol. Bottom row: intraoperative measurement protocol. BET is our *Blur-Estimating Tracking* module and BMF is our *Blur-Model Fitting*. They allow us to robustly detect the IB, the *Infocus-Breakpoint*, whose *reference depth* d_0 is calibrated preoperatively and used to resolve scale intraoperatively.

Zisserman, 2004). Calibration provides the matrix K of internal camera parameters and the sensor's pixel density α which relates pixels and mm. The important optical distortion of colonoscopes may directly bias the measurement of neoplasias (see section 6.3) if uncorrected with an increased bias away from the image center (Brauer-Burchardt et al., 2006). Both the geometric and IB calibration are done simultaneously using BET and BMF. A checkerboard is manually held approximately frontoparallel to the colonoscope and moved toward its distal end until it touches it. The video is recorded and geometric calibration is solved (Bouguet, 2008). For each frame of the video, the pattern is tracked and blur is estimated with BET. The BMF module is then applied to detect the IB and the corresponding reference depth d_0 is recorded. The overall preoperative calibration process generally takes less than a minute and a few minutes at most.

Intraoperative measurements. In the intraoperative course, the endoscopist first spots a neoplasia whose size should be measured. They put the neoplasia in full visibility and infocus, and manually select a polygonal area containing the neoplasia in the image by means of mouse clicking. They then move the colonoscope toward the neoplasia until it touches it. Our system automatically tracks the neoplasia with BET and estimates the relative defocus information of the tracked area simultaneously. The frames are recorded and form a video which is processed for scale recovery. The offline treatment of the recorded video takes less than a minute. Our system detects the IB by fitting our optical blur model with BMF. It then displays an image of the neoplasia where the endoscopist may interactively measure its length by clicking pairs of points and its area by positioning ellipses.

4. Image Formation Modeling and the Infocus-Breakpoint

The goal of modeling here is to define a cost function which will allow us to match the color of corresponding pixels in two different images in spite of motion and blur, and thus to later implement BET and BMF to robustly detect the IB.

4.1. Image Formation Modeling

Image formation modeling aims at defining the photometric part of the image formation process, including defocus.

Geometric optical modeling. Geometric optical modeling relies on the laws of Snell-Descartes which describe the formation of an image through a thin lens (Pentland, 1987). Let A_d be a 3D point located at a distance d from the camera. We define the infocus plane as orthogonal to the lens' axis at a distance $\hat{d}$. We shall show in section 4.2 that $\hat{d} = d_0$, in other words, that the infocus plane lies at the reference depth. The image of A_d is a point if it lies on the infocus plane at distance $\hat{d} = \frac{ef}{e-f}$, where e is the distance between the lens and the image sensor and f is the lens' focal length (see figure 6). For $d = \hat{d}$, the focal plane[2] coincides with the sensor. In any other case (for $d \neq \hat{d}$), A_d will be imaged as a circle of confusion whose radius $R(d)$ is given by:

[2]For a fixed object plane, the focal plane is defined as the infocus image plane; if the image sensor coincides with the focal plane, the image of the fixed object plane is infocus.

$$R(d) = r \left| \frac{e}{f} - \frac{e}{d} - 1 \right| \tag{1}$$

where r is the lens' radius (see figure 7). Note that $R(\hat{d}) = R\left(\frac{ef}{e-f}\right) = 0$.

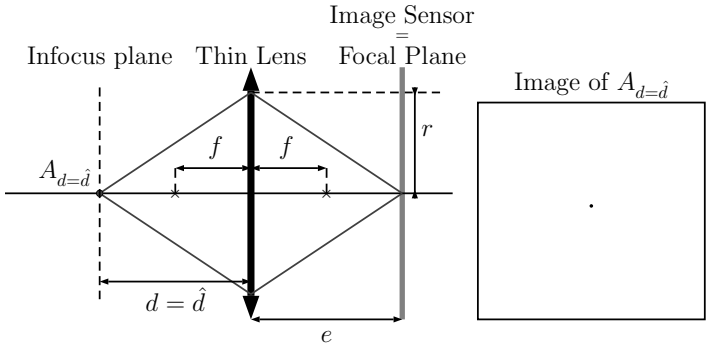


Figure 6: Image formation by a thin lens. The image of a point A_d is a point if $d = \hat{d}$, meaning that A_d lies on the infocus plane.

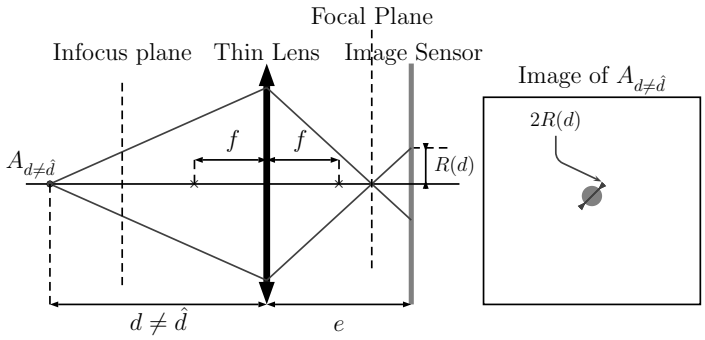


Figure 7: Image formation by a thin lens. The image of a point $A_{d\neq\hat{d}}$ which lies off the infocus plane is a spot of radius $R(d)$.

Modeling defocus. The Point Spread Function (PSF) is the impulse response of the optical system. In geometric optical modeling the predicted intensity is constant within the blur circle. This defines a PSF by the pillbox function whose radius is then related to the depth of the imaged point. Physical

optical modeling (modeling phenomena including light diffraction, lens properties, sensor integration) leads to more advanced PSF models. (Claxton and Staunton, 2008) showed that the generalized Gaussian is a suitable model because it encompasses the classical pillbox and 2D Gaussian models. It has also been experimentally observed that (Pentland, 1987), considering a polychromatic light, the PSF could be approximated by a 2D isotropic Gaussian $g_{\sigma(d)}$. Our approach relies on this standard assumption, holding for most cameras, and widely used in DfD and DfF (Pentland, 1987; Namboodiri and Chaudhuri, 2007). For a prime lens, the standard deviation $\sigma(d)$ of the Gaussian (the PSF's spread) is proportional to the distance between the 2D image point and the infocus object plane:

$$\sigma(d) \propto R(d) \tag{2}$$

The coefficient of proportionality directly depends on the camera's internal parameters and can be calibrated. Defining $\hat{I} = I_{\hat{d}}$ as the infocus image of some surface, in the absence of any geometric distortion, the image function I_d for a surface at depth d is then given by:

$$I_d = \hat{I} * g_{\sigma(d)} \tag{3}$$

For the typical prime lens optical colonoscopic systems, the optical blur model given by equations (1), (2) and (3) only depends on the local depth. Local optical blur measurement is however unstable because of its high sensitivity to image noise.

4.2. The Infocus-Breakpoint

Figure 8 is an instance of the optical blur model (1) and (2). It is worth to note that there exist a fixed point corresponding to the sharp/blur transition: the IB. The IB occurs at the reference depth d_0 which can be calibrated. Theoretically, d_0 is defined by:

$$d_0 = \underset{d}{\operatorname{argmin}}\, R(d) = \frac{ef}{e-f} \tag{4}$$

In order to robustly detect the IB within a colonoscopic video, we propose to use an approximate colonoscope motion toward a neoplasia. We here assume a constant speed motion of the colonoscope. This allows us to fit the continuous geometric optical blur model (1) to the discrete blur measurement made for each frame of the colonoscopic video. Equivalently, a backward motion could be applied.

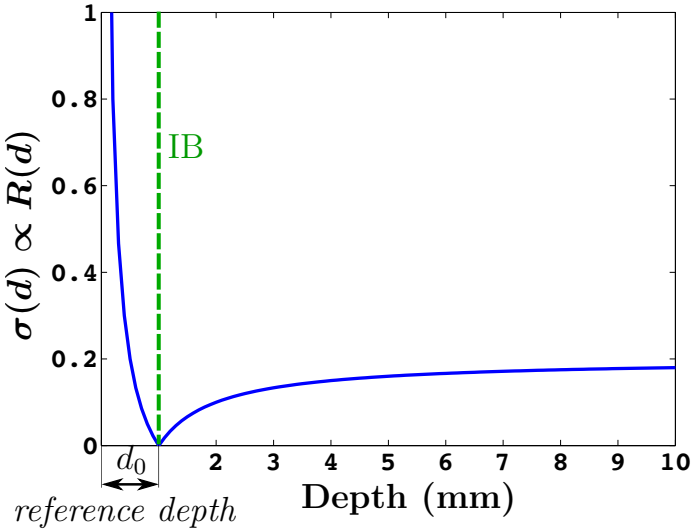


Figure 8: The optical blur model of equations (1) and (2) reveals the existence of a fixed point which corresponds to a sharp/blur transition we call the *Infocus-Breakpoint* (IB). For a prime lens optical system, it corresponds to a unique distance d_0 between the camera and the tissue. This is here called the reference depth and is calibrated preoperatively.

5. 3D Length Estimation and Image Registration Modeling

Consistently estimating the image blur related to some image area, here the neoplasia, requires one to track this area in a video stream. Practically, our experiments showed that a neoplasia has either a sharp boundary or is sufficient textured to be tracked in the video (see section 7.3). The aim of geometric modeling here is to define *(i)* the inference model used to measure a 3D length from a 2D image and *(ii)* the image deformation model used in tracking.

5.1. Measurements: from Image to 3D Lengths

Once the IB is detected within a video, the sharpest frame can be extracted and undistorted. Because we assume the neoplasia's boundaries to be planar and frontoparallel to the colonoscope's distal end, the reference depth allows us to infer the length d_{3D} of the neoplasia in mm from its length d_{2D} in pixels in the image using:

$$d_{3D} = \frac{d_0}{\alpha f} d_{2D} \tag{5}$$

with α the sensor's pixel density in pixels per mm. This is illustrated in figure 9. Preoperative geometric calibration gives the product αf (see section 3).

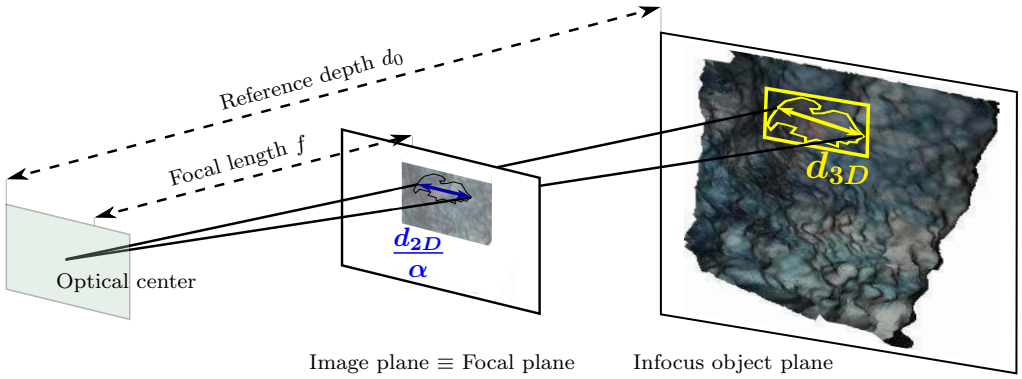


Figure 9: Intraoperative size measurement. d_{2D} and d_{3D} are respectively the size of the imaged neoplasia in pixels and the real size of the neoplasia in mm. α is the sensor's pixel density.

Any other image from the video could be used for making interactive measurements. Indeed, since the neoplasia has been tracked along the colonoscopic video, the manually selected points defining the distance to be measured can be transferred to the image corresponding to the IB.

We monitor the evolution of defocus in a local area corresponding to a neoplasia assumed to be planar. There may however be blur variations along the neoplasia's boundary particularly for pedonculated ones, for which the depth may vary along its boundary. For such cases, we comput a mean depth bounded by the depth of the neoplasia's boundaries. Our experiments showed that the equifocal assumption did not affect significantly the measurement quality (see section 7).

5.2. Image Registration Modeling

We have used two inter-frame deformation models. The first one is a simple affine transformation:

$$\mathcal{W}_{\text{AFF}}(\mathbf{q}) = \mathbf{A}\mathbf{q} + \mathbf{t} \tag{6}$$

where $\mathbf{A} \in \mathbb{R}^{2\times 2}$ is a linear map (embedding anisotropic scaling, rotation and shearing), $\mathbf{t} \in \mathbb{R}^{2\times 1}$ is a translation vector and $\mathbf{q} = (x, y)^T$ is the coordinate vector of an image point.

The second deformation model we used is a Thin-Plate Spline (TPS) warp (Bookstein, 1989). It is obtained by stacking two $\mathbb{R}^2 \to \mathbb{R}$ TPS funtions sharing their deformation centres and internal regularization weight. We use a feature-driven parameterization (Bartoli et al., 2010) defined as:

$$\mathcal{W}_{\text{TPS}}(\mathbf{q}) = \mathcal{M}\boldsymbol{\ell}_{\mathbf{q}} \text{ where } \mathcal{M}^\top = \mathbf{E}_\lambda \mathbf{P}' \tag{7}$$

This warp is parametrized by l deformation centres $\mathbf{c}'_j \in \mathbb{R}^{2\times 1}$ (with $j \in [1; l]$) in the target image. They are stacked in matrix $\mathbf{P}' \in \mathbb{R}^{l\times 2}$. The TPS kernel function for the squared distance is given by $\rho(d^2) = d^2 \log d^2$, and defines $\boldsymbol{\ell}_{\mathbf{q}}^\top = (\rho(d^2(\mathbf{q}, \mathbf{c}_1)) \cdots \rho(d^2(\mathbf{q}, \mathbf{c}_l))\ \mathbf{q}^\top\ 1)$. Finally, $\mathbf{E}_\lambda$ is a fixed design matrix built from the source deformation centres $\mathbf{c}_j \in \mathbb{R}^{2\times 1}$ (with $j \in [1; l]$) and the internal regularization weight $\lambda = 10^{-5}$.

The affine deformation model is global and stiff while the TPS deformation model is local and flexible. We expect the TPS model to be robust to regular colonoscopic artifacts and to model colon deformations. We experimentally compared the accuracy and robustness of these two deformation models as well as their influence on blur estimation when integrated to the BET module (see section 7.1).

6. Blur-Estimating Tracking, Blur-Model Fitting and IB Detection

6.1. Tracking in Colonoscopic Videos

There exists a large body of work on image tracking through registration. However, very few papers have investigated the simultaneous estimation of both registration and optical blur. The approach of (Deschênes et al., 2004) which extends (Myles and da Vitoria Lobo, 1998) estimates simultaneously a spatial shift (stereo disparities, 2D motion, and/or zooming disparities) and an optical blur difference between two images. However, this window-based approach is sensitive to illumination variations, noise and textureless areas. Feature detection and description methods such as SIFT (Lowe, 2004) have been shown to perform poorly in colonoscopy (Alcantarilla et al., 2013) (see figure 1).

6.2. Proposed Tracking Strategy

We propose to use a direct registration approach, using the color of all pixels in the ROI. Image registration processes two images at a time. One is the source, the other the target. In registering a whole video, a possible strategy is to choose a fixed frame for the source. This strategy performs poorly here, due to the important changes of the image contents occuring during the colonoscope's forward motion. We thus register consecutive frames I^k and I^{k+1}, where k is the frame index. Relative optical blur difference $\delta^k = s^k - t^k$, where s^k and t^k corresponds respectively to the blur applied to the source and target image for their registration, is computed simultaneously for each pair of images as described in the next section. The proposed BET module is illustrated in figure 10 and summarized in algorithm 1.

Blur-Estimating Tracking (BET)

Input:

- Video $(I^1, \ldots, I^N)$
- Initial ROI (ROI^1)

Algorithm:

For $k = 1$ to $N - 1$

$(\mathcal{W}^k, \delta^k) \leftarrow$ Register (I^k, I^{k+1}, ROI^k)

$ROI^{k+1} \leftarrow \mathcal{W}^k(ROI^k)$

End For

Output:

- Relative blur $(\delta^1, \ldots, \delta^{N-1})$
- (optional) Warps $(\mathcal{W}^1, \ldots, \mathcal{W}^{N-1})$
- (optional) ROIs $(ROI^1, \ldots, ROI^N)$

Algorithm 1: Our BET module simultaneously estimates the geometric deformation and the relative defocus information between successive images in a video stream.

This strategy implies that relative blur measurements be combined between consecutive frames of the colonoscopic video to obtain an absolute blur measurement up to a global additive constant; this is discussed in section 6.4.

6.3. Image Pair Registration

We propose to incorporate the defocus model, defined in section 4.1, in a direct registration method. From equation (3), the following relationships

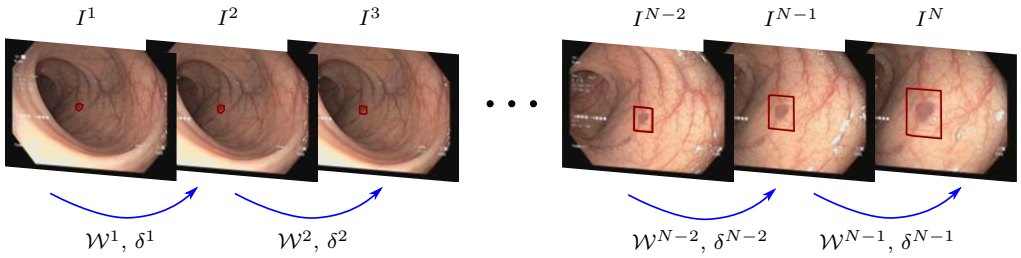


Figure 10: Illustration of the tracking strategy used in BET for a real colonoscopic video. Our method relies on the simutaneous estimation of a geometric deformation $\mathcal{W}^k$ and a relative optical blur δ^k for each pair of consecutive images $\left(I^k,\ I^{k+1}\right)$ for $k \in [1; N-1]$.

hold:

$$\begin{cases} I^k = \hat{I} * g_{\sigma(d^k)} \\ I^{k+1} = \hat{I} * g_{\sigma(d^{k+1})} \end{cases} \tag{8}$$

where d^k and d^{k+1} are the depth of the ROI respectively in images k and $k+1$. It is worth to note that equation (8) does not model geometric distortions. Those will be discussed when formulating the final registration problem shortly below. The infocus image $\hat{I}$ is unknown and has to be eliminated from the equations. The idea is to use the sharpest image as a 'predictor' of the other one. However, which one of the two images I^k and I^{k+1} is the sharpest is a priori unknown (see figure 8). This is actually given by the sign of $\Delta^k = \sigma(d^k) - \sigma(d^{k+1})$ (where Δ^k is an absolute blur difference); if $\Delta^k > 0$, I^k is the sharpest. We thus rewrite equation (8) as:

$$\begin{cases} I^k * g_{\sigma(d^{k+1})} &= \hat{I} * g_{\sigma(d^k)} * g_{\sigma(d^{k+1})} \\ I^{k+1} * g_{\sigma(d^k)} &= \hat{I} * g_{\sigma(d^{k+1})} * g_{\sigma(d^k)} \end{cases} \tag{9}$$

Because convolution with Gaussian functions is commutative, we obtain the following relationship:

$$I^k * g_{\sigma(d^{k+1})} = I^{k+1} * g_{\sigma(d^k)} \tag{10}$$

Note that this equation has a degree of freedom meaning that realative blur can only be computed up to a constant g_κ which cannot be neglected, especially in the presence of noise or motion blur artifacts. For $\kappa > 0$, it may be

rewritten as:

$$I^k * \underbrace{g_{\sigma(d^{k+1})} * g_\kappa}_{g_{t^k}} = I^{k+1} * \underbrace{g_{\sigma(d^k)} * g_\kappa}_{g_{s^k}} \tag{11}$$

Here, g_{s^k} and g_{t^k} correspond to relative blur measurements respectively in the source and target frames I^k and I^{k+1}. Hence, image pair registration as illustrated in figure 11 involves resolving the following non-linear least-squares problem:

$$\min_{(\zeta^k, s^k, t^k, \mathcal{W}^k)} \sum_{\mathbf{q} \in ROI} \rho^k(\mathbf{q})[I^k * g_{t^k}(\mathbf{q}) - \zeta^k(I^{k+1} * g_{s^k}(\mathcal{W}^k(\mathbf{q})))]^2 + \lambda \min(s^k, t^k)^2 \tag{12}$$

The penalty $\min(s^k, t^k)^2$ prevents overblurring: it ensures $\kappa \approx 0$ and thus that $t^k \approx \sigma(d^{k+1})$ and $s^k \approx \sigma(d^k)$. The mask ρ^k allows us to discard saturated areas and ζ^k is a gain that adjusts global intensity variation due to the colonoscope's light delivery channel being partially hidden behind folds of the colonic wall and the light intensity altered by auto-adjustment of the light source module. We solve problem (12) with the Levenberg-Marquardt method (Björk, 1996) with as initialization: $\zeta^k = 1$, $s^k = t^k = \varepsilon$ and $\mathcal{W}^k$ the identity warp.

Geometric distortion is not introduced in our BET module. This is because undistorting the images would break the constancy of the blur kernel. However the images used to infer size measurements are undistorted (see section 5.1). We noticed no significant impact on tracking and blur estimation even with neoplasias evaluated in real colonoscopic conditions imaged off the optcal axis.

6.4. Blur-Model Fitting and IB Detection

Fitting by integration. In order to extract the IB, it is necessary to retrieve the absolute blur measurement along the video. A natural approach is to integrate the value of δ^k and extract the IB thanks to sign changes. However, this method is particularly sensitive to noise and erroneous blur estimation. It does not allow one to accurately extract the IB, introducing ambiguities as illustrated in figure 12. For instance, when the colonoscope is close to the neoplasia or when the part of the colon explored by the endoscopist is difficult to access, motion blur artifacts frequently arise in the video and cause the integration method to fail.

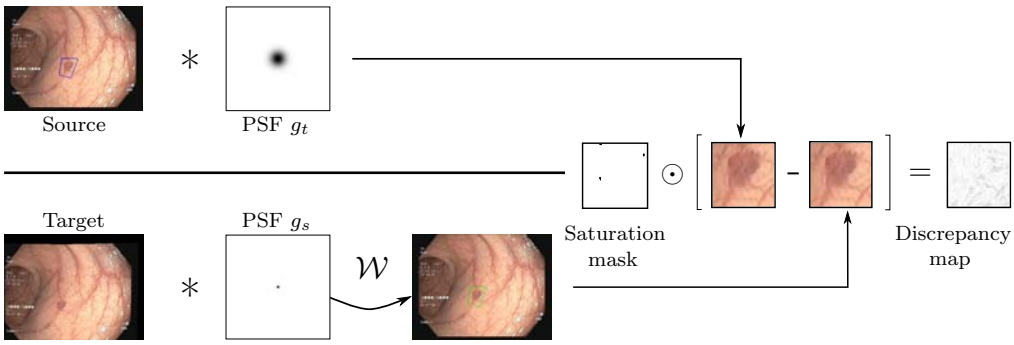


Figure 11: Illustration of our *Blur-Estimating Tracking* (BET) module for a pair of colonoscopic images. In this exemple, the source image is sharper than the target image. Our goal is to minimize the norm of the discrepancy map obtained by subtracting the blurred source and the blurred and warped target for a ROI containing the neoplasia to be measured.

Register

Input:

- Image pair (I^k, I^{k+1})
- ROI in I^k (ROI^k)

Algorithm:

Compute $(\zeta^k, s^k, t^k, \mathcal{W}^k)$ by solving equation (12)
Set $\delta^k \leftarrow s^k - t^k$

Output:

- Warp $\mathcal{W}^k$
- Relative blur δ^k

Algorithm 2: Image pair registration algorithm. We simultaneously compute warp parameters and relative between a source and a target image.

Proposed Blur-Model Fitting method. At the beginning of the colonoscopic video, when $d > d_0$, the target frame is sharpest than the source frame and so $\Delta^k < 0$ and s^k represents the absolute blur difference (while t^k should theoretically be equal to 0). Similarly, when $d \leq d_0$, $\Delta^k > 0$; the source frame is sharpest than the target image and so t^k is the absolute blur difference (while

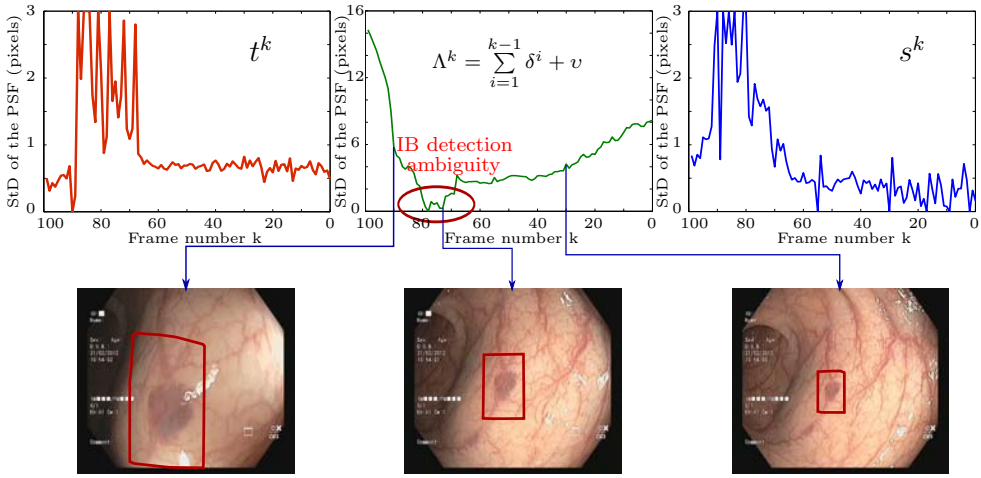


Figure 12: Illustration of the integration method to solve *Blur-Model Fitting* (BMF) and the detection of the *Infocus-Breakpoint* (IB). This figure illustrates the IB detection ambiguity due to errors in relative blur estimation. These may be caused by motion blur which occurs close to the neoplasia typically around the IB. The integration constant υ, which is unknown, has been defined such that the absolute defocus at the IB equals 0.

s^k should theoretically be equal to 0). We obtain the following relationship:

$$\Delta^k \approx \begin{cases} -s^k & \text{for} \quad d > d_0 \\ t^k & \text{for} \quad d \leq d_0 \end{cases} \tag{13}$$

Hence, our problem is to define a reliable approximation of the IB to integrate absolute depth measurements which will be refined by fitting the optical blur model (1). We use the fitting by integration method defined at the previous section and choose the index $k_0 \in [1; N-1]$ corresponding to the minimum of $\Lambda^k = \sum_{i=1}^{k-1} \delta^i + \upsilon$ as a first approximation of the IB. The integration constant υ corresponding to the absolute defocus of the first frame of the video stream is unknown. As the value of the standard deviation of the Gaussian defocus theoretically vanishes at the IB, we define $\upsilon = \left| \min_{j=2,\dots N} \sum_{l=1}^{j-1} \delta^l \right|$. Moreover, most of the artifacts arise when the colonoscope is close to the neoplasia. Owing to the forward motion assumption,

we always chose the largest frame index as IB in case of an ambiguity (in the case of a backward motion, we would choose the smallest frame index as IB approximation). The absolute blur measurement is then computed by integrating the PSF standard deviation derivatives (see figure 13):

$$\Psi^k \stackrel{\text{def}}{=} \sigma(d^k) \approx \begin{cases} -\sum_{i=0}^{k} (s^i) + \upsilon & \text{for } k < IB \ (d > d_0) \\ \left(\Psi^{IB-1}\right)^2 + \sum_{i=k_{IB}}^{N} (t^i) + \upsilon & \text{for } k \geq IB \ (d \leq d_0) \end{cases} \tag{14}$$

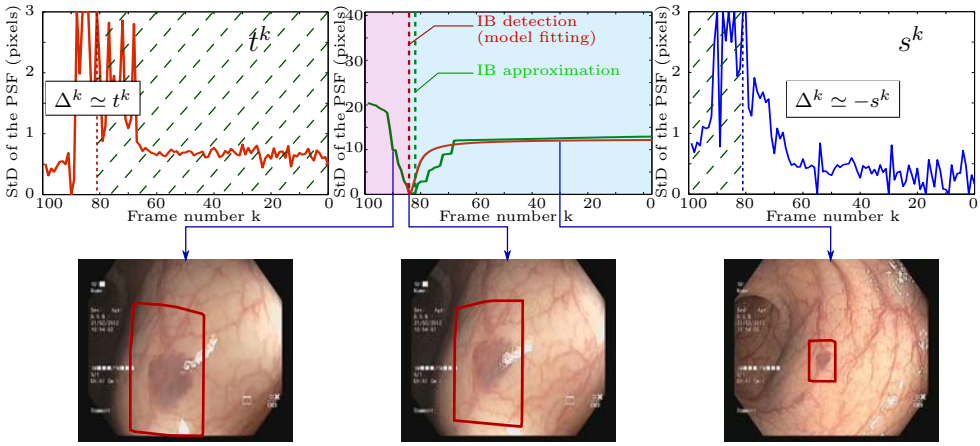


Figure 13: *Blur-Model Fitting* (BMF) and *Infocus-Breakpoint* (IB) detection. The top left and right graphs show relative blur estimation respectively for the source and target frames along a colonoscopic video. The central graph shows the absolute blur estimation obtained by integrating relative blur measurements in the source and target frame relatively to an initial IB approximation. The optical blur model, here in red is then fitted to the absolute blur estimate to accurately extract the IB. The botton images show the colonoscopic video and the tracking results obtained by our TPS-based Blur-Estimating Tracking (TPS-BET) method.

The parametric model (1) can then be fitted to the absolute blur measurements assuming a constant speed β (in pixels per frame):

$$\min_{r,e,f,\beta} \sum_{k=1}^{k_0} \left[\Psi^k - \left(r \left| \frac{e}{f} - 1 - \frac{e}{\beta k} \right| \right) \right]^2 \tag{15}$$

The non-linear minimisation is solved using the Levenberg-Marquardt method. In order to compute the derivatives of equation (15), we replaced the absolute value of the parametric model (2) by the pseudo-Huber norm (Huber, 1964): $L(\left|\frac{e}{f} - 1 - \frac{e}{\beta k}\right|) = \sqrt{1 + \left(\frac{e}{f} - 1 - \frac{e}{\beta k}\right)^2} - 1$. The parameters are initialized from geometric calibration values (see section 5). It is worth to note that the theoretical model can only be partially fit to the data belonging to the range $[1 \dots k_0]$. Indeed, when the colonoscope's tip is very close to the neoplasia, the optical blur measurement saturates and the thin lens assumption breaks down. Thus, only the depth greater or equal to the predicted IB were considered for non-linear optimization.

Because the computation of the absolute blur is relative to an initial guess k_0 for the IB, we apply algorithm 3 to robustly refine it.

Blur-Model Fitting (BMF)

Input:

- Relative blur $(\delta^1, \dots, \delta^{N-1})$

Algorithm:

residual, IB $\leftarrow 0$

Compute predicted IB: $k_0 \leftarrow \underset{i}{\text{argmin}} \sum_{i=1}^{N-1} \delta^i$

For $k_{IB} = k_0 - 10$ to $k_0 + 10$

  Compute absolute blur relative to the IB index k_{IB} using equation (14)

  Obtain current_IB and current_residual by solving equation (15)

  If current_residual $<$ residual

    IB, residual $\leftarrow$ current_IB, current_Residual

  End If

End For

Output:

- IB
- (optional) BMF residual error

Algorithm 3: The integration of our BET relative blur estimates allows us to compute absolute blur along the video stream and a first approximation of the *Infocus-Breakpoint* (IB). The optical blur model (14) can then be fit to the abslute blur estimates which allows us to accurately extract the IB.

The BMF method allows us to reliably refine the first IB approximation and handles an arbitrary colonoscope motion. Because the IB corresponds to a precalibrated depth, the size of the neoplasia can finally be inferred in the undistorted IB frame $I^{IB} \approx \hat{I}$ (see section 5).

7. Experimental Validation

The proposed algorithm was evaluated on three types of datasets:

- **Simulation:** a simulated dataset was used to evaluate the robustness of the proposed algorithm to noise, blur and motion amplitude.
- **Phantom:** a phantom model was designed in order to evaluate our method in conditions similar to real colonoscopy but with ground truth. The phantom videos were made with a GIF type Q160 Olympus® colonoscope.
- **In-vivo:** a real colonoscopy dataset was used to compare the performance of the proposed algorithm with measurement guesses made by endoscopists. The videos were made with two high definition GIF Type N180 and XP190N Olympus® colonoscopes.

We implemented two versions of BET: AFF-BET and TPS-BET. The former uses a rigid affine warp $\mathcal{W}_{\text{AFF}}$; the latter uses a TPS warp $\mathcal{W}_{\text{TPS}}$ (see section 5). Four TPS centers were used to carry out the experiments. They were placed on a regular grid inside the ROI. The behaviour of TPS-BET handles typical deformations arising in regular colonoscopic conditions.

7.1. Simulated Data

Setup. We created a simulated dataset which reproduces the slight deformations of the colonic cavity observed in real colonoscopic images. To this end, we computed a random warp which was applied to 10 reference images extracted from real colonoscopy examination (see figure 15). The image generation process was based on the following steps (see figure 14). We first defined a virtual sphere whose center lies on the optical axis of a simulated colonoscope's optical system at a depth of 4 cm. A subset of the sphere's points was projected on the simulated endoscopic image to define 2D target deformation centers. The 3D sphere was then randomly translated and

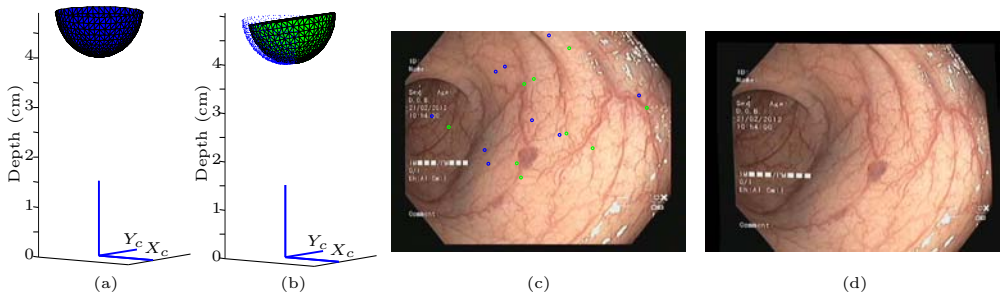


Figure 14: Illustration of the synthetic image generation process. (a) and (b): A sphere centered on the optical axis of a simulated colonoscope's optical system is translated and rotated in 3D. (c): A subset of the 3D points of the 'reference' and 'target' sphere is projected in the image plane and a warp is computed using the projected points as deformation centers. (d): Finally a new image is generated by applying the randomly computed warp to an image of a real colonoscopy.

rotated. The subset of 3D points was reprojected in the image plane to generate the corresponding 2D source deformation centers. A warp was then computed and a new image was generated by warping the reference images.

Three parameters, noise, blur and camera speed, were varied on a predefined range as illustrated in figure 16. The camera speed is here modeled by the translation magnitude of the 3D sphere. The random sphere rotation was chosen small to generate slight deformations of the reference image. For each value of the range, the dataset was made up of 100 random deformations of a source image.

The default setup for each of the experimental parameters was the following:

Motion amplitude: the influence of the camera speed was evaluated according to the relative translational displacement of the sphere on the range $[0; 6]$ cm. Because the sphere center was placed at a close depth of the simulated colonoscope, this corresponds practically to a relative displacement of the deformation centers which lied in $[0; 70]$ pixels. The sphere's rotation belonged to the range $([\frac{\pi}{20}; \frac{\pi}{15}]$. The Gaussian blur standard deviation was of 1 pixel and the noise was of 1%.

Blur: the Gaussian blur standard deviation was evaluated on the range $[0; 10]$ pixels. The sphere motion's amplitude was of 1 cm (which corre-

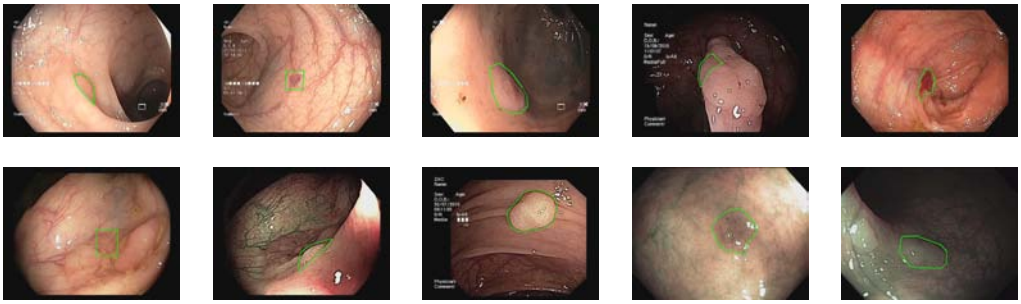


Figure 15: Various colonic neoplasias used to generate the synthetic dataset. For each of the experimental parameters, 100 random transformations were applied to each of these images to generate a set of 1000 images.

sponds to a displacement of the deformation centers of approximately 10 pixels) and the noise was of 1%.

Noise: the noise was evaluated on the range $[0; 10\%]$. The motion's amplitude was of 1 cm and the Gaussian blur standard deviation of 1 pixel.

Statistical analysis. The results are visible in figure 17. TPS-BET is significantly more robust to geometric deformations. AFF-BET is robust to small displacements when the sphere translation is less than 2 cm. TPS-BET handles more important displacements (translation up to 3.5 cm) which is a significant benefit considering the difficult manipulation of a real colonoscope close to a neoplasia. Beyond these values, the tracking drifts and makes blur estimates unstable (as illustrated by the standard deviation of the errors). Such blur estimation errors could lead to ambiguities which prevent our method from accurately determining the IB.

TPS-BET and AFF-BET have shown similar behaviour regarding the impact of noise and blur. The registration error is less than 0.5 pixels even in the presence of an important amount of blur as illustrated in figure 16. The blur estimation error is less than 0.5 pixels when the real blur is less than 6 pixels. Beyond this value, the blur estimation error linearly increases for both AFF-BET and TPS-BET so as the standard deviation of the errors. Practically, such an amount of blur was not observed during our experiments on real colonoscopy.

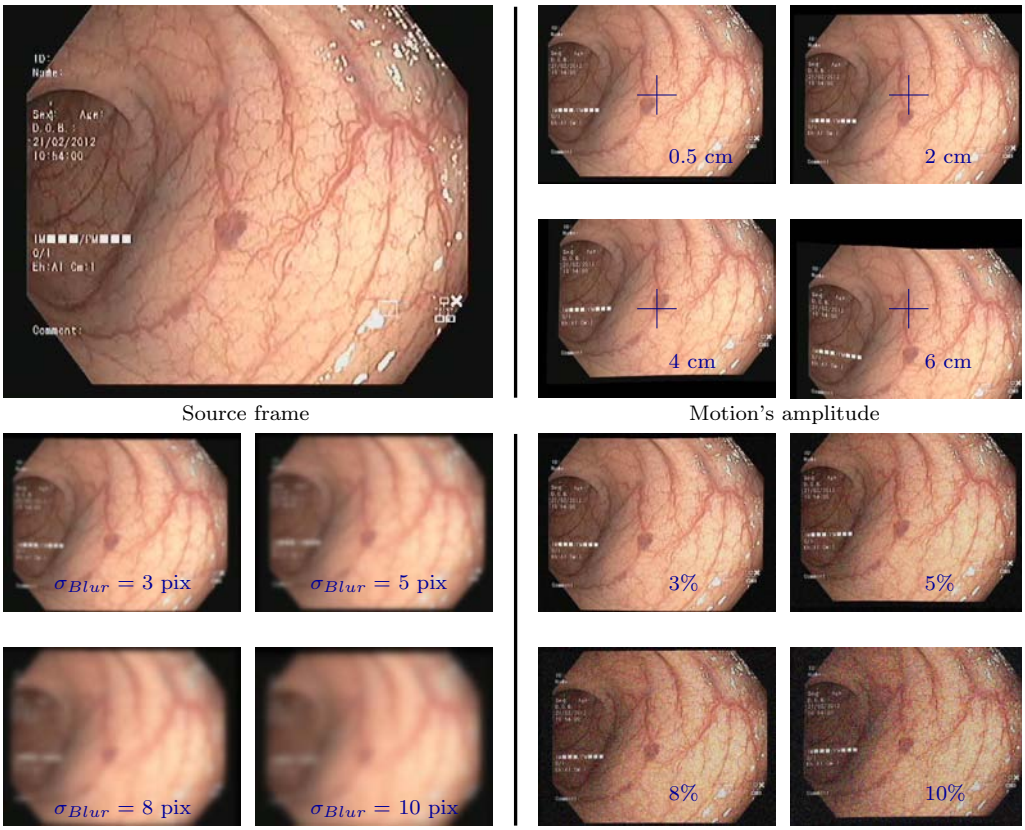


Figure 16: Simulated datasets. The source image (top left) has been transformed to evaluate the influence of the motion's amplitude (top right), the blur (bottom left) and the noise (bottom right) on the robustness of our method.

The noise has a slight influence on the registration error when it is lower than 6%. For more important values, the registration error increases rapidly even if TPS-BET is more robust and stable in the presence of such artifacts. Blur estimation is more affected by the presence of noise. The error is less than 1 pixel when noise is less than 3%. Beyond this value the blur estimation error linearly increases which may lead to IB detection ambiguities.

The synthetic evaluation has shown that TPS-BET is significantly more robust to important motion speed of the colonoscope. Moreover, it has showed that blur estimation is slightly more robust to noise artifacts. Finally, while blur estimation is biased when image blur is important, it will be

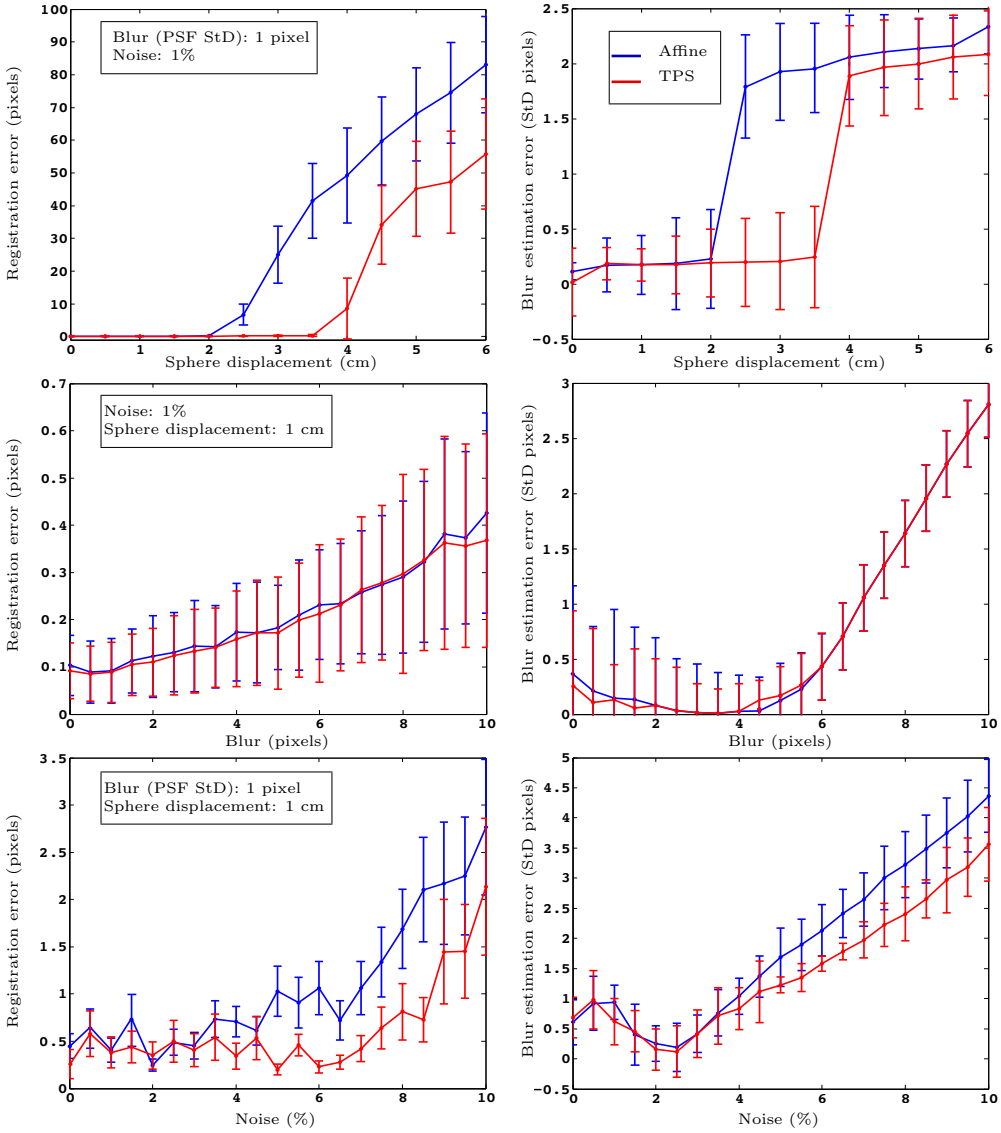


Figure 17: The left column shows the affine (blue) and TPS (red) *Blur-Estimating Tracking* registration errors according to camera speed, blur and noise. The right column shows the blur estimation error. Each row corresponds to a specific experimental parameter. Error bars represent the standard deviation of errors.

shown with phantom and real colonoscopy experiments that blur estimation

errors can be handled well by our BMF module.

7.2. Phantom Data

Our phantom model (figure 18) is a tube within which a pig's colon was inserted. Small objects from 1.6 to 15 mm were placed inside the colon cavity and wrapped up in colonic tissue. The pig's colon is frequently used for endoscopists training models due to its anatomical similarity to the human colon. However, its internal appearance is significantly different, with less texture and landmarks. Blood vessels cannot be observed as they are not irrigated.

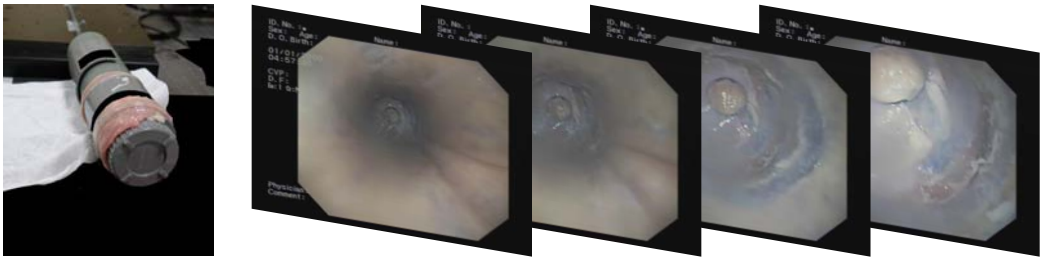


Figure 18: Left: external view of our phantom model. It is composed of a tube within which a pig's colon was inserted. Right: internal views of our phantom model. Several marbles of known size were wrapped up in colonic tissue and placed inside the phantom model.

Motivation. It is practically impossible to obtain the ground truth of a neoplasia's size in-vivo. Excision may cause neoplasias to shrink. Our phantom model allows us to reproduce colonoscopic conditions while having the true size of measured objects. Thus, artifacts such as motion blur or light saturation may arise in phantom videos. The colonoscope's motion has been controlled both manually (8 videos) and robotically (7 videos).

Figure 19 shows the root mean square measurement error for the AFF-BET and TPS-BET modules. Figure 20 illustrates the benefits of TPS-BET over AFF-BET in tracking, particularly for videos where the colonoscope's motion was manually controlled.

Results. The results show a significant improvement of measurement accuracy and a more reliable behaviour particularly for the videos made by manually controlling the colonoscope's motion. For this dataset, TPS-BET has a mean error of 6.9% ($\approx$ 0.8 mm) while AFF-BET has an error of 26%

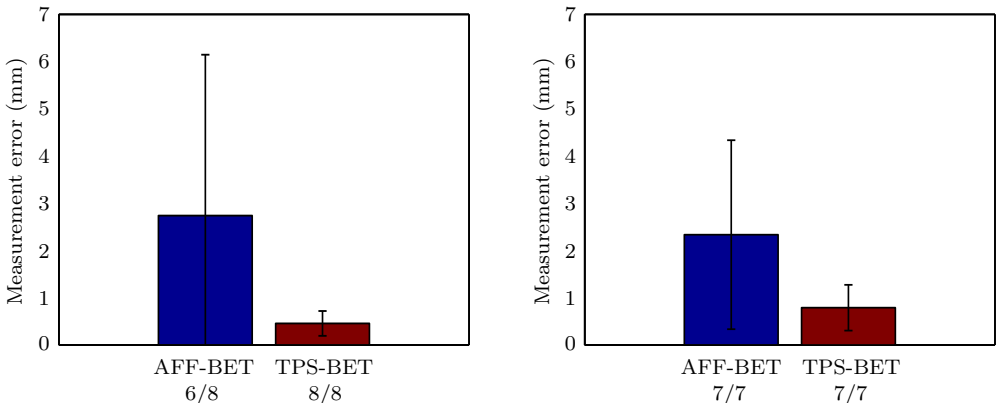


Figure 19: Measurement error analysis on phantom dataset. Left: robotically controlled motion of the colonoscope. Right: handheld colonoscope. The last row indicates for each method and each experimental setup, the number of videos which were successfully processed.

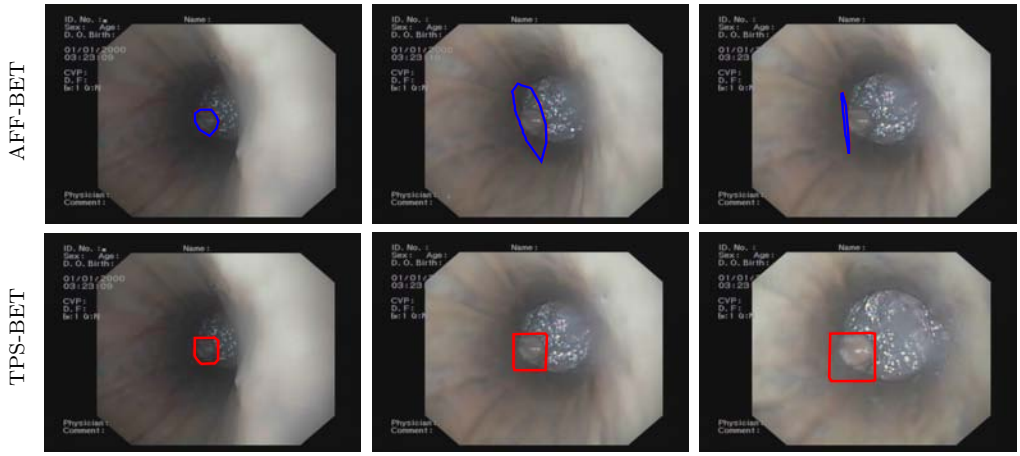


Figure 20: Exemple of Affine *Blur-Estimating Tracking* (AFF-BET) drift. Top: the manual displacement of the colonoscope generates motion blur which could not be handled by this method. Bottom: the TPS *Blur-Estimating Tracking* (TPS-BET) was successful.

(≈ 2.8 mm). This error is notably due to inaccurate tracking which implies IB extraction ambiguities. Moreover, two videos of this dataset could not be processed by AFF-BET because of motion blur while TPS-BET was able to

process all the videos.

For the subset of videos made by robotically controlling the colonoscope, experiments have shown that both AFF-BET and TPS-BET perform well. However, the results obtained by AFF-BET, with a mean measurement error of 24% ($\approx$ 2.4 mm), were degraded by artifacts such as moisture on the colonoscope's lens. For this second subset of videos, TPS-BET has a mean measurement error of 4.9% ($\approx$ 0.5 mm). TPS-BET handles motion blur and artifacts much more efficiently than AFF-BET.

7.3. In-vivo Data

Images in the operating room. A set of 5 colonoscopic videos was made by two experienced endoscopists during colonoscopic interventions. The videos were made following the protocol described in section 3 and illustrated in figure 21. In order to avoid strong motion blur artifacts and to make sure that the neoplasia is always visible within the colonoscopic videos, the endoscopists were asked to force as much as possible a linear motion of the colonoscope. The two endoscopists achieved our clinical protocol without difficulty. The results obtained with AFF-BET and TPS-BET were compared with the visual estimation of the endoscopists (using a surgical tool as reference) which was considered as the gold standard measurement. Figure 22 shows the root mean square measurement error for the two BET methods. In-vivo videos and TPS-BET measurements are presented in figure 23.

Results. Experiments have shown similar results for both of the evaluated algorithms; with a mean error of 4.5 % ($\approx$ 0.2 mm) compared to the visual estimation for TPS-BET and 6.5 % ($\approx$ 0.3 mm) for AFF-BET. However, similarly to the evaluation on our phantom model, one video could not be processed by AFF-BET. Because of motion blur artifacts, the affine tracking and blur measurement was inaccurate. These ambiguities did not allow us to robustly estimate the IB. TPS-BET was more robust to such artifacts.

8. Discussion

Experimental results, and more particularly results obtained on phantom videos, validate the applicability of our IB detection technique in colonoscopy. The proposed method relies on a simple medical protocol which has proved to be well adapted to real conditions of colonoscopy (both in terms of method efficiency and convenience of the medical protocol).

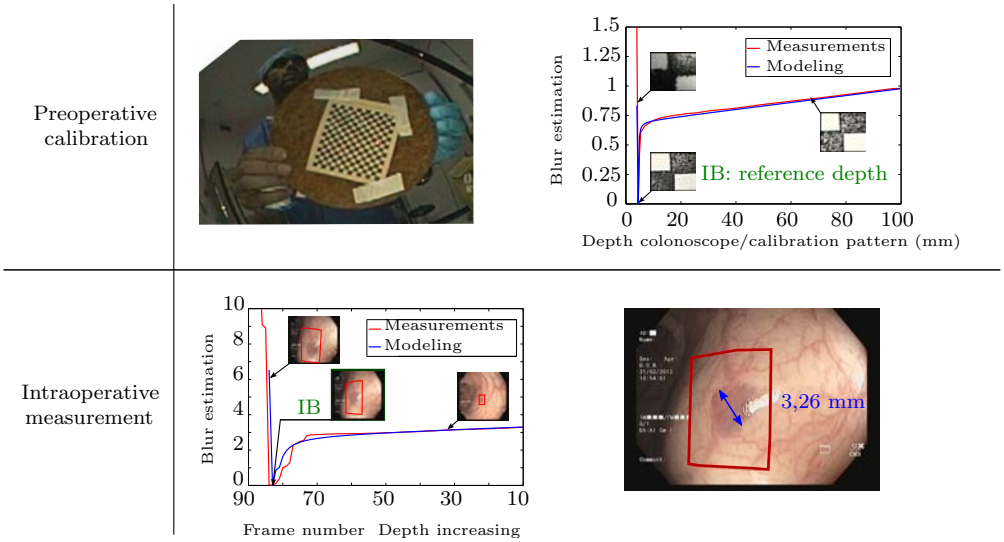


Figure 21: Top row: preoperative calibration for the GIF type N180 colonoscope. Bottom row: intraoperative measurement of a neoplasia (see figure 5 for our detailed protocol).

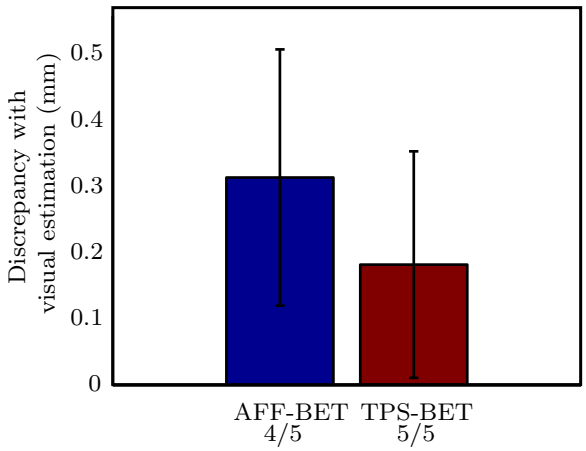


Figure 22: Measurement error analysis. The measurements are compared with the endoscopist' visual estimates.

The TPS-BET method we proposed shows promising results which outperform the AFF-BET method in robustness and accuracy. Indeed, minor

artifacts and more particularly noise (due to optic degradation) and motion blur were well handled by TPS-BET while AFF-BET could not accurately track neoplasias and thus measure a coherent blur along the colonoscopic video. Our method was mainly designed to measure small neoplasias (less than 1.5 cm) since the size of these lesions is an essential diagnostic criterion. The measurement error accuracy is less than a millimeter which fits the endoscopist's needs. It represents a percentage of error lower than 7% (see section 7.2). This significantly outperforms the visual estimation method curently used by endoscopists for which (Chaptini et al., 2014) showed that erroneous estimations greater than 20% arise in 52% of the 2812 studied cases.

9. Conclusion

Colonoscopy is the reference medical examination for the diagnosis and treatment of colonic diseases. It allows the endoscopists to remove neoplasias which could turn into malignant tumors. Besides the shape and appearance of these lesions, the size is an essential criterion for diagnostics and surveillance intervals time. While no practical solutions allows the endoscopists to reliably estimate the size of neoplasias, we here proposed a passive size measurement system, based on the *Infocus-Breakpoint* (IB) detection method, which can handle regular colonoscopic conditions.

The IB is the sharp/blur breakpoint which arises while the colonoscope moves toward the colonic tissue. Because most of the current colonoscopes host a prime lens optics, the IB corresponds to a fixed *reference depth* which can be calibrated. We have proposed a TPS *Blur-Estimating Technique* (BET) which simultaneously tracks a neoplasia along a video and estimates the optical defocus. Combined with an optical blur model, this method allows one to accurately extract the IB and infer measurements of a neoplasia, assuming the lesion is planar and frontoparallel to the colonoscope's distal end.

Pilot trials are needed in order to validate our system with regards to its influence on surveillance intervals and inter-operator variability. Technically, our work could be extended in two ways. It could be combined with an automatic neoplasia detection method. A fine segmentation could improve the simultaneous tracking and blur estimation step although this would restrict the extend of applicability (our system allows an endoscopist to measure any area of the colonic environment). Our work could also be combined with

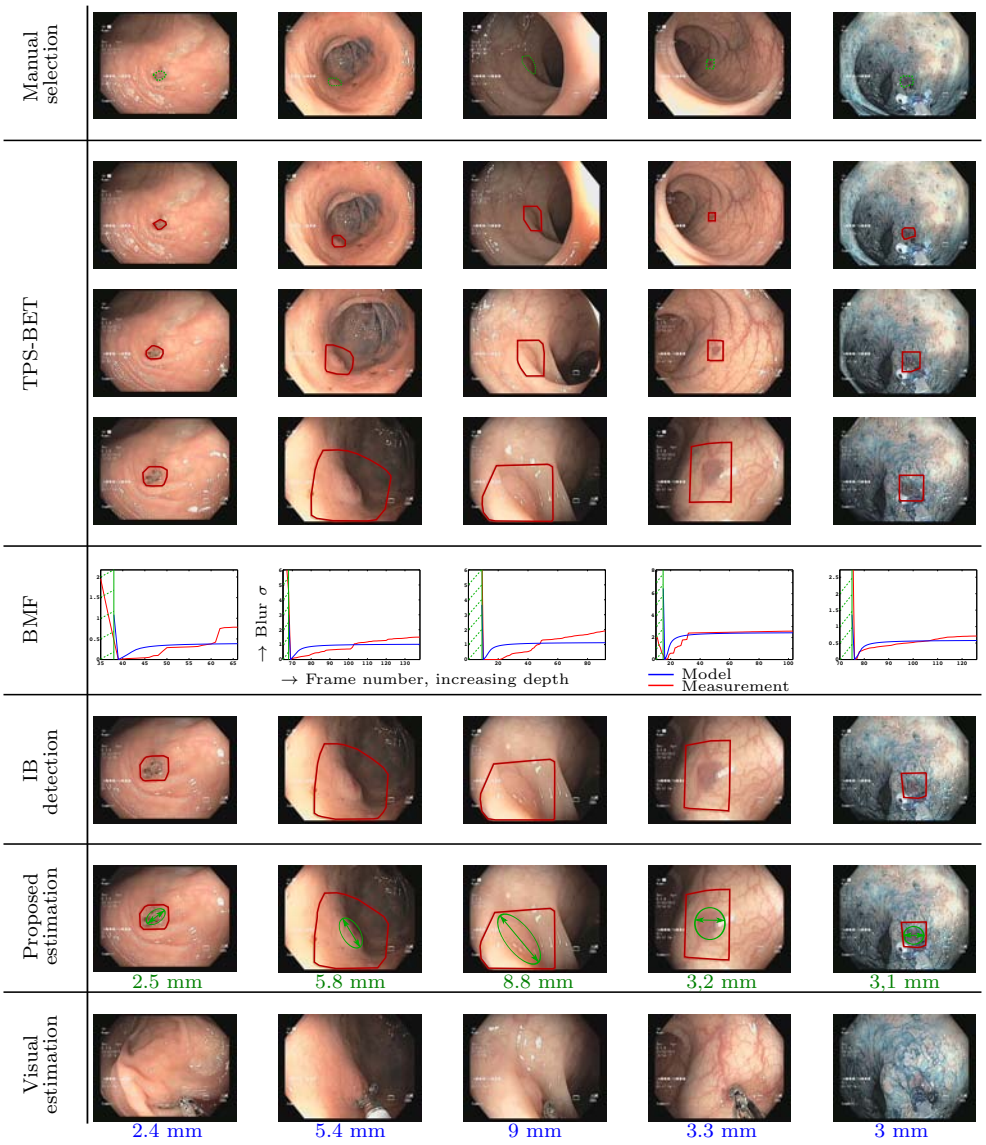


Figure 23: In-vivo dataset evaluation. Five colonoscopic videos were evaluated (in columns). Our system was compared with the visual estimation made by endoscopists (last row). Each step of the method is detailled up to the neoplasia size estimation. The last colonoscopic video (last column) used indigo carmine as contrast dye.

SfM (Alcantarilla et al., 2013; Chen et al., 2010; Koppel et al., 2007) to obtain a scaled 3D reconstruction of the colonic environment. Such a combination could lead to an accurate 3D reconstruction of suspicious depressed neoplasias which spread within the colonic wall.

Acknowledgment

This work has been supported by the French National Research Agency (ANR) through the TecSan program. The authors thank Yansys Medical®, as well as Professor Thierry Ponchon from the gastroenterology department of Edouard Herriot Hospital, Doctor Vincent Lepilliez and Doctor Sylvain Beorchia, private gastroenterologists.

References

Alcantarilla, P., Bartoli, A., Chadebecq, F., Tilmant, C., Lepilliez, V., 2013. Enhanced imaging colonoscopy facilitates dense motion-based 3D reconstruction, IEEE Engineering in Medicine and Biology Conference.

Bartoli, A., Perriollat, M., Chambon, S., 2010. Generalized thin-plate spline warps. International Journal of Computer Vision 88, 85–110.

Björk, H., 1996. Numerical Methods for Least Squares Problems. SIAM.

Bookstein, F.L., 1989. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 567–585.

Bouguet, J.Y., 2008. Camera calibration toolbox for matlab URL: `http://www.vision.caltech.edu/bouguetj/calib_doc/.`

Brauer-Burchardt, C., Heinze, M., Munkelt, C., Khmstedt, P., Notni, G., 2006. Distance dependent lens distortion variation in 3D measuring systems using fringe projection, British Machine Vision Conference.

Chadebecq, F., Tilmant, C., Bartoli, A., 2012. Measuring the size of neoplasias in colonoscopy using depth-from-defocus, IEEE Engineering in Medicine and Biology Conference.

Chadebecq, F., Tilmant, C., Bartoli, A., 2013. Using the infocus-breakpoint to estimate the scale of neoplasia in colonoscopy, IEEE International Symposium on Biomedical Imaging.

Chang, C., Chiu, H., Wang, H., Lee, C., Tai, J., Tu, C., Tai, C., Chiang, T., Huang, J., Chang, D., Lin, J., 2010. An endoscopic training model to improve accuracy of colonic polyp size measurement. International Journal of Colorectal Disease 25, 655–660.

Chaptini, L., Chaaya, A., Depalma, F., Hunter, K., Peikin, S., Laine, L., 2014. Variation in polyp size estimation among endoscopists and impact on surveillance intervals. Journal of Gastrointestinal Endoscopy 0, 1–8.

Chen, C., Sargent, D., Wang, Y., 2010. Modeling tumor, polyp, lesion structure in 3D for computer-aided diagnosis in colonoscopy, SPIE Medical Imaging.

Claxton, C., Staunton, R., 2008. Measurement of the point-spread function of a noisy imaging system. Journal of the Optical Society of America A 25, 159–170.

Deschênes, F., Ziou, D., Fuchs, P., 2004. An unified approach for a simultaneous and cooperative estimation of defocus blur and spatial shifts. Image and Vision Computing 22, 35–57.

Endoscopic Classification Review Group, 2005. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. Journal of Endoscopy 37, 570–578.

Favaro, P., Soatto, S., Burger, M., Osher, S.J., 2008. Shape from defocus via diffusion. IEEE Transaction on Pattern Analysis and Machine Intelligence 30, 518–531.

Field, M., Clarke, D., Strup, S., Seales, W., 2009. Stereo endoscopy as a 3D measurement tool, IEEE Engineering in Medicine and Biology Conference.

Hartley, R.I., Zisserman, A., 2004. Multiple view geometry in computer vision – 2nd Edition. Cambridge University Press.

Hensley, H., Merkel, C., Chang, W., Devarajan, K., Cooper, H., Clapper, M., 2009. Endoscopic imaging and size estimation of colorectal adenomas in the multiple intestinal neoplasia mouse. Journal of Gastrointestinal Endoscopy 69, 742–749.

Huber, P., 1964. Robust estimation of a location parameter. The Annals of Mathematical Statistics 35, 73–101.

Hyun, Y., Han, D., Bae, J., Park, H., Eun, C., 2011. Graduated injection needles and snares for polypectomy are useful for measuring colorectal polyp size. Journal of Digestive and Liver Disease 43, 391–394.

Kaufman, A., Wang, J., 2008. Visualization in Medicine and Life Sciences. Springer Berlin Heidelberg.

Kim, D., Pickhardt, P., Taylor, A., Leung, W., Winter, T., Hinshaw, J., Gopal, D., Reichelderfer, M., Richard, H., Pfau, P., 2007. CT colonography versus colonoscopy for the detection of advanced neoplasia. The New England Journal of Medicine 357, 1403–1412.

Koppel, D., Chen, C., Wang, Y., Lee, H., Gu, J., Poirson, A., Wolters, R., 2007. Toward automated model building from video in computer-assisted diagnoses in colonoscopy, SPIE Medical Imaging.

Kudo, S., Hirota, S., Nakajima, T., Hosobe, S., Kusaka, H., Kobayashi, T., Himori, M., Yagyuu, A., 1994. Colorectal tumours and pit pattern. Journal of Clinical Pathology 47, 880–885.

Lee, Y.C., Moller, K., Suthakorn, J., 2010. Non-radiological colonoscope tracking, IEEE Conference on Robotics Automation and Mechatronics.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110.

Mannan, S., Choi, T.S., 2011. An unorthodox approach towards shape from focus, IEEE Conference on Image Processing.

Margulies, C., Krevsky, B., Catalano, M., 1994. How accurate are endoscopic estimates of size? Journal of Gastrointestinal Endoscopy 40, 174–176.

Mirota, D., Ishii, M., Hager, G., 2011. Vision-based navigation in image-guided interventions. Annual Review of Biomedical Engineering 13, 297–319.

Moug, S., Vernall, N., Saldanha, J., McGregor, J., Balsitis, M., Diament, R., 2010. Endoscopists estimation of size should not determine surveillance of colonic polyps. International Journal of Colorectal Disease 7, 646–650.

Myles, Z., da Vitoria Lobo, V., 1998. Recovering affine motion and defocus blur simultaneously. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 652–658.

Namboodiri, V., Chaudhuri, S., 2007. On defocus, diffusion and depth estimation. Pattern Recognition Letters 28, 311–319.

Parot, V., Lim, D., González, G., Traverso, G., Nishioka, N., Vakoc, B., Durr, N., 2013. Photometric stereo endoscopy. Journal of Biomedical Optics 18, 1–8.

Pentland, A., 1987. A new sense for depth of field. IEEE Transactions on Pattern Analysis and Machine Intelligence 9, 523–531.

Reichl, T., Gardiazabal, J., Navab, N., 2013. Electromagnetic servoing - a new tracking paradigm. IEEE Transaction on Medical Imaging 32, 1526–1535.

Rubio, C., Jnasson, J., Nesi, G., Mazur, J., Olafsdttir, E., 2010. The size of colon polyps revisited: intra- and inter-observer variations. International Institute of Anticancer Research 6, 2419–2423.

Schmalz, C., Forster, F., Schick, A., Angelopoulou, E., 2012. An endoscopic 3D scanner based on structured light. Medical Image Analysis 16, 1063–1072.

Schoen, R., Gerber, L., Margulies, C., 1997. The pathologic measurement of polyp size is preferable to the endoscopic estimate. International Journal of Gastrointestinal Endoscopy 46, 492–496.

Shim, S.O., Choi, T.S., 2010. A novel iterative shape from focus algorithm based on combinatorial optimization. Pattern Recognition 43, 3338–3347.

Summers, R., 2010. Polyp size measurement at ct colonography: what do we know and what do we need to know?

Takeshita, T., Nakajima, Y., Kim, M., Onogi, S., Mitsuishi, M., Matsumoto, Y., 2009. 3D shape reconstruction endoscope using shape from focus, Conference on Computer Vision Theory and Applications.

Talbot, I., 1996. Redetection and growth of colorectal polyps. International Journal of Gastroenterology and Hepatology 39, 492–492.

Trucco, E., Verri, A., 1998. Introductory Techniques for 3D Computer Vision. Prentice Hall PTR.

Wei, Y., Dong, Z., Wu, C., 2009. Global depth from defocus with fixed camera parameters, IEE Conference on Mechatronics and Automation.

Wohler, C., d'Angelo, P., Kruger, L., Kuhl, A., Groß, H.M., 2009. Monocular 3D scene reconstruction at absolute scale. Journal of Photogrammetry and Remote Sensing 64, 529–540.

World Health Organization, 2008. World Cancer Report 2008. International Agency for Research on Cancer. URL: http://www.who.int/cancer/publications/world_cancer_report2008.

Zhou, C., Cossairt, O., Nayar, S., 2010. Depth from diffusion, IEEE Conference on Computer Vision and Pattern Recognition.

Zhuo, S., Sim, T., 2009. On the recovery of depth from a single defocused image. Computer Analysis of Images and Patterns 5702, 889–897.