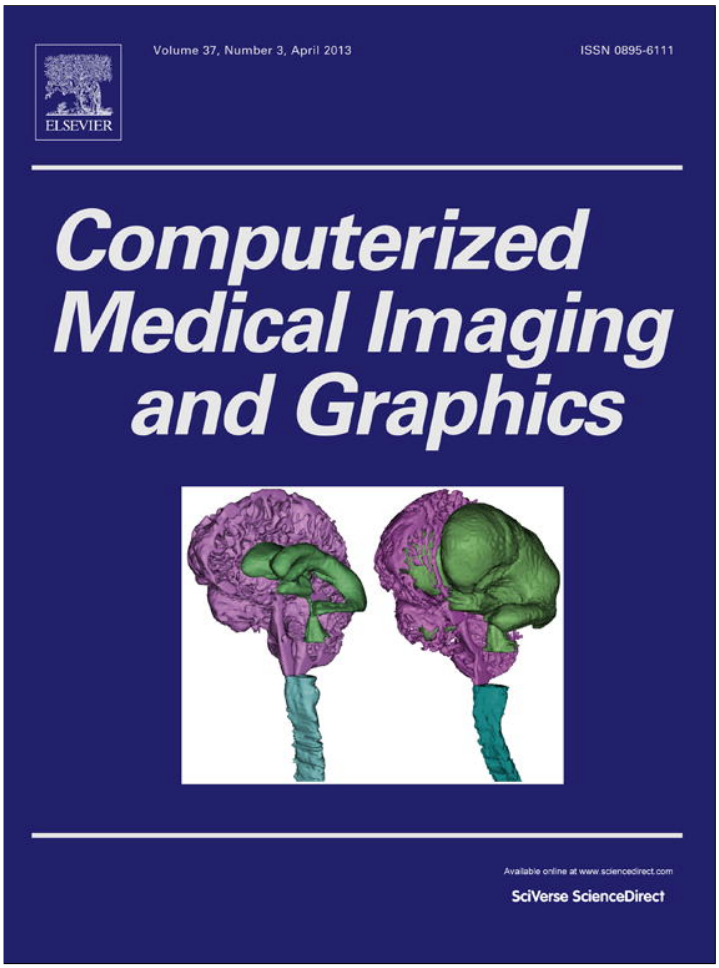


Computerized Medical Imaging and Graphics 37 (2013) 207–223

Contents lists available at SciVerse ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

An optical flow approach to tracking colonoscopy video

Jianfei Liu[a], Kalpathi R. Subramanian[b,*], Terry S. Yoo[c]

[a] Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD 20892, USA
[b] Department of Computer Science, The University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[c] Office of High Performance Computing and Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

ARTICLE INFO

Article history:
Received 25 June 2012
Received in revised form 18 January 2013
Accepted 25 January 2013

Keywords:
Colonoscopy
Tracking
Optical flow
Egomotion

ABSTRACT

We can supplement the clinical value of an optical colonoscopy procedure if we can continuously co-align corresponding virtual colonoscopy (from preoperative X-ray CT exam) and optical colonoscopy images. In this work, we demonstrate a computer vision algorithm based on optical flow to compute egomotion from live colonoscopy video, which is then used to navigate and visualize the corresponding patient anatomy from X-ray CT data. The key feature of the algorithm lies in the effective combination of sparse and dense optical flow fields to compute the focus of expansion (FOE); FOE permits independent computation of camera translational and rotational parameters, directly contributing to the algorithm's accuracy and robustness. We performed extensive evaluation via a colon phantom and clinical colonoscopy data. We constructed two colon like phantoms, a straight phantom and a curved phantom to measure actual colonoscopy motion; tracking accuracy was quantitatively evaluated by comparing estimated motion parameters (velocity and displacement) to ground truth. Thirty straight and curved phantom sequences were collected at 10, 15 and 20 mm/s (5 trials at each speed), to simulate typical velocities during colonoscopy procedures. The average error in velocity estimation was within 3 mm/s in both straight and curved phantoms. Displacement error was under 7 mm over a total distance of 287–288 mm in the straight and curved phantoms. Algorithm robustness was successfully demonstrated on 27 optical colonoscopy image sequences from 20 different patients, and spanning 5 different colon segments. Specific sequences among these were chosen to illustrate the algorithm's decreased sensitivity to (1) recording interruptions, (2) errors in colon segmentation, (3) illumination artifacts, (4) presence of fluid, and (5) changes in colon structure, such as deformation, polyp removal, and surgical tool movement during a procedure.

1. Introduction

Colorectal cancer caused 49,380 deaths in the United States in 2011 [52]. *Optical colonoscopy* (OC) is an important screening tool to detect and treat colon cancer. An OC procedure involves guiding a flexible endoscope into the colon, permitting visual inspection and removal of inflamed tissue, abnormal growth, and ulcers. OC is an exploratory procedure, dependent on the physician's skills and experience, and can miss cancer causing polyps [32,4]. *Virtual colonoscopy* (VC) [51,27] is an alternate screening tool, capable of providing fully interactive views of the interior of the colon for surgical planning and diagnosis. VC has limitations: lesions smaller than 5 mm cannot be reliably detected [25], and currently, there are no automatic techniques to simultaneously track conventional optical colonoscopy images and the pre-segmented virtual colonoscopy images.

Recent studies [4] have revealed that OC procedures miss more polyps than expected, especially in the ascending colon. Studies on polyp detection rates using OC and VC techniques [54,16,28] have also shown they are comparable. Thus, tools that automatically coalign OC and VC images can be useful in reducing the chance of missing colorectal lesions, as they provide better spatial perspective in the vicinity of the colonoscope location. Additionally, when pre-detected polyp information is available from the virtual images, their locations can be integrated into our tracking system, and serve to warn the gasteroenterologist of an 'oncoming' polyp during the procedure. Fig. 1 illustrates the overall scope of this work.

Similar to the use of GPS devices in navigation applications, a colonoscope can be augmented with location sensors; magnetic sensors have been used by several researchers [46,14,12]. However, these methods have two disadvantages: first, the colonoscopy has to be modified to include the sensors, as well as calibrating them. Second, the colon is a highly deformable organ; rigid transformation parameters measured from sensors do not compensate for the

* Corresponding author. Tel.: +1 704 687 8579; fax: +1 704 687 3516.
E-mail addresses: jianfei.liu@nih.gov (J. Liu), krs@uncc.edu (K.R. Subramanian), yoo@nlm.nih.gov (T.S. Yoo).

0895-6111/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compmedimag.2013.01.010

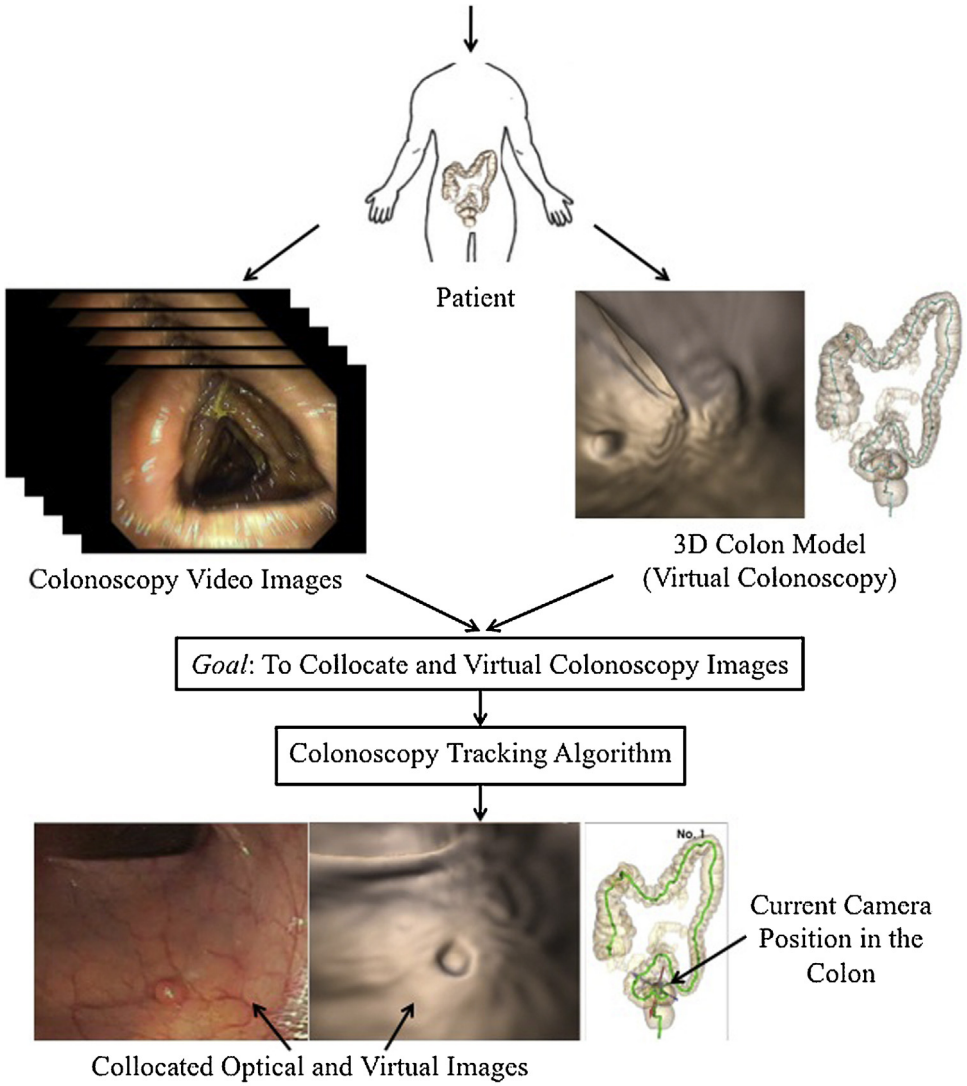


Fig. 1. Co-aligning optical and CT colonoscopy images. Optical colonoscopy and CT colonoscopy are two technologies to screen the colon. The goal of this work is to automatically co-align corresponding images so as to simultaneously present these to the gastroenterologist. Optical colonoscopy (on the left) produces a continuous stream of images. These are analyzed to determine the camera location and orientation, a process known as egomotion determination. This information is used in adjusting the virtual camera of the 3D reconstructed CT volume (on the right), resulting in the bottom view: co-aligned OC and VC images are shown, as well as the location of the virtual camera in the colon.

deformation, complicating the tracking system in co-aligning OC and VC images. In contrast, our approach is to exploit computer vision techniques to develop a robust tracking system without using *external sensors*.

In this paper, we present a tracking algorithm based on optical flow [24,43,8] to co-align optical and virtual colonoscopy images. The lack of topological information in colonoscopy images (the colon is a tubular structure without any bifurcations, unlike the bronchi), makes optical flow a reasonable choice for tracking image motion. Second, since the image motion between successive frames is relatively small, optical flow is a good choice for accurately representing this motion. Earlier work [40] presented the mathematical aspects of this method. Here we build on this work and present the following contributions:

1. We designed two colon-like phantoms (straight and curved) in the shape of a tunnel to generate ground truth. Unlike our earlier phantom [40], the goal here was to validate our tracking algorithm, by experiments (1) measuring colonoscope velocity and distance traveled, and (2) to ensure the results are repeatable. The colonoscope was displaced at speeds of 10, 15 and 20 mm/s, that are representative of colonoscopy procedures. At each speed, five trials were performed to analyze tracking accuracy. The average velocity error was within 3 mm/s in both straight and curved phantoms at all three speeds. Displacement error was less than 7 mm over a total distance of 287–288 mm in both phantoms.
2. We tested the colonoscopy tracking algorithm on 27 optical colonoscopy image sequences from 20 patients, and spanning 5 different colon segments.[1] These experiments demonstrate the robustness of our method with respect to the following:
   - interruptions in video recording;
   - structural changes of the colon, specifically relating to colon folds;
   - presence of fluid and/or blurry images;
   - structural changes due to surgical removal of polyps;

[1] Part of the Walter Reed Army Medical Center training dataset archive from the National Cancer Institute.

- use of surgical tools for polyp removal, resulting in motion of both the colonoscope, and the tool.

3. We used a cylinder of constant radius to approximate the virtual colon for computing depth values that were needed for egomotion determination. The sensitivity of the tracking pipeline was assessed on optical image sequences. Results indicated that this was sufficient for tracking colonoscopy data. More significantly, *this generalizes the algorithm to other applications, such as real-world navigation*, where accurate depth information is generally not available and must be approximated.

2. Related work

We review relevant work on image motion computation, egomotion estimation, and bronchoscopy tracking.

2.1. Image motion computation

Image motion inside video streams can be classified into two types [59]: (1) optical flow, and (2) corresponding pair computation.

Optical flow [24,5,43,49,50,30,17,58,8,53] is the distribution of apparent velocities of movement of intensity patterns. Optical flow is usually estimated with the intensity constancy model, in conjunction with other constraints [5] such as spatial coherence. Earlier methods assumed that optical flow remained locally constant [43,58] and estimated optical flow by minimizing intensity variance within a small neighborhood of successive frames. As the size of a local region is hard to determine, the resulting flow field is usually not consistent. Other approaches explicitly integrated the intensity constancy model with spatial smoothing constraints, resulting in a globally smooth flow field [24]. However, this can result in oversmoothing motion boundaries. Nagel [49,50] performed *oriented smoothness* by smoothing along directions where the intensity variation is small. Black [5] used *robust statistics* to reduce estimation outliers to retain the fidelity of the motion boundary. Lai and Vemuri [30] explicitly introduced a zero-crossing constraint to prevent smoothing along intensity discontinuities. Unfortunately, intensity discontinuities are not always the same as motion discontinuities. Recently, Brox [8,53], presented a variational model to combine and optimize several motion constraints, and obtained encouraging results.

Optical flow approaches can estimate small image motion quite well, but usually fail for large displacement, such as two images that bridge a blurry image sequence; in this case, the temporal derivatives are poorly estimated or undefined. Lowe [42] investigated this issue and proposed a scale invariant feature transform (SIFT), which involves a scale invariant region detector and a descriptor based on the gradient distribution in the detected regions. The region descriptor is insensitive to scale changes, as it can successfully search for corresponding pairs between two objects with large variation in size. Inspired by Lindeberg's affine scale-space work [35,36], a number of researchers [45,63,44,62,26] computed affine-invariant image patches and defined affine-invariant descriptors on them to eliminate affine distortion caused by the projection of widely separated cameras. However, these descriptors can achieve locally affine invariance, but are not globally invariant to image deformation [42]. All region descriptors are good at finding corresponding points with large image velocities, but are highly sensitive to the small image motion, since the region difference is not obvious within a one or two pixel displacement. Optical flow is thus a reasonable choice for small motion detection between successive frames, while region based image matching approaches take advantage of estimating large displacement caused by correspondingly large motion.

2.2. Egomotion determination

Optical flow computation techniques have led to tracking algorithms, generally referred to as *egomotion* determination. Bruss and Horn [11] proposed a least-square minimization scheme to search the 3D motion parameters that best approximate the measured flow field. In order to be less sensitive to inaccuracies and ambiguities in optical flow, Adiv [1,2] proposed a decomposition scheme to compute the motion parameters according to an estimated residual. However, these methods are still sensitive to the accuracy of the underlying flow field. Other researchers used motion parallax to compute invariant properties of the flow field [29,41,33], such as the *focus of expansion (FOE)*, in order to improve the robustness of the tracking algorithms. The FOE is defined as the projection of the camera's translational axis on the image plane. Detection of the focus of expansion permits independent estimation of translational and rotational velocities [21,57,60]. Their comparison can be found in Tian [61]. Our tracking algorithm will also utilize the FOE.

2.3. Bronchoscopy tracking

There has been a large effort focused on registering optical and virtual bronchoscopy images [7,56,13,48,46,14,15,23]. Bricault and Ferretti [7] pioneered this work and proposed a multi-stage tracking algorithm, making full use of anatomical marks, bifurcations, and repeated 2D and 3D registrations to align optical and virtual images. However, this method relies on anatomical features to reconstruct 3D shape. Mori [47] proposed a two-stage bronchoscopy tracking algorithm. Endoscope motion parameters were first estimated through optical-flow-based epipolar geometry between consecutive optical bronchoscopy images, and they were then refined through matching optical and virtual bronchoscopy images. Rai [56,55] assumed similarity in depth values of virtual and optical bronchoscopy images if the two views were aligned, and applied pose estimation approaches to match virtual and optical images. This method is sensitive to depth discontinuities, since depth accuracy is dependent on the sampling rate of the Z-buffer. This also restricts the method to applications that have access to a Z-buffer for computing depth values. Instead, in our method, depth values are computed from an approximation of the virtual colon. We use a cylinder of constant radius (we use the averaged radius of the virtual colon) built from the virtual colon centerline. Deguchi [13] utilized sum-of-square differences as a cost metric to measure the similarity of intensity distribution between virtual and optical images to estimate camera motion. Nagao [48] used a Kalman filter to linearly predict camera motion by combining registration results from previous frames, and reduced the search space of the motion parameters. Recent trends in bronchosocopy tracking integrate a sensor device with registration algorithms for robustness and accuracy [46,14,15].

Registration algorithms work well, since they exploit anatomical marks, bifurcations and other structural features that characterize bronchoscopy data. However, the colon has no bifurcations or other significant topological features that can be exploited. Second, an important goal of aligning virtual and optical bronchoscopy images is to assist needle biopsy procedures, where accuracy within a few millimeters is required. On the contrary, the goal of co-aligning virtual and optical colonoscopy images is for *image-guided navigation*, and it is sufficient to be in the vicinity of relevant features, such as a polyp or a particular colon fold. Third, in contrast to bronchoscopy data, colon images contain artifacts due to illumination (Fig. 12, row 2, area marked B), presence of fluid (Fig. 12, row 1, area marked A), blurry images due to the camera pointed towards the wall, or fast motion of endoscope (Fig. 12, row 3, optical image is very blurry), and colon deformation and movement (Fig. 11, distinction in the top row images between OC and VC

representing the same part of the colon). Registration algorithms to track colon images in the context of these artifacts can be quite challenging.

Ching [12] introduced a commercially available electromagnetic sensor to track an optical colonoscope. Validation experiments using their system on rigid phantoms showed promising results. A key advantage of using sensors in tracking systems is their ability to capture large camera motion, in contrast to image-based methods, as rigid motion parameters are easily obtained without the reliance on accurate image motion estimation. However, as observed in the bronchoscopy tracking methods [46,14,15], sensor-based methods have difficulty in accurately estimating non-rigid motion in the bronchia; the problem is worse in the deformable colon. Image registration is thus a necessary component of these methods to compensate for local deformation. Our approach is image-based, and can produce sufficiently accurate tracking, as detailed in the following sections.

3. Relationship between colonoscope motion and optical flow

We begin by describing the mathematical relationship between colonoscope motion and optical flow and derive the governing equations for egomotion determination.

A video stream generated by the motion of a colonoscope is defined as $I(x, y, t) : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}$. The primary goal of colonoscopy tracking is to compute from $I(x, y, t)$ the camera trajectory $\Upsilon(t) = [\Upsilon_X(t), \Upsilon_Y(t), \Upsilon_Z(t)] : \mathbb{R}^+ \to \mathbb{R}^3$, and the camera orientation $\Theta(t) = [\Theta_X(t), \Theta_Y(t), \Theta_Z(t)] : \mathbb{R}^+ \to \mathbb{R}^3$, in order to drive the virtual colonoscopy camera.

$$\begin{aligned} \Upsilon(t) &= \Upsilon(0) + \int_0^t \vec{T}^W(\tau)\, d\tau \\ \Theta(t) &= \Theta(0) + \int_0^t \vec{R}^W(\tau)\, d\tau \end{aligned} \tag{1}$$

where $\vec{T}^W(t) = [T_X^W(t), T_Y^W(t), T_Z^W(t)] : \mathbb{R}^+ \to \mathbb{R}^3$ and $\vec{R}^W(t) = [R_X^W(t), R_Y^W(t), R_Z^W(t)] : \mathbb{R}^+ \to \mathbb{R}^3$ are translational and rotational velocities in the world coordinates, $\Upsilon(0) = (0, 0, 0)$ and $\Theta(0) = (0, 0, 0)$; thus, at $t = 0$, the world and camera coordinates are exactly aligned.

The motion information inside $I(x, y, t)$ is a spatial–temporal motion field, $\mathcal{F}(\vec{v}) = \vec{v}(x, y, t) : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}^3$, caused by motion of the camera relative to the object. However, the exact motion field is usually unknown, and optical flow, $\mathcal{F}(\vec{u}) = \vec{u}(x, y, t) : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}^3$, can be considered as an approximate motion field. The problem of estimating camera translational and rotational velocities $\vec{T}^W(t)$ and $\vec{R}^W(t)$ is thus formulated as minimizing the difference between the motion field and optical flow,

$$\min_{\vec{T}^W(t), \vec{R}^W(t)} ||\mathcal{F}(\vec{u}) - \mathcal{F}(\vec{v})|| = \min_{\vec{T}^W(t), \vec{R}^W(t)} \iint ||\vec{u}(x, y, t) - \vec{v}(x, y, t)||^2 \, dx\, dy \tag{2}$$

Next we analyze the projective relationship between the camera motion velocities $\vec{T}^W(t)$ and $\vec{R}^W(t)$ and the motion field $\mathcal{F}_v$ at time t. As it is more convenient to deduce this relationship in the instantaneous camera coordinates, let $\vec{T}^C(t) = [T_X^C(t), T_Y^C(t), T_Z^C(t)]$ and $\vec{R}^C(t) = [R_X^C(t), R_Y^C(t), R_Z^C(t)]$ represent camera velocities in the camera coordinates. We can now write

$$\vec{T}^W(t) = \mathbf{M}^T(t)\vec{T}^C(t) \quad \vec{R}^W(t) = \mathbf{M}^R(t)\vec{R}^C(t) \tag{3}$$

where $\mathbf{M}^T$ and $\mathbf{M}^R$ represent the affine transform between two coordinate systems [18].

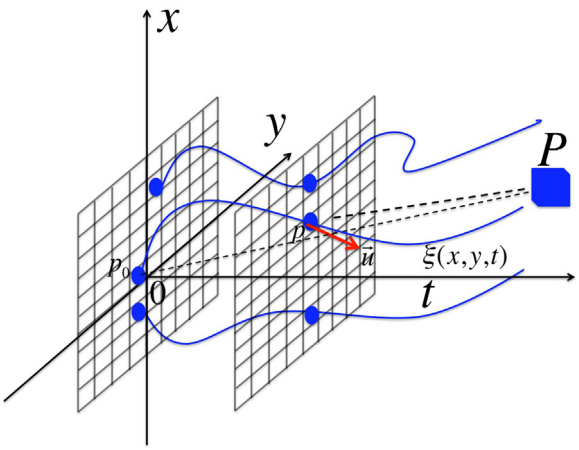


Fig. 2. Optical flow as a function of time. Optical flow estimation is based on the assumption that intensities of image points, such as p_0 and p (filled blue circles), projected from the same object point P remain invariant. These projected points form a profile curve, $\xi(x, y, t)$; several such curves are illustrated. The optical flow vector $\vec{u}$ at time t is the tangent vector of $\xi(x, y, t)$, indicated by the red arrow. (For interpretation of colour in the artwork, the reader is referred to the web version of the article.)

$\vec{T}^C(0) = \vec{T}^W(0)$ and $\vec{R}^C(0) = \vec{R}^W(0)$ because $\mathbf{M}^T(0) = \mathbf{I}$ and $\mathbf{M}^R(0) = \mathbf{I}$ at $t = 0$. Let $P = (X, Y, Z)$ be an object point in the camera coordinate and its projection point on the image plane is $p = (x, y) = (fX/Z, fY/Z)$, where f is the focal length of the camera. If P is observed from the camera coordinate, it moves towards the camera at the speed of $-\vec{T}^C(t)$ and $-\vec{R}^C(t)$ caused by the actual camera movement. Its motion vector [11] at point p is given by

$$\begin{aligned} \vec{v}(x, y, t) &= \begin{bmatrix} v_x(x, y, t) \\ v_y(x, y, t) v_t(x, y, t) \end{bmatrix} \\ &= \begin{bmatrix} \frac{T_Z^C(t)}{Z(x, y, t)} \left(x - \frac{f T_X^C(t)}{T_Z^C(t)} \right) + R_X^C(t) \frac{xy}{f} - R_Y^C(t) \left(f + \frac{x^2}{f} \right) + R_Z^C(t) y \\ \frac{T_Z^C(t)}{Z(x, y, t)} \left(y - \frac{f T_Y^C(t)}{T_Z^C(t)} \right) + R_X^C(t) \left(f + \frac{y^2}{f} \right) - R_Y^C(t) \frac{xy}{f} - R_Z^C(t) x \\ \alpha \end{bmatrix} \end{aligned} \tag{4}$$

where α is a constant temporal component. $Z(x, y, t)$ is the depth value, and it varies with each image point. Thus, Eq. (4) defines $\mathcal{F}_v$, the motion field at time t. If $\mathcal{F}_v$ is known, $\vec{T}^C(t)$ and $\vec{R}^C(t)$ can be determined, and using Eq. (3), $\vec{T}^W(t)$ and $\vec{R}^W(t)$.

Optical flow is used to approximate the motion field, assuming that intensities of image points remain invariant if these points are projected from the same object. For instance, in Fig. 2, all the projection points of a point P at varying times are of the same intensity (the filled blue circles along each curve). These points form a profile curve $\xi(x, y, t)$ from $[0, t]$. Three such curves are shown in Fig. 2. Assuming p is the projection point at t, its optical flow vector, $\vec{u} = (u_x, u_y, u_t)$,[2] is the tangent vector of ξ (the red arrow in Fig. 2). Let $p_0 = (x_0, y_0)$ be the projection point at $t = 0$, thus

$$\xi = (x_0, y_0, 0) + \int_0^t \vec{u}\, d\tau = \begin{bmatrix} x_0 + \int_0^t u_x\, d\tau \\ y_0 + \int_0^t u_y\, d\tau \\ \int_0^t u_t\, d\tau \end{bmatrix} \tag{5}$$

[2] To simplify the notation, (x, y, t) of all functions is ignored, thus, $\vec{u}$ will denote $\vec{u}(x, y, t)$.

and

$$\frac{dI(\xi)}{dt} = 0 \Rightarrow \partial_x I u_x + \partial_y I u_y + \partial_t I u_t = 0 \tag{6}$$

However, determining $\vec{u}$ is an under-constrained problem. For instance, Horn [24] assumed the optical flow vector at p is similar to its neighborhood and proposed a smoothness constraint, expressed as the square of the magnitude of the gradient of the optical flow vector,

$$S(x,y) = |\nabla u_x|^2 + |\nabla u_y|^2 \tag{7}$$

Combining Eq. (6) and Eq. (7),

$$\iint [(\partial_x I u_x + \partial_y I u_y + \partial_t I u_t)^2 + \beta(|\nabla u_x|^2 + |\nabla u_y|^2)]\, dx\, dy \tag{8}$$

Calculus of variations is applied to minimize Eq. (8) and determine the optical flow field, $\mathcal{F}_u$. Here, β is a constant. Additional constraints such as gradient constancy [8,53] or robust statistics to avoid over-smoothing along edges [5] can be used to further improve optical flow accuracy.

After optical flow is determined, Eq. (2) can be converted into a 6×6 linear system [22,11]. $\vec{T}^C(t)$ and $\vec{R}^C(t)$ are first computed, followed by $\vec{T}^W(t)$ and $\vec{R}^W(t)$ using Eq. (3). However, there are errors in the optical flow estimation that make it impractical to directly solve this linear system. In our earlier work [40], matrix perturbation theory was used to illustrate the numerical issues and the resulting instabilities using this approach.

Thus, instead of this direct approach, we compute the *focus of expansion (FOE)* [41] to separate the computation of camera translation and rotation velocities. The intersection between $\vec{T}^C(t)$ and the image plane is defined as the *focus of expansion* when the camera moves towards the object, and as the the *focus of contraction*, when it moves away from it. If $\vec{T}^C(t)$ is parallel to the image plane, the intersection is at infinity. FOE makes it possible to separate translational and rotational components from the motion field, $\mathcal{F}_v$ since it is determined solely by the translational velocity $\vec{T}^C(t)$. This makes it possible to decompose the original 6×6 system into two 3×3 systems, improving the numerical characteristics of the solver.

We can split $\vec{v}$ into two vectors, $\vec{v}^T$ and $\vec{v}^R$, corresponding to the camera translation and rotation (and using Eq. (4)),

$$\vec{v} = \vec{v}^T + \vec{v}^R \tag{9}$$

where

$$\vec{v}^T = \begin{bmatrix} v_x^T \\ v_y^T \\ v_t^T \end{bmatrix} = \begin{bmatrix} \frac{T_Z^C}{Z}\left(x - f\frac{T_X^C}{T_Z^C}\right) \\ \frac{T_Z^C}{Z}\left(y - f\frac{T_Y^C}{T_Z^C}\right) \\ 0 \end{bmatrix} \tag{10}$$

and

$$\vec{v}^R = \begin{bmatrix} v_x^R \\ v_y^R \\ v_t^R \end{bmatrix} = \begin{bmatrix} R_X^C \frac{xy}{f} - R_Y^C\left(f + \frac{x^2}{f}\right) + R_Z^C y \\ R_X^C\left(f + \frac{y^2}{f}\right) - R_Y^C \frac{xy}{f} - R_Z^C x \\ \alpha \end{bmatrix} \tag{11}$$

It can be seen from Eq. (10) that the spatial components of $\vec{v}^T$ intersect at $(fT_X^C/T_Z^C, fT_Y^C/T_Z^C)$, which is the FOE. In other words, spatial components of translational motion vector $\vec{v}^T$ are always parallel to the 2D vector joining the current feature point p to the FOE. The idea of motion parallax proposed by Longuet [41] can be used to compute the FOE. It can be determined by searching for pairs of adjacent points near depth discontinuities.

Let $\vec{d} = [d_x, d_y]$ be a 2D vector joining the current feature point p to FOE and $\vec{d}_\perp = [d_{\perp x}, d_{\perp y}]$ is perpendicular to $\vec{d}$. Including the temporal component, let $\vec{e} = [d_x, d_y, 0]$ and $\vec{e}_\perp = [d_{\perp x}, d_{\perp y}, 0]$. As $\vec{d}$ is parallel to $\vec{v}^T$ from Eq. (10), $\vec{v}^T \cdot \vec{e}_\perp = 0$. We can eliminate translational velocity as follows:

$$\vec{v}^R \cdot \vec{e}_\perp = \vec{v}^R \cdot \vec{e}_\perp + \vec{v}^T \cdot \vec{e}_\perp = (\vec{v}^T + \vec{v}^R) \cdot \vec{e}_\perp = \vec{v} \cdot \vec{e}_\perp = \vec{u} \cdot \vec{e}_\perp \tag{12}$$

Substituting Eq. (11) into Eq. (12), we obtain

$$\begin{bmatrix} R_X^C \frac{xy}{f} - R_Y^C\left(f + \frac{x^2}{f}\right) + R_Z^C(t) y \\ R_X^C\left(f + \frac{y^2}{f}\right) - R_Y^C \frac{xy}{f} - R_Z^C x \\ \alpha \end{bmatrix} \cdot \vec{e}_\perp = \vec{u} \cdot \vec{e}_\perp \tag{13}$$

A sequence of linear equations is constructed over the motion field through Eq. (13), and rotational velocity $\vec{R}^C$ is estimated.

Assume depth values Z are known from another source, such as the Z-buffer of virtual colonoscopy viewer, as described in Section 4.3.2. Substituting $\vec{R}^C$ into Eq. (11) to compute $\vec{v}^R$, and letting ε be the square of the difference between the motion field and optical flow components (caused only by translation),

$$\begin{aligned} \varepsilon &= \iint ||\vec{v}^T - (\vec{u} - \vec{v}^R)||^2\, dx\, dy \\ &= \iint \left\| \begin{bmatrix} \frac{T_Z^C}{Z}\left(x - f\frac{T_X^C}{T_Z^C}\right) \\ \frac{T_Z^C}{Z}\left(y - f\frac{T_Y^C}{T_Z^C}\right) \\ 0 \end{bmatrix} - (\vec{u} - \vec{v}^R) \right\|^2 dx\, dy \end{aligned} \tag{14}$$

Set $\partial/\partial \vec{T}^C \varepsilon = 0$ to minimize Eq. (14), and a 3×3 linear system is obtained,

$$\mathbf{A}\vec{T}^C = \vec{b} \tag{15}$$

where

$$\mathbf{A} = \begin{bmatrix} -\iint \frac{f}{Z}\, dx\, dy & 0 & \iint \frac{x}{Z}\, dx\, dy \\ 0 & -\iint \frac{f}{Z}\, dx\, dy & \iint \frac{y}{Z}\, dx\, dy \\ -\iint \frac{xf}{Z}\, dx\, dy & -\iint \frac{yf}{Z}\, dx\, dy & \iint \frac{(x^2+y^2)}{Z}\, dx\, dy \end{bmatrix} \tag{16}$$

and

$$\vec{b} = \begin{bmatrix} \iint (u_x - v_x^R)\, dx\, dy \\ \iint (u_y - v_y^R)\, dx\, dy \\ \iint (x(u_x - v_x^R) + y(u_y - v_y^R))\, dx\, dy \end{bmatrix} \tag{17}$$

$\vec{T}^C$ can be computed from this linear system.

After $\vec{T}^C$ and $\vec{R}^C$ are obtained, $\vec{T}^W$ and $\vec{R}^W$ are computed using Eq. (3). Substituting them into Eq. (1), the camera trajectory and orientation are given by

$$\begin{aligned} \Upsilon(t) &= \Upsilon(0) + \int_0^t \mathbf{M}^T(\tau)\vec{T}^C(\tau)\, d\tau \\ \Theta(t) &= \Theta(0) + \int_0^t \mathbf{M}^R(\tau)\vec{R}^C(\tau)\, d\tau \end{aligned} \tag{18}$$

Thus, $\Upsilon(t)$ and $\Theta(t)$ are incrementally recovered and used to drive virtual colonoscopy camera.

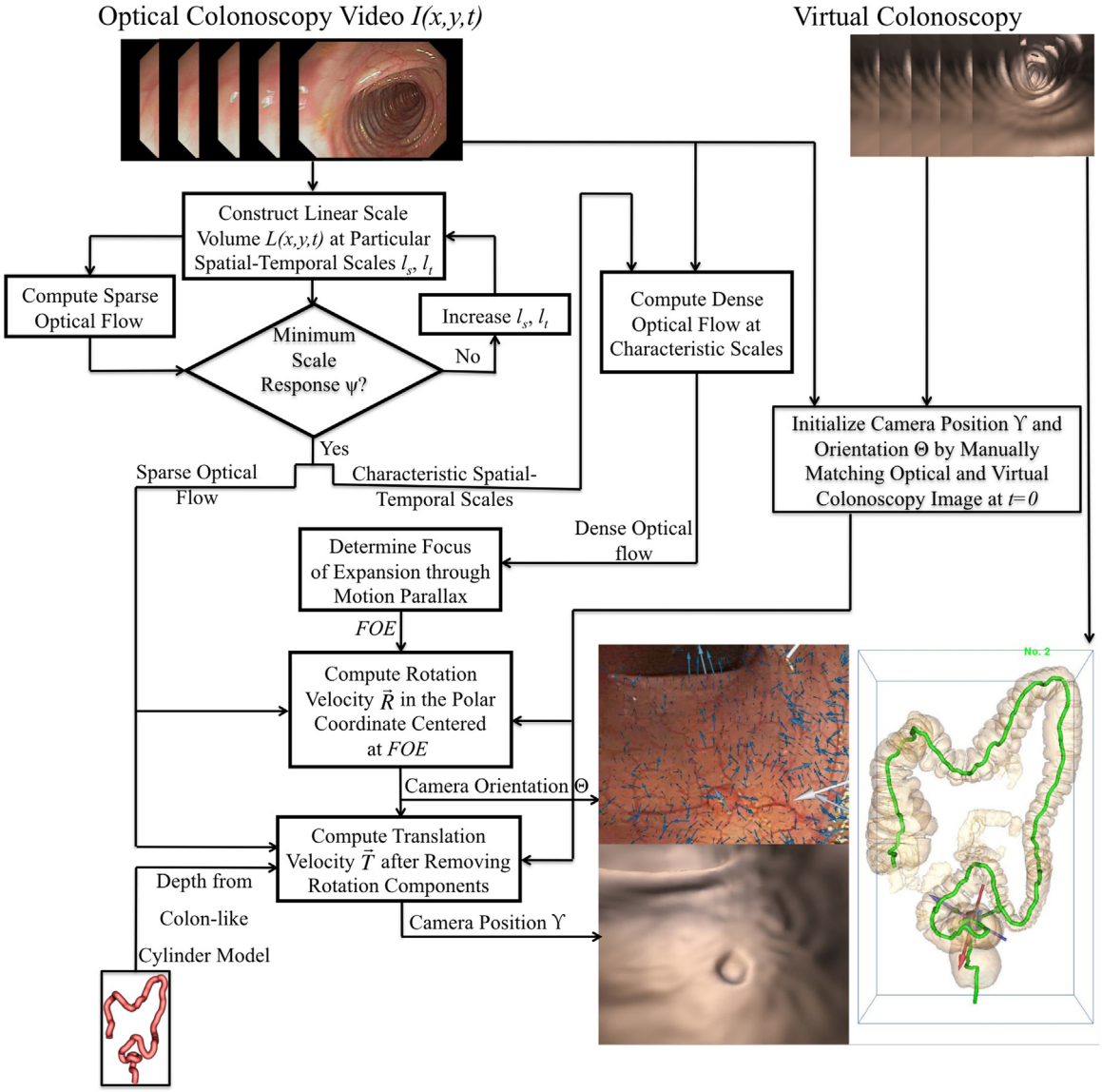


Fig. 3. The colonoscopy tracking algorithm. The input consists of the optical colonoscopy video stream and the CT images that have been segmented and reconstructed into a 3D volume. Scale space analysis is performed to compute the optimal spatial–temporal scales for each image, prior to computation of the sparse optical flow field using the Harris metric. These characteristic scales are used in computing the dense flow field, which in turn determines the focus of expansion (FOE). The FOE and sparse flow field are used to determine the camera rotational velocity. After removal of the rotational velocity from the optical flow field, translational velocity is determined, using depth values from a colon model. The camera parameters are transformed into CT volume coordinates to adjust the virtual colonoscopy camera, as illustrated in the bottom right.

4. The colonoscopy tracking algorithm

Fig. 3 shows the framework of our colonoscopy tracking algorithm. It begins by identifying a small set of accurate sparse optical flow vectors and determining *characteristic spatial–temporal scales* for each colonoscopy image. These are then employed to compute a dense optical flow field, from which we compute the FOE. The FOE and the sparse flow field are then used to estimate the rotational velocity of the camera. Finally, the translational velocities are computed through elimination of the rotational components from the flow field. Depth values used in this computation are derived from a cylindrical model of the virtual colon.

4.1. Sparse optical flow and scale selection

Our tracking algorithm begins by identifying a relatively small set of stable feature points and their corresponding optical flow, resulting in a *sparse* optical flow field. The flow field is determined using a *multi-scale* approach in order to stabilize the process of flow computation as well as reduce the influence of noise.

An anisotropic Gaussian scale-space representation $L : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+^2 \to \mathbb{R}$ [31] of an image sequence is constructed by convolution of $I(x, y, t)$ with a Gaussian kernel with distinct spatial and temporal scale parameters l_s, l_t,

$$L(x, y, t; l_s, l_t) = G(x, y, t; l_s, l_t) \cdot I(x, y, t) \tag{19}$$

where

$$G(x, y, t; l_s, l_t) = \frac{e^{((-(x^2+y^2)/2l_s)-(t^2/2l_t))}}{\sqrt{(2\pi)^3 l_s^2 l_t}} \tag{20}$$

and the semicolon in $G(x, y, t; l_s, l_t)$ implies that the convolution is performed only over x, y, t, while l_s and l_t indicate the spatial and temporal scale parameters. Since optical flow $\vec{u}$ is represented in the spatial–temporal domain, the anisotropic Gaussian kernel is applied to account for the differential sampling rates across the spatial and temporal dimensions. Unlike methods that consider spatial

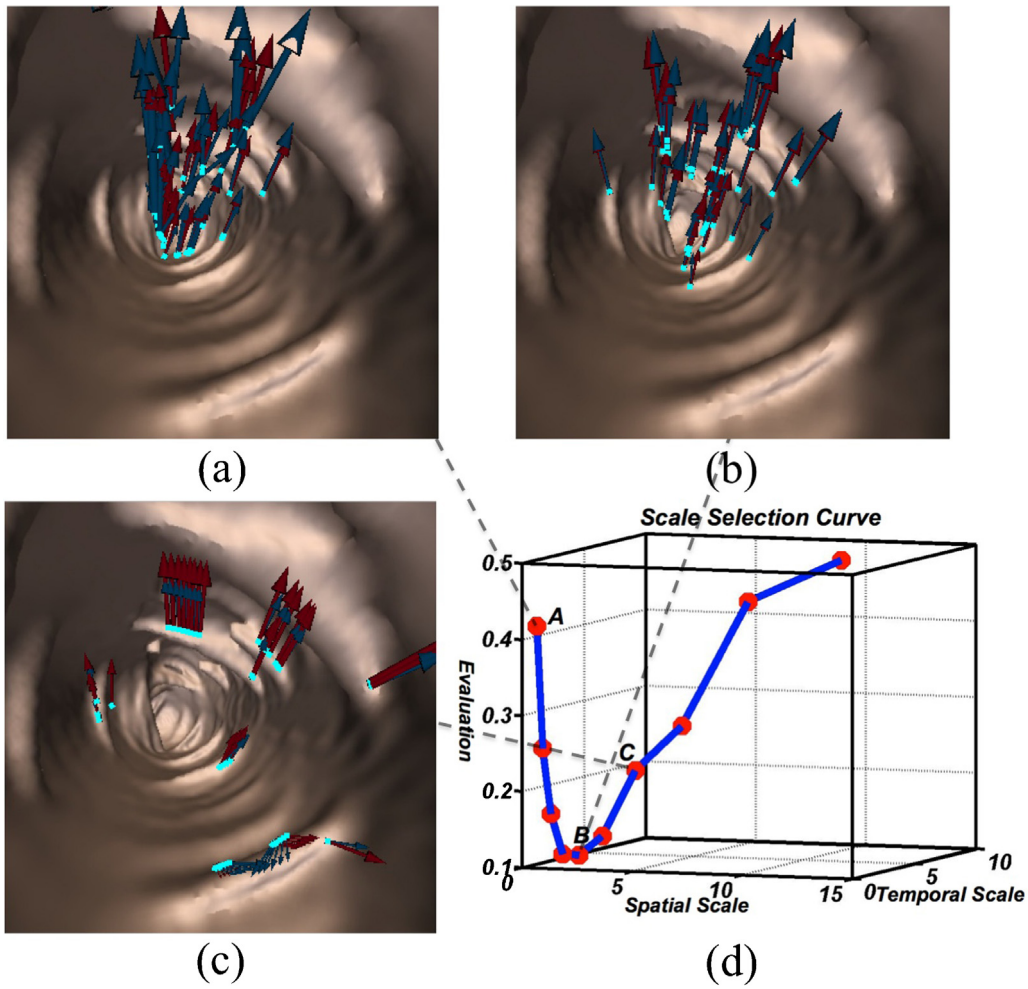


Fig. 4. Illustration of spatial–temporal scale selection in optical flow computation. Ground-truth flow vectors are in red and estimated flow vectors are in blue. Green cubes represent interest point positions: (a) fine spatial and temporal scales, (b) optimal spatial and temporal scales, (c) coarse scales, (d) the response curve between spatial–temporal scales and the metric values. The scale values at points A, B and C correspond to images (a), (b) and (c), respectively. (For interpretation of colour in the artwork, the reader is referred to the web version of the article.)

[34] or the temporal scale [64] individually, our approach is targeted toward determining the optimal spatial and temporal scales for optical flow computation. During implementation, 11 consecutive colonoscopy images centered at the current frame are buffered to formulate a local temporal sequence. The anisotropic scale space is then built over the sequence to preserve local details in the video stream and keep computational costs reasonable.

In order to reduce the ambiguities in corresponding point-pairs, we are interested in feature correspondences that exhibit maximum variance in the spatial domain and minimum temporal difference. *Interest points* [20] are considered good feature candidates in the spatial domain, as there are at least two dominant edge directions in their neighborhood; they are detected by the *Harris matrix*, defined as

$$\mathbf{H} = G(x, y; l_w) \cdot \begin{bmatrix} (\partial_x L)^2 & (\partial_x L)(\partial_y L) \\ (\partial_x L)(\partial_y L) & (\partial_y L)^2 \end{bmatrix} \tag{21}$$

where $\partial_x L = \partial/\partial x\,(L(x, y, t; l_s, l_t))$, $\partial_y L = \partial/\partial y\,(L(x, y, t; l_s, l_t))$, and

$$G(x, y; l_w) = \frac{e^{(-(x^2+y^2)/2l_w)}}{2\pi l_w} \tag{22}$$

is a Gaussian window function, with $l_w = \sqrt{2} l_s$. The distinctness of an interest point can be measured as

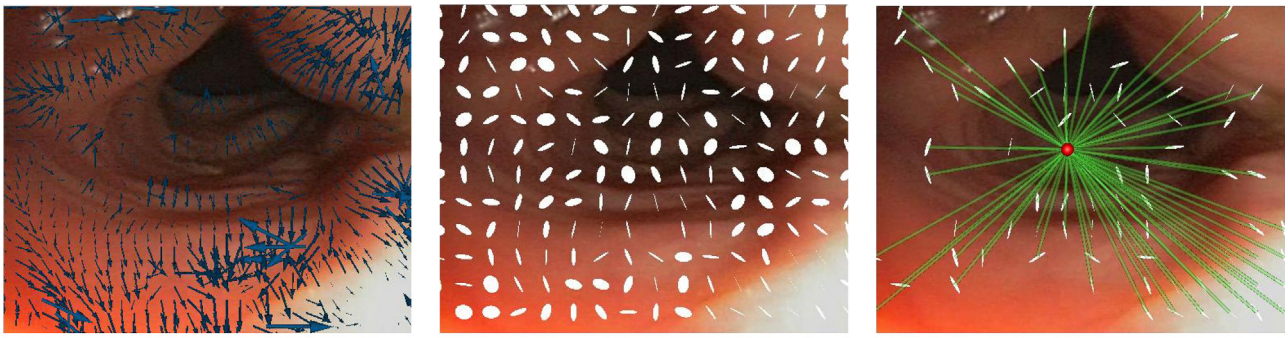
$$C(x, y, t; l_s, l_t) = \det(\mathbf{H}) - \gamma \mathrm{Trace}^2(\mathbf{H}) \tag{23}$$

where γ is a constant.

By combining Eqs. (19) and (23) and setting the temporal component α in Eq. (4) equal to 1, we propose the following scale-space metric for computing sparse optical flow,

$$\begin{aligned} &N(x, y, t; l_s, l_t) \\ &= \frac{G(x, y; l_w) \cdot [L(x, y, t; l_s, l_t) - L(x + u_x, y + u_y, t + 1; l_s, l_t)]^2}{\sqrt{|C(x, y, t; l_s, l_t)|} + 1.0} \\ &\approx \frac{G(x, y; l_w) \cdot [(\partial_x L)u_x + (\partial_y L)u_y + (\partial_t L)]^2}{\sqrt{|C(x, y, t; l_s, l_t)|} + 1.0} \end{aligned} \tag{24}$$

The numerator in Eq. (24) represents the similarity between corresponding pairs, while the denominator measures how distinct the selected features are in their local neighborhood. Good corresponding point-pairs should make the numerator (temporal difference) as small as possible and the denominator (spatial distinctiveness) as large as possible. Thus, the smaller the response of $N(x, y, t; l_s, l_t)$, the better the match. A critical property of the scale-selection metric is that it is invariant with respect to changes



(a) (b) (c)

Fig. 5. Determining the *focus of expansion*: (a) dense optical flow. (b) Anisotropy of the covariance matrix of Eq. (26) in each region of the grid, indicated by ellipse; principal orientation within each region is indicated by the long axes of the ellipses. (c) FOE (intersection of the green lines) is determined by least-squares fitting by choosing high confidence regions. Most of these regions are near depth discontinuities. (For interpretation of colour in the artwork, the reader is referred to the web version of the article.)

in scale, and characteristic scale is defined as the scale over its local extrema. In Eq. (24), its numerator consists of multiple combinations of derivative operators including ∂_x^2, ∂_y^2, ∂_t^2, $\partial_x\partial_y$, $\partial_x\partial_t$ and $\partial_y\partial_t$, while the denominator is the root of derivative combination of ∂_x^4, ∂_y^4 and $\partial_x^2\partial_y^2$. Thus, the derivative combinations of both numerator and denominator are of order 2. The response of Eq. (24) therefore remains invariant to scale changes, and is the basis for spatial and temporal scale selection. Fig. 4(d) illustrates the scale response of $N(x, y, t; l_s, l_t)$ for an example, with the minimum *characteristic spatial and temporal scales* corresponding to point B.

Sparse optical flow is determined by computing the first derivative of the numerator of Eq. (24) with respect to (u_x, u_y), and setting them to zero to obtain its minimum value. A 2×2 linear system is obtained,

$$G(x, y; l_w) \cdot \begin{bmatrix} (\partial_x L)u_x + (\partial_y L)u_y + (\partial_t L)(\partial_x L) \\ (\partial_x L)u_x + (\partial_y L)u_y + (\partial_t L)(\partial_y L) \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{H} \begin{bmatrix} u_x \\ u_y \end{bmatrix} = G(x, y; l_w) \cdot \begin{bmatrix} (\partial_t L)(\partial_x L) \\ (\partial_t L)(\partial_y L) \end{bmatrix} \tag{25}$$

In Appendix A, Algorithm 1 illustrates the pseudocode on scale selection and sparse optical flow computation.

An image sequence from virtual colonoscopy was used to examine the effectiveness of the scale selection metric. Scale selection results are illustrated in Fig. 4. Fig. 4(d) shows a response curve plotted as a function of the spatial (l_s) and temporal (l_t) scale parameters. It can be seen that the response curve first decreases to a local minimum, and then gradually increases. There are also three navigation images overlaid with ground-truth flow vectors (in red) and estimated flow vectors (in blue). The small green cubes indicate the positions of the chosen feature points. Fig. 4(a), corresponding to point A in Fig 4(d) shows the results with fine spatial and temporal scales, where large vectors deviate from the ground truth because the scales are not large enough to eliminate the noise or large intensity variation; in Fig 4(c), which corresponds to point C in Fig 4(d), small vectors diverge because the chosen scales are too coarse and small areas with varying motion are merged. Spatial–temporal scales at the local minim are a means to balance between these two extremes, and as seen in Fig 4(b) (point B in Fig 4(d)), generate flow vectors close to the ground truth.

4.2. Dense optical flow and focus of expansion (FOE)

In order to accurately locate the position of the FOE, the motion information in the entire image is used. The image sequence is smoothed with the chosen spatial and temporal scales from the previous step and Horn's method [24] is used to compute the dense optical flow.

Longuet [41] demonstrated that the direction of spatial optical flow vector difference of two adjacent points at the depth discontinuity would point to the FOE, as their vector difference is caused primarily by the translational component. We use an adaptive subdivision method similar to Reiger [57], to find these points, as illustrated in Fig. 5. Fig. 5(a) illustrates the dense optical flow field at a particular frame. The image plane is subdivided into rectangular regions and white ellipses are located at their centers in Fig. 5b). The estimated spatial optical flow vector difference $\Delta\vec{u} = (\Delta u_x, \Delta u_y) = (u_x(p_c) - u_x(p), u_y(p_c) - u_y(p))$ between the center point $p_c = (x_c, y_c)$ and its neighbors $p = (x, y)$ are tabulated in each sub-region. The covariance matrix, $\mathbf{C}(\Delta\vec{u})$ in each sub-region S is thus formed,

$$\mathbf{C}(\Delta\vec{u}) = \begin{bmatrix} \sum_S (\Delta u_x)^2 & \sum_S \Delta u_x \Delta u_y \\ \sum_S \Delta u_x \Delta u_y & \sum_S (\Delta u_y)^2 \end{bmatrix} \tag{26}$$

The eigenvector (represented by the major axis of white ellipses in Fig. 5(b)) corresponding to the largest eigenvalue is the direction joining the center point to the FOE. The lengths of minor and major axes of all white ellipses are determined by two eigenvalues. The lower the eigen-ratio of this matrix $\delta = ||\lambda_{small}/\lambda_{large}||$, the higher the confidence of the computed direction. We can threshold this ratio to select the subregions with high confidence corresponding to line-like ellipses in Fig. 5c. A line-fitting procedure is performed on the selected regions and the intersection of these lines (shown in green) is the estimated FOE (filled red circle) in Fig. 5c. Note that most of the selected subdivision regions are near colon folds, which are areas of depth discontinuity.

4.3. Motion parameters

After the location of the FOE is determined, translational and rotational velocities are estimated separately.

4.3.1. Rotational velocity

The sparse optical flow field is used to determine the rotational velocity, due to its accuracy. Substituting all sparse optical flow

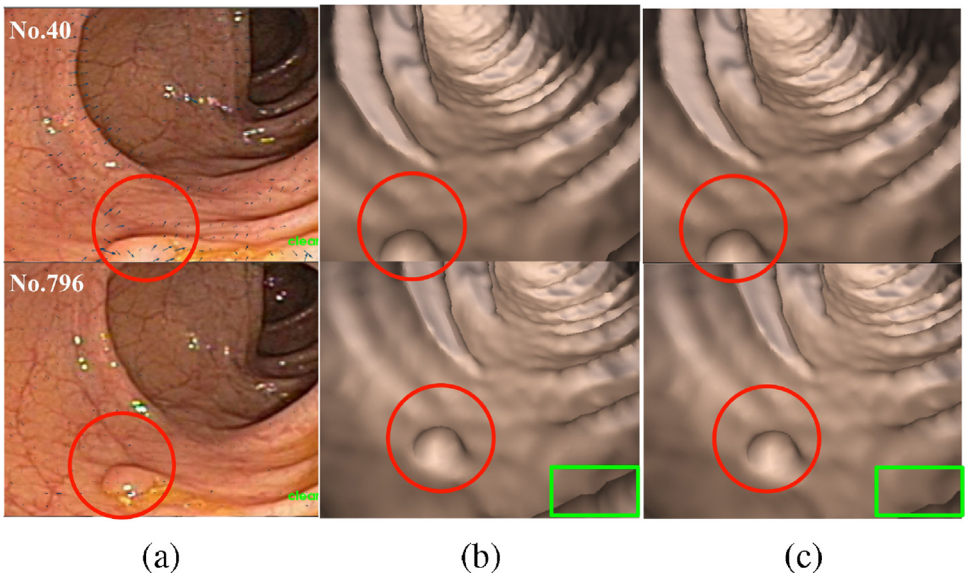


Fig. 6. (a–c) Comparison of tracking results by using depth values from a cylinder-like colon model and the actual segmented colon. To generalize the tracking algorithm, we use a cylinder like model derived from the 3D virtual colon. Results shown at two different frames. Left column illustrate the optical images, middle column shows results using depth from the virtual colon, and right column shows results using the colon model. A round polyp (red circle) is used as a reference to evaluate the tracking accuracy. Tracking results are comparable by using different depth sources, as the polyp is tracked well. (For interpretation of colour in the artwork, the reader is referred to the web version of the article.)

vectors into Eq. (13) leads to a sequence of linear equations for each sparse flow vector,

$$\begin{cases} \begin{bmatrix} R_X^C \frac{x_1 y_1}{f} - R_Y^C \left(f + \frac{x_1^2}{f}\right) + R_Z^C y_1 \\ R_X^C \left(f + \frac{y_1^2}{f}\right) - R_Y^C \frac{x_1 y_1}{f} - R_Z^C x_1 \\ 1 \end{bmatrix} \cdot \vec{e}_{\perp 1} = \vec{u}(x_1, y_1) \cdot \vec{e}_{\perp 1} \\ \vdots \\ \begin{bmatrix} R_X^C \frac{x_n y_n}{f} - R_Y^C \left(f + \frac{x_n^2}{f}\right) + R_Z^C y_n \\ R_X^C \left(f + \frac{y_n^2}{f}\right) - R_Y^C \frac{x_n y_n}{f} - R_Z^C x_n \\ 1 \end{bmatrix} \cdot \vec{e}_{\perp n} = \vec{u}(x_n, y_n) \cdot \vec{e}_{\perp n} \end{cases} \tag{27}$$

where n is the number of sparse optical flow vectors. Singular value decomposition is applied to compute $\vec{R}^C$.

4.3.2. Translational velocity

Discretizing Eq. (16) and Eq. (17), we obtain

$$\mathbf{A} = \sum_{i=1}^{n} \begin{bmatrix} -\frac{f}{Z_i} & 0 & \frac{x_i}{Z_i} \\ 0 & -\frac{f}{Z_i} & \frac{y_i}{Z_i} \\ -\frac{x_i f}{Z_i} & -\frac{y_i f}{Z_i} & \frac{(x_i^2 + y_i^2)}{Z_i} \end{bmatrix} \tag{28}$$

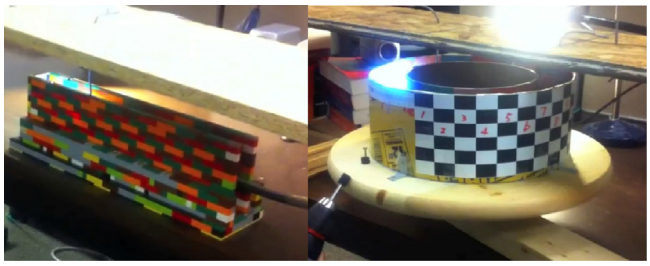
$$\vec{b} = \sum_{i=1}^{n} \begin{bmatrix} u_x(x_i, y_i) - v_x^R(x_i, y_i) \\ u_y(x_i, y_i) - v_y^R(x_i, y_i) \\ x_i(u_x(x_i, y_i) - v_x^R(x_i, y_i)) + y_i(u_y(x_i, y_i) - v_y^R(x_i, y_i)) \end{bmatrix} \tag{29}$$

where Z_i is the depth value of ith feature point.

Computation of the translation velocities requires the knowledge of Z_i. The depth value can be computed from the virtual colonoscopy viewer (via Z-buffer) based on the assumption that optical and virtual colonoscope cameras have similar depth maps if optical and virtual colonoscopy images are well aligned [56]. In our earlier work [40] we followed this strategy. The depth value of a feature point is computed by choosing the median value of an image region centered at the current feature location in the Z-buffer. The depth map is changing with the movement of the virtual camera. However, the assumption of continuously aligned optical and virtual images is not always valid, given that the colon can deform or undergo structural changes during the procedure. Also, note that Eq. (28) and Eq. (29) are somewhat insensitive to depth value errors, as they use averaged quantities in determining camera motion parameters. Thus, approximate depth values are sufficient to estimate accurate motion parameters. Moreover, using the Z-buffer from the segmented colon of a particular patient CT image restricts the generality of the tracking algorithm. A better alternative is to use a model of the colon in place of the virtual colon data.

We use a tube-like model to approximate the colon, as can be seen in lower-left of Fig. 3. The core of the colon model is the centerline from the virtual colon and the radius is the average distance of all centerline points to the colon boundary. Instead of using patient specific parameters, an alternate is to use a radius that is averaged over patients, further generalizing the model, or a model with varying radii that typically represent the different segments of the colon anatomy.

The sensitivity of the cylinder-like colon model is evaluated by comparing the tracking results using depth values from the actual CT colon model with the cylinder-like colon model. Fig. 6(a) shows two instances from a sequence of 796 OC images containing a polyp (marked inside red circles) in the descending colon. We compare tracking results between the use of depth values from the virtual colon (Fig. 6b) and the cylinder model (Fig. 6c). In both cases, tracking results are quite reasonable at frame 40 (top row). At frame 796, the only noticeable difference between the OC and VC images is the appearance of the fold, located in the bottom right corner (green squares). It is smaller using the cylinder model (bottom right image) versus the colon model (bottom center image). While a detailed study on this issue is beyond the scope of this work, experiments



(a) (b)

Fig. 7. (a) The straight phantom experiment and (b) the curved phantom experiment.

on additional clinical datasets illustrates no major differences or errors introduced through the use of depth values from a cylinder model.

Translational velocity $\vec{T}^C$ can thus be obtained through solving $\mathbf{A}\vec{T}^C = \vec{b}$. After $\vec{T}^C$ and $\vec{R}^C$ are determined, their corresponding velocities $\vec{T}^W$ and $\vec{R}^W$ in world coordinate are computed using Eq. (3). The affine transform matrices $\mathbf{M}^T$ and $\mathbf{M}^R$ describing the relationship between camera and world coordinates are estimated by deriving the spatial transformation between virtual camera (corresponding to the current camera coordinate) and the reference world coordinate. Finally, the camera position Υ and orientation Θ are determined and used to drive the virtual colonoscope camera.

5. Phantom validation

In order to evaluate our tracking algorithm, we constructed two colon-like phantoms, straight and curved, and performed extensive experiments to test the accuracy of the algorithm by comparison to ground truth. Specifically, the colonoscope's velocity and displacement were validated in these experiments.

The phantom design was driven by two goals, the ability (1) to accurately compare the estimated velocity and displacement with *measurable ground truth*, and (2) for the experiments to be *repeatable*, so that the results are statistically reliable. There have been some earlier work in computer vision on using graphical models [3] to generate phantom images. Such synthetic and 'clean' images are not a good fit, since colonoscopy images contain numerous artifacts, and such factors as fish-eye effects cannot be modeled or evaluated. In bronchoscopy tracking, bronchi-shaped phantoms have been used to produce phantom images. The tracking accuracy is measured either by comparing estimated motion parameters to measurements from magnetic sensor devices [46,14], or by visual inspection between optical and virtual bronchoscopy images [56,55,48]. Similarly, we used a curved phantom in our earlier work [40], using a hose with ridges, and artificial polyps glued along the inside wall. The main difficulty with these designs was repeatability, and to reliably and quantitatively measure motion parameters.

5.1. Straight phantom setup

LEGO bricks were used to build a straight-tunnel phantom (Fig. 7(a)) to take advantage of the many edges and corners for optical flow computation. LEGO bricks also facilitate accurate measurement of the displacement of the endoscope, as it passes each LEGO brick. The interior of the straight-tunnel phantom is 105 mm × 32 mm × 384 mm.

Fig. 7(a) illustrates the straight phantom setup. The straight phantom is driven by a motor at a constant speed, ensuring repeatability in the experiments. A straight iron wire is attached to a long wooden board to suspend the colonoscope in the tunnel. The phantom is placed under the board, with the colonoscope in the tunnel. A wooden box is fastened to a table by a clamp, and a steel rod is fixed inside this box, and one end of a fish wire is wound around the axis. The fish wire then passes through a small hole in the wooden box and the other end is connected to the straight phantom. A drill controlled by a power supply rotates the axis at a constant speed, pulling the phantom. An external video camera points to the straight phantom. The images acquired by the external video camera and colonoscope are recorded.

5.2. Curved phantom setup

Two concentric sheets (thick cardboard) of radii 158.5 mm and 102.5 mm were used to build a curved phantom (Fig. 7b). The height of each sheet is 125 mm. Textured (color squares) patterns coated the inside of the two curved sheets to simulate LEGO bricks, generating visual cues for optical flow computation. The size of each colored square is 54 mm × 28 mm.

Fig. 7(b) shows the setup of the curved phantom. Instead of translating the curved phantom, a small wheel of radius 0.6 mm is attached to the end of the drill and used to rotate the turntable. Based on this speed reduction, the colonoscope can move as slow as 10 mm/s, while the drill rotates at high speeds.

5.3. Data collection

In the straight phantom experiments, image sequences were collected at speeds of 10, 15 and 20 mm/s, which are typical speeds used during a colonoscopy procedure. Twenty-five trials were conducted at each speed. Of these, five sequences were selected such that (1) the phantom's displacement divided by the total displacement time approximates the desired speed, within a margin of 2 mm/s, (2) the total displacement time between the five trials is within a margin of 0.3 s. In all, 15 exterior and interior straight phantom image sequences were collected at 10, 15 and 20 mm/s.

Similar to the straight phantom experiments, five exterior and five interior curved phantom image sequences were acquired at each of the three speeds.

5.4. Validation results

Exterior phantom image sequences were used to determine the actual colonoscope motion, and interior image sequences were used to estimate the colonoscope's motion by the proposed tracking algorithm. Thus, the accuracy of the tracking algorithm was analyzed by comparing the ground truth to the estimated colonoscope motion. All computations on the phantom image sequences and the subsequent clinical image sequences are carried out on a Linux machine with a four-core 2.5 GHz Intel Xeon CPU and 8 GB memory executing C/C++code.

5.4.1. Ground-truth motion determination

In each straight phantom sequence, 19 candidate locations at boundaries between LEGO bricks were selected for determining the ground-truth camera velocities. Assuming the motorized pullback maintained a constant speed between consecutive locations, the ground-truth velocities were calculated by dividing the distance traveled (16 mm of one brick length) by the elapsed time. The ground-truth colonoscope displacements were then computed by integrating the product between the ground-truth velocities and the elapsed time.

A similar strategy was used to measure the ground-truth camera motion in the curved phantom except that a checkerboard pattern wrapped on the outer walls were used in place of LEGO bricks. Each block is 29 mm × 19 mm, as seen in Fig. 7b. Assuming constant motor speed inside the square, the colonoscope's instantaneous speed was determined by dividing the square length

(29 mm) over the elapsed time. Colonoscope displacements were then determined by integrating the product between the ground-truth velocities and the elapsed time.

5.4.2. Tracking results

Exterior phantom images were employed to determine the ground-truth camera motion by using the approaches described above. Our tracking algorithm was performed on the corresponding interior video streams to estimate camera motion parameters. We also analyzed the effect of the distorted interior images caused by the fish-eye effect of the colonoscope camera. The Matlab Toolbox [6] was used to calibrate the camera and remove the distortion from the optical colonoscopy images. The comparison of the tracking results on the same colonoscopy image sequences, with and without camera calibration helps us to understand the influence of the fish-eye effect on our tracking algorithm.

Fig. 8(a) illustrates the camera velocity plots on five straight phantom trials at a speed of 10 mm/s; the solid curve in the middle extents of these plots represents the average velocity of the five trials, and the upper and lower curves denote the maximum and minimum velocities across the five trials. The red plot represents the ground-truth velocities, while the green and blue plots represent estimated camera velocity curves on the original and calibrated image sequences (five each). At 10 mm/s, the velocity error range is (0.01–7.2) mm/s on the original phantom image sequences after 750 images, and (0.01–7.4) mm/s on the calibrated sequences in the straight phantom. Fig. 8(b) shows the camera displacement plots. The displacement error range is (0.0–14.65) mm on the original sequences and (0.0–14.57) mm on the calibrated sequences.

Fig. 9(a) shows the camera velocity curves on five curved phantom trials at a speed of 10 mm/s. The error range of camera velocity is (0–6.08) mm/s on the original sequences and (0–7.42) mm/s on the calibrated sequences. Fig. 9(b) shows the corresponding camera displacement plots. The range of displacement error is (0–4.22) mm on the original sequences, and it is (0–10.2) mm on the calibrated sequences. Velocity and displacement results at speeds of 15 mm/s and 20 mm/s are similar and consistent with those at 10 mm/s in both straight and curved phantoms. Further details may be found in our recent work [39].

Validation experiments on straight and curved phantoms generated the following findings.

1. Average velocity error is under 3 mm/s on the original and calibrated phantom image sequences at speeds of 10, 15 and 20 mm/s in both straight and curved phantom experiments. Average displacement error is less than 7 mm over a total displacement of 287–288 mm in the straight and curved phantoms.
2. Our tracking algorithm is robust to fish-eye effects of colonoscopy cameras, as tracking accuracy is comparable between the original and calibrated phantom image sequences in both straight and curved phantom experiments.
3. We have also validated our tracking algorithm on the phantom image sequences at speeds of 15 mm/s and 20 mm/s. Based on our observation, the number of tracked colonoscopy images is the primary reason for the camera displacement errors, while the actual speed of the colonoscope has a smaller role because the tracking errors are reducing with the increase of camera velocities.

6. Clinical data evaluation

We randomly selected 20 patients from the WRAMC virtual colonoscopy training data archive of the National Cancer Institute to evaluate the proposed tracking algorithm on optical colonoscopy sequences. Each patient underwent OC and VC examination, and OC

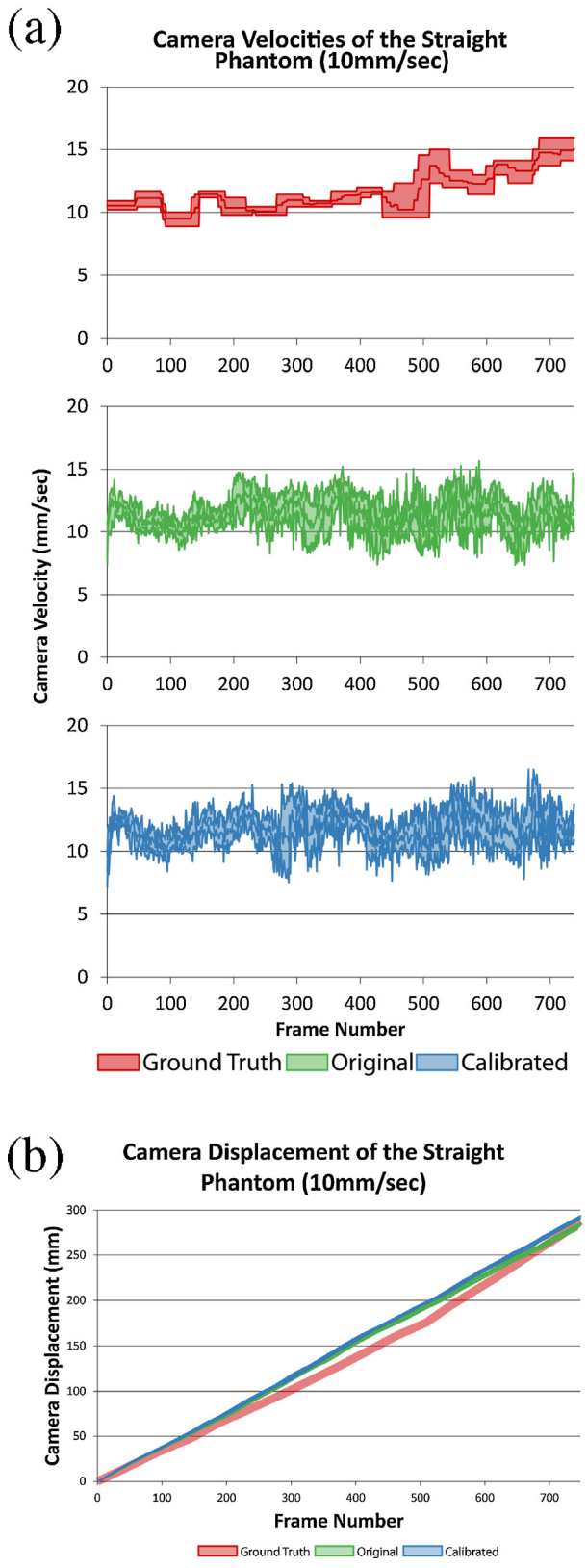


Fig. 8. Velocity and displacement plots of the endoscope camera at 10 mm/s in the straight phantom: (a) camera velocity plots in the straight phantom and (b) displacement plots in the straight phantom. (For interpretation of colour in the artwork, the reader is referred to the web version of the article.)

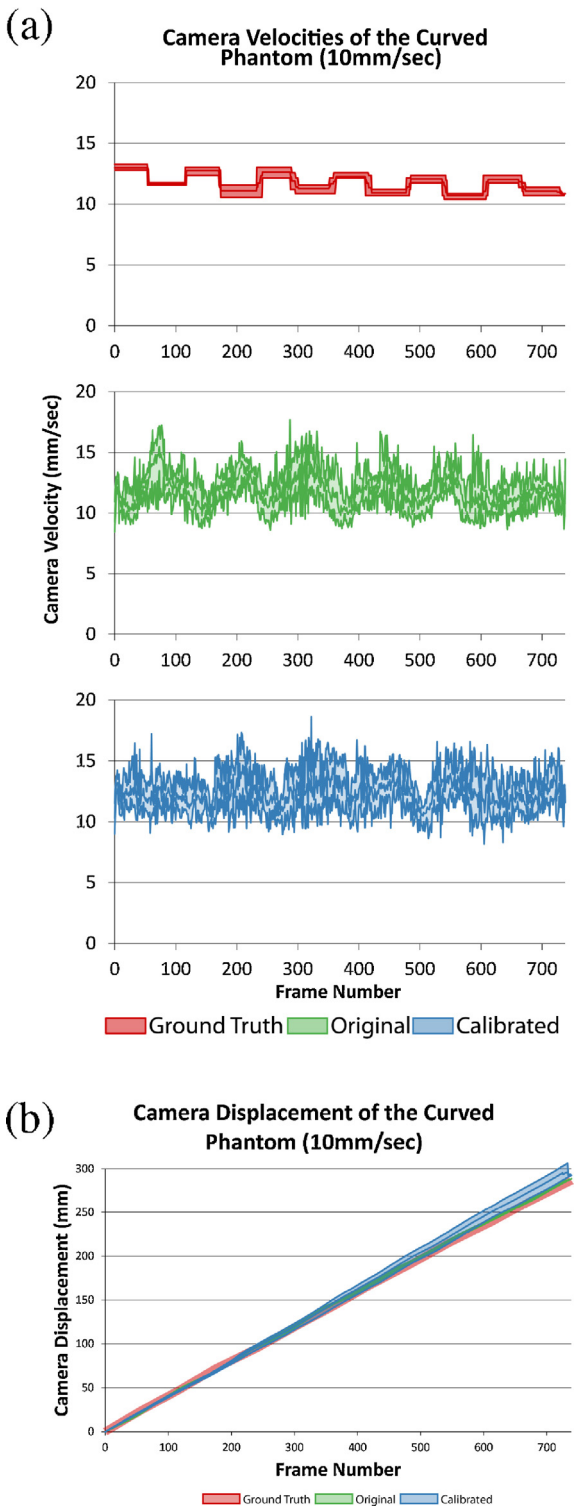


Fig. 9. (a and b) Velocity and displacement plots of the endoscope camera at 10 mm/s in the curved phantom.

and VC reports recorded the polyp information, including polyp size and location. As can be seen from Table 1, 15 of the patients had at least one polyp; distribution of polyps among them include 5 polyps in the rectum, 7 in the sigmoid colon, 8 in the ascending colon, 4 in the descending colon, and 2 in the transverse colon. Twenty-seven sequences from these 20 patients were extracted to analyze our tracking algorithm. They include 10 ascending colon sequences, 5 transverse colon sequences, 7 descending colon sequences,

Table 1
Clinical data evaluation. Twenty-seven colonoscopy image sequences from 20 patients were used to evaluate the tracking algorithm, containing up to 4 polyps. Sequences ranged from 240 to 3630 optical images.

Patient	# Colon segments	# Tracked frames	# Polyps
1	Rectum	3630	3
2(a)	Ascending	1317	2
2(b)	Descending	1890	2
3	Sigmoid	797	2
4	Descending	933	4
5	Transverse	1309	1
6(a)	Ascending	432	1
6(b)	Transverse	875	1
6(c)	Rectum	800	1
7	Descending	1362	0
8	Sigmoid	2018	1
9(a)	Descending	322	0
9(b)	Descending	400	0
9(c)	Ascending	933	0
10	Descending	716	0
11	Sigmoid	1443	3
12	Ascending	1044	1
13	Descending	770	0
14(a)	Transverse	240	0
14(b)	Sigmoid	360	0
14(c)	Ascending	435	0
15	Ascending	1101	4
16	Ascending	2234	3
17	Transverse	882	1
18	Rectum	3528	1
19	Ascending	1369	1
20	Transverse	250	0

4 sigmoid colon sequences, and 3 rectum sequences. These sequences had anywhere from 240 to 3630 images and were successfully tracked by our method. Table 1 details the data characteristics.

It can be seen from Table 1 that the number of tracked images is related to the existence of polyps in the dataset (datasets 9, 14, 20 vs. 3, 4, 5); our current algorithm relies on good initialization and the polyp locations in VC (marked by experts) make it easier and more accurate to align OC and VC images at the first frame. In the absence of polyps or other clear landmarks, we need to rely on fold similarity between OC and VC images, which is prone to error since the colon's tubular structure and fold shapes can be misleading. In addition, deformation is generally more pronounced in the sigmoid and transverse colon, which makes tracking a significantly more challenging problem in these regions. Sharp turns taken by the colonoscope appear to flatten the colon in these regions. In the dataset archive that was selected here, few polyps exist in the transverse colon (2 out of 27). Hence the colonoscope is withdrawn faster in these regions (datasets 14(a) and 20). Future work will need to focus on more accurate and automatic methods to initialize the tracking system.

Besides initialization, colonoscopy images exhibit a number of artifacts that pose significant challenges to any tracking system. These include recording interruptions, deformation and structural changes due to patient movement, or simply from the fact that OC and VC are separate acquisitions over time. Image artifacts include specularities (extremely bright regions), blurriness due to the endoscope facing a wall or very fast motion. Pre- and post-surgery images are another instance we will consider. We illustrate the robustness of our tracking algorithm under many of these conditions in the following sections,

6.1. Recording interruption

A sequence of 800 OC images containing a polyp in the rectal segment was used for evaluation. Fig. 10 shows 5 frames from this sequence. The left column illustrates the OC images and right

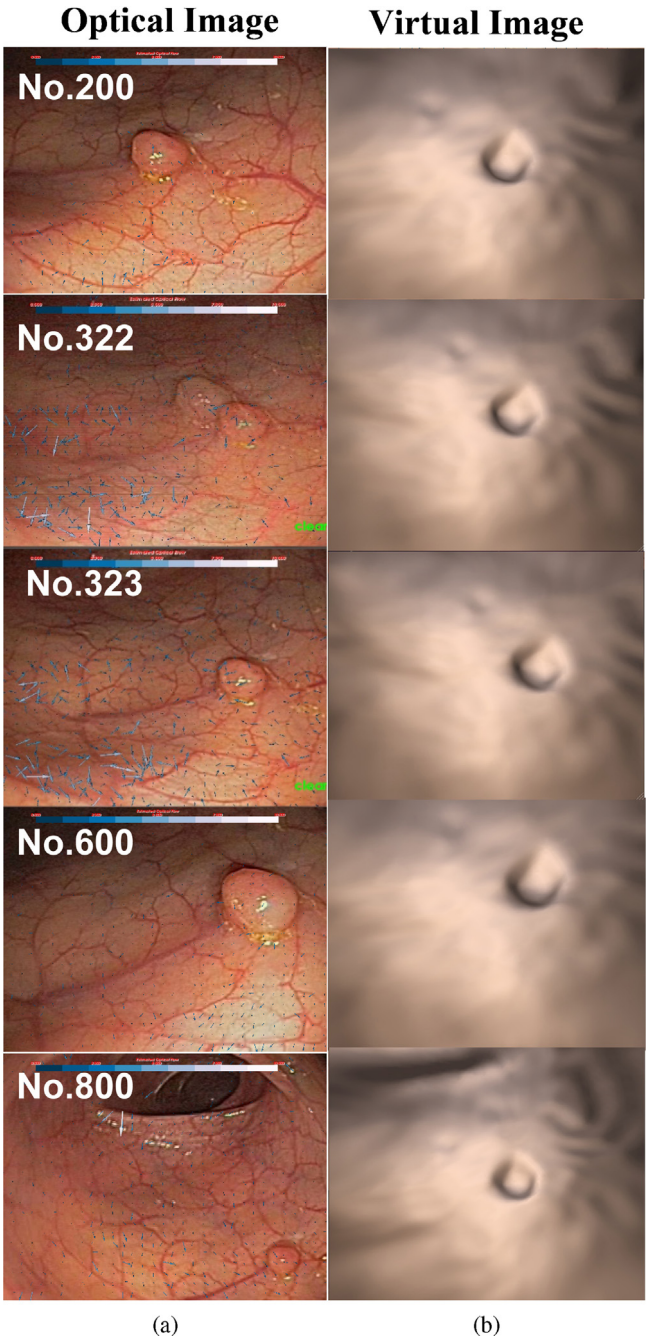


Fig. 10. (a and b) Robustness evaluation: fast camera motion and recording interruption on a rectal colonoscopy image sequence. Column 1: OC images; column 2: tracked VC images. The polyp is always tracked even if colonoscope moves significantly or video recording is interrupted.

column shows the tracked VC images. From frame 1 to 200, the colonoscope moves toward the polyp and slightly rotates around it. OC image is tracked well. From frame 200 to frame 400, the colonoscope is initially stationary due to a recording interruption. After 44 still frames, recording is suddenly resumed (second and third rows). Then the colonoscope slightly rotates and leaves the polyp. The VC image shows that the motion of OC is also well tracked. After frame 400, the colonoscope moves closer to the polyp and rotates around it again. The tracked VC image (fourth row) shows that the VC camera also moves towards the polyp. From frame 600 to 800, the colonoscope retreats from the polyp and has a significant rotation. As our system is currently unable to handle large

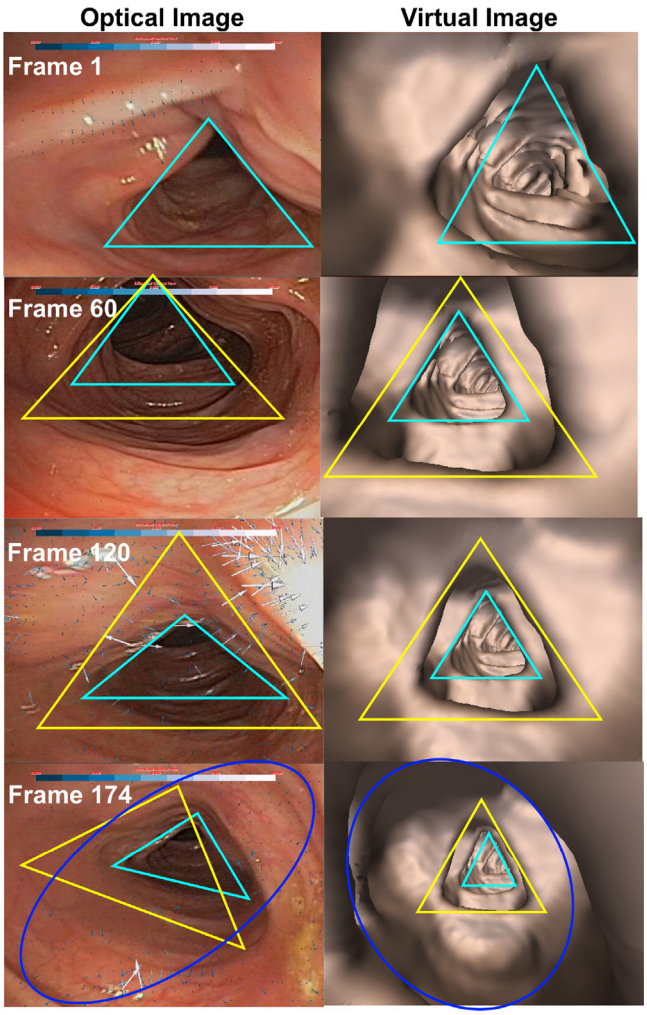


Fig. 11. Robustness evaluation: colon deformation on a transverse colonoscopy sequence. Row 1: first fold marked with a cyan triangle in both OC and VC images; row 2: colonoscope at second fold (yellow triangle) in both OC and VC. Fold shape in OC is elliptical, but triangular in VC; row 3: fold in OC becomes triangular with colonoscope near the colon's wall and VC still stays near the second fold; row 4: both VC and OC arrive at the third fold, marked by blue ellipse. (For interpretation of colour in the artwork, the reader is referred to the web version of the article.)

changes of motion (this happens at frame 750, and especially with rotation), the tracking is stopped (the tracking camera is frozen) as the changes exceed set thresholds (currently, 10 mm for translation and 6° for rotation). Exceeding these thresholds is considered a tracking failure in the current system. Tracking recovery will be investigated in future work on the system. Notice that a fold appears in the top OC image, while it partially shows in the corresponding VC image. The tracking error gradually increases, but the polyp is always tracked.

6.2. Colon deformation

A sequence of 174 OC images were acquired in the transverse colon to illustrate and evaluate the impact of deformation. Fig. 11 shows the OC (left column) and VC (right column) images of four frames from this sequence. VC and OC images are manually aligned at the first fold marked with a cyan triangle (row 1). Row 2 indicates the colonoscope arriving at the second fold, marked by the yellow triangle; the fold becomes *fat* and is elliptical in the OC image, while it is still triangular in the VC image. In row 3, OC and VC images show the fold reverting to a triangular shape. The images in the fourth row show the colonoscope at the third fold, labeled by the blue

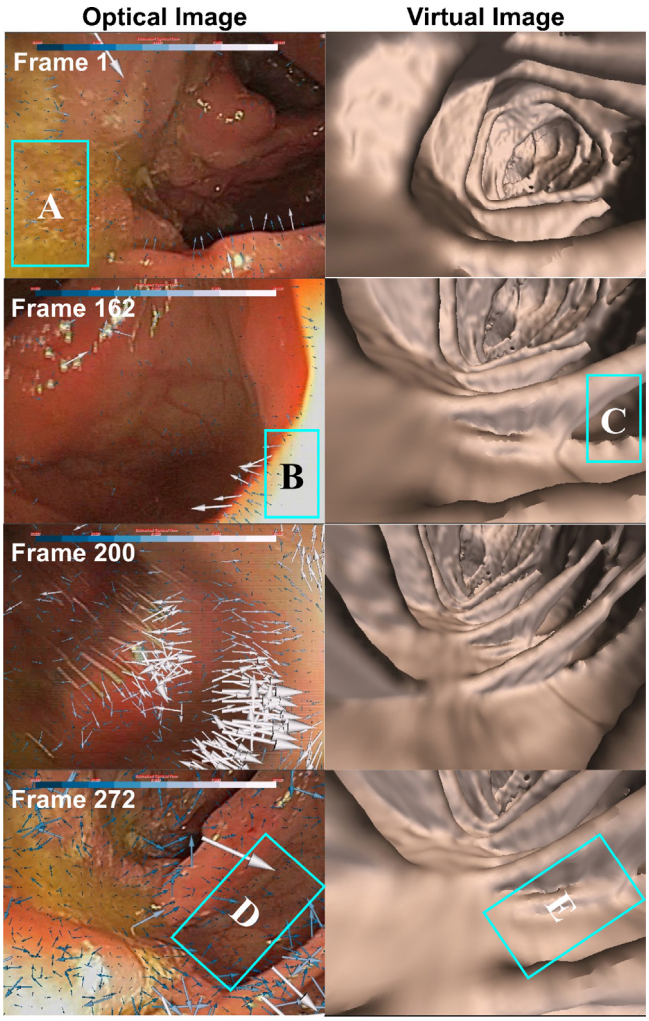


Fig. 12. Robustness evaluation: fluid and illumination artifacts, blurry images on an ascending image sequence. It contains fluid presence (area marked A), illumination band (area marked B), segmentation error results in artificial hole (area marked C), blurry image (row 3, left column). At the end of the sequence, images are tracked well, as shown by corresponding areas D and E in the OC and VC images of column 4. (For interpretation of colour in the artwork, the reader is referred to the web version of the article.)

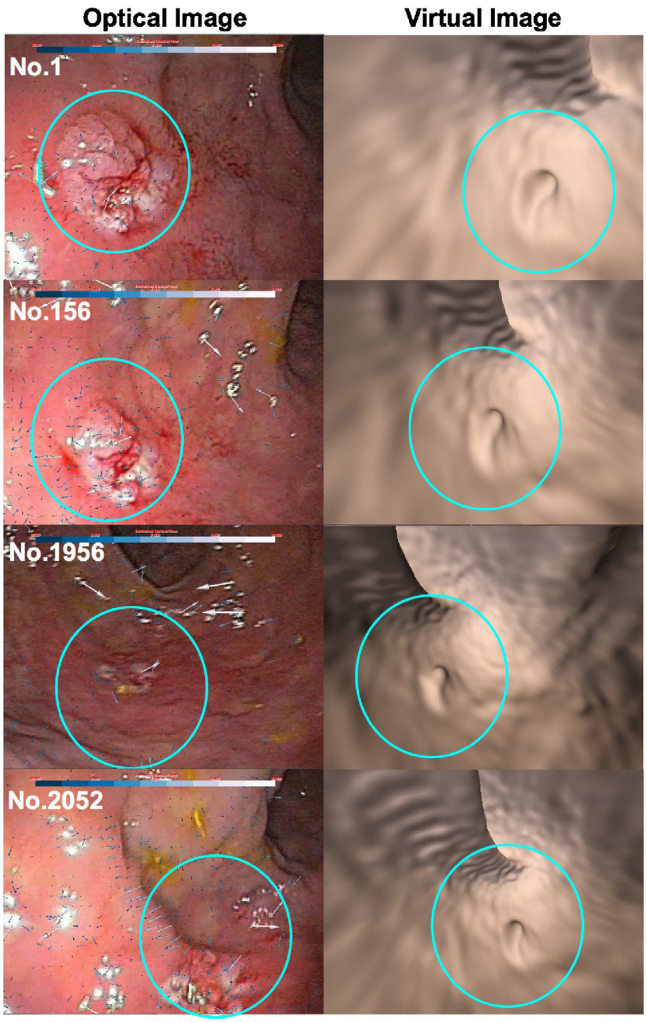


Fig. 13. Robustness evaluation: surgery-related structural changes. Illustration of tracking an image sequence containing pre- and post-polyp removal in the rectum colon. The top two rows show the tracking results of two frames before the polyp is removed, while the bottom two rows show the results after polyp removal. Polyp positions are marked by the cyan circle. (For interpretation of colour in the artwork, the reader is referred to the web version of the article.)

ellipse. Here, the orientation of the marked triangle and ellipse do not necessarily represent camera rotation, since the deformation also contributes to the change in shape of the fold. Although this case is difficult to evaluate, the results demonstrate that our algorithm is not very sensitive to the fold or other structural changes in the colon, since the tracking system is able to keep the OC and VC images in sync, *i.e.*, they reach the same fold.

6.3. Fluid and illumination artifacts, blurriness

A sequence of 272 OC images were captured between two folds in the ascending colon. This sequence contained images with fluid and illumination artifacts, as well as blurry frames. These artifacts are not present in the VC images, as they have been segmented out. Fig. 12 shows four frames from this sequence. Yellow fluid (region marked A in row 1), a strong illumination band (colon moving close to the wall, region marked B in row 2), and blurriness (colon moving fast, region marked C in row 3) are some of the difficulties that face the tracking system, while the corresponding VC images are devoid of these artifacts. Despite these artifacts, it can be seen in row 4 that the colonoscope is close to the second fold (area marked D), and in sync with the VC image (area marked E). Also note the artificial hole in the VC image of row 2 (area marked C), a segmentation error. But it does not influence our tracking results. This is important since perfect segmentations are almost never achievable [19].

6.4. Surgery induced structural changes

Optical colonoscopy is both a screening and treatment procedure; removal of polyps changes the structure of the colon. Fig. 13 illustrates an example. The left column OC image shows the circled polyp, while the OC image in row 4 shows the area where the polyp was removed. It is important that a tracking algorithm continue to perform under these conditions. In this example, we have selected an image sequence acquired before and after the removal of the polyp in the rectal colon. The top 2 rows of images in Fig. 13 show the tracking results of two images of the sequence before the polyp removal, while the bottom two rows illustrate the results after the polyp removal. The positions of the polyp are marked in both OC and VC images with cyan circles. It can be seen that the tracking algorithm continues to function despite these structural changes induced by surgery.

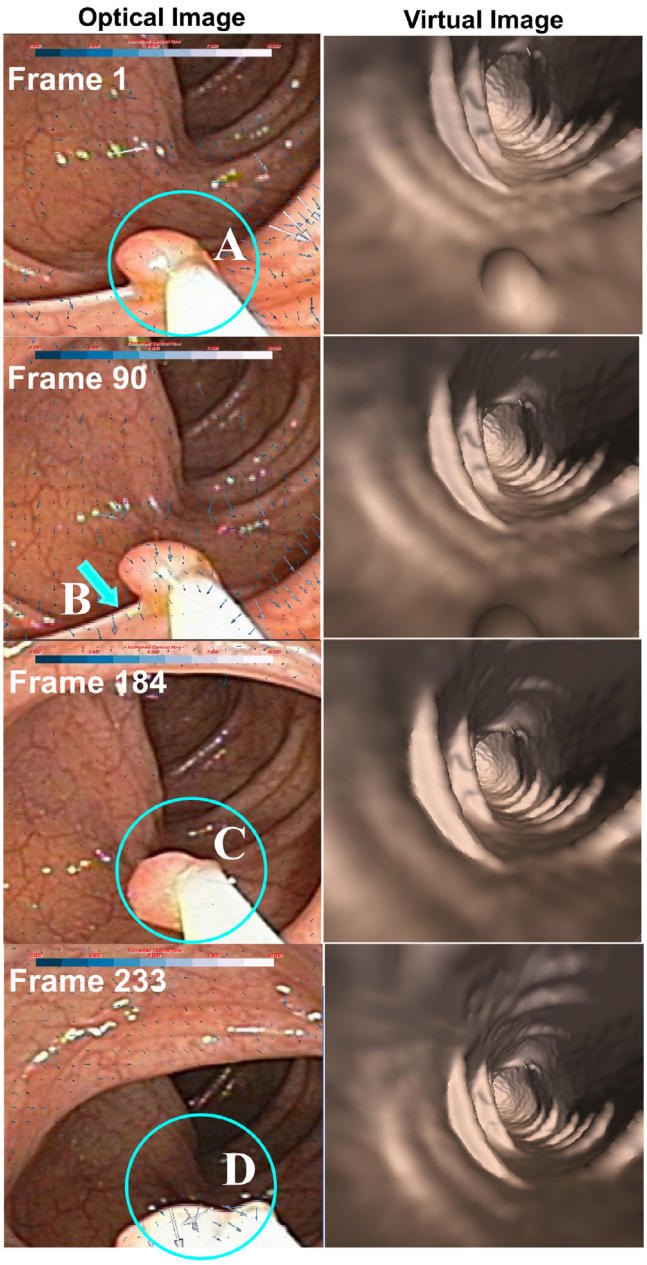


Fig. 14. Robustness evaluation: simultaneous motion of colonoscope and surgical tools in flow field. An image sequence in the descending colon involving simultaneous motion of colonoscope and surgical tools, breaking the basic assumption of egomotion computation. Row 1: snare inserted into colon to circle polyp, marked A; row 2: polyp is lifted up for removal, VC image continues to track motion; row 3: polyp is in the OC, but disappears in the VC image due to colonoscope's motion; row 4: withdrawal of snare and polyp, VC image continues to track the optical image.

6.5. *Multi-object motion induced by surgical tools*

During intervention, surgical tools will appear in the optical images. An example is illustrated in the left column images of Fig. 14. Both the colonoscope and the tool are simultaneously influencing the motion field captured by the optical flow. Theoretically, this breaks the condition of egomotion determination. However, if the affected region is relatively small and localized, then the optical colonoscope can perhaps be successfully tracked. We attempted to test this with a sequence of optical images captured in the descending colon. As illustrated in Fig. 14 (top row), a snare is inserted into the colon to enclose the polyp, shown in the area marked A. VC image is initially aligned based on the polyp's position. Since the tissue near the polyp (area indicated by the arrow B) is stretched, the gastroenterologist has to lift the polyp in order to remove it. The polyp in the tracked VC image is partly hidden because the egomotion in the OC image is translating along the $-Y$ direction. In the area marked C (third row), the polyp is removed and attached by the snare, and the polyp disappears in the tracked VC of the third column. However, the VC still follows the actual egomotion of the OC, while the polyp continues to stay in the OC image. The images in the fourth row show both the tool and the polyp withdrawn from the colon, however, this does not affect the tracking and VC continues to follow the egomotion of the optical sequence.

7. Conclusions and future work

We have presented an optical flow based tracking algorithm to co-align optical and virtual colonoscopy images. Based on our discussions with medical practitioners, such a system can be a useful navigation aid during endoscopic procedures. For instance, a polyp that is visible within the virtual images can provide spatial context or better localization during an endoscopic procedure. However, formal studies of the use of the system in the operating room are needed for a full evaluation of the costs and benefits.

We chose optical flow as the means to estimate object motion in colonoscopy images since there are few geometric or topological cues in the data. Second, the motion between successive frames is usually quite small; thus, region based matching techniques lack the needed accuracy and are inappropriate. Additionally, optical flow is less sensitive to lighting and small changes in shape.

Our optical flow based method uses a combination of sparse and dense optical flow and the use of the FOE in achieving a robust and stable tracking algorithm. Optimal spatial–temporal scales are determined for each image during the sparse flow computation procedure, which are used to compute the dense flow field. The dense flow field is employed to compute the FOE, utilizing the full motion information in the field. The FOE permits separation of the rotational and translational velocities contributing to the mathematical robustness of the algorithm. Motion parameters are estimated using the sparse flow field and the FOE.

We have performed extensive experiments, on (1) 30 straight and curved phantom image sequences at speeds of 10, 15 and 20 mm/s over a distance of about 288 mm, for comparison to ground truth, and (2) 27 clinical colonoscopy image sequences from 20 patients. As shown in Figs. 8 and 9, average estimated velocity error is less than 3 mm/s on the original and calibrated phantom image sequences at the three speeds for both phantoms. Average displacement error is less than 7 mm over a total translated distance of about 287–288 mm, for both phantoms. Also, it was seen that the tracking algorithm is not very sensitive to camera calibration, a highly desirable result. Specific challenges posed by optical colonoscopy data include recording interruption, the presence of fluid, illumination and blurred images, deformation of colon tissue due to both patient position changes between the virtual and optical images, as well as changes that can occur during surgery. Through five image sequences (from five different patients), we showed the robustness of our tracking algorithm under these conditions.

There are several issues that need to be addressed before this method can see application in clinical practice. First, our current implementation is not real-time, requiring about 2.1 s to track each frame, primarily due to multi-scale optical flow computation. Bruhn [9,10] has developed a multi-grid strategy to achieve real-time optical flow computation. We are currently investigating this strategy to improve multi-scale optical flow computation. Second, large motion displacements between successive frames cannot be accurately estimated, as the temporal derivatives are

based on finite differences. In colonoscopy images, this happens when the endoscope touches the colon wall or a polyp, followed by a rapid withdrawal or a rotation to move away from the wall. Thus the images bridging a blurry image sequence will exhibit significant changes in motion. In this case, region-based methods such as SIFT [42] or affine-invariant interest point detector [45] might be more appropriate. We recently have developed region flow [37] and temporal volume flow [38] to address the issue of large image motion and obtained some encouraging results, reported elsewhere. Third, a strategy is needed to reset and align the camera with the virtual images when there is significant change or deformation in the colon structure; colon folds can stretch or contract, S turns in the VC images can straighten out or disappear during the procedure. Finally, automatic initialization of the tracking system (starting images of OC and VC must be accurate and close) and re-initialization of the system with large drift errors are necessary for successfully tracking an entire colonoscopy procedure. Temporal volume flow [38] is a possible approach to reinitialize the tracking system. It can search for an OC image pair with similar visual features. The camera motion parameters of one OC image with large tracking errors can be re-computed by exploiting the other OC image with less errors. The drift errors can be reduced by using the newly computed camera motion parameters.

Acknowledgments

The authors would like to thank the National Cancer Institute for the Walter Reed Army Medical Center colonoscopy training datasets and the National Library Medicine for the Medical Informatics Training Program in 2006 and 2007.

Appendix A. Multi-scale optical flow computation

Algorithm 1. Multi-scale sparse optical flow computation

Input: Video stream $I(x, y, t)$.
Output: Sparse optical flow vectors (u_x, u_y) and characteristic spatial-temporal scales l_s and l_t.
Initialize $l_s = 0.5$ *and* $l_t = 0.3$;
while $(u_x, u_y, l_s, l_t) \notin \arg\min_{u_x,u_y,l_s,l_t} N(x, y, t; l_s, l_t)$ **do**
  Construct anisotropic scale space at spatial-temporal scales l_s and l_t (as per Eq."(19)";
  Select interest points with the $C(x, y, t; l_s, l_t)$ (as per Eq."(23)";
  Estimate $\vec{u}$ (as per Eq."(25)";
  $l_s \leftarrow \sqrt{2.0} l_s, l_t \leftarrow \sqrt{2.0} l_t$.;
end

References

[1] Adiv G. Determining three-dimensional motion and structure from optical flow generated by several moving objects. IEEE Trans Pattern Anal Mach Intell 1985;7(4):384–401.
[2] Adiv G. Inherent ambiguities in recovering 3D motion and structure from a noisy flow field. IEEE Trans Pattern Anal Mach Intell 1989;11(5):477–89.
[3] Aodha OM, Brostow GJ, Pollefeys M. Segmenting video into classes of algorithm-suitability. In: Proceedings of IEEE conference on computer vision and pattern recognition. 2010. p. 1054–61.
[4] Baxter N, Goldwasser M, Paszat L, Saskin R, Urbach D, Rabeneck L. Association of colonoscopy and death from colorectal cancer. Ann Intern Med 2009;150:1–8.
[5] Black, M.J., 1992. Robust incremental optical flow. PhD thesis, Yale University.
[6] Bouguet J-Y. Camera calibration toolbox for Matlab; 2010 www.vision.caltech.edu/bouguetj/calib_doc/index.html
[7] Bricault I, Ferretti G, Cinquin P. Multi-level strategy for computer-assisted transbronchial biopsy. In: Proceedings of 1th international conference on medical image computing and computer-assisted intervention. 1998. p. 161–268.
[8] Brox T, Bruhn A, Papenberg N, Weickert J. High accuracy optical flow estimation based on a theory for warping. In: Proceedings of 8th European conference on computer vision, vol. 4. 2004. p. 25–36.
[9] Bruhn A, Weickert J. Towards ultimate motion estimation: combining highest accuracy with real-time performance. In: Proceedings of IEEE international conference on computer vision. 2005. p. 749–55.
[10] Bruhn A, Weickert J, Kohlberger T, Schnšrr C. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. Int J Comput Vis 2006;70(3):257–77.
[11] Bruss AR, Horn BKP. Passive navigation. Comput Vis Graph Image Process 1983;21:3–20.
[12] Ching L, Moller K, Suthakorn J. Non-radiological colonoscope tracking image guided colonoscopy using commercially available electromagnetic tracking system. In: Proceedings of IEEE conference on robotics automation and mechatronics (RAM). 2010. p. 62–7.
[13] Deguchi D, Suenaga K, Hasegawa Y, Toriwaki J, Batake J, Natori HTH. New image similarity measure for bronchoscope tracking based on image registration. In: Proceedings of 6th international conference on medical image computing and computer-assisted intervention. 2003. p. 399–406.
[14] Deligianni F, Chung A, Yang GZ. Non-rigid 2D–3D registration with catheter tip EM tracking for patient specific bronchoscope simulation. In: Proceedings of 9th international conference on medical image computing and computer-assisted intervention. 2006. p. 281–8.
[15] Deligianni F, Chung A, Yang GZ. Non-rigid 2D/3D registration for patient specific bronchoscopy simulation with statistical shape modeling. IEEE Trans Med Imaging 2006;25(11):1462–71.
[16] Ferrucci JT. Colonoscopy: virtual and optical another look, another view. Radiology 2005;235:13–6.
[17] Fleet DJ, Jepson AD. Computation of component image velocity from local phase information. Int J Comput Vis 1990;5(1):77–104.
[18] Foley JD, van Dam A, Feiner SK, Hughes JF. Computer graphics: principles and practice in C. 2nd ed. Addison-Wesley Professional; 1995.
[19] Franaszek M, Summers RM, Pickhardt P, Choi J. Hybrid segmentation of colon filled with air and opacified fluid for CT colonography. IEEE Trans Med Imaging 2006;25(3):358–68.
[20] Harris C, Stephens MJ. A combined corner and edge detector. In: Proceedings of the Alvey vision conference. 1988. p. 147–52.
[21] Heeger D, Jepson A. Subspace methods for recovering rigid motion 1: algorithm and implementation. Int J Comput Vis 1992;7(2):95–117.
[22] Helferty JP, Higgins WE. Combined endoscopic video tracking and virtual 3D CT registration for surgical guidance. In: Proceedings of IEEE conference on image processing. 2002. p. 961–4.
[23] Higgins WE, Helferty JP, Lu K, Merritt SA, Rai L, Yu K-C. 3D CT-video fusion for image-guided bronchoscopy. Comput Med Imaging Graph 2007;32:159–73.
[24] Horn B, Schunck B. Determining optical flow. Artif Intell 1981;17(3):185–203.
[25] Johnson DA. CTC screening (virtual colonoscopy): is it virtually ready to replace optical colonoscopy? Medsc Gastroenterol 2008 http://www.medscape.com/viewarticle/580949
[26] Kadir T, Zisserman A, Brady M. An affine invariant salient region detector. In: Proceedings of the European conference on computer vision. 2004. p. 404–16.
[27] Kaufman AE, Lakare S, Kreeger K, Bitter I. Virtual colonoscopy. Commun ACM 2005;48(2):37–41.
[28] Kim DH, Pickhardt PJ, Taylor AJ, Leung WK, Winter TC, Hinshaw JL, Gopal DV, Reichelderfer M, Hsu RH, Pfau PR. CT colonography versus colonoscopy for the detection of advanced neoplasia. N Engl J Med 2007;357:1403–12.
[29] Koenderink J, Doorn AJV. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. Opt Acta 1975;22(9):773–91.
[30] Lai S-H, Vemuri BC. Reliable and efficient computation of optical flow. Int J Comput Vis 1998;29(2):87–105.
[31] Laptev I, Lindeberg T. Space–time interest points. In: Proceedings of the ninth IEEE international conference on computer vision. 2003. p. 432–9.
[32] Lieberman D. Quality and colonoscopy: a new imperative. Gastrointest Endosc 2005;61(3):392–4.
[33] Lim J, Barnes N. Estimation of the epipole using optical flow at antipodal points. Comput Vis Image Understand 2009;114(2):245–53.
[34] Lindeberg T. A scale selection principle for estimating image deformations. Image Vis Comput 1998;16(14):961–77.
[35] Lindeberg T, Garding J. Shape-adapted smoothing in estimation of 3D depth cues from affine distortions of local 2D structure. In: Proceedings of the 3rd European conference on computer vision. 1994. p. 389–400.
[36] Lindeberg T, Garding J. Shape-adapted smoothing in estimation of 3D shape cues from affine deformations of local 2D brightness structure. Image Vis Comput 1997;15(6):415–34.
[37] Liu J, Subramanian K, Yoo T. Region flow: a multi-stage method for colonoscopy tracking. In: Proceedings of MICCAI 2010. 2010. p. 505–13.
[38] Liu J, Subramanian K, Yoo T. Temporal volume flow: an approach to tracking failure recovery. In: Proceedings of SPIE medical imaging. 2011.
[39] Liu J, Subramanian K, Yoo T. A phantom design for validating colonoscopy tracking. In: Proceedings of SPIE medical imaging. 2012.
[40] Liu J, Subramanian K, Yoo T, Uitert R. A stable optic-flow based method for tracking colonoscopy images. In: Proceedings of mathematical methods in biomedical image analysis. 2008. p. 1–8.
[41] Longuet-Higgins H, Prazdny K. The interpretation of a moving retinal image. In: Proceedings of the Royal Society of London. 1980. p. 385–97.
[42] Lowe D. Distinctive image features from scale-invariant keypoints. Int J Comput Vis 2004;60(2):91–110.

[43] Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. In: Proceedings of international joint conference on artificial intelligence. 1981. p. 281–8.
[44] Matas J, Chum O, Urban M, Pajdla T. Robust wide baseline stereo from maximally stable extremal regions. In: Proceedings of the British machine vision conference. 2002. p. 384–93.
[45] Mikolajczyk K, Schmid C. Scale and affine invariant interest point detectors. Int J Comput Vis 2004;60(1):63–86.
[46] Mori K, Deguchi D, Akiyama K, Kitasaka Jr T, Suenaga CRM, Takabatake Y, Mori H, Natori MH. Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: Proceedings of 8th international conference on medical image computing and computer-assisted intervention. 2005. p. 543–55.
[47] Mori K, Deguchi D, Sugiyama J, Suenaga Y, Toriwaki Jr J, Takabatake CM, Natori HH. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. Med Image Anal 2002;6(3):321–36.
[48] Nagao J, Mori K, Enjouji T, Deguchi D. Fast and accurate bronchoscope tracking using image registration and motion prediction. In: Proceedings of 7th international conference on medical image computing and computer-assisted intervention. 2004. p. 551–8.
[49] Nagel H-H. Constraints for the estimation of displacement vector fields from image sequences. In: Proceedings of international joint conference on artificial intelligence. 1983. p. 945–51.
[50] Nagel H-H. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. In: Proceedings of the first European conference on computer vision. 1990. p. 139–48.
[51] Nain D, Haker S, Grimson Jr W, Wells EC, Ji WM, Kikinis H, Westin RCF. Intrapatient prone to supine colon registration for synchronized colonoscopy. In: Proceedings of 5th international conference on medical image computing and computer-assisted intervention. 2002. p. 573–80.
[52] NCI. Colon and rectal cancer. National Cancer Institute; 2010 http://www.cancer.gov/cancertopics/types/colon-and-rectal
[53] Papenberg N, Bruhn A, Brox T, Didas S, Weickert J. Highly accurate optic flow computation with theoretically justified warping. Int J Comput Vis 2006;67(2):141–58.
[54] Pickhardt PJ, Hassan C, Halligan S, Marmo R. Colorectal cancer: CT colonography and colonoscopy for detection systematic review and meta-analysis. Radiology 2011;259:393–405.
[55] Rai L, Helferty J, Higgins W. Combined video tracking and image-video registration for continuous bronchoscopic guidance. Int J Comput Assisted Radiol Surg 2008;3(3–4):315–29.
[56] Rai L, Merritt SA, Higgins WE. Real-time image-based guidance method for lung-cancer assessment. In: Proceedings of IEEE conference on computer vision and pattern recognition. 2006. p. 2437–44.
[57] Reiger J, Lawton D. Processing differential image motion. J Opt Soc Am A 1985;2(2):354–9.
[58] Singh A. An estimation-theoretic framework for image-flow computation. In: Proceedings of the third IEEE international conference on computer vision. 1990. p. 168–77.
[59] Stiller C, Konrad J. Estimating motion in image sequences. IEEE Signal Process Mag 1999;16:70–91.
[60] Sundareswaran V. Egomotion from global flow field data. In: Proceedings of the IEEE workshop on visual motion. 1991. p. 140–5.
[61] Tian T, Tomasi C, Heeger D. Comparison of approaches to egomotion computation. In: Proceedings of IEEE conference on computer vision and pattern recognition. 1996. p. 315–20.
[62] Tuytelaars T, Gool LV. Matching widely separated views based on affine invariant regions. Int J Comput Vis 2004;59(1):61–85.
[63] Tuytelaars T, Mikolajczyk K. Local invariant feature detectors: a survey. 1st ed. Now Publishers Inc.; 2008.
[64] Yacoob Y, Davis L. Temporal multi-scale models for flow and acceleration. Int J Comput Vis 1999;32(2):147–63.