A Stable Optic-Flow Based Method for Tracking Colonoscopy Images

Jianfei Liu, Kalpathi Subramanian
Charlotte Visualization Center, Department of Computer Science, University of North Carolina at Charlotte
9201 University City Blvd, Charlotte, NC 28223
jliu1,krs@uncc.edu

Terry Yoo, Robert Van Uitert
National Library of Medicine, National Institutes of Health
8600 Rockville Pike, Bethesda, MD 20894
tyoo@mail.nih.gov, robert.vanuitert@gmail.com

Abstract

In this paper, we focus on the robustness and stability of our algorithm to plot the position of an endoscopic camera (during a colonoscopy procedure) on the corresponding pre-operative CT scan of the patient. The colon has few topological landmarks, in contrast to bronchoscopy images, where a number of registration algorithms have taken advantage of features such as anatomical marks or bifurcations. Our method estimates the camera motion from the optic-flow computed from the information contained in the video stream. Optic-flow computation is notoriously susceptible to errors in estimating the motion field. Our method relies on the following features to counter this, (1) we use a small but reliable set of feature points (sparse optic-flow field) to determine the spatio-temporal scale at which to perform optic-flow computation in each frame of the sequence, (2) the chosen scales are used to compute a more accurate dense optic flow field, which is used to compute qualitative parameters relating to the main motion direction, and (3) the sparse optic-flow field and the main motion parameters are then combined to estimate the camera parameters. A mathematical analysis of our algorithm is presented to illustrate the stability of our method, as well as comparison to existing motion estimation algorithms. We present preliminary results of using this algorithm on both a virtual colonoscopy image sequence, as well as a colon phantom image sequence.

1. Introduction

Colorectal cancer is a leading cause of cancer and cancer-related mortality in the United States[1]. The survival rate can be significantly improved through early cancer detection and treatment. *Optical Colonoscopy(OC)* is a minimally invasive screening and cancer detection tool, that involves guiding a long, flexible endoscope into the colon, allowing visual inspection and removal of inflamed tissue, abnormal growth (also known as a polyp), and ulcers. However, OC is an exploratory procedure and depends on the physician's skills and experience, and can miss polyps[23]. A newer technology, *Virtual Colonoscopy (VC)*[30, 21] is capable of providing interactive views of the interior of the colon for surgery planning and diagnosis. VC has limitations: lesions less than 5mm cannot be detected, and, currently, there is no means to track the conventional colonoscopy images and the pre-segmented virtual colonoscopy images. This is necessary for effective and convenient use of VC images during a colonoscopy procedure.

The past decade has seen a considerable body of work in registering optical and virtual bronchoscopy images [20, 18, 31, 10, 29, 11, 28, 12]. However, this task exploits anatomical marks, bifurcations and other structural features, as well as repeated 2D and 3D registrations to align the virtual and optical images. Unlike the bronchi of the lungs, the colon has no bifurcations or other anatomical features for navigation. As the goal of our work is to track the endoscopic camera so as to be in the vicinity of landmarks such as polyps or a particular fold, the requirements are somewhat more relaxed, unlike the registration algorithms typically used in bronchoscopy tracking to guide needle biopsy procedures. Given these considerations, *optic flow* based schemes are a reasonable approach in the absence of other visual cues.

Optic flow represents the distribution of apparent velocities of brightness patterns in an image[19], and is used to estimate the projected motion of the relative displacement between the camera and the objects. There are a large number of methods to compute optical flow, including differential techniques[19, 27, 36, 7], region-based matching[4, 25], and phase-based methods[13, 15]. Detailed reviews of optic flow computation can be found in [5, 6, 14]. Optic flow computation techniques have led to tracking algorithms, generally referred to as *egomotion* determination. Bruss[8] proposed a linear-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 optic flow fields, Adiv[2, 3] proposed a decomposition scheme to compute the motion parameters according to an estimated residual. Nevertheless, these methods are still sensitive to the accuracy of the underlying flow field. Matrix perturbation theory[33] can be used to illustrate the sensitivity of the estimation matrix used in computing motion parameters.

A number of researchers have used motion parallax to compute invariant properties of the flow field[22, 26], such as the *focus of expansion (FOE)*, in order to improve the robustness of the tracking algorithms. The focus of expansion is defined as the projection of the camera's translation axis on the image plane. Detection of the focus of expansion permits independent estimation of translation and rotation parameters, and the approach taken in [17, 32, 34].

In this work, we focus on the stability and robustness of our proposed algorithm[**?**]for tracking optical and virtual colonoscopy images. Fig. 1 illustrates the steps involved in our method. The key to our approach is to be able to compute an accurate flow field, *given the limitations and difficulties associated with colonoscopy images.* Specifically, our method has the following 3 features that results in a stable tracking algorithm:

1. **[Multi-Scale Approach:]** We choose a small set of robust feature points (corner points) to compute a sparse optic flow field; an iterative scheme and a scale selection metric is proposed to compute optimal spatial and temporal scales for each image frame.

2. **[Dense Flow Field and FOE Computation:]** The computed scales are used to determine the dense optic flow field, which in turn is used to obtain the *focus of expansion* using a subdivision based method; qualitative motion information relating to the main motion direction is extracted. We illustrate comparisons of our approach to the direct approach to computing motion parameters from the accurate sparse flow field. We demonstrate mathematically and experimentally its sensitivity to flow field errors.

3. **[Determining Camera Motion Parameters:]** The qualitative features derived from the dense flow field and the accurate sparse flow field (step 1) are used to estimate the camera motion parameters. We illustrate the robustness of our tracking algorithm on two example datasets.

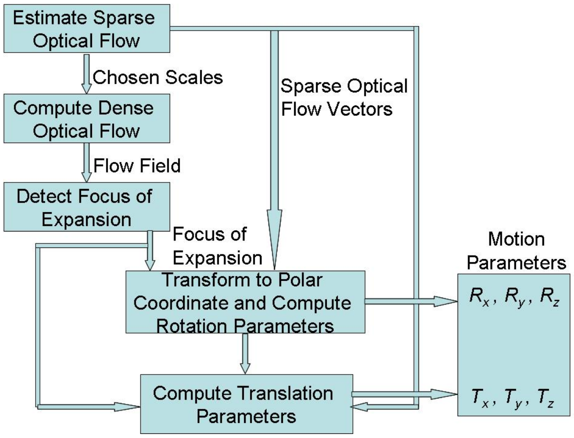

Estimate Sparse Optical Flow
Chosen Scales
Sparse Optical Flow Vectors
Compute Dense Optical Flow
Flow Field
Detect Focus of Expansion
Focus of Expansion
Motion Parameters
Transform to Polar Coordinate and Compute Rotation Parameters
R_x, R_y, R_z
Compute Translation Parameters
T_x, T_y, T_z

Figure 1. The colonoscopy tracking algorithm.

2. Methods

2.1. Optic-Flow Computation: A Multi-scale Approach

Our tracking algorithm begins by identifying a relatively small set of stable feature points and their corresponding optical flow, resulting in a *sparse* optic flow field. Moreover, the flow field is determined using a *multi-scale* approach.

Let (u_x, u_y) define the flow vector at an image pixel (x, y) at time t. Then

$$L(x,y,t) = L(x+u_x, y+u_y, t+1) \tag{1}$$

where $L(x, y, t)$ is the scale-space representation of the original image $I(x, y, t)$, σ, τ are the spatial and temporal scale parameters, and

$$\begin{aligned} L(x,y,t) &= g(x,y,t;\sigma^2,\tau^2) * I(x,y,t) \quad (2) \\ g(x,y,t;\sigma^2,\tau^2) &= \frac{e^{\left(\frac{-(x^2+y^2)}{2\sigma^2} - \frac{t^2}{2\tau^2}\right)}}{\sqrt{(2\pi)^3\sigma^4\tau^2}} \quad (3) \end{aligned}$$

Anisotropic Gaussian is applied to account for the differential sampling rates across the spatial and temporal dimensions. Unlike methods that consider spatial [24] or the temporal scale [37] individually, our approach is targeted at determining the optimal spatial and temporal scales for optic flow computation.

In order to reduce the ambiguities in corresponding pairs, *corner points* are chosen as feature candidates. Corner points are detected by the Harris matrix[16] defined as

$$\mu = \begin{bmatrix} L_x^2 & L_x L_y \\ L_x L_y & L_y^2 \end{bmatrix} \tag{4}$$

where

$$(L_x, L_y, L_t) = (\partial_x(L(x,y,t)), \partial_y(L(x,y,t)), \partial_t(L(x,y,t))) \tag{5}$$

We are interested in corresponding pairs that exhibit maximum variance in the spatial domain and minimum difference along the temporal direction. We propose the following scale-space metric to achieve this,

$$\Theta = \frac{\int\int_W G(x,y)|L(x,y,t) - L(x+u_x, y+u_y, t+1)|^2}{\int\int_W G(x,y)\sqrt{|det(\mu) - \alpha * Trace^2(\mu)|} + \beta} \qquad (6)$$

where α and β are constants. The measurement is performed within a window of size W to avoid the aperture problem[19], and $G(x,y)$ is its window function.

The numerator in Eq. 6 represents the similarity between corresponding pairs, while the denominator measures how distinct the selected features are in their local neighborhood. The smaller the response of Θ, the better the match. We use Eq. 6 as the basis for spatial and temporal scale selection. We argue that these *characteristic spatial and temporal scales* should also make Θ a local minimum in scale space. For implementation, the numerator can be converted into the iterative Lucas-Kanade algorithm [27] (Taylor series approximation), while the denominator is related to the Harris matrix, resulting in the following approximation to Eq. 6,

$$\Theta \approx \frac{\int\int_W G(x,y)|L_x u_x + L_y u_y + L_t|^2}{\int\int_W G(x,y)\sqrt{|det(\mu) - \alpha * Trace^2(\mu)|} + \beta} \qquad (7)$$

An image sequence from virtual colonoscopy was used to examine the effectiveness of the scale selection metric. Ground truth, consisting of the exact motion field (Eq. 9) and depth values were respectively calculated from the known motion of the virtual camera and the Z-buffer. Scale selection results are illustrated in Fig. 2 and Table 1. Fig. 2d shows a response curve plotted as a function of the two spatial and temporal 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 (red) and the estimated flow vectors(blue). The small green cubes indicate the positions of the chosen feature points. Fig. 2(a), corresponding to point A in 2(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 variance; in 2(c), which corresponds to point C in 2(d), small vectors diverge because the chosen scales are too coarse and small areas with varying motion are merged. Spatio-temporal scales at the local minima are a means to balance between these two extremes, and as seen in 2b (point B in 2(d)), generate flow vectors close to the groundtruth.

Table. 1 shows numerical results of the errors between the groundtruth and estimated flow vectors for various combinations of spatial and temporal scales. These scale values correspond to the points on the response curve in Fig. 2(d). 40 feature pairs were selected in this example. Their average, minimum, and maximum differences, in magnitude and direction between the estimated optical flow and ground truth, were calculated. In this example, the scale pair, (1.41, 0.71) or $(2.0, 1.0)$ are the best choices and are around the local minima (Fig. 2d).

(σ, τ)	Error	Error Measurements		
	Type	Average	Minimum	Maxium
(0.5, 0.25)	ε_{magn}	0.1033	0.0011	0.5521
	ε_{dir}	4.6374	0.1586	25.32
(0.71, 0.35)	ε_{magn}	0.081	0.0002	0.5792
	ε_{dir}	3.6221	0.04483	23.19
(1.0, 0.5)	ε_{magn}	0.0384	0.0004	0.1491
	ε_{dir}	2.3875	0.3323	6.7625
(1.41, 0.71)	ε_{magn}	0.0346	0.0022	0.1222
	ε_{dir}	1.449	0.0131	7.0472
(2.0, 1.0)	ε_{magn}	0.0584	0.0042	0.1487
	ε_{dir}	1.4337	0.0104	8.4133
(2.82, 1.41)	ε_{magn}	0.1266	0.062	0.1979
	ε_{dir}	1.598	0.0345	7.8617
(4.0, 2.0)	ε_{magn}	0.1947	0.1095	0.2939
	ε_{dir}	5.4366	0.1977	23.6694
(5.66, 2.83)	ε_{magn}	0.272	0.1699	0.5248
	ε_{dir}	8.0157	0.013	31.497
(8.0, 4.0)	ε_{magn}	0.385	0.266	0.682
	ε_{dir}	15.48	0.635	90.08
(11.3, 5.66)	ε_{magn}	0.7954	0.292	5.075
	ε_{dir}	33.55	0.822	138.4

Table 1. Comparison between the groundtruth and the estimated optic flow vectors, in terms of magnitude and direction. Error measurements are in units of relative magnitude and degrees(see Eqn. 8).

$$\begin{aligned} \varepsilon_{magnitude} &= \frac{\|\mathbf{u} - \mathbf{v}\|}{\|\mathbf{v}\|} \\ \varepsilon_{direction} &= \frac{|\mathbf{u} \cdot \mathbf{v}|}{\|\mathbf{u}\|\|\mathbf{v}\|} \end{aligned} \qquad (8)$$

where $\mathbf{u}$ and $\mathbf{v}$ are estimated and groundtruth flow vectors.

2.2. Dense Flow Field and FOE Computation:

Given the characteristic spatial and temporal scales, they are used to compute a dense optic flow field, which is more accurate than if a single scale was chosen throughout the image sequence. We use Horn's method [19] on the smoothed image sequence, using the optimal spatial and temporal smoothing parameters. Using the full flow field of the image also leads to a more robust algorithm to detect the focus of expansion. We use a subdivision based method[32] to detect the focus of expansion, which we describe next, preceded by a mathematical basis for our approach.

Fig. 3 shows the coordinate system of a moving camera. $\mathbf{T} = (T_x, T_y, T_z)$ and $\mathbf{R} = (\omega_x, \omega_y, \omega_z)$ represent the

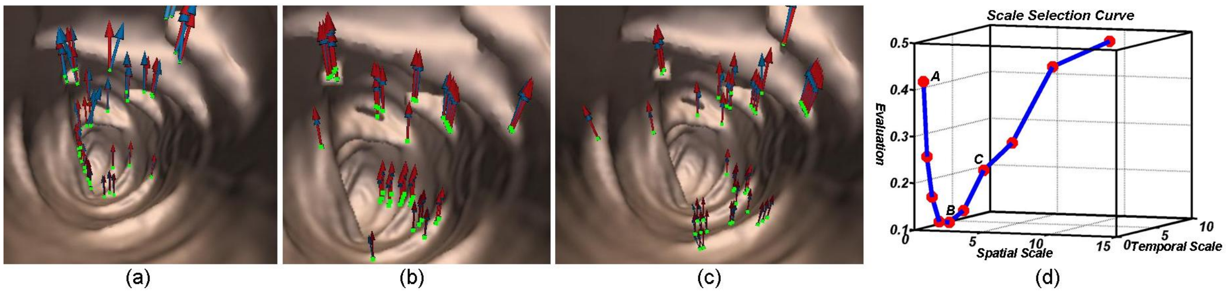


Figure 2. The relationship between spatio-temporal scale and the scale metric. Groundtruth flow vectors are in red and estimated flow vectors are in blue. Green cubes represent the selected feature point positions, (a) Results with relatively fine spatial and temporal scales, (b) Results with optimal spatial and temporal scales, (c) Result with relatively coarse scales, (d) The response curve between spatio-temporal scales and the scale metric; the scale values at points A, B and C correspond to images (a), (b), and (c) respectively.

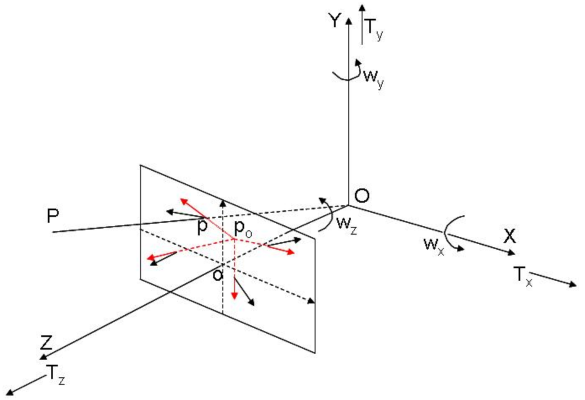


Figure 3. The motion coordinate system, where the camera is put at point O and its optical axis is along the Z-axis.

translation and rotation vectors along the spatial dimensions Let the instantaneous coordinates of an object point P in camera coordinates be (X, Y, Z) and its projection point p in the image plane be (x, y) Geometrically, its optical flow is [19, 26]

$$\begin{aligned} u_x &= \frac{-T_x f + T_z x}{Z} + \omega_x \frac{xy}{f} - \omega_y (f + \frac{x^2}{f}) + \omega_z y \\ u_y &= \frac{-T_y f + T_z y}{Z} + \omega_x (f + \frac{y^2}{f}) - \omega_y \frac{xy}{f} - \omega_z x \end{aligned} \tag{9}$$

The motion parameters can be directly estimated from these equations, and the approach taken by Bruss and Horn[8], approximating the motion parameters by the estimated optic flow. This leads to the following linear system

$$A\mathbf{x} = b \tag{10}$$

where A is a 6×6 system shown below, formed from integrating individual equations formed from each chosen feature candidate(using Eq. 9), $\mathbf{x}$ is the unknowns, $\mathbf{T} = (T_x, T_y, T_z)$, $\mathbf{R} = (\omega_x, \omega_y, \omega_z)$, and b are constants.

$$\begin{aligned} A &= \iint \begin{pmatrix} \frac{f}{Z} & 0 & -\frac{x}{Z} & -\frac{xy}{f} & (f + \frac{x^2}{f}) & -y \\ 0 & \frac{f}{Z} & -\frac{y}{Z} & -(f + \frac{y^2}{f}) & \frac{xy}{f} & x \\ \frac{xf}{Z} & \frac{yf}{Z} & -\frac{x^2+y^2}{Z} & -y(f + \frac{x^2+y^2}{f}) & x(f + \frac{x^2+y^2}{f}) & 0 \\ \frac{xy}{Z} & \frac{f^2+y^2}{Z} & -\frac{y}{Z}(\frac{x^2+y^2}{f} + f) & -((\frac{xy}{f})^2 + (f + \frac{y^2}{f})^2) & \frac{xy}{f}(2f + \frac{x^2+y^2}{f}) & fx \\ \frac{f^2+x^2}{Z} & \frac{xy}{Z} & -\frac{x}{Z}(\frac{x^2+y^2}{f} + f) & -\frac{xy}{f}(2f + \frac{x^2+y^2}{f}) & ((f + \frac{x^2}{f})^2 + (\frac{xy}{f})^2) & -yf \\ \frac{fy}{Z} & -\frac{fx}{Z} & 0 & fx & fy & -(x^2 + y^2) \end{pmatrix} \\ \mathbf{X} &= (T_x, T_y, T_z, \omega_x, \omega_y, \omega_z)^T \\ b &= \iint (-u_x, -u_y, -xu_x - yu_y, -(f + \frac{y^2}{f})u_y - \frac{xy}{f}u_x, -\frac{xy}{f}u_y - (f + \frac{x^2}{f})u_x, xu_y - yu_x)^T \end{aligned}$$

However, there are errors in the optical flow estimation that makes it impractical to directly solve this linear system. In Appendix A, perturbation theory is used to show the numerical problems and the resulting instabilities in using this approach. Fig. 4 shows the relationship between the absolute translation error of the first 150 images of a virtual colonoscopy image sequence. Notice the large translation errors along X and Y at the points marked **A, B**; these points represent numerical instabilities in the characteristics of the linear system and the chosen feature points. Refer to Appendix A for details.

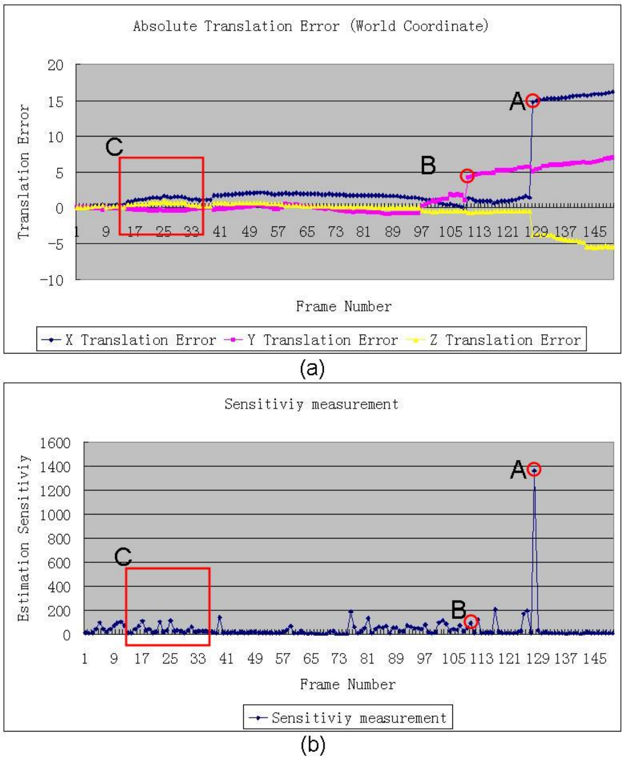
Thus, it is important to enforce some constraints on the motion parameters to reduce the sensitivity of estimation procedure. Note that we can split the camera's translation and rotation, rewriting Eq. 9 as

$$\begin{aligned} u_x &= u_x^T + u_x^R \\ u_y &= u_y^T + u_y^R \end{aligned} \tag{11}$$

where

$$\begin{aligned} u_x^T &= \frac{T_z}{Z}(x - \frac{fT_x}{T_z}) \\ u_y^T &= \frac{T_z}{Z}(y - \frac{fT_y}{T_z}) \\ u_x^R &= \omega_x \frac{xy}{f} - \omega_y (f + \frac{x^2}{f}) + \omega_z y \\ u_y^R &= \omega_x (f + \frac{y^2}{f}) - \omega_y \frac{xy}{f} - \omega_z x \end{aligned} \tag{12}$$

From Eq. 12, we note that the translation components of all optical flow vectors, intersect at the *focus of expansion* at $p_0 = (\frac{fT_x}{T_z}, \frac{fT_y}{T_z})$. This is depicted by the red vectors in Fig. 3. We can thus compute the two translation ratios



(a)

(b)

Figure 4. The relationship between absolute translation errors and the sensitivity measurement $\zeta = \frac{\mu}{|\lambda_1|}$ of the estimation system. (a) The absolute estimated translation errors of 150 frames of a CT colonoscopy sequence. (b) Corresponding sensitivity measurement .

$\frac{T_x}{T_z}, \frac{T_y}{T_z}$), given the focal length. We use this to estimate the rotation and translation parameters independently.

Longuet-Higgins [26] showed in theory that the vector joining two object points having different depth values and projecting on to the same pixel will point toward the *focus of expansion(FOE*. We use this property and a subdivision based method similar to [32], to detect the focus of expansion.

Fig. 5 shows the process of determining focus of expansion. The image plane is subdivided into rectangular regions (region lattice indicated by the dark dots in 5b and c) and flow vector differences between the region center and its neighbors are tabulated. A system of linear equations (the dense optic flow field is used) is formed, as given by Eq. 13.

$$\sum_W \begin{bmatrix} \Delta u_x^2 & \Delta u_x \Delta u_y \\ \Delta u_x \Delta u_y & \Delta u_y^2 \end{bmatrix} \tag{13}$$

where $(\Delta u_x, \Delta u_y) = (u_{x1} - u_{x2}, u_{y1} - u_{y2})$ represents the flow vector difference between the two image points. Eigen ratio of this matrix $\delta = \|\lambda_{small}/\lambda_{large}\|$ is computed, and thresholded. A line fitting algorithm is run on each region and the intersection of these lines completes the FOE calculation.

2.3. Determining Camera Motion Parameters

As stated earlier, the translation and rotation components of the camera are determined independently. Unlike most subdivision methods [32, 17, 34] that use the dense flow field, we use only the *sparse flow field* vectors, as they represent stable feature points (corner points). This further contributes to the accuracy of camera parameter estimation. There are two steps:

Rotation Parameters. All flow vectors are transformed into polar coordinates with the *focus of expansion* at the origin. This permits the translation component to be eliminated as follows:

$$u \cdot e_\perp = (u_T + u_R) \cdot e_\perp = u_R \cdot e_\perp \tag{14}$$

where $e_\perp$ is perpendicular to translation component (that joins the feature point to the focus of expansion).

Translation Parameters. The rotation parameters are next substituted into Eq. 12 to remove the rotation components from the optical flow vector for each selected feature. The following formulation is obtained,

$$Z/T_z = \|d\|/\|u_T\| \tag{15}$$

where Z is obtained from the Z-Buffer. Thus, the removal of the rotation parameters results in a 3×3 linear system, in contrast to the 6×6 system used in Eq. 9. To reduce the sensitivity to depth discontinuities, the mean depth value within a local neighborhood of the feature point is used. Similarly, a sequence of T_zs corresponding to different feature points is computed. The median of these is chosen and outliers removed based on a threshold. The mean of the remaining values is the T_z estimate. Based on the position of the *focus of expansion*, T_x and T_y are then determined.

Fig. 6 compares our method to that of Bruss and Horn[8] on a 750 frame virtual colonoscopy (CT) image sequence consisting. As reviewed in [35], Bruss and Horn's method was considered to be one of the more superior methods, and a typical representative linear-square estimation method. It can be seen that the first 127 frames produce very little error; at this point, error starts significantly increasing (magenta curve) to about 80mm at the end of the sequence, while the error (blue curve) using our method remains around 10mm.

Fig. 7 shows a second experiment on a 652 frame colon phantom image sequence. A bent tube with artificial polyps glued to its interior surface was imaged using both CT and an endoscope. Fig. 7(a) shows the camera being tracked at frame 652, while in Fig. 7(b), Bruss and Horn's method is shown at the 10th frame. It can be seen that the camera has moved out of the colon phantom (external view on the right

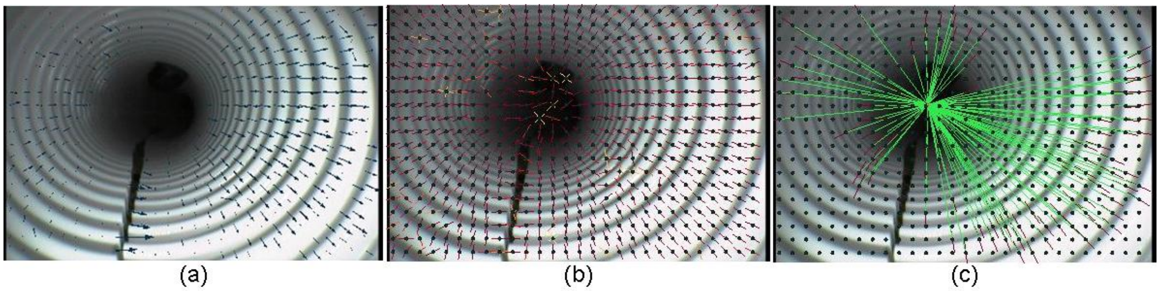


Figure 5. Determining *focus of expansion*: (a) Dense optical flow field; (b) Subdivision lattice, indicated by the dark dots; the prinicipal orientation within each region is indicated by the red lines, (c) focus of expansion (intersection of the green lines) is determined by least-squares fitting.

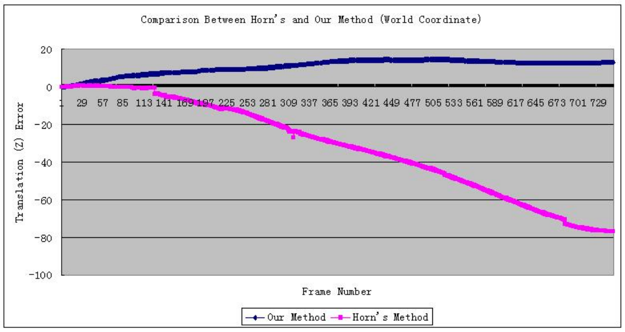


Figure 6. Comparison to Bruss and Horn's method on a 750 frame virtual colonoscopy sequence. Absolute error along Z axis. Our method results in an accumulated error of about 10mm, while Horn's method increases to 80mm.

shows the camera at the boundary of the colon phantom). In this sequence, our proposed algorithm was able to track the colon phantom images between the first and second polyp (about 700 frames). Note the second polyp displayed for the optical and the virtual images, as marked by red arrows in Fig. 7(a). As ground truth is not available for real data, we use polyps as landmarks to evaluate our method.

3. Conclusions

In this work, we have attempted to justify the robustness and stability of our proposed tracking algorithm. Our interest focuses on tracking optical colonoscopy images, which are mostly devoid of anatomical features that can otherwise be exploited by traditional registration algorithms. Another key distinction is the less stringent requirements on accuracy; generally, the tracking algorithm needs to place the virtual images in the vicinity of the endoscopic camera, as polyps and other abnormal tissue can easily be seen during the procedure. The VC images provide a kind of navigation 'roadmap' aid to the physician during the procedure.

Our method demonstrated the importance of both feature selection as well as scale selection in computing optic flow fields; a scale selection metric and an iterative algorithm are proposed to compute the optimal scale. Use of an optimal spatio-temporal scale provides a robust scheme to determining the FOE, which helps us to independently compute the rotation and translation parameters of the camera. Direct methods of computing these parameters were shown using perturbation theory to be numerically unstable, and validated by experiments on both VC nad OC image sequences. Finally, we have compared our approach to Bruss and Horn's algorithm using a VC and an OC image sequence.

Our next steps include more detailed testing of our algorithm on real colonoscopy data [1]. Additionally, the accuracy of our algorithm could be improved by reducing initialization errors. We also need to look at reinitialization, to recover from tracking failures (for instance, a long sequence of blurry frames) before this method can see application in clinical practice.

Appendix A

Here we show the difficulties in directly trying to solve the system described in Eqn. 10. The errors in optic flow can be modeled as

$$A\mathbf{x} = b + \delta \tag{16}$$

where δ represents a perturbation vector. In terms of matrix perturbation theory[33, 9], the bound of the relative error is

$$\frac{1}{n}\kappa(A)\frac{||b||}{||A||||\mathbf{x}||}\epsilon_b\mu \le \frac{||\overline{\mathbf{x}} - \mathbf{x}||}{||\mathbf{x}||} \le \sqrt{n}\kappa(A)\frac{||b||}{||A||||\mathbf{x}||}\epsilon_b \tag{17}$$

where $\kappa(A)$ is the condition number and $\epsilon_b = \frac{||\delta||}{||b||}$. Assume $A^{-1} = (r_1^T \dots r_n^T)^T$ and ψ_i is the angle between r_i and δ, then $\mu = \max_i\{\|r_i\||cos\psi_i|\}/max_k\|r_k\|$. $\overline{\mathbf{x}}$ is the estimated value of $\mathbf{x}$.

Let $\lambda_1 \le \lambda_2 \le \dots \le \lambda_n$ be the eigenvalues of A, the

[1] http://www.virtualcolonoscopy.nci.nih.gov

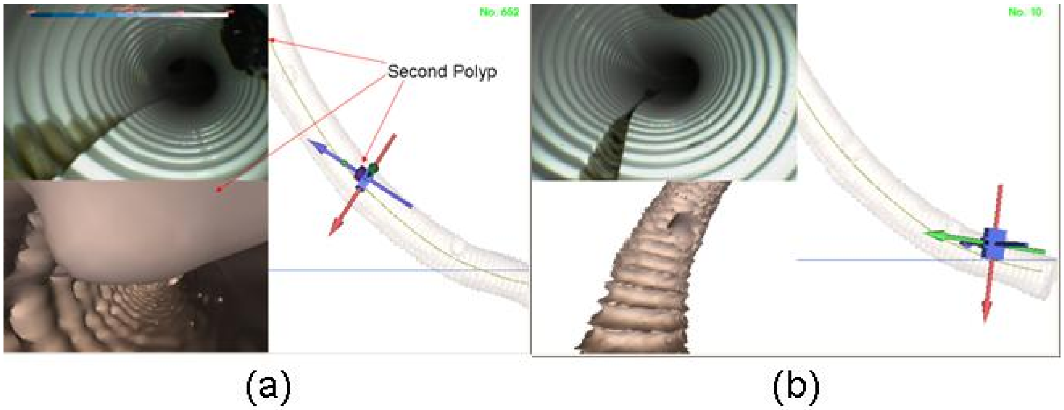


Figure 7. Comparison between our method and Bruss and Horn's method on phantom data. (a) Tracking results at frame 652, showing the camera successfully reaching the second artificial polyp, (b) Bruss and Horn's algorithm results at frame 10, where the camera is out of the virtual phantom. Phantom images are the top images and the virtual CT images are the lower images.

lower bound of Eq. 17 can be converted into

$$\begin{aligned}\frac{1}{n}\kappa(A)\frac{\|b\|}{\|A\|\|\mathbf{x}\|}\epsilon_b\mu &= \frac{1}{n}\frac{|\lambda_n|}{|\lambda_1|}\frac{\|b\|}{|\lambda_n|\|\mathbf{x}\|}\epsilon_b\mu \\ &= \frac{1}{n}\frac{\mu}{|\lambda_1|}\frac{\|b\|}{\|\mathbf{x}\|}\epsilon_b \qquad (18)\end{aligned}$$

as $\kappa(A) = |\lambda_n|/|\lambda_1|$ and $\|A\| = |\lambda_n|$[33]. $\frac{\|b\|}{\|\mathbf{x}\|}$ can be treated as a constant since $\mathbf{x}$ and b are the actual input and output signals, and do not affect the estimation process. ϵ_b relies on the measured signal and the output signal. Therefore, $\zeta = \frac{\mu}{|\lambda_1|}$ is solely related to the linear system. If it is stable, the estimated error will be small even if the perturbation ratio ϵ_b is high. Fig. 4 shows the relationship between the absolute translation error of the first 150 images of a virtual colonoscopy image sequence and ζ. Large translation errors are seen in X and Z on the frames close to the point marked **A** because $\lambda_1 \approx 3 \times 10^{-4}$, which increases the lower bound, although $\epsilon_b = 0.31$ is small. At the point **B**, a small error along Y-axis is due to the perturbation error, $\epsilon_b = 0.85$ and $\zeta = 89$. Although it might be possible to model the optical flow estimation error it is considerably harder to model ζ as it is dependent on the distribution of the feature points as well as the relationship between the perturbation vector and the estimation matrix. In addition, the perturbation effect is difficult to predict, as seen in point C, where the X translation error initially increases, then decreases near **B**.

References

[1] A snapshot of colorectal cancer, 2002. http://planning.cancer.gov/disease/Colorectal-Snapshot.pdf. 1

[2] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(4):384–401, 1985. 2

[3] G. Adiv. Inherent ambiguities in recovering 3d motion and structure from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):477–489, 1989. 2

[4] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989. 2

[5] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994. 2

[6] S. S. Beauchemin and J. L. Barron. The computation of optical flo. *ACM Computing Surveys*, 27(3):433–466, 1995. 2

[7] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. *Proceedings of 8th European Conference on Computer Vision*, 4:25–36, 2004. 2

[8] A. R. Bruss and B. K. P. Horn. Passive navigation. *Computer Vision, Graphics and Image Processing*, 21:3–20, 1983. 2, 4, 5

[9] S. Chandrasekaran and I. C. F. Ipsen. On the sensitivity of solution components in linear systems of equations. *SIAM Journal on Matrix Analysis and Applications*, 16(1):93–112, 1995. 6

[10] D. Deguchi, K. Mori, Y. Suenaga, J. Hasegawa, J. Toriwaki, H. T. Batake, and H. Natori. New image similarity measure for bronchoscope tracking based on image registration. *Proceedings of 6th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 399–406, 2003. 1

[11] F. Deligianni, A. Chung, and G. Z. Yang. Predictive camera tracking for bronchoscope simulation with condensation. *Proceedings of 8th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 910–917, 2005. 1

[12] F. Deligianni, A. Chung, and G. Z. Yang. Non-rigid 2d-3d registration with catheter tip em tracking for patient specific

bronchoscope simulation. *Proceedings of 9th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 281–288, 2006. 1

[13] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104, 1990. 2

[14] B. Galvin, B. McCane, K. Novins, D. Mason, and S. Mills. Recovering motion fields: An evaluation of eight optical flow algorithms. *Proceedings of the British Machine Vision Converence (BMVC)'98*, pages 195–204, 1998. 2

[15] T. Gautama and M. V. Hulle. A phase-based approach to the estimation of the optical flow field using spatial filtering. *IEEE Transactions on Neural Networks*, 13(5):1127–1136, 2002. 2

[16] C. Harris and M. J. Stephens. A combined corner and edge detector. *Proceedings of the Alvey Vision Conference*, pages 147–152, 1988. 2

[17] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion 1: Algorithm and implementation. *International Journal of Computer Vision*, 7(2):95–117, 1992. 2, 5

[18] J. P. Helferty, A. J. Sherbondy, A. P. Kiraly, and W. E. Higgins. System for live virtual-endoscopic guidance of bronchoscopy. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 68–68, 2005. 1

[19] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17(3):185–203, 1981. 1, 2, 3, 4

[20] P. C. I. Bricault, G. Ferretti. Multi-level strategy for computer-assisted transbronchial biopsy. *Proceedings of 1th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 161–268, 1998. 1

[21] A. E. Kaufman, S. Lakare, K. Kreeger, and I. Bitter. Virtual colonoscopy. *Communications of the ACM*, 48(2):37–41, 2005. 1

[22] J. Koenderink and A. J. V. Doorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22(9):773–791, 1975. 2

[23] D. Lieberman. Quality and colonoscopy: a new imperative. *Gastrointestinal endoscopy*, 61(3):392–394, 2005. 1

[24] T. Lindeberg. A scale selection principle for estimating image deformations. *Image and Vision Computing*, 16(14):961–977, 1998. 2

[25] J. Little, H. Bulthoff, and T. Poggio. Parallel optical flow using local voting. *Proceedings of 2th IEEE International Conference on Computer Vision*, pages 454–459, 1988. 2

[26] H. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London*, pages 385–397, 1980. 2, 4, 5

[27] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proceedings of International Joint Conference on Artificial Intelligence*, pages 281–288, 2006. 2, 3

[28] K. Mori, D. Deguchi, K. Akiyama, T. Kitasaka, C. R. M. Jr., Y. Suenaga, H. Takabatake, M. Mori, and H. Natori. Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. *Proceedings of 8th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 543–555, 2005. 1

[29] J. Nagao, K. Mori, T. Enjouji, and D. Deguchi. Fast and accurate bronchoscope tracking using image registration and motion prediction. *Proceedings of 7th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 551–558, 2004. 1

[30] D. Nain, S. Haker, W. Grimson, E. C. Jr., W. M. Wells, H. Ji, R. Kikinis, and C. F. Westin. Intra-patient prone to supine colon registration for synchronized colonoscopy. *Proceedings of 5th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 573–580, 2002. 1

[31] L. Rai, S. A. Merritt, and W. E. Higgins. Real-time image-based guidance method for lung-cancer assessment. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2437–2444, 2006. 1

[32] J. Reiger and D. Lawton. Processing differential image motion. *Journal of the Optical Society of America A*, 2(2):354–359, 1985. 2, 3, 5

[33] G. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, 1990. 2, 6, 7

[34] V. Sundareswaran. Egomotion from global flow field data. *Proceedings of the IEEE Workshop on Visual Motion*, pages 140–145, 1991. 2, 5

[35] T. Tian, C. Tomasi, and D. Heeger. Comparison of approaches to egomotion computation. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 315–320, 1996. 5

[36] A. Verri, F. Girosi, and V. Torre. Differential techniques for optica flow. *Journal of the Optical Society of America A*, 7(5):912–922, 1990. 2

[37] Y. Yacoob and L. Davis. Temporal multi-scale models for flow and acceleration. *International Journal of Computer Vision*, 32(2):147–163, 1999. 2