Temporal Volume Flow: An Approach to Tracking Failure Recovery

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

[a,b]The University of North Carolina at Charlotte
Charlotte Visualization Center, Dept. of Computer Science,
9201 University City Blvd, Charlotte, NC, 28223;
[c]Office of High Performance Computing and Communications,
National Library of Medicine, National Institutes of Health,
Bethesda, MD, 20892

ABSTRACT

The simultaneous use of pre-segmented CT colonoscopy images and optical colonoscopy images during routine endoscopic procedures provides useful clinical information to the gastroenterologist. Blurry images in the video stream can cause the tracking system to fail during the procedure, due to the endoscope touching the colon wall or a polyp. The ability to recover from such failures is necessary to continually track images, and goes towards building a robust tracking system. Identifying similar images before and after the blurry sequence is central to this task.

In this work, we propose a *Temporal Volume Flow(TVF)* based approach to search for a similar image pair before and after blurry sequences in the optical colonoscopy video. TVF employs nonlinear intensity and gradient constancy models, as well as a discontinuity-preserving smoothness constraint to formulate an energy function; minimizing this function between two temporal volumes before and after the blurry sequence results in an estimate of TVF. A voting approach is then used to determine an image pair with the maximum number of point correspondences. Region flow algorithm[10] is applied to the selected image pair to determine camera motion parameters.

We applied our algorithm to three optical colonoscopy sequences. The first sequence had 235 images in the ascending colon, and 12 blurry images. The image pair selected by TVF decreases the rotation error of the tracking results using the region flow algorithm. Similar results were observed in the second patient in the descending colon, containing 535 images and 24 blurry images. The third sequence contained 580 images in the descending colon and 172 blurry images. Region flow method failed in this case due to improper image pair selection; using TVF to determine the image pair allowed the system to successfully recover from the blurry sequence.

Keywords: colonoscopy tracking, temporal volume flow, egomotion

1. INTRODUCTION

The simultaneous use of pre-segmented CT colonoscopy images and optical colonoscopy images during routine endoscopic procedures provides useful clinical information to the gastroenterologist. Automatic tracking systems facilitate this capability, providing a clear and unambiguous spatial context to the location of the endoscope within the anatomical structure. However, as illustrated in Fig. 1, the appearance of blurry images that lack sufficient features in the optical video stream can cause tracking systems to fail. Blurry images are common in endoscopic sequences, for eg., endoscope touching the colon wall, fluid immersion, extreme brightness conditions, etc.

Further author information: (Send correspondence to Kalpathi R. Subramanian)
Jianfei Liu: E-mail: jliu1@uncc.edu, Telephone: 1 704 687 8641
Kalpathi R. Subramanian: E-mail: krs@uncc.edu, Telephone: 1 704 687 8579
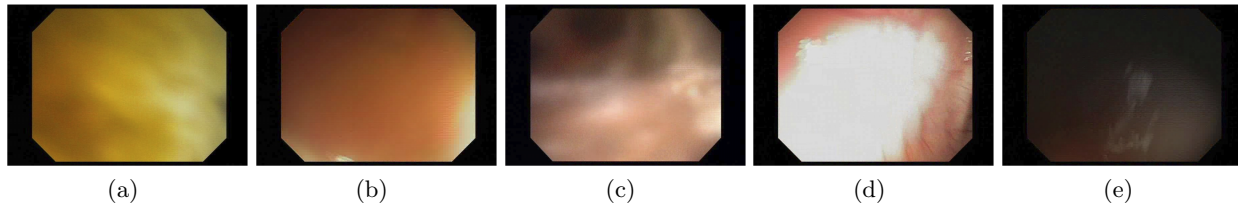Terry S. Yoo: E-mail:yoo@nlm.nih.gov, Telephone: 1 301 435 3268

Figure 1. Five different types of blurry images in optical colonoscopy video, (a) colonoscope immersed in fluid, (b) colonoscope touching colon wall, (c) colonoscope's lens covered by the water, (d) extremely bright regions, and (e) dark areas.

Although automatic tracking failure recovery is critical for endoscopy tracking systems, this issue is seldom investigated. Rai[14] enumerated several types of bronchoscopy images that are hard to track using image registration algorithms. However, no solutions were proposed to recover from these situations. Magnetic sensors have been used to assist tracking bronchoscopy images,[2, 12, 13] which help predict the endoscope position during blurry image sequences; image registration algorithms were used to improve tracking accuracy. The problem is considerably harder with colonoscopy video, since the colon can deform. In our earlier work, we used a region flow based approach[10] to recover motion parameters. In this method, region descriptors are used to model invariant structure information for matching, and are superior to point descriptors since the image pair bridging a blurry image sequence undergoes significant image motion. Region flow is computed to match all possible image regions between the image pair, and accurately identify sparse feature correspondences. However, the performance of the region flow based algorithm depends on the extent of the similarity of the selected pair. For instance, consider Fig. 2(a), that shows an image pair bridging a blurry sequence of 53 images used by region flow algorithm. The two images display the same fold, but with different intensity distributions and different scales.
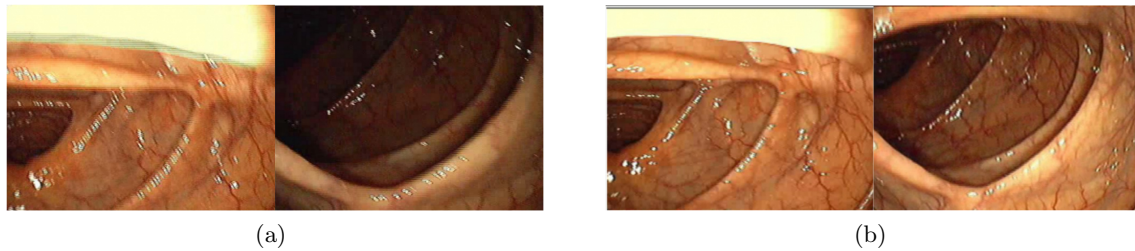

Figure 2. The comparison of the selected image pair used by region flow (a)) and temporal volume flow(b), respectively. The image pair selected by the region flow method differ in their intensity distribution as well as size of folds. On the contrary, the image pair selected by temporal volume flow is comparatively more similar.

Temporal volume matching, the subject of this paper, is an efficient means to identify temporal coherence, that will be used in determining the most similar image pair that bridges a blurry sequence. Laptev[7, 8] extended the multi-scale Harris point detector[9] to identify some interest points in the video stream, and developed a temporal feature descriptor for every point. Thus, temporal volumes can be matched by comparing temporal feature descriptors. There are many other temporal feature detectors.[6, 16, 17] However, all these algorithms detect a set of sparse feature points, and are mainly used for identifying significant transition points or events in the video stream. Our *Temporal Volume Flow(TVF)* approach is to densely match temporal volumes. Estimating temporal volume flow results in a robust and intelligent selection of an image pair for maximizing point correspondences. Fig. 2(b) shows an image pair selected by TVF. Compared with Fig. 2(a), there are two improvements, (1) intensity distribution is more similar, and (2) reduced scale variation.

2. METHODOLOGY

2.1. Temporal Volume Flow Computation

Assume there is a colonoscopy video stream $I(x, y, t)$ with a blurry sequence in the interval, (t_1, t_2). The purpose of TVF computation is to search for an image pair from two video segments, $\tau_1 = (t_1 - \Delta t, t_1)$ and $\tau_2 = (t_2, t_2 + \Delta t)$. Without loss of generality, we define τ as the segment time. A three-dimensional video stream is converted into a continuous *four-dimensional temporal volume*, $V(x, y, t, \tau)$. TVF computation is to densely match $V(x, y, t, \tau)$ at time $\tau = \tau_1$ and $\tau = \tau_2$.

Let $\overrightarrow{u} = (u_x, u_y, u_t, 1)$ be the flow vector at a point $\boldsymbol{p} = (x, y, t, \tau)$. We start with intensity and gradient constancy models.[1,5]

$$
\begin{aligned}
V(\boldsymbol{p} + \overrightarrow{u}) &= V(\boldsymbol{p}) \\
\nabla V(\boldsymbol{p} + \overrightarrow{u}) &= \nabla V(\boldsymbol{p})
\end{aligned}
\tag{1}
$$

where $\nabla = (\partial_x, \partial_y, \partial_t)^{\mathsf{T}}$. The linearized formulation of Eq. 1 is

$$
\begin{aligned}
V_x u_x + V_y u_y + V_t u_t + V_\tau &= 0 \\
V_{xx} u_x + V_{xy} u_y + V_{xt} u_t + V_{x\tau} &= 0 \\
V_{xy} u_x + V_{yy} u_y + V_{yt} u_t + V_{y\tau} &= 0 \\
V_{xt} u_x + V_{yt} u_y + V_{tt} u_t + V_{t\tau} &= 0
\end{aligned}
\tag{2}
$$

In conjunction with smoothness constraint, we can formulate an energy function to estimate TVF.

$$
\begin{aligned}
E(u_x, u_y, u_t) = \iint \Psi(|V(\boldsymbol{p} + \overrightarrow{u}) - V(\boldsymbol{p})|^2 + \alpha|\nabla V(\boldsymbol{p} + \overrightarrow{u}) - \nabla V(\boldsymbol{p})|^2) \\
+ \beta \Psi(|\nabla u_x|^2 + |\nabla u_y|^2 + |\nabla u_t|^2) dx dy dt
\end{aligned}
\tag{3}
$$

where $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}, \epsilon = 0.001$ is a modified $L1$ norm. α and β are two constants to balance different components in Eq. 3. The Euler-Lagrange equations of Eq. 3 are

$$
\begin{aligned}
&\Psi'(V_\tau^2 + \alpha(V_{x\tau}^2 + V_{y\tau}^2 + V_{t\tau}^2))(V_x V_\tau + \alpha(V_{xx} V_{x\tau} + V_{xy} V_{y\tau} + V_{xt} V_{t\tau})) \\
&\quad - \beta \text{div}\left(\Psi'(|\nabla u_x|^2 + |\nabla u_y|^2 + |\nabla u_t|^2)\nabla u_x\right) = 0 \\
&\Psi'(V_\tau^2 + \alpha(V_{x\tau}^2 + V_{y\tau}^2 + V_{t\tau}^2))(V_y V_\tau + \alpha(V_{xy} V_{x\tau} + V_{yy} V_{y\tau} + V_{yt} V_{t\tau})) \\
&\quad - \beta \text{div}\left(\Psi'(|\nabla u_x|^2 + |\nabla u_y|^2 + |\nabla u_t|^2)\nabla u_y\right) = 0 \\
&\Psi'(V_\tau^2 + \alpha(V_{x\tau}^2 + V_{y\tau}^2 + V_{t\tau}^2))(V_t V_\tau + \alpha(V_{xt} V_{x\tau} + V_{yt} V_{y\tau} + V_{tt} V_{t\tau})) \\
&\quad - \beta \text{div}\left(\Psi'(|\nabla u_x|^2 + |\nabla u_y|^2 + |\nabla u_t|^2)\nabla u_t\right) = 0
\end{aligned}
\tag{4}
$$

where the derivatives related to $V_{*\tau}$ are defined as the temporal difference.

$$
\begin{aligned}
V_\tau &= V(\boldsymbol{p} + \overrightarrow{u}) - V(\boldsymbol{p}) \quad V_{x\tau} = V_x(\boldsymbol{p} + \overrightarrow{u}) - V_x(\boldsymbol{p}) \\
V_{y\tau} &= V_y(\boldsymbol{p} + \overrightarrow{u}) - V_y(\boldsymbol{p}) \quad V_{t\tau} = V_t(\boldsymbol{p} + \overrightarrow{u}) - V_t(\boldsymbol{p})
\end{aligned}
$$

However, it is non-trival to compute Eq. 4 because it is a non-convex and nonlinear function. Two strategies are exploited to handle this issue.

1. **Multi-resolution scheme:** A coarse-to-fine temporal volume pyramid is built to initialize the minimization process near the actual minima with the sampling rate equal to 0.75. Let $\overrightarrow{u}^k = (u_x^k, u_y^k, u_t^k, 1)$ be the TVF vector at the pyramid level k and $\overrightarrow{u}^0 = (0, 0, 0, 1)$. The minimization process starts from the coarsest level and is gradually propagated towards the finest level.

2. **Sequential linearization strategy:** We employed two nested fixed point iterations[1] to remove non-linearity in Eqn. 4. Let l denote the outer iteration index, and the x-component of Eq. 4 can be rewritten as

$$
\begin{aligned}
&\Psi'((V_\tau^{k,l+1})^2 + \alpha((V_{x\tau}^{k,l+1})^2 + (V_{y\tau}^{k,l+1})^2 + (V_{t\tau}^{k,l+1})^2))(V_x^{k,l}V_\tau^{k,l+1} \\
&\quad + \alpha(V_{xx}^{k,l}V_{x\tau}^{k,l+1} + V_{xy}^{k,l}V_{y\tau}^{k,l+1} + V_{xt}^{k,l}V_{t\tau}^{k,l+1})) \\
&\quad - \beta\mathrm{div}\left(\Psi'(|\nabla u_x^{k,l+1}|^2 + |\nabla u_y^{k,l+1}|^2 + |\nabla u_t^{k,l+1}|^2)\nabla u_x^{k,l+1}\right) = 0
\end{aligned}
\tag{5}
$$

At iteration $l+1$, we can approximate

$$
\begin{aligned}
V_\tau^{k,l+1} &\approx V_\tau^{k,l} + V_x^{k,l}du_x^{k,l} + V_y^{k,l}du_y^{k,l} + V_t^{k,l}du_t^{k,l} \\
V_{x\tau}^{k,l+1} &\approx V_{x\tau}^{k,l} + V_{xx}^{k,l}du_x^{k,l} + V_{xy}^{k,l}du_y^{k,l} + V_{xt}^{k,l}du_t^{k,l} \\
V_{y\tau}^{k,l+1} &\approx V_{y\tau}^{k,l} + V_{xy}^{k,l}du_x^{k,l} + V_{yy}^{k,l}du_y^{k,l} + V_{yt}^{k,l}du_t^{k,l} \\
V_{t\tau}^{k,k+1} &\approx V_{t\tau}^{k,l} + V_{xt}^{k,l}du_x^{k,l} + V_{yt}^{k,l}du_y^{k,l} + V_{tt}^{k,l}du_t^{k,l}
\end{aligned}
\tag{6}
$$

through Taylor expansion. Here, $u_x^{k,l+1} = u_x^{k,l} + du_x^{k,l}$, $u_y^{k,l+1} = u_y^{k,l} + du_y^{k,l}$, and $u_t^{k,l+1} = u_t^{k,l} + du_t^{k,l}$. We define the following two terms to abbreviate the description in Eqs. 5,6.

$$
\begin{aligned}
\Psi'^{k,l}_D &= \Psi'((V_\tau^{k,l} + V_x^{k,l}du_x^{k,l} + V_y^{k,l}du_y^{k,l} + V_t^{k,l}du_t^{k,l})^2 + \alpha((V_{x\tau}^{k,l} + V_{xx}^{k,l}du_x^{k,l} + V_{xy}^{k,l}du_y^{k,l} + V_{xt}^{k,l}du_t^{k,l})^2 \\
&\quad + (V_{y\tau}^{k,l} + V_{xy}^{k,l}du_x^{k,l} + V_{yy}^{k,l}du_y^{k,l} + V_{yt}^{k,l}du_t^{k,l})^2 + (V_{t\tau}^{k,l} + V_{xt}^{k,l}du_x^{k,l} + V_{yt}^{k,l}du_y^{k,l} + V_{tt}^{k,l}du_t^{k,l})^2)) \\
\Psi'^{k,l}_S &= \Psi'(|\nabla(u_x^{k,l} + du_x^{k,l})|^2 + |\nabla(u_y^{k,l} + du_y^{k,l})|^2 + |\nabla(u_t^{k,l} + du_t^{k,l})|^2)
\end{aligned}
\tag{7}
$$

So Eq. 5 becomes

$$
\begin{aligned}
&\Psi'^{k,l}_D(V_x^{k,l}(V_\tau^{k,l} + V_x^{k,l}du_x^{k,l} + V_y^{k,l}du_y^{k,l} + V_t^{k,l}du_t^{k,l}) + \alpha(V_{xx}^{k,l}(V_{x\tau}^{k,l} + V_{xx}^{k,l}du_x^{k,l} + V_{xy}^{k,l}du_y^{k,l} + V_{xt}^{k,l}du_t^{k,l}) \\
&\quad + V_{xy}^{k,l}(V_{y\tau}^{k,l} + V_{xy}^{k,l}du_x^{k,l} + V_{yy}^{k,l}du_y^{k,l} + V_{yt}^{k,l}du_t^{k,l}) + V_{xt}^{k,l}(V_{t\tau}^{k,l} + V_{xt}^{k,l}du_x^{k,l} + V_{yt}^{k,l}du_y^{k,l} + V_{tt}^{k,l}du_t^{k,l}))) \\
&\quad - \beta\mathrm{div}\left(\Psi'^{k,l}_S\nabla(u_x^{k,l} + du_x^{k,l})\right) = 0
\end{aligned}
\tag{8}
$$

Another inner iteration is introduced to remove non-linearity in Ψ'_D and Ψ'_S. Let m be the iteration index, Eq. 8 can be linearized as

$$
\begin{aligned}
&\Psi'^{k,l,m}_D(V_x^{k,l}(V_\tau^{k,l} + V_x^{k,l}du_x^{k,l,m+1} + V_y^{k,l}du_y^{k,l,m+1} + V_t^{k,l}du_t^{k,l,m+1}) \\
&\quad + \alpha(V_{xx}^{k,l}(V_{x\tau}^{k,l} + V_{xx}^{k,l}du_x^{k,l,m+1} + V_{xy}^{k,l}du_y^{k,l,m+1} + V_{xt}^{k,l}du_t^{k,l,m+1}) \\
&\quad + V_{xy}^{k,l}(V_{y\tau}^{k,l} + V_{xy}^{k,l}du_x^{k,l,m+1} + V_{yy}^{k,l}du_y^{k,l,m+1} + V_{yt}^{k,l}du_t^{k,l,m+1}) \\
&\quad + V_{xt}^{k,l}(V_{t\tau}^{k,l} + V_{xt}^{k,l}du_x^{k,l,m+1} + V_{yt}^{k,l}du_y^{k,l,m+1} + V_{tt}^{k,l}du_t^{k,l,m+1}))) \\
&\quad - \beta\mathrm{div}\left(\Psi'^{k,l,m}_S\nabla(u_x^{k,l} + du_x^{k,l,m+1})\right) = 0
\end{aligned}
\tag{9}
$$

where $(du_x^{k,l,0}, du_y^{k,l,0}, du_t^{k,l,0}) = (0, 0, 0)$. Following the same procedure, we can derive equations corresponding to y and t components. Thus, each voxel has three linear equations, which leads to a massive sparse linear system to compute TVF. The system is solved by using the successive over-relaxation(SOR) method[18] in the innermost iteration. Let n indicate the index for this iteration.

$$
\begin{aligned}
(du_x)_p^{k,l,m,n+1} &= (1-\omega)(du_x)_p^{k,l,m,n} + \omega[\sum_{q\in\mathcal{N}^+(p)} (\Psi'_S)_{p\sim q}^{k,l,m}\left((u_x)_q^{k,l} + (du_x)_q^{k,l,m,n}\right) \\
&\quad + \sum_{q\in\mathcal{N}^-(p)} (\Psi'_S)_{p\sim q}^{k,l,m}\left((u_x)_q^{k,l} + (du_x)_q^{k,l,m,n+1}\right) - \sum_{q\in\mathcal{N}(p)} (\Psi'_S)_{p\sim q}^{k,l,m}(u_x)_p^{k,l}
\end{aligned}
$$

$$- \frac{1}{\alpha} (\Psi'_D)^{k,l,m}_{\boldsymbol{p}} \left((V_x^{k,l})_{\boldsymbol{p}} \left((V_y^{k,l})_{\boldsymbol{p}} (du_y)^{k,l,m,n}_{\boldsymbol{p}} + (V_t^{k,l})_{\boldsymbol{p}} (du_t)^{k,l,m,n}_{\boldsymbol{p}} + (V_\tau^{k,l})_{\boldsymbol{p}} \right) \right.$$

$$+ \gamma (V_{xx}^{k,l})_{\boldsymbol{p}} \left((V_{xy}^{k,l})_{\boldsymbol{p}} (du_y)^{k,l,m,n}_{\boldsymbol{p}} + (V_{xt}^{k,l})_{\boldsymbol{p}} (du_t)^{k,l,m,n}_{\boldsymbol{p}} + (V_{x\tau}^{k,l})_{\boldsymbol{p}} \right)$$

$$+ \gamma (V_{xy}^{k,l})_{\boldsymbol{p}} \left((V_{yy}^{k,l})_{\boldsymbol{p}} (du_y)^{k,l,m,n}_{\boldsymbol{p}} + (V_{yt}^{k,l})_{\boldsymbol{p}} (du_t)^{k,l,m,n}_{\boldsymbol{p}} + (V_{y\tau}^{k,l})_{\boldsymbol{p}} \right)$$

$$+ \gamma (V_{xt}^{k,l})_{\boldsymbol{p}} \left. \left((V_{yt}^{k,l})_{\boldsymbol{p}} (du_y)^{k,l,m,n}_{\boldsymbol{p}} + (V_{tt}^{k,l})_{\boldsymbol{p}} (du_t)^{k,l,m,n}_{\boldsymbol{p}} + (V_{t\tau}^{k,l})_{\boldsymbol{p}} \right) \right)]$$

$$/ [\sum_{\boldsymbol{q} \in \mathcal{N}(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} + \frac{1}{\alpha} (\Psi'_D)^{k,l,m}_{\boldsymbol{p}} \left((V_x^{k,l})^2_{\boldsymbol{p}} + \gamma \left((V_{xx}^{k,l})^2_{\boldsymbol{p}} + (V_{xy}^{k,l})^2_{\boldsymbol{p}} + (V_{xt}^{k,l})^2_{\boldsymbol{p}} \right) \right)]$$

$$(du_y)^{k,l,m,n+1}_{\boldsymbol{p}} = (1 - \omega)(du_y)^{k,l,m,n}_{\boldsymbol{p}} + \omega [\sum_{\boldsymbol{q} \in \mathcal{N}^+(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} \left((u_y)^{k,l}_{\boldsymbol{q}} + (du_y)^{k,l,m,n}_{\boldsymbol{q}} \right)$$

$$+ \sum_{\boldsymbol{q} \in \mathcal{N}^-(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} \left((u_y)^{k,l}_{\boldsymbol{q}} + (du_y)^{k,l,m,n+1}_{\boldsymbol{q}} \right) - \sum_{\boldsymbol{q} \in \mathcal{N}(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} (u_y)^{k,l}_{\boldsymbol{p}}$$

$$- \frac{1}{\alpha} (\Psi'_D)^{k,l,m}_{\boldsymbol{p}} \left((V_y^{k,l})_{\boldsymbol{p}} \left((V_x^{k,l})_{\boldsymbol{p}} (du_x)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_t^{k,l})_{\boldsymbol{p}} (du_t)^{k,l,m,n}_{\boldsymbol{p}} + (V_\tau^{k,l})_{\boldsymbol{p}} \right) \right.$$

$$+ \gamma (V_{xy}^{k,l})_{\boldsymbol{p}} \left((V_{xx}^{k,l})_{\boldsymbol{p}} (du_x)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{xt}^{k,l})_{\boldsymbol{p}} (du_t)^{k,l,m,n}_{\boldsymbol{p}} + (V_{x\tau}^{k,l})_{\boldsymbol{p}} \right)$$

$$+ \gamma (V_{yy}^{k,l})_{\boldsymbol{p}} \left((V_{xy}^{k,l})_{\boldsymbol{p}} (du_x)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{yt}^{k,l})_{\boldsymbol{p}} (du_t)^{k,l,m,n}_{\boldsymbol{p}} + (V_{y\tau}^{k,l})_{\boldsymbol{p}} \right)$$

$$+ \gamma (V_{yt}^{k,l})_{\boldsymbol{p}} \left. \left((V_{xt}^{k,l})_{\boldsymbol{p}} (du_x)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{tt}^{k,l})_{\boldsymbol{p}} (du_t)^{k,l,m,n}_{\boldsymbol{p}} + (V_{t\tau}^{k,l})_{\boldsymbol{p}} \right) \right)]$$

$$/ [\sum_{\boldsymbol{q} \in \mathcal{N}(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} + \frac{1}{\alpha} (\Psi'_D)^{k,l,m}_{\boldsymbol{p}} \left((V_y^{k,l})^2_{\boldsymbol{p}} + \gamma \left((V_{xy}^{k,l})^2_{\boldsymbol{p}} + (V_{yy}^{k,l})^2_{\boldsymbol{p}} + (V_{yt}^{k,l})^2_{\boldsymbol{p}} \right) \right)]$$

$$(du_t)^{k,l,m,n+1}_{\boldsymbol{p}} = (1 - \omega)(du_t)^{k,l,m,n}_{\boldsymbol{p}} + \omega [\sum_{\boldsymbol{q} \in \mathcal{N}^+(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} \left((u_t)^{k,l}_{\boldsymbol{q}} + (du_t)^{k,l,m,n}_{\boldsymbol{q}} \right)$$

$$+ \sum_{\boldsymbol{q} \in \mathcal{N}^-(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} \left((u_t)^{k,l}_{\boldsymbol{q}} + (du_t)^{k,l,m,n+1}_{\boldsymbol{q}} \right) - \sum_{\boldsymbol{q} \in \mathcal{N}(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} (u_t)^{k,l}_{\boldsymbol{p}}$$

$$- \frac{1}{\alpha} (\Psi'_D)^{k,l,m}_{\boldsymbol{p}} \left((V_t^{k,l})_{\boldsymbol{p}} \left((V_x^{k,l})_{\boldsymbol{p}} (du_x)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_y^{k,l})_{\boldsymbol{p}} (du_y)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_\tau^{k,l})_{\boldsymbol{p}} \right) \right.$$

$$+ \gamma (V_{xt}^{k,l})_{\boldsymbol{p}} \left((V_{xx}^{k,l})_{\boldsymbol{p}} (du_x)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{xy}^{k,l})_{\boldsymbol{p}} (du_y)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{x\tau}^{k,l})_{\boldsymbol{p}} \right)$$

$$+ \gamma (V_{yt}^{k,l})_{\boldsymbol{p}} \left((V_{xy}^{k,l})_{\boldsymbol{p}} (du_x)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{yy}^{k,l})_{\boldsymbol{p}} (du_y)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{y\tau}^{k,l})_{\boldsymbol{p}} \right)$$

$$+ \gamma (V_{tt}^{k,l})_{\boldsymbol{p}} \left. \left((V_{xt}^{k,l})_{\boldsymbol{p}} (du_x)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{yt}^{k,l})_{\boldsymbol{p}} (du_y)^{k,l,m,n+1}_{\boldsymbol{p}} + (V_{t\tau}^{k,l})_{\boldsymbol{p}} \right) \right)]$$

$$/ [\sum_{\boldsymbol{q} \in \mathcal{N}(\boldsymbol{p})} (\Psi'_S)^{k,l,m}_{\boldsymbol{p} \sim \boldsymbol{q}} + \frac{1}{\alpha} (\Psi'_D)^{k,l,m}_{\boldsymbol{p}} \left((V_t^{k,l})^2_{\boldsymbol{p}} + \gamma \left((V_{xt}^{k,l})^2_{\boldsymbol{p}} + (V_{yt}^{k,l})^2_{\boldsymbol{p}} + (V_{tt}^{k,l})^2_{\boldsymbol{p}} \right) \right)] \qquad (10)$$

Here, $\mathcal{N}(\boldsymbol{p})$ is the 6-neighborhood[15] of $\boldsymbol{p}$. $\mathcal{N}^+(\boldsymbol{p})$ denotes the neighbors $\boldsymbol{q}$ of $\boldsymbol{p}$ with the indices of $\boldsymbol{q}$ larger than that of $\boldsymbol{p}$, and $\mathcal{N}^-(\boldsymbol{p})$ with the indices of $\boldsymbol{q}$ is smaller than that of $\boldsymbol{p}$. $\omega \in (0, 2)$ is the relaxation parameter that affects the convergence of the linear system. As suggested by Young,[18] values close to 2 gives the best performance. We use $\omega = 1.99$ in our implementation.

The TVF computation is summarized in Algorithm 1, and the left image of Fig. 3 illustrates the results of TVF. Note the movement of the bottom right fold between two volumes; all flow vectors capture this main motion.

2.2. Tracking Failure Recovery

The computed volume flow between the two temporal volumes describes voxel correspondences between images in the two volumes. Thus, if there are m and n images respectively in the two volumes, there are mn pairs of images that will be considered. We select the image pair that has the largest number of voxel correspondences.

Algorithm 1: Temporal Volume Flow Computation

Data: $V(x,y,t,\tau)$ at τ_1 and τ_2

Result: TVF field $\overrightarrow{u}(x,y,t)$.

1 Build K level temporal volume pyramids for $V(x,y,t,\tau_1)$ and $V(x,y,t,\tau_2)$;

2 **for** $k \leftarrow 1$ **to** K **do**

3   Initialize k-th level TVF field;

4   **for** $l \leftarrow 1$ **to** L **do**

5   Compute derivatives $V_\tau^{k,l}, V_{x\tau}^{k,l}, V_{y\tau}^{k,l}$ and $V_{t\tau}^{k,l}$ through bilinear interpolation;

6   **for** $m \leftarrow 1$ **to** M **do**

7   Compute diffusion terms $\Psi_D'^{k,l,m}$ and $\Psi_S'^{k,l,m}$;

8   Initialize increment TVF vector $d\overrightarrow{u}^{k,l,m} = 0$;

9   **for** $n \leftarrow 1$ **to** N **do**

10   Update increment TVF vector $d\overrightarrow{u}^{k,l,m,n}$ in Eq. 10 at every voxel through SOR method;

11   Update TVF field $\overrightarrow{u}^{k,l,m} = \overrightarrow{u}^{k,l,m} + d\overrightarrow{u}^{k,l,m}$;
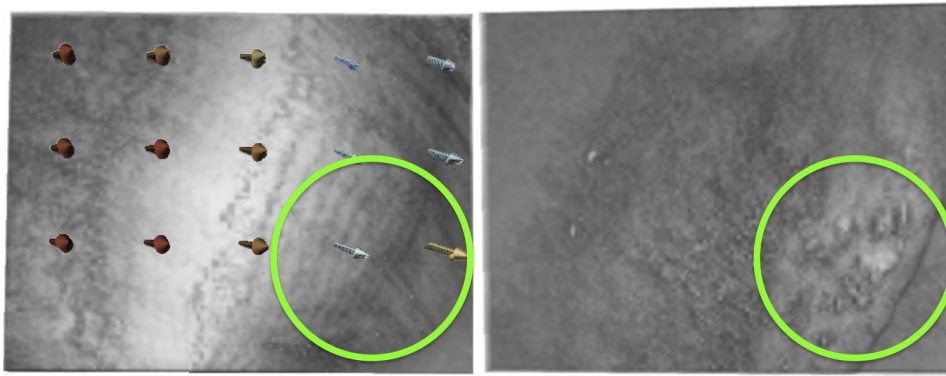

Figure 3. The temporal volume flow between two image sequences; left and right images are before and after a blurry sequence. Corresponding folds are highlighted by green circles.The fold in the left image moves to the bottom right corner in the right image. Clearly, computed flow vectors follow this main motion.

Next, the region flow algorithm[10] is used to recover the camera motion parameters. In this method, we compute the normalized cross-correlation metric on all 5×5 sized image regions between the two images. Max-product loop belief propagation algorithm[3] is employed to minimize the resulting graph energy. The minimization results in a set of region flow vectors that provide a good estimate of the image motion. SIFT detector[11] is used to detect two sets of feature points in the image pair, and these are matched by using region flow vectors to guide and limit the search space. The camera motion parameters are computed by identifying the the visual angle changes between any two SIFT features.[19]

3. EXPERIMENTAL RESULTS

We have tested our method on three optical colonoscopy sequences, from three patients.

Sequence 1: Ascending Colon. This sequence contained 235 images in the ascending colon and had 12 blurry images between 77 and 88. Region flow method chose frames 73 and 89 to compute motion parameters, while TVF selected 68 and 109. Both region flow and temporal volume flow are able to continue tracking, because corresponding folds appear in OC and VC images, as seen in the yellow lines in Fig. 4. However, the region flow method resulted in a large rotation error, in contrast to the TVF method, as illustrated in Figs. 4(a) and 4(b).

Sequence 2: Descending Colon. Similar results are observed in a 535 image sequence with a rounded polyp in the descending colon, as shown in Fig. 5. This sequence contains 24 blurry images(caused by fluid) from frame 130 to 153. The colon also undergoes contraction as well as expansion in this sequence. Both methods continue
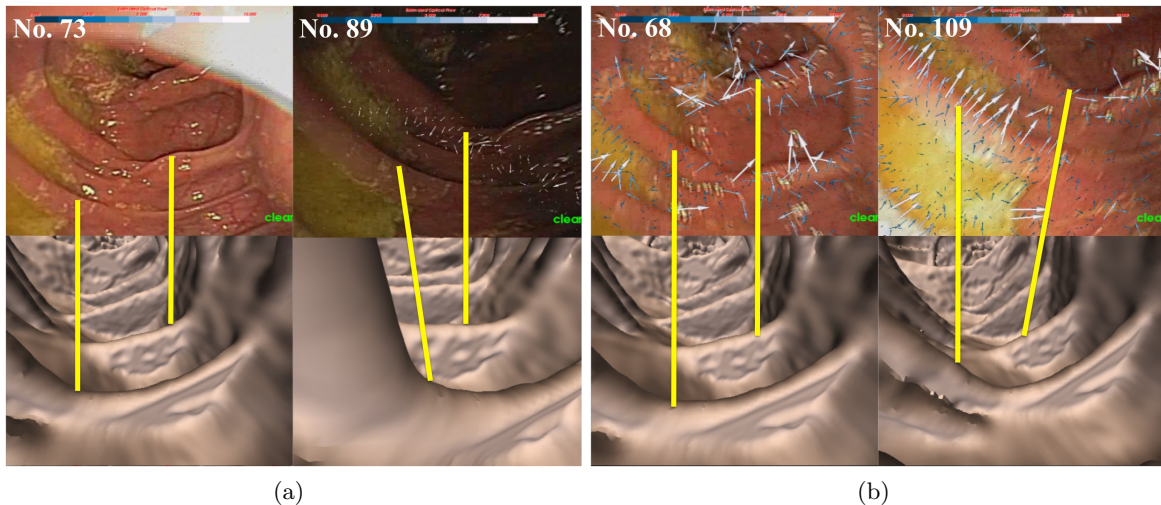
$$(a) \qquad\qquad (b)$$

Figure 4. Comparison of tracking results using region flow vs. temporal volume flow in the ascending colon. (a) Region flow image pair (73 and 89), (b) TVF image pair (68 and 109). Corresponding folds are connected by yellow lines in OC and VC images. There is significant rotation error in the left image, in contrast to the right image.

to successfully track after the blurry sequence. At frame 535, although the polyp is visible in both OC and VC images, TVF had decreased error, since the polyp in the VC image is more similar to that in the OC image in Fig. 5(f).

Sequence 3: Descending Colon. Fig. 6 illustrates a 580 image sequence in the descending colon after polyp removal. There is a long blurry image sequence from 82 to 353. The locations of the polyp are highlighted by red rectangles in Fig. 6(a). Region flow fails to recover the motion parameters because there are insufficient feature correspondences from the selected frames, 30 and 356. On the contrary, TVF successfully continues to track (selecting frames 30 and 374), as seen in Fig. 6(c). Note the same folds inside the rectangles appear in both OC and VC images.

4. CONCLUSION AND FUTURE WORK

In this work, we have proposed a temporal volume flow approach to continually track colonoscopy video sequences that encounter blurry image sequences. Our method employs nonlinear intensity and gradient constancy models, which are combined into an energy function. The energy function is minimized through coarse-to-fine and sequential linearization schemes. The image pair with the most similar features before and after a blurry sequence is identified and used to compute motion parameters.

Three clinical sequences were chosen to test the TVF algorithm. The first two clinical sequences showed that accuracy can be improved through the TVF algorithm. The third sequence demonstrated that an improper choice of image pair can cause tracking to fail, as was the case with the region flow approach. On the contrary, the image pair chosen by TVF contained sufficient corresponding features, and the system continued to track images till the end of the sequence.

Future work will involve looking into improving the performance of the temporal volume flow computation, as ultimately real-time performance will be required for application in clinical practice. Comparison of our method to image registration approaches[4] also needs to be investigated, to better understand the benefits of exploiting temporal coherence.
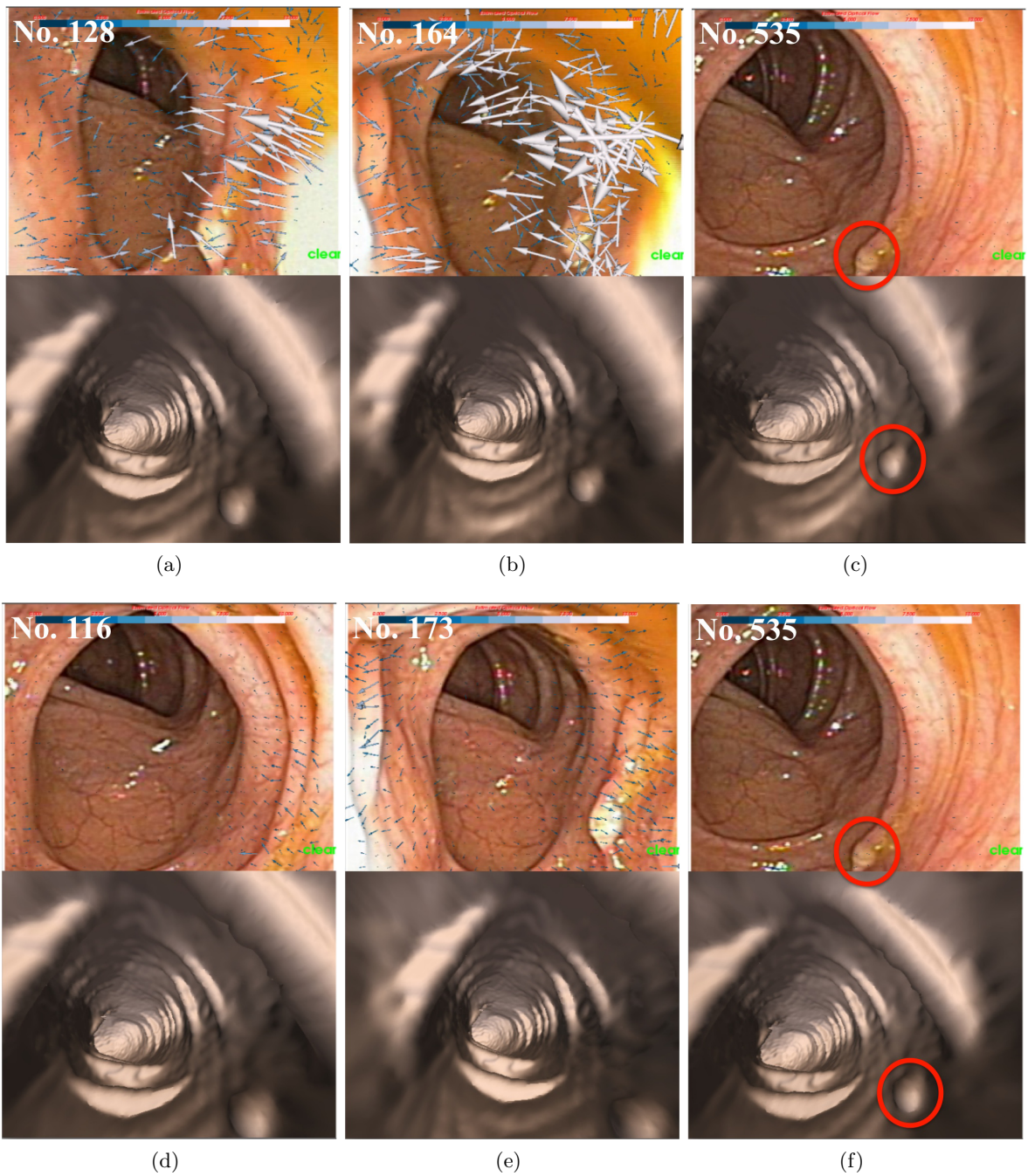
Figure 5. Comparison of the tracking results using region flow(top row, a-c) and temporal volume flow(bottom row, d-f) in the descending colon with a rounded polyp. (a,b) Tracking results of the selected image pair at frames 128 and 164, (c) Tracking results at frame 535, polyp highlighted by red circle in OC and VC images, (d,e) Frames 116 and 173 selected by volume flow, (f) Tracking results at frame 535. Tracking accuracy is improved since polyp is close to the bottom of image.
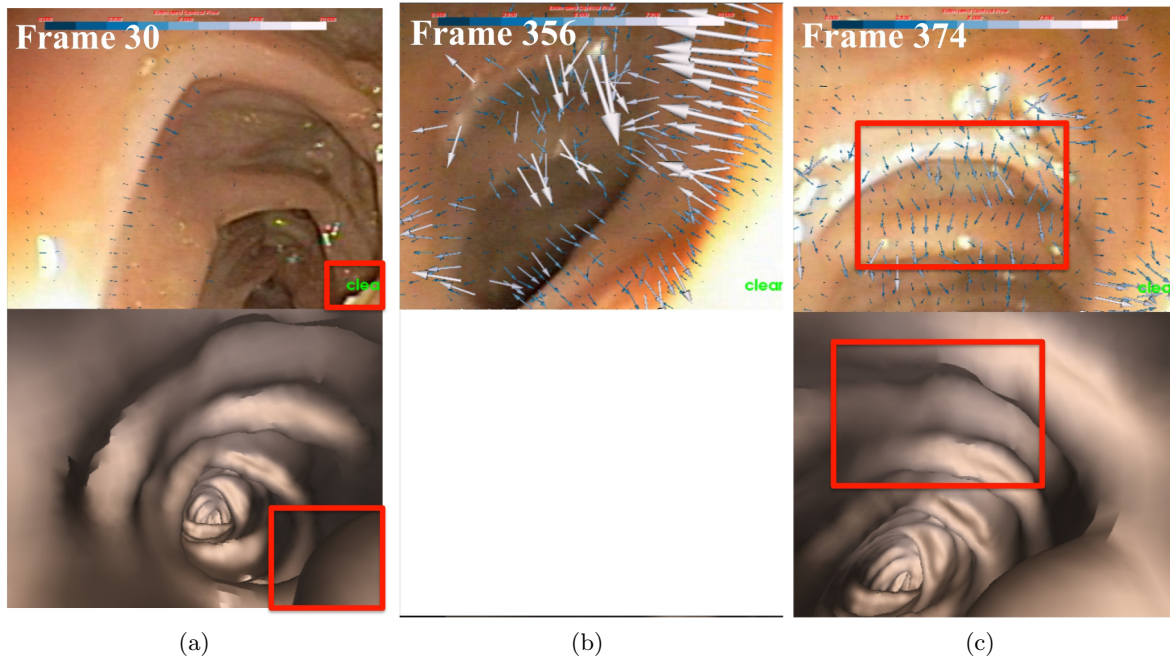
Figure 6. Comparison of the tracking results using region flow and temporal volume flow in the descending colon after polyp removal. (a) OC and VC images at frame 30 before a blurry sequence; Red rectangles indicate polyp locations, (b) Region flow fails to track after the blurry sequence, selecting frame 356 to match frame 30, (c)Temporal volume flow chooses frame 374 to successfully continue tracking, because the same folds (red rectangles) appear in both OC and VC images.

REFERENCES

1. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of ECCV*, pages 25–36, 2004.

2. F. Deligianni, A. J. Chung, and G.Z. Yang. Non-rigid 2d-3d registration with catheter tip em tracking for patient specific bronchoscope simulation. In *Proc. of MICCAI*, pages 281–288, 2006.

3. P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1):41–54, 2006.

4. J. P. Helferty, A. J. Sherbondy, A. P. Kiraly, and W. E. Higgins. System for live virtual-endoscopic guidance of bronchoscopy. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 68–75, 2005.

5. B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(3):185–203, 1981.

6. Yan Ke, Rahul Sukthankar, and Martial Hebert. Efficient visual event detection using volumetric features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 166–173, 2005.

7. I. Laptev. On space-time interest points. *International Journal of Computer Vision*, 64(2):107–123, 2005.

8. I. Laptev and T. Lindeberg. Space-time interest points. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 432–439, 2003.

9. T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):77–116, 1998.

10. Jianfei Liu, K.R. Subramanian, and T.S. Yoo. Region flow: A multi-stage method for colonoscopy tracking. In *Proc. of MICCAI*, 2010. September 20-24, Beijing, China.

11. D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

12. K. Mori, D. Deguchi, K. Aklyama, K. Kitasaka, CR. Jr Maurer, Y. Suenaga, H. Takabatake, M. Mori, and H. Natori. Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In *Proc. of MICCAI*, pages 543–550, 2005.

13. K. Mori, D. Deguchi, K. Ishitani, T. Kitasaka, Y. Suenaga, Y. Hasegawa, K. Imaizumi, and H. Takabatake. Bronchoscope tracking without fiducial markers using ultra-tiny electromagnetic tracking system and its evaluation in different environments. In *Proc. of MICCAI*, pages 644–651, 2007.

14. Lav Rai, James Helferty, and William Higgins. Combined video tracking and image-video registration for continuous bronchoscopic guidance. *International Journal of Computer Assisted Radiology and Surgery*, 3(3-4):315–329, 2008.

15. Azriel Rosenfeld. Three-dimensional digital topology. *Information and Control*, 50(2):119–127, 1981.

16. Paul Scovanner, Saad Ali, and Mubarak Shah. A 3-dimensional sift descriptor and its application to action recognition. In *Proceedings of the 15th international conference on Multimedia*, pages 357–360, 2007.

17. Geert Willems, Tinne Tuytelaars, and Luc Van Gool. An efficient dense and scale-invariant spatio-temporal interest point detector. In *Proceedings of the 10th European Conference on Computer Vision*, pages 650–663, 2008.

18. David M. Young. *Iterative Solution of Large Linear Systems (Computer Science and Applied Mathematics)*. Academic Pr, 1rd edition, 1971.

19. John Zhang. *Computing Camera Heading: A Study*. PhD thesis, Sanford University, 1999.