Paper Presentation

Alperen Degirmenci

April 19, 2011

Liu, J.; Subramanian, K.; Yoo, T.; Van Uitert, R.

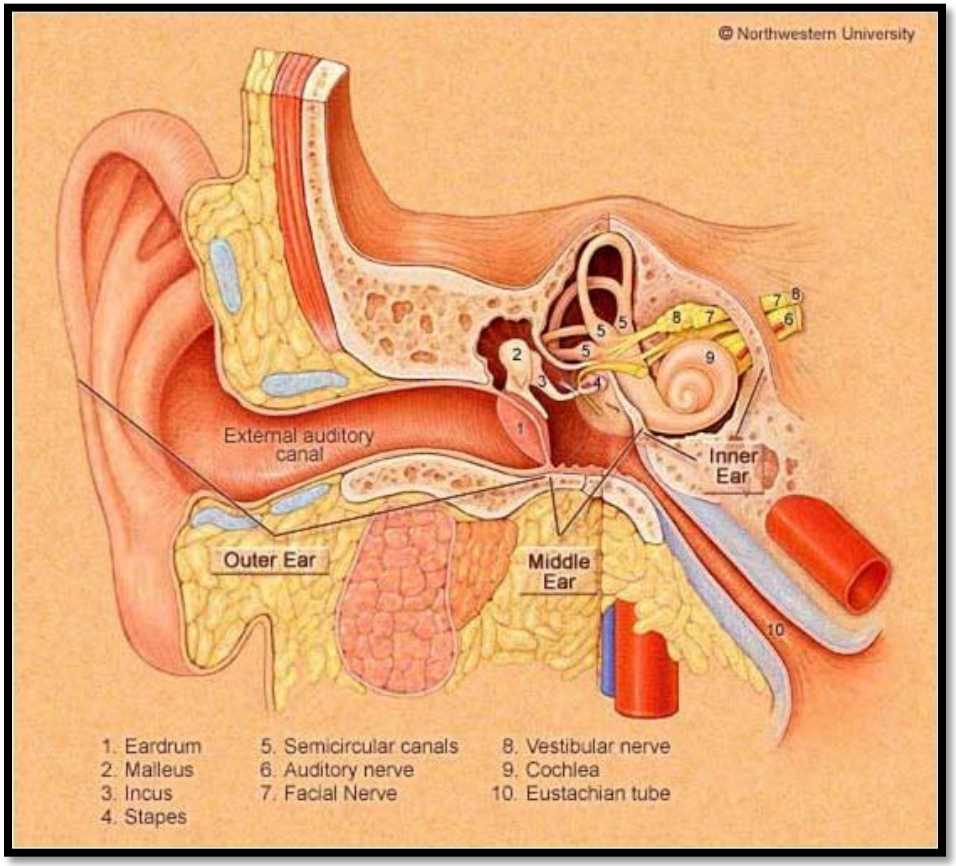
“A stable optic-flow based method for tracking colonoscopy images,”

Computer Vision and Pattern Recognition Workshops, 2008 IEEE Computer Society Conference: pp. 1-8, June 2008.

JOHNS HOPKINS
UNIVERSITY

Our Project

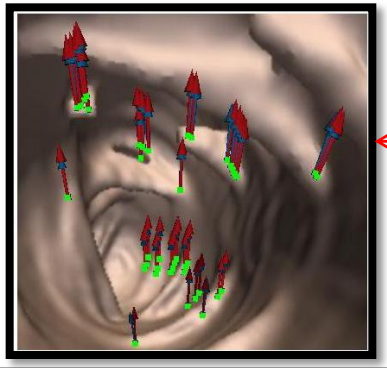
Using a borescope for generating a 3D reconstruction of the cochlear canal and a safe insertion path via virtual fixtures.

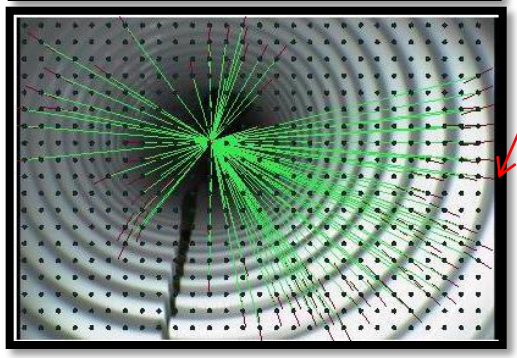


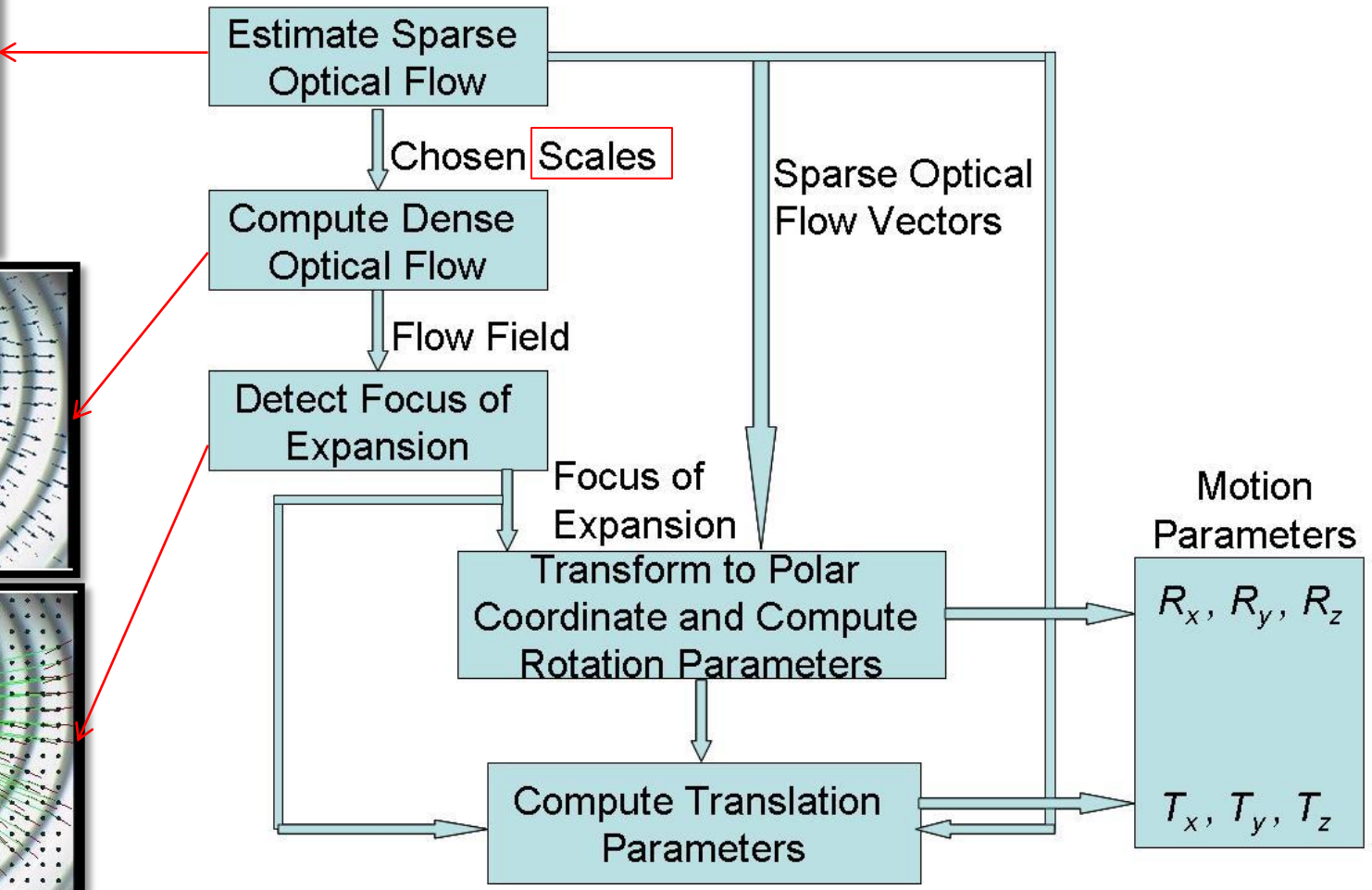
Background and Motivation

- A better algorithm for determining the position of the endoscopic camera during colonoscopy
- Uses optical flow
- Can be used as a part of our 3D reconstruction algorithm

Algorithm Overview

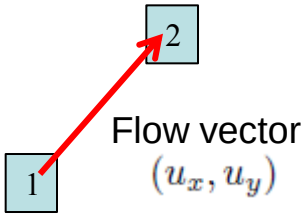






Figures taken from Liu et al.

Sparse Optical Flow Field



Pixel location (x,y)
Time t

If L is the scale-space representation of image I, then

$$L(x, y, t) = L(x + u_x, y + u_y, t + 1)$$

Scale-space image | Anisotropic Gaussian | Image

$$L(x, y, t) = g(x, y, t; \sigma^2, \tau^2) * I(x, y, t)$$

$$g(x, y, t; \sigma^2, \tau^2) = \frac{e^{\left(\frac{-(x^2+y^2)}{2\sigma^2} - \frac{t^2}{2\tau^2}\right)}}{\sqrt{(2\pi)^3 \sigma^4 \tau^2}}$$

Anisotropic Gaussian is applied to account for the differential sampling rates across the spatial and temporal dimensions.

σ,and τ are the spatial and temporal scale parameters

Sparse Optical Flow Field

- By applying the following Harris matrix, corners in the image are detected

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

where

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

- Will be used as features for tracking

Sparse Optical Flow Field

We are interested in corresponding pairs that exhibit maximum variance in the spatial domain and minimum difference along the temporal direction.

Similarity between corresponding pairs

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

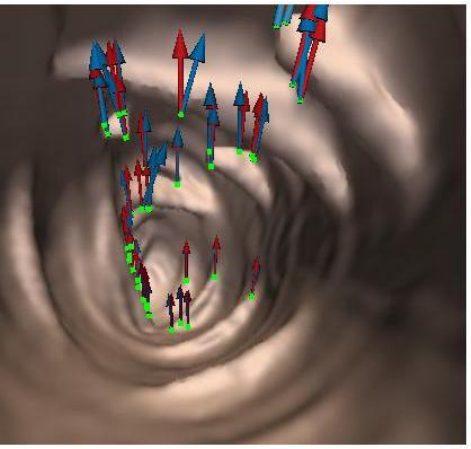
How distinct the selected features are in their local neighborhood

Smaller the Θ, better the match

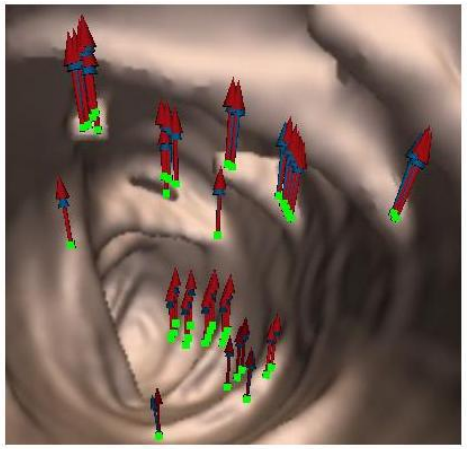
The numerator can be converted into the iterative Lucas-Kanade algorithm using Taylor series approximation

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

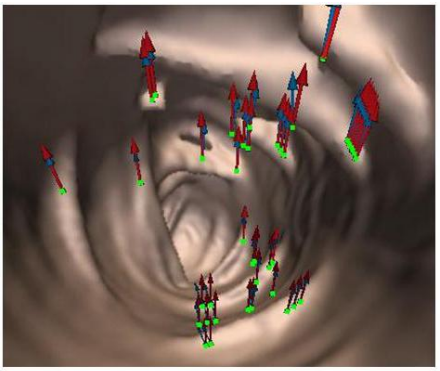
Sparse Optical Flow Field



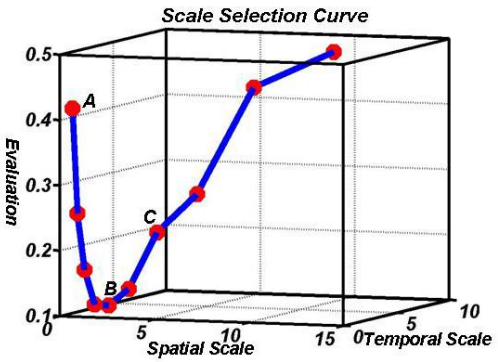
(a)



(b)



(c)



(d)

Figures taken from Liu et al.

- The relationship between spatio-temporal scale and the scale metric.

- Ground truth flow vectors are in red
- Estimated flow vectors are in blue

- Green cubes represent the selected feature point positions

- (a) Results with fine spatial and temporal scales
- (b) Results with optimal spatial and temporal scales
- (c) Result with relatively coarse scales
- (d) The response curve between spatiotemporal scales and the scale metric

- The scale values at points A, B and C correspond to images (a), (b), and (c) respectively.

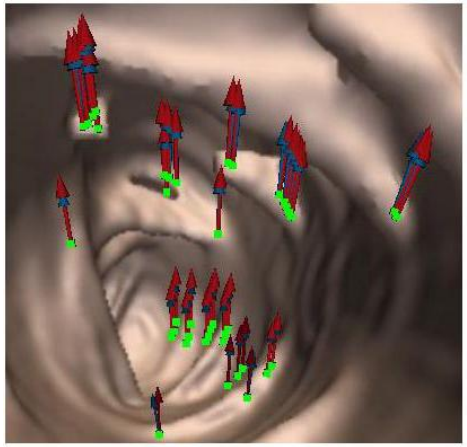
Sparse Optical Flow Field

(σ, τ)	Error Type	Error Measurements		
		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

$$\varepsilon_{magnitude} = \frac{\|\mathbf{u} - \mathbf{v}\|}{\|\mathbf{v}\|}$$

$$\varepsilon_{direction} = \frac{|\mathbf{u} \cdot \mathbf{v}|}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

where u and v are estimated and groundtruth flow vectors.

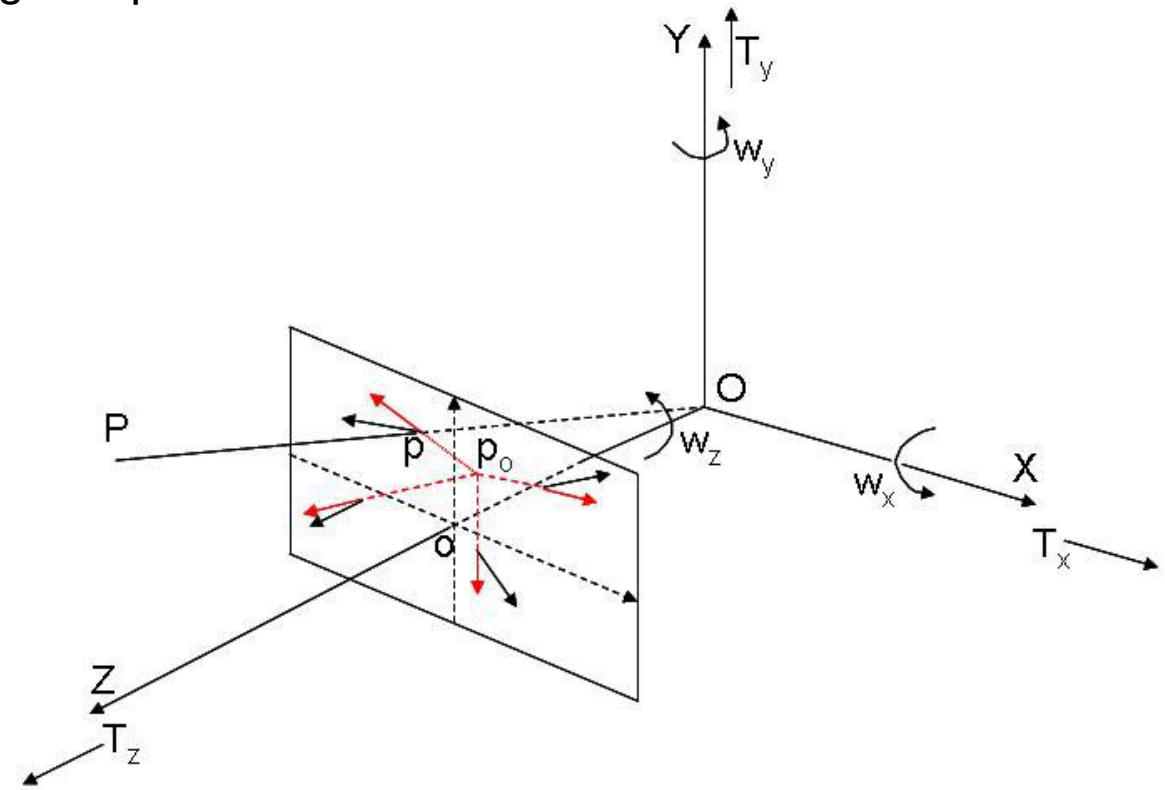


Figures taken from Liu et al.

Dense Flow Field and FOE

- The characteristic spatial and temporal scales are used to compute a dense optic flow field.

- More accurate than if a single scale was chosen throughout the image sequence.



Figures taken from Liu et al.

Dense Flow Field and FOE

- Object point P: in camera coordinates (X,Y,Z)
- Projection of P, point p: in the image plane (x, y)
- Geometrically, its optical flow is:

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

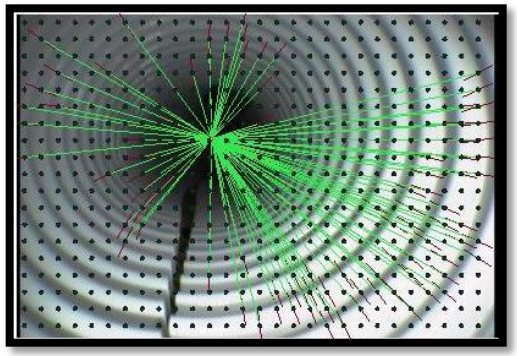
Translation Rotation

$$\mathbf{T} = (T_x, T_y, T_z) \text{ and } \mathbf{R} = (\omega_x, \omega_y, \omega_z)$$

Figures taken from Liu et al.

Determining Camera Motion Parameters

$$u_x = u_x^T + u_x^R$$
$$u_y = u_y^T + u_y^R$$



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

The vector joining two object points having different depth values and projecting on to the same pixel will point toward the focus of expansion.

Instead of solving a 6x6 linear system, the flow vectors can be transformed into polar coordinates with the focus of expansion at the origin. This gives us the rotation term, since the translation term is eliminated.

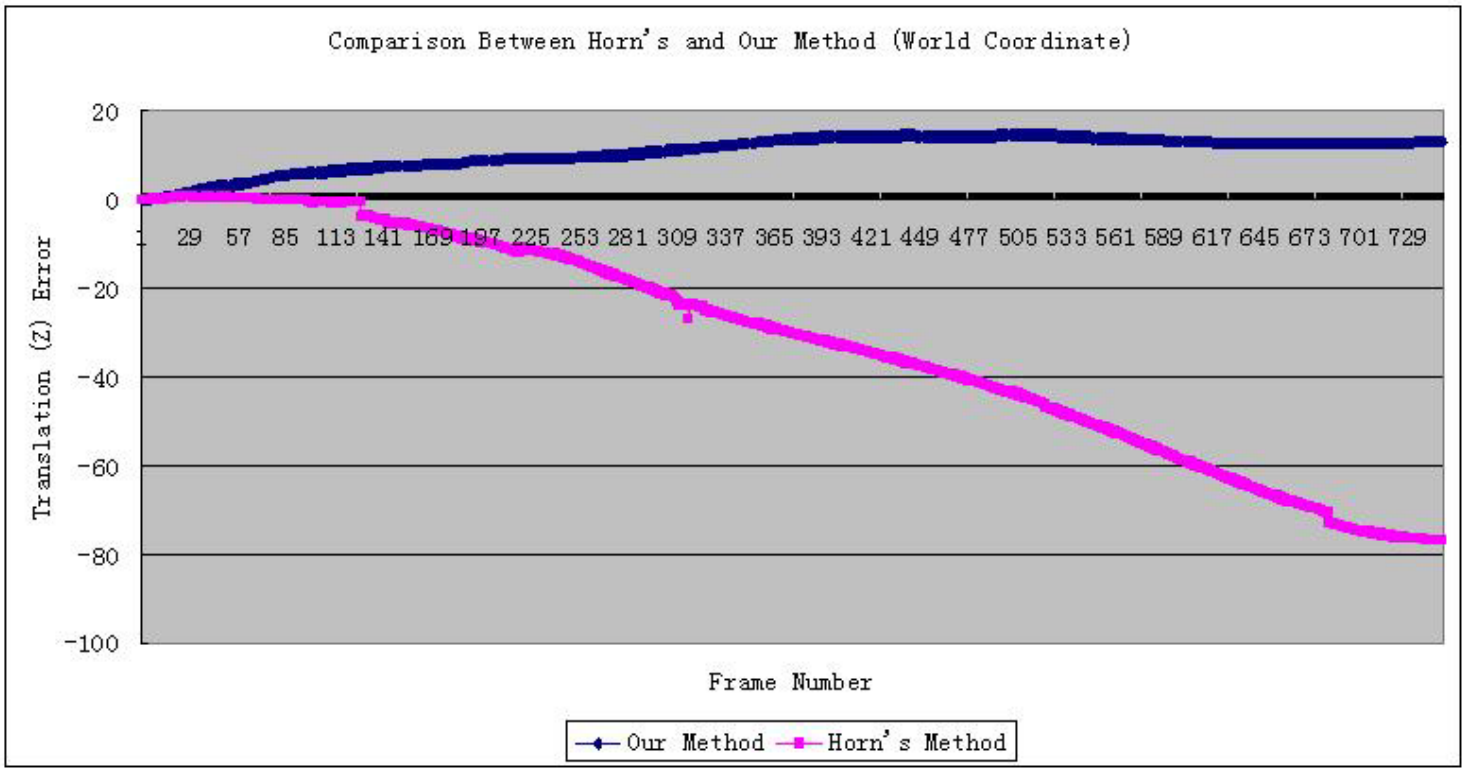
$$u \cdot e_\perp = (u_T + u_R) \cdot e_\perp = u_R \cdot e_\perp$$

Figures taken from Liu et al.

Improving Accuracy

A sequence of Tz's 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 Tz estimate. Based on the position of the focus of expansion, Tx and Ty are then determined.

$$Z/T_z = \|d\|/\|u_T\|$$



Figures taken from Liu et al.

Improving Accuracy

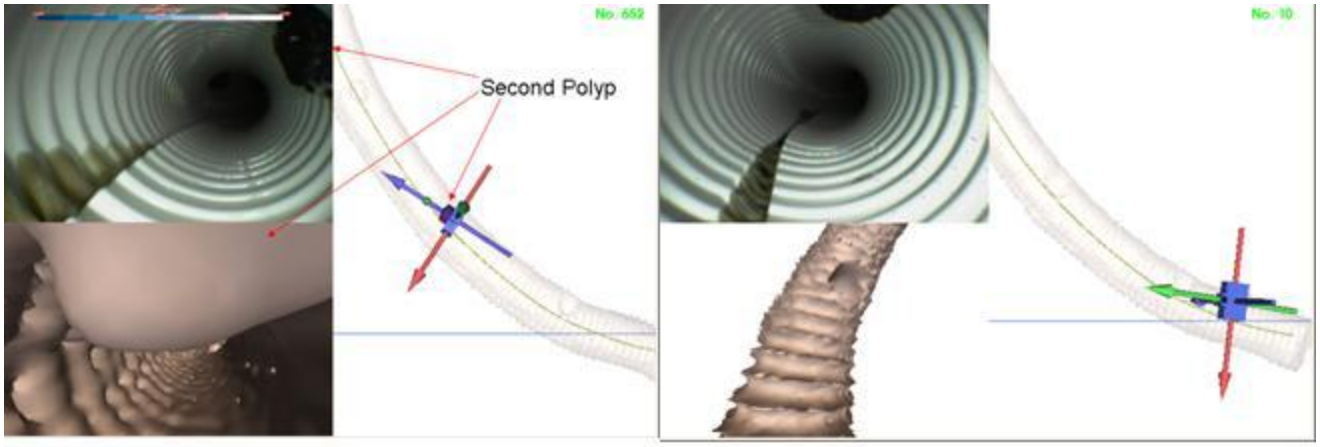


Figure 7. Comparison between our method and Bruss and Horn's method on phantom data. (a) Tracking results at frame 652, showing the camera successfully reaching the second artificial polyp, (b) Bruss and Horn's algorithm results at frame 10, where the camera is out of the virtual phantom. Phantom images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

Figures taken from Liu et al.

Critique

- Neglected to specify the thresholds used for eliminating outliers and for choosing good tracking features.
- Didn’t explain how they have registered the phantom images, CT scans, and the endoscopic camera, or how they generated these graphics.
- How computational speed of the algorithm compares to other methods’. In order for this algorithm to be feasible, it has to run in real-time.

Relevance to Our Project

- Get scale factors, and deform images for 3D reconstruction.

- If using a non-rigid probe, can get the position of the tip.

Questions?