Towards Designing an Optical-Flow Based Colonoscopy Tracking Algorithm: A Comparative Study

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

[a]Imaging Biomarkers and Computer-Aided Diagnosis Laboratory,
Radiology and Imaging Sciences,
National Institutes of Health Clinical Center,
Bethesda, MD, 20892;
[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

Automatic co-alignment of optical and virtual colonoscopy images can supplement traditional endoscopic procedures, by providing more complete information of clinical value to the gastroenterologist. In this work, we present a comparative analysis of our optical flow based technique for colonoscopy tracking, in relation to current state of the art methods, in terms of tracking accuracy, system stability, and computational efficiency.

Our optical-flow based colonoscopy tracking algorithm starts with computing multi-scale dense and sparse optical flow fields to measure image displacements. Camera motion parameters are then determined from optical flow fields by employing a Focus of Expansion (FOE) constrained egomotion estimation scheme. We analyze the design choices involved in the three major components of our algorithm: dense optical flow, sparse optical flow, and egomotion estimation. Brox's optical flow method,[1] due to its high accuracy, was used to compare and evaluate our multi-scale dense optical flow scheme. SIFT[6] and Harris-affine features[7] were used to assess the accuracy of the multi-scale sparse optical flow, because of their wide use in tracking applications; the FOE-constrained egomotion estimation was compared with collinear,[2] image deformation[10] and image derivative[4] based egomotion estimation methods, to understand the stability of our tracking system.

Two virtual colonoscopy (VC) image sequences were used in the study, since the exact camera parameters(for each frame) were known; dense optical flow results indicated that Brox's method was superior to multi-scale dense optical flow in estimating camera rotational velocities, but the final tracking errors were comparable, viz., 6mm vs. 8mm after the VC camera traveled 110mm. Our approach was computationally more efficient, averaging 7.2 sec. vs. 38 sec. per frame. SIFT and Harris affine features resulted in tracking errors of up to 70mm, while our sparse optical flow error was 6mm. The comparison among egomotion estimation algorithms showed that our FOE-constrained egomotion estimation method achieved the optimal balance between tracking accuracy and robustness. The comparative study demonstrated that our optical-flow based colonoscopy tracking algorithm maintains good accuracy and stability for routine use in clinical practice.

Keywords: Colonoscopy Tracking, Optical Flow, Egomotion Estimation

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

Medical Imaging 2013: Image-Guided Procedures, Robotic Interventions, and Modeling, edited by David R. Holmes, Ziv R. Yaniv, Proc. of SPIE Vol. 8671, 867103 · © 2013 SPIE · CCC code: 1605-7422/13/$18 · doi: 10.1117/12.2007956

1. INTRODUCTION

The co-alignment of pre-segmented virtual and optical colonoscopy images can bring pre-detected polyp information in the virtual colonoscopy data during an optical colonocopy procedure, thus supplementing the clinical value of routine endoscopic procedures. We have developed an optical flow based colonoscopy tracking algorithm[5] to address this problem. The objective of this work is to validate our algorithm design through a comparative study with current state of the art approaches.

Tracking accuracy validation is critical to demonstrate the feasibility of an endoscopy tracking algorithm for routine clinical use. Existing algorithms[3,8,9] developed phantom experiments to serve this purpose. However, phantom validation only provides overall tracking accuracy while the influence of different algorithm components on tracking accuracy is unknown. We examine this carefully in this work, so as to get a deeper understanding of the performance of the tracking algorithm and for the future improvements to the technique.

Our tracking algorithm consists of three major components: multi-scale dense optical flow computation, sparse optical flow computation, and a FOE-constrained egomotion estimation. Brox's method[1] was chosen to compare and evaluate our choice of multi-scale dense optical flow; SIFT[6] and affine-Harris[7] features were selected to understand the choice of sparse optical flow; collinear,[2] image deformation,[10] and image derivative[4] based egomotion estimation methods were compared with our FOE-constrained approach. The comparative study demonstrated that the optical-flow based colonoscopy tracking algorithm made reasonable choices for tracking colonoscopy images, with respect to tracking accuracy, computational efficiency, and system stability. We briefly describe our optical flow based colonoscopy tracking algorithm, and the algorithms used in the comparative study.

2. METHODOLOGY

In this section, we briefly describe our optical flow based colonoscopy tracking algorithm, and the state of the art algorithms used in the comparative study.

2.1. Optical-flow Based Colonoscopy Tracking

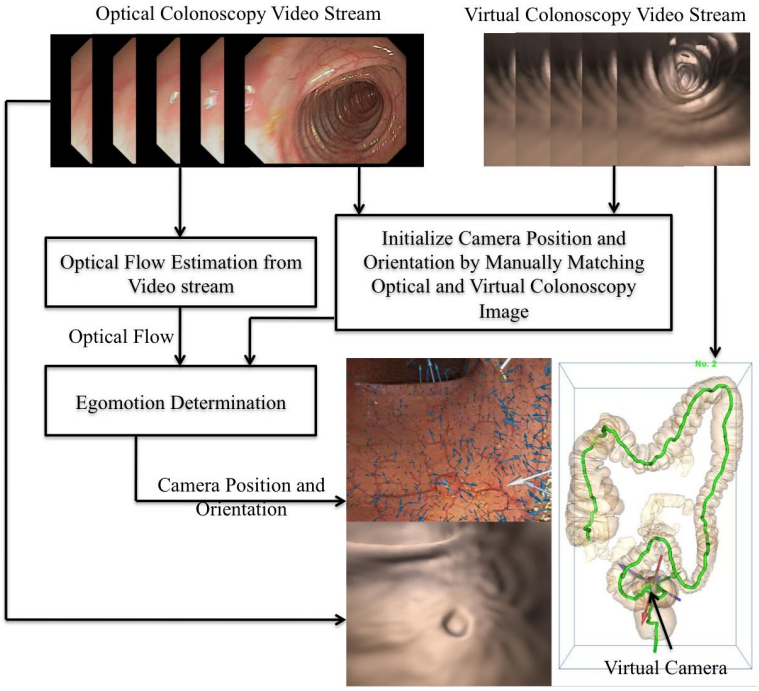


Figure 1. The colonoscopy Tracking Algorithm.

Fig. 1 illustrates the optical-flow based colonoscopy tracking algorithm. Optical and virtual colonoscopy images are initialized by manually adjusting position and orientation of the virtual camera. When a new optical colonoscopy image is imported into the tracking system, a multi-scale strategy is used to compute sparse and dense optical flow fields to measure image displacements. Dense optical flow is then subdivided into several image regions and the flow difference in each region points to the Focus of expansion(FOE). Our egomotion estimation is thus also a subdivision method. Camera translation and rotation parameters are then separately estimated by using the FOE and sparse optical flow field. Optical and virtual colonoscopy images are finally aligned by employing the estimated camera motion parameters to displace the virtual camera. Complete details of the colonoscopy tracking algorithm can be found in our earlier work.[5]

In this work, the comparative study is concerned with assessing multi-scale dense optical flow, multi-scale sparse optical flow, and the FOE-constrained egomotion estimation.

2.2. Dense Optical Flow

Brox's algorithm[1*] was chosen for comparison because of its high accuracy. The computational framework was established upon a variational function, and optical flow was estimated by using the Euler-Lagrange equation to minimize this function. To achieve high accuracy, the variational function employs (1) intensity and gradient constancy models, (2) a coarse-to-fine computation strategy, (3) non-quadratic minimization method, and (4) an image warping technique to estimate optical flow between successive images.

2.3. Sparse Optical Flow

SIFT features[6] and affine-Harris features[7] were chosen for sparse optical flow evaluation because they are widely used in applications of tracking and object recognition.

The SIFT algorithm begins by detecting a set of local 3D extrema in the scale-space pyramid built with difference-of-Gaussian (DoG) filters; this is followed by removing feature candidates with strong edge response. The descriptor for each detected feature is represented by a 3D histogram of gradient locations and orientations in a region centered at the current feature point, and each bin is weighted by gradient magnitude. Thus, two sets of SIFT feature points can be detected in two colonoscopy images, and SIFT descriptors are built in their neighborhoods. Nearest-neighborhood search[6] is used to determine corresponding pairs between the two feature point sets.

The affine-Harris detector begins by first detecting a set of interest points through the Harris matrix and the image scale is determined by the Laplace-of-Gaussian (LoG). The local region's affinity of each detected feature point is estimated through the second moment matrix. Then the local region is iteratively warped into an affine-invariant image region using the estimated affinity. Finally, the SIFT descriptor of each feature is built on the transformed region instead of original image patch. Affine-Harris based image matching is similar to SIFT, except each SIFT feature descriptors are build on the affine-invariant regions.

2.4. Egomotion Estimation

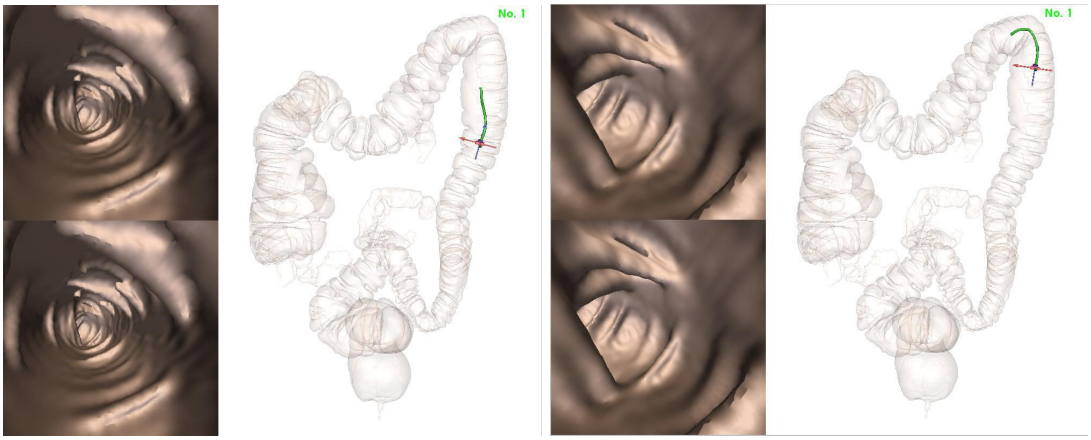
Collinear,[2] image deformation,[10] and image derivative based egomotion estimation methods[4] were chosen to evaluate the stability of our FOE constrained egomotion estimation.

Collinear method is also a FOE based egomotion estimation. It is based on the observation that the FOE stays at a line where the sum of optical flow vectors of all points is zero. FOE can thus be determined by searching a set of lines with such a property and choosing their intersection. After FOE is determined, camera translation and rotation parameters can be estimated. Image deformation method exploits visual angle changes to measure the image motion. Camera translation and rotation computation can be naturally isolated since visual angle changes are only dependent on the camera translation. After camera translational velocity is computed, camera rotational velocity can be sequentially estimated. Image derivative based egomotion estimation method imports camera motion parameters into the image intensity constancy model to avoid feature computation, and it directly estimates camera motion parameters by measuring image derivative changes.

*Received the Longuet-Higgins Best Paper Award, ECCV 2004.

3. COMPARISON RESULTS

Two virtual colonoscopy sequences shown in Fig. 2 were used for comparative study, (facilitating comparison to ground truth), (a) a 420 frame sequence in the descending colon, with mostly translational motion along Z, and (b) a 480 frame sequence that sharply rotates from the descending to transverse colon, at about 40° per sec.



(a) (b)

Figure 2. The first frame of two example VC sequences. (a) A descending colonoscope image sequence is used to evaluate the tracking algorithm when the dominant motion is translation. Here, the top left image means the input image, the bottom left image is the tracked image, and the right image is the external view. A green centerline segment in this view represents the trajectory of the actual camera movements. (b) A colonoscopy sequence with a sharp turn between descending and transverse colons is chosen to validate tracking algorithms when the main motion is rotation.

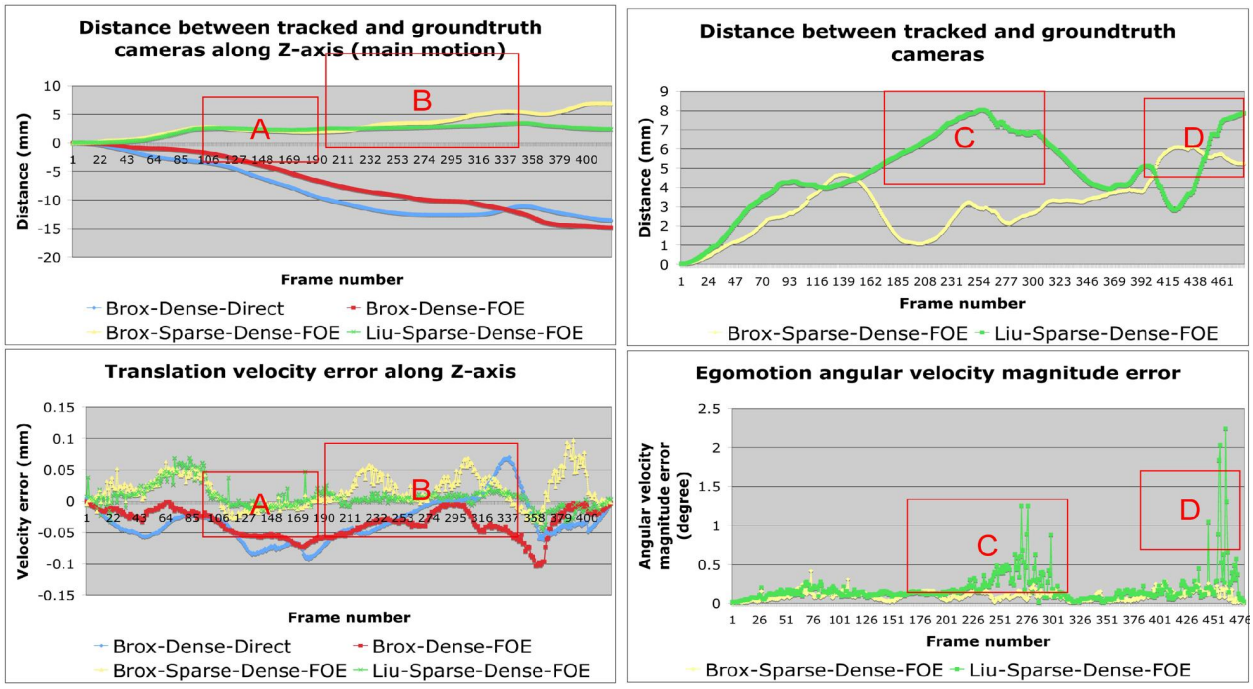
3.1. Dense Optical Flow: Sensitivity and Accuracy

Four different scenarios were used for dense optical flow comparison,

1. Dense optical flow using Brox's algorithm, directly solve camera translation and rotation parameters.
2. Dense optical flow using Brox's algorithm, followed by FOE determination.
3. Dense optical flow using Brox's algorithm, sparse feature points using our method, use dense flow at only the sparse feature points to compute motion parameters.
4. Our optical flow based tracking algorithm(Fig. 1).

Figs. 3(a) and 3(b) illustrate error curves for the two sequences. The accumulated Euclidean distance between the tracked and ground truth cameras (top images), and velocity magnitude errors (bottom images), are plotted over each sequence. Fig. 3(a) shows that methods 3 and 4 (yellow, green curves) have the smallest accumulated error (results are similar for the second sequence, but red, blue curves not shown). This is directly attributed to the *use of sparse flow*, which has significant impact on accuracy. Use of the entire dense flow field, or directly solving camera translation and rotation parameters results in large accumulated errors(20mm after 400 frames); see also corresponding translational velocity error curves. Also region B(yellow curves) shows the velocity error cycling; this is due to optical flow errors in homogeneous regions; Brox's algorithm estimates flow incorrectly in these regions, while our subdivision method ignores these regions. Fig. 3(b) illustrates a sequence moving around a sharp turn. The larger errors in our method directly arises from the rotational flow error; Brox's method keeps the rotational errors very low (regions C, D). However, the tracking errors are still within 10mm for both methods after 480 frames.

Finally, use of the more accurate optical flow method of Brox *et al.,* comes with a high computational cost: 38 sec. vs. 7.2 sec. per frame for our method.

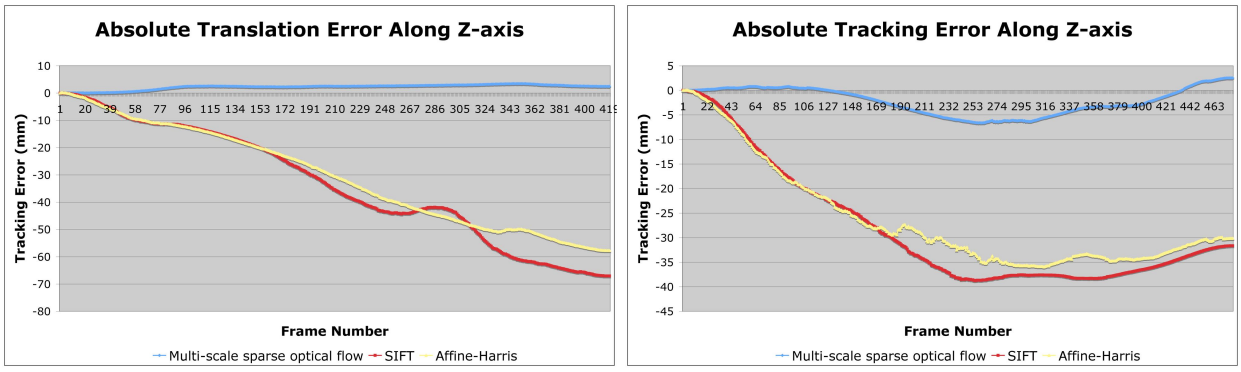


(a) (b)

Figure 3. Impact of accuracy of dense optical flow on the tracking algorithm. (a) A sequence with mostly Z translation. Use of sparse flow from interest points(green, yellow curves) helps keep errors small, (b) Sequence with a sharp rotation of about 40°. High errors (green curve) due to rotation estimation errors in our method compared to Brox's method (yellow curve). Error is within 10mm for both methods at end of sequence.

3.2. Sparse Optical Flow: Sensitivity and Accuracy

The comparison of sparse optical flow in Fig. 4 shows that the tracking errors using multi-scale sparse optical flow (along the Z direction in this case) are within 5-8mm, while SIFT and the affine-Harris based methods are 40-70mm.



(a) (b)

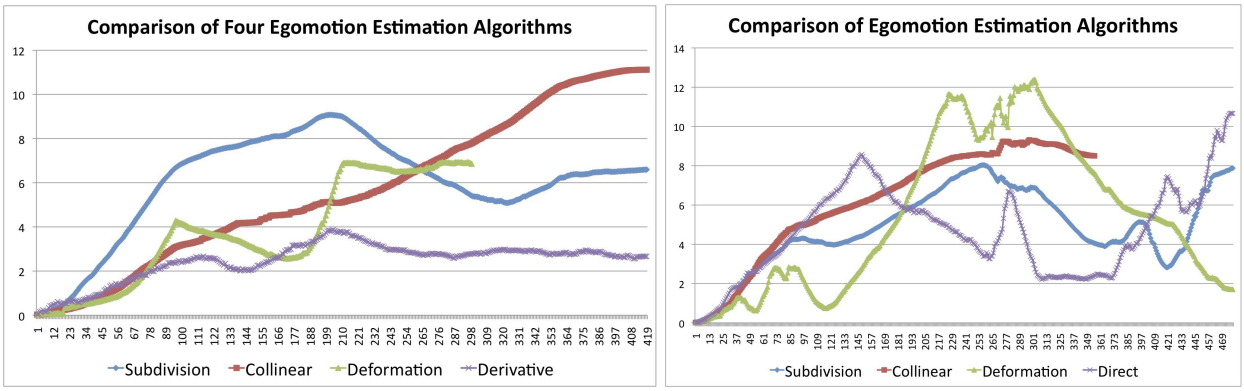
Figure 4. Impact of sparse optical flow on the tracking algorithm. Multi-scale sparse flow (green curves) helps to keep tracking errors within 10mm on two example sequences, while the image matching strategy from SIFT (red curves) and affine Harris descriptors (yellow curves) produces large translation errors: up to 70mm and 40mm respectively on the two sequences.

There are 3 primary reasons:

1. Multi-scale sparse optical flow is a differential technique that can achieve sub-pixel accuracy.
2. SIFT descriptors built on a certain sized image patch are good for large image motion estimation, but fail in case of small displacement(1 or 2 pixels), due to misclassification of feature points that are close to each other.
3. There is no guarantee of finding true corresponding points based on SIFT feature matching.

3.3. Egomotion Estimation: Stability and Accuracy

The comparison of egomotion estimation methods are presented in Fig. 5. In Fig. 5(a), image derivative based egomotion estimation generates the smallest error because image intensity constancy model is well satisfied for small translations of the camera. However, it generates the largest error in Fig. 5(b), as large camera rotation causes significant intensity variance. Image deformation based method requires sparse image features to be uniformly distributed in the image plane, so the visual angles can be accurately formulated through Delaunay triangulation. But feature points are clustered near the image center at frame 300, which results in tracking failure. Collinear method searches a set of lines where the sum of flow vectors equals zero. The intersection of these lines is *FOE*. However, rotation dominates the second sequence, making it more difficult to find such lines. Consequently, collinear method becomes unstable at frame 358.



(a) (b)

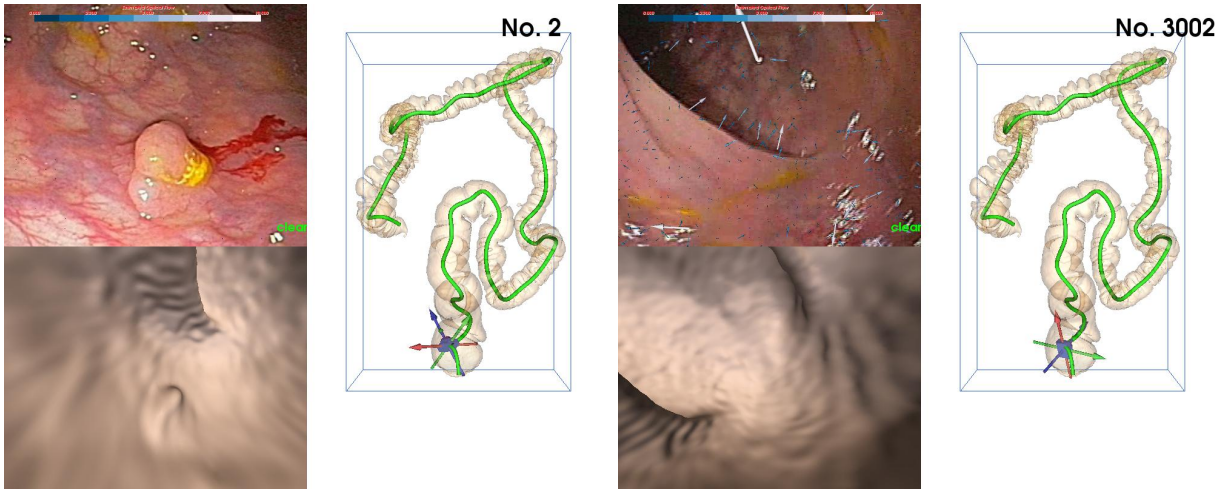
Figure 5. Comparison of camera displacement errors between collinear, image deformation, image derivative based egomotion estimation and our methods.

The second sequence also shows that the subdivision method generates the smallest rotation error. We can conclude that the subdivision method is more accurate and stable, for estimating slow and fast camera motion.

4. CONCLUSIONS AND DISCUSSION

In this work, we performed a comparative study to understand the design choices of our optical flow based colonoscopy tracking algorithm, with respect to multi-scale dense optical flow, multi-scale sparse optical flow, and FOE-constrained egomotion estimation. Brox's method[1] was chosen to understand the influence of dense optical flow on the final tracking results due to its high accuracy. The comparative study indicated that tracking errors of both methods were under 10mm after the camera traveled a distance of 110mm. However, our method was computationally more efficient, taking 7.8s vs. 38s per frame. The results of sparse optical flow showed that the tracking errors of SIFT[6] and affine-Harris[7] features were up to 70mm while sparse optical flow error was about 8mm. The comparison between our FOE-constrained egomotion estimation and collinear,[2] image deformation,[10] and image derivative based egomotion estimation methods[4] showed that our method achieved an optimal balance between tracking accuracy and stability.

The comparative study has demonstrated that our optical flow based approach is an appropriate choice for co-aligning optical and virtual colonoscopy images, based on tracking accuracy, computational efficiency, and system stability. Fig. 6 gives the tracking results on an optical colonoscopy image sequence with 3000 frames in the rectum colon. Note that both polyps and colon folds are accurately tracked in this sequence, thanks to multi-scale sparse and dense optical flow as well as subdivision egomotion estimation constrained by focus of expansion. These results demonstrate that our design is a robust and accurate scheme to track optical colonoscopy images.



(a) (b)

Figure 6. Tracking results on an actual rectum colonoscopy image sequence.

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. N. da Vitoria Lobo and J.K. Tsotsos. Computing egomotion and detecting independent motion from image motion using collinear points. *Computer Vision and Image Understanding*, 64:21–52, 1996.
3. F. Deligianni, A. J. Chung, and GZ Yang. Non-rigid 2d-3d registration with catheter tip em tracking for patient specific bronchoscope simulation. In *Proc. of MICCAI*, pages 281–288, 2006.
4. Berthold K.P. Horn and E.J. Weldon Jr. Direct methods for recovering motion. *Intrernational Journal of Computer Vision*, 2:51–76, 1988.
5. J. Liu, K.R. Subramanian, T. Yoo, and R. Uitert. A stable optic-flow based method for tracking colonoscopy images. In *Proc. of Mathematical Methods in Biomedical Image Analysis*, pages 1–8, 2008. June 27-28, Anchorage, Alaska.
6. D. Lowe. Distinctive image features from scale-invariant keypoints. *Intrernational Journal of Computer Vision*, 60(2):91–110, 2004.
7. K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *Intrernational Journal of Computer Vision*, 60(1):63–86, 2004.
8. 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.
9. L. Rai, J. Helferty, and W. 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.
10. C. Tomasi and J. Shi. Image deformations are better than optical flow. *Mathematical and Computer Modelling Journal*, 24(5/6):165–175, 1996.