

Region Flow: A Multi-stage Method for Colonoscopy Tracking

Jianfei Liu¹, Kalpathi R. Subramanian¹, and Terry S. Yoo²

¹ Department of Computer Science,
The University of North Carolina at Charlotte, Charlotte, NC, 28223, USA
jliu1@uncc.edu, krs@uncc.edu

² Office of High Performance Computing and Communications,
National Library of Medicine, NIH, Bethesda, MD, 20892, USA
yoo@nlm.nih.gov

Abstract. Co-located optical and virtual colonoscopy images provide important clinical information during routine colonoscopy procedures. Tracking algorithms that rely on image features to align virtual and optical images can fail when they encounter blurry image sequences. This is a common occurrence in colonoscopy images, when the endoscope touches a wall or is immersed in fluid. We propose a *region-flow* based matching algorithm to determine the large changes between images that bridge such interruptions in the visual field. The region flow field is used as the means to limit the search space for computing corresponding feature points; a sequence of refining steps is performed to identify the most reliable and accurate feature point pairs. The feature point pairs are then used in a deformation based scheme to compute the final camera parameters. We have successfully tested this algorithm on four clinical colonoscopy image sequences containing anywhere from 9-57 consecutive blurry images. Two additional tabletop experiments were performed to quantitatively validate the algorithm: the endoscope was moved along a slightly curved path by 24 mm and along a straight path by 40 mm. Our method reported errors within 1-5% in these experiments.

Keywords: colonoscopy, tracking, failure recovery, image matching.

1 Introduction

The simultaneous use of pre-segmented virtual and optical colonoscopy images during routine endoscopic procedures provides useful clinical information to the gastroenterologist. Tracking algorithms must be employed to keep both image sequences aligned throughout the procedure. In medical images this presents a number of challenges: images can become blurry (endoscope touching a wall, fluid immersion), bright areas or tools may appear, etc. In these instances, endoscopic images over short periods of time may be devoid of features, causing tracking algorithms to fail. The goal of this work is to investigate new methods to skip such interruptions in the visual field and seamlessly continue tracking. Optical colonoscopy image sequences (our application of interest here) are particularly

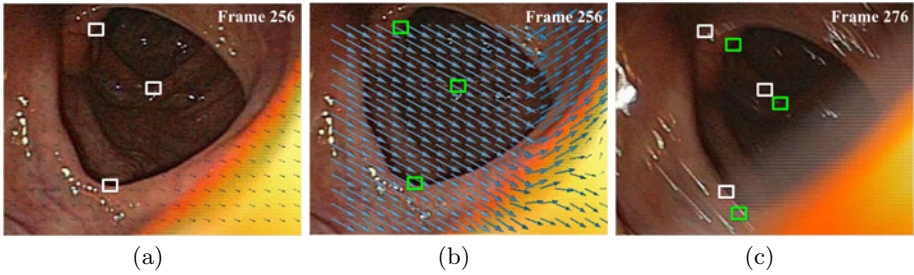


Fig. 1. Region flow vs. optical flow for describing large motion, (a) source image with overlaid optical flow vectors, (b) source image with overlaid region flow vectors, (c) target image after a 20 frame blurry sequence. White and green squares in the target image represent 3 selected regions in the image, and correspond to the white and green squares in the source images, after application of optical and region flow vectors. Region flow does a better job tracking the image motion. The lengths of the vectors in the source images represent the magnitude of the motion velocity.

challenging, due to deformation and other artifacts, and necessitate very robust algorithms.

The problem we study in this work can be stated as follows: given tracked images prior to a blurry image sequence, determine camera parameters (translation and rotation velocities) after the sequence, so as to continue tracking. There have been two general approaches to handling blurry images in endoscopic sequences: the use of a magnetic sensor[1,2] for tracking bronchoscopy images, and recovery from failures. It is unclear that this will work well with colonoscopy images, due to the more severe deformation effects, making sensor calibration a difficult task. An alternate approach is to use computer vision algorithms to locate and match corresponding features along the temporal dimension[3]. Termed wide-baseline matching[4,5], these methods find temporal correspondence through local comparison of feature descriptors, and results depend on the distinctness of the image features. Optical flow has also been used to track images[6], but typically does not work well for large changes between images. Fig. 1 illustrates an example, with Figs. 1(a),1(c) representing optical images bridging a blurry image sequence. The white squares in these images represent corresponding pairs generated by the optical flow field; they do not match up with the green squares, which roughly represent the positions of the true corresponding pairs.

In this work we present a computer vision algorithm to accurately match corresponding features representing large motion between images, to improve the robustness of tracking algorithms. Central to this approach is the use of *region flow*, a dense feature matching strategy that provides a basis and framework for understanding large motion. As indicated by the work of Brox et al.[7] and Liu et al.[8], dense feature correspondences can improve the accuracy of many vision applications, such as structure-from-motion, object recognition, and image retrieval. In the problem addressed here, region flow computation is a key and novel step of our method, for two reasons, (1) permitting the algorithm to limit the search space for accurately identifying corresponding features, and

(2) point-to-point correspondence relies on the intensity constancy model, which is generally not true for large motion; instead, invariance of a region's intensity distribution is a more reasonable assumption. The region flow matching step is followed by a sequence of refinements, that lead towards accurate computation of corresponding feature point pairs, and involves region-to-region and point-to-point matching steps, and false feature rejection. Finally, an image deformation based egomotion estimation method is used to recover camera parameters; we constrain the 2D-2D image deformation based egomotion determination problem[9] to a 3D-2D pose estimation problem by using depth values from a colon model. Fig. 2 details the various steps of our region flow based approach to handling large motion and thereby provide the means to skip blurry image sequences.

We demonstrate initial results using this method on four clinical colonoscopy sequences containing blurry images; two of these are in the sigmoid colon and the remaining two in the ascending colon. A table top validation experiment was also performed to quantify the accuracy of the method by acquiring two sequences in which the endoscope was moved 24 mm along a slightly curved path and 40 mm along a straight path.

2 Methods

The flow of our algorithm for recovering motion parameters after a blurry image sequence is described in Fig. 2. We describe in more detail the major steps of our method.

2.1 Region Flow Computation

Computing region flow is the key to efficiently determining feature point correspondences between images representing large motion.

Let $I_1(x, y)$ and $I_2(x, y)$ be a pair of normalized images, with $\vec{u} = (u_x, u_y)$ representing the region flow vector at point (x, y) . The similarity between two regions of $I_1(x, y)$ and $I_2(x, y)$ with relative displacement \vec{u} can be measured by Normalized Cross-Correlation(NCC) and given by

$$NCC(x, y, u_x, u_y) = \iint I_2(x + u_x, y + u_y)I_1(x, y)dxdy \quad (1)$$

Similar to optical flow computation[10], we use a global energy function to compute region flow, within a minimization framework.

$$E(u_x, u_y) = \iint \underbrace{\min(|1.0 - NCC(x, y, u_x, u_y)|, \alpha)}_{\text{Data constraint}} + \lambda \underbrace{\min((|\nabla u_x| + |\nabla u_y|), \beta)}_{\text{Smoothness constraint}} dxdy \quad (2)$$

where α and β are truncation values to prevent over-smoothing. λ is a parameter to balance data and smoothness constraints.

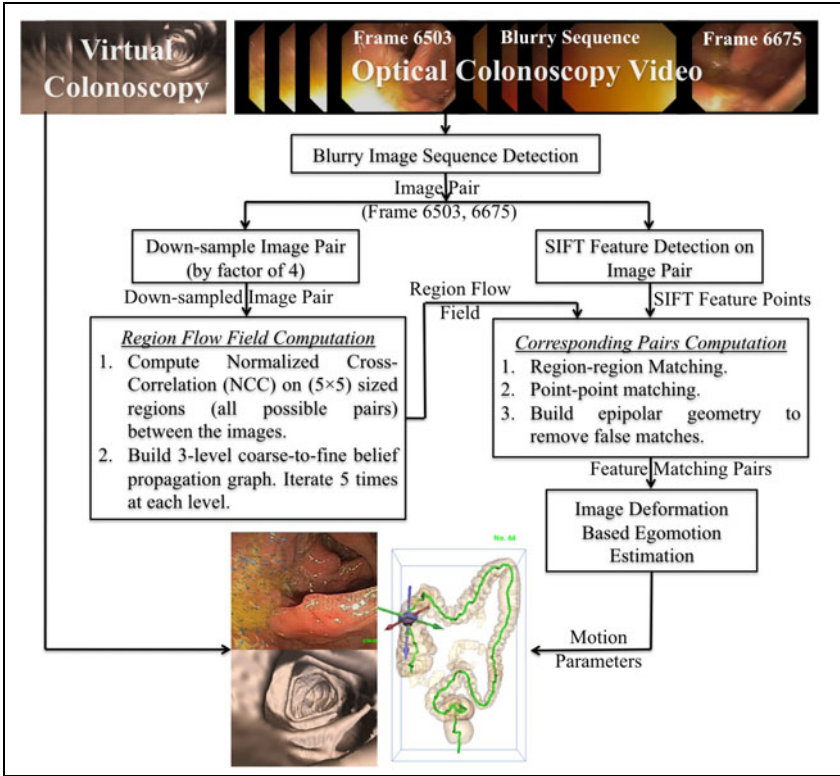


Fig. 2. Region flow based algorithm for recovering motion parameters after a blurry image sequence

Implementation. In general, region flow computation involves matching regions in the source image with regions in the target image, at every pixel, an $O(n^4)$ computation, for $N \times N$ sized images. In our implementation, we reduce the computation by (1) downsampling images by a factor of 4 in each dimension, and (2) restrict the largest image motion (we allow up to 150 pixel displacement on a 500 by 400 image), so that the corresponding search space in the target image is reduced. The minimization procedure of Eq. 2 is performed using the coarse-to-fine belief propagation procedure described in Felzenwalb[11]. The minimization results in a set of region flow vectors that provide a good approximation to the image motion. Fig. 1(b) illustrates an example colonoscopy image with overlaid region flow vectors. The region flow vectors follow the image motion between the two images, Figs. 1(b) and 1(c).

2.2 Corresponding Pairs Computation

Region-to-Region Matching. In this step, corresponding regions are identified using the region flow field and a local matching procedure. A set of stable feature points are detected by the SIFT algorithm[5] in the source and target

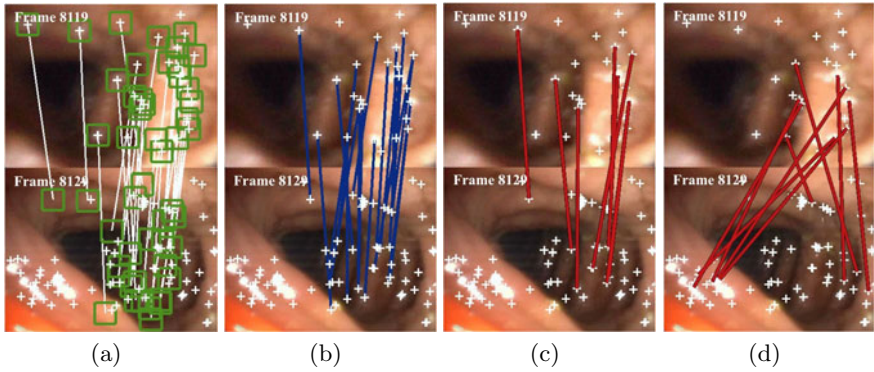


Fig. 3. Corresponding Pairs Computation. Top and bottom images represent images before and after the blurry image sequence, (a) Region-to-Region matching. Green squares indicate the matched regions using the region flow field. Local search using NCC is performed to find the best region pair (b) Point-to-Point feature matching. Using SIFT descriptor as a metric, the best SIFT feature point pair is determined between source and target regions. (c) False feature match rejection using epipolar geometry, (d) local matching using only (locally defined) SIFT feature descriptors, illustrating significant errors.

images. The corresponding regions in the target image are identified using the region flow vectors and a local neighborhood search. In Fig. 3(a), the green squares joined by the white lines represent corresponding regions containing at least one SIFT feature point in the source image and 0 or more SIFT feature points in the target region. In the implementation, the mapped region is locally adjusted using NCC as a metric to find the best region match.

Point-to-Point Feature Matching. In this step, each corresponding region pair is refined to a corresponding point pair. If the target region does not contain a SIFT feature point, it is removed. For target regions with multiple SIFT feature point candidates, the candidate with the closest SIFT descriptor (a distance metric) is chosen as the best candidate. Fig. 3(b) illustrates the selected feature point pairs after this step.

False Feature Match Rejection. With the chosen feature point pairs, epipolar geometry is built using the RANSAC algorithm[4]. Outliers that do not satisfy the epipolar geometry constraints are removed, as seen in Fig. 3(c).

Finally, Fig. 3(d) illustrates the same example using just SIFT feature point matching. It can be clearly seen that the lack of global motion information results in significant mismatches.

2.3 Image Deformation Based Egomotion Estimation

In the final step, we estimate the camera motion parameters of the image after the blurry images. We use a deformation based method.

In Fig. 4, the visual angle θ_1 of two world coordinate points P and Q from the camera projection center O_1 is defined as the angle between the projection rays of P and Q . If $\vec{v}_{o_1p_1}$ and $\vec{v}_{o_1q_1}$ are the normalized projection rays, the disparity between θ_1 and θ_2 can be expressed in terms of vectors orthogonal to $\vec{v}_{o_1p_1}$ and $\vec{v}_{o_1q_1}$ as follows(see [9]):

$$\theta_2 - \theta_1 = \vec{T} \cdot (dP \vec{v}_{o_1p_1}^\perp + dQ \vec{v}_{o_1q_1}^\perp) \quad (3)$$

where $dP = \frac{1}{|P-O_1|}$, and $dQ = \frac{1}{|Q-O_1|}$ and $\vec{T} = (T_x, T_y, T_z)$ is the translation velocity. Eq. 3 depends only on \vec{T} . As depth values can be obtained from the virtual colon model, the computation of \vec{T} is linearized. Once \vec{T} is known, we can compute the *Focus of Expansion*, $FOE = (fT_x/T_z, fT_y/T_z)$. Rotation velocities $\vec{R} = (R_x, R_y, R_z)$ are then computed by embedding the FOE in a polar coordinate system, as described by Reiger[12].

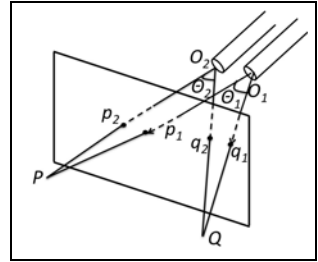


Fig. 4. Egomotion Estimation

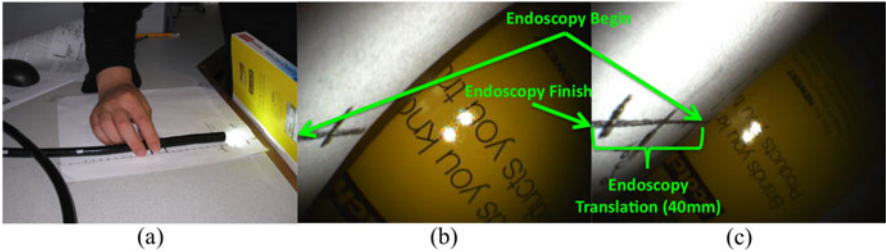


Fig. 5. Validation. A table top experiment was performed to move the endoscope a fixed distance. The resulting video was used to test the algorithm. (a) The table top experimental setup: the endoscope was placed on a surface with distance markings and moved along a predefined path, (b,c) Acquired images at the beginning and end of the sequence.

3 Experimental Results

We have tested our algorithm on four clinical colonoscopy sequences, distributed in different segments of the colon. We have also performed a validation experiment to quantify the accuracy of the algorithm.

3.1 Validation Experiment

Our goal in this experiment was to move the endoscope precisely a certain distance and acquire images between the end points. Since depth of the colonoscope from its starting point was needed, a flat surface was held at right angles at the edge of the desk, thus the depth is the same for all points in the projected image.

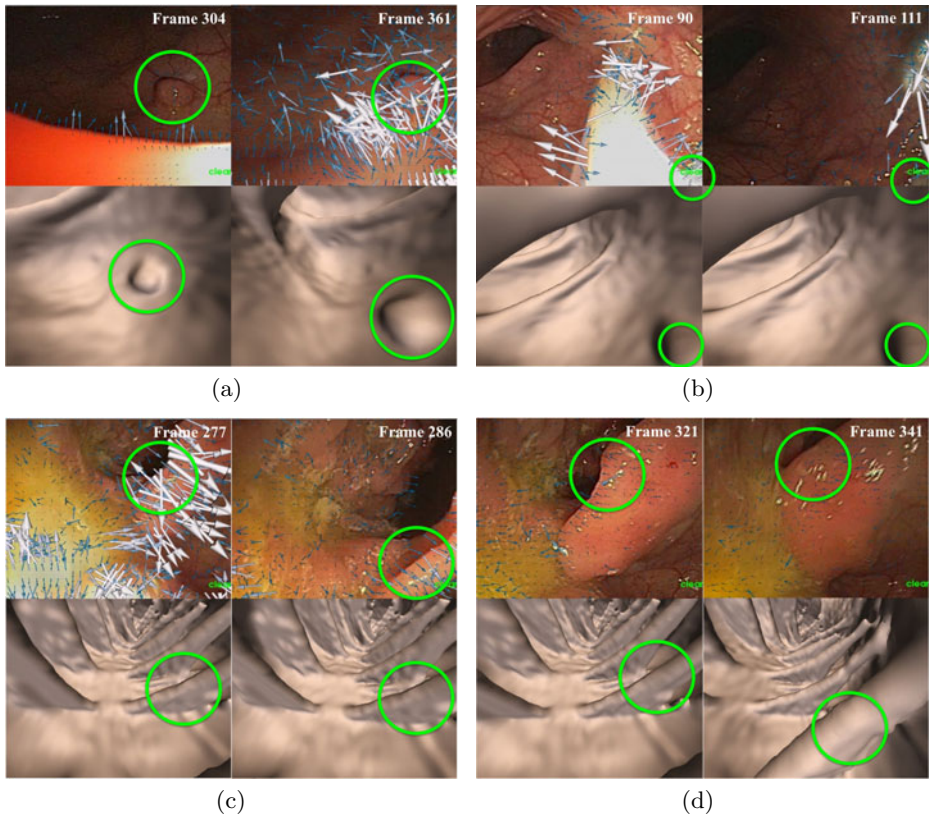


Fig. 6. Results on 4 Colonoscopy Sequences. OC-VC image pair before and after blurry sequences, (a) 520 polyp surgery sequence in sigmoid colon with a 57 image blurry sequence (b) 160 image polyp removal sequence in sigmoid colon with a 21 image blurry sequence, (c,d) 450 image sequence with 2 blurry image sequences of 9 and 19 images. The tracking system tracked through both blurry images sequences successfully.

Fig. 5 illustrates the experimental setup and image pair of one of the acquired sequences. Two sequences were tested with this setup, (1) colonoscope moved along a straight path by 40 mm, (2) colonoscope moved along a slightly curved path, with the end to end (Euclidean) distance of 24mm. Analyzing these two sequences using our algorithm resulted in distances of 39.6mm and 22.96mm respectively.

3.2 Clinical Colonoscopy Experiments

Fig. 6 illustrates 4 example colonoscopy sequences with blurry image sequences. The top rows illustrate the optical images before and after the blurry images. The corresponding virtual colonoscopy images are in the bottom rows. Regions marked by green circles indicate corresponding features to establish accuracy.

Experiment 1: Polyp Surgery in the Sigmoid Colon. This sequence contains 520 images, with a blurry image sequence from frame 304 to 361, due to the colonoscope touching the colon wall. In Fig. 6(a) the polyp can be clearly seen in the OC and VC images, including scale changes in the polyp. The fold in the virtual image is likely due to deformation.

Experiment 2: Polyp Removal in the Sigmoid Colon. This sequence represents the removal of the polyp, and contains 160 images, with a blurry image sequence between 90 and 111. Injection of water (bright area) in the vicinity of the removed polyp caused the blurry image sequence. Though somewhat harder to see, the green circles estimate the location of the polyp quite well in the OC and VC images.

Experiments 3,4: Ascending Colon. This sequence in the ascending colon contains 450 images and contained two blurry sequences, from 277 to 286 and 321-340; in both cases, the colonoscope was very close to a fold. Our algorithm was able to track continuously through the two blurry sequences, as seen by the well aligned OC and VC images in Figs. 6(c),6(d).

Our initial results are very promising. Despite the large changes in images in these sequences and the artifacts (especially deformation) in colonoscopy images, the region flow field accurately captures the global motion characteristics, easing the corresponding pairs computation. In all of these experiments, it is possible to identify features (folds, polyps, etc.) that provides qualitative accuracy and confidence in the tracking system. It is worthy to note that our tracking system tracked continuously through the two blurry image sequences in Figs. 6(c),6(d) without interruption.

4 Conclusions

We have presented a region flow based algorithm to handle large motion induced changes in colonoscopy video; this frequently happens when the colonoscope touches a wall or fold, or is immersed in fluid. The region flow field provides the computational basis for accurate and robust corresponding pairs computation, which in turn permits estimating camera parameters. We show through a validation experiment and four clinical colonoscopy sequences the effectiveness of our algorithm to keep the tracking system functioning as it encounters blurry image sequences; in our experiments, blurry image sequences ranged from 9 to 57 consecutive images. We are currently looking at two issues in improving our method, (1) carefully examine the computational considerations in computing region flow, and (2) improve the robustness of the egomotion estimation.

References

1. Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 543–555. Springer, Heidelberg (2005)

2. Deligianni, F., Chung, A., Yang, G.Z.: Non-rigid 2d-3d registration with catheter tip EM tracking for patient specific bronchoscope simulation. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 281–288. Springer, Heidelberg (2006)
3. Sullivan, J., Carlsson, S.: Recognizing and tracking human action. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 629–644. Springer, Heidelberg (2002)
4. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. *International Journal of Computer Vision* 60(1), 63–86 (2004)
5. Lowe, D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
6. Liu, J., Subramanian, K., Yoo, T., Uitert, R.: A stable optic-flow based method for tracking colonoscopy images. In: Proc. of MMBIA, Anchorage, Alaska (2008)
7. Brox, T., Bregler, C., Malik, J.: Large displacement optical flow. In: Proc. of CVPR 2009, Miami, FL (June 2009)
8. Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: Sift flow: Dense correspondence across different scenes. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 28–42. Springer, Heidelberg (2008)
9. Zhang, J.: Computing Camera Heading: A Study. PhD thesis, Stanford University (1999)
10. Horn, B., Schunck, B.: Determining optical flow. *Artif. Intell.* 17(3), 185–203 (1981)
11. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. *International Journal of Computer Vision* 70(1), 41–54 (2006)
12. Reiger, J., Lawton, D.: Processing differential image motion. *Journal of the Optical Society of America A* 2(2), 354–359 (1985)