

## Reconstruction of a 3D Virtual Colon Structure and Camera Motion for Screening Colonoscopy

### Authors:

Wallapak Tavanapong <sup>1)</sup>

Dong Ho Hong <sup>1)</sup>

Johnny Wong <sup>1)</sup>

Piet C. de Groen <sup>2)</sup>

Jung Hwan Oh <sup>3)\*</sup>

<sup>1)</sup> Computer Science Department, Iowa State University, Ames, IA 50011, U.S.A.

<sup>2)</sup> Department of Medicine, Division of Gastroenterology, Hepatology and Nutrition, University of Minnesota, Minneapolis, MN 55455, U.S.A.

<sup>3)</sup> Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203, U.S.A.

### \*Correspondence to:

Jung Hwan Oh

Department of Computer Science and Engineering, University of North Texas,  
1155 Union Circle #311366  
Denton, TX 76203, U.S.A.  
Tel: 940-369-7790  
Fax: 940-565-2799

### E-mail addresses:

tavanapo@iastate.edu

wong@iastate.edu

Junghwan.Oh@unt.edu:

**Keywords:** Endoscopic image analysis; 3D reconstruction algorithms; visualization; screening colonoscopy

### Abstract

#### Background/Purpose

Optical colonoscopy is the gold standard for detection and prevention of Colorectal cancer. However, any lesion outside the field of view of the camera during colonoscopy is missed, and can develop from a polyp into a cancer or from a small cancer into a large, metastatic cancer. Hence, viewing the colon mucosa as much as possible is desirable. The ultimate goal is to use the reconstruction results to mark the area likely unobserved during colonoscopy where polyps may be hidden.

#### Materials and Methods:

This paper presents a novel method for reconstructing (i) a 3D virtual colon structure from a sequence of 2D colonoscopic images and (ii) the endoscope camera navigation path. Unlike existing work that focus on reconstruction of accurate colon surface for a computer-aided surgery, this work focuses on estimating the alignment of the colon haustral folds, thickness and protrusion of the folds, and the camera path for computer-aided screening colonoscopy to inspect as much as possible of the colon mucosa.

#### Results:

On endoscopy video of a synthetic colon model, we achieved at least 73% and 90% accuracy in estimating the directions of camera translation and rotation motions, respectively. The average percentage of the depth error is about 10% (8.5/85.8) of the average depth of all the folds seen in all the image sequences. The average percentage of the circumference error is about 10% (17/178.2) of the average circumference of all the folds seen in all the image sequences. The results are promising and give further insight to address this challenging problem.

#### Conclusions:

We present the work on reconstruction of a colon structure and endoscope motion from a colonoscopic image sequence for screening colonoscopy. We overcame several challenges such as the scale difference of consecutive reconstructed models, estimating closed fold contours from tracked fold edges, and registering the models. As future work, we will investigate a multi-frame registration approach to further reduce the depth and circumference estimation errors by registering nearby frames that are more similar in fold shape together.

## 1. Introduction

Colorectal cancer (CRC) develops in the mucosa of the largest hollow organ, the colon. Despite the fact that over 14 million colonoscopies are performed each year in the US [1] still about 49,000 people die annually of CRC [2]. The vast majority of these CRCs are thought to develop from missed polyps, especially adenoma-type polyps. The colon is a complex, non-rigid hollow organ with multiple sharp angulations. It is about 150 cm long nested inside the abdominal cavity. Although the core shape is tubular, the organ contains mucosal folds, called haustrae, and a number of angulations, called flexures, that create a shape that is anything but a simple tube. In addition, haustrae and flexures are not evenly disturbed along the longitudinal axis of the colon; haustrae are larger and deeper in the right colon whereas most of the angulations are present in the left colon. Polyps or small cancers are difficult to view when located between or behind haustrae.

Optical colonoscopy (OC) is the gold standard for detection and prevention of CRC. During the insertion phase of colonoscopy, a flexible endoscope with a fish-eye camera at the tip is gradually inserted into the anus and advanced as far as possible inside the colon, preferably all the way to the cecum and the terminal ileum of the small intestine. During the withdrawal phase, the endoscopist should gradually withdraw the endoscope to see as much as possible of the colon mucosa to find abnormalities.

A therapeutic operation (e.g., polypectomy) can be performed during the procedure using instruments. Any lesion outside the field of view (FoV) of the camera during the entire examination is missed and can develop from a polyp into a cancer or from a small cancer into a large, metastatic cancer. Hence, viewing the colon mucosa as much as possible is desirable.

Although the endoscopist has a global concept about the likely shape of the colon, the actual shape and structure of a specific patient's colon can vary widely from this global concept; therefore the endoscopist does not have prior knowledge of a specific patient's colon structure before colonoscopy.

Our ultimate goal is to be able to reconstruct the entire colon structure on the fly and provide feedback of the area likely missed once each major section of the colon has been inspected. The endoscopist can then visit the area. We realize that the ultimate goal is ambitious since there are many frames that cannot be used such as blurry frames, frames showing instruments, stool, and water injection. Camera motion is not constant. The colon has many twists and turns. Due to these challenges, we focus on reconstruction from a short sequence of clear images with slow camera motion in this paper.

Our research problem in a broad sense is similar to the problem of recovering an architectural structure of an indoor environment (e.g., a 3D floor plan of rooms) using images inside the rooms. The recent solution in that domain requires 3D scanning of objects inside the room provided that the distance of the objects from the 3D scanner is at a known fixed position [3]. The solution for this problem is not constrained by execution time. Furthermore, the camera position is known.

Previously, we introduced *Iterative Geometric Region Expansion* [4], a hybrid technique based on image processing and geometric interpolation, to detect closed contours of colon folds from a single colonoscopic image with the lumen seen in the image. Because of the geometric interpolation, our method can handle part of the fold that is flat or partially occluded, which is common. We proposed two methods to reconstruct a 3D virtual colon from a single colonoscopic image [5, 6].

Depth-from-shape (DfS) estimates the depth (the distance from the camera) of the fold contour from shape [5]. DfS requires the fold shape to be close to a perfect circle. We then investigated the depth from brightness intensity (DfI) [6] method, which does not require such assumptions. Like other existing works in colon surface reconstruction, we assume that the colon surface except at specular spots is Lambertian. Prior to the application of DfI, we calibrated the brightness intensities of the pixels on the detected fold contours to account for an uneven distribution of brightness intensities for the pixels of the object surface at the same depth.

In this paper, we present a novel method to reconstruct a 3D virtual colon from a sequence of colonoscopic images. We focus on estimating the alignment of the colon haustral folds, the fold thickness (width) and fold protrusion (height) of the folds, and the camera path, for computer-aided screening colonoscopy to inspect as much as possible of the colon mucosa. Existing 3D reconstruction methods for generic objects such as structure-from-motion [7-14], multiview stereo [15, 16], shape-from-shading [17-39] determine the structure/surface of an object from images of the external view of the structure. For computer-aided surgery, shape-from-shading (SfS) [40], shape-from-motion (SfM) [41], combined SfS and SfM [42], and a motion-based technique [43] were proposed to reconstruct the colon surface, not the overall structure of the colon. Recently, surface reconstruction methods using a hybrid approach have also been proposed [44, 45, 60]. In screening colonoscopy, camera motion is not constant, which makes it difficult to for existing shape-from-motion methods to perform well. Furthermore, commonly used colonoscopes have a single fish-eye camera, not stereo camera. Therefore, reconstruction from stereo views is not possible. Estimation

of the depth of the fold contour from the camera based on intensity is our only solution but it is full of challenges. For instance, points on the same fold contour may not be on the same plane. Different illumination levels happen across subsequent frames. With reconstruction from brightness intensity, we have different sizes (scales) of the reconstructed models that need to be registered into a single virtual colon.

Our contribution is a new reconstruction method by (i) tracking edges of fold contours across frames, which helps reducing errors due to motions unrelated to camera movement such as water bubble or instrument motions, (ii) application of Umeyama's least-squares estimation of transformation parameters [46] to derive transformation matrices of camera translations and rotations as well as scaling factors to register the local reconstructed models of varying scales into a single global model, and (iii) merging the reconstructed models into a single global model, which deals with the flat and occluded part of the fold contours.

On endoscopy video of a synthetic colon model, we achieved promising results of at least 73% and 90% accuracy in estimating the directions of camera translation and rotation motions, respectively. The mean normalized errors of fold depths among three sections of the synthetic colon are from 3% to 20.75%. The mean normalized errors of fold circumferences are from 7.5% to 14.58%. The performance is sensitive to how well the fold contour shape matches the real contour shape. Because it relies on tracking, the method cannot handle very fast motion or abrupt change between successive frames.

The remainder of the paper is organized as follows. Section 2 summarizes the related work. Section 3 presents the proposed method. Section 4 discusses

experimental design and results. Sections 5 and 6 provide discussions and the conclusion and the description of the future work.

## 2. Related Work

In recent years, there are several advances in endoscopic image analysis for different types of endoscopy procedures and for different purposes such as abnormality detection, quality measurements, and content-based retrieval. We limit our discussion to those related to 3D reconstruction and categorize them as follows.

### 2.1. *Virtual colonoscopy assisted colonoscopy*

This category can be divided into two sub-categories. (1) Use of discretized points on the centerline of a reconstructed virtual colon by CT Colonography (CTC) to find corresponding discretized points in the shortest path of the endoscope [47, 48]. The major assumption is that the endoscope during colonoscopy follows the shortest path and the colon shape during CTC and OC is the same. (2) A tracking method of colon folds [49] to align VC and OC images. Temporal volume flow, region flow (matching of pixel blocks instead of individual pixel matching), and incremental egomotion estimation were used. The evaluation was done on rigid colon models built from LEGOs. The method is capable of tracking colon folds in about 3,600 frames and took 2-5 minutes.

### 2.2. *Colon surface reconstruction*

Methods for colon surface reconstruction based on shape-from-shading (SfS) [40], shape-from-motion (SfM) [41], combined SfS and SfM [42], and a motion-based technique [43] are not suitable to reconstruct a tubular colon structure. SfS

would wrongly express the lumen as relatively far surface. SfM requires consistent slow motion and rigid objects to be effective, which is not common during colonoscopy.

### 2.3. *Cylinder structure reconstruction*

The work in [50] models a colon segment as a circular cylinder of 3D circles (not necessarily corresponding to colon folds). Alignment of neighboring circles and distance among them are determined by optimizing smoothness energy difference between the projected flow of the neighboring circles and the actual optical flows between neighboring images. The technique can generate a small 3D colon segment, but has the following inherent limitations. (i) Some colon segments (e.g., transverse colon) are not circular. (ii) The technique assumes that the neighboring circles are not occluded; however, partial occlusion is typical. (iii) It does not model colon fold thickness and fold protrusion, which is important for identifying areas likely unobserved during colonoscopy.

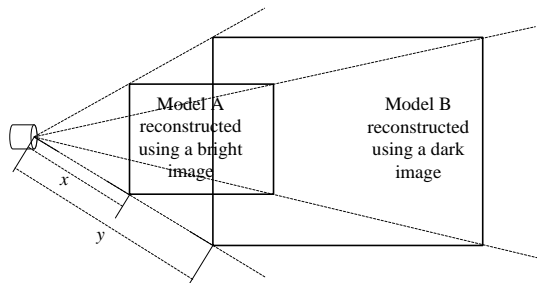
In summary, except our prior work, there is no existing hardware or software technology to create from 2D colonoscopic images a hollow colon structure with important details: fold thickness, protrusion, depth, and alignment among folds.

### 2.4. *Our Prior Work: Depth estimation from intensity*

We assume that the colon surface is Lambertian surface except at specular points. The depth estimation consists of two key steps: (i) brightness intensity calibration and (ii) estimation of the distance from the camera (depth) based on the brightness intensity of each pixel on a fold contour. Intensity calibration is required because surface at the center of the image is brighter than that at the border of the image even

when the surface is at the same distance from the camera [6].

The implication of Dfl is that when an image has lower illumination, the reconstructed colon model is larger since the distance from the camera is farther and the reverse projection vectors of points on the fold contour expand more. Fig. 1 illustrates two different models A and B reconstructed from images with different illumination. Therefore, our 3D registration method must be able to register colon models of different scales, which is atypical for 3D registration.



**Fig. 1.** Impact of illumination on the scales of the colon models

### 2.5. Our Prior Work: 3D virtual colon rendering and the unseen area identified

After the 3D coordinate of each point on each fold contour is estimated, we build two types of surface: (i) surface of colon folds and (ii) surface between folds (or wall surface) using several features extracted from the images and geometrical relations between neighboring folds. We use Cubic Bézier curves [51] for interpolation of the surface and creation of the triangle mesh model.

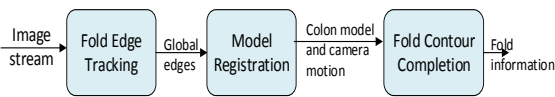
While we can use OpenGL to render the virtual colon easily, OpenGL does not have an API to query for faces of the triangles outside the field of view of the camera. These faces form the unobserved areas. At each frame, given the camera position, up vector, and look at vector, we

identify the unobserved area as follows. For each triangle face of the colon model that has not been marked as inspected in the previous frames, we check three conditions: (i) if the face is facing the camera (at the origin of the coordinate system), (ii) if the face is within the camera field-of-view, and (iii) if there are no obstacles from the face to the endoscope. The first condition is satisfied if the angle difference (in the range of  $[0, 180]$ ) between the normal of the triangle face and the vector from the camera to the center of the triangle face (the line of sight of the face) is greater than  $90^\circ$ . The second condition is satisfied if the angle difference between the line of sight of the face and the camera direction is at most half of FOV of the camera. The last condition is satisfied if the distance of the triangle face from the camera is same as the OpenGL-provided depth (Z-value) of the corresponding pixel in the viewport. Every pixel in a viewport has its own depth value indicating the distance of the closest object from the camera in that direction. If all the three conditions are satisfied, the triangle face is marked as inspected. If any of the three conditions is not satisfied, we count that polygon as an unobserved polygon. The percentage of the unobserved areas of the model is the percentage of the number of unobserved triangle faces to the total number of faces in the model.

## 3. Materials and Methods

We focus on reconstruction of a virtual colon structure and camera motions from a sequence of consecutive endoscopic images with low motion. Fig. 2 shows the three major components: fold edge tracking, model registration, and fold contour completion. The fold edge tracking step establishes the correspondence of edges across frames and assigns them the same global edge ID. The edges are later combined to lengthen the contour of the

colon fold to which the global edge is assigned. This is to leave a small missing portion of the contour as much as possible for estimation. The model registration step derives the camera motion for each frame from the tracked edges. The fold contour completion step forms the closed contour by estimating the missing part of it using computational geometry. After all the fold contours and information are completed, we use our prior methods to render the model and unobserved areas [6].



**Fig. 2.** Components of the system for obtaining fold contour information and camera motions

### 3.1. Fold Edge Tracking

For each frame, we preprocess the image using the pre-processing step of reference [4] that removes images that are unlikely to have fold contours, cuts edges with sharp turns, removes edges that are too short, and keeps only the inner edge of a thick fold. Due to limited space, refer to [4] for more details. We do not use images without colon fold contours for reconstruction because these images do not contribute to the understanding of the colon structure. The output of this step is the estimated center of the innermost fold and edges that are likely fold edges. To track a fold edge across successive frames, we sample a number of pixels along each fold edge of the current frame and find the corresponding points (so called “tracked points”) with similar characteristics in the next frame using Pyramid Lucas-Kanade (LK) tracking algorithm [52]. Other optical-flow tracking algorithms with low significant processing overhead may be used. A good feature point to track for LK tracking is a point with strong derivative

values in two orthogonal directions. Linear edges are not good candidates for tracking because they have strong derivative values in one direction only. However, fold edges are typically curve edges and for each edge, we sample pixels at an equal distance apart to reduce the possibility of having collinear sampled pixels for tracking. We select new points to track every frame and use them to locate corresponding points (tracked points) only in the subsequent frame since the tracked points may not locate on any fold edge of the frame. We use the tracking results to assign a global edge ID to each edge, depending on whether the edge in the current frame corresponds to an edge in its previous frame as follows. Because the tracked points may not fall exactly on the edge of the current frame  $i$ , we first label each tracked point and its eight connected neighbors in frame  $i$  with the global edge ID of the corresponding edge in frame  $i-1$  (i.e., global edge ID of the edge used to find these tracked points). If any edge pixel in frame  $i$  is at the same pixel as one of the labels, the edge pixel is assigned that edge label. An edge is either assigned the lowest value (i.e., lowest global edge ID indicating older age) among the labels assigned to its edge pixels or is given a new unique global edge ID if its edge pixels are not assigned any labels (i.e., not overlapping with any edges in the previous frame). Fold edge tracking across frames continues until the tracking error exceeds a threshold  $TH\_TRACK\_ERR$ . Any motion vectors whose length in a perspective view is greater than a threshold  $TH\_VALID\_MV$  are discarded.

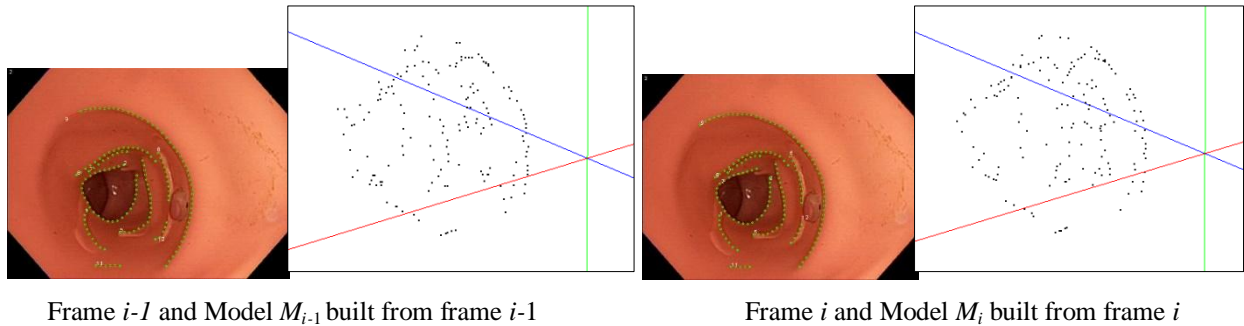
### 3.2. Model Registration

This process involves five steps. We define a local colon model for each frame as a set of 3D points located relatively from the camera in a local 3D coordinate system with the camera position at the origin.

### 3.2.1. Build a local model

These points are first estimated in the first step by the application of DfI on the pixels on the detected fold contours in the frame using an algorithm for building a local model of a frame. Fig. 3 shows example results of the local models. Therefore, there are as many local colon models as the number of frames in a video segment. The global coordinate system is derived from the first input frame. The local model of the subsequent frame is registered with the global model reconstructed from all its prior frames using the correspondence among

points of the models from fold edge tracking. The challenges are as follows. The local models do not have identical shape even for the same part of the colon due to noise in input images, *DfI* estimation errors, and colon mobility. Each local model may cover slightly different folds. Recall that DfI gives different model scales for different illumination in images. To overcome these challenges, we use the least-squares approach by Umeyama [46] to estimate transformation parameters of the local models to the global model such that the transformation error is minimized.



**Fig. 3.** Two consecutive images of the inside of an Olympus synthetic colon and the corresponding local models. Green dots show the sample points used for tracking in frame  $i-1$  and the corresponding tracked points in frame  $i$ . Red, green, and blue axes represent  $x$ ,  $y$ , and  $z$  axes, respectively. Black dots are the green dots mapped to corresponding 3D coordinates.

### 3.2.2. Calculation of transformation matrices

This is done in two steps: (1) calculating transformation parameters by registering a local model of frame  $i$  ( $M_i$ ) to its previous local model ( $M_{i-1}$ ) and (2) computing transformation parameters to register  $M_i$  to the global model by accumulating all the transformation parameters from the first frame up to frame  $i$ . Therefore, we can obtain the camera motions between consecutive frames all the way up to frame  $i$ .

Let  $M_i = \{p_{ij}\}$  where  $p_{ij}$  is a point  $j$  on a model  $i$  where  $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ .

The local model  $M_i$  is reconstructed from frame  $i$  using DfI. We assume that between two consecutive frames, there is at most one camera transformation. We use Umeyama's solution [46] that gives the least squared error to derive the camera transformation since this method supports different model scales and robust against corrupted data points. Note that other 3D rigid registration methods that do not restrict the two models being registered to be the same scale can be used. While there are recent development in non-rigid 3D registration methods [53-59], but these methods cannot handle the significant difference between the model sizes.

Instead of finding the transformation matrix between adjacent frames, one can employ a multi-frame registration approach by considering a range of frames to find the best model with the least transformation error. For example, we can find the best matching frames that may not be consecutive frames such as when the camera looks away to the side and comes back to look down the colon lumen. This multi-frame registration approach should produce a better colon model with the trade-off of increased processing time. We would still need to derive the camera motion between consecutive frames. We leave this for future work.

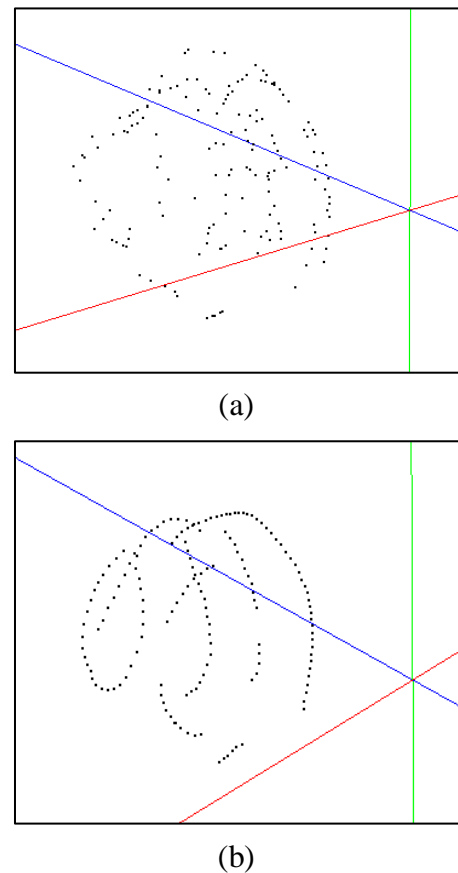
### 3.2.3. Camera motion calculation

We obtain the camera position at frame  $i$  ( $C^L_i$ ) from repeated applications of the translation matrices from registering adjacent frames starting from the first frame to the current frame  $i$ . We obtain the camera view vector ( $C^V_i$ ) and the camera up vector ( $C^U_i$ ) at frame  $i$ . The camera up vector is perpendicular to the camera view vector and can express camera torque motion. Let  $C^V_1 = (0, 0, -1)$  be the initial camera view vector; let  $C^U_1 = (0, 1, 0)$  be the initial camera up vector; and  $C^L_1 = (0, 0, 0)$  is the initial camera location in the local coordinate system.

### 3.2.4. Merging $M^G_i$ to a single global model $G$

This step is needed as the global model of each frame is not exactly the same due to transformation errors although we have attempted to minimize the errors. We merge edges of the same global edge ID of the two models in three steps. First, we find a 3D representative curve (RC) for each edge in a given model  $M^G_i$  because the points on the same edge may not necessarily be at the same depths from the camera due to depth estimation errors and

transformation errors. Starting from the first point to the last point on the edge, we replace the depth of the point with the average depth between the depth of this point and that of its immediate neighboring point on the edge. Fig. 4(b) shows the result of this step to the model  $M^G_i$  in Fig. 4(a). Second, each RC of  $M^G_i$  is assigned the global edge ID of the edge from which the RC is derived.



**Fig. 4.** Example of derived representative curves (RCs). (a) Model  $M^G_i$ , (b) RCs in Model  $M^G_i$

Third, we merge the RCs of  $M^G_i$  with those of the global model  $G$  using an algorithm for merging the local model  $M^G_i$  to the global model  $G$ . For each pair of RCs with the same global edge ID, one from  $M^G_i$  and the other from the global model  $G$ , the reference RC is first chosen to be the longer edge of the pair so that fewer points on the



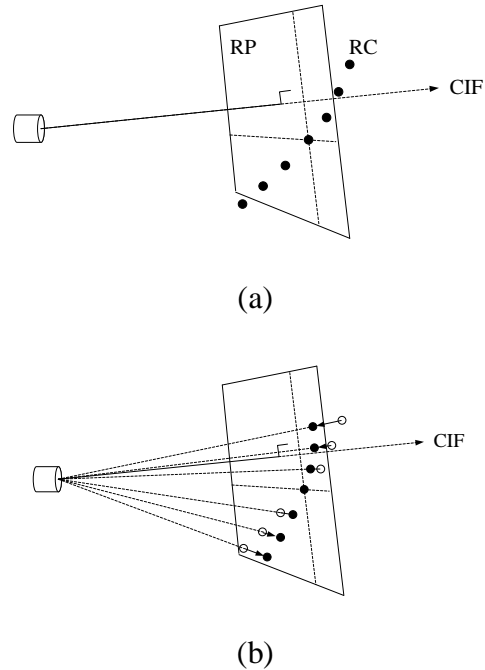
colon fold contour need to be estimated in the subsequent step.

### 3.2.5. Completing fold contours

There are several challenges to complete a fold contour in 3D space. For instance, short RCs are likely mis-detected blood vessel edges or are the results of wrong tracking. Several RCs are also located near each other. To overcome these challenges, we detect and discard noise RCs in the global model  $G$  before grouping the remaining ones and connecting the RCs in the same group to generate a closed fold contour for each group.

*Step 1: Select important RCs:* We discard any RC shorter than a threshold,  $TH\_VALID\_RC\_LENGTH$ . The length of an RC is the sum of the Euclidean distance between 3D coordinates of the neighboring points on the RC. In decreasing order of RC lengths, for each  $RC_i$ , we find an  $RC_j$  (where  $j \neq i$ ) closer to  $RC_i$  than a threshold,  $TH\_DIST\_BTW\_RC$ , and discard this shorter  $RC_j$  since we already have the longer representative curve  $RC_i$ . To calculate the distance between  $RC_i$  and  $RC_j$ , for each point in  $RC_i$ , we find the nearest point in the  $RC_j$  based on the Euclidean distance among them. The average of these distances is the distance between  $RC_i$  and  $RC_j$ .

*Step 2: Make all points on the same selected RC on the same representative plane (RP):* Let the center of the RP be at the mid-point of its RCs and the RP be perpendicular to the vector from the center of the innermost fold (CIF) to the center of RP as shown in Fig. 5(a). Next, we project all the points on the RC on the RP along their reverse projection to CIF as illustrated in Fig. 5(b).



**Fig. 5.** Diagram showing a representative plane (RP) of a given RC

*Step 3: Find an average representative plane of a group of RPs:* Let the current RP be the RP with the lowest global ID (with the oldest edge). The oldest edge is the same edge detected in the earliest frame. Recall that the 3D reconstruction is aimed to be used during the withdrawal phase when a careful inspection of the colon is expected. During the withdrawal phase, the camera is expected to move more backward than forward. The older edge appears closer to the camera in an earlier frame and is gradually seen as an inner fold in a later frame until it is disappeared. We apply our DfI using the edge information detected in its oldest frame as our DfI performs better when the edge is an outer edge closer to the camera. We group the RP whose Euclidean distance from the current RP is within a threshold,  $TH\_DIST\_BTW\_RP$ . If at least two RPs are in the group, we compute the average RP of the group. Next, we project all the points on the RPs in the same group to this average RP along the reverse projection vector of each point. The RPs in the group are excluded from further

grouping. We repeat the same process until all RPs are considered. Note that several RCs may fall on the same average RP plane, possibly reducing the unknown portion of a colon fold on this plane. This and the previous step can also be done together.

*Step 4: Create a closed fold contour:* Each  $RC_j$  has its start tip  $s_j$  and the end tip  $e_j$ . To create a smooth curve to close a gap between a start tip of one RC and the end tip of the other RC, we draw a Cubic Bézier curve if the gap is small as follows. We first compute the estimated center  $c$  of the fold contour as the average location of all the end points of all the RCs in the RP. We start pairing  $RC_j$  with  $RC_i$  that has the smallest angle  $\angle e_i c e_j$  between the end tip  $e_i$  of the  $RC_i$ ,  $c$ , and the start tip  $s_j$  of the  $RC_j$ .

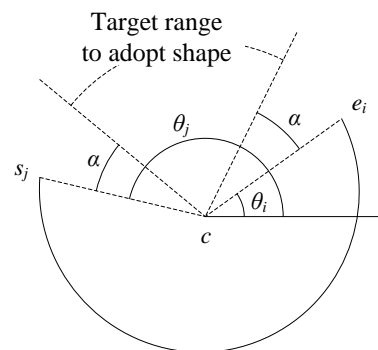
To draw a Bézier curve between  $e_i$  of  $RC_i$  and  $s_j$  of  $RC_j$ , we compute the tangent lines: one for  $e_i$  and the other for  $s_j$ , using the tip and the  $k^{th}$  point on the RC from that tip. The value of  $k$  was determined empirically. We select a control point on each tangent line at the distance  $d$  from the tip where  $d = control\_prop * \text{Euclidean distance between } e_i \text{ and } s_j$ . If the Euclidean distance between the two control points is at most the distance between  $s_j$  and  $e_i$ , we draw a Bézier curve to connect the two tips using the two calculated control points. On the other hand, if the distance is larger than that between  $s_j$  and  $e_i$ , we do not draw a Bézier curve in order to prevent an unsmooth contour because the two tangential lines do not meet. In this case, we attempt to adopt the shape of the nearest fold contour by assuming that the fold shape is quite similar. The details of this method are as follows.

*Step 5: Shape adoption:* We calculate the range (polar angle  $[0, 360^\circ)$ ) uncovered by  $RC_i$  which is  $[\theta_i, \theta_j]$  where  $\theta_i$  is the polar angle of  $e_i$  and  $\theta_j$  is the polar angle of  $s_j$  as shown in Fig. 6. We select the closest RC in other RPs that overlap with the range  $[\theta_i + \alpha, \theta_j - \alpha]$ . Then, we project the points on the

overlapping segment of the selected RC on this RP on their reverse projection vector. The margin  $\alpha$  is to allow a smooth curve between  $RC_i$  and the adopted curve segment where  $\alpha = (\theta_j - \theta_i) * margin\_prop$ . Next, we connect the adopted segment and the two tips of the  $RC_i$  using Bézier curves to close the gap. If there is no RC in any other RPs to adopt the shape from or the uncovered polar angle after the shape adoption is still at least the threshold angle, we discard this RC from fold contour completion. The values of *control\_prop* and *margin\_prop* and thresholds were determined experimentally.

We repeat this process until we can no longer pair any remaining RC in this RP. If we cannot generate a closed fold contour on the RP, we discard the RP. We continue processing the remaining RPs until all the RPs have been considered.

Note that RP is perpendicular to the vector from the camera to CIF. We reproject the fold contour completed on an RP to the plane that the fold is actually lying on by fitting all the points on real RCs (not the Bézier curves or shape adoption) from the contour to a plane using the least squares fitting algorithm. Once the 3D fold information is obtained, we use our previous methods [6] for creating colon walls between fold contours and for rendering the virtual colon when a user invokes.



**Fig. 6.** Shape adoption;  $s_j$  and  $e_i$  denote the start tip of  $RC_j$  and end tip of  $RC_i$  respectively

#### 4. Experimental Design and Results

We developed our software in C++ with OpenCV 2.1 and OpenGL. We evaluated the effectiveness of our technique on estimating the directions of reconstructed camera motions and the reconstructed colon structures.

**Ground truth:** Because there are no real 3D data of the colon during colonoscopy, we used an Olympus synthetic colon as in [6]. The synthetic colon has three sections: two straight sections and one curvy section in between. Our domain expert measured the depth of folds and the fold circumference ground truth from the outside of the synthetic colon. The domain expert inspected the inside of the synthetic colon by repeatedly performing a full insertion (from the beginning to the end of the synthetic colon) followed by a full withdrawal (from the end to the beginning of the synthetic colon). We used 562 and 456 images of the synthetic colon for evaluation of camera motion reconstruction and reconstruction of the colon structure, respectively. These images were captured at the frame rate of 29.97 fps at the resolution of 720x480 pixels using an Olympus colonoscope. To obtain the ground truth for evaluation of the camera motion direction in x, y, z axes and scaling direction, we manually reviewed the videos and noted the camera motion directions and scaling direction as the ground truth.

Note that the measurements taken from the outside of the synthetic colon may be different from those taken from the inside of the synthetic colon during the inspection but these measurements are closest measurements attainable. Although the curvy section of the colon (section 2) has several colon folds, most images of the inside of this section are blurry or either show only the colon mucosa or distorted folds due to the endoscope movement into this section. For the images with the lumen

in the curvy section, only two folds were seen.

**Parameters and values:** We used the parameter values for DfI as in [6] and used the following parameter values:

TH\_TRACK\_ERR=1500,  
TH\_VALID\_MV=50,  $k = 5$ , TH\_DIST = 0.005, TH\_VALID\_RC\_LENGTH = 0.4,  
TH\_DIST\_BTW\_RC = 0.07,  
TH\_DIST\_BTW\_RP = 0.1, *control\_prop* = 0.4, *margin\_prop* = 0.25.

These values were determined based on experiments to give good results using a different image set not included in the image set for evaluation.

##### 4.1. Evaluation of camera motion direction estimation

Camera motion consists of translation and rotation. Camera translation and rotation are obvious during colonoscopy. Scaling, however, is not because of the actual change in the colon scale, but is the result of DfI where the depth of the reconstructed colon fold depends on the brightness intensity as mentioned previously.

In Tables 1-3, 'ID' denotes the start frame of the image sequence and '#frames' denotes the number of frames used in a study case. For a sequence of 10 frames, there are 9 pairs of frames to derive camera motion from as the first frame is the reference. We originally wanted to use at least 30 frames per case; however, we found that 30 frames include more than one type of camera motion in some cases. Therefore, in our evaluation, we only selected cases where each sequence shows a single camera motion. 'GT' denotes the ground truth camera direction only since we cannot measure the exact amount of camera motion. Our performance metric is *accuracy defined as the percentage of correct estimations by the software over the total number of*

*estimations.* For each case, let score (S) be the total number of correct estimations of camera motion directions (in x, y, and z axes including no motion) by our software.

#### 4.1.1. Evaluation on camera translation direction

For the z-axis, the camera direction is either backward (B/+), forward (F/-), or no motion (N). For the y-axis, the direction is either up (U/+), down (D/-), or no motion (N). For the x-axis, the direction is right (R/+), left (L/-), or no motion (N). Table 1 shows the normalized camera motion in the x-axis, y-axis, and z-axis, and the score counting the total number of correct estimations in the three axes. For each case, the score is between 0 and 3. With small motion, it is difficult to observe the motion direction from watching the image sequence.

Based on observation, we treated the case that the absolute normalized camera motion along an axis less than 0.2 (20% of the distance the camera moved) as no motion.

Table 1 shows that the technique gave 22 out of 30 correct camera translation directions under the set threshold for no motion. The accuracy is 73% (22/30). For case ID 13253, the camera direction in each of the three axes is correct. The score is, therefore, three. The absolute normalized score in the x-axis is 0.03 less than 0.2; the software determines that there is no clear motion, which is also correct according to the ground truth. Only in the z-axis, the software found a significant camera forward movement, which is the same as the ground truth. Five out of ten cases show correct camera directions in all three axes.

**Table 1.** Effectiveness of estimation of camera translation directions on 10 sequences of 239 frames of synthetic colon images. Nor. x, y, and z capture the normalized distance (Cal. Distance /Cal. Abs travel distance omit from the table) with the direction of the translation in x-axis, y-axis, and z-axis, respectively. The camera direction in the x-axis is denoted with right (+)/left(-), in the y-axis with up(+)/down(-), and in the z-axis with backward(+)/forward(-)

ID, #frames	GT (x/y/z)	Nor. X	Nor. Y	Nor. Z	Score
4984,18	R/U/N	<b>0.24 (R)</b>	<b>0.22 (U)</b>	-0.95 (F)	2
6165,31	L/N/B	<b>-0.47 (L)</b>	<b>0.01 (N)</b>	<b>0.88 (B)</b>	3
6846,21	N/N/B	<b>-0.18 (N)</b>	0.74 (U)	-0.66 (F)	1
12521,19	N/U/F	<b>-0.01 (N)</b>	<b>0.62 (U)</b>	<b>-0.78 (F)</b>	3
12915,31	N/N/B	0.85 (R)	0.29 (U)	<b>0.44 (B)</b>	1
13253,31	N/N/F	<b>-0.03 (N)</b>	<b>0.06 (N)</b>	<b>-1.00 (F)</b>	3
13284,31	N/N/F	<b>-0.11 (N)</b>	<b>-0.17 (N)</b>	<b>-0.98 (F)</b>	3
13437,14	R/N/F	<b>0.42 (R)</b>	0.28 (U)	<b>-0.87 (F)</b>	2
16074,30	L/D/B	<b>-0.22 (L)</b>	0.68 (U)	-0.70 (F)	1
17355,13	N/D/B	<b>-0.18 (N)</b>	<b>-0.21 (D)</b>	<b>0.96 (B)</b>	3
<b>Accuracy</b>			<b>22/30=73%</b>		

#### 4.1.2. Evaluation on camera rotation direction

We decompose the rotation matrix into an individual rotation matrix for each axis.

For this evaluation, the GT of the rotation about the x-axis (Up, Down), the y-axis (Left, Right), and the z-axis (Counter Clockwise, Clockwise), respectively are determined for each image sequence. When

the camera rotation about an axis is less than about 7°, it is difficult to observe the rotation about that axis from watching the corresponding image sequence. Therefore, when the absolute computer generated score

is less than 7°, we treated such a rotation as no rotation. Tables 2 shows results of individual sequences and the accuracy of 90% (27/30).

**Table 2.** Effectiveness of estimation of the camera rotation directions on 10 sequences of 190 frames of synthetic colon images in the x-axis: Up (+)/ Down (-), in the y-axis: Left (+)/Right (-), in the z- axis: counter-clockwise denoted as CC (+)/clockwise denoted as C (-); *N*: no camera motion

ID, #frames	GT	x	y	z	Score
6080,31	D/L/CC	<b>-45.7 (D)</b>	<b>36.3 (L)</b>	-17.8 (C)	2
6937,9	N/R/N	<b>3.5 (N)</b>	<b>-26.5 (R)</b>	<b>-2.4 (N)</b>	<b>3</b>
6973,28	N/L/CC	<b>2.4 (N)</b>	<b>11.7 (L)</b>	<b>16.6 (CC)</b>	<b>3</b>
12017,11	D/L/N	<b>-7.8 (D)</b>	<b>9.8 (L)</b>	<b>1.4 (N)</b>	<b>3</b>
12143,8	U/R/N	<b>20.6 (U)</b>	<b>-28.5 (R)</b>	-64.3 (C)	2
12230,16	D/L/N	<b>-31.2 (D)</b>	<b>27.5 (L)</b>	<b>2.5 (N)</b>	<b>3</b>
12521,19	U/L/N	<b>25.74 (U)</b>	<b>8.7 (L)</b>	<b>0.1 (N)</b>	<b>3</b>
12805,31	U/R/C	<b>16.9 (U)</b>	<b>-10.4 (R)</b>	<b>-17.1 (C)</b>	<b>3</b>
12863,11	D/R/N	<b>-28.8 (D)</b>	<b>-45.9 (R)</b>	11.2 (CC)	2
12872,26	U/N/C	<b>10.3 (U)</b>	<b>2.6 (N)</b>	<b>-26.9 (C)</b>	<b>3</b>
<b>Accuracy</b>		<b>27/30=90%</b>			

#### 4.1.3. Evaluation of scaling

When an image has low illumination, the reconstructed colon model is large since the distance from the camera is far and the reverse projection vectors expand. For a brightly illuminated image, the reconstructed model is small. When we have an image sequence showing images with lower illumination to brighter illumination, we have a small reconstructed colon model (reconstructed from the brighter frame that comes later in time such as model A in Fig. 1 to register with a large reconstructed colon model (reconstructed from the darker frame that happens earlier in time such as model B in Fig. 1; the scale increases in this case (scale>1). We denote this case as Dark-Bright (D-B). Similarly, we need to scale down (scale<1) when the image sequence showing images changing from bright to dark, denoted as (B-D). Table 3 shows the accuracy of 100%, which is a very good result despite the fact that the total amount

of light put out by an endoscope processor is typically automatically adjusted to be within a range (auto lamp illumination mode).

We also evaluated the effectiveness of the algorithms on endoscopic images of the real colons. Out of 30 directions in 10 sequences (3 directions per sequence) of 201 images total, the software detected 26 correct rotation directions, resulting in an accuracy of 86.6%. The performance on images from the synthetic colon is better than that on the real colon images because images of the synthetic colon are easier to be processed. Folds in the synthetic colon are not as complex as those in the real colon images. In terms of scaling direction, our software calculated scaling direction correctly for 90% (9/10) for 135 real colon-images. Only one sequence out of ten cases that the software detected no direction but the ground truth has a positive scaling factor (Dark-Bright). The individual results for rotation directions and scaling on real colon

images are omitted to keep the article in reasonable length.

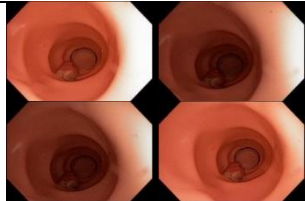
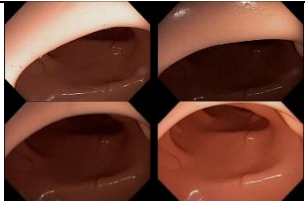

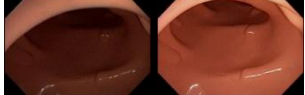
4.2. Evaluation of the reconstructed colon structures



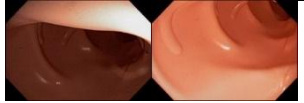


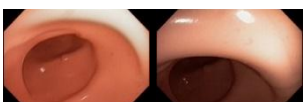
We consider the reconstructed depth and the reconstructed circumferences for three different sections of the synthetic colon. We use MAD calculated as the mean absolute difference of measurements between the reconstructed model and the ground truth considering all the folds seen in the image sequence. We first selected the physical scaling factor that gave the least MAD against the ground truth and scaled every fold in the reconstructed model to the ground truth using the same scaling factor. Note that this physical scaling factor is to scale the reconstructed model in logical units to the physical synthetic colon in millimeters. It is different from the scaling factor calculated to register two reconstructed models reconstructed from images with different illuminations. For instance, in Table 4, the physical scaling factor matching fold 2 to its corresponding fold in the synthetic colon gave the least MAD and was chosen to scale all the folds in the section. Tables 4-6 show the evaluation results for one image sequence in the first straight section, the curve section, and the last straight section, respectively. In

Table 4, the first fold has the largest normalized error in both depth estimation (16/25) and circumference estimation (50/166). Fold 2 has the least error because it was used to derive the scaling factor. Later folds tend to have lower normalized errors as well.

Table 7 shows the results with more sequences from each section. We observed a better recall of fold edges in the straight part of the colons than our previous technique that builds a virtual colon from a single endoscopic frame. The average depth MAD and circumference MAD on the synthetic colon are 8.5 mm and 17 mm, respectively and 94.7% of the folds seen in colonoscopy can be reconstructed. During the curvy section of the colon (section 2), only a few folds are shown in the corresponding image sequence, resulting in a small depth MAD and circumference MAD. Although 100% of the folds seen are reconstructed from the image sequences in this section, this section of the synthetic colon actually has more folds. These folds were not seen continuously in one long image sequence since the model was deformed by the endoscope during withdrawal. The conditions are different from real-life where a more pliable colon wall would allow more folds to be discovered.

**Table 3.** Effectiveness of scaling on 10 sequences of 163 frames of synthetic colon images; GT ground truth---Bright to Dark (B-D) or Dark to Bright; the calculated scale (Cal. Scale) for B-D (<1) and for D-B (>1)

ID, #frames	GT	Start frame / End frame	Cal. Scale	ID, #frames	GT	Start frame / End frame	Cal. Scale
4815,31	B-D		<b>0.43 (B-D)</b>	12038,8	B-D		<b>0.65 (B-D)</b>
4885,31	D-B		<b>1.09 (D-B)</b>	12054,9	D-B		<b>1.59 (D-B)</b>

12521,1 9	D- B		<b>2.16</b> <b>(D-B)</b>	12075,8 B- D		<b>0.75</b> <b>(B-D)</b>
11937,1 9	D- B		<b>1.74</b> <b>(D-B)</b>	12110,8 D- B		<b>1.07</b> <b>(D-B)</b>
11963,2 0	B- D		<b>0.66</b> <b>(B-D)</b>	12145,8 B- D		<b>0.57</b> <b>(B-D)</b>
<b>Accuracy</b>			<b>10/10=100%</b>			

**Table 4.** Case 12900 (section 1 – straight); Circf represents circumferences. The depth and circumference of the reconstructed model after scaling to physical units was done. \* denotes the fold used to give the scaling factor for the entire model.

	<b>Ground truth (mm)</b>		<b>Reconstructed model</b>		<b>Absolute Difference</b>	
	Depth	Circf.	Depth	Circf.	Depth	Circf.
Fold 1	25	166	41	216	16	50
Fold 2*	51	190	51	190	0	0
Fold 3	75	184	65	198	10	14
Fold 4	118	160	96	158	22	2
Fold 5	138	160	124	132	14	28
<b>MAD (mm)</b>					<b>13</b>	<b>19</b>

**Table 5.** Case 12320 (section 2 - curve); Circf represents circumferences. The depth and circumference of the reconstructed model after scaling was done. \*denotes the fold used to obtain the scaling factor for the entire model.

	<b>Ground truth (mm)</b>		<b>Reconstructed model</b>		<b>Absolute Difference</b>	
	Depth	Circf.	Depth	Circf.	Depth	Circf.
Fold 1	53	160	55	184	2	24
Fold 2*	85	160	87	160	2	0
<b>MAD (mm)</b>					<b>2</b>	<b>12</b>

**Table 6.** Case 16203 (section 3 - straight); Circf represents circumferences. The depth and circumference of the reconstructed model after scaling was done. \*denotes the fold used to obtain the scaling factor for the entire model.

	<b>Ground truth (mm)</b>		<b>Reconstructed model</b>		<b>Absolute Difference</b>	
	Depth	Circf.	Depth	Circf.	Depth	Circf.
Fold 1*	41	180	50	180	9	0
Fold 2	67	190	61	155	6	35
Fold 3	93	185	99	147	6	39
Fold 4	112	168	109	166	4	2
Fold 5	140	100	121	145	19	45
<b>MAD (mm)</b>					<b>9</b>	<b>24</b>

**Table 7.** Overall results; MAD represents the mean absolute difference in mm. considering all the folds seen in each image sequence; recall of folds represents the ratio of the number of reconstructed folds and the number of actual folds seen in each image sequence. MAD is in mm.

Section ID	Case ID	Depth MAD	Circf. MAD	Recall of folds	# Frames
1	6955	13	33	5/5	35
	6990	10	30	5/5	38
	12900	13	19	5/5	145
	20399	9	25	3/5	77
	19024	5	12	4/4	58
2	12320	2	12	2/2	22
	12117	8	0	1/1	29
	18753	12	11	3/3	20
3	16203	9	24	5/5	18
	15940	4	4	3/3	24
<b>Avg.</b>		<b>8.5</b>	<b>17</b>	<b>36/38</b> <b>(94.7%)</b>	

The average percentage of the depth error is about 10% (8.5/85.8) of the average depth of all the folds seen in all the image sequences. The average percentage of the circumference error is about 10% (17/178.2) of the average circumference of all the folds seen in all the image sequences. The errors are considered small. We cannot quantify the performance of the technique using endoscopy images of real colons since there is no available ground truth that can be measured. Fig. 7 shows an example of reconstruction on a sequence of endoscopic images from a real colon where some colon folds have triangular shape.

### 5. Discussion and Limitation

Overall, the test result shows that the proposed technique produces directions of endoscope translation motion in at least 73% in agreement with the ground truth. The technique performs much better at 90% accuracy for estimating rotation directions. For colon structure reconstruction evaluation, we tested our software with

several sequences of images with overall illumination changes across the images in the sequence. The illumination changes impact the change in the scale of the local reconstructed colon models that form the global colon model. Our experiments show that, when considering the directions of scaling (up or down), we can scale the models correctly in at least 90% of the studied cases. The correct camera translation and rotation estimations and the colon structure with details fold protrusion and width are important for determining which area of the colon has been inspected and which has not, which has significant implication on the ability to give direct feedback to the endoscopist.

We found that the fold edge tracking step plays an important role to the performance of the reconstruction of the colon structures. Even though we eliminate several noise edges, some noise edges from instrument, polyp, etc. remain. If the remaining edge being tracked is not an edge of a fold, but of other noises, we get a wrong colon shape as a result. We will revise the



preprocessing step of our previous work to better remove noise edges. In our current implementation, many steps such as the fold edge tracking and labeling tracked edges are not optimized for speed. Because out-of-focus frames happen frequently, interrupting the reconstruction, multiple small 3D virtual colon segments are expected and they are to be registered together. Without any additional hardware, the registration will require matching of salient features in the images in the two nearest virtual colon segments and registration of these segments.

## 6. Conclusion and Future Work

We present the work on reconstruction of a colon structure and endoscope motion from a colonoscopic image sequence for screening colonoscopy. Given the difficulty of the problem, we still have to address a few more problems as discussed in Section 5. Nevertheless, we overcame several challenges such as the scale difference of consecutive reconstructed models, estimating closed fold contours from tracked edges, and registering the models. We evaluated our proposed technique in two aspects: endoscope motion reconstruction and colon structure reconstruction. For the evaluation of motion reconstruction, we compared the software generated camera translation and rotation directions with ground truth motion directions. As future work, we will investigate a multi-frame registration approach to further reduce the depth and circumference estimation errors

by registering nearby frames that are more similar in fold shape. We have not found any existing colonoscopy simulators that provide the measurements of the inside circumference of the colon fold and the shape of the fold. Developing such a realistic synthetic ground truth has been a challenging problem by itself. A 3D colon model from existing CT Colonography data is unlikely to match well with the colon structure during colonoscopy since the colon poses are different. Using a cylinder with flat surface without colon fold protrusion or fold thickness does not give a realistic colon model to develop computer-assisted technology to improve colonoscopy quality in practice. Under consultation with the domain expert, we will investigate 3D printing of a 3D colon model where the model is reconstructed using manually labeled fold contours of colonoscopy images of the real colon. The reconstruction should give us relative fold alignment, fold depth, fold protrusion, and fold thickness at each point on a fold, which we can scale the model to a physical unit and print the model.

## Conflict of Interest Statement

Oh, Wong, and Tavanapong have management roles and own stock in EndoMetric Corp. Piet de Groen and Mayo Clinic Rochester have a financial interest and hold an equity position in EndoMetric Corp, a company that offers products for real-time analysis and feedback of quality of colonoscopy.

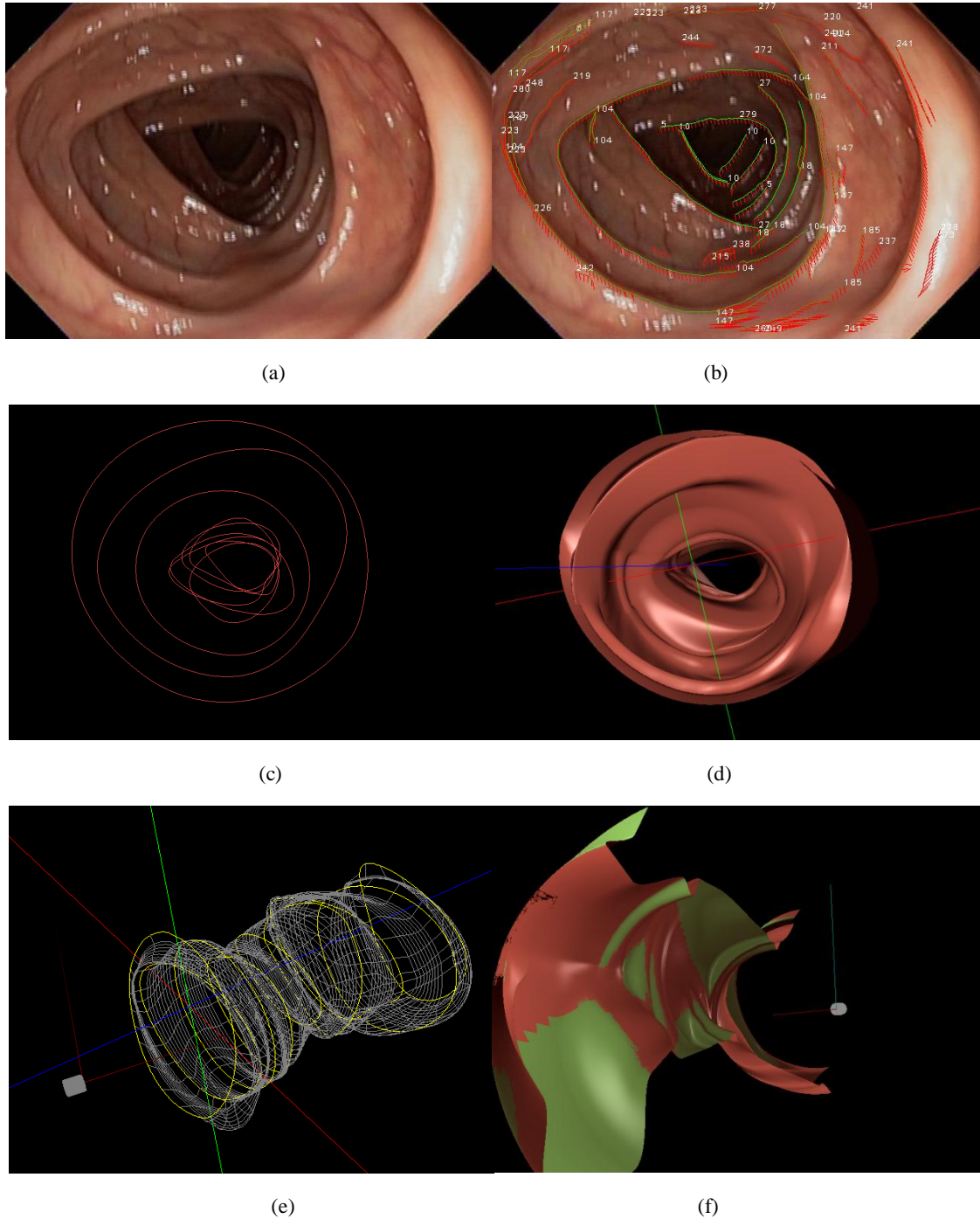


Fig. 7. (a) One of the input images; (b) Green edges indicate that the fold edge is tracked through several previous frames. Short red lines indicate the motion of the tracked points. White numbers are global edge ID assigned to each edge. (c) Front view of the completed contour  $G$  in the global coordinate system reconstructed from multiple frames. (d) Rendered  $G$ ; red, green and blue lines represents x-, y- and z-axes of the global coordinate system, respectively. (e) Wireframe of  $G$ ; fold contours are marked with yellow color. Gray cylinder represents the endoscope. (f) Green area represents the unobserved area under the reconstructed camera motion on the reconstructed colon model (e); red and blue lines from the endoscope indicate the line of sight and the up vector, respectively.

## REFERENCES

1. Levin, T.R., *Colonoscopy Capacity*. Gastroenterology, 2004: p. 1841-1849.
2. American Cancer Society *Colorectal Cancer Facts and Figures*. 2016.
3. Mura, C., et al., *Automatic room detection and reconstruction in cluttered indoor environments with complex room layouts*. Computer and Graphics, 2014. **44**: p. 20-32.
4. Hong, D., et al. *Colon Fold Contour Estimation for 3D Visualization of Colon Structure from 2D Colonoscopy Images*. in *Int'l Symp. on Biomedical Imaging: From Nano to Macro (ISBI 2011)*. 2011. Chicago, IL, USA.
5. Hong, D., et al. *3D Reconstruction of Colon Segments from Colonoscopy Images*. in *Proc. of IEEE Int'l Conf. on Bioinformatics and Bioengineering*. 2009. Taiwan.
6. Hong, D., et al., *3D Reconstruction of Virtual Colon Structures from Colonoscopy Images*. Computerized Medical Imaging Graphics Journal, 2013. **38**(1): p. 22-33.
7. Hajder, L. and D. Chetverikov, *Weak-perspective structure from motion for strongly contaminated data*. Pattern Recognition Letters, 2006. **27**(14): p. 1581-1589.
8. Hajder, L., D. Chetverikov, and I. Vajk, *Robust structure from motion under weak perspective*. 2nd International Symposium on 3d Data Processing, Visualization, and Transmission, Proceedings, 2004: p. 828-835.
9. Hajder, L., A. Pernek, and C. Kazo, *Weak-perspective structure from motion by fast alternation*. Visual Computer, 2011. **27**(5): p. 387-399.
10. Kazo, C. and L. Hajder, *Rapid Weak-perspective Structure from Motion with Missing Data*. 2011 Ieee International Conference on Computer Vision Workshops (Iccv Workshops), 2011.
11. Taylor, C.J. and D.J. Kriegman, *Structure and Motion from Line Segments in Multiple Images*. Ieee Transactions on Pattern Analysis and Machine Intelligence, 1995. **17**(11): p. 1021-1032.
12. Tomasi, C. and T. Kanade, *Shape and Motion from Image Streams - a Factorization Method*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(21): p. 9795-9802.
13. Zhang, L., et al., *Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo*. Ninth Ieee International Conference on Computer Vision, Vols I and II, Proceedings, 2003: p. 618-625.
14. Chandraker, M. *What Camera Motion Reveals about Shape with Unknown BRDF*. in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. 2014.
15. Cremers, D. and K. Kolev, *Multiview Stereo and Silhouette Consistency via Convex Functionals over Convex Domains*. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2011. **33**(6): p. 1161-74.

16. Furukawa, Y. and J. Ponce, *Accurate, Dense, and Robust Multiview Stereopsis*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. **32**(8): p. 1362-1376.
17. Ahmed, A.H. and A.A. Farag. *Shape from shading under various imaging conditions*. in *IEEE Conf. on Computer Vision and Pattern Recognition*. 2007. MN.
18. Ahmed, A.H. and A.A. Farag. *Shape from shading for hybrid surfaces*. in *IEEE Conf. on Image Processing (ICIP 2007)*. 2007. Atlanta, GA.
19. Christou, C.G. and J.J. Koenderink, *Light source dependence in shape from shading*. Vision Research, 1997. **37**(11): p. 1441-1449.
20. Horn, B.K.P., *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. 1970, MIT.
21. Horn, B.K.P. and M.J. Brooks, *The Variational Approach to Shape from Shading*. Computer Vision Graphics and Image Processing, 1986. **33**(2): p. 174-208.
22. Hougen, D.R. and N. Ahuja, *Estimation of the Light-Source Distribution and Its Use in Integrated Shape Recovery from Stereo and Shading*. Fourth International Conference on Computer Vision : Proceedings, 1993: p. 148-155.
23. Ikeda, O., *Shape-from-shading algorithm for oblique light source*. Advances in Visual Computing, Proceedings, Pt 2, 2007. **4842**: p. 357-366.
24. Ikeuchi, K. and B.K.P. Horn, *Numerical Shape from Shading and Occluding Boundaries*. Artificial Intelligence, 1981. **17**(1-3): p. 141-184.
25. Kondo, S., et al., *Shape and Source from Shading Using Zero Crossings*. International Conference on Pattern Recognition, 1992. **I**: p. 534-537.
26. Lee, C.H. and A. Rosenfeld, *Improved Methods of Estimating Shape from Shading Using the Light-Source Coordinate System*. Artificial Intelligence, 1985. **26**(2): p. 125-143.
27. Ming, X., R.C. Zhao, and P. Maria, *Solving self-shadow problem of shape from shading in light source projected system*. Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004: p. 334-337.
28. Oliensis, J. and P. Dupuis, *A Global Algorithm for Shape from Shading*. Fourth International Conference on Computer Vision : Proceedings, 1993: p. 692-701.
29. Prados, E. and O. Faugeras, *Shape from shading: a well-posed problem?* 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, Proceedings, 2005: p. 870-877.
30. Rashid, H.U. and P. Burger, *Differential Algorithm for the Determination of Shape from Shading Using a Point Light-Source*. Image and Vision Computing, 1992. **10**(2): p. 119-127.
31. Sato, T. and K. Hosokawa, *Visual rather than proprioceptive information contribute more to shape-from-shading when the light-source was actively moved*. Perception, 2009. **38**: p. 33-33.

32. Visentini-Scarzanella, M.S., D.; Guang-Zhong Yang. *Metric depth recovery from monocular images using Shape-from-Shading and specularities*. in *Image Processing (ICIP)*. 2012.
33. Wada, T., H. Ukida, and T. Matsuyama, *Shape from shading with interreflections under a proximal light source: Distortion-free copying of an unfolded book*. *International Journal of Computer Vision*, 1997. **24**(2): p. 125-135.
34. Wu, C., S.G. Narasimhan, and B. Jaramaz, *A multi-image shape-from-shading framework for near-lighting perspective endoscopes*. *International Journal of Computer Vision*, 2010. **86**(2-3): p. 211-228.
35. Xu, M., R.C. Zhao, and M. Petrou, *Solving self-shadow problem of shape from shading in light source projected system*. 2004 7th International Conference on Signal Processing Proceedings, Vols 1-3, 2004: p. 1235-1238.
36. Yang, L. and J.-Q. Han, *A Perspective Shape-from-Shading Method using Fast Sweeping Numerical Scheme*. *Optica Applicata*, 2008. **38**(2): p. 387-398.
37. Zhang, L., et al., *A Unified Framework for Document Restoration using In-painting and Shape-from-Shading*. *Pattern Recognition*, 2009. **42**(11): p. 2961-2978.
38. Zhang, R., et al., *Shape from Shading: A Survey*. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1999. **21**(8): p. 690-706.
39. Zheng, Q. and R. Chellappa. *Estimation of Illuminant Direction, Albedo, and Shape from Shading*. in *IEEE Computer Vision and Pattern Recognition*. 1991.
40. Deguchi, K. *Shape Reconstruction from Endoscope Image by its Shadings*. in *IEEE/SICE/RSJ Int'l Conf. on Multisensor Fusion and Integration for Intelligent Systems*. 1996.
41. Koppel, D., et al. *Toward Automated Model Building from Video in Computer-Assisted Diagnoses in Colonoscopy*. in *Proc. of SPIE Medical Imaging Conference*. 2007. San Diego, CA, USA.
42. Kaufman, A. and J.N. Wang, *3D surface reconstruction from endoscopic videos*. *Visualization in Medicine and Life Sciences*, 2008: p. 61-+.
43. Sun, D., et al., *Surface Reconstruction from Tracked Endoscopic Video Using the Structure from Motion Approach*, in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions: 6th International Workshop, MIAR 2013 and 8th International Workshop, AE-CAI 2013, Held in Conjunction with MICCAI 2013, Nagoya, Japan, September 22, 2013. Proceedings*, H. Liao, et al., Editors. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 127-135.
44. Qingyu Zhao, True Price, Stephen Pizer, Marc Niethammer, Ron Alterovitz, and Julian Rosenman. "The Endoscopogram: a 3D Model Reconstructed from Endoscopic Video Frames." *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016.

45. Saad Nadeem and Arie Kaufman. "Depth Reconstruction and Computer-Aided Polyp Detection in Optical Colonoscopy Video Frames," *arXiv preprint arXiv:1609.01329*, 2016.
46. Umeyama, S., *Least-Squares Estimation of Transformation Parameters between 2 Point Patterns*. Ieee Transactions on Pattern Analysis and Machine Intelligence, 1991. **13**(4): p. 376-380.
47. Marino, J., F. Qiu, and A. Kaufman. *Registration of Virtual and Optical Colonoscopy Views*. in *Proc. of MICCAI 2008 Workshop: Computational and Visualization Challenges in the New Era of Virtual Colonoscopy*. 2008.
48. Marino, J., F. Qiu, and A. Kaufman, *Virtually Assisted Optical Colonoscopy*, in *SPIE Medical Imaging 2008: Physiology, Function, and Structure from Medical Images*. 2008: San Diego, CA, USA. p. 69160J-1.
49. Liu, J., K.R. Subramanian, and T.S. Yoo, *A robust method to track colonoscopy videos with non-informative images*. International Journal of Computer Assisted Radiology and Surgery, 2013. **8**(4): p. 575-592.
50. Zhou, J., et al. *Circular generalized cylinder fitting for 3D reconstruction in endoscopic imaging based on MRF*. in *Proc. of Computer Vision and Pattern Recognition Workshops*. 2008. Anchorage, Alaska, USA.
51. Sederberg, T. *BYU Bézier curves*.
52. Lucas, B.D. and T. Kanade. *An Iterative Image Registration Technique with an Application to Stereo Vision*. in *IJCAI'81 Proceedings of the 7th international joint conference on Artificial intelligence*. 1981. San Francisco, CA, USA.
53. Tektonidis, M. and K. Rohr, *Diffeomorphic Multi-Frame Non-Rigid Registration of Cell Nuclei in 2D and 3D Live Cell Images*. IEEE Transactions on Image Processing, 2017. **PP**(99): p. 1-13.
54. Yang, J., et al., *Sparse non-rigid registration of 3D shapes*. Proceedings of the Eurographics Symposium on Geometry Processing, 2015: p. 89 - 99.
55. Ge, S. and G. Fan, *Sequential non-rigid point registration for 3D human pose tracking*. IEEE International Conference on Image Processing (ICIP), 2015. **1105 - 1109**.
56. Aktar, N., J. Alam, and M. Pickering, *A Non-Rigid 3D Multi-Modal Registration Algorithm Using Partial Volume Interpolation and the Sum of Conditional Variance*. 2014 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2014: p. 1 - 7.
57. Matinfar, B. and L. Zagrochev, *Non-rigid Registration of 3D Ultrasound Images Using Model-Based Segmentation*. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014: p. 323 - 328.
58. Serradell, E., et al., *Robust non-rigid registration of 2D and 3D graphs*. IEEE Conference on Computer Vision and Pattern Recognition, 2012: p. 996 - 1003.

59. Cheng, Z.-Q., et al., *Non-rigid Registration in 3D Implicit Vector Space*. 2010 Shape Modeling International Conference, 2010: p. 37 - 46.
60. Rui Wang, True Price, Qingyu Zhao, Jan-Michael Frahm, Julian Rosenman, and Stephen Pizer. "Improving 3D Surface Reconstruction from Endoscopic Video via Fusion and Refined Reflectance Modeling." *SPIE Medical Imaging*, pp. 101330B-101330B, 2017