Visibility Map: A New Method in Evaluation Quality of Optical Colonoscopy

Mohammad Ali Armin[1,2], Hans De Visser[2], Girija Chetty[1], Cedric Dumas[2], David Conlan[2], Florian Grimpen[3], and Olivier Salvado[2]

[1] Department of Computer Science, University of Canberra, Australia
m.a.armin@gmail.com
[2] CSIRO Digital Productivity Flagship, Australia
Olivier.Salvado@csiro.au
[3] Department of Gastroenterology and Hepatology, Royal Brisbane and Women's Hospital

Abstract. Optical colonoscopy is performed by insertion of a long flexible endoscope into the colon. Inspecting the whole colonic surface for abnormalities has been a main concern in estimating quality of a colonoscopy procedure. In this paper we aim to estimate areas that have not been inspected thoroughly as a quality metric by generating a visibility map of the colon surface. The colon was modeled as a cylinder. By estimating the camera motion parameters between each consecutive frame, circumferential bands from the cylinder of the colon surface were extracted. Registering these extracted band images from adjacent video frames provide a visibility map, which could reveal uncovered areas by clinicians from colonoscopy videos. The method was validated using a set of realistic videos generated using a colonoscopy simulator for which the ground truth was known, and by analyzing results from processing actual colonoscopy videos by a clinical expert. Our method was able to identify 100% of uncovered areas on simulated data and achieved with sensitivity of 96% and precision of 74% on real videos. The results suggest that visibility map can increase clinicians' awareness of uncovered areas, and would reduce the chance of missed polyps.

Keywords: Optical colonoscopy, Visibility map, Colonoscopy Quality, Uncovered area, Camera motion parameters.

1 Introduction

Colorectal or Bowel cancer is the second cause of cancer related death after lung cancer, in Australia and the Western world [1]. Early diagnosis of bowel cancer can increase the chance of survival for patients by up to 90%. Colonoscopy is the gold standard method for detection and removal of colonic polyps. The efficiency of a colonoscopy procedure is influenced by many factors, including the amount of the colon surface that is inspected for the presence of polyps by the clinician. Studies have reported that even experienced gastroenterologists can still miss up to 33% of polyps [2, 3]. This is in part due to polyps, in particular flat lesions such as sessile serrated adenomas, not being recognized even though they are in view, but also due to polyps not being viewed because they were never inspected by the camera.

N. Navab et al. (Eds.): MICCAI 2015, Part I, LNCS 9349, pp. 396–404, 2015.
DOI: 10.1007/978-3-319-24553-9_49

One way to address this problem is to increase clinicians' skills through training; another way is to provide assistance to clinicians during the intervention by developing assistive technologies. There are several techniques which can measure the quality of a colonoscopy inspection by metrics such as withdrawal time [4], number of informative frames [5], and uncovered areas [6]. Hoang et al. proposed a method to identify colon folds to reconstruct individual 3D colon segments for every frame and determine their visible areas, without providing a combined feedback [6]. We aim to develop a technology that estimates uncovered areas by generating a visibility map of the flattened colon surface. To do this, we modeled the colon as a cylinder, and we hypothesized that circumferential bands of the colon surface could be extracted by projecting a 3D cylinder onto each frame. To know where on a frame to project the 3D cylinder, camera motion parameters between successive frames were estimated through epipolar geometry analysis [7], which has demonstrated high accuracy in endoscope camera pose estimation [8, 9]. Others have used a combination of different techniques to increase the robustness of camera motion estimation [10] but at the cost of extra computation time. Finally, each extracted band was registered to the one from the previous frame, and rastered to build a visibility map of the colon internal surface. A diagram of our method is shown in Fig. 1 and described in the next section.

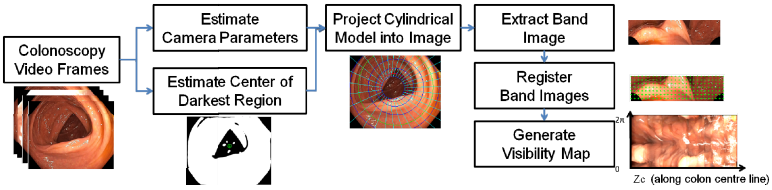

Fig. 1. Main processing steps of our proposed method

2 Method

2.1 Camera Motion Estimation

In this section, first we briefly describe the mathematical model for a colonoscope camera, then explain in detail the epipolar geometry based algorithm [7] to obtain the camera location and orientation relative to the colon lumen.

Colonoscope Camera Model. The camera of the colonoscope used in our setup had a fisheye lens (190HD Olympus endoscope), introducing image deformation. We used a mathematical model proposed by Scaramuzza et al. [11] to model fisheye lens camera projection. Using this model a 3D point (X,Y,Z) can be projected into image point (u,v) through the following equation:

$$\lambda \begin{bmatrix} u \\ v \\ a_0 + a_1\rho + \cdots + a_{n-1}\rho^{n-1} \end{bmatrix} = R. \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T \tag{1}$$

where λ is scaling factor, $\rho = \sqrt{(u-u_0)^2 + (v-v_0)^2}$ indicates distance from image center (u_0, v_0) and $a_0, a_1, \ldots, a_n$ are intrinsic parameters used to correct the

deformation in the colonoscopy images. R and T represent extrinsic camera parameters (rotation and translation).

Since the colonoscopy camera motion is parallax in most of the scene (each consecutive frame pair can be assumed as a stereo pair), an epipolar geometry based algorithm was employed to estimate the camera motion as follows:

Camera Calibration and Distortion Correction. An offline calibration specific to each colonoscope was performed using a fisheye model described in [11] to obtain the intrinsic camera parameters and correct image deformation. To reduce the computing time, only features' positions on the image were corrected for distortion.

Feature Detection and Uninformative Frame Removal. A set of feature points was automatically extracted and matched between each pair of consecutive frames using the Kanade-Lucas-Tomasi (KLT) algorithm [12]. Features consisted mostly of blood vessel patterns and soft tissue structures. Frames with no features (e.g. blurry frames) were excluded from further computation and assumed to be uninformative.

The RANdom Sample Consensus (RANSAC) [13] and deviation of distances between the corresponding points were employed to remove unreliable matches (outliers) and increase robustness of the motion estimation algorithm.

Extrinsic Camera Parameters Estimation. A combination of five and eight point algorithms explained in [7, 14] were used to calculate extrinsic camera parameters.

The 3D positions of tracked features were calculated using triangulation and extrinsic camera parameters. The distance between the coordinate of 3D points projected into the image plane and the feature points defined a reprojection error obtained by

$$E = \sum_{i=1}^{M} \sum_{l=1}^{N} \left|\left| p_{il} - \acute{p}_i \right|\right|^2 \tag{2}$$

where N is the number of features on M images and $\acute{p}_i$ is the projection of a 3D point P_i estimated through triangulation and camera parameters to the image plane, and p_{il} is the correspondence tracked feature on the image.

First we estimated the reprojection error of camera parameters computed by five and eight point algorithms, then the camera parameters with a lower reprojection error were optimized using a Levenberg-Marquardt technique [15] by minimizing the cost function defined in Eqn. (2). The optimized extrinsic camera parameters (camera pose) inferred from consecutive frames was then concatenated to estimate camera position [16].

Kalman Filtering. To predict the camera pose in presence of uninformative frames, it is assumed that acceleration was constant, and the standard Kalman filter (Eqn. (3)) was used to predict each camera pose x_k from previous optimized camera pose x_{k-1}

$$x_k = Ax_{k-1} + Gw_{k-1}, \qquad z_{k-1} = Hx_{k-1} + \mathrm{v}_{k-1} \tag{3}$$

with A the state matrix similar to [17], z_{k-1} the observation at time k, $G = (I_6\ I_6\ I_6)$ and $H = (I_6\ 0_6\ 0_6)$ the driving and measurement matrices, where 0_6 and I_6 are

6×6 zero and identity matrices. w_{k-1} and v_{k-1} are process and Gaussian noise (1e-4 and 0.06 determined experimentally for best qualitative results).

2.2 Generating Visibility Map

In this section we describe how circumferential bands of the colon surface (band image) were extracted from a video frame (area between green and red circle in Fig. 2) and merged to build a visibility map of a colon segment.

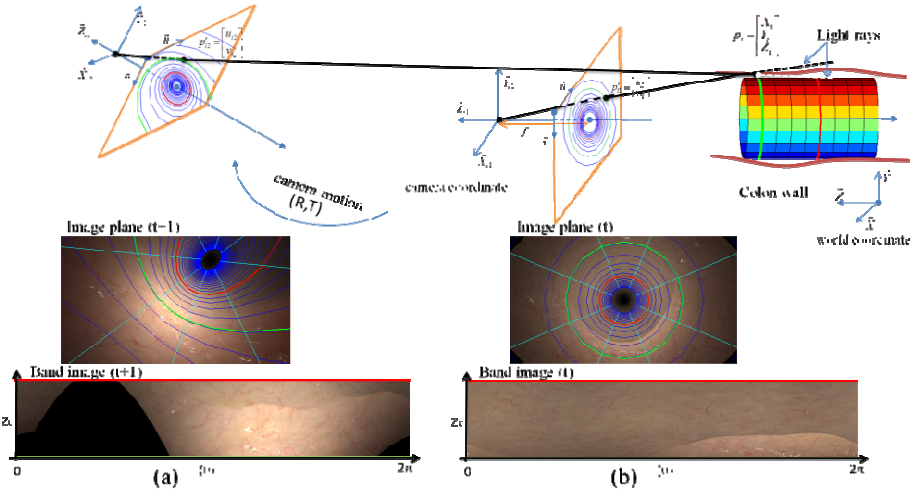

Fig. 2. Projection of cylinder into video frames and extracted band images (area between green and red perimeters), when camera moving backward from time (t) to (t+1); camera looking toward center of colon (b), the camera location and orientation change (a).

Cylindrical Model. We assumed that a colon can be modeled as a cylinder. Fig. 2 represents the cylindrical model; this cylinder was made of concentric 3D circles. Every point $P_i = (X_i, Y_i, Z_i)$ on a 3D circle was calculated by:

$$X_i = r.cos(\theta_i), \quad Y_i = r.sin(\theta_i), \quad Z_i = Z_{ci} + \Delta Z \tag{4}$$

where r is the radius of cylinder. We assumed an average colon diameter of 4cm (anatomical average), and we found that values from 3cm to 5cm did not change substantially the results. Z_{ci} is distance of cylinder from camera, ΔZ is the distance between each two circle on the cylinder, θ_i is the radian in a 3D circle. θ_i ranged from 0 to 2π with a step size $\Delta\theta$ related to the number of points (NP) on a 3D circle calculated by ($2\pi/NP$).

Cylindrical Model Projection and Band Image Extraction. From the camera parameters (intrinsic and pose), the 3D cylinder was projected into the image plane, using the camera model defined by Equation 1.

The orientation of the camera was updated by forcing its line of sight to go through the center of dark region (deepest area) of each frame, which was computed as barycenter of the darkest class of a two class segmentation using Otsu thresholding similar to [18].

The cylinder model had two perimeters, defined between Z_{ci}+1 cm and Z_{ci}+3 cm, from the camera defining a band on the cylinder circumference. We chose those values based on testing using the simulator and feedback from clinicians as to where the image quality is best and the lighting optimal. These perimeters defined a circumferential band that was moved at each frame depending on the camera pose. For example, when the camera was facing the wall, the perimeters could be seen on each side of the camera field of view. The image segment defined by that band was unwrapped and formed a rectangular patch (band image). The part of the band image outside the field of view was rastered in black.

Visibility Map Generation. Each extracted band image was assumed to overlap with the one from the previous frame, and two consecutive band images were registered by Local Weight Mean (LWM) registration algorithm [19] and merged on a plan corresponding to the internal colon surface that we call the visibility map. Fig. 4 shows an example of a visibility map.

3 Experiments and Results

3.1 Simulated Video

First we validated our algorithm on videos generated by a colonoscopy simulator. The simulator consists of a computer simulation of the colon model and a haptic device that allows insertion of an instrumented colonoscope to drive the simulation [20]. The colon model designed for this simulator has a parametric mathematical model of the colon allowing the generation of realistic human colon geometry. The simulator could also generate the ground truth camera poses and uncovered areas that were used for validation. We validated our method on ten different realistic videos from different parts of the colon generated by the simulator (each video on average covered 20cm of a colon length). Errors of the extrinsic camera parameters (orientation and translation) were computed between the ones estimated from the videos and those used by the simulator. Errors were averaged over each of the ten videos resulting in 10 errors for each parameter shown as boxplot in Fig. 3.

Translation errors between consecutive frames were less than 2 mm and rotation errors were less than 0.6 degrees. This corresponded to scenario without lens or colon obstructions, and in simulated videos using constant acceleration. It is expected that in real conditions, those errors might be higher.

3.2 Extracted Visibility Map from Simulated Videos

One example of implementation of our method is shown in Fig. 4, where dark spots represent uncovered areas. During any given actual colonoscopy the physician continuously

moves the colonoscope back and forth, naturally resulting in bands overlapping each other with different views (e.g. some might be darker/brighter due to lighting differences). Fig. 5 shows a case when the camera was moving forward and facing one side of the colon, and then moving backward while facing another side of the colon. The final composite image correctly filled in the internal surface, which Zc is the traveled distance along colon center line.

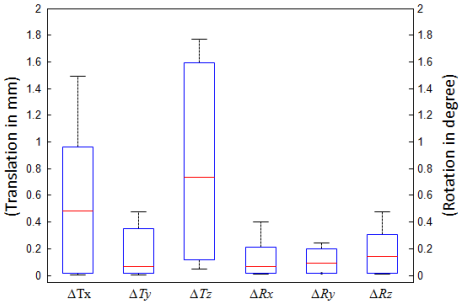

Fig. 3. The root mean square (RMS) error of camera motion parameters estimated on realistic videos generated by the simulator.

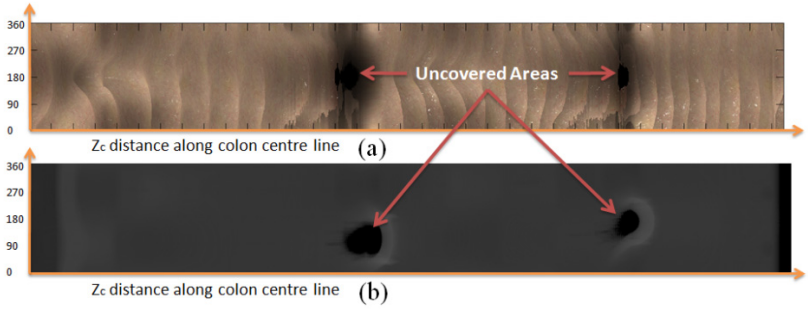

Fig. 4. Visibility map from a simulated colonoscopy video by our method (a), and its ground truth generated by the simulator (b), the black areas indicate uncovered areas.

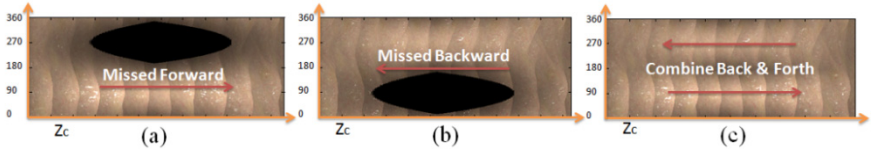

Fig. 5. Visibility map generated from a segment of the simulated colon: Parts of the colon are missed in forward motion (a), and backward motion in same place (b). Combination of two maps by max of pixel value shows the final coverage area (c).

3.3 Application to Actual Colonoscopy Video

We implemented our algorithm on nine segments of around 20 cm were inspected from five different videos of a 190HD Olympus endoscope, with 50 frame/sec and each frame sized 1856×1044 pixels. Two experiments were conducted. First, the

expert noted the uncovered areas from the video extracts. Second, the expert noted again the uncovered areas while seeing the visibility maps when the videos were shown. In the first case he identified nineteen uncovered areas, whereas in the second experi ment he noted twenty-four uncovered areas. The latter was used as the ground truth to validate our automated method by computing sensitivity and precision [21]. The sensitivity was 96% and precision was 74%. Fig. 6 represents the visibility map of a segment of a colon. Mean typical processing time was 1.5 min per frame using a standard PC, Matlab, and non-optimized scripts.

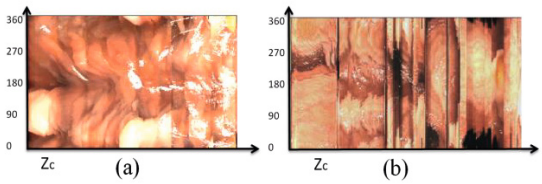

Fig. 6. Visibility map generated from a segment of a real colon, which is fully covered (a), and a map from another part of colon with some dark spot as uncovered areas (b).

4 Discussion and Conclusion

In this paper, we present a new method to estimate uncovered areas during optical colonoscopy. To achieve this, we modeled the colon as a cylinder, the camera parameters were estimated using epipolar geometry analysis, a 3D cylinder projected into colonoscopy image. The band image that could be extracted was then rastered on a visibility map showing the internal colon wall.

Estimating camera parameters from a colonoscopy video is challenging because of the non-rigidity of the colon anatomy, the presence of uninformative frames (e.g. a dirty lens), and low feature content of some tissue. Those issues were addressed in our technique by use of epipolar geometry to estimate camera pose, and a Kalman filter to estimate the camera parameters in the presence of uninformative frames.

On the visibility map of the colon surface, uncovered areas, or areas with lack of focus (removed uninformative frames) showed up as black areas. However, errors in the camera parameters translated into error of the position of the band images on the visibility map. This was reduced by registering successive band images, but could be improved by correcting the camera parameters with the registration results, thereby reducing drift in estimating the camera position, for example. This will be part of our future work, which will also include the identification and characterization of haustral folds to detect suspicious folds where potential polyps could be hidden.

Another challenge while extracting band images is the structure of colon that might depart from the cylinder model that we used. In our test this created problems at the flexure (high curvature). In future work, we will model the colon as a generalized cylinder around the center line defined by the camera trajectory.

By making the clinicians aware of uncovered areas post-procedurally, awareness of challenging areas as well as automated reports could lead to improved inspection efficacy and reduce the chance of missing polyps. This was already seen during our tests when the expert changed his opinion after seeing the visibility map and added uncovered areas that he thought he had missed before. The processing time of the proposed method is still too slow to achieve real-time assistance, however code optimization was not the focus of our research, and processing time could be significantly accelerated by using compiled code, particularly for graphical processing unit (GPU). Real-time computation would allow providing feedback during the procedure.

References

1. Australian Institute of Health and Welfare. http://www.aihw.gov.au/
2. Pickhardt, P.J., et al.: Location of adenomas missed by optical colonoscopy. Ann. Intern. Med. 141, 352–359 (2004)
3. Rex, D.K.: Who is the best colonoscopist? Gastrointest. Endosc. 65, 145–150 (2007)
4. Nawarathna, R., Oh, J., Muthukudage, J., Tavanapong, W., Wong, J., de Groen, P.C.: Real-time phase boundary detection for colonoscopy videos using motion vector templates. In: Yoshida, H., Hawkes, D., Vannier, M.W., et al. (eds.) Abdominal Imaging 2012. LNCS, vol. 7601, pp. 116–125. Springer, Heidelberg (2012)
5. Arnold, M., et al.: Indistinct frame detection in colonoscopy videos. In: MVIPC, pp. 47–52 (2009)
6. Hong, D., et al.: 3D Reconstruction of Virtual Colon Structures from Colonoscopy Images. Comput. Med. Imaging Graph. 38, 22–33 (2013)
7. Hartley, R., et al.: Multiple view geometry in computer vision. Cambridge University Press, Cambridge (2003)
8. Wang, H., et al.: Robust motion estimation and structure recovery from endoscopic image sequences with an adaptive scale kernel consensus estimator. In: CVPR, pp. 1–7 (2008)
9. Puerto-Souza, G.A., Staranowicz, A.N., Bell, C.S., Valdastri, P., Mariottini, G.-L.: A comparative study of ego-motion estimation algorithms for teleoperated robotic endoscopes. In: Luo, X., Reich, T., Mirota, D., Soper, T., et al. (eds.) CARE 2014. LNCS, vol. 8899, pp. 64–76. Springer, Heidelberg (2014)
10. Liu, J., et al.: A robust method to track colonoscopy videos with non-informative images. Int. J. Comput. Assist. Radiol. Surg. 8, 575–592 (2013)
11. Scaramuzza, D., et al.: A flexible technique for accurate omnidirectional camera calibration and structure from motion. In: ICVS, pp. 45–45 (2006)
12. Shi, J., et al.: Good features to track. In: CVPR, pp. 593–600 (1994)
13. Fischler, M.A., et al.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. CACM 24, 381–395 (1981)
14. Nister, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. 26, 756–770 (2004)
15. More, J.: The Levenberg-marquardt algorithm: implementation and theory. In: Cleaveland, W.R. (ed.) CONCUR 1992. LNCS, vol. 630, pp. 105–116. Springer, Heidelberg (1992)
16. Scaramuzza, D., et al.: Visual Odometry. IEEE Robot. Autom. Mag. 18, 80–92 (2011)

17. Nagao, J., et al.: Fast and accurate bronchoscope tracking using image registration and motion prediction. In: Barillot, C., Haynor, D.R., Hellier, P., et al. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 551–558. Springer, Heidelberg (2004)
18. Zhen, Z., et al.: An Intelligent Endoscopic Navigation System, MA, pp. 1653–1657 (2006)
19. Goshtasby, A.: Image registration by local approximation methods. J. IVC. 6, 255–261 (1988)
20. De Visser, H., et al.: Developing A Next Generation Colonoscopy Simulator. Int. J. Image Graph. 10, 203–217 (2010)
21. Han, J.: Data mining: concepts and techniques. Elsevier Morgan Kaufmann, Amsterdam, Boston (2006)