PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

3D reconstruction of gastrointestinal regions using shape-from-focus

Bilal Ahmad, Ivar Farup, Pål Anders Floor

Bilal Ahmad, Ivar Farup, Pål Anders Floor, "3D reconstruction of gastrointestinal regions using shape-from-focus," Proc. SPIE 12701, Fifteenth International Conference on Machine Vision (ICMV 2022), 127011J (7 June 2023); doi: 10.1117/12.2679345

SPIE.

Event: Fifteenth International Conference on Machine Vision (ICMV 2022), 2022, Rome, Italy

3D Reconstruction of Gastrointestinal Regions Using Shape-from-Focus

Bilal Ahmad, Ivar Farup, and Pål Anders Floor

Department of Computer Science,
Norwegian University of Science & Technology, 2815 Gjøvik, Norway

ABSTRACT

3D shape reconstruction from images is an active topic in computer vision. Shape-from-Focus (SFF) is an important approach which requires image stack in a focus controlled manner to infer the 3D shape. In this article, 3D reconstruction of synthetic gastrointestinal regions is done using SFF. Image stack is generated in Blender software with focus controlled camera. A color focus measure is applied for shape recovery followed by a weighted L2 regularizer to estimate for inaccurate depth values. A precise comparison is done between recovered shape and ground truth data by measuring the depth error and correlation between them. Results shows that SFF technique will be practical for 3D reconstruction of GI regions with focus and motion controlled pillcams which is technologically feasible to implement.

Keywords: 3D reconstruction, Pillcams, Shape-from-Focus, One-view

1. INTRODUCTION

It has always been challenging to diagnose gastrointestinal diseases such as, crohn disease, small bowel cancer, ulcerative colitis and other disorders due to difficulty in accessing such a complex environment in human body. Endoscopy is usually carried out to examine such complex interior of human body which is a discomforting and painful procedure for the patients. Wireless capsule endoscope (WCE) were introduced in year 2000 by Given Imaging [1], as an alternate of the regular colonoscopy which is a patient-friendly, noninvasive and painless procedure to examine gastrointestinal (GI) regions . WCE is a pill-sized capsule that the patient swallows. It consists of camera on board, capturing images of the GI system while travelling through it. 3D reconstruction is not commonly used in intestinal screening but a 3D model based on WCE frames can be helpful better diagnose, visualize or analyze the areas of interests.

3D reconstruction is an inverse problem which can be rectified by applying different techniques to the images [2]. It is vital to get information regarding the 3D structure or the scene's depth since most tasks are completed in the 3D world. The concept of depth estimation involves using various approaches or algorithms to attain the spatial information of the object or to acquire the distances of all the points present in the scene, with respect to a specific chosen point.

Vision-based depth estimation methods are generally classified into different categories. Some methods comprise of the usage of special devices for depth estimation [3]. Examples of these technique are ultrasonic and optical time-of-flight estimation in which measured energy beam is first transported and then reflected energy is detected [4]. Other methods do not make use of any artificial source of energy and natural outdoor scenes fall under its category. Various monocular image-based techniques such as texture gradient analysis and photometric methods are used. Other methods are hinged on the motion or multiple relative positions of the camera [5]. Some methods aslo use contour matching techniques for depth enstimation [6]. 3D reconstruction has numerous applications in measurement systems, robotics, medical applications including diagnostics, video surveillance and monitoring etc. [7, 8].

Shape-from-Focus (SFF) is one of the 3D reconstruction technique which recovers shape of the objects using several images of the same scene. Images are captured either by changing the distance between the camera and the scene or by exploiting the focus settings of the camera. SFF identifies the best focused pixels in image stack via Focus Measure (FM) operator and utilize them as cue for depth estimation.

Further author information: (Send correspondence to Bilal Ahmad)
Bilal Ahmad: E-mail: bilal.ahmad@ntnu.no
Ivar Farup: E-mail: ivar.farup@ntnu.no
Pål Anders Floor: E-mail: paal.anders.floor@ntnu.no

Fifteenth International Conference on Machine Vision (ICMV 2022), edited by Wolfgang Osten, Dmitry Nikolaev, Jianhong Zhou, Proc. of SPIE Vol. 12701, 127011J

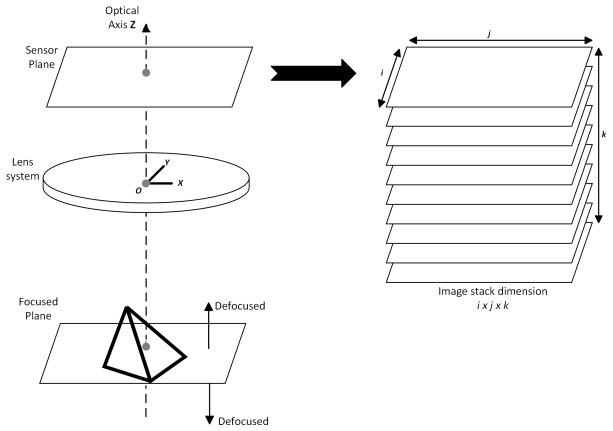


Figure 1: Image Acquisition in Shape-from-Focus.

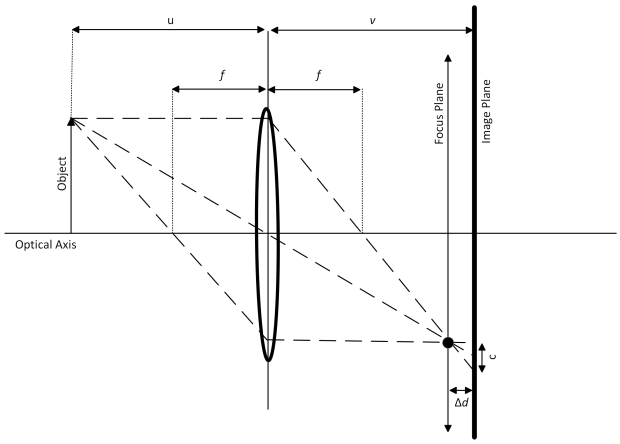


Figure 2: Focusing via Gaussian Lens law.

SFF was first discussed by Nayyar [9], who computed the focus value of the pixels by taking the laplacian of image as a first step and then uses gaussian interpolation method for shape recovery. Since then, a variety of FM operators have been proposed in the literature. These FM are divided into six different categories (Gradient, Laplacian, Wavelet, Statistics, Discrete cosine transform and miscellaneous) based on their working principle [10]. These FM's consider gray scale images for depth recovery. Recently, color FM are also proposed which uses RGB information in the image to evaluate the quality of a pixel [11, 12].

In real world applications, SFF is handy where camera settings can be controlled. SFF could be applied for 3D reconstruction of GI regions using capsule endoscopic images. But currently available pillcams are passive devices which are driven by natural peristalsis [13]. Many researchers are working to develop devices which allows capsule endoscopy to be performed in a controlled manner [14]. If such product is developed with focus controlled cameras, then SFF could be applicable.

In this paper, a color based SFF technique is applied to reconstruct synthetic GI regions to show the usefulness of the approach in pillcam application. The models are imported in Blender* and then modified for true comparison. Images stack from different regions of GI are also generated to test SFF algorithm. A weighted L2 regularizer is added as an additional step because focus values in some smooth regions are misleading and therefore, can not be trusted. With regularizer, position of correct pixels in the depth map are retained and rectified for inaccurate ones. Finally, depth Error and correlation are measured between recovered shape and the ground truth. The modified approach is also compared with most common one-view method, namely Shape-from-Shading (SfS) [15].

The remainder of this article is organized as follows. Section 2 explains regularized shape-from-Focus method. Results are compared and discussed in Section 3 and Section 4 concludes the article.

2. SHAPE FROM FOCUS

In SFF, images are taken either by changing the distance between the camera and the object in small steps with size Δ_{step}, or by changing the focus settings of the camera. Images I_n are stored in the image stack where, $1 \leq n \leq k$ and k is the total number of images in the image stack as shown in Fig. 1. An image is stored in each step and the total number of images are given by, $k = U/\Delta_{step}$, where, U is the toal displacement of the object. Some of the factors regarding the image sequence that can affect the quality of 3D reconstruction include total number of images, step size, surface details and imaging conditions. Step size, Δ_{step}, is very critical among others. A very small step size may confuse the focus information whereas a larger Δ_{step} results in less number of images which ultimately leads to loss of depth information. Therefore, it should be taken proportional to depth of field [16].

SFF measures focus quality of each pixel in the image stack to identify the best focused pixels. The quality of pixels is determined through focus measure (FM) operator which suppresses the defocused pixels and enhances the focused pixels.

*https://www.blender.org/

These focus pixels are then used to recover the depth of the scene. Gaussian lens law can be used to describe focusing on every pixel in the image sequence. If the distance between the object and the lens is such that the focus plane is shifted by a distance Δd from the image plane, a circle of confusion (c) is formed on the image plane as shown in Fig. 2. If the focus plane lies on image plane then $\Delta d = 0$ and image will be highly focused. The object distance (u) and image distance (v) from the lens is defined as,

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \tag{1}$$

where f is focal length of the camera. u and v are object and image plane distances from lens respectively. After image acquisition, next step is to measure the focus value of each pixel in the image stack. A color focus measure is applied on RGB images to compute the focus value of each pixel [12].

2.1 Color Focus Measure

An FM operator acts as a high-pass-filter which separates the high frequency content from low frequency content by enhancing the focused pixels and suppressing the defocused pixels. It computes the sharpness of a pixel by selecting a local window. Object points captured with different focus settings are then compared to identify the best focused pixel for depth estimation.

In order to compute the focus value of pixels, a color focus measure is applied. In the first step, the color difference between the neighbouring pixels and center pixels are computed and summed together in a local $\omega = 3 \times 3$ window. It is then followed by calculating their spread. The sum and the spread can be combined together as [12],

$$FM_c(i,j,k) = \sigma_\Delta^2 \sum_{r=1}^{\omega^2-1} \delta_r, \tag{2}$$

where δ_r is the difference between the center and the neighbouring pixels stacked together in Δ. σ_Δ^2 is the variance of Δ. After computing the focus value of each pixel, the depth map is obtained by finding the position of the best focused pixel which can be written as,

$$D_o(i,j) = \arg\max_k (FM_c(i,j,k)). \tag{3}$$

2.2 Weighted L2 regularizer

Initial depth map, D_o, obtained from the focus values contains many inaccurate depth points. This is due to the fact that some areas in the synthetic GI are smooth and therefore, resulting images had a low frequency variation in those areas. The focus values obtained in those regions were erroneous resulting in incorrect depth points.

A weighted L2 regularizer is introduced in which the focus value of each depth point is used as a fidelity term. Depth points containing the higher focus values were trusted and therefore, retained to their actual positions. However, depth points containing smaller focus values were mistrusted and therefore, neighbourhood depth values were given more weight to alter their position. In this way, incorrect depth points were successively move closer to their true depth values.

This problem is solved by minimizing the error function E_D, which can be computed as,

$$E_D = \int_\Omega \mid \nabla D \mid^2 + \lambda FM_c \mid D - D_o \mid^2 d\Omega. \tag{4}$$

Equation (4) is solved with gradient descent such as,

$$\frac{\partial D}{\partial t} = \nabla^2 D - \lambda FM_c (D - D_o), \tag{5}$$

where λ is a weighting factor between fidelity term and smoothness term. A small time step, Δt, is added to ensure stability with higher value of λ.

Depth map D obtained from weighted L2 regularizer can be used to measure corresponding world coordinates (x,y) for accurate recovered shapes mapping to perspective view. Therefore, (x,y), in perspective projection can be written as,

$$x = \widetilde{x}\frac{D}{f} \qquad y = \widetilde{y}\frac{D}{f}, \tag{6}$$

where $(\widetilde{x},\widetilde{y})$ are the image coordinates.

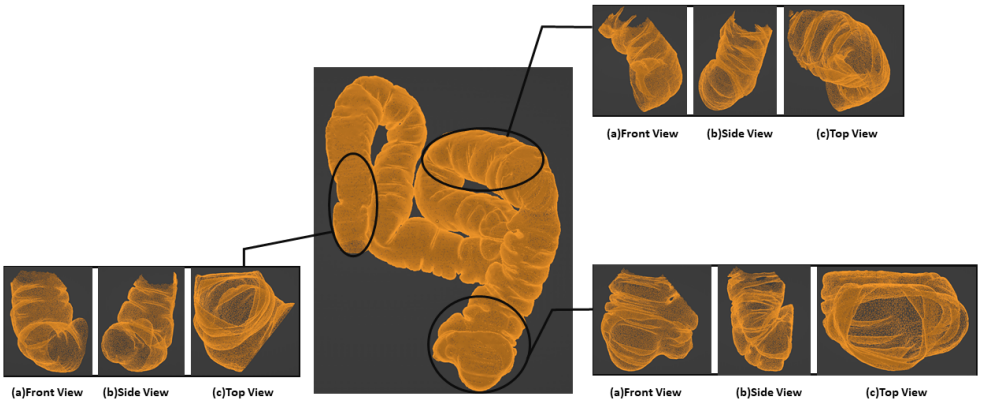


Figure 3: GI model.

3. RESULTS & DISCUSSION

3.1 Ground Truth Models

Shape-from-Focus algorithm is tested on different areas of synthetic GI regions. These models are constructed from CT scans of human patients and the texture on the model is based on analysis of real endoscopic videos [17]. Highlighted regions utilized for 3D reconstruction along with different views are shown in Fig. 3.

3D reconstruction is usually done in medical application without the availability of ground truth data. Therefore, Blender is chosen to have a ground truth scenario where precise comparison can be done between reconstructed surface and the ground truth data. Different other parameters such as distance between the camera and object or focus settings of the camera, can also be controlled in Blender which are essential to create an image stack for SFF.

In order to compare the reconstructed surface with the ground truth models, they are modified using Python API in Blender. When a model is placed under a perspective camera some of the occluded vertices/areas are not viewed by the camera. Therefore, to test the accuracy of SFF algorithm, it is necessary to remove all the occluded vertices and build the model consisting of only those vertices which are inside the camera frustum as well as viewed by the camera. The modified model is then exported in an obj file and then finally imported in MATLAB. The side and top view of ground truth models are shown in Fig. 5(a, c, e) and Fig. 6(a, c, e) respectively.

3.2 Image Acquisition for SFF

In order to generate image stack of each GI region, an environment is created similar to Fig. 1 in Blender. A simplified camera is selected and placed at $(0,0,0)$ in world space. Such camera is selected because distortions are not important to test the algorithm and it has to be removed in any case from the images by calibrating the camera. A point light source is also placed at camera center to imitate the light sources used in the pillcams. Different regions of GI model are placed below the camera. Models are also wrapped with texture to appear analogous to human GI regions.

Images of size 200×200 are acquired by changing the focus distance of the camera with a step size of $5mm$. In each image certain area of the scene is kept in focus while rest remain defocused. More than 150 images of each region are generated and stored in Portable Network Graphics (PNG) file format. Different samples in image stack for all three regions are shown in Fig. 4.

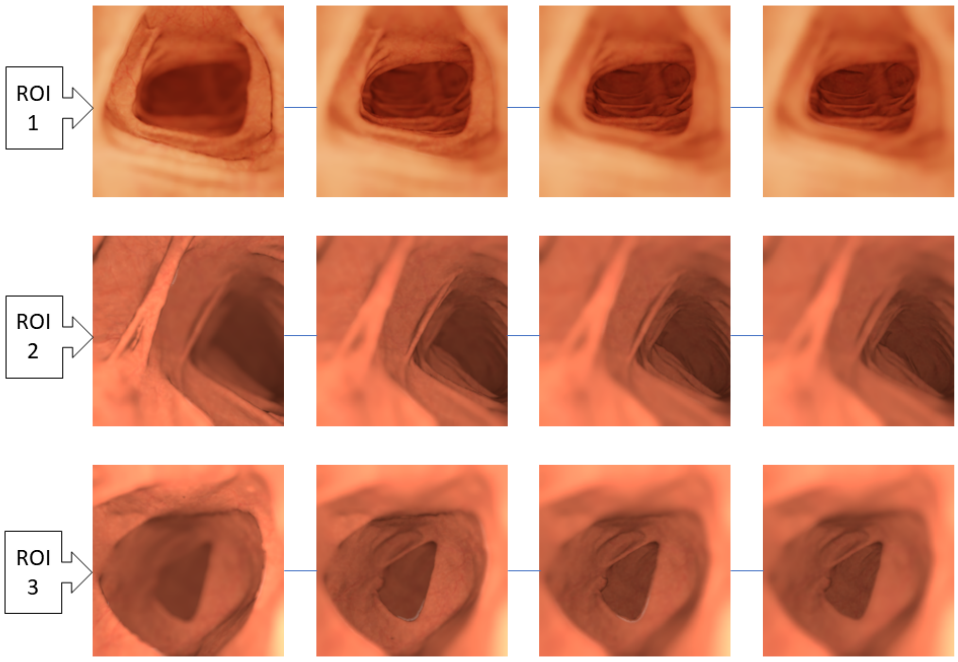


Figure 4: Image samples with different focus for all three regions investigated for 3D reconstruction.

3.3 3D reconstruction

Color image stack of each region is utilized to compute the focus value of each pixel using equation (2). Initial depth map D_o is reconstructed by finding the position of best focused pixel in image stack. D_o along with its focus value is then utilized in equation (5) to correct for inaccurate depth points. The value of λ is different for different cases and empirical in our experiment. Finally, the world points (x, y) are computed from depth point using equation (6). The side and top view of reconstructed regions are shown in Fig. 5(b, d, f) and Fig. 6(b, d, f) respectively.

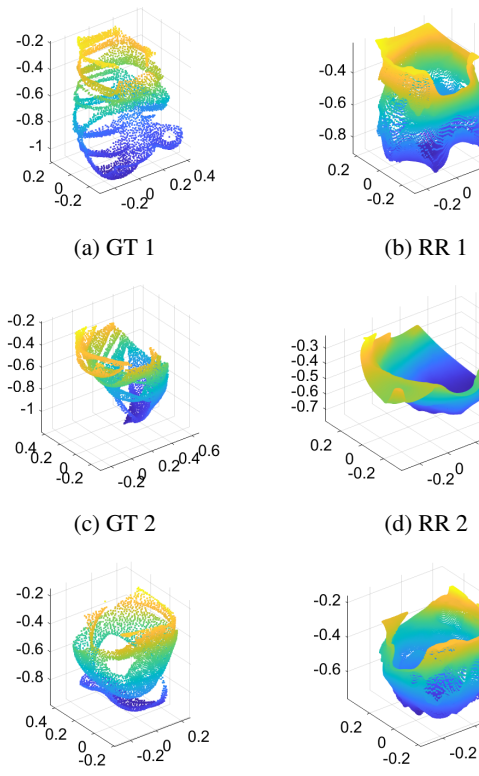
In order to evaluate the quality of 3D reconstruction, the reconstructed surfaces are compared with ground truth by measuring correlation and depth error. These methods are chosen to assess different features of the reconstructed surfaces. Correlation evaluates the shape of the reconstructed surface independent of scale and position and computed by estimating variance and covariance of the recovered shape and the ground truth data. Depth error (e_d) correctly evaluates the geometric deformation of the reconstructed shape and is given by,

$$e_d = \frac{1}{\Omega} \sum \sum_{i,j \in \Omega} \left| \frac{D_{i,j} - \widehat{D}_{i,j}}{\widehat{D}_{i,j}} \right|, \tag{7}$$

where $\widehat{D}$ is the ground truth and D is recovered 3D shape. Ω represents the region of the 3D model considered for error estimation.

Weighted L2 SFF is compared with SFS which is a most common one-view method. Results with SFS are generated by rendering single image of the same regions shown in Fig. 4 [18]. Table 1 shows the correlation and depth error for weighted L2 SFF and SFS, when compared with ground truth models. The proposed method shows higher correlation and lower depth error for all three cases. With weighted L2 regularizer, incorrect depth points are rectified which has significantly improved the results.

The shapes recovered with regular SFF for all three regions are shown in Fig. 7. It is evident from the figure that the depth map is corrupted with regular SFF and incorrect depth points have estimated the world coordinates (x, y) inaccurately leading to incorrect mapping to perspective view.

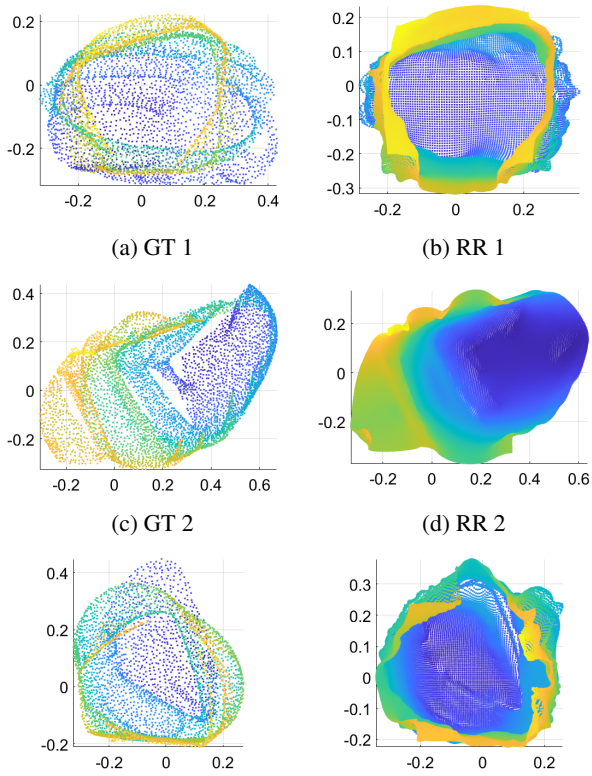


(a) GT 1 (b) RR 1

(c) GT 2 (d) RR 2

(e) GT 3 (f) RR 3

Figure 5: Side view of ground truth (GT) and recovered regions (RR) with weighted L2 SFF.



(a) GT 1 (b) RR 1

(c) GT 2 (d) RR 2

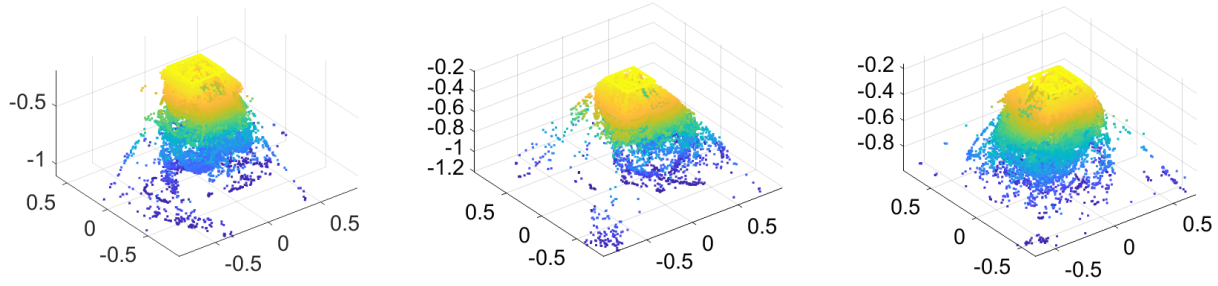
(e) GT 3 (f) RR 3

Figure 6: Top view of ground truth (GT) and recovered regions (RR) with weighted L2 SFF

3.4 Feasibility Study

SFF is a very simple and practical approach but it requires cameras with focus control settings to obtain images with different focus settings. Image stack can be generated either by changing the distance between the camera and the scene or by changing the focus settings of the camera.

Currently, available pillcams do not have such focusing cameras, and are driven by natural peristalsis inide human GI. Different prototypes for controlling the movement of pillcams have been designed and discussed in the literature. Glass *et al.* presented a mechanism to anchor the capsules to intestinal walls [19]. Karagozler *et al.* also presented a six legged endoscopic capsule to mimic the crawling motion [20]. These prototypes could be a step towards locomotive controlled pillcams.



(a) RR 1 (b) RR 2 (c) RR 3

Figure 7: Side view of recovered regions (RR) with regular SFF.

Table 1: Comparison between recovered shapes and ground truth models.

G1 region	Weighted L2 Shape-from-Focus		Shape-from-Shading	
	Correlation	Depth Error	Correlation	Depth Error
ROI 1	0.9217	0.2302	0.8883	0.4470
ROI 2	0.9707	0.1535	0.7927	0.3878
ROI 3	0.9302	0.2080	0.8725	0.3726

It is also technologically possible to insert a mechanical system inside pillcam to control the distance between the lens and sensor to control the focus by moving the lens. Modern mobile phones with such a narrow thickness have this technology. Therefore, it is also possible to introduce the mechanism for pillcams.

SFF requires an image stack for depth recovery. For an accurate depth estimation it is required to have large number of images with a constant step size Δ_{step}. Battery power is very important in pillcams and capturing an image stack with different focus settings may create power problem. Therefore, image capturing process can be optimized by varying the step size as depth of field (DOF) has a direct relation with object distance (u) for a given f-number (N), which is as follows [21],

$$DOF = \frac{2u^2Nc}{f^2}. \tag{8}$$

According to equation (8), DOF is proportional to u^2. Therefore, capturing images with varying Δ_{step} would be optimal for both power and SFF algorithm. That being said, if the technology is available in the future, SFF method would be practical for 3D reconstruction of human GI regions captured from pillcams as it shows very promising results when applied on synthetic GI regions.

4. CONCLUSION

In this article, SFF method is applied on different GI regions. Images are acquired by controlling the focus settings of the camera. Color focus measure is then applied for initial depth estimation followed by a weighted regularization step to correct for inaccurate depth points. Reconstructed surfaces are then compared with ground truth models by measuring average depth error and correlation. Result shows that SFF can be handy for 3D reconstruction of real pillcam images if motion and focus controlled pillcams will be available in the future.

ACKNOWLEDGMENTS

Funding was provided by the Research Council of Norway under the project CAPSULE no. 300031.

REFERENCES

[1] Iddan, G., Meron, G., Glukhovsky, A., and Swain, P., "Wireless capsule endoscopy," *Nature* **405**(6785), 417–417 (2000).

[2] Ham, H., Wesley, J., and Hendra, H., "Computer vision based 3d reconstruction: A review," *International Journal of Electrical and Computer Engineering* **9**(4), 2394 (2019).

[3] Lai, X.-b., Wang, H.-s., and Xu, Y.-h., "A real-time range finding system with binocular stereo vision," *International Journal of Advanced Robotic Systems* **9**(1), 26 (2012).

[4] Steckel, J. and Peremans, H., "Batslam: Simultaneous localization and mapping using biomimetic sonar," *PloS one* **8**(1), e54076 (2013).

[5] Özyeşil, O., Voroninski, V., Basri, R., and Singer, A., "A survey of structure from motion*.," *Acta Numerica* **26**, 305–364 (2017).

[6] Hadi, N. A., Ibrahim, A., Yahya, F., and Ali, J. M., "Centroid based on branching contour matching for 3d reconstruction using beta-spline," *Journal of Image and Graphics* **1**(3), 138–142 (2013).

[7] Koulaouzidis, A., Iakovidis, D. K., Yung, D. E., Mazomenos, E., Bianchi, F., Karagyris, A., Dimas, G., Stoyanov, D., Thorlacius, H., Toth, E., et al., "Novel experimental and software methods for image reconstruction and localization in capsule endoscopy," *Endoscopy International Open* **6**(02), E205–E210 (2018).

[8] Lee, H. M. and Choi, W. C., "Algorithm of 3d spatial coordinates measurement using a camera image," *J Image Graphics* **3**(1) (2015).

[9] Nayar, S. K. and Nakagawa, Y., "Shape from focus," *IEEE Transactions on Pattern analysis and machine intelligence* **16**(8), 824–831 (1994).

[10] Pertuz, S., Puig, D., and Garcia, M. A., "Analysis of focus measure operators for shape-from-focus," *Pattern Recognition* **46**(5), 1415–1432 (2013).

[11] Ahmad, B., Mutahira, H., Li, M., and Muhammad, M. S., "Measuring focus quality in color space," in [*2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE)*], 115–119, IEEE (2019).

[12] Mutahira, H., Ahmad, B., Muhammad, M. S., and Shin, D. R., "Focus measurement in color space for shape from focus systems," *IEEE Access* **9**, 103291–103310 (2021).

[13] Ciuti, G., Menciassi, A., and Dario, P., "Capsule endoscopy: from current achievements to open challenges," *IEEE reviews in biomedical engineering* **4**, 59–72 (2011).

[14] Yang, Y. J., "The future of capsule endoscopy: The role of artificial intelligence and other technical advancements," *Clinical Endoscopy* **53**(4), 387 (2020).

[15] Wu, C., Narasimhan, S. G., and Jaramaz, B., "A multi-image shape-from-shading framework for near-lighting perspective endoscopes," *International Journal of Computer Vision* **86**(2-3), 211–228 (2010).

[16] Muhammad, M. and Choi, T.-S., "Sampling for shape from focus in optical microscopy," *IEEE transactions on pattern analysis and machine intelligence* **34**(3), 564–573 (2012).

[17] İncetan, K., Celik, I. O., Obeid, A., Gokceler, G. I., Ozyoruk, K. B., Almalioglu, Y., Chen, R. J., Mahmood, F., Gilbert, H., Durr, N. J., et al., "Vr-caps: a virtual environment for capsule endoscopy," *Medical image analysis* **70**, 101990 (2021).

[18] Ahmad, B., Floor, P. A., and Farup, I., "3d reconstruction of gastrointestinal regions from single images. (accepted)," in [*Colour and Visual Computing Symposium (CVCS)*], (2022).

[19] Glass, P., Cheung, E., and Sitti, M., "A legged anchoring mechanism for capsule endoscopes using micropatterned adhesives," *IEEE Transactions on Biomedical Engineering* **55**(12), 2759–2767 (2008).

[20] Karagozler, M. E., Cheung, E., Kwon, J., and Sitti, M., "Miniature endoscopic capsule robot using biomimetic micro-patterned adhesives," in [*The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006. BioRob 2006.*], 105–111, IEEE (2006).

[21] Allen, E. and Triantaphillidou, S., [*The manual of photography*], CRC Press (2012).