3D Reconstruction of Gastrointestinal Regions from Single Images

Bilal Ahmad*, Pål Anders Floor, Ivar Farup and Milan Kresović

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

Abstract

3D shape reconstruction from images is one of the problem under investigation in the field of computer vision. Shape-from-shading (SfS) is an important approach which requires the reflectance properties of surface and light source position to infer the 3D shape. SfS is usually tested in medical applications without the availability of the ground truth data which makes the conclusion dubious. In this article SfS is applied on synthetic gastrointestinal regions and precise comparison is done between recovered shape and ground truth data by measuring the depth error and correlation between them. Results shows that SfS can recover the shapes quite well if penalized correctly.

Keywords

3D reconstruction, Capsule endoscopy, Shape-from-shading,

1. Introduction

With the advancement of medical field, the current trend is to make surgery ever less invasive. This implies smaller and smaller cuts in the patient's skin, which do not give the surgeons a direct view of their work. They only leave enough space for small cameras to be introduced into the patient's body. If the resulting images in such conditions are of poor quality [1], it can make the surgeon's work even harder. 3D reconstruction can be helpful in such cases to 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 the 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 photo-metric methods are used. Other methods are hinged on the motion or multiple relative positions of the camera [5]. 3D reconstruction has numerous applications in robotics, medical applications including diagnostics, video surveillance and monitoring etc. [6].

Shape-from-shading (SfS) is one of the many computer vision techniques to reconstruct the 3D shape of an object. It is distinct from other methods because it requires only one image for 3D reconstruction. SfS consists of two steps. In the first step, a reflection model is developed based on reflectance properties of the surface, position of the camera and light source. In the second step, a numerical scheme is designed to solve the image irradiance equation (IIE) which is constructed either by using partial differential equation (PDE) or optimization methods.

SfS was first discussed by Horn and Brooks [7], who developed an iterative scheme based on nonlinear first-order PDE by relating 3D shape to intensity variation in its image. Kimmel *et al.* [8] solved the SfS problem using fast marching method. Tankus *et al.* re-examines the SfS problem by solving the IIE under perspective projection so that it could be treated on a broader set of real world cases. Wu *et al.* [9] also solve the IIE under perspective projection with multiple light sources around camera center.

In real world applications, SfS is useful in situations where one shot of the scene is available. One of the recent application of SfS in today's era is capsule endoscopy [10] where usually position of light sources are known which are essential for this method. Additionally, rapid movement of capsule in certain areas of GI makes SfS a preferable choice because those areas might be captured once. In recent years, SfS has been applied on endoscopic images for 3D reconstruction [11, 12]. Although results seem promising, the conclusion is still uncertain because SfS methods are mostly applied without the availability of the ground truth data.

In this paper, a precise comparison is done between the recovered 3D shape and the ground truth data of synthetic models of Gastrointestinal (GI) regions developed by [13]. The models are imported in Blender[1] and then modified for true comparison. SfS was implemented with anisotropic diffusion as a smoothness constraint to preserve details in the recovered geometry. It is also novel in this work because L2 regularizer is typically used as a smoothness constraint which is unable to preserve the edges. Depth Error and correlation is then measured between recovered shape and the ground truth to estimate the quality of 3D reconstruction.

The remainder of this article is organized as follows. Section 2 explains perspective SfS model with anisotropic diffusion. Results are compared and discussed in Section 3 and Section 4 concludes the article.

2. Point Light Source Perspective SfS Model

This section briefly explains the SfS model under point light and perspective projection, where the light source is placed at the center of the camera projection as shown in Figure 1. Under the assumption of diffuse surface, radiance emitted by the surface element $\mathbf{S}$ can be computed
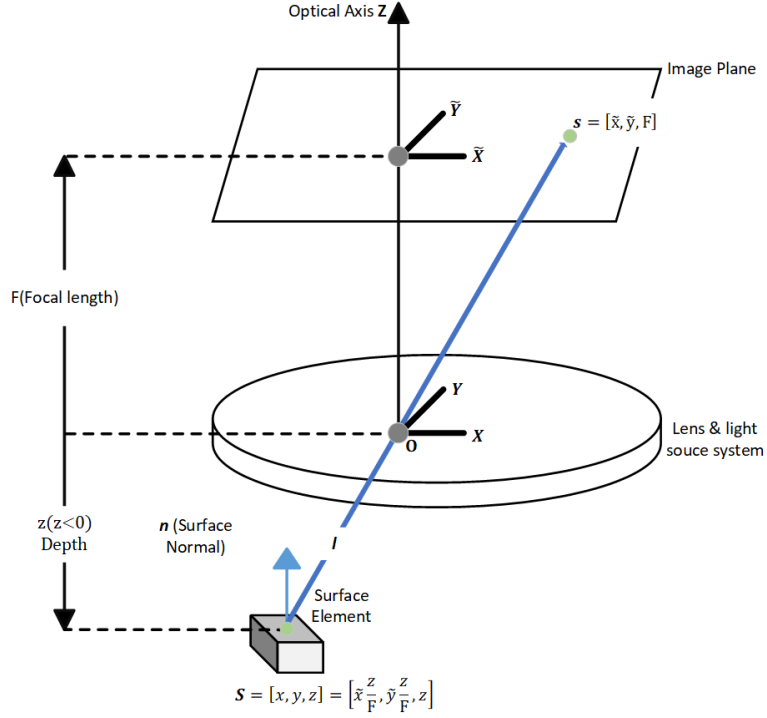
---

[1]https://www.blender.org/

Figure 1: Perspective shape from shading model with Light source at the camera projection center **O**. Camera coordinate system (x, y, z) is centered at **O**. z-axis is parallel to optical axis pointing towards image plane.

according to Lambert's cosine Law and inverse square fall of law of point light source [9],

$$R(\widetilde{x}, \widetilde{y}, z, p, q) = I_O \rho \left(\frac{\mathbf{n}(\widetilde{x}, \widetilde{y}, z, p, q) \cdot \mathbf{l}(\widetilde{x}, \widetilde{y}, z)}{r(\widetilde{x}, \widetilde{y}, z)^2} \right), \tag{1}$$

where I_O is the light intensity and ρ is the surface albedo. $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$ are the components of surface gradients. $\mathbf{n}$ is the surface unit normal and $\mathbf{l}$ is a unit vector representing the direction of the light ray incident at point **S**. r^2 is inverse square distance fall-off law of isotropic point light. The light source is considered at the camera center, but can easily be extended to multiple point light source not necessarily at the center [9].

The surface normal $\mathbf{n}$ can be represented in terms of partial derivatives of the depth z with respect to x and y [7]:

$$\mathbf{n} = \frac{[-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1]}{\sqrt{(\frac{\partial z}{\partial x})^2 + (\frac{\partial z}{\partial y})^2 + 1}}, \tag{2}$$

where (x, y, z) are camera coordinates. Under perspective projection we have,

$$x = \widetilde{x}\frac{z}{F} \qquad y = \widetilde{y}\frac{z}{F}, \tag{3}$$

where F is the focal length and $(\widetilde{x}, \widetilde{y})$ are image coordinates, and camera is pointing in the negative z-direction as depicted in Figure 1.

According to Horn and Brooks [7], the IIE is,

$$R(\widetilde{x}, \widetilde{y}, z, p, q) = I(\widetilde{x}, \widetilde{y}). \tag{4}$$

Equation (4) is solved to estimate z by minimizing the difference between image irradiance $I(\widetilde{x}, \widetilde{y})$ and reflectance map $R(\widetilde{x}, \widetilde{y}, z, p, q)$. Optimization is done on depth z whereas p and q are updated by taking the gradient of updated z. The relevant optimization problem is given by,

$$\arg\min_{z} E(z) = \lambda e_i(z) + (1 - \lambda)e_s(z), \tag{5}$$

where e_i is the irradiance error and e_s represents smoothness constraint. λ is the weighting factor between e_i and e_s.

$e_i(z)$ can be computed over the image domain $(\Omega \subset \mathbb{R})$ as,

$$e_i(z) = \int_{\Omega} (I(\widetilde{x}, \widetilde{y}) - R(\widetilde{x}, \widetilde{y}, z, p, q))^2 d\Omega, \tag{6}$$

$e_s(z)$ is solved with anisotropic diffusion [14], which is a non-linear, space-variant technique utilized to reduce the noise from the surface without smoothing edges, lines or other details which are important to interpret the surface. It is then combined with Equation (6) which is then solved with gradient decent. A small time step Δt is introduces to ensure stability with higher values of λ.

To impose anisotropic diffusion as a smoothness constraint, a 2×2 structure tensor is derived as a first step from the gradient of the depth z which is given as,

$$S_{i,j} = \frac{\partial z}{\partial x^i}\frac{\partial z}{\partial y^j}. \tag{7}$$

Afterwards, corresponding eigenvalues (λ_+, λ_-) and eigenvectors (θ_+, θ_-) are derived similar to [15]. From (λ_+, λ_-) and (θ_+, θ_-), diffusion tensor $\mathbf{D}$ is derived such as,

$$\mathbf{D} = \frac{\partial\psi}{\partial\lambda_+}\theta_+\theta_+^T + \frac{\partial\psi}{\partial\lambda_-}\theta_-\theta_-^T. \tag{8}$$

In terms of (λ_+, λ_-), langrangian density ψ can be written as [14],

$$e_s(z) = \int_{\Omega} \psi(\lambda_+, \lambda_-)d\Omega. \tag{9}$$

Equations (6) and (9) are combined together in Equation (5) and can be written as,

$$\arg\min_{z} E(z) = \int_{\Omega} (\lambda(I - R)^2 + (1 - \lambda)\psi(\lambda_+, \lambda_-))d\Omega, \tag{10}$$
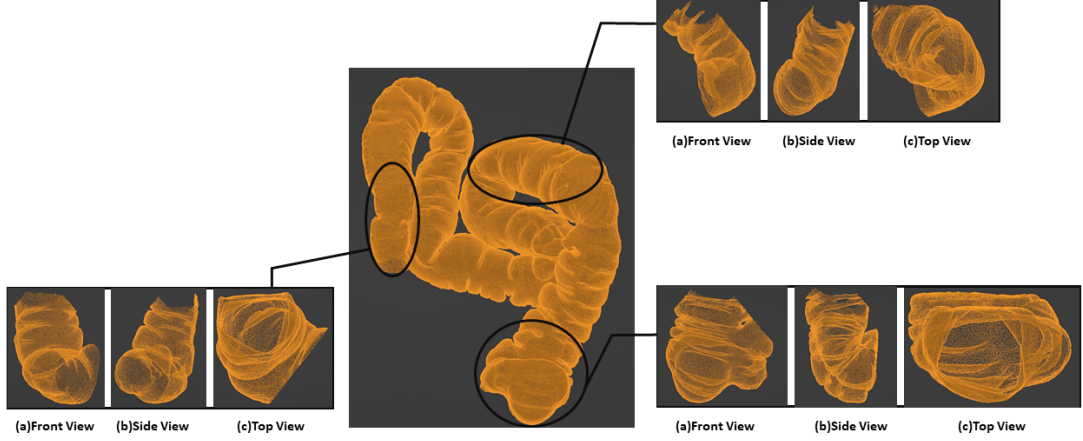
Figure 2: GI model.

The solution to Equation (10) is given by Euler-Lagrange PDE,

$$\lambda(I - R)\frac{\partial R}{\partial z} + (1 - \lambda)\nabla \cdot (\mathbf{D}\nabla z) = 0, \tag{11}$$

which we numerically solve by,

$$\frac{\partial z}{\partial t} = \nabla \cdot (\mathbf{D}\nabla z) + \frac{\lambda}{1 - \lambda}(I - R)\frac{\partial R}{\partial z}. \tag{12}$$

For simplicity and linearizing the diffusion tensor to compute it once and for all, structure tensor in Equation (8) is derived from the gray scale image $\upsilon(\widetilde{x}, \widetilde{y})$.

3. Results & Discussions

3.1. Ground Truth Models

Shape-from-shading algorithm is tested on different areas of synthetic GI regions [13]. The model is imported in Blender to render images of different areas of the model. Highlighted regions utilized for 3D reconstruction along with the model are shown in Figure 2. Blender is chosen not only to construct a ground truth scenario but also controlling different parameters like light intensity (I), focal length (F), which are needed for 3D reconstruction using SfS.

An environment is created similar to Figure 1. The camera is placed at $(0, 0, 0)$. F of the camera is set to 25mm. A point light source is also placed at camera center. Point light source is
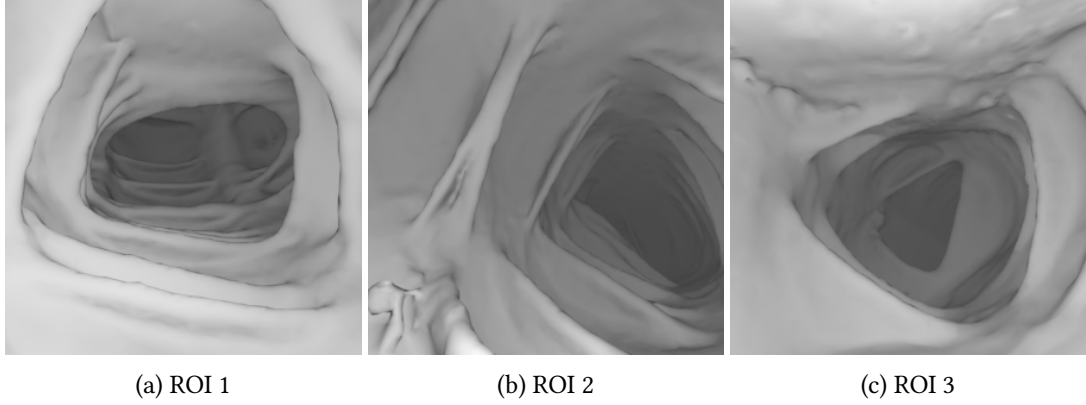
(a) ROI 1　　　　　　　　　(b) ROI 2　　　　　　　　　(c) ROI 3

Figure 3: Rendered images of GI regions

selected to imitate the illumination mechanism of pillcams which has four light sources around camera center. GI model is now cut into different regions of interest and placed under the camera at $z < 0$. Material properties of the model is set to be Diffuse BSDF with a constant albedo $\rho = 1$.

For a true comparison between the reconstructed surface and the ground truth, these respective regions 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, 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 as an obj format and then finally imported in MATLAB. These ground truth models are shown in Figure 4(a), 4(c), 4(e).

3.2. Assessment Criteria

In order to evaluate the quality of 3D reconstruction, the reconstructed surfaces are compared with ground truth by measuring depth error and correlation. These methods are chosen to assess different features of the reconstructed surfaces. Depth error is chosen because it correctly evaluates the geometric deformation of the reconstructed shape. Correlation is chosen because it evaluates the shape of the reconstructed surface independent of scale and position.

Correlation is computed by estimating variance and covariance of the recovered shape and the ground truth data. Whereas, Geometric deformation is investigated by measuring the average depth error (e_d) between the recovered shapes and the ground truth and is given by,

$$e_d = \frac{1}{\Omega} \sum \sum_{i,j \in \Omega} \left| \frac{\widehat{D}_{i,j} - D_{i,j}}{D_{i,j}} \right|,$$ (13)
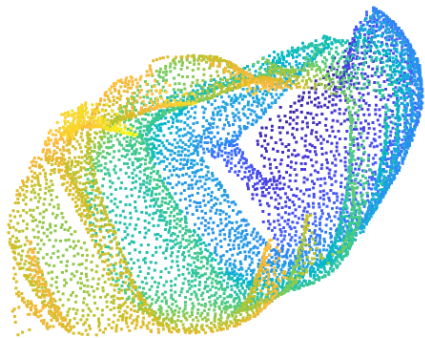
where D is the ground truth and $\widehat{D}$ is recovered 3D shape. Ω represents the region of the 3D model considered for error estimation.
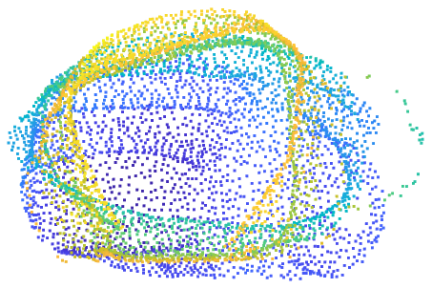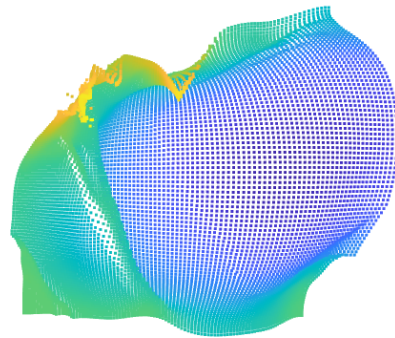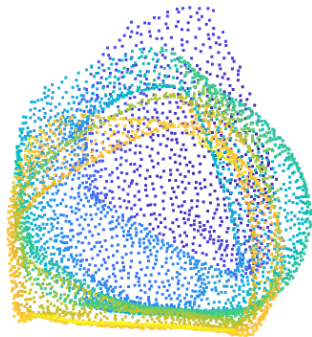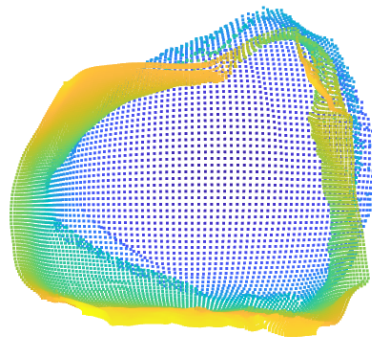
(a) GT 1
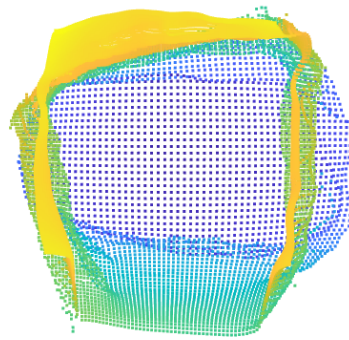
(b) RS 1

(c) GT 2

(d) RS 2

(e) GT 3

(f) RS 3

Figure 4: Ground truth (GT) comparison with recovered shapes (RS)

Table 1
Comparison between recovered shape and ground truth model.

G1 Regions	Depth Error	Correlation
ROI 1	0.4470	0.8883
ROI 2	0.3878	0.7927
ROI 3	0.3726	0.8725

3.3. Image Irradiance from Rendered Images

Images of size 100×100 of each model is rendered as shown in Figure 3. $I(\widetilde{x}, \widetilde{y})$ falling on the camera sensor is related to gray scale image $\upsilon(\widetilde{x}, \widetilde{y})$ via camera response function $r(\cdot)$ [9],

$$I(\widetilde{x}, \widetilde{y}) = \frac{r^{-1}[\upsilon(\widetilde{x}, \widetilde{y})]}{M(\widetilde{x}, \widetilde{y})}, \tag{14}$$

where $M(\widetilde{x}, \widetilde{y})$ is anisotropy of the light source. Point lights are perfectly isotropic by definition and so $M(\widetilde{x}, \widetilde{y}) = 1$. Images are saved in Portable Network Graphics (PNG) file format and therefore, image irradiance is just the gamma correction $\gamma = 2.2$ of the gray scale image i.e.,

$$I(\widetilde{x}, \widetilde{y}) = \upsilon^{\gamma}(\widetilde{x}, \widetilde{y}). \tag{15}$$

$I(\widetilde{x}, \widetilde{y})$ is also converted from pixel units to physical units in order to have correspondence between $I(\widetilde{x}, \widetilde{y})$ and R. Conversion to physical units is given by,

$$I_p(\widetilde{x}, \widetilde{y}) = \frac{I(\widetilde{x}, \widetilde{y}) - \min I(\widetilde{x}, \widetilde{y})}{\max I(\widetilde{x}, \widetilde{y}) - \min I(\widetilde{x}, \widetilde{y})} \times \left(\frac{I_O \cos \theta_1}{r_1^2} - \frac{I_O \cos \theta_2}{r_2^2} \right) + \frac{I_O \cos \theta_2}{r_2^2}, \tag{16}$$

where $I_p(\widetilde{x}, \widetilde{y})$ represents the physical value of the image irradiance and (θ_1, r_1) and (θ_2, r_2) decides the upper and lower bound of $I_p(\widetilde{x}, \widetilde{y})$. (θ_1, θ_2) are the angles between surface normal and light ray at the maximum and minimum point on the surface respectively. (r_1, r_2) are the distance from the light source to the maximum and minimum point on the surface respectively. These points are chosen by identifying the maximum and minimum lit area in the ground truth model and then computing the angles and the distances from the light source.

3.4. 3D Reconstruction

A flat surface is given as an initial condition for all three cases in order to test the robustness of the method. Initial reflectance map is then computed using Equation (1). Updated z values are calculated by solving Equation (12). The value of λ is different for different cases and empirical in our experiment but more weight is given on irradiance term to have a better reconstruction. 3D reconstruction of different areas of GI is shown in Figure 4(b), 4(d), 4(f).

Correlation and depth error are computed between recovered shapes and ground truth models for all three regions and shown in Table 1. Due to occlusion and dim light on lower areas of GI, some part of the recovered surfaces were smoothened and therefore, could not recover precisely. In spite of that, results are quite plausible because a high value of correlation and a low value

of depth error is attained for all three cases. Although, certain simplifying assumptions are made because authors are interested in proof of concept of using SfS to reconstruct complex GI geometry. However, considering a flat surface as an initial condition and reaching to the solution is quite reassuring to apply SfS technique on real capsule images.

4. Conclusion and Future Work

In this article, near light source perspective SfS method is applied on different GI regions. Given a reflection model, numerical scheme is formulated with anisotropic diffusion. Shape for each region is recovered and then compared with ground truth by measuring the average depth error and correlation. Result shows that SfS can handle complex geometries if penalized correctly.

In future work, SfS method will be applied on real capsule endoscopic images where we will have to deal with different textures, specularities, occlusion and distorted images. In addition to that, brightest and dimmest image points in physical units will be estimated for the right scale between $I(\widetilde{x}, \widetilde{y})$ and R. Radiometric calibration will be needed to compute the image irradiance and intensity of the light sources will also be measured. All of this will be essential to correctly implement SfS technique on real capsule images.

Acknowledgments

References

[1] D. E. Yung, J. N. Plevris, R. Leenhardt, X. Dray, A. Koulaouzidis, E. S. B. R. W. Group, et al., Poor quality of small bowel capsule endoscopy images has a significant negative effect in the diagnosis of small bowel malignancy, Clinical and Experimental Gastroenterology 13 (2020) 475.

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

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

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

[5] P. A. Floor, I. Farup, M. Pedersen, 3d reconstruction of the human colon from capsule endoscope video. (accepted), in: Colour and Visual Computing Symposium (CVCS), 2022.

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

[7] B. K. Horn, M. J. Brooks, The variational approach to shape from shading, Computer Vision, Graphics, and Image Processing 33 (1986) 174–208.

[8] R. Kimmel, J. A. Sethian, Optimal algorithm for shape from shading and path planning, Journal of Mathematical Imaging and Vision 14 (2001) 237–244.

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

[10] G. Iddan, G. Meron, A. Glukhovsky, P. Swain, Wireless capsule endoscopy, Nature 405 (2000) 417–417.

[11] A. Koulaouzidis, A. Karargyris, Three-dimensional image reconstruction in capsule endoscopy, World journal of gastroenterology: WJG 18 (2012) 4086.

[12] V. S. Prasath, I. N. Figueiredo, P. N. Figueiredo, K. Palaniappan, Mucosal region detection and 3d reconstruction in wireless capsule endoscopy videos using active contours, in: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2012, pp. 4014–4017.

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

[14] D. Tschumperlé, R. Deriche, Vector-valued image regularization with pdes: A common framework for different applications, IEEE transactions on pattern analysis and machine intelligence 27 (2005) 506–517.

[15] G. Sapiro, D. L. Ringach, Anisotropic diffusion of multivalued images with applications to color filtering, IEEE transactions on image processing 5 (1996) 1582–1586.