Solving Well-posed Shape from Shading Problem Using Implicit Neural Representations

Wanxin Bao
Department of Precision Engineering
The University of Tokyo
Tokyo, Japan
baowanxin@robot.t.u-tokyo.ac.jp

Ren Komatsu
Department of Precision Engineering
The University of Tokyo
Tokyo, Japan
komatsu@robot.t.u-tokyo.ac.jp

Atsushi Yamashita
Department of Precision Engineering
The University of Tokyo
Tokyo, Japan
yamashita@robot.t.u-tokyo.ac.jp

Hajime Asama
Department of Precision Engineering
The University of Tokyo
Tokyo, Japan
asama@robot.t.u-tokyo.ac.jp

***Abstract*—We propose a method for solving well-posed shape from shading problem by using implicit neural representations. We build an image irradiance equation and solve the equation by a sinusoidal representation network called SIREN, which is proposed by Sitzmann et al. in 2020. Object surface is expressed by Oren-Nayar model and a perspective projection model with light source located at the optical center is considered. Based on the above models, image irradiance equation is constructed, which is a partial differential equation (PDE). We introduce a neural network SIREN to solve this PDE, where implicit neural representations use the sine as a periodic activation function. Experiments are performed on three synthetic images and two real images. Results demonstrate that our proposed method performs with much higher accuracy.**

***Index Terms*—Shape from shading, implicit neural representations, partial differential equation, 3D reconstruction**

I. Introduction

To reconstruct the 3D shape of an object from its 2D image has been a fundamental task in computer vision. Techniques like simultaneous localization and mapping (SLAM) and structure from motion (SfM) can realize reconstruction of the environment. ORB-SLAM [1] was a feature-based SLAM and it used ORB features, which allowed real-time performance without GPUs, and showed good invariance to viewpoint. This system can reconstruct the environment in real time, in small and large, indoor and outdoor environments. Schoenberger et al. [2] proposed an incremental SfM, which can improve the robustness and accuracy of the reconstruction process while reducing the computational cost. However, such feature-based methods, which use multiple images to reconstruct the scene, may cause problems when there is little light in the environment or there is not enough texture for the surface. The reason is that feature extraction becomes difficult and then it is hard to correspond feature points. To overcome this difficulty, shape from shading (SfS) is an effective technique, which we will demonstrate in this study, because it is to use single intensity image to build the 3D environment.

SfS deals with reconstruction of the shape from a gradual variation of shading in the image [3]. Given a gray-scale image, SfS is to recover light source, surface normal, depth and albedo at each pixel of the image. To analyse the components of an image is not simple due to concave or convex ambiguities [4]. As a result, SfS problem has been regarded as an ill-posed problem for a long time. Even based on some assumptions that the surface is perfectly Lambertian surface and light direction is already known, solving the function between the brightness and shape is still a tough problem. SfS was introduced by Horn et al. [5] who first derived an equation describing the relationship between the shape of a surface and its corresponding brightness. The early days of SfS approaches [6] are based on the assumptions that (1) object surface is Lambertian surface, (2) light source is located at infinity, (3) the camera model is orthographic. Based on the above assumptions, they built an image irradiance equation and calculated the numerical solution. These assumptions simplified the complicated shading and imaging process, but in this way, reconstructed result had bad accuracy because it is impossible that object surface is perfect diffuse reflection. Moreover, most cameras are pin-hole cameras and light source is usually near the object, not at infinity. To improve the accuracy of SfS problem, Lee et al. [7] proposed a more realistic model. They assumed a new imaging model where a perspective camera projection and light source located at the optical center were performed. Okatani et al. [8] used an endoscope camera and proposed a notion of equal distance contour and obtained the equation for this contour.

Recently, many learning-based SfS have been proposed and they have achieved high accuracy. The 3D object reconstruction method based on deep learning is to train neural networks to learn the mapping relationship between 2D image and 3D object. Choy et al. [9] proposed a recurrent neural network architecture, which takes in one or multiple images of arbitrary viewpoints and outputs a 3D occupation grid result. It shows good performance even when the object has few texture. Wu et al. [10] represented 3D shape by using convolutional deep belief network. Their model learned complex 3D shape from raw CAD data and it realized shape completion from 2.5D depth maps.

In this study, our problem setting is that there is an object in a dark environment. A camera with flash light is used to take a photo and there is no other light source in the environment. As a result, we use the perspective projection model and assume the light source is located at the optical center. We think the object surface perform diffuse reflection and apply Oren-Nayer model [11] in our method. An attenuation term $1/r^2$ is also considered because it is also one of the plus to guarantee a well-posed SfS problem. After that, we derive an non-linear PDE and calculate its solution by sinusoidal representations network SIREN [12] where we use ADAM optimizer [13] for updating parameters of the network. Compared with previous works, our proposed method solve a well-posed SfS problem and has higher accuracy in calculating the depth of the objects.

II. RELATED WORKS

A. Well-posed SfS problem with numerical solution

There are many methods that deal SfS problems with numerical solution. Ahmed et al. [14] used a linear combination of Lambertian model and Ward [7] model to express the hybrid surface. Ikeda [15] applied a linear approximation to hybrid reflectance map composed of Lambertian model and Phong model [16]. Considering the reflectance model and projective camera model, they could obtain a partial differential equation (PDE) and they calculated its numerical solution. Wang et al. [6] proposed a hybrid reflectance model composed of a linear combination of Oren-Nayar model [11] and Ward model [17]. Incorporating perspective camera model and light source located at optical center, they built an image irradiance equation. It is also a Hamilton-Jacobi equation and they calculated its viscosity solution of the equation by using iterative sweeping method.

B. Learning-based SfS problem

Learning-based SfS problem deals with decomposing image components, including reflectance, surface normal, lighting, shading etc.. Sengupta et al. [18] proposed learning-based inverse rendering which jointly estimated albedo, surface normal and lighting. Their original residual appearance renderer (RAR) performed self-supervised learning using reconstruction loss. Janner et al. [19] proposed a shared convolutional encoder and three separated decoders for lighting, shape and reflectance. After that, shape and lighting predictions were used to train shading function. But they only used synthetic image dataset to finish their experiment. Li et al. [20] used synthetic data to train CNN-based intrinsic image models and generalized from synthetic data to real world images. However, in their method calculation cost was high. Generally speaking, learning-based SfS problem is still solving an ill-posed problem, which means it may cause ambiguities in their reconstruction process.

C. Gradient descent-based PDE solution

Deep learning has revolutionized fields such as images, text, and speech recognition. These fields required statistical approaches which can model non-linear functions of high-dimensional inputs. Deep learning, by using multi-layer neural networks, has proven effective in practice for many tasks. Deep learning can also be used for solving partial differential equation. Deep Galerkin method (DGM) [21] calculated high-dimensional PDE by deep neural network. By minimizing squared error, their method could solve high-dimensional Hamilton-Jacobi-Bellman PDE and Burger's equation. Deep Ritz method (DRM) [22] used two residual connections in their network, which guaranteed avoiding vanishing gradient problem.

In this paper, a photo of an object located in a dark environment is taken by a camera with flash light. We think the object surface performs diffuse reflection and apply Oren-Nayer model [11] in our method. We also use the perspective projection model and assume the light source is located at the optical center. The attenuation term $1/r^2$ is also considered because it is one of the plus to guarantee a well-posed SfS problem. Compared to [6], after obtaining the non-linear PDE, our method calculates its solution by implicit neural representations SIREN where ADAM optimizer is used for updating parameters of the network.

III. METHOD

Our study is to solve well-posed SfS problem by using implicit neural representations. The advantage is eliminating the ambiguity in SfS problem and improving accuracy for depth estimation. In this section, we summarized our method in two subsections. The first section introduces how to build the image irradiance equation which describes the relationship between light source, surface reflection and projection. The second half is about implicit neural representations for solving this equation.

A. Construct image irradiance partial differential equation

1) Perspective camera model: As we think a camera performs perspective projection and light source is located at the optical center [23], this can be modeled in Fig. 1.

The scene is represented by a surface $S(\boldsymbol{x})$. f is the focal length of the camera.

$$S(\boldsymbol{x}) = \frac{f u(\boldsymbol{x})}{\sqrt{f^2 + ||\boldsymbol{x}||^2}} \begin{pmatrix} \boldsymbol{x} \\ -f \end{pmatrix}, \tag{1}$$

where

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \Omega. \tag{2}$$

Ω is the given image domain. $u(\boldsymbol{x})$ represents the depth along the projection direction. For such a surface $S(\boldsymbol{x})$, we want to calculate the normal vector at the point $S(\boldsymbol{x})$. The

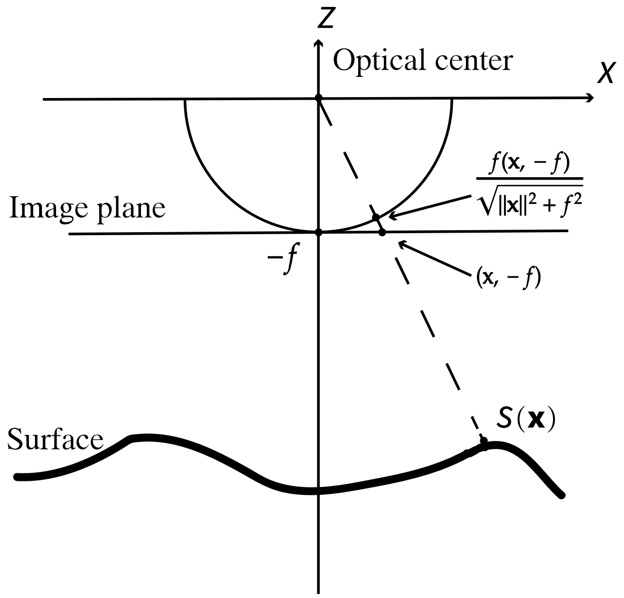

Fig. 1. Camera model with perspective projection and light source located at optical center.

normal vector can be obtained by compute tangent vectors in x_1 and x_2 directions.

$$\boldsymbol{n}(\boldsymbol{x}) = \begin{pmatrix} f\nabla u(\boldsymbol{x}) - \frac{f\nabla u(\boldsymbol{x})}{f^2+||\boldsymbol{x}||^2}\boldsymbol{x} \\ \boldsymbol{x}\nabla u(\boldsymbol{x}) + \frac{f\nabla u(\boldsymbol{x})}{f^2+||\boldsymbol{x}||^2} f \end{pmatrix}. \tag{3}$$

The single point light source is located at optical center, so the unit light vector $\boldsymbol{L}(\boldsymbol{x})$ can be described as

$$\boldsymbol{L}(\boldsymbol{x}) = \frac{1}{\sqrt{f^2+||\boldsymbol{x}||^2}} \begin{pmatrix} -\boldsymbol{x} \\ f \end{pmatrix}. \tag{4}$$

The term $\cos\theta_i$ is the dot product between $\boldsymbol{L}(\boldsymbol{x})$ and $\boldsymbol{n}(\boldsymbol{x})$, using the change of variables $v(\boldsymbol{x})$=ln$u(\boldsymbol{x})$

$$\begin{aligned} \theta_i &= \arccos \frac{\boldsymbol{n}(\boldsymbol{x})}{||\boldsymbol{n}(\boldsymbol{x})||}\boldsymbol{L}(\boldsymbol{x}) \\ &= \arccos \frac{Q(\boldsymbol{x})}{\sqrt{f^2||\nabla v(\boldsymbol{x})||^2 + (\boldsymbol{x}\nabla v(\boldsymbol{x}))^2 + Q^2(\boldsymbol{x})}}, \end{aligned} \tag{5}$$

where $Q(\boldsymbol{x}) = \frac{f}{\sqrt{f^2+||\boldsymbol{x}||^2}}$.

2) Oren-Nayar reflection model: We try to find a more realistic reflection model to express diffuse surfaces instead of applying Lambertian surface model in our method. Oren-Nayar model [11] was a comprehensive model for roughly diffuse object surface. It fully considered geometric and radiometric phenomena such as masking, shadowing and interreflections between points of the surface. This model followed the bidirectional reflectance distribution function (BRDF) and it assumed that the surface was composed of V-cavities. Each cavity consisted of two planar facets. Considering a Gaussian distribution of facet slopes and using the geometric relationship shown in Fig. 2, a simplified expression for the reflected radiance can be summarized as follows.

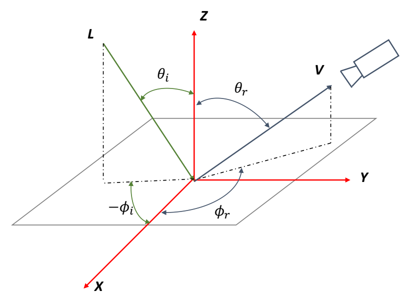

Fig. 2. Reflection geometry model. (θ_i, ϕ_i) is the incident direction $\boldsymbol{L}$, (θ_r, ϕ_r) is the reflected direction $\boldsymbol{V}$.

$$\begin{aligned} L_r(\theta_i, \phi_i; \theta_r, \phi_r) &= \frac{\rho}{\pi} I_0 \cos\theta_i \times \\ (A + B\max[0, \cos(\phi_r - \phi_i)] &\sin\alpha\tan\beta), \end{aligned} \tag{6}$$

where $A = 1 - 0.5\frac{\sigma^2}{\sigma^2+0.33}$, $B = 0.45\frac{\sigma^2}{\sigma^2+0.09}$, $\alpha = \max[\theta_i, \theta_r]$, $\beta = \min[\theta_i, \theta_r]$. Albedo ρ represents the fraction of incident energy that is reflected by the surface. I_0 means the intensity of point light source. The parameter σ is used to measure surface roughness.

As we assume the perspective camera model, $\theta_i = \theta_r = \alpha = \beta$, $\phi_i = \phi_r$. We also consider the attenuation term $1/r^2$ because it is another plus to guarantee a well-posed SfS problem, the reflected radiance equation becomes

$$L_r = \frac{\rho}{\pi}\frac{I_0}{r^2}(A\cos\theta_i + B\sin^2\theta_i). \tag{7}$$

The relationship between image brightness and surface radiance [8] is

$$E_i = L_r \frac{\pi}{4}\left(\frac{d}{f}\right)^2 \cos^4\chi, \tag{8}$$

where E_i is image irradiance, which is considered to be equal to image brightness. d is the diameter of the lens and f is the focal length. The term $\cos^4\chi$ implies nonuniform brightness even for uniform illumination, but actual optical system is designed to correct it. As a result, we may consider that image brightness is proportional to surface radiance:

$$E_i = \lambda L_r. \tag{9}$$

If we denote $I = \frac{E_i}{\lambda I_0}$, $k = \rho$, the brightness equation becomes

$$I = \frac{1}{r^2}\frac{k}{\pi}(A\cos\theta_i + B\sin^2\theta_i). \tag{10}$$

3) Build the image irradiance equation: Considering the above conditions, we can combine Eqs. (5) and (10) and build image irradiance equation.

$$-e^{-2v(\boldsymbol{x})} + \frac{f^2 I(\boldsymbol{x})}{D(\boldsymbol{x}, \nabla v)} = 0, \forall \boldsymbol{x} \in \Omega, \tag{11}$$

where
$D(\boldsymbol{x}, \nabla v) = \frac{\rho}{\pi}\left[A\frac{Q(\boldsymbol{x})}{\sqrt{F(\boldsymbol{x},\nabla v)+Q^2(\boldsymbol{x})}} + B\frac{F(\boldsymbol{x},\nabla v)}{F(\boldsymbol{x},\nabla v)+Q^2(\boldsymbol{x})}\right]$,
$F(\boldsymbol{x}, \nabla v) = f^2||\nabla v(\boldsymbol{x})||^2 + (\boldsymbol{x}\nabla v(\boldsymbol{x}))^2$,
$r = fu(\boldsymbol{x}) = fe^{v(\boldsymbol{x})}$.

B. Implicit neural representations to solve PDE

We used sinusoidal representation networks (SIREN) [12] to calculate the solution of Eq. (11). SIREN is a powerful paradigm which can represent implicitly defined differential signals. For function $\Phi(\boldsymbol{x})$ that satisfy equations of form

$$F(\boldsymbol{x}, \Phi(\boldsymbol{x}), \nabla\Phi(\boldsymbol{x}), ...) = 0, \tag{12}$$

we can learn a neural network that parameterizes $\Phi(\boldsymbol{x})$ to map $\boldsymbol{x}$ while satisfying the constrained condition in Eq. (12). The input for the network is the spatial coordinate $\boldsymbol{x} \in R^m$. $\Phi(\boldsymbol{x})$ is implicitly defined by the function F.

We parameterize Φ_θ as fully connected neural network with parameters θ, and solve the optimization problem by gradient descent. Our neural network is a four layers network with one layer of input, three hidden layers. All layers are multi-layer perceptrons (MLPs) with sinusoidal function as the activation function. For pixel coordinate $\boldsymbol{x}_i = (x_i, y_i)$, we construct the loss function

$$J(\Phi) = || - e^{-2\Phi_\theta(\boldsymbol{x}_i)} + \frac{f^2 I(\boldsymbol{x}_i)}{D(\boldsymbol{x}_i, \nabla\Phi)}||^2_\Omega. \tag{13}$$

$J(\Phi)$ measures how well the function $\Phi_\theta(\boldsymbol{x})$ satisfies the PDE differential operator. If $J(\Phi) = 0$, then $\Phi_\theta(\boldsymbol{x})$ is a solution to PDE.

The goal is to find a set of parameters θ so that $\Phi_\theta(\boldsymbol{x})$ minimizes the loss function $J(\Phi)$. If the loss $J(\Phi)$ is small, then $\Phi_\theta(\boldsymbol{x})$ will closely satisfy the PDE. In our proposed method, parameters θ are updated by using the well-known ADAM algorithm, which was proved to be a very effective optimizer in machine learning. It yields fast and robust convergence in our work and the hyper-parameters for ADAM have intuitive interpretations and typically require little tuning.

IV. Simulation Experiment

To evaluate our proposed method, a simulation experiment was carried out. We compared our work with Wang's method [6]. All algorithms were implemented by using Python code. Our proposed method was performed on Google Colab.

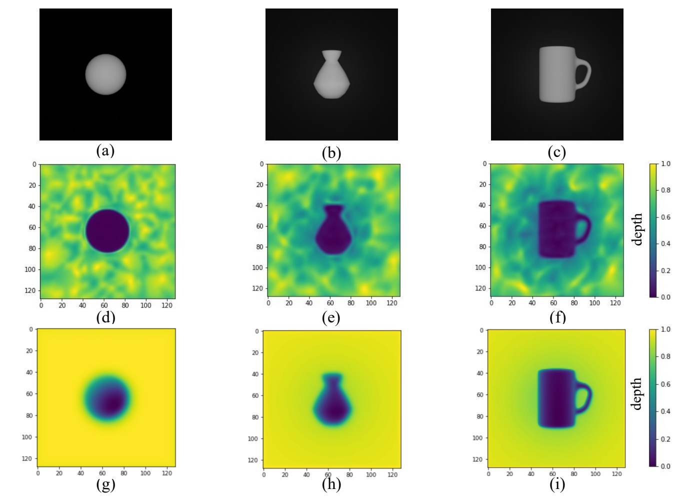

Fig. 3. Experiment results: (a)-(c) shading synthetic images of sphere, vase and mug; (d)-(f) normalized depth maps by proposed method correspondingly; (g)-(i) normalized depth maps by Wang's method correspondingly.

A. Experimental setup

We used an open-source software Blender to create 3D models of a sphere, a vase and a mug, which are shown in Fig. 3(a), Fig. 3(b) and Fig. 3(c), respectively. These models are classical objects in SfS study [6] [14] [23]. We set the roughness σ of all three models as 0.5. All images are presented with a resolution of 128×128 pixels. Point light source and focal center of the pin-hole camera sit at the same spatial location.

B. Results and discussion

Figures 3(d)(e)(f) are depth maps by our proposed method for sphere, vase and mug, respectively. We could see that our proposed method reconstruct objects' shape successfully and the shape of the objects look clearly. However, there seems to be many noises in the surrounding pixels of objects. We also implemented Wang's method [6] to see how well-posed SfS problem performed by using numerical method to solve its solution. As illustrated in Fig. 3(g), Fig. 3(h) and Fig. 3(i), they are depth maps by Wang's method for sphere, vase and mug respectively. We find that their method shows good performance for the surrounding pixels of objects. However, the edge of the objects looks blurred.

In order to further evaluate the performance of the proposed method, a quantitative comparison between our method and Wang's method is employed by using mean absolute error (MAE) and root mean square error (RMSE) for three synthetic images. We obtained the ground truth of depth maps for three synthetic images by Blender. We used a mask to cover the surrounding pixels of the original images and calculated MAE and RMSE by using the pixels where objects occupied. TABLE I lists the comparison of depth estimation for our proposed method and Wang's method. For the same iterations, our proposed method shows more superiority than Wang's

TABLE I
MAE AND RMSE COMPARISON OF OUR PROPOSED METHOD AND WANG'S METHOD FOR THREE SYNTHETIC IMAGES.

Images	Iterations	Proposed method		Wang's method	
		MAE	RMSE	MAE	RMSE
Fig. 3(a)	50	0.0857	0.1515	0.2183	0.2740
Fig. 3(b)	40	**0.0466**	0.1243	0.1577	0.2070
Fig. 3(c)	100	0.0471	**0.1075**	0.0903	0.1309

method in the reconstructed errors for all three tested images. The MAE is given by

$$MAE = \frac{1}{n} \times \sum_{i=1}^{n} |\hat{D}_i - D_i|, \tag{14}$$

where $\hat{D}_i$ is estimated value, D_i is actual value.
The RMSE is given by

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T} (\hat{D}_t - D_t)^2}{T}}, \tag{15}$$

where $\hat{D}_t$ is estimated value, D_t is actual value.

In our proposed method, we used ADAM [13] optimizer for iteration, which was proved to be a very effective optimizer in machine learning. It yields fast and robust convergence in our work and the hyper-parameters for ADAM have intuitive interpretations and typically require little tuning. On the other hand, in Wang's method, they applied Newton method to iterate. Their method was also based on 2D numerical Hamiltonian and fixed-point iterative sweeping method. However, their method was dependent on initial value and some artificial viscosities, which may influence the solution of the PDE. Moreover, due to the fixed-point iterative method, value of every pixel was dependent on its neighboring value result, which may be the reason why the edge of the objects looks blurred in their depth maps.

V. REAL EXPERIMENT

A. Experimental setup

To prove the effectiveness of our proposed method, we also used iPhone XS to take two real images of vase and David's statue for experiment which are shown in Fig. 4(a) and Fig. 4(b), respectively. The material of the vase and David's statue is plaster. In a real dark environment, we used iPhone's flash as the only light source for taking photos. Resolution of two real images are also 128×128 pixels.

B. Results and discussion

Figures 4(c) and (d) show depth maps by our proposed method for vase and David's statue. Figures 4(e) and (f) illustrate depth maps by Wang's method for vase and David's statue. Similar to the results on synthetic images, depth maps by our proposed method show clear edge for both images, but there are some noises for the surrounding pixels. Wang's method shows good performance for surrounding pixels while edge of the objects looks blurred. As the ground truth is

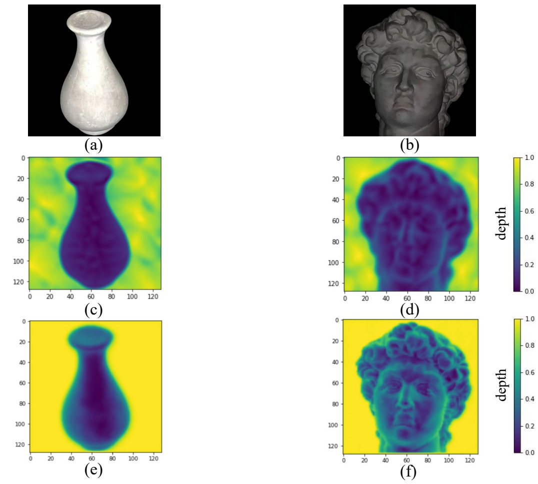

Fig. 4. Experiment results: (a)-(b) real images of vase and David's statue obtained by iPhone; (c)-(d) normalized depth maps by proposed method correspondingly; (e)-(f) normalized depth maps by Wang's method correspondingly.

unavailable for two real images, we cannot calculate the MAE or RMSE for these two real images.

VI. CONCLUSIONS

In this paper, we proposed a method of solving well-posed SfS problem by using implicit neural representations. We explained building the image irradiance equation and solve the equation by SIREN. Object surface is expressed by Oren-Nayar model and a perspective projection model with light source located at the optical center is considered. Based on the above models, image irradiance equation is constructed, which is a partial differential equation. We introduce a neural network SIREN to solve this PDE, where implicit neural representations use the sine as a periodic activation function. Experiments are performed on three synthetic images and two real images. Results demonstrate that our proposed method performs with high accuracy when estimating depth of the objects.

Possible real world application for SfS study includes face reconstruction [24] where shape, light and reflectance could be obtained and [25] where depth map of faces could be estimated. [8] also provided another application that SfS study of using endoscopic camera estimates stomach wall shape so that more quantitative evaluations and morphological studies can be performed further.

Some noises occur for the surrounding pixels in our proposed approach. Eliminating surrounding noises may be one of our future works. Future work may also include integrating our result into learning-based SfS approach. By solving an ill-posed problem, image components including reflectance, lighting, shading are obtained in learning-based method. We think our work of solving well-posed SfS by implicit neural representations could improve the accuracy performance of learning-based SfS methods.

References

[1] R. Mur-Artal, J. M. M. Montiel and J. D. Tardos: "ORB-SLAM: a versatile and accurate monocular SLAM system," IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147–1163, 2015.

[2] J. L. Schonberger and J. M. Frahm: "Structure-from-motion revisited," Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104–4113, 2016.

[3] R. Zhang, P. S. Tsai, C. E. Cryer and M. Shah: "Shape from shading: A survey," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 8, pp. 690–706, 1999.

[4] P. N. Belhumeur, D. J. Kriegman and A. L. Yuille: "The bas-relief ambiguity," International Journal of Computer Vision, vol. 35, no. 1, pp. 33–44, 1999.

[5] B. K. P. Horn: "Shape from shading: a method for obtaining the shape of a smooth opaque object from one view," PhD Thesis, Department of Electrical Engineering and Computer Science Massachusetts Institute of Technology, 1970.

[6] G. Wang and J. Cheng: "Three-dimensional reconstruction of hybrid surfaces using perspective shape from shading," Optik - International Journal for Light and Electron Optics, vol. 127, no. 19, pp. 7740–7751, 2016.

[7] K.M. Lee and C. C. J. Kuo: " Shape from shading with a generalized reflectance map model," Computer Vision and Image Understanding, vol. 67, no. 2, pp. 143–160, 1997.

[8] T. Okatani and K. Deguchi: " Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center," Computer Vision and Image Understanding, vol. 66, no. 2, pp. 119–131, 1997.

[9] C. B. Choy, D. Xu, J. Gwak, K. Chen and S. Savarese: "3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction," Lecture Notes in Computer Vision, pp. 628–644, 2016.

[10] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao: "3D ShapeNets: A Deep Representation for Volumetric Shapes," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1912–1920, 2015.

[11] M. Oren and S. K. Nayar: "Generalization of the Lambertian model and implication in machine vision," Internation Journal of Machine Vision, vol. 14, pp. 227–251, 1995.

[12] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell and G. Wetzstein: "Implicit neural presentations with periodic activation functions," Advances in Neural Information Processing Systems 33, 2020.

[13] D. P. Kingma and J. L. Ba: "ADAM: A method for stochastic optimization," Proceedings of the International Conference for Learning Presentations, 2015.

[14] A. Ahmed and A. Farag: "Shape from shading for hybrid surfaces," Proceedings of the 2007 IEEE International Conference on Image Processing, pp. 525–528, 2007.

[15] O. Ikeda: "Shape distortion analysis of the shape-from-shading algorithm using Jacobi iterative method," Proceedings of the 1st International Symposium on 3D Data Processing, Visualization and Transmission, pp. 396–399, 2002.

[16] B. T. Phong: "Illumination for computer generated pictures," Communications of the ACM, vol. 18, no. 6, pp. 311–317, 1975.

[17] G. J. Ward: "Measuring and modeling anisotropic reflection," Proceedings of the 19th annual conference on Computer Graphics and Interactive Techniques, vol. 26, no. 265–272, 1992.

[18] S. Sengupta, J. Gu, K. Kim, G. Liu, D. W. Jacobs and J. Kautz: "Neural Inverse Rendering of an Indoor Scene from a Single Image," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8598–8607, 2019.

[19] M. Janner, J. Wu, T.D. Kulkarni, I. Yildirim and J. B. Tenenbaum: "Self-supervised intrinsic image decomposition," Advances in Neural Information Processing System, 2017.

[20] Z. Li and N. Snavely: "CGIntrinsics: Better intrinsic image decomposition through physically-based rendering," Proceedings of the European Conference on Computer Vision, 2018.

[21] J. Sirignano and K. Spiliopoulos: "DGM: A deep learning algorithm for solving partial differential equations," Journal of Computational Physics, vol. 375, pp. 1339–1364, 2018.

[22] W. E and B. Yu: "The Deep Ritz Method: A Deep Learning-Based Numerical Algorithm for Solving Variational Problems," Communications in Mathematics and Statistics, vol. 6, no. 1, pp. 1–12, 2018.

[23] E. Prados and O. Faugeras: "Shape from shading: a well-posed problem?" Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 870–877, 2005.

[24] S. Sengupta, A. Kanazawa, C. D. Castillo and D. W. Jacobs: "SfSNet: Learning shape, reflectance and illumination of faces 'in the wild'," Proceedings of the IEEE on Computer Vision and Pattern Recognition, pp. 6296–6305, 2018.

[25] R. Kimmel, K. Siddiqi, B. B. Kimia and A. M. Bruckstein: "Shape from shading: level set propogation and viscosity solutions," International Journal of Computer Vision, vol. 16, pp. 107–133, 1995.