Anisotropic Diffusion for Depth Estimation in Shape from Focus Systems

Bilal Ahmad[a], Ivar Farup[b] and Pål Anders Floor[c]
Department of Computer Science, Norwegian University of Science & Technology, 2815 Gjøvik, Norway
{*bilal.ahmad, ivar.farup, paal.anders.floor*}*@ntnu.no*

Keywords: Anisotropic Diffusion, Shape from Focus, 3D Reconstruction.

Abstract: Shape from focus is a monocular method that uses the camera's focus as the primary indicator for depth estimation. The initial depth map is usually improved by penalizing the L2 regularizer as a smoothness constraint, which tends to smoothen the structural details due to linear diffusion. In this article, we propose an energy minimization-based framework to improve the initial depth map by utilizing a nonlinear, spatial technique, called anisotropic diffusion as a smoothness constraint, which is combined with a fidelity term that incorporates the focus values of the initial depth to enhance structural aspects of the observed scene. Experiments are conducted on synthetic and real datasets which demonstrate that the proposed method can significantly improve the depth maps.

1 INTRODUCTION

3D reconstruction is a challenging problem within the domain of computer vision. It can be effectively addressed by employing different techniques or algorithms to obtain spatial information about an object or a scene. Vision-based depth estimation methods are generally categorized into different approaches. Some methods rely on monocular image analysis techniques such as texture gradient analysis (Verbin and Zickler, 2020), and photometric methods (Ahmad et al., 2022). Furthermore, some methods leverage multiple images, relying on camera motion or various relative positions (Özyeşil et al., 2017), while others involve using image focus as a cue to determine the depth of the scene (Ali and Mahmood, 2021). The widespread application of 3D reconstruction can be found in various fields, including measurement systems, robotics, medical diagnostics, video surveillance, and monitoring, among others (He et al., 2022), (Ahmad et al., 2023b).

Shape from focus (SFF) is one of the passive monocular techniques that recovers the depth or 3D shape of an object through the analysis of an image sequence captured by manipulating the focus settings of the camera. In SFF, the key step involves identifying the sharpest and best-focused pixels from an image sequence using a specialized operator known as the Focus Measure (FM) operator. These selected focused pixels serve as a main cue for estimating the depth information of the objects.

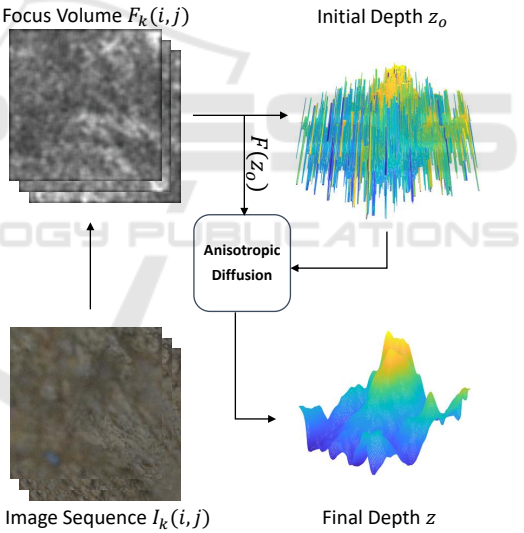


Figure 1: Proposed SFF system.

The SFF method was introduced by Nayar *et al.* (Nayar and Nakagawa, 1994), who computed the focus values of pixels by applying Laplacian operations on the images and subsequently improved the depth map using a Gaussian distribution method. Ali *et al.* (Ali and Mahmood, 2022) proposed a method to optimize focus volume by designing an energy minimization function which contains smoothness and structural similarity along with the data term. Some deep-

[a] https://orcid.org/0000-0001-8811-0404
[b] https://orcid.org/0000-0003-3473-1138
[c] https://orcid.org/0000-0001-6328-7414

Ahmad, B., Farup, I. and Floor, P.
Anisotropic Diffusion for Depth Estimation in Shape from Focus Systems.
DOI: 10.5220/0012303400003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 4: VISAPP*, pages 85-89
ISBN: 978-989-758-679-8; ISSN: 2184-4321

learning-based methods have also been proposed to estimate the depth maps using the SFF method (Mutahira et al., 2021b), (Dogan, 2023).

A variety of techniques have been proposed to improve the initial depth map by enforcing a smoothness constraint. Tseng *et al.* (Tseng and Wang, 2014) introduced a maximum a posteriori framework, which integrates a prior spatial consistency model for depth reconstruction. Moeller *et al.* (Moeller et al., 2015) proposed a variational method which include non-convex data fidelity term and a convex non-smooth regularization term to remove noise and improve depth maps. The smoothness constraint is usually employed with the L2 regularization, which tends to smoothen the sharp edges and other structural details due to linear diffusion. Therefore, in this paper, we emphasize on preserving the structural edges and fine details in enhanced depth maps.

This article provides an efficient method for improving the initial depth map using an energy minimization-based framework. The framework introduces a nonlinear, spatial technique, known as anisotropic diffusion (AD) (Perona and Malik, 1990), to serve as a smoothness constraint. The primary goal of AD in our application is to reduce surface noise while preserving the edges, lines, and other critical details necessary for accurate surface interpretation in the SFF systems. AD can be considered as a fusion of L2 and L1 regularization techniques. Similar to the L2 regularization, it promotes smoothing within regions of an image with similar intensity values, preserving image structure and reducing noise. At the same time, it shares similarities with L1 regularization, as it selectively preserves sharp edges and critical features. This combination allows AD to strike a balance between retaining fine details while effectively minimizing unwanted noise. The AD is combined with a fidelity term consisting of the focus values of the initial depth to iteratively converge the incorrect depth points to their true depth values. The proposed method is rigorously evaluated using both real-world and synthetic datasets and also compared with the L2 regularizer.

The remainder of the article is organized as follows. Section 2 explains the proposed methodology. In Section 3 both real and synthetic datasets are discussed. Initial and improved depth maps are also compared with each other, and lastly, Section 4 concludes the article.

2 PROPOSED METHOD

The proposed SFF system is depicted in Figure 1. In the first step, an initial depth map is computed by applying the traditional SFF method (Pertuz et al., 2013). For this purpose, an FM operator is applied to the image sequence. An FM operator serves as a high-pass filter, effectively isolating the high-frequency content from the low-frequency content. It achieves this by enhancing the focused pixels while suppressing the defocused ones. The operator calculates the sharpness of each pixel within the image sequence $I(i,j)$ to obtain an image focus volume $F(i,j)$, which is given as,

$$F_k(i,j) = FM * I_k(i,j), \tag{1}$$

where, $*$ is a 2D convolution operator and k represents an image index. Using the image focus volume, a depth map is constructed by traversing along the optical axis and identifying the image numbers that correspond to the maximum focus value for each object point. The resulting depth map is called the initial depth map z_o, which is given as,

$$z_o(i,j) = \arg\max_k(F_k(i,j)). \tag{2}$$

We propose the improvement of the z_o through the formulation of an energy minimization-based framework. The energy model consists of a smoothness term and a data fidelity term, which is given as follows,

$$E(z,\nabla z) = E_s(\nabla z) + \lambda E_f(z), \tag{3}$$

where, λ is a weighting factor between the fidelity term and the smoothness term. The objective is to obtain an optimized depth z that minimizes the energy function provided in Equation (3). The fidelity term $E_f(z)$ is designed using z_o and focus values of initial depth points $F(z_o)$. $E_f(z)$ is computed over the image domain ($\Omega \subset \mathbb{R}^2$) which is given as,

$$E_f(z) = \int_\Omega F(z_o) \mid z - z_o \mid^2 d\Omega. \tag{4}$$

The Equation (4) is designed to prioritize the best-focused regions in the image, contributing to provide a more reliable prior within the minimization framework presented in Equation (3).

To impose AD as a smoothness constraint, a 2×2 structure tensor $\mathbf{S}$ is derived (Di Zenzo, 1986), as a first step from the gradient of the initial depth z_o, and then updated with subsequent z's in each iteration. The $\mathbf{S}$ is given as,

$$\mathbf{S} = \nabla z \otimes \nabla z, \tag{5}$$

where, $\otimes$ represents the tensor product. Subsequently, we obtain the eigenvalues (λ_+, λ_-) and eigenvectors

(θ_+, θ_-) of the **S**, similar to (Sapiro and Ringach, 1996). The diffusion tensor **D** is derived from (λ_+, λ_-) and (θ_+, θ_-), which is given as,

$$\mathbf{D} = \frac{\partial \psi}{\partial \lambda_+} \theta_+ \otimes \theta_+ + \frac{\partial \psi}{\partial \lambda_-} \theta_- \otimes \theta_-. \quad (6)$$

In terms of (λ_+, λ_-), the Lagrangian density ψ can be written as (Tschumperlé and Deriche, 2005),

$$E_s(\nabla z) = \int_\Omega \psi(\lambda_+, \lambda_-) d\Omega. \quad (7)$$

Equation (7) is combined with Equation (4), which can be written as,

$$E(z, \nabla z) = \int_\Omega \psi(\lambda_+, \lambda_-) + \lambda F(z_o) \mid z - z_o \mid^2 d\Omega. \quad (8)$$

The solution to Equation (8) is given by Euler-Lagrange PDE,

$$\nabla \cdot (\mathbf{D} \nabla z) - \lambda F(z_o)(z - z_o) = 0. \quad (9)$$

Equation (6) is solved with gradient descent such as,

$$\frac{\partial z}{\partial t} = \nabla \cdot (\mathbf{D} \nabla z) - \lambda F(z_o)(z - z_o), \quad (10)$$

using explicit Euler time integration. It is to be noted that the proposed method is not restricted to any particular FM operator. Instead, it is designed to be generic, allowing the utilization of any FM operator to compute the initial depth z_o.

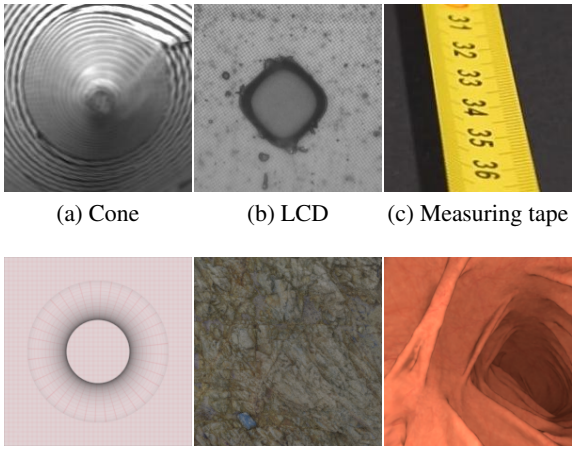
3 EXPERIMENTS

3.1 Datasets

The proposed method is tested on both synthetic and real datasets, shown in Figure 2. The real dataset contains the images of Real cone, LCD-TFT filter, and Measuring tape. Real cone and LCD-TFT filter are taken from (Mutahira et al., 2021a), whereas Measuring tape is taken from (PureMoCo, 2017). For synthetic data, three distinct models are selected: a torus, a mountain, and a colon region. The torus and the mountain models are constructed using Blender[1] software. Meanwhile, the colon model is taken from (İncetan et al., 2021) to demonstrate the viability of the SFF method as a potential 3D reconstruction approach for future Wireless Capsule Endoscope (WCE) applications, especially with the prospect of focus-controlled cameras becoming available in the future (Ahmad et al., 2023b).

The synthetic models are placed in front of a focus controlled camera and images are taken by changing the focus distance of the camera with a constant step size. In each image, a certain area of the scene is kept in focus, while the rest remain defocused. For each object, a series of 15 to 20 images are captured and stored in Portable Network Graphics (PNG) file format.

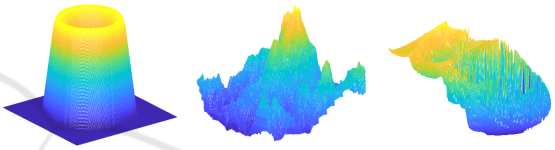
[1]https://www.blender.org/



(a) Cone (b) LCD (c) Measuring tape

(d) Torus (e) Mountain (f) Colon

Figure 2: Sample images from three real (a-c) and three synthetic (d-f) image sequences.



(a) Torus (b) Mountain (c) Colon

Figure 3: Ground truth depth maps for synthetic image sequences.

To establish a meaningful comparison between the reconstructed surfaces and the ground-truth models, the Python API in Blender is used to modify the models accordingly. When a model is positioned under a perspective camera, certain vertices or areas may become occluded, falling outside the camera's field of view. Therefore, to accurately assess the accuracy of the SFF algorithm, it becomes imperative to exclude all occluded vertices and construct a model that comprises only those vertices situated within the camera frustum and visible to the camera. The modified ground truth models of synthetic data are shown in Figure 3.

3.2 Results

The modified Laplacian method (Nayar and Nakagawa, 1994) is applied to the image stack to compute the focus volume. The initial depth map z_o is reconstructed using Equation (2). Both z_o and $F(z_o)$ are

Table 1: Quantitative evaluation for synthetic dataset.

Objects	RMSE			Correlation		
	Initial	L2	AD	Initial	L2	AD
Torus	0.0050	0.0036	**0.0032**	0.9599	0.9797	**0.9835**
Mountain	0.0167	0.0081	**0.0073**	0.7782	0.9521	**0.9612**
Colon	0.1227	0.1010	**0.0947**	0.8760	0.9179	**0.9281**

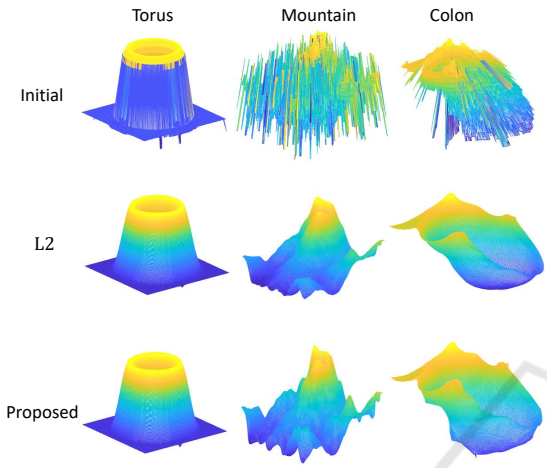


Figure 4: Reconstructed depth maps for synthetic data.

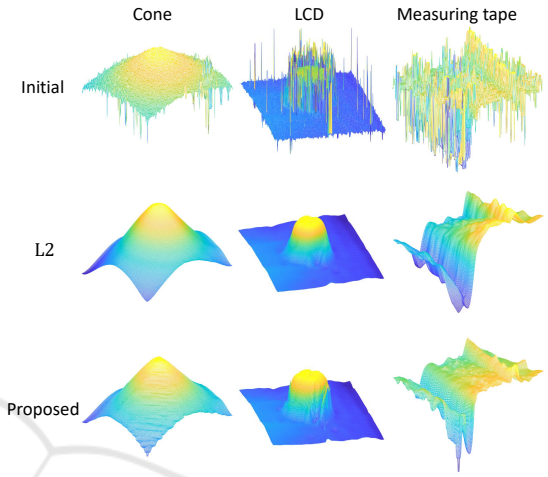


Figure 5: Reconstructed depth maps for real data.

subsequently employed in Equation (10) to rectify inaccurate depth points. The value of λ is different for different cases and empirical in our experiment. The proposed method is also compared with the L2 regularizer (Ahmad et al., 2023a) to show the effectiveness of the method. The surfaces reconstructed from synthetic and real data are shown in Figures 4 and 5, respectively, where the first row shows the initial depth, while the second and third rows show the depth recovered using the L2 and the proposed method, respectively.

The depth maps for synthetic data are compared with the ground-truth models by measuring the RMSE and the correlation. The selection of these methods has been made to evaluate different aspects of the reconstructed surfaces. The correlation is chosen to assess the quality of the reconstructed shapes, independent of the scale and position. RMSE is scale-dependent and evaluates the geometric deformation of the reconstructed shapes. Table 1 shows the quantitative evaluation of the synthetic dataset. The proposed method achieves higher correlation and lower RMSE for all three objects. The proposed method successfully addresses incorrect depth points, resulting in significant improvements in reconstruction.

The initial depth map, obtained from the focus values, exhibits numerous inaccuracies, possibly due to low-frequency variations in certain areas of the objects. As a consequence, the focus values acquired in those regions are erroneous, leading to incorrect depth points. With the proposed method, depth points with higher focus values are trusted and retained in their original positions. Conversely, depth points with lower focus values are mistrusted and neighboring depth values are given more significance to adjust their depth values. This iterative procedure facilitates the gradual convergence of erroneous depth points towards their true depth values, thereby yielding a refined and more accurate depth map. The proposed method improves the overall precision and reliability of the results, as can be confirmed by examining Table 1 and by visual analysis of Figure 4 and Figure 5.

The proposed method is also compared with the L2 regularizer. Although the L2 regularizer has demonstrated an improvement in the initial depth map, it is observed that the application of the L2 regularizer resulted in smoothing of the intricate details within the structure, as can be confirmed by visually inspecting Figure 4 and Figure 5. On the other hand, the proposed method manages to preserve a significant portion of these fine details. Furthermore, the proposed method exhibits an overall increase in accuracy of almost 10% over the L2 regularizer in terms of RMSE.

4 CONCLUSION

This article presents an energy minimization-based framework to improve the depth map for the SFF method. The framework has been formulated with AD as a smoothness constraint and a fidelity term, which incorporates the focus value of the initial depth to improve the overall structure of the scene. Experiments are conducted with real and synthetic datasets. For synthetic dataset, both z_o and z are also compared with ground truth by measuring RMSE and correlation. The results indicate that the proposed method can significantly improve the accuracy of the depth map by removing noise and preserving the structural details of the scene. The proposed method is also compared with the L2 regularizer, demonstrating a substantial 10% improvement in terms of RMSE over it.

ACKNOWLEDGEMENTS

This work was supported by the Research Council of Norway through the project CAPSULE under Grant 300031.

REFERENCES

Ahmad, B., Farup, I., and Floor, P. A. (2023a). 3D reconstruction of gastrointestinal regions using shape-from-focus. In *Fifteenth International Conference on Machine Vision (ICMV 2022)*, volume 12701, pages 463–470. SPIE.

Ahmad, B., Floor, P. A., and Farup, I. (2022). A comparison of regularization methods for near-light-source perspective shape-from-shading. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3146–3150. IEEE.

Ahmad, B., Floor, P. A., Farup, I., and Hovde, Ø. (2023b). 3D reconstruction of gastrointestinal regions using single-view methods. *IEEE Access*.

Ali, U. and Mahmood, M. T. (2021). Robust focus volume regularization in shape from focus. *IEEE Transactions on Image Processing*, 30:7215–7227.

Ali, U. and Mahmood, M. T. (2022). Energy minimization for image focus volume in shape from focus. *Pattern Recognition*, 126:108559.

Di Zenzo, S. (1986). A note on the gradient of a multi-image. *Computer vision, graphics, and image processing*, 33(1):116–125.

Dogan, H. (2023). A higher performance shape from focus strategy based on unsupervised deep learning for 3d shape reconstruction. *Multimedia Tools and Applications*, pages 1–24.

He, M., Zhang, J., Shan, S., and Chen, X. (2022). Enhancing face recognition with self-supervised 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4062–4071.

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. (2021). VR-Caps: A virtual environment for capsule endoscopy. *Medical image analysis*, 70:101990.

Moeller, M., Benning, M., Schönlieb, C., and Cremers, D. (2015). Variational depth from focus reconstruction. *IEEE Transactions on Image Processing*, 24(12):5369–5378.

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

Mutahira, H., Muhammad, M. S., Li, M., and Shin, D.-R. (2021b). A simplified approach using deep neural network for fast and accurate shape from focus. *Microscopy Research and Technique*, 84(4):656–667.

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

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

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639.

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

PureMoCo (2017). PureMoCo. *Focus-zoom-unit (motorized follow focus) 1:15–2:23. YouTube*. https://www.youtube.com/watch?v=KFryXjYbTJc. Accessed: 2020-11-06.

Sapiro, G. and Ringach, D. L. (1996). Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE transactions on image processing*, 5(11):1582–1586.

Tschumperlé, D. and Deriche, R. (2005). Vector-valued image regularization with pdes: A common framework for different applications. *IEEE transactions on pattern analysis and machine intelligence*, 27(4):506–517.

Tseng, C.-Y. and Wang, S.-J. (2014). Shape-from-focus depth reconstruction with a spatial consistency model. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(12):2063–2076.

Verbin, D. and Zickler, T. (2020). Toward a universal model for shape from texture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 422–430.