



Improving Shape Transformations for RGB Cameras Using Photometric Stereo

H. I. Wahhab¹, A. N. Alanssari², Ahmed L. Khalaf³, Ravi Sekhar^{4,*}, Pritesh Shah⁵, Jamal F. Tawfeq⁶

¹ College of Computer Science & Information Technology, University of Kerbala, Kerbala, 56002, Iraq

² Faculty of Basic education, Department of Mathematics, University of Kufa, Najaf, Iraq

³ College of Engineering, Al-Iraqia University, Baghdad, 10054, Iraq

^{4,5} Symbiosis Institute of Technology (SIT) Pune Campus, Symbiosis International (Deemed University) (SIU), Pune, 412115, Maharashtra, India

⁶ Department of Medical Instrumentation Technical Engineering, Medical Technical College, Al-Farahidi University, Baghdad, 10070, Iraq

Emails: haider.wahhab@uokerbala.edu.iq; alaan.azeez@uokufa.edu.iq; ahmed.l.khalaf@aliraqia.edu.iq; ravi.sekhar@sitpune.edu.in; pritesh.shah@sitpune.edu.in; jamaltawfeq55@gmail.com

Abstract

The emergence of low-cost red, green, and blue (RGB) cameras has significantly impacted various computer vision tasks. However, these cameras often produce depth maps with limited object details, noise, and missing information. These limitations can adversely affect the quality of 3D reconstruction and the accuracy of camera trajectory estimation. Additionally, existing depth refinement methods struggle to distinguish shape from complex albedo, leading to visible artifacts in the refined depth maps. In this paper, we address these challenges by proposing two novel methods based on the theory of photometric stereo. The first method, the RGB ratio model, tackles the nonlinearity problem present in previous approaches and provides a closed-form solution. The second method, the robust multi-light model, overcomes the limitations of existing depth refinement methods by accurately estimating shape from imperfect depth data without relying on regularization. Furthermore, we demonstrate the effectiveness of combining these methods with image super-resolution to obtain high-quality, high-resolution depth maps. Through quantitative and qualitative experiments, we validate the robustness and effectiveness of our techniques in improving shape transformations for RGB cameras.

Keywords: RGB cameras; photometric stereo; RGB ratio model; multi-light model; computer vision; RGB-Fusion

1. Introduction

This paper presents novel methods for improving shape transformations for RGB cameras using photometric stereo. While low-cost RGB cameras have become increasingly important in computer vision tasks, the depth maps captured by these cameras often lack intricate object details, contain noise, and have missing information. Existing depth refinement methods struggle to distinguish shape from complex albedo, resulting in visible artifacts in the refined depth maps and affecting the quality of 3D reconstruction. To address these limitations, we propose two innovative methods. The first method, the RGB ratio model, addresses the nonlinearity problem in previous approaches and provides a closed-form solution. It effectively preserves geometric intricacies and surpasses the performance of state-of-the-art models. The second method, the robust multi-light model, retrieves accurate shape from imperfect depth data without relying on regularization. It outperforms other depth refinement methods both quantitatively and qualitatively. Additionally, we combine the proposed methods with image super-resolution, enhancing the resolution of low-resolution images to acquire high-quality, high-resolution depth maps. The

robustness and effectiveness of our techniques are demonstrated through quantitative and qualitative experiments. These novel methods significantly contribute to the field of shape transformations for RGB cameras and overcome the limitations of previous works. In this paper, with the emergence of affordable RGB cameras, many research areas in modern computer vision, computer graphics, and robotics have been significantly boosted, such as 3D modeling [1] and reconstruction [2, 3], human motion capture [4], and visual SLAM [5,6]. While low-cost commercial RGB sensors provide RGB images along with corresponding depth information, the quality of the depth data often proves unsatisfactory. Figure 1 depicts the task of depth refinement. As illustrated in Fig. 1(b), the depth map exhibits substantial noise and a quantization effect, resulting in missing depth values. Additionally, some fine-grained details of the statue are absent in the rough depth representation. Conversely, Fig. 1(c) showcases the depth with significantly enhanced quality following refinement through our proposed method, which effectively preserves geometric intricacies. Fig. 1(d) demonstrates a notable improvement in the forehead regions, wherein the details of wrinkles and eyes are retrieved from the raw depth data. [7] reported that noisy and imperfect depth measurements can compromise the quality of 3D reconstruction, while the estimation of the camera trajectory may experience severe deviations due to accumulated errors stemming from the imprecise depth data. Although certain methods, such as Nikolas Engelhard's [8], have made efforts to recover such details by combining depth data from multiple views, the achieved detail recovery remains limited. The depicted depths are presented as 3D surfaces to enhance visualization. The enhancements in in-depth detail are evident through our RGB-Fusion Like method, as demonstrated in Fig. 1.

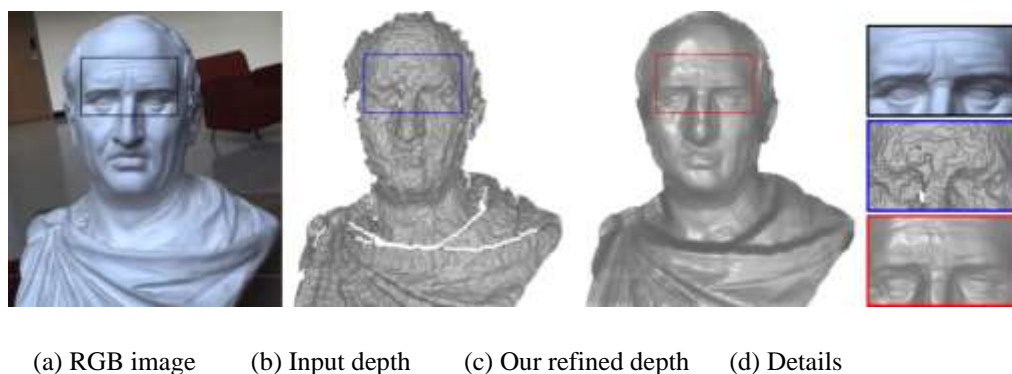


Figure 1: Illustrations for the depth-refinement task. The depths are plotted as 3D surfaces for the sake of better visualization. It is obvious to notice the improvement in the depth details using our RGB-Fusion Like method.

Problem Statement

In this paper, our focus is on exploring the refinement of a single-depth map. To enhance the depth information obtained from consumer depth sensors, we delve into intrinsic object details through the integration of RGB images. This involves investigating factors such as the spatial positions of scene illuminations, their corresponding impacts on object shading, and the object's material reflectance. Shape from shading (SFS) deals with recovering shape from gradual shading variations in images, while photometric stereo (PS) is a computer vision technique that estimates object surface normal by observing the object under diverse lighting conditions. These fundamental methods can estimate intrinsic image properties, including layout patterns and geometric details. Consequently, they have been effectively incorporated into numerous state-of-the-art depth enhancement techniques. Shape refinement approaches based on SFS or PS generally utilize the Lambertian reflectance model. However, refining depth using this model entails a nonlinear inverse problem. Previous methods, such as those discussed in [9, 10], directly applied nonlinear optimization algorithms like Levenberg Marquardt or the alternating direction method of multipliers to solve the problem, resulting in runtime challenges. An alternative method [11] employed a fixed-point scheme by utilizing outcomes from previous iterations to mitigate nonlinearity, yet this approach sometimes led to optimization divergence. Moreover, many shading-based depth refinement methods assume either uniform or constant albedo [4], while others impose piecewise smoothness constraints on albedo to tackle this inverse problem. Considering that depth maps from consumer RGB sensors typically lack satisfactory resolution, we leverage higher-resolution RGB images as references and successfully integrate super-resolution into our depth refinement technique. This allows the final refined depth to achieve the same resolution as the high-resolution RGB images, capturing all intricate details. Notably, this introduces a novel shading-based depth super-resolution approach. In this paper, we introduce the RGB ratio model, where red, green, and blue LED lights are positioned in various directions to construct a lighting model for each channel of the color image captured by RGB sensors. This proposed setup and model

effectively address nonlinearity and offer a closed-form solution. The primary contributions of this paper are:

- Proposing a new RGB ratio model to resolve the nonlinearity and achieve accuracy like the state-of-the-art methods.
- Introducing a robust multi-light method that outperforms other depth refinement methods both quantitatively and qualitatively.
- Presenting the image super-resolution with the proposed method and exhibiting the high-quality and high-resolution depth.

2. Related Works

This section in the paper provides an overview of previous studies and research that are relevant to the current study on improving shape transformations for RGB cameras using photometric stereo. The proposed methods in the current study, the RGB ratio model and the robust multi-light model, aim to overcome the limitations of previous approaches. By highlighting the limitations of previous works and showcasing the advancements of the proposed methods, the authors establish the novelty and significance of their research in the field of shape transformations for RGB cameras.

The widely recognized shape from shading (SFS) problem was initially introduced by Christian Kerl in 2013 [12], and a plethora of literature quickly emerged to advance the field. The concept underlying RGB imaging is that, with knowledge of the light source's position, it becomes possible to estimate an object's shape or surface from just a single grayscale image. This inverse problem is inherently ill-posed. From a mathematical standpoint, the luminance of the RGB camera can be decomposed as follows:

$$I = \rho S \quad (1)$$

where I is an intensity image, ρ is the reflectance (albedo) of the surface, and S is the shading image. An example of such an image decomposition is shown in Fig. 3. SFS methods assume that the observed object follows Lambert's cosine law [11], based on which Eq. 1 can be reformulated to the Lambertian reflectance model means is the property that defines an ideal "matte" or diffusely reflecting surface. The apparent brightness of a Lambertian surface to an observer is the same regardless of the observer's angle of view:

$$I = \rho l > n \quad (2)$$

where we can notice that the shading S is the inner product of the light direction and the surface normal. Thus, the task of SFS is to retrieve the shape (surface normal) from the shading based on the Lambertian reflectance model. Moreover, many state-of-the-art shape or depth refinement methods use an extension of the Lambertian model called spherical harmonics [13] which can represent the illumination more realistically. The first-order SH model can account for 87.5% of real-world light and consequently means as a result. Grandfather had sustained a broken back while working in the mines. Consequently, he spent the rest of his life in a wheelchair. we applied it throughout the whole paper:

$$I = \rho (l > n + \phi) \quad (3)$$

where ϕ is the ambient light parameter. In recent years, there have been many studies focused on depth or shape refinement based on either SFS (use only one single image) or PS (multiple images with different illuminations). We will discuss these two streams respectively in the following sections as shown in Fig. 2.

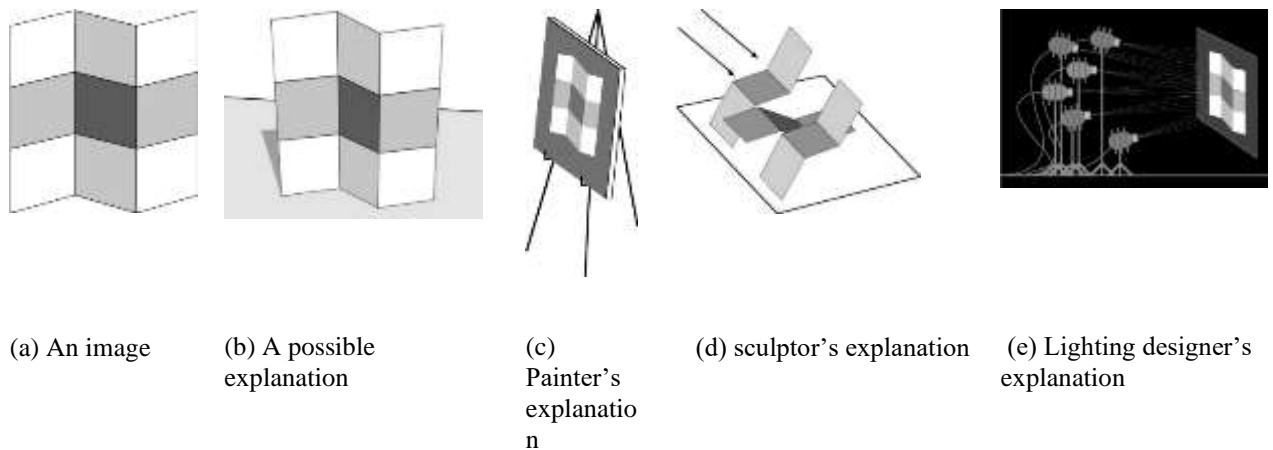


Figure 2: Various explanations for a twice-bent surface. Illustrations for the ambiguity suffered by SFS.



Figure 3: The input RGB and depth image of a vase. The depth map is visualized using colors from blue (near) to yellow (far), where the missing areas and quantization effects can be noticed.

2.1 SFS Method

Because SFS techniques rely on a single image to estimate depth or shape, they are vulnerable to inherent ambiguities that can constrain their accuracy, such as occlusion ambiguity and texture ambiguity, even when the light source and albedo are specified. While a rough depth map may be provided, the illumination and albedo are unknown, requiring the imposition of regularization terms to obtain a valid solution to the inverse problem. Han et al. [14] introduced an approach that integrates a global lighting model, utilizing the provided color and depth (as depicted in Fig. 3), with a spatially varying local lighting model. This approach facilitates more precise and detailed lighting modeling. To enforce the integrality constraint on smooth surface orientations, they applied a penalty on the curl of local neighbors. However, their method assumes uniform albedo. To accommodate objects with multiple albedo values, they had to employ another intrinsic image decomposition algorithm [7] and employ k-means clustering to group albedo into areas with consistent values. Such a framework is both unrealistic and computationally intensive, rendering it unsuitable for numerous real-world applications.

2.2 PS-Based Method

A PS-based method constitutes another category of shading-based depth refinement. By utilizing multiple images captured from diverse illuminations, these approaches can address the ambiguities inherent to SFS methods and exhibit superior performance in separating albedo and surface normals. Haque et al. [15] introduced a technique for shape reconstruction and depth refinement using an IR camera, eliminating the necessity for an RGB camera.

However, akin to various multi-view photometric reconstruction approaches, they assumed uniform albedo, making them unsuitable for objects with multiple albedo values. In [14], the authors employed a standard photometric stereo approach to decompose input images captured under distinct illuminations. They adopted an iterative reweighted method to approximate the Rank 3 radiometric brightness matrix, subsequently factorizing it into corresponding lighting, albedo, and surface normals. While this method can handle multi-objects, it still requires IR images instead of RGB images. In this scenario, at least one additional infrared light source is invariably necessary, whereas in our approach, a low-cost LED light or even the flashlight on a phone suffices.

3. Methodology

In this paper, two novel methods based on photometric stereo are proposed to improve the quality of depth maps obtained from consumer RGB cameras. The first method, known as the RGB Ratio Model, addresses the nonlinearity problem commonly encountered in depth enhancement methods. It aims to accurately estimate the depth and recover intricate details without introducing artifacts. This approach demonstrates the potential for achieving high-resolution 3D reconstruction from affordable RGB cameras. Many computer vision applications, such as 3D object reconstruction or visual SLAM, require depth information from RGB cameras. However, the outcomes of these applications are often constrained by the subpar quality of depth acquisition using consumer RGB cameras. In this paper, to enhance flawed depth maps, we propose two shape refinement methods based on uncalibrated photometric stereo. Like [16], we initially employed red, green, and blue LEDs for active illuminations, allowing us to treat each channel of the acquired color image as an intensity image illuminated from a distinct direction. An additional method we introduce requires just one white LED. Maintaining a fixed angle of view, we manually moved the LED lights while capturing images. Furthermore, the utilization of shading-based methods for depth map super-resolution is a novel approach. We have effectively adapted our methods for depth super-resolution, yielding superior outcomes.

In more precise terms, we first introduce certain preprocessing techniques to fill missing areas, reduce noise, and mitigate quantization effects in the input depth image. Subsequently, we provide a comprehensive description of a state-of-the-art depth refinement method introduced in [16], which we selected as a foundation for implementation. Following this, we present a method based on an RGB ratio model aimed at eliminating nonlinearity, commonly encountered in contemporary depth enhancement methods. Finally, we introduce another method that bypasses the need for regularization terms. This technique demonstrates the capacity to handle objects with intricate albedos and extend into depth super-resolution. Our robust multi-light approach possesses the ability to capture the shape of synthetic objects with highly complex albedo. In contrast, the one-image state-of-the-art depth enhancement methods perform well for many simple colored objects yet struggle to disentangle intricate albedo from the true object shape, resulting in severe artifacts in the final estimated depth.

3.1 Preprocessing

Due to the hardware limitation, there are usually holes with missing values on the depth images. Moreover, the depth data is often noisy, and consequently, denoising and acquiring a relatively smooth surface is usually required. In these subsections, we describe the depth inpainting and denoising algorithms, respectively, used in our preprocessing component.

a. Depth Inpainting

Image inpainting itself is a very mutual area and has been widely applied as a useful tool for many modern computer vision applications, e.g., restoring the damaged parts of ancient paintings, and removing unwanted texts or objects in photography [14]. Since the idea of image inpainting is to automatically replace the lost or undesired parts of an image with the neighboring information by an interpolation operation, it is sufficient to apply it to fill in the missing areas (see Fig. 4).

b. Depth Denoising

The depth images acquired from low-cost RGB cameras usually contain various types of noise as well as quantization effects. As a standard preprocessing technique, image denoising has been applied to our inpainted input depth map. In line with previously proposed depth refinement methods [14], we employ bilateral filtering [7] as our depth preprocessing smoother. A key advantage of the bilateral filter is its ability to reduce noise while preserving edges in the input image. Like a standard Gaussian smoothing filter, which relies on the difference in image values (depth in our case) between center pixels and neighbors, the bilateral filter additionally incorporates spatial differences to establish the weighting function, as depicted in Fig. 4.

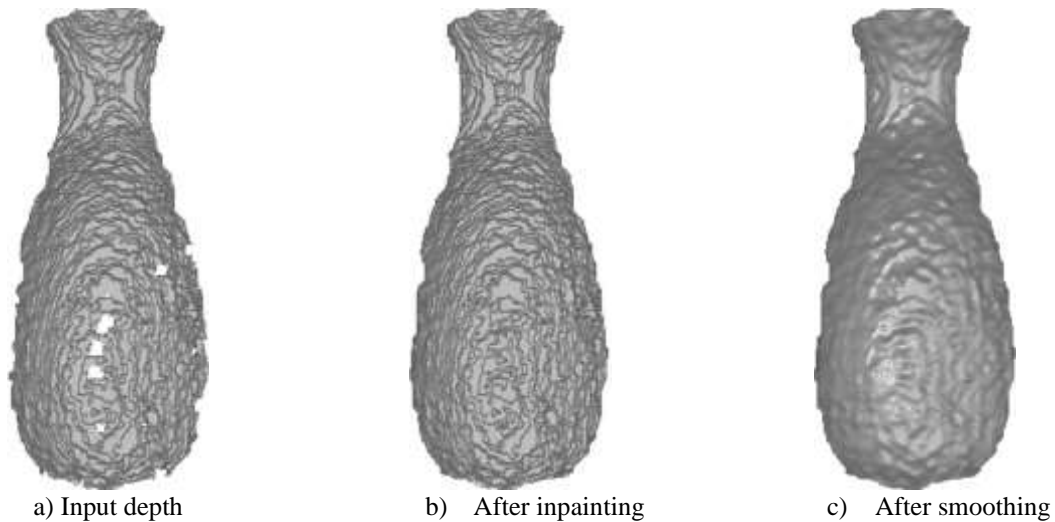


Figure 4: Illustrations for the pre-processing on the depth image of the vase.

3.2 RGB Fusion Method

RGB-Fusion is a cutting-edge depth recovery method proposed by Jaesik Park et al. [16]. This method is effective for natural scene illumination and can significantly expedite depth map enhancement in comparison to other approaches. In this paper, instead of individually estimating pixel-wise ambient light through a separate energy function, we jointly calculate all four first-order spherical harmonics parameters (3 for point-source light direction and 1 for ambient light) using fast least squares. This modification reduces the number of tuning parameters from 8 to 5, while the resultant outcomes exhibit only negligible differences. Throughout the entire estimation process encompassing light, albedo, and depth, we exclusively utilize information confined within the provided mask, thereby further accelerating the algorithm. This rationale is why we refer to our first method as the "RGB-Fusion" method.

3.3 Proposed Method I: RGB Ratio Model

RGB cameras have found wide application in numerous modern computer vision fields, including 3D reconstruction [1, 2], visual odometry, mapping on quadcopters [17], and visual SLAM algorithms [4, 5]. An RGB camera captures a color image usually represented in the RGB color space, along with a depth map, where each pixel reflects the real-world distance between the camera and the corresponding pixel position. Depending on the depth measurement technologies employed, RGB cameras can be classified into passive and active types. A passive RGB camera typically consists of two RGB cameras with a known translation between them. After capturing an image from each camera, matching features in the two images is performed, followed by triangulation to calculate depth (see Fig. 2). Active technologies emit light into the environment, enabling depth image acquisition even in dark indoor settings. These active methods can further be categorized as Time of Flight (ToF) or structured light approaches. Many active stereo cameras utilize infrared (IR) light projection, but they are limited to indoor environments due to the reliance on sources of infrared light like the sun.

The RGB color model operates on an additive color system, creating a broad color space by blending red, green, and blue with varying intensities. The name of the model comes from the first letters of the words "red," "green," and "blue," which represent the additive primary hues [18, 19]. The RGB color model is extensively employed in electronic systems for image sensing, representation, and display, as well as in traditional photography. This color model is based on a theory predating electronic devices, rooted in how humans perceive color. Color components (like phosphors or dyes) and their reactions to different levels of red, green, and blue can vary among manufacturers and even change over time within the same device. Without color management, the same RGB value may not yield the same color on different devices. Common RGB input devices encompass color televisions, video cameras, image scanners, and digital cameras. Output devices for RGB include conventional and flat-panel televisions, computer and smartphone screens, video projectors, multicolor LED displays, and large screens such as Jumbotrons. However, color printers employ a subtractive approach and often use the CMYK color model, thus diverging from RGB devices [20, 21]. We will demonstrate the robustness of the depth estimation from the

proposed multi-light method for objects with intricate albedo. Subsequently, we will showcase the performance evaluation for specular objects.

Limitations

The RGB ratio model method can estimate the albedo and the depth better than our RGB-Fusion Like method in some cases based on the assumption that the non-linearity optimization problem in the RGB-Fusion has been solved. Nevertheless, to avoid undesirable results, the following points must be taken into consideration:

- Three LED lights must be set up far away from each other, where the albedo refinement may fail if the lights are set too close. This can lead to some inconvenience, such as the requirement of a relatively larger space to set up the system than the RGB-Fusion method.
- The active red, green, and blue LEDs are likely to bring extra specularity. For example, if there is a need to refine the depth of a specular object, the specular reflection will be from not only the natural scene illumination but also RGB lights from 3 directions, which will make the refined results even worse.
- Auto white balance (AWB) has a big impact on the refined results. This is because the success of our method highly relies on the difference between 3 channels in a color image. Also, AWB will mix up the information among the channels, where it is necessary to turn it off. This impedes the generalization of our method, because AWB has been set as a default in many modern inexpensive cameras.

3.4 Proposed Method II: Robust Multi-Light Model

The regularization settings that prioritize discontinuous smoothness exert a significant influence on albedo estimation for both our RGBD-like technique and the RGB ratio model. This approach represents the conventional manner of determining albedo in the majority of existing depth-refinement methods. While whites themselves with substantial pattern variations and minimal dominant color can appear striking, complex-colored and intricately patterned objects pose challenges, and not all these methods prove effective for albedo estimation. The success of the final depth refinement outcome is contingent upon the success of whiteness estimation. Furthermore, as the values of regularization parameters often need to differ for distinct objects, configuring these parameters becomes a time-consuming and laborious task, as depicted in Fig. 5.

Therefore, it is reasonable to propose a novel approach capable of eliminating all regularization terms and accurately estimating intricate albedo, thus avoiding overfitting issues and mitigating ambiguities when applying only the shading term. We have included several color images of a stationary object illuminated from various angles, wherein all these images can be collectively utilized to estimate the object's albedo. As a result, unlike depth refinement using a solitary image, ambiguities stemming from the shading term have been resolved, obviating the need for regularization. Figure 6 demonstrates the robustness of our proposed estimation method.



Figure 5: Illustrations for the obtained color images of a vase from various light directions with a white LED light. Even the phone flashlight is sufficient for giving various lighting.

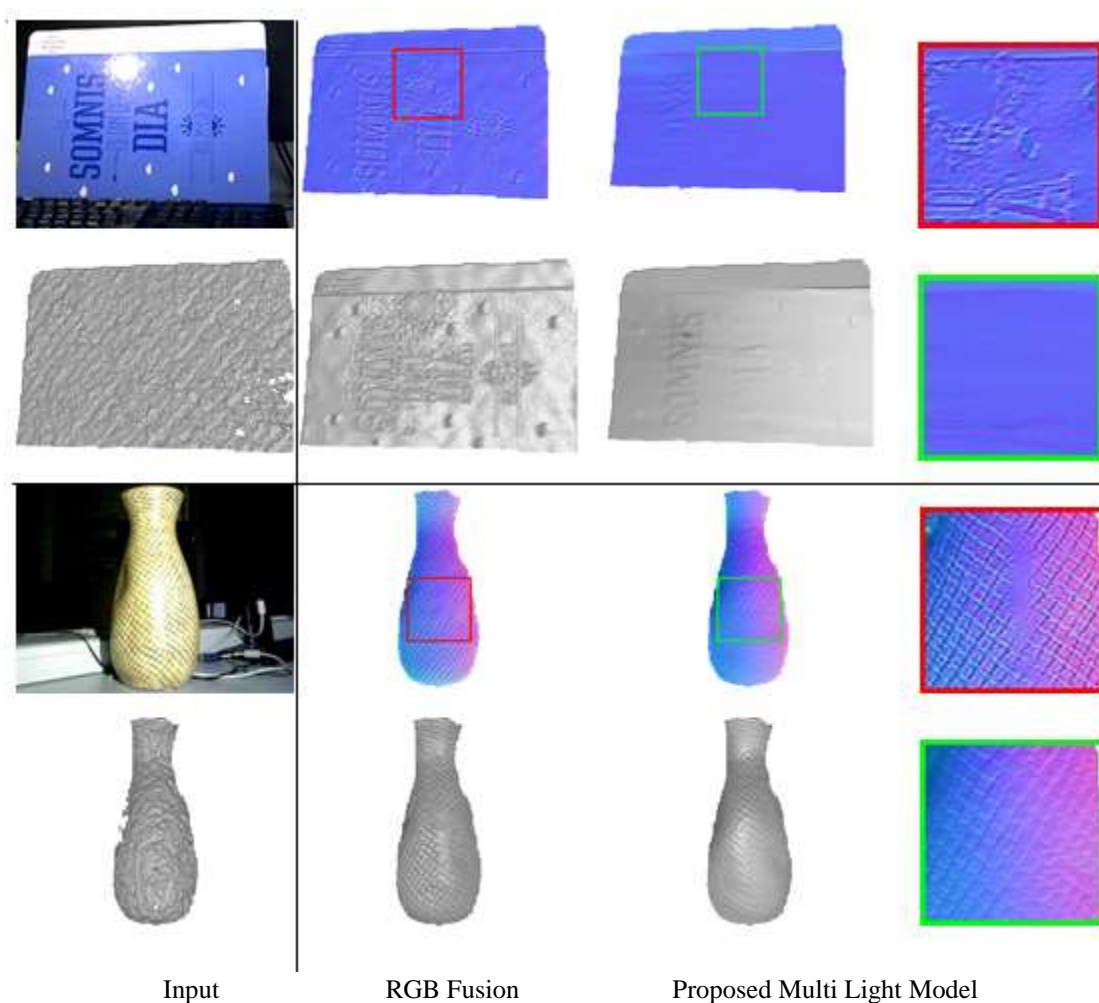


Figure 6: Comparisons between our multi-light method and RGB-Fusion for two specular objects. The RGB images in the first column are among 4 various illuminations. The first and third rows correspond to the surface normal, while the second and fourth are the refined depths. We can notice the RGB-Fusion method has strong artifacts on the refined depth in the specular part, while our method can still correctly acquire all the correct details under the specularity.

Due to our utilization of a single white light instead of controlling lighting with red, green, and blue LEDs, the ratio model is inapplicable in this context. We chose to revert to the standard 1st-order SFS method for constructing input color images. Similar to the aforementioned two methods, our proposed algorithm comprises three key components: light estimation, which provides inputs to rendering algorithms and shaders to ensure the natural appearance of shading, shadows, and reflections under diverse conditions; albedo estimation, which quantifies diffuse reflection of solar radiation relative to total solar radiation, measured on a scale from 0 (indicating a black body that absorbs all incident radiation) to 1 (indicating a body that reflects all incident radiation); and depth enhancement, referring to any technique that improves the visual interpretation of an image. We need to iteratively update all three of these components, unlike the RGB ratio model approach, where prior initial estimations of light and albedo are necessary. In contrast, we can assume the albedo to be 1 everywhere initially, and the illuminations can be set to frontal directions at the outset.

The RGB ratio model addresses the nonlinearity problem often encountered in depth enhancement methods. It utilizes red, green, and blue LEDs as active lights and employs iterative depth map refinement using ratio Lambertian models for each channel pair of the input RGB image. This method provides a closed-form solution and aims to improve the accuracy of depth estimation. The robust multi-light model, on the other hand, utilizes multiple images captured with different illumination angles to collectively estimate depth, albedo, and lighting conditions. This method does not rely on regularization terms and is designed to handle objects with complex albedo. It aims to accurately estimate the depth and recover intricate details without introducing artifacts. The proposed methods offer novel approaches to improve the quality of depth maps obtained from consumer RGB

cameras.

Limitations

In this section, we present the limitations that could occur when dealing with specific criteria, to avoid them.

- The scale factor between the RGB and depth image is around 2 for the RGB camera. Consequently, we can enlarge our map twice larger than the original size, which means that the refined depth resolution will be 1280×960 . Moreover, we assume that the input depth image has been registered well, such that the up-sampled depth is aligned with the large RGB image after simple interpolation.
- After having the estimated light and the color albedo, we can continue refining the depth and hence we need to rearrange the energy function with the depth as the argument.

4. Results

In this section, we present the results of our experiments and provide a detailed analysis of each scenario to demonstrate the effectiveness of our proposed depth refinement methods. The Root Mean Square Error (RMSE) between the ground truth and the estimated depth was adopted to quantify the success of depth clarification. This is in addition to the Mean Angular Error (MAE) between the ground truth and the estimated normal directions. The RMSE reflects the global quality of the refined depth (low frequency), while the MAE assesses the precision of the recovered depth (high frequency). As demonstrated by Table 1, the following observations were made:

- Adding the Laplacian smoothness term in the depth enhancement energy of the RGB-Fusion method makes a huge improvement on the refined results.
- Single-depth image refinement methods (RGB-Fusion and RGB ratio model) have a chance to acquire satisfactory results only when the albedo is simple with several big color patches. However, they will fail and give even worse results than the input depth in terms of RMSE and MAE when the albedos get complicated. Most of the small details on the albedo of the 'Pattern' and 'Complicate Pattern' cannot be acquired by these methods, which produce the wrong depth estimation with artifacts. A possible reason is that their albedo estimations highly rely on the regularization terms which prefer piecewise smoothness, but this does not meet the condition of most real-world objects.
- It can be effortlessly noticed that our robust multi-light method has a strong ability to handle cases with extremely complicated albedo. Instead of using any regularization terms, our method uses only one shading term to estimate the albedo with extra images illuminated from various light directions. Compared to the albedo estimated by other methods, the albedo from our multi-light method could recover most of the details.

Table 1: Quantitative evaluations among 4 methods. RMSE and MAE are in millimeters and degrees respectively. “No smooth” means no Laplacian smoothness term in-depth enhancement.

Method	Simple RGB		Pattern		Complicated	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Input reference	3.3305	16.3096	3.3305	16.3096	3.3305	16.3096
RGB-Fusion [11] (no smooth)	3.3418	18.9115	3.3872	27.0026	3.3411	25.6574
RGB-Fusion [11]	3.1751	17.2197	3.1890	18.4722	3.1708	18.0850
Fusion-Like (not smooth)	3.3475	17.5911	3.3459	23.4808	3.3898	35.2610
RGB ratio model	1.9437	5.0574	2.9116	17.5238	3.1006	21.2286
Robust multi-light model	2.3125	3.8708	1.5794	1.7368	1.8424	2.6815

4.2 Comparison in Terms of Runtime

We compare the runtime for 4 methods, which are listed in Table 2. It is noticeable that our RGB-Fusion method uses fewer parameters (5 against 8) and less runtime (7s against 21s) than RGB-Fusion [11], while the accuracy of our implementation is quite like the original one. We attribute this to two factors: First, our implementation only considers pixels within the mask, whereas the original implementation processes the entire image. Second, rather than estimating the ambient light for each pixel individually, we simply treat it as a single parameter inside the

first-order spherical harmonics so the ambient light can be obtained along with lighting directions. It has been shown in Fig. 7 that when the number of images increases, the runtime for each iteration has a linear ascent, while two errors decrease and reach a platform. This is reasonable because the details on a certain part of the object can be retrieved when there exists light on it. Therefore, the more various lightings we have, the more details we can obtain. If we consider all possibilities, 10 ~ 20 would be the suitable number for illuminations.

Table 2: The comparison of the runtime between RGB-fusion method, our implementation RGB-fusion like method, the proposed RGB ratio model, and the robust multi-light method in synthetic data.

Method	Runtime
RGB-Fusion [11]	21.64
RGB-Fusion Like	7.75
Proposed I: RGB Ratio	49.33
Proposed II: Multi-Light	52.82

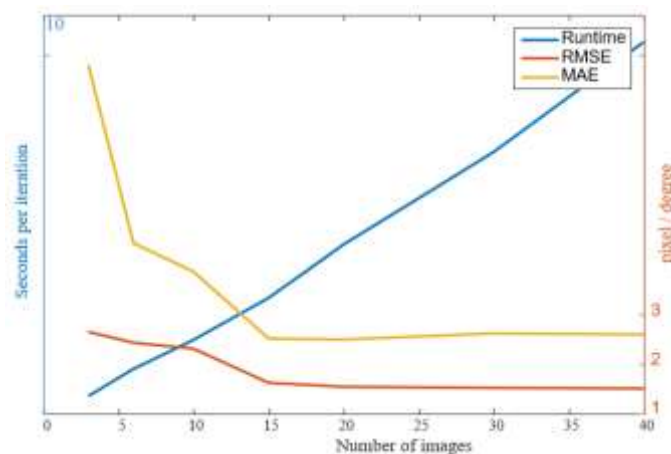


Figure 7: Illustrations for the runtime, RMSE, and MAE in various illuminations for the proposed robust multi-light method. 10 ~ 20 is the suitable number for different lighting.

5. Discussion

To explain why our method works effectively with specular objects, we initially assume the presence of 4 input color images. We are aware that the specularity in each image varies due to the shifting active LED light, indicating that the specularity visible in one image is highly likely to not be specular in the other 3 images. Simultaneously, our algorithm utilizes all 4 images for albedo and depth refinement, rather than relying solely on the image with specularity. Consequently, the other three images in the least squares computation aid in rectifying the impact of specularity in that region. Parallel to the analysis conducted in [18], we conduct a comparison using the synthetic dataset of vibrant patterns from the previous section. As evident in Fig. 8, the depth enhancement achieved with our method closely aligns with the ground truth. The outcome obtained from the LDR-PS method appears to resemble the ground truth from the frontal perspective (though it seems darker due to the LDR-PS method's occasional depth estimation inaccuracies). Nevertheless, when the reconstructed map is rotated sideways, it becomes evident that the discontinuity between the head and body has been excessively smoothed, and the pedestal has suffered distortion. This phenomenon is referred to as the "generalized bas-relief ambiguity".

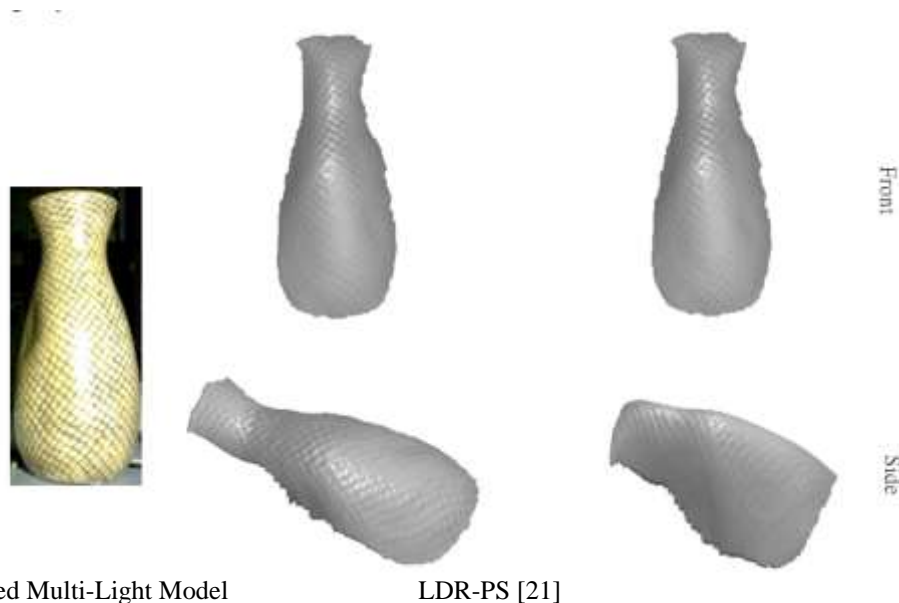


Figure 8: Comparison of the recovered depth between our multi-light method and the un-calibrated photometric stereo method LDR-PS on the real-world vase. Our method is able to recover the real shape, while the LDR-PS could not acquire the correct shape because of the GBR ambiguity.

6. Conclusion

The proposed methods for depth refinement based on photometric stereo offer promising solutions to enhance the quality and accuracy of depth maps obtained from consumer RGB cameras. The experimental evaluation confirms the robustness and effectiveness of the proposed methods in enhancing the quality and accuracy of depth maps. These methods have the potential to significantly benefit various computer vision applications, including 3D modeling, reconstruction, and visual SLAM, by providing high-quality depth information from consumer RGB cameras. It is important to note that the specific values obtained in the experimental evaluation may vary depending on the dataset and implementation details. However, the consistent trend of superior performance demonstrated by the proposed methods compared to existing approaches highlights their potential for practical applications in the field of computer vision.

7. Future Work

There are various possible directions for future work on our depth or shape refinement research:

- It has been shown in [22] that the noise level of depth acquisition from low-cost depth sensors grows quadratically concerning the increasing distance. Hence, the refined depth should be theoretically more accurate if every depth pixel is weighted according to the corresponding measurement noise.
- As aforementioned, the depths for complicated objects refined by single image-based methods contain artifacts, since the designed constraints on the albedo are not practical so the estimated depth may be affected by the inaccurate albedo.
- The existing 3D object reconstruction/modeling methods are subject to low-resolution and bad-quality depths. It is promising to integrate them with our shading-based depth super-resolution method, which will potentially improve the reconstruction accuracy.
- As with other methods, one prerequisite of our methods is that the depth image needs to be registered to the RGB image. It is possible to integrate the depth image registration within the refinement framework implicitly such that we can directly acquire depth and color images from the RGB cameras without any other third-party software [23].

Acknowledgment: The authors would like to thank The Ministry of Higher Education and Scientific Research, Iraq, and Symbiosis Institute of Technology (SIT) Pune Campus, Symbiosis International (Deemed University) (SIU) in India.

Funding Statement: The authors received no specific funding for this study.

Availability of Data and Materials: The data used in this paper can be requested from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] K. A Tychola, Tsimperidis and G. A. Papakostas, "On 3D reconstruction using RGB-D cameras," *Digital*, vol.2, no.3, pp. 401-421, 2022.
- [2] M. Zollhöfer, P. Stotko, A. Görlitz, C. Theobalt, M. Nießner *et al.*, "State of the art on 3D reconstruction with RGB-D cameras," *Computer Graphics Forum*, vol. 37, no. 2, pp. 625-652, 2018.
- [3] C. Delasse, H. Lafkiri, R. Hajji, I. Rached and T. Landes, "Indoor 3D reconstruction of buildings via azure Kinect RGB-D camera," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 2–22, 2022.
- [4] A. Boonsongsrikul and J. Eamsaard, "Real-time human motion tracking by Tello EDU drone," *Sensors*, vol. 23, no. 897, pp.21-25, 2023.
- [5] A. Tourani, H. Bavle, J. L. Sanchez-Lopez and H. Voos, "Visual SLAM: what are the current trends and what to expect?," *Sensors*, vol. 22, no. 23, pp. 9297, 2022.
- [6] I. A. Kazerouni, L. Fitzgerald, G. Dooly and D. Toal, "A survey of state-of-the-art on visual SLAM," *Expert Systems with Applications*, vol. 205, pp. 117734, 2022.
- [7] M. Robert, "Out-of-core bundle adjustment for 3d workpiece reconstruction," Master's thesis, Technische Universität München, Germany, September 2013.
- [8] N. Engelhard, F. Endres, J. Hess, J. Sturm and W. Burgard, "Real-time 3D visual SLAM with a hand-held RGB-D camera," in *Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum*, Vasteras, Sweden, Vol. 180, pp. 1-15, April 2011.
- [9] C. Wu, M. Zollhöfer, M. Nießner, M. Stamminger, S. Izadi *et al.*, "Real-time shading-based refinement for consumer depth cameras," *ACM Transactions on Graphics*, vol. 33, no. 200, pp. 1-10, 2014.
- [10] R. Or-El, R. Hershkovitz, A. Wetzler, G. Rosman, A. M. Bruckstein *et al.*, "Real-time depth refinement for specular objects," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, Nevada, USA, pp. 4378-4386, 27-30 June 2016.
- [11] R. Or-El, G. Rosman, A. Wetzler, R. Kimmel and A. M. Bruckstein, "RGBD-fusion: real-time high precision depth recovery," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, vol. 3, no. 4, pp. 59–59, 7-12 June 2015.
- [12] C. Kerl, J. Sturm and D. Cremers, "Dense visual SLAM for RGB-D cameras," in *Proc. of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Tokyo, Japan, vol. 11, no. 4, pp. 48–59, 3-7 Nov. 2013.
- [13] J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.37, pp. 1670-1687, 2015.
- [14] Y. Han, J. Y. Lee and I. S. Kweon, "High-quality shape from a single RGB-D image under uncalibrated natural illumination," in *Proc. of the IEEE International Conference on Computer Vision*, Sydney, Australia, pp. 1617-1624, 1-8 Dec. 2013.
- [15] M. Haque, A. Chatterjee and V.M. Govindu, "High-quality photometric reconstruction using a depth camera," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, USA, pp. 2275-2282, 23-28 June 2014.
- [16] J. Park, S. N. Sinha, Y. Matsushita, Y. W. Tai and I. S. Kweon, "Multiview photometric stereo using planar mesh parameterization," in *Proc. of the IEEE International Conference on Computer Vision*, Sydney, Australia, vol. 3, no. 4, pp. 110–119, 1-8 Dec. 2013.
- [17] C. Kerl "Odometry from rgb-d cameras for autonomous quadcopters," Ph.D. dissertation, Fakultät für Informatik der Technischen Universität München, Germany, 2012.

- [18] A. Trémeau, S. Bianco and R. Schettini, "RGB-D- λ : 3D Multispectral Acquisition with Stereo RGB Cameras," in *Proc. of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 77, no. 45, pp. 9–20, 2016.
- [19] M. H. Brill, "How the CIE 1931 color-matching functions were derived from Wright-Guild data," *Color Research & Application*, vol. 23, no. 4, pp. 259–259, Aug. 1998.
- [20] O. Ikeda, "Shape reconstruction for color objects using segmentation and photometric stereo," in *Proc. of the International Conference on Image Processing*, vol. 33, no. 44, pp. 12–25, 2004.
- [21] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, *et al.*, "Visual odometry and mapping for autonomous flight using an RGB-D camera," in *Robotics Research: The 15th International Symposium ISRR*, vol. 1, no. 1, pp. 235-252, 2017.
- [22] Y. Quéau, J. Mérou, J. D. Durou and D. Cremers, "Dense multi-view 3d-reconstruction without dense correspondences," arXiv preprint arXiv:1704.00337, vol. 5, no. 6, pp. 110–130, 2017.
- [23] S. Alsubai, M. Hamdi, S. Abdel-Khalek, A. Alqahtani, A. Binbusayyis, and R. F. Mansour, "Bald eagle search optimization with deep transfer learning enabled age-invariant face recognition model," *Image and Vision Computing*, vol. 126, p. 104545, Oct. 2022, doi: <https://doi.org/10.1016/j.imavis.2022.104545>.