Article

SREVAS: Shading Based Surface Refinement under Varying Albedo and Specularity

Zhihua Hu [1,*], Yaolin Hou [1], Pengjie Tao [1] and Jie Shan [2]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; houyaolin@whu.edu.cn (Y.H.); pjtao@whu.edu.cn (P.T.)
[2] Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA; jshan@purdue.edu
* Correspondence: huzhihua@whu.edu.cn

check for updates

Abstract: Shape-from-shading and stereo vision are two complementary methods to reconstruct 3D surface from images. Stereo vision can reconstruct the overall shape well but is vulnerable in texture-less and non-Lambertian areas where shape-from-shading can recover fine details. This paper presents a novel, generic shading based method to refine the surface generated by multi-view stereo. Different from most of the shading based surface refinement methods, the new development does not assume the ideal Lambertian reflectance, known illumination, or uniform surface albedo. Instead, specular reflectance is taken into account while the illumination can be arbitrary and the albedo can be non-uniform. Surface refinement is achieved by solving an objective function where the imaging process is modeled with spherical harmonics illumination and specular reflectance. Our experiments are carried out using images of indoor scenes with obvious specular reflection and of outdoor scenes with a mixture of Lambertian and specular reflections. Comparing to surfaces created by current multi-view stereo and shape-from-shading methods, the developed method can recover more fine details with lower omission rates (6.11% vs. 24.25%) in the scenes evaluated. The benefit is more apparent when the images are taken with low-cost, off-the-shelf cameras. It is therefore recommended that a general shading model consisting of varying albedo and specularity shall be used in routine surface reconstruction practice.

Keywords: shape-from-shading; multi-view stereo; surface refinement; specularity; albedo

1. Introduction

Reconstruction of 3D surface from multi-view images is of great interest in recent decades. The combination of structure from motion (SfM) [1–3] and multi-view stereo (MVS) [4–7] can reconstruct the 3D shape of a scene with multiple images. SfM helps to estimate the camera parameters including interior orientation parameters (focal length, principal point position and lens distortion parameters) and exterior orientation parameters (camera locations and orientations), while MVS attempts to reconstruct the 3D shape by searching corresponding pixels or other features from images. According to [8], the major challenges of the current MVS algorithms are texture-poor objects, thin structures, and non-Lambertian surfaces. Besides, it also makes MVS reconstruction harder when the images are captured with varying illumination conditions. As a result, reliable and accurate correspondences are difficult to establish and fine details cannot be well restored.

Unlike MVS, shape-from-shading (SfS) [9] can recover the detailed shape from a single image under proper assumptions, for example, known illumination, constant surface albedo and ideal reflectance model. Through modeling the imaging process as an interaction of illumination, surface albedo and surface normal, SfS can recover the detailed surface shape through determining the surface normal (vector). Since the effects of illumination, surface albedo and surface normal are multiplied,

the illumination and the surface albedo are often considered known or constant to make the model solvable. As a variation of the SfS method, photometric stereo, makes use of images taken at a fixed location but under varying illumination conditions [10,11].

Considering the above reasons, SfS and MVS are complementary to each other. Several methods using shading to refine the surface generated by MVS have been proposed [12–16]. Generally, MVS is used to create the initial surface and then shading is applied to refine the initial surface. Besides keeping the illumination and surface albedo constant [17,18], the surface reflection model is often assumed to be Lambertian, which is an ideal diffuse reflection model that reflects equally in all directions. However, in reality, a lot of objects violate the Lambertian assumption and the imaging conditions might be very different such that the SfS methods mentioned above would underperform. There are several methods considering the specular reflectance in recent years [19–22]. Nehab et al. [19] considered the dense reconstruction of specular objects with controllable light sources. Or-El et al. [22] proposed an early depth refinement framework that explicitly accounts for specular reflectance when refining the depth of specular objects with an infrared image. Liu et al. [20] considered the specular reflectance when using SfS and improved the quality of 3D reconstruction of dynamic objects captured by a single camera.

Based on the above observation, we intend to achieve surface refinement under varying albedo and specularity (SREVAS). To this end, we explicitly model the specular reflectance by adding a specular component to the Lambertian reflectance model. The method will be implemented to refine surfaces with diverse multi-view images, including Internet images. SREVAS allows using images of a scene with non-uniform albedo under arbitrary illumination, which means an image may have one or more light sources and the illumination conditions of different images can be different. Furthermore, surfaces with considerable specular reflectance can be recovered due to the use of a comprehensive and realistic reflectance model. Based on the physics and geometry of the imaging process, we introduce an objective function that considers the surface normal, surface albedo and illumination. The method is tested with four benchmark datasets: the DTU Robot Image dataset [23] with multiple light sources (indoor), a synthetic dataset of Joyful Yell [24] with only pure Lambertian reflectance, a multi-view stereo dataset with ground truth [25] (outdoor), and an Internet photo dataset [26] with very different illumination conditions (outdoor). Experiments show that the proposed method can significantly recover more surface details in all these cases than the recently reported shading based surface refinement and reconstruction methods, namely, MVIR [24] and SMVS [27].

The rest of the paper is organized as follows: Section 2 briefly reviews the related work. Section 3 formulates the proposed SREVAS method, while Section 4 presents the experiments on all the datasets. Section 5 concludes our work.

2. Related Works

Considering the underlying surface models, MVS algorithms can be roughly divided into four types [5]: voxel-based [28,29], deformable polygonal mesh-based [30,31], depth map-based [32–34], and patch-based [35–37] methods. The objective of MVS is to find the relevant pixels of the same object in multiple images and to reconstruct the surface. As stated above, high-frequency surface details may not be well recovered by MVS algorithms in texture-less or non-Lambertian surfaces since the image similarity is difficult to determine in those areas without using a prior. The use of particle swarm optimization can achieve better accuracy and robustness in texture-poor or specular surfaces compared to other MVS algorithms [38]. However, the detailed surfaces are still hard to be well recovered with MVS, especially when the images are captured under very different illumination conditions.

In contrast to MVS, SfS recovers the surface normal through modeling the imaging process with one image or multiple images captured in a fixed position under varying illumination conditions. By recovering the surface based on the (surface) normal field, SfS can achieve better surface details since they are intrinsically embedded in the surface normal. However, SfS is an ill-posed problem. As such, assumptions about the illumination conditions and surface albedo are usually imposed to make the problem solvable. For example, uniform albedo and known illumination were

assumed [39,40]. In recent years, the requirements for illumination conditions have relaxed, but the surface albedo still needs to be uniform [41,42]. The photometric stereo algorithm is another way to relax the restriction in traditional SfS by capturing multiple images at a fixed position under different lighting conditions [10]. Due to recent efforts [43–46], photometric stereo methods can handle un-calibrated natural lighting and non-Lambertian reflectance. For example, the introduction of spherical harmonics [47,48] allows un-calibrated natural lighting. The use of non-Lambertian reflectance models, for example, microfacet-based reflectance model can help photometric stereo to deal with highly specular surfaces [49].

MVS requires a well-textured surface whereas SfS generally can deal with texture-less surfaces better. It is therefore of great interest to take advantage of these two complementary methods to best reconstruct the surface. Wu et al. refine the initial MVS surface based on shading under un-calibrated illuminations represented by spherical harmonics [47]. To achieve the refinement, the albedo is assumed to be constant while the illumination is assumed to be fixed and distant in their method. There are researchers trying to decompose the reflectance from shading [50] and reconstruct the surfaces of texture-less objects with the combination of photo-consistency and shading [51,52]. However, the reflectance model is assumed to be Lambertian. There are also some methods based on the photometric stereo [44,53–55]. Nehab et al. [56] proposed a method that can effectively recover the detailed surface with the normal determined by the photometric stereo. However, the images used in photometric stereo need to be captured in a fixed position under varying illumination conditions, which is hard to achieve, especially under natural illuminations. As another effort to relax the above limitation of the photometric stereo, instead of using the original images, Shi et al. [57] used the images created from a fixed position under varying illumination conditions by 3D warping the depth map generated from SfM and MVS. With the development of RGB-D sensors, SfS is also used to refine the depth map [58–63]. To reconstruct the surface in high quality, the visual hull is used to constrain partial vertices [64]. Although structured light can be used to reconstruct detailed surfaces well with proper consideration of surface normal [65], it needs specific equipment, which limits its applications. Kim et al. [24] proposed a method that refines the initial surface from MVS under arbitrary illumination and albedo through solving an imaging model represented by the spherical harmonics. However, it can only recover the surface detail under the Lambertian assumption. To better recover the surface, there is a need to explicitly model both specular and diffuse reflectance especially when the materials of the surface exhibit a mixture of both. In recent years, several methods that directly add shading to the image matching procedure are proposed [27,51,66]. Similarly to Kim et al. [24], they all assumed the Lambertian reflectance and are prone to fail when specular reflectance exists.

To sum up, many methods that use SfS to refine the initial surface from MVS or RGBD images have been proposed [16,17,53,60]. However, most of the previous methods can only refine the surface under limited conditions such as uniform albedo [16], known illumination [67], constant illumination [18] or Lambertian reflectance [24]. Different from the previous methods, the proposed method can recover more detailed surfaces under specularity and varying albedo by extending the existing imaging models.

3. The SREVAS Method

As shown in Figure 1, with the multiple input images, the camera parameters are estimated using SfM and an initial surface is reconstructed with MVS firstly. Then the initial surface is refined with the proposed SREVAS method. In our experiments, we use the VisualSFM [68] to estimate the camera parameters and CMPMVS [6] to recover the initial surface as a mesh model. To assure enough point density for texture-less areas, the initial surface is densified by recursively subdividing the triangles in the mesh until a preset maximum size for every triangle is reached. Through modeling the imaging process, the rendered image intensity can be calculated by the illumination, the albedo and normal of the surface. Since the wrong shape will increase the inconsistencies between the observed and rendered image intensities, the surface can be refined by solving an objective function of data terms and regularization terms about the illumination, the albedo and normal of the surface. This section describes

the proposed SREVAS method in detail including the imaging model, the data term, the geometry term, the diffuse reflectance smoothness term and the specular reflectance smoothness term.
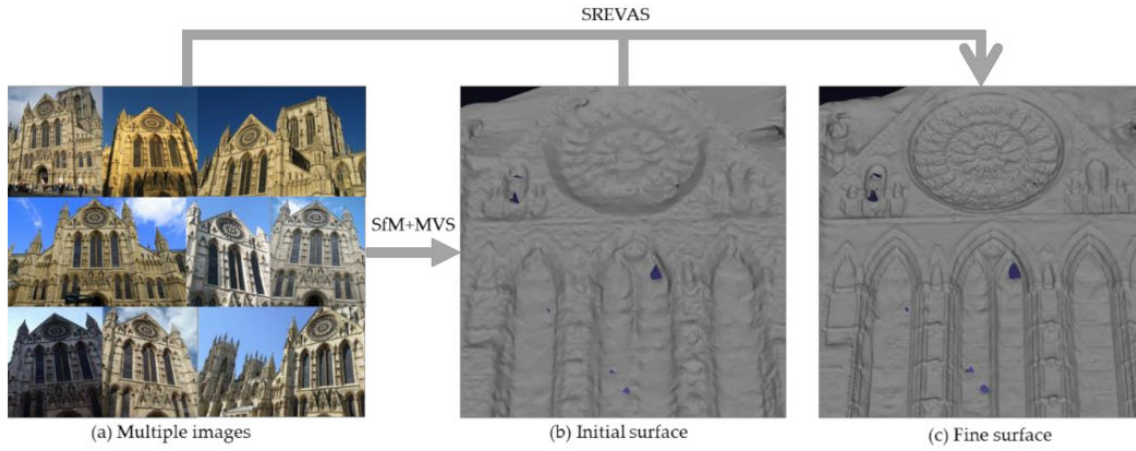

Figure 1. An overview of the proposed SREVAS. (**a**) The input multiple images, (**b**) the initial surface reconstructed through SfM and MVS, (**c**) the fine surface refined with the proposed SREVAS.

When modeling the imaging process, the reflectance model has to be considered. Most of the previous works [18,24,69] assume that the reflectance is perfectly diffuse, in other words, the Lambertian assumption. However, this is often violated. Instead, we assume that the reflectance model is a mixture of diffusion and specularity. We add a specular component to the Lambertian reflectance model to consider the properties of non-Lambertian surfaces. For the Lambertian part of the reflectance model, we follow Basri and Jacobs' work [47] and approximate the illumination with second-order spherical harmonic basis functions. This is particularly suitable for representing complex illumination and has been commonly used [18,24,70]. Based on the consideration described above, the imaging process is modeled as:

$$I_i = R_i \sum_{n=1}^{N} L_n h_n + S_i \tag{1}$$

$$
\begin{aligned}
&h_1 = \tfrac{1}{\sqrt{4\pi}},\ h_2 = \sqrt{\tfrac{3}{4\pi}} n_z,\ h_3 = \sqrt{\tfrac{3}{4\pi}} n_x,\ h_4 = \tfrac{1}{\sqrt{4\pi}} n_y,\ h_5 = \tfrac{1}{2}\sqrt{\tfrac{5}{4\pi}}\left(3n_z^2 - 1\right), \\
&h_6 = 3\sqrt{\tfrac{5}{12\pi}} n_x n_z,\ h_7 = 3\sqrt{\tfrac{5}{12\pi}} n_y n_z,\ h_8 = \tfrac{3}{2}\sqrt{\tfrac{5}{12\pi}}\left(n_x^2 - n_y^2\right),\ h_9 = 3\sqrt{\tfrac{5}{12\pi}} n_x n_y
\end{aligned}
\tag{2}
$$

where I_i is the corresponding pixel value of the i-th vertex, R_i is the per-vertex albedo (also often noted as ρ_i), (n_x, n_y, n_z) is the per-vertex unit normal (vector) at the i-th vertex, h_1–h_9 are the spherical harmonic bases, L_1–L_9 are the per-image coefficients of the spherical harmonic bases or simply lighting coefficients and N is the number of the lighting coefficients. In our experiments N is 9 since the second-order spherical harmonic bases are used and S_i is the specular component which can vary for different vertices and images. In order to allow varying illumination conditions, the lighting coefficients are image variable and the specular components are pixel variable. In the meantime, allowing different surface locations to have different albedos makes the proposed method more general for various complex scenes.

Based on the imaging model above, we build our objective function with the surface albedo, surface normal, lighting coefficients and specular component.

$$\underset{G,R,L,S}{argmin}\{E_{data}(G,R,L,S) + \alpha^2 E_{gsm}(G) + \beta^2 E_{rsm}(R) + \gamma^2 E_{ssm}(S)\} \tag{3}$$

where G is the geometry (position) displacement of the vertex along its normal; R is the albedo of the vertex, varying band to band, like red, green and blue in our experiments; L is the lighting coefficients in Equation (1) and S is the specular component. α, β and γ are used to balance the data term E_{data},

geometry smoothness term E_{gsm}, diffuse reflectance smoothness term E_{rsm} and specular reflectance smoothness term E_{ssm}. The above objective function is essential to determine the best set of G, R, L and S under the constraints of rendering difference, geometry smoothness, diffuse reflectance smoothness and specular reflectance smoothness.

The data term E_{data} is measured by the difference between the observed and rendered pixel values (intensities).

$$E_{data}(G, R, L, S) = \sum_{i=0}^{m} \sum_{c \in V_i} \frac{\|I_o^{i,c}(O_i + G_i) - I_r^{i,c}(O_i + G_i, R, L, S)\|^2}{N_{V_i}} \tag{4}$$

where m is the number of vertices, V_i is the visible camera set for the i-th vertex, $I_o^{i,c}$ is the observed pixel value of i-th vertex in the image c, O_i is the initial position of the vertex while $I_r^{i,c}$ is the rendered pixel value and N_{V_i} is the number of images in V_i. The visibility of every vertex is computed with ray-triangle intersections between the ray from a camera to the vertex and all the triangles in the mesh. If the ray from a camera to the vertex is not occluded by any other triangle in the mesh, the vertex is regarded as visible in the camera. $I_o^{i,c}$ will be re-computed with the change of vertex displacement G_i. $I_r^{i,c}$ is defined in Equation (1) with the surface albedo, surface normal, illumination conditions and specular component. The normal (n_x, n_y, n_z) of a vertex is computed using the vertex position O_i and its displacement G_i by averaging the normal of adjacent faces at the vertex, where the spherical harmonic basis h_1 to h_9 will be recomputed as well.

The geometry term E_{gsm} encourages the surface to be smooth. To this end, we calculate the weighted mean distance between the vertex and its neighbor vertices.

$$E_{gsm}(G) = \sum_{i=0}^{m} \sum_{j \in A_i} w_{i,j}^{gsm} \frac{\|(O_i + G_i) - (O_{i,j} + G_{i,j})\|^2}{N_{A_i} \times l_i} \tag{5}$$

where A_i is the set of adjacent vertices of i-th vertex, $(O_i + G_i)$ are the new coordinates of i-th vertex while $(O_{i,j} + G_{i,j})$ is the new coordinates of j-th adjacent vertex at i-th vertex, N_{A_i} is the number of vertices in the set A_i and l_i is the average edge length between adjacent vertices and its centroid. The bilateral filter [71] weight computed from the pixel value difference and vertex coordinate difference is used to compute $w_{i,j}^{gsm}$.

$$b_{i,j} = exp\left(-\frac{\|(O_i + G_i) - (O_{i,j} + G_{i,j})\|^2}{2 \times \delta_i^d} - \frac{\|\varphi_i - \varphi_j\|^2}{2 \times \delta_i^w}\right) \tag{6}$$

$$w_{i,j}^{gsm} = \frac{b_{i,j}}{\sum_{j \in A_i} b_{i,j}} \tag{7}$$

where $\varphi_i = \sum_{j \in A_i} \sum_{c \in V_i} \frac{\|I_o^{i,c} - I_o^{j,c}\|^2}{N_{A_i} \times N_{V_i}}$, $\delta_i^d = \sqrt{\sum_{j \in A_i} \frac{\|(O_i + G_i) - (O_{i,j} + G_{i,j})\|^2}{N_{A_i}}}$, $\delta_i^w = \sqrt{\sum_{j \in A_i} \frac{\|\varphi_i - \varphi_j\|^2}{N_{A_i}}}$. This smoothness term with the bilateral filter weight is set to encourage the surface to be smooth while preserving sharp edges in images. Different from the usage of the bilateral filter in image filtering, the area of the filter kernel is defined with the neighbor of the vertex instead of a regular window in image space.

Since the albedo of the surface is allowed to be varying in our objective function, we set a diffuse reflectance smoothness term to better separate it from the lighting coefficients. To decompose the ambiguity, in other words, separating the albedo from shading, the diffuse reflectance smoothness term E_{rsm} is calculated based on the assumption often used in intrinsic image decomposition [72] that vertices having similar albedo should have similar color values in each input image.

$$E_{rsm}(R) = \sum_{i=0}^{n} \sum_{j \in A_i} w_{i,j}^{rsm} \frac{\|R_i - R_j\|^2}{N_{A_i}} \tag{8}$$

where $w_{i,j}^{rsm} = exp\left(-k\|\bar{I}_{i,c}^o - \bar{I}_{j,c}^o\|\right)$, $\bar{I}_{i,c}^o$ is the mean color value in all visible images of the *i*-th vertex and k is a constant value.

Inspired by Liu et al. [20], the specular reflectance smoothness term E_{ssm} is set to prevent the rendered value from being considered only as specular and to encourage the specular component to be spatially smooth.

$$E_{ssm}(S) = \sum_{i=0}^{m} \left(w^{s1}\|S_i\|^2 + \sum_{j \in A_i} w^{s2}\|S_i - S_j\|^2 \right) \tag{9}$$

where w^{s1} and w^{s2} are constant values. Additionally, in order to regularize the illumination scale ambiguity, we select a dominant camera that has the largest view frustum and constrains the squared sum of its lighting coefficients to be unit similar to [24].

In the objective function, four types of variables: the lighting coefficients, the surface albedo, the specular components and the vertex displacements are meant to model the imaging process. Through optimizing the objective function described above by the Levenberg–Marquardt implementation from the Ceres Solver [73], the best sets of the four types of variables are determined and the surface can be refined since the positions of vertices will be updated by their displacements. The lighting coefficients are the same for one image while the albedo and the vertex displacement of one surface point are the same in different images. For the specular components, they can be different for different surface points and images. In addition, the constraints on the rendering difference, the geometry smoothness, diffuse reflectance smoothness and specular reflectance smoothness make the objective function solvable and robust.

Similarly to many shading based surface refinement methods [16,24,64], we achieved surface refinement through optimization of an objective function of data terms and regularization terms. The data term is measured with the rendering difference while the geometry smoothness and reflectance smoothness are added as constraints to keep the surface smooth while preserving sharp edges in images. The main difference is that the proposed method considers the specular, a common phenomenon in real-world situations. Due to the introduction of the specular component into the proposed method, a new specular reflectance smoothness constraint is designed to robustly solve the objective function. In [24], the geometry smoothness constraint is an image intensity weighted local surface curvature, while we use the edge preserving bilateral filter kernel as the weight.

4. Experiments and Discussion

Two groups of experiments were designed. The first one was meant to evaluate the effectiveness of the specular component and its solution robustness. For this purpose, we excluded all the specular related terms in the proposed method, in other words, a weak SREVAS without specularity, or named SREVA, to refine the initial surface. This was tested with the DTU Robot Image dataset [23] and the synthetic dataset of Joyful Yell [24]. The DTU dataset was collected in the laboratory (indoor) with several controllable lighting sources and has obvious specular reflections in some images. As such, it is suitable to evaluate the effectiveness of the specular component. Similarly, the experiment on the dataset of Joyful Yell was designed to evaluate the robustness of SREVAS since the dataset is synthetic (computer-generated) under perfect Lambertian without specular reflection.

The second group of tests was meant to understand the performance of the SREVAS method compared with the initial surface reconstruction method CMPMVS, and two representative shading based surface refinement and reconstruction methods, MVIR [24] and SMVS [27]. MVIR can recover detailed surfaces with the SfS technique under arbitrary illumination and albedo, while SMVS combines stereo and SfS in a single optimization scheme. As for the experiment data, we use the Herz-Jesu-P8 [25] and an Internet dataset [26]. All the images of the Herz-Jesu-P8 and Internet datasets are captured under real-world (outdoor) conditions different from the datasets in the first group of experiments. The Herz-Jesu-P8 dataset has calibrated camera parameters and a ground truth model. The images were captured under nearly the same illumination condition. In contrast, there are no calibrated camera parameters and ground truth models for the Internet dataset, where the images were captured under very different illumination conditions.

The DTU dataset, synthetic dataset and Herz-Jesu-P8 dataset have camera parameters. CMPMVS can, therefore, be directly applied to recover the initial surface. For the Internet dataset, SfM is used to estimate the camera parameters first. After the initial surface is recovered with CMPMVS, MVIR, our SREVAS and SREVA are applied to refine it, respectively. For MVIR, an executable program provided by Kim et al. [24] is used. Since the optimal parameters for the synthetic and Internet datasets in MVIR were provided in Kim et al.'s work [24], MVIR is only applied to the synthetic and Internet datasets in our study. For SMVS [27], the source code is provided and it is applied to the Herz-Jesu-P8 and Internet datasets with the default parameters.

Throughout the experiments, we use the same values for all optimization parameters in SREVAS: $\alpha = 0.1$, $\beta = 0.3$, $\gamma = 0.5$, $k = 10$, $w^{s1} = 0.01$ and $w^{s2} = 0.3$. When experimenting with SREVA, α is 0.15 and β is 0.4.

4.1. Specular Component in SREVAS

Firstly, the Buddha model in the DTU dataset is used to evaluate the effectiveness of the specular component in the proposed method. According to [23], the dataset was generated under seven different lighting conditions from 49 or 64 positions in the laboratory. Calibrated camera parameters and ground truth points generated by structured light scanning were provided. The scene we choose contains 64 images of 1600×1200 pixels. To evaluate the effectiveness of our specular component, we choose 7 images with some specular areas shown in Figure 2.

Figure 2. The 7 images of Buddha in the DTU dataset.

As shown in Figure 3, the initial surface (second column) generated by CMPMVS lacks the fine and sharp structures and is over-smoothed. In the area with specular reflection, SREVA (third column) creates many artifact details, whereas SREVAS (fourth column) can keep the surface smooth as shown in the second row. However, both SREVA and SREVAS can recover fine details in many areas such as the one shown in the third row. This result demonstrates that the proposed method can not only recover fine details but also keep smooth in specular reflectance areas.


(a) Ground truth　　(b) CMPMVS　　(c) SREVA　　(d) SREVAS
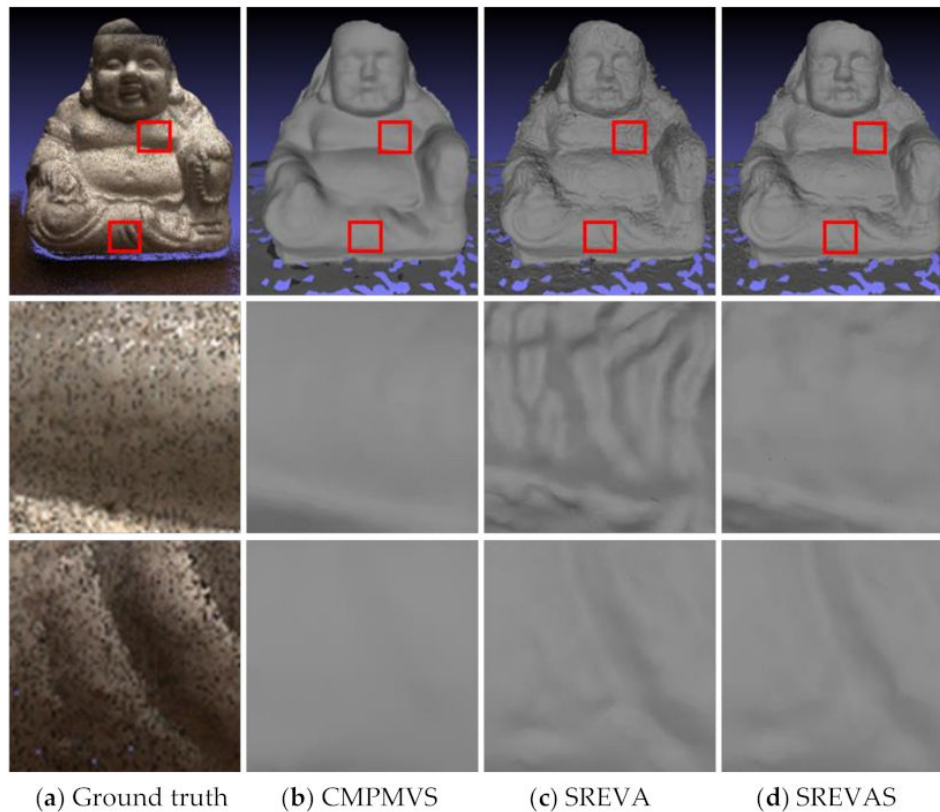
Figure 3. Buddha in the DTU dataset. (**a**) The ground truth, (**b**) the initial surface from CMPMVS, (**c**) the surface refined by SREVA and (**d**) the surface refined by SREVAS. The second and third rows are respectively for the top and bottom small squares in the scene.

To quantitatively evaluate the surfaces from CMPMVS, SREVA and SREVAS, we use the algorithm provided by Jensen et al. [23]. The results are evaluated based on the accuracy and completeness [23], where the accuracy is measured as the distance from the results to the structured light reference and the completeness is measured from the reference to the results. The mean value, median value and root mean square value of the distances are computed. Table 1 shows that the proposed SREVAS method performs the best compared to its specular-free SREVA version and CMPMVS, yielding an improvement of 1.1–12.8% in position accuracy compared to the initial surface. The improvement is calculated through dividing the accuracy difference between SREVAS and CMPMVS by the accuracy of CMPMVS. Dropping the specular component from the SREVAS model has caused quality deterioration in the resultant surfaces. This demonstrates the necessity of considering the specular component for surface refinement.

Table 1. Quantitative evaluation for the Buddha in DTU dataset (in mm).

Method	Completeness			Accuracy		
	Mean	Medium	RMSE	Mean	Medium	RMSE
CMPMVS	0.52	0.39	0.75	0.47	0.23	0.91
SREVA	0.52	0.40	0.75	0.52	0.25	0.99
SREVAS	**0.46**	**0.34**	**0.68**	**0.45**	**0.21**	**0.90**

To evaluate the influence of the specular component in our approach when the actual reflectance is perfect Lambertian (i.e., no specular reflectance exists), the following test is designed. We use the dataset created by [24] with the well-known synthetic surface model "Joyful Yell". According to [24], a total of 37 input images of 2048 × 1536 pixels were generated with the CG rendering software (https://www.blender.org/). As shown in Figure 4, each image was under a single color but randomly generated light source, while the albedo of the model was colored to be non-constant with the CG software (https://www.blender.org/). For the reflectance model, the object was assumed to have a perfect Lambertian surface.
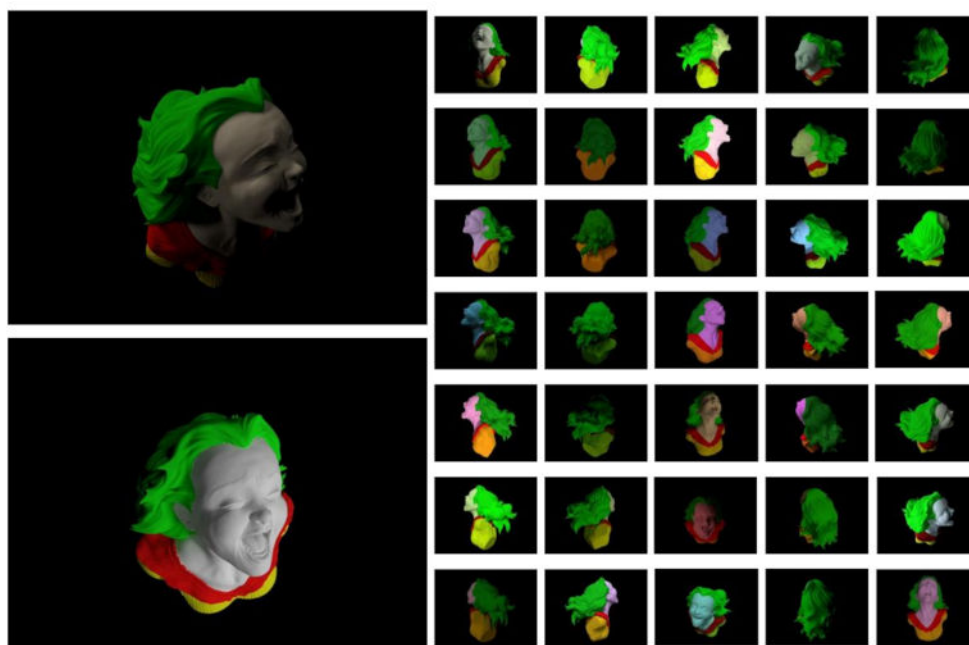
Figure 4. The 37 images of the "Joyful Yell" dataset.

As shown in Figure 5, the results from multi-view stereo CMPMVS lack fine details in the face, ear, hair and clothes of the model, while MVIR and the proposed method, either without or with the specular component, can recover fine and sharp details in those places. Besides, the CMPMVS surface is very rough while the results of the proposed method and MVIR are smooth. As mentioned above, it is hard to find good correspondence in texture-less areas for MVS. Besides, since images were generated under very different illumination conditions, the bad performance of CMPMVS is not unexpected. As for the proposed method either with or without specular component, shading is used to recover the fine details and geometry constraint is applied to assure the surface be smooth. It should be noted that our model is general, in other words, the specular component is included so as to prove the capability and generalization of our approach. The results demonstrate that our solution technique is quite stable even when the model has certain redundancy, such as the specular reflection parameters. The potential effect of the over-parameterization problem by introducing the specular component is minimal and can be ignored in practice.
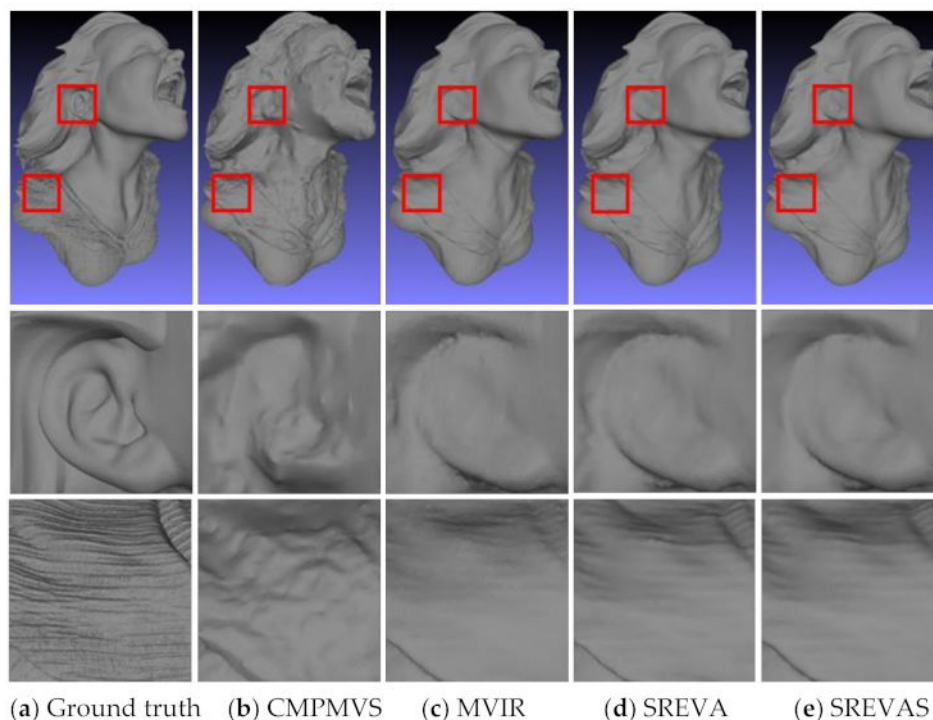
(a) Ground truth   (b) CMPMVS   (c) MVIR   (d) SREVA   (e) SREVAS

Figure 5. The Joyful Yell dataset. (**a**) The ground truth, (**b**) the initial surface from CMPMVS, (**c**) the surface refined by MVIR, (**d**) the surface refined by SREVA and (**e**) the surface refined by SREVAS. The second and third rows are respectively for the top and bottom small squares in the scene.

To quantitatively evaluate the results of different methods, the depth and surface normal of all the results and the ground truth are computed. Examples of the relative depth (divided by a mean depth of the ground truth) and surface normal errors are shown in Figures 6 and 7. As shown in Figure 6, no matter the smooth area (the left small square) or the area with many detailed shapes (the right small square), MVIR, SREVA and SREVAS significantly reduce the depth errors while the result of SREVA is slightly better. When we compare the surface normal, Figure 6 shows that MVIR, SREVA and SREVAS greatly improve the accuracy of the surface normal from CMPMVS in regions with either many detailed shapes (the left small square) or smooth shapes (the right small square). The result demonstrates that our method can refine the surface normal (i.e., surface shape) and improve the accuracy of the surface at the same time.

The overall root mean square (RMS) of relative depth errors and normal errors are computed for all images shown in Table 2. The table shows that MVIR and both versions of the proposed method clearly improve the accuracy in both depth and surface normal comparing to the input surface generated by CMPMVS. SREVAS improves 10.1% of the depth accuracy and 17.6% of the normal accuracy. Since the reflectance model of surface is assumed to be pure Lambertian, SREVA achieves the highest accuracy in depth and performs slightly better than SREVAS, whose result is still acceptable and slightly better than MVIR. Figure 8 further shows the accumulated distribution of depth errors of the surface reconstructed by CMPMVS and refined by MVIR, SREVA and SREVAS. It can be observed that all methods can improve the number of accurate pixels while SREVA and SREVAS are better than MVIR, especially when the relative depth error is less than 0.4%.
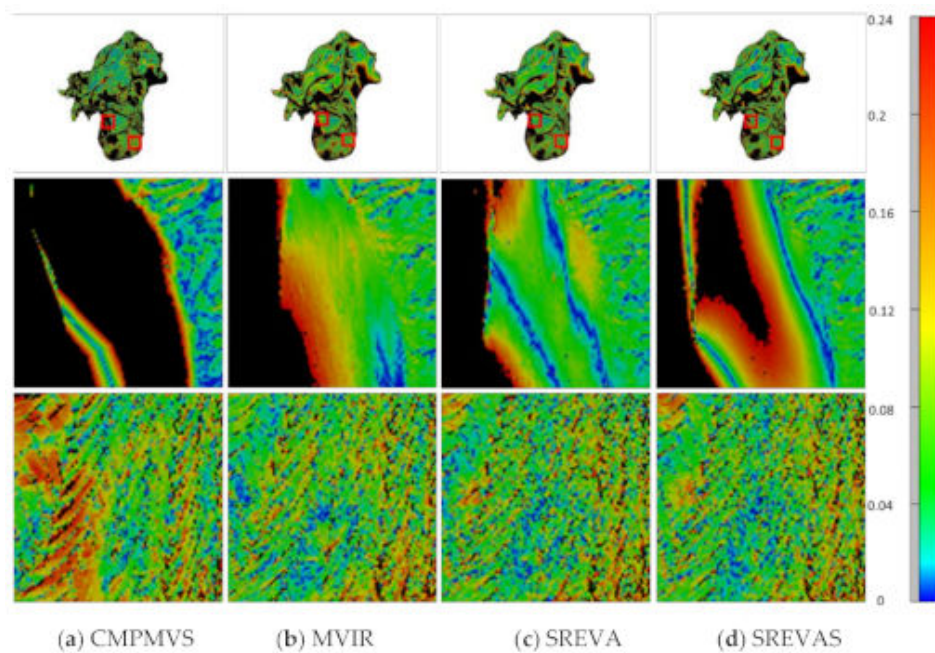
Figure 6. Distribution of the relative depth errors for the Joyful Yell dataset. (**a**) The relative depth errors of the initial surface from CMPMVS, (**b**) the surface refined by MVIR, (**c**) the surface refined by SREVA and (**d**) the surface refined by SREVAS. The second and third rows are respectively for the top and bottom small squares in the scene. The black pixels encode the relative depth error larger than 0.24%.
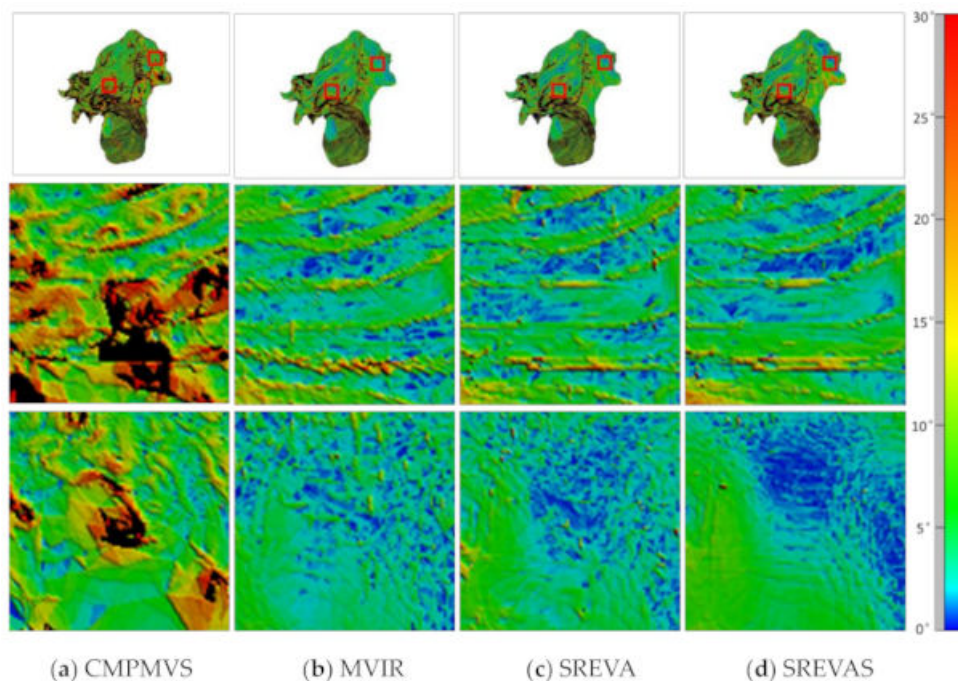

Figure 7. Distribution of the normal errors for the Joyful Yell dataset. (**a**) The normal error of the initial surface from CMPMVS, (**b**) the surface refined by MVIR, (**c**) the surface refined by SREVA and (**d**) the surface refined by SREVAS. The second and third rows are respectively for the left and right small squares in the scene. The black pixels encode the angular error larger than 30°.

Table 2. Quantitative evaluation on the 3D reconstruction results for the Joyful Yell dataset.

Method	RMS Relative Depth Error (%)	RMS Normal Error (Degree)
CMPMVS	1.29	25.23
MVIR	1.17	21.03
SREVA	**1.11**	20.80
SREVAS	1.16	**20.78**
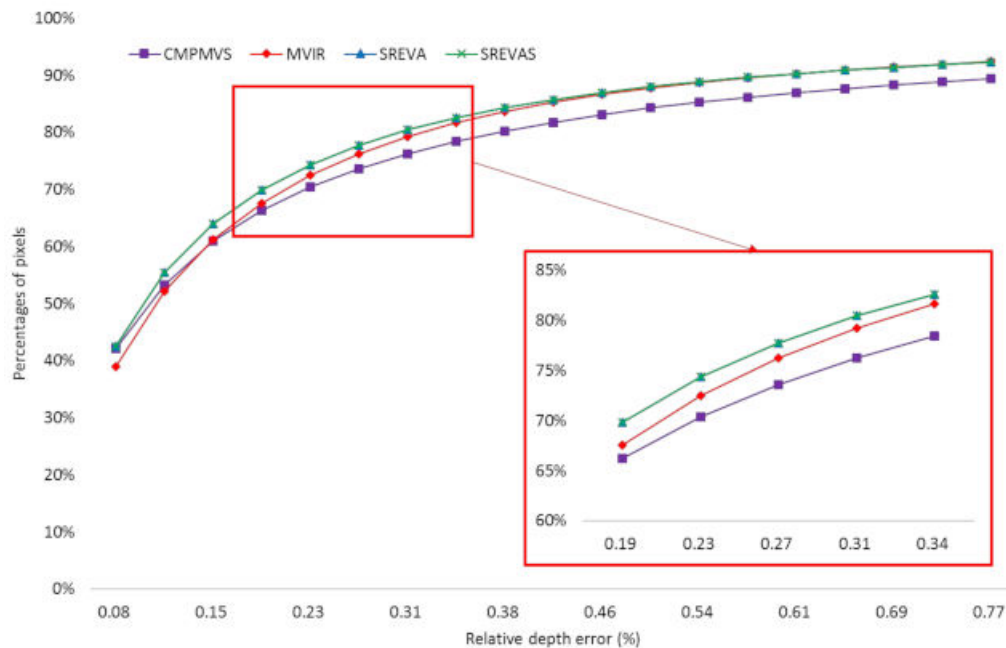
Figure 8. Accumulative distribution of relative depth errors of the surfaces generated by CMPMVS and refined by MVIR, SREVA and SREVAS on the Joyful Yell dataset.

The experiments above have shown that the proposed SREVAS method can recover fine details that CMPMVS missed and improve the accuracy of the surface no matter if the surface reflection model is a mixture of Lambertian and specular or perfect Lambertian. SREVA can perform slightly better when the actual surface reflection model is set to be pure Lambertian, but poorly when there is a mixture of Lambertian and specular reflections. Therefore, SREVAS is a better and reliable choice unless we are certain that the surface reflection model is pure Lambertian.

4.2. Performance of SREVAS

To further evaluate the performance of the proposed method, experiments on the Herz-Jesu-P8 dataset are first conducted. The images of the Herz-Jesu-P8 dataset [25] were taken under natural illumination with accurate camera parameters. There were 8 images of 3072 × 2028 pixels in the Herz-Jesu-P8 dataset. The ground truth 3-D model was obtained using accurate light detection and ranging (laser scanning).

Figure 9 shows, from the top row, the ground truth, the initial surface generated by CMPMVS, the refined surface by SREVAS and the surface reconstructed by SMVS. The illumination conditions of all the images in the Herz-Jesu-P8 dataset are nearly the same, as such the overall shape is reconstructed well by CMPMVS whereas some sharp details are over-smoothed or missed. As for SMVS, the surface is over-smoothed and there are many void areas. In contrast, Figure 8 depicts that the surface recovered by SREVAS has more fine details in the left and right squares; the shapes recovered by SREVAS can significantly better represent the ground truth.
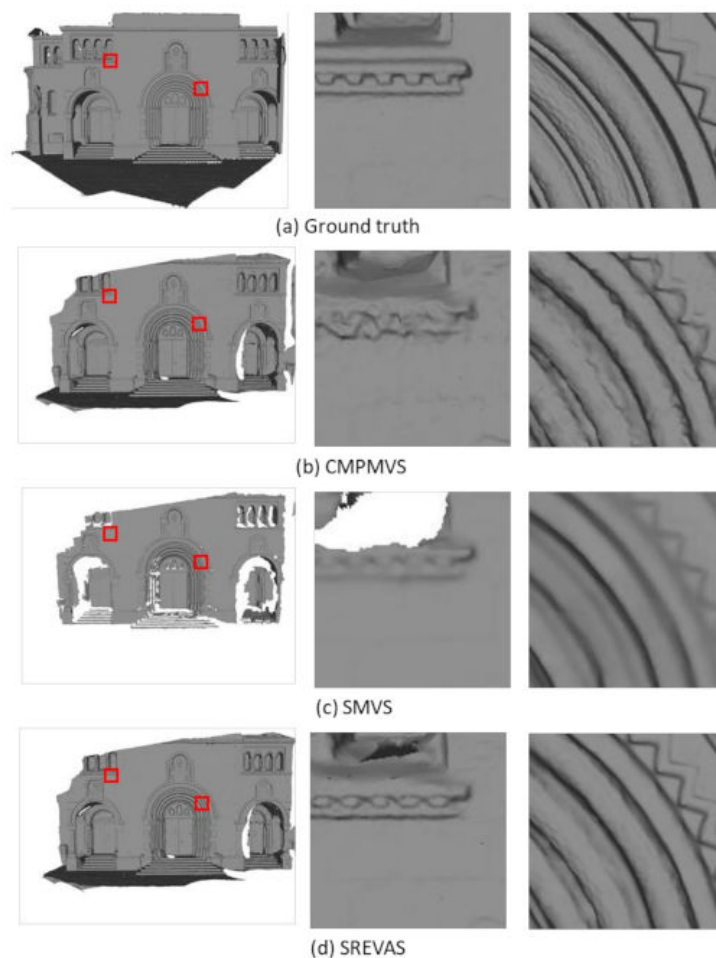
Figure 9. Results of the Herz-Jesu-P8 dataset. (**a**) The ground truth, (**b**) initial surface by CMPMVS, (**c**) the surface reconstructed by SMVS and (**d**) the surface refined by SREVAS. The second and third columns are respectively for the left and right small squares in the scene.

As shown in Table 3, SREVAS yields slight improvement (0.4%) in the accuracy of depth compared to the initial surface from CMPMVS. As for the surface normal, the proposed method still shows improvement (2.5%) comparing to the initial surface. SMVS combines shading with MVS in a single optimization scheme and achieves the best for accuracy of depth and normal among the three methods. However, as shown in Figure 8, there are many void areas in the surface reconstructed by SMVS because it discards many poor points after reconstruction [27]. Therefore, we calculate the omission rate which is the ratio of the number of pixels not existing in the reconstructed surface compared to the ground truth. As shown in Table 3, the omission rate of SMVS is the highest (24.25%). SREVAS can achieve the best balanced performance in terms of details, position accuracy, normal accuracy and omission rate. It is shown that SREVAS is very dependent on the initial surface. Although a more detailed shape can be recovered by SREVAS as shown in Figure 8, the accuracy in depth and normal has not been improved that much to be comparable with SMVS.

Table 3. Quantitative evaluation on the Herz-Jesu-P8 dataset.

Method	RMS Depth Error (cm)	RMS Normal Error (Degree)	Omission Rate (%)
CMPMVS	21.97	24.21	6.12
SMVS	**18.57**	**22.05**	24.25
SREVAS	21.89	23.60	**6.11**

Below we present a qualitative evaluation by using the Fountian-P11 dataset. Similar to the Herz-Jesu-P8 dataset, the images of the Fountian-P11 dataset [25] were also taken under natural illumination with accurate camera parameters. There were eleven 3072×2048 images. Figure 10 shows the reconstructed surfaces, including ground truth (from laser scanning), the initial surface from CMPMVS, the refined surface by SREVAS and the surface reconstructed by SMVS, respectively. The overall shapes are well reconstructed by all three methods. However, there are apparent void areas in the surface reconstructed by SMVS. It also results in an over-smoothed surface without many details, whereas the surface from CMPMVS is noisy. In contrast, results from SREVAS are sharper and more similar to the ground truth than CMPMVS and SMVS.
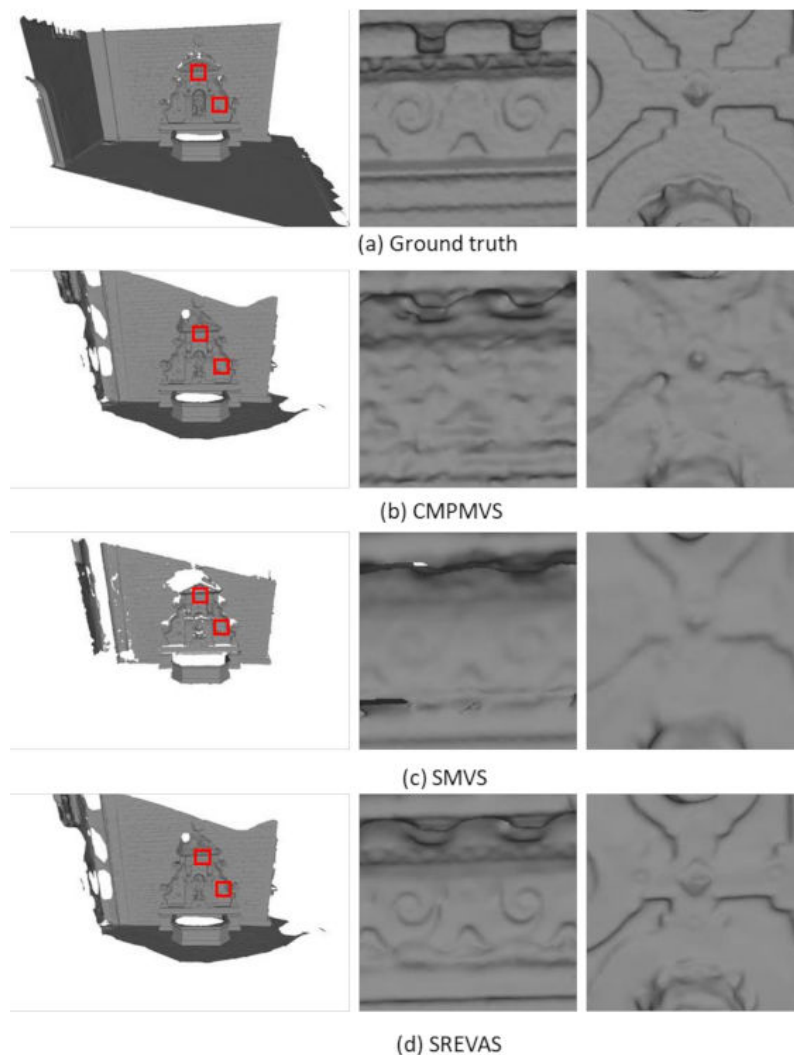

Figure 10. Results of the Fountain-P11 dataset. (**a**) The ground truth, (**b**) initial surface by CMPMVS, (**c**) the surface refined by SREVAS and (**d**) the surface reconstructed by SMVS. The second and third columns are respectively for the top and bottom small squares in the scene.

After experimenting on the datasets with camera parameters, we also evaluate the proposed SREVAS on an Internet dataset without camera parameters. As shown in Figure 1a, the Internet dataset used is the Yorkminster in the 1DSfM dataset [26]. Similarly to the work of [24], the same 9 images are used for our experiment. As we can see from Figure 10, the illumination conditions among the 9 images are very different. Since there are no camera parameters for the images, VisualSFM is first used to estimate the camera parameters and then CMPMVS is applied to recover the initial surface for further refinement.

Figure 11 shows the initial surface from CMPMVS, the surface refined by MVIR and SREVAS and the surface reconstructed by SMVS. Considering the large difference in illumination conditions and the absence of camera parameters, it is hard to achieve a fine surface with CMPMVS, which gives us some room to refine its result. Both MVIR and the proposed SREVAS clearly improve the details of the surface compared to the initial surface. However, it is obvious that SREVAS outperforms MVIR, such as in the places marked with a red rectangle. Compared to MVIR, the shapes refined by SREVAS become much more realistic to the ones shown in images. In the last red rectangle, MVIR recovers some details while generating a lot of artifact details as well. As for SREVAS, the shapes are well recovered and similar to the shapes in the original images. Considering that the Lambertian assumption is often violated and there are many shadows in the images, the unsatisfactory performance of MVIR can be explained. For the proposed SREVAS, the modeling of specular reflectance is more practical and can better recover the detailed shape even when shadows are visible. As for SMVS, some detailed shapes can be reconstructed comparing to the surface reconstructed by CMPMVS, however, there are still many void areas similar to the result of the Herz-Jesu-P8 dataset. In contrast, SREVAS can recover much more detailed shape than SMVS as shown in the zoom-in views of the red rectangles.
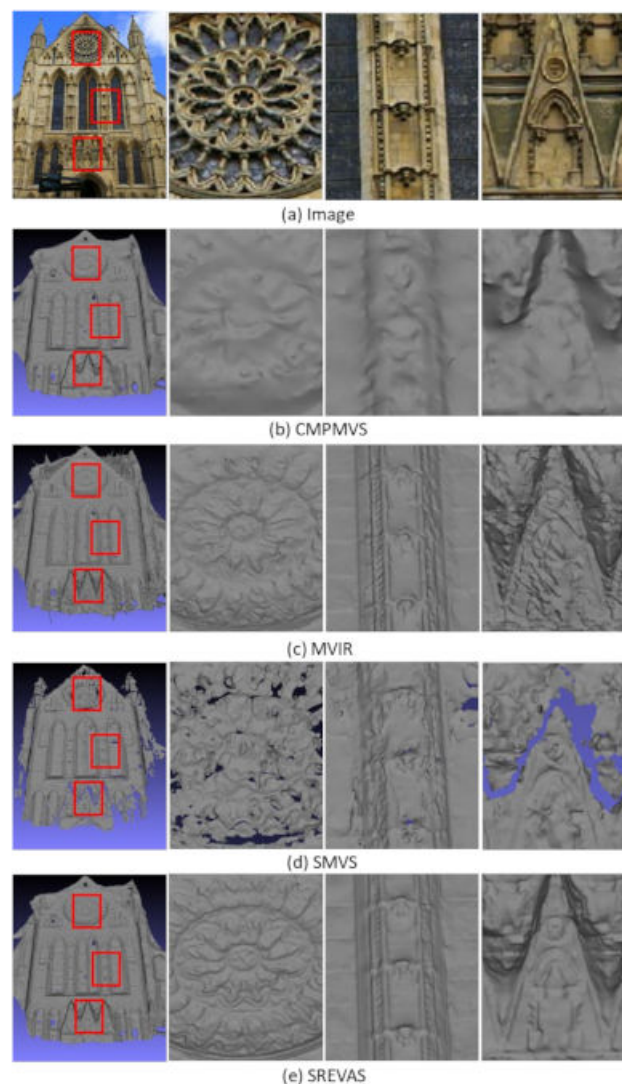

Figure 11. The Yorkminster dataset. (**a**) The reference image, (**b**) the initial surface, (**c**) the surface refined by MVIR, (**d**) the surface reconstructed by SMVS and (**e**) the surface refined by SREVAS. The second, third and fourth columns are respectively for the top, middle and bottom small squares in the scene.

4.3. Susceptibility of the Parameters

In our experiments, the parameters are set with the basic rule that there should not be much difference among the terms in the objective function and mostly only α and β are tuned. Therefore, the susceptibility of α and β is evaluated with the DTU dataset here. As shown in Table 4, the accuracy and completeness [23] of the recovered surface do not change much with different α and β, which means the parameters are not susceptible.

Table 4. Susceptibility of α and β evaluated on the Buddha in DTU dataset (in mm).

α	β	Mean Completeness	Medium Completeness	Mean Accuracy	Medium Accuracy
0.05	0.2	0.48	0.36	0.47	0.23
0.1	0.2	0.47	0.35	0.46	0.22
0.15	0.2	0.49	0.37	0.46	0.22
0.05	0.3	0.46	0.35	0.47	0.23
0.1	0.3	**0.46**	**0.34**	**0.45**	**0.21**
0.15	0.3	0.47	0.35	0.46	0.22
0.05	0.4	0.47	0.35	0.48	0.23
0.1	0.4	0.47	0.35	0.46	0.22
0.15	0.4	0.47	0.35	0.46	0.22

4.4. Runtime

The proposed method is implemented using C++ with external dependencies: Ceres Solver [73] and OpenCV [74]. We experiment on a standard Windows 10 computer with an Intel Xeon CPU of 64 GB memory without GPU optimization. As shown in Table 5, SREVAS has the highest computational cost (1.3×–3.0×) for the Herz-Jesu-P8 dataset since the variables in our framework are the most and no accelerated process is applied. Nevertheless, SREVAS is able to produce the most points (1.2–3.1× more) among all three methods, a necessity to achieve fine details in surface reconstructions. It should also be noted that the computational efficiency is about the same across the three methods, considering the runtime per point.

Table 5. Runtime of the MVIR, SMVS and SREVAS for the Herz-Jesu-P8 dataset.

Method	# Images	Image Size	# Points	Runtime (Hour)
MVIR	8	3072 × 2028	2,021,305	3.45
SMVS	8	3072 × 2028	820,905	1.50
SREVAS	8	3072 × 2028	2,523,050	4.50

4.5. Limitations

There are several limitations observed for SREVAS. It has model bias since the spherical harmonics assume distant lighting and convex objects. Nevertheless, our experiences show that SREVAS can still refine the results from multi-view stereo but without achieving its best performance. Furthermore, the low omission rate of surface reconstruction with SREVAS can sometimes be at the price of relatively low geometry accuracy, comparing to SMVS where shape-from-shading and multi-view stereo are combined in a single optimization scheme. Finally, the performance of SREVAS is dependent on the quality of the initial surface. As shown in Figure 12, some details on the surface cannot be well recovered due to the existence of occlusions and shadows in the input images. Similarly, when the input images are taken under nearly the same illumination and can be well reconstructed by image matching, SREVAS can only slightly improve the position and normal accuracy, despite being able to yield more fine details in texture-weak regions.
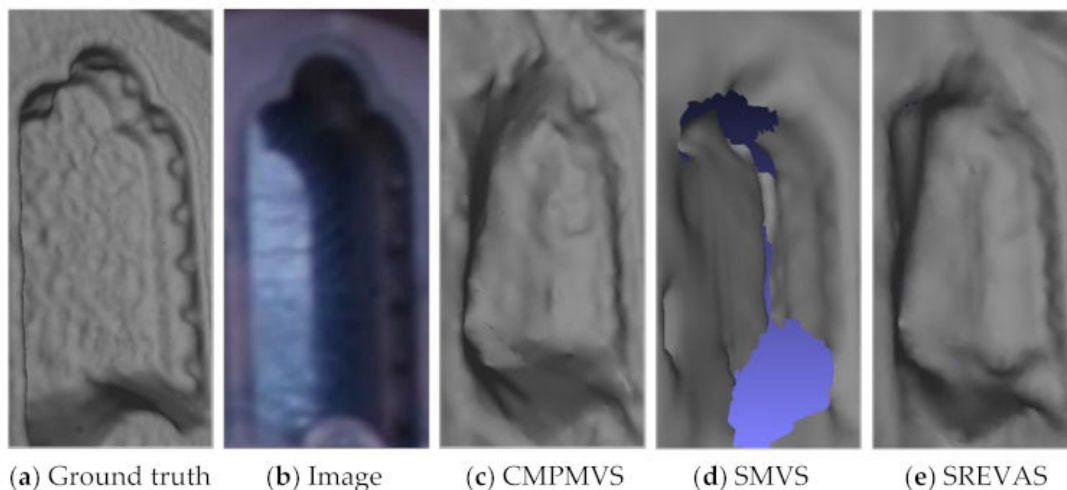
(a) Ground truth　　**(b)** Image　　**(c)** CMPMVS　　**(d)** SMVS　　**(e)** SREVAS

Figure 12. Difficult regions due to shadow and occlusion in the Herz-Jesu-P8 dataset. (**a**) The ground truth, (**b**) the reference image, (**c**) the surface reconstructed by CMPMVS, (**d**) the surface reconstructed by SMVS, (**e**) the surface refined by the proposed SREVAS.

5. Conclusions

We have proposed a shading-based surface refinement SREVAS method, which can be used for reconstructing surfaces with varying albedo and specular reflection. Starting from this imaging model, we use an objective function to refine the initial surface generated by MVS. All our experiments demonstrate that this method can refine the surface initially created from multi-view image matching. It is shown that with varying illumination conditions, SREVAS can improve the accuracy up to 10.1% in surface position and 17.6% in the surface normal with lower omission rate compared to the initial surface. To be specific, our investigations achieve the following concluding remarks.

For ideal Lambertian (i.e., no specular reflection) surface (e.g., the synthetic dataset Joyful Yell), the use of SREVAS can still recover fine details with high accuracy, though its final results are slightly worse than the ones from SREVA, which best fits the data under these circumstances. Since a certain magnitude of specular reflection in reality always exists, it is recommended that SREVAS should be used as a common practice.

For scenes with obvious specular reflections (e.g., the DTU dataset) or scenes with a mixture of Lambertian and specular reflections (e.g., the Herz-Jesu-P8 dataset), SREVAS can recover realistic surface details and keep the smoothness of the reconstructed surface. On the contrary, ignoring the specular component would lead to a lot of artifacts in the reconstructed surface. The study demonstrates the necessity and effectiveness of the specular component for shape-from-shading. With an appropriate illumination model and effective solution technique, shading is able to improve the surface resultant from multi-view image matching, especially under the circumstance of specular reflection and weak-texture.

When there are no accurate camera parameters for the input images (e.g., the Yorkminster dataset), the proposed method generates surfaces significantly better than some existing ones, such as CMPMVS and SMVS. This finding suggests that the shape-from-shading technique, in general, can contribute more to surface reconstruction from low cost, off-the-shelf images.

It should be noted that the proposed method assumes the same overall lighting of all pixels in one image. Future work may extend the current model to consider patch-wise lighting conditions in one image. Besides, continuing investigation on how to handle shadows and occlusions in the images is of necessity.

References

1. Agarwal, S.; Snavely, N.; Seitz, S.M.; Szeliski, R. Bundle adjustment in the large. In *European Conference on Computer Vision, Crete, Greece, 5–11 September 2010*; Springer: Berlin/Heidelberg, Germany, 2010.

2. Beardsley, P.; Tort, P.; Zisserman, A. 3D model acquisition from extended image sequences. In *European Conference on Computer Vision, Cambridge, UK, 14–18 April 1996*; Springer: Berlin/Heidelberg, Germany, 1996; Volume 1065, pp. 683–695.

3. Wu, C.; Agarwal, S.; Curless, B.; Seitz, S.M. Multicore bundle adjustment. In Proceedings of the CVPR 2011, Providence, RI, USA, 20–25 June 2011; pp. 3057–3064.

4. Seitz, S.M.; Diebel, J.; Scharstein, D.; Szeliski, R. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 519–528.

5. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1362–1376. [CrossRef]

6. Jancosek, M.; Pajdla, T. Multi-view reconstruction preserving weakly-supported surfaces. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 20–25 June 2011.

7. Vu, H.H.; Labatut, P.; Pons, J.P.; Keriven, R. High accuracy and visibility-consistent dense multiview stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 889–901. [CrossRef]

8. Furukawa, Y.; Hernández, C. *Multi-View Stereo: A Tutorial*; Now Publishers: Hanover, MA, USA, 2015; Volume 9, ISBN 978-3-319-05557-2.

9. Horn, B.K.P. *Shape from Shading*; Horn, B.K.P., Brooks, M.J., Eds.; MIT Press: Cambridge, MA, USA, 1989; pp. 123–171. ISBN 0-262-08183-0.

10. Woodham, R.J. Photometric method for determining surface orientation from multiple images. *Opt. Eng.* **1980**, *19*, 139–144. [CrossRef]

11. Ikehata, S.; Wipf, D.; Matsushita, Y.; Aizawa, K. Photometric stereo using sparse bayesian regression for general diffuse surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1816–1831. [CrossRef]

12. Blake, A.; Zisserman, A.; Knowles, G. Surface descriptions from stereo and shading. *Image Vis. Comput.* **1985**, *3*, 183–191. [CrossRef]

13. Jin, H.; Cremers, D.; Wang, D.; Prados, E.; Yezzi, A.; Soatto, S. 3-D reconstruction of shaded objects from multiple images under unknown illumination. *Int. J. Comput. Vis.* **2008**, *76*, 245–256. [CrossRef]

14. Park, J.; Sinha, S.N.; Matsushita, Y.; Tai, Y.W.; Kweon, I.S. Robust Multiview Photometric Stereo Using Planar Mesh Parameterization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1591–1604. [CrossRef]

15. Samaras, D.; Metaxas, D.; Fua, P.; Leclerc, Y.G. Variable albedo surface reconstruction from stereo and shape from shading. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662), Hilton Head Island, SC, USA, 15 June 2000; Volume 1, pp. 480–487.

16. Wu, C.; Wilburn, B.; Matsushita, Y.; Theobalt, C. High-quality shape from multi-view stereo and shading under general illumination. In Proceedings of the CVPR 2011, Providence, RI, USA, 20–25 June 2011; pp. 969–976.

17. Wu, C.; Liu, Y.; Dai, Q.; Wilburn, B. Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. *IEEE Trans. Vis. Comput. Graph.* **2011**, *17*, 1082–1095.

18. Zollhöfer, M.; Dai, A.; Innmann, M.; Wu, C.; Stamminger, M.; Theobalt, C.; Nießner, M. Shading-based refinement on volumetric signed distance functions. *ACM Trans. Graph.* **2015**, *34*, 1–14. [CrossRef]

19. Nehab, D.; Weyrich, T.; Rusinkiewicz, S. Dense 3D reconstruction from specularity consistency. In Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.

20. Liu-Yin, Q.; Yu, R.; Agapito, L.; Fitzgibbon, A.; Russell, C. Better Together: Joint Reasoning for Non-rigid 3D Reconstruction with Specularities and Shading. Available online: https://arxiv.org/abs/1708.01654 (accessed on 22 October 2020).

21. Song, Y.; Shan, J. Color correction of texture images for true photorealistic visualization. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 308–315. [CrossRef]

22. Or-El, R.; Hershkovitz, R.; Wetzler, A.; Rosman, G.; Bruckstein, A.M.; Kimmel, R. Real-Time Depth Refinement for Specular Objects. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.

23. Jensen, R.; Dahl, A.; Vogiatzis, G.; Tola, E.; Aanaes, H. Large scale multi-view stereopsis evaluation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.

24. Kim, K.; Torii, A.; Okutomi, M. Multi-view inverse rendering under arbitrary illumination and albedo. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016.

25. Strecha, C.; Von Hansen, W.; Van Gool, L.; Fua, P.; Thoennessen, U. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.

26. Wilson, K.; Snavely, N. Robust global translations with 1DSfM. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014.

27. Langguth, F.; Sunkavalli, K.; Hadap, S.; Goesele, M. Shading-aware multi-view stereo. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016.

28. Faugeras, O.; Keriven, R. *Variational principles, surface evolution, PDE's, level set methods, and the stereo problem*; IEEE Transaction on Image Processessing: Berder Island, France, 1998; Volume 7, pp. 336–344.

29. Heise, P.; Jensen, B.; Klose, S.; Knoll, A. Variational patchmatch multiview reconstruction and refinement. *Proc. IEEE Int. Conf. Comput. Vis.* **2015**, *2015 Inter*, 882–890.

30. Esteban, C.H.; Schmitt, F. Silhouette and stereo fusion for 3D object modeling. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Nice, France, 14–17 October 2003; pp. 46–53.

31. Furukawa, Y.; Curless, B.; Seitz, S.M.; Szeliski, R. Manhattan-world stereo. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1422–1429.

32. Kolmogorov, V.; Zabih, R. Computing visual correspondence with occlusions using graph cuts. In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 508–515.

33. Goesele, M.; Curless, B.; Seitz, S.M. Multi-View Stereo Revisited. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; Volume 2, pp. 2402–2409.

34. Kwon, H.; Tai, Y.W.; Lin, S. Data-driven depth map refinement via multi-scale sparse representation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 159–167.

35. Habbecke, M.; Kobbelt, L. Iterative multi-view plane fitting. In Proceedings of the International Fall Workshop of Vision, Modeling, and Visualization, Aachen, Germany, 22–24 November 2006.

36. Gallup, D.; Frahm, J.M.; Pollefeys, M. Piecewise planar and non-planar stereo for urban scene reconstruction. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010.

37. Hou, Y.; Peng, J.; Hu, Z.; Tao, P.; Shan, J. Planarity constrained multi-view depth map reconstruction for urban scenes. *ISPRS J. Photogramm. Remote Sens.* **2018**, *139*, 133–145. [CrossRef]

38. Saouli, A.; Babahenini, M.C. Towards a stochastic depth maps estimation for textureless and quite specular surfaces. In ACM SIGGRAPH 2018 Posters. In Proceedings of the 45th International Conference & Exhibition on Computer Graphics and Interactive Techniques, Vancouver, BC, Canada, 12–16 August 2018.

39. Ikeuchi, K.; Horn, B.K.P. Numerical shape from shading and occluding boundaries. *Artif. Intell.* **1981**, *17*, 141–184. [CrossRef]

40. Zhang, R.; Tsai, P.S.; Cryer, J.E.; Shah, M. Shape from shading: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 690–706. [CrossRef]

41. Forsyth, D.A. Variable-source shading analysis. *Int. J. Comput. Vis.* **2011**, *91*, 280–302. [CrossRef]

42. Johnson, M.K.; Adelson, E.H. Shape estimation in natural illumination. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2011, Providence, RI, USA, 20–25 June 2011.

43. Shen, L.; Tan, P. Photometric stereo and weather estimation using internet images. In Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009.

44. Ackermann, J.; Langguth, F.; Fuhrmann, S.; Goesele, M. Photometric stereo for outdoor webcams. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.

45. Oxholm, G.; Nishino, K. Multiview shape and reflectance from natural illumination. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2163–2170.

46. Mo, Z.; Shi, B.; Lu, F.; Yeung, S.K.; Matsushita, Y. Uncalibrated Photometric Stereo under Natural Illumination. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 19–21 June 2018.

47. Basri, R.; Jacobs, D. Lambertian reflectances and linear subspaces. *IEEE Int. Conf. Comput. Vis.* **2001**, *25*, 383–390.

48. Basri, R.; Jacobs, D. Photometric stereo with general, unknown lighting. *Int. J. Comput. Vis.* **2001**, *72*, 239–257. [CrossRef]

49. Chen, L.; Zheng, Y.; Shi, B.; Subpa-Asa, A.; Sato, I. A Microfacet-Based Reflectance Model for Photometric Stereo with Highly Specular Surfaces. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.

50. Mélou, J.; Quéau, Y.; Durou, J.D.; Castan, F.; Cremers, D. Beyond multi-view stereo: Shading-reflectance decomposition. In *International Conference on Scale Space and Variational Methods in Computer Vision, Kolding, Denmark, June 4–8, 2017*; Springer: Cham, Switzerland, 2017; pp. 694–705.

51. Quéau, Y.; Mélou, J.; Durou, J.-D.; Cremers, D. Dense Multi-view 3D-reconstruction Without Dense Correspondences. Available online: https://arxiv.org/abs/1704.00337 (accessed on 22 October 2020).

52. Mélou, J.; Quéau, Y.; Castan, F.; Durou, J.D. A Splitting-Based Algorithm for Multi-view Stereopsis of Textureless Objects. In *International Conference on Scale Space and Variational Methods in Computer Vision*; Springer: Cham, Switzerland, 2019; pp. 51–63.

53. Yu, L.F.; Yeung, S.K.; Tai, Y.W.; Terzopoulos, D.; Chan, T.F. Outdoor photometric stereo. In Proceedings of the 2013 IEEE International Conference on Computational Photography, Cambridge, MA, USA, 19–21 April 2013.

54. Xie, W.; Dai, C.; Wang, C.C.L. Photometric stereo with near point lighting: A solution by mesh deformation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.

55. Tsiotsios, C.; Davison, A.J.; Kim, T.K. Near-lighting Photometric Stereo for unknown scene distance and medium attenuation. *Image Vis. Comput.* **2017**, *57*, 44–57. [CrossRef]

56. Nehab, D.; Rusinkiewicz, S.; Davis, J.; Ramamoorthi, R. Efficiently combining positions and normals for precise 3D geometry. *ACM Trans. Graph.* **2005**, *24*, 536–543. [CrossRef]

57. Shi, B.; Inose, K.; Matsushita, Y.; Tan, P.; Yeung, S.K.; Ikeuchi, K. Photometric stereo using internet images. In Proceedings of the 2014 International Conference on 3D Vision, Tokyo, Japan, 8–11 December 2014.

58. Barron, J.T.; Malik, J. Intrinsic scene properties from a single RGB-D image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*. [CrossRef]

59. Han, Y.; Lee, J.Y.; Kweon, I.S. High quality shape from a single RGB-D image under uncalibrated natural illumination. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013.

60. Wu, C.; Zollhöfer, M.; Nießner, M.; Stamminger, M.; Izadi, S.; Theobalt, C. Real-time shading-based refinement for consumer depth cameras. *ACM Trans. Graph.* **2014**, *33*, 1–10. [CrossRef]

61. Yu, L.F.; Yeung, S.K.; Tai, Y.W.; Lin, S. Shading-based shape refinement of RGB-D images. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013.

62. Choe, G.; Park, J.; Tai, Y.W.; Kweon, I.S. Exploiting shading cues in kinect IR images for geometry refinement. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.

63. Choe, G.; Park, J.; Tai, Y.W.; Kweon, I.S. Refining Geometry from Depth Sensors using IR Shading Images. *Int. J. Comput. Vis.* **2017**, *122*, 1–16. [CrossRef]

64. Xu, D.; Duan, Q.; Zheng, J.; Zhang, J.; Cai, J.; Cham, T.J. Shading-Based Surface Detail Recovery under General Unknown Illumination. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 423–436. [CrossRef]

65. Weinmann, M.; Ruiters, R.; Osep, A.; Schwartz, C.; Klein, R. Fusing structured light consistency and Helmholtz normals for 3D reconstruction. In Proceedings of the British Machine Vision Conference, Guildford, UK, 3–7 September 2012; pp. 1–12.

66. Maurer, D.; Ju, Y.C.; Breuß, M.; Bruhn, A. Combining Shape from Shading and Stereo: A Joint Variational Method for Estimating Depth, Illumination and Albedo. *Int. J. Comput. Vis.* **2018**, *126*, 1342–1366. [CrossRef]

67. Shan, Q.; Adams, R.; Curless, B.; Furukawa, Y.; Seitz, S.M. The visual turing test for scene reconstruction. In Proceedings of the 2013 International Conference on 3D Vision—3DV 2013, Seattle, WA, USA, 29 June–1 July 2013; pp. 25–32.

68. Wu, C. VisualSFM: A Visual Structure from Motion System. Available online: Ccwu.me/vsfm (accessed on 22 October 2020).

69. Wu, C. Inverse Rendering for Scene Reconstruction in General Environments. Ph.D Thesis, Saarland University, Saarbrücken, Germany, 2014.

70. Barron, J.T.; Malik, J. Shape, albedo, and illumination from a single image of an unknown object. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition; Providence, RI, USA, 16–21 June 2012.

71. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, India, 7 January 1998; pp. 839–846.

72. Jeon, J.; Cho, S.; Tong, X.; Lee, S. Intrinsic image decomposition using structure-texture separation and surface normals. In *European Conference on Computer Vision*; Sringer: Cham, Switzerland, 2014.

73. Agarwal, S.; Mierle, K. Others Ceres Solver. Available online: http://ceres-solver.org (accessed on 22 October 2020).

74. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools.* **2000**, *25*, 120–125.