An RGBN Benchmark

Sema Berkiten Szymon Rusinkiewicz
Princeton University

Abstract

A variety of algorithms in both computer vision and graphics use datasets of an object or scene captured with fixed camera but varying illumination. Evaluating these algorithms is frequently challenging because of the lack of ground truth on the one hand, and insufficiently realistic and varied synthetic datasets on the other. In this work, we present a synthetic benchmark for applications such as photometric stereo, and justify it by comparing to real-life objects and their rendered models. Additionally, we introduce a system that allows the user to create scenes by combining arbitrary 3D models, materials, and light configurations. The system outputs physically-based renderings as well as dense ground-truth maps of quantities such as normals, height map, BRDF specifications, and albedo. We present a number of synthetic datasets which will be available online, and we provide a few photometric datasets of real-life objects. Our work demonstrates that real objects can be simulated well enough so that the conclusions about accuracy drawn from our synthetic datasets match those based on real objects. The paper also demonstrates a use case for this RGBN benchmark: the evaluation of photometric stereo algorithms. We present a taxonomy of photometric stereo techniques, investigate the causes of errors in several of them, and propose a photometric stereo variant that iteratively estimates shadowing.

1. Introduction

Photometric datasets have been used by various applications in computer vision and graphics including BRDF acquisition [17], photometric stereo [34], (non)photo-realistic rendering, illumination representations such as polynomial texture mapping [24], linear image subspaces [7], and digitization of cultural heritage. Despite the fact that photometric datasets are used in a variety of fields, the literature lacks a realistic benchmark with ground truth. This may be due to the fact that it is tedious to create a benchmark that (1) is based on real-world data with ground truth, and (2) contains several versions of the same scene, with one parameter (e.g., noise level, lighting configuration, material, etc.) varied while everything else is kept constant.

We propose a synthetic but realistic (physically-based) and validated (against ground-truth data) benchmark for evaluation of photometric algorithms. Each benchmark dataset includes a collection of images of a scene under different lighting, together with ground-truth surface normals, height map, BRDF attributes, and albedo. The datasets are organized into "experiments," in which one parameter of the scene is varied while others are kept constant.

There are two major benefits to using synthetic data for our benchmark. First, the ground truth available with each dataset is *real* ground truth. An alternative would have been to use real images, but obtain "ground truth" using, say, a 3D scanner. This has a major problem: the precision and the calibration of the scanning system, alignment of partial scans to form the whole model, and alignment of the 3D model to the photometric data are all sources of error in the "ground truth" geometry or normal maps. Even with high-precision scanners, the user of the benchmark might not know whether the biases introduced in the "ground truth" might favor some algorithms over others.

The second major benefit of using synthetic data is that it is possible to isolate the effects of noise, lighting, materials, etc. By conducting "experiments" in which individual factors vary in isolation, users of the benchmark are able to make more precise conclusions about the relative strengths and weaknesses in the algorithms they are evaluating.

There are also two potential drawbacks to using synthetic data, and we have attempted to mitigate them. First, the renderings might not be realistic. To accurately simulate as many light-transport phenomena as possible, we generate the benchmark images using Mitsuba Renderer [21]. Mitsuba Renderer is a physically-based rendering tool that supports various surface materials, subsurface-scattering, glittering effects etc. It also takes all global illumination into account during rendering.

Even with a physically-based renderer, there may be a concern that conclusions about algorithms based on synthetic images might not match those based on real-world images. Therefore, to justify the use of synthetic data, we conduct an experiment comparing the error patterns of a photometric stereo algorithm on several real objects and

their 3D scanned models in Section 4.

In addition to the benchmark itself, we provide a flexible but simple user interface that allows users to create their own datasets while controlling variables such as light configuration, light intensity, noise level, amount or characteristics of ambient lighting. Finally, we explore one of the applications of the benchmark: evaluation of various photometric stereo algorithms. Photometric stereo is the general name of the algorithms that try to estimate surface properties, most notably normal maps and surface color, from multiple images under varying illumination.

In summary, our contributions in this work are 3-fold:

- Publicly available photometric datasets as well as the software to create new photometric datasets with ground truth (Section 3);
- A new calibrated photometric stereo algorithm which estimates the surface normals and the albedo with an iterated least squares method by updating "pixel confidence" values (Section 6);
- An example usage of the proposed benchmark: evaluation of various photometric stereo algorithms (Section 6) along with a taxonomy of photometric stereo algorithms (Section 5).

2. Related Work

In the literature, there are several publicly available photometric datasets under different illuminations. For example, [4], [20], [35], [3], [37], and [32] provide photometric datasets that they used in their work. However, there is no precise ground truth for these datasets to enable quantitative comparisons between algorithms.

The most relevant benchmark there exist is proposed by [1] . In their work, they present a robotic system to capture single-view images of an object under varying illuminations as well as a 3D scan via a structured-light scanner. They use the 3D scans to generate the ground truth, however the 3D point cloud is not dense in order to form a high quality and accuracy depth map or normal map. Another disadvantage of this benchmark is the limited number of lights per view (20 LED lights in a planar arrangement) and it is not possible to expand the benchmark by other users.

Other related benchmarks are for intrinsic image algorithms, published by Grosse et al.[18] and Bell et al.[8]. The datasets in these benchmarks are a subset of all the photometric datasets we focus on this paper. For intrinsic image estimations, they only need one image as the input data and to create ground truth Grosse et al.[18] captured 10 more images, while Bell et al.[8] used crowd-sourcing to obtain the ground truth. In summary, these benchmarks provide one image as input and *relative* shading and reflectance per scene, as opposed to the dense ground truth data for normal map, material, and height map provided by our benchmark.

For other applications in computer graphics/vision, there are several reputable benchmarks. For example, Scharstein and Szeliski published a comprehensive benchmark on two-frame stereo [28], and subsequently a number of works used their system and datasets to validate their work. A benchmark for 3D object segmentation was presented by [14], and [25] published a database for 2D image segmentation. One common property of all these benchmarks is that they are manually created. In the literature, there are not only real-world datasets which are manually labeled by users, but also some synthetic benchmarks. The most recent benchmark which proposes using synthetic datasets is for 3D surface reconstruction [9].

The importance of a well-prepared benchmark is undeniable. When a field lacks of a reference to compare different approaches, not only the users but also the researchers in the field have difficulties to determine the best approach/algorithm for their problems. As a result, some problems such as photometric stereo do not have consensus best-of-breed solutions, even though they are well-studied. We believe that a benchmark with high precision will help us to understand the problem better and explore the reasons why some algorithms actually fail.

3. Benchmark Overview

Our system is designed to create realistic photometric data along with their corresponding ground truth from arbitrary 3D scenes. We use the Mitsuba Renderer [21] to produce photo-realistic renderings. We also capture a few real-world objects to use in justification of the synthetic datasets.

3.1. Synthetic Data

In this section, we introduce the process of synthetic dataset creation from a single viewpoint under different illuminations. Although our system is designed for single-view photometric data, it can be used for multi-view photometric data applications, such as [19], without any modification.

The production of the synthetic benchmark is composed of two parts: i) creation of an arbitrary scene, ii) producing the photometric data and ground truth.

3.1.1 User Interface

We provide a simple user interface to create scenes for our benchmark. Although general modeling packages, such as Blender, could be used for this, our interface is customized to produce photometric data: it provides fewer options than a package such as Blender, and it can create variations of the same scene automatically.

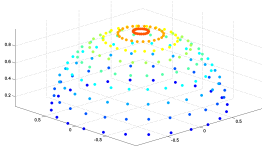
The user interface is customized to create arbitrary scenes from 3D models (refer to the supplementary mate-

rial for details). The user can set each object in the scene, and assign either preset materials or her own material specifications (limited to the materials supported by the Mitsuba Renderer). These materials can be as simple as diffuse single-color albedo or as complex as human skin with both texture and subsurface scattering.

3.1.2 Images and Ground Truth Data

We produce two types of output: photometric data (realistically rendered images) and ground truth data.

Photometric Data: physically-based renderings of the scene under various light directions are produced by Mitsuba Renderer [21]. A typical light configuration is shown at right; however the user can turn on/off different light-rings and right/left hemispheres, as well as change the number of lights and the radius of the light dome.

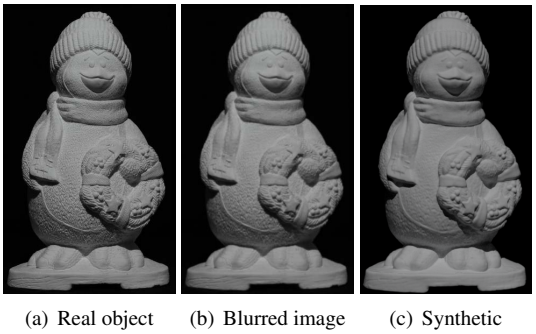


While the resulting renderings include a great deal of global illumination complexity, it is our goal to understand the extent to which that complexity affects the estimation of normals or other photometric quantities. For this reason, we output a set of "clean" ground truth images consisting of:

- **Normal Map:** Per pixel-surface normals are produced as RGB images, using an embedding that maps $[-1..1]$ to $[0..255]$ in each channel.
- **Depth Map:** Euclidean distance in 3D from the camera center to the surface for rays passing through each pixel is saved as floating-point EXR images.
- **BRDF:** Diffuse RGB albedo per pixel is rendered using the Mitsuba Renderer's utility functions. For non-Lambertian materials, the full BRDF specifications can be obtained from the scene files.
- **Light direction/position per image:** For a directional light model, light directions are saved in camera coordinates. For a point light source model, light positions are saved in camera coordinates.

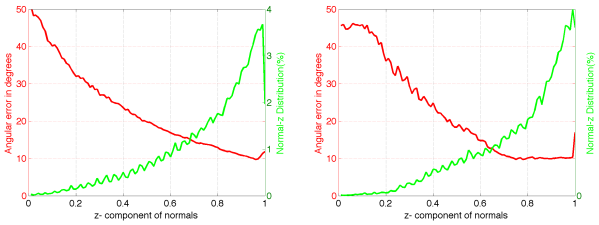
3.2. Real-world Data

One potential concern about a synthetic benchmark is that it might not capture the full complexity of real-world images. Therefore, despite the limitations of scanned 3D data, we perform a real-world evaluation to indicate the plausibility of our synthetic benchmark. We captured several objects using the same setup as [32]: fixed camera on a tripod and a hand-held flash. We took around 36 images per object and one image without flash to compensate for the ambient lighting. We also scanned all the objects with a NextEngine laser scanner to obtain a 3D model, and use this to extract ground truth data. Please refer to the supplementary material for images.



(a) Real object (b) Blurred image (c) Synthetic

(d) Real normals (e) Smoothed (f) 3D model



(g) Real Object: 14.4 (h) Synthetic Object: 13.1

Figure 1. **Verification of the synthetic dataset: (a)** Photograph of the real object, **(b)** Gaussian filtered image of (a), **(c)** Rendering of the 3D model of the object, which is scanned with a NextEngine laser scanner.**(d)**, **(e)**, and **(f)** are the normal maps estimated from real photos, smoothed real normal map, and the ground truth respectively. **(g)** and **(h)** show the angular error distribution of the normal maps recovered from the real photographs and the synthetic renderings respectively. Normal maps are estimated by least squares with shadow/specularity thresholding, and average angular errors are indicated under each plots.

4. Justification of the Synthetic Data

To verify that we can simulate a real-life object, we conducted an experiment that compares photometric stereo results of a real dataset to one created synthetically. First, we pick an object with a single color and relatively diffuse surface but fairly complicated geometry, as shown in Figure 1.a. We obtain photographs of the object, and obtain "ground truth" from a scanned 3D model. The latter

Figure 2. **Albedo acquisition of the Penguin:** An X-rite colorcheckerboard is used for color calibration and albedo estimation.

is created by scanning the object with a NextEngine laser scanner, aligning partial scans with ICP, and reconstructing a final mesh with the Poisson surface reconstruction algorithm [22]. The resolution of the voxel size for surface reconstruction is 0.25 mm, and the image resolution is set accordingly to have a similar resolution. We would like to point out that because of the imperfections and smoothing effects in the reconstruction algorithm, the final 3D model lacks many of the high-frequency details of the real object.

For material acquisition, a flat region of the object, which is the bottom of the object in this case, is photographed with an X-rite color checker as shown in Figure 2. A color profile is extracted from the calibration board and the image is corrected accordingly. Assuming that the object has a single color, we average the color values on the flat region, which is manually selected, of the object. This way, a single color value is obtained for the Penguin. To determine the specularity and roughness of the object, we experiment with varying materials and parameters on the Mitsuba Renderer to get the most realistic image. We use the *roughplastic* material with the diffuse reflectance of *RGB(225,226,228)*, specularity of *0.4*, and *0.1* roughness.

A ground truth normal map of the object is created as follows: the normal map of the object is estimated from the real photographs, the 3D model is aligned to the normal map estimated from the real photographs by using the alignment algorithm in [10], and the normals of the 3D model are rendered in the aligned camera coordinates as shown in Figure 1.d, and f.

Since the 3D model does not have very high-frequency details because of the smoothing effect of the reconstruction, the estimated normal map of the real object is smoothed, as shown in Figure 1.e. Since we do not know how much the reconstruction introduces smoothing to the 3D scan, we determine the amount by comparing the rendered object in Figure 1.c to the real photographs smoothed with different amounts. In Figure 1.b the closest blurred photograph to the rendered image is shown. The plots in Figure 1.g and h show the angular errors in normal maps recovered from the real photographs (1.e) and the synthetic renderings (1.f). The average angular errors are comparable: 14.4 and 13.1 in degrees, respectively. The error distributions over the z- component of the ground truth normals also demonstrate a very similar pattern. Therefore, we conclude that we can create synthetic benchmarks which are comparable to real-life scenarios. We also include a few more similar experiments in the supplementary material.

For the rest of the paper, we will use only synthetic data for evaluations unless otherwise stated.

5. A Taxonomy of Photometric Stereo

Photometric stereo (PS) algorithms can be divided into two categories, calibrated and uncalibrated, based on the availability of information about the illumination conditions. In the same fashion, one can call algorithms (un)calibrated based on other factors such as knowledge of the BRDF, shadow maps, and possibly geometric priors; however we will use knowledge of lighting conditions as the main factor to categorize the algorithms, because it has the biggest effect on the results, as we will discuss in Section 6. The sub-factors that shape the PS algorithms can be categorized broadly into four: illumination, material, outlier, and geometry. We sub-categorize the (un)calibrated PS algorithms based on those factors bellow.

5.1. Calibrated Photometric Stereo (CPS)

Acquisition setup: can be designed for different purposes with the following options:

- Fixed light positions with respect to the camera. For example, light stage [33] or light domes [24] with built in lights and camera(s) on it are commonly used techniques to capture photometric data. They are robust in terms of acquisition of the light directions/positions and repeatability. Drawbacks of these setups are: they are expensive, hard to maintain, take a lot of space, and are not easily scalable.

- Hand-held light sources, with calibration objects such as chrome spheres used to estimate light position and diffuse white spheres to measure the intensity and the color of the light source [32]. This setup supports mobile light sources as opposed to the fixed lights in light domes.

- The light emitters can be modeled as point sources or as area lights. While LED lights usually act as a point light sources, a flash of a camera behaves as an area light source that emits almost parallel light rays.

Illumination: Instead of a single light source per frame, Brostow et al. [11] used 3 colored light- red, green, blue- to perform fast dense-normal recovery on video sequences. Abrams et al. [2] used the outdoor day-light images by using the sun as their light source which requires calibrations

based on the time of the day and the day of the year in a similar way to Ackermann et al. [3]. Other than a point or a directional light source, Ma et al. [23] proposed using a polarized spherical gradient illumination so that they only need 4 images to recover the surface normals.

Material: Hertzmann and Seitz [20] used a known geometry (sphere) in the scene with the same material of the object to create a normal lookup table. However, in real-life it is hard to find or create these calibration objects. By assuming that most objects are composed of a few fundamental materials, Goldman et al. [17] proposed solving the normals for a few BRDFs with varying weights per pixel. Tan et al. [30] assumed that the surface is composed of micro-facets instead of assuming Lambertian surface per pixel to recover higher resolution normals than the original images.

Outliers: Drew et al. [15] handled shadows and specularity by using radial basis function interpolation. Chandraker et al. [12] proposed an algorithm to segment different shadows regions. On the other hand, Wu et al. [36] assumed that the outliers are sparse so they optimize for a low-rank observation matrix plus a sparse outlier matrix.

Geometry: Basri and Jacobs [7] suggested to use a 9D linear subspace using spherical harmonics to map the per-pixel surface properties of convex Lambertian objects.

5.2. Uncalibrated Photometric Stereo (UPS)

Acquisition setup: does not require any special setup other than a fixed camera position and varying illumination.

Illumination: Zhou and Tan [39] solved uncalibrated PS by adding some constraints on the light configuration even though the light positions are not known.

Material: Alldrin et al. [5] assumed that the number of distinct albedos is low, and they solve the generalized bas-relief (GBR) ambiguity by minimizing the entropy of the albedo distribution. Under the Lambertian assumption, Favaro and Papadhimitri [16] exploited maximal Lambertian reflections to solve the GBR ambiguity. Shi et al. [29] clustered the pixels to leverage the strong correlation between the surface normals and the intensity profiles of pixels.

Outliers: Tan et al. [31] leveraged the non-Lambertian reflectance components to solve the Generalized-Bas-Relief (GBR) ambiguity. Similarly, Chandraker et al. [13] used the inter-reflections on the non-convex surfaces to solve the GBR ambiguity. Mukaigawa et al. [26] proposed using the photometric linearization by introducing a classification-based criterion for specular and diffuse reflections, and cast and attached shadows.

Geometry: Basri et al. [6] solved for the GBR ambiguity by using only a few images under the assumption of observed objects' being convex. Queau et al. [27] used total variation regularization to solve for a piece-wise smooth surface. Yuille and Snow [38] showed that the surface integrability constraint reduces the GBR ambiguity to a problem with 3 unknowns.

6. Evaluation of Photometric Stereo

In this section, we conduct various experiments on several photometric stereo algorithms. These experiments are designed to reveal the effects of various factors such as image noise, shadows, inter-reflections, uncompensated ambient illumination, material type, and number of images. We evaluate the following photometric stereo algorithms:

Calibrated Photometric Stereo (CPS):

1. (Lsq) We implemented a least-squares minimization with outlier rejection based on hard thresholds for shadows (40 in the range of $[0, 255]$) and specularity (254 in the range of $[0, 255]$) [34],

2. (RPCS) Low-Rank Matrix Completion and Recovery [36]: solves the problem of recovering a low-rank matrix with both missing and corrupted entries, which models all types of non-Lambertian effects such as shadows and specularities,

3. (Iter Lsq) Motivated by the sensitivity of the simple least-squares algorithm to shadow estimation, we propose an iterative weighted least-squares minization that repeatedly assigns per-pixel confidence values to each image (initially based the intensities in all observed images), computes normals and albedos using those confidences, then updates the confidences based on whether the pixel values agree with (Lambertian) re-renderings of the estimated normals and albedos.

4. (HSH) Hemispherical harmonics [15]: proposes a robust version of PTM [24] by identifying both specular and shadow pixels and then modeling their contribution, separately or together, using a radial basis function interpolation.

Uncalibrated Photometric Stereo (UPS):

1. (SCPS) Self Calibrating Photometric Stereo [29]: automatically determines a radiometric response function and resolves the generalized bas-relief ambiguity by analyzing color/intensity profiles in the RGB and irradiance-time domains – we have tested on only a few cases,

2. (LDR) Lambertian diffuse reflectance [16]: solves the generalized bas relief ambiguity by exploiting points where the Lambertian reflection is maximal,

3. (Entropy) Entropy minimization [5]: proposes a new prior on the albedo distribution that the entropy of the distribution should be low,

		CPS				UPS		
		Lsq	RPCS	Iter Lsq	HSH	LDR	Entropy	TV
Ambient	med	18.73	20.74	**6.95**	26.78	24.64	65.92	**17.19**
	std	16.12	7.27	**6.62**	**5.07**	21.27	24.62	**2.79**
Inter-reflection	med	44.67	44.57	**44.51**	55.40	87.65	**83.16**	93.33
	std	13.76	12.14	14.30	**8.37**	**10.69**	41.85	24.27
Material	med	3.33	8.17	**3.09**	11.79	**13.75**	56.32	18.42
	std	4.58	0.55	**0.51**	1.09	6.31	31.94	**1.68**
Number of Images	med	**0.87**	4.76	**0.83**	13.43	**5.87**	8.52	18.51
	std	**0.88**	13.25	**0.94**	23.30	20.25	35.41	**8.10**
Noise	med	0.96	6.74	**0.80**	91.00	**10.20**	60.95	17.76
	std	0.45	0.39	**0.23**	4.04	0.95	22.28	**0.26**
Ring Position	med	**2.97**	10.84	**3.39**	68.00	**29.80**	60.94	45.79
	std	**13.67**	15.03	**12.40**	16.35	34.76	34.70	**25.67**
Roughness	med	5.74	7.21	**3.77**	14.95	**13.53**	46.37	27.65
	std	1.73	0.46	**0.11**	10.39	**2.87**	26.53	7.39
Shadow	med	**13.70**	51.94	43.24	58.16	99.58	**53.65**	67.39
	std	**5.94**	20.14	27.95	23.22	14.50	20.84	**7.54**
Specularity	med	2.67	6.78	**1.16**	10.74	**7.50**	70.08	13.33
	std	1.76	**0.07**	**0.13**	0.63	**0.79**	3.20	0.86

Figure 3. **Evaluations:** Angular error statistics (median values and standard deviations) of estimated normal maps for various experiments. Errors are measured in degrees.

4. (TV) Total Variation [27]: solves the GBR ambiguity by performing a total variation regularization on both the estimated normal field and albedo.

We conduct experiments on the following isolated factors on the algorithms mentioned above:

- Uncompensated ambient illumination;
- Inter-reflections: tested on cones with the same radius but varying heights;
- Material: tested on materials of different combinations of diffuse, specular, subsurface scattering, and textured materials;
- Number of images: tested on different number of images ranging from 3 to 36;
- Image noise: tested by adding artificial Poisson image noise to renderings;
- Light-ring position: tested by changing the height of the light ring over the dome;
- Amount of shadows: tested on a wave shape with different depths;
- Surface roughness and specularity.

Median values and standard deviations of angular errors per experiment are shown in Figure 6 for both CPS and UPS algorithms. We mark the best results with blue. We draw the following conclusion from these statistics:

- In general, CPS algorithms outperform UPS algorithms in all experiments as expected, which implies that the illumination conditions have the greatest effect on the PS algorithms;
- Among the CPS algorithms, our iterative method either outperforms or performs as well as the others;
- Among the UPS algorithms, LDR produces less error than other methods on average, while TV is more robust to variations in the isolated factors, producing less standard deviation in errors;
- Inter-reflections and shadows are the factors that have the strongest influence on the errors. These two factors are hard to separate in the experiments even though we tried to have a minimum amount of self-shadowing in inter-reflections experiment by adjusting the light positions accordingly; however it is unavoidable to get rid of the self-shadowing completely and vice versa;
- On the other hand, image noise and surface specularity have the least effect on errors in the PS algorithms;
- Increasing the number of images is only important up to a point for each algorithm, which varies between 5 and 10 images. See the supplementary material for the detailed graphs;
- The material type or the surface roughness, although tested on a limited subset, does not affect the normal estimation as much as shadows or inter-reflections.

We also include full statistics of the angular errors in the supplementary material as well as the full error graphs for each experiment. We also explain details on how we conducted the experiments.

Furthermore, we do error analysis over normal-z distribution and in frequency domains as follows:

Error Distribution over Normal-z Direction: In Figure 5, we show average errors in both normals and albedo on 3 datasets for various CPS and UPS algorithms mentioned above. The x-axis of the plot is the z-component of normal vectors, while the y-axis is the average angular error for portions of the surface having a normal with that z-component. As the z-component approaches 1, the surface approaches facing to the camera. The histogram of z-components is shown with the blue curve (legend on the right y-axis). The first row shows an image from the datasets, while the 2nd and 3rd rows are angular and albedo error curves for calibrated photometric stereo algorithms, respectively. The last two rows are the error plots for uncalibrated photometric stereo in the same fashion. Total average errors and the valid pixel ratios (some algorithms fail to estimate normals/albedo at some parts of the objects)

are indicated in the legends per algorithm from left to right. The conclusion of this test is that there is a high correlation between the error and the angle between the camera and the true surface normal.

Frequency Domain Analysis: We also analyze the frequency distribution of errors, and separate the low-frequency and high-frequency errors. Low-frequency and high-frequency error images for normal maps are calculated as follows:

$$D = abs(N_{gt} - N_e) \quad (1)$$

$$L = \text{cross\_bilateral\_filter}(D, N_{gt}) \quad (2)$$

$$H = abs(D - L) \quad (3)$$

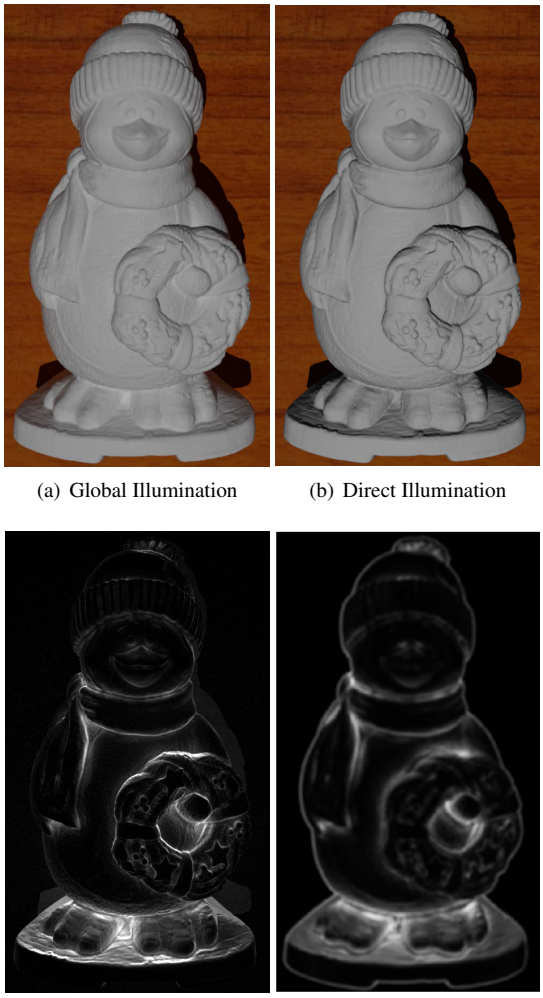
where N_{gt} is the ground truth normal map, N_e is the estimated normal map, and L and H are the low and high frequency components of the error respectively.

In Figure 4, the penguin is rendered with and without global illumination, shown on the top row. We observe that the difference of these two images resembles to the low-frequency component of the angular errors as shown in the bottom row. We do not show the high-frequency component of the angular error here because it is not as interesting, high-frequency errors concentrated around the silhouette of the objects. Given that the object has a diffuse material, the difference between global and direct illuminations is mostly caused by inter-reflections evidenced by the concavities being the areas with the biggest difference. Thus, we can conclude that the low-frequency errors are dominated by global illumination which is not accounted for by the algorithms.

7. Conclusion

This paper presents an RGBN benchmark including a number of synthetic datasets along with an interface which enables the user to create arbitrary scenes and a few photometric datasets of real-life objects. We will make all the code and the data available online with the hope that researchers will make use of this benchmark in the future. We think that a synthetic benchmark would be very useful to understand which algorithms, using photometric data, work better under which conditions.

Limitations and Drawbacks: Although we discussed that the synthetic datasets are comparable to the real ones, they are limited by the physically-based rendering system. Even though the today's rendering machines are very powerful and capable, they have limitations. For example, Mitsuba Renderer cannot simulate the wave properties of light, does not account for polarization, and the accuracy of any rendering system relies on the underlying floating point computations [21]. We also did not intend to create a benchmark with all the varieties but we would like to let it grow by accepting submissions from other users of our interface.



(a) Global Illumination (b) Direct Illumination

(c) Difference between a and b (d) CPS low frequncey error

Figure 4. **First row:** show the rendered images with and without global illumination. **(c)** is the difference image between (a) and (b). **(d)** is the low frequency component of the angular error of the normal map calculated with iterative least squares. (Difference images are scaled for better visualization.)

Future Work: We hope that our evaluations inspire improved photometric stereo algorithms that are robust to many types of effects. While our iterative least-squares photometric stereo algorithm provides a first step by exhibiting increased resistance to shadowing, both the calibrated and uncalibrated cases would benefit from targeted improvements based on the test cases we have developed as part of this benchmark.

References

[1] H. Aanæs, A. L. Dahl, and K. Steenstrup Pedersen. Interesting interest points. *Int. J. Comput. Vision*, 97(1):18–35, Mar. 2012. 2

[2] A. Abrams, C. Hawley, and R. Pless. Heliometric stereo:

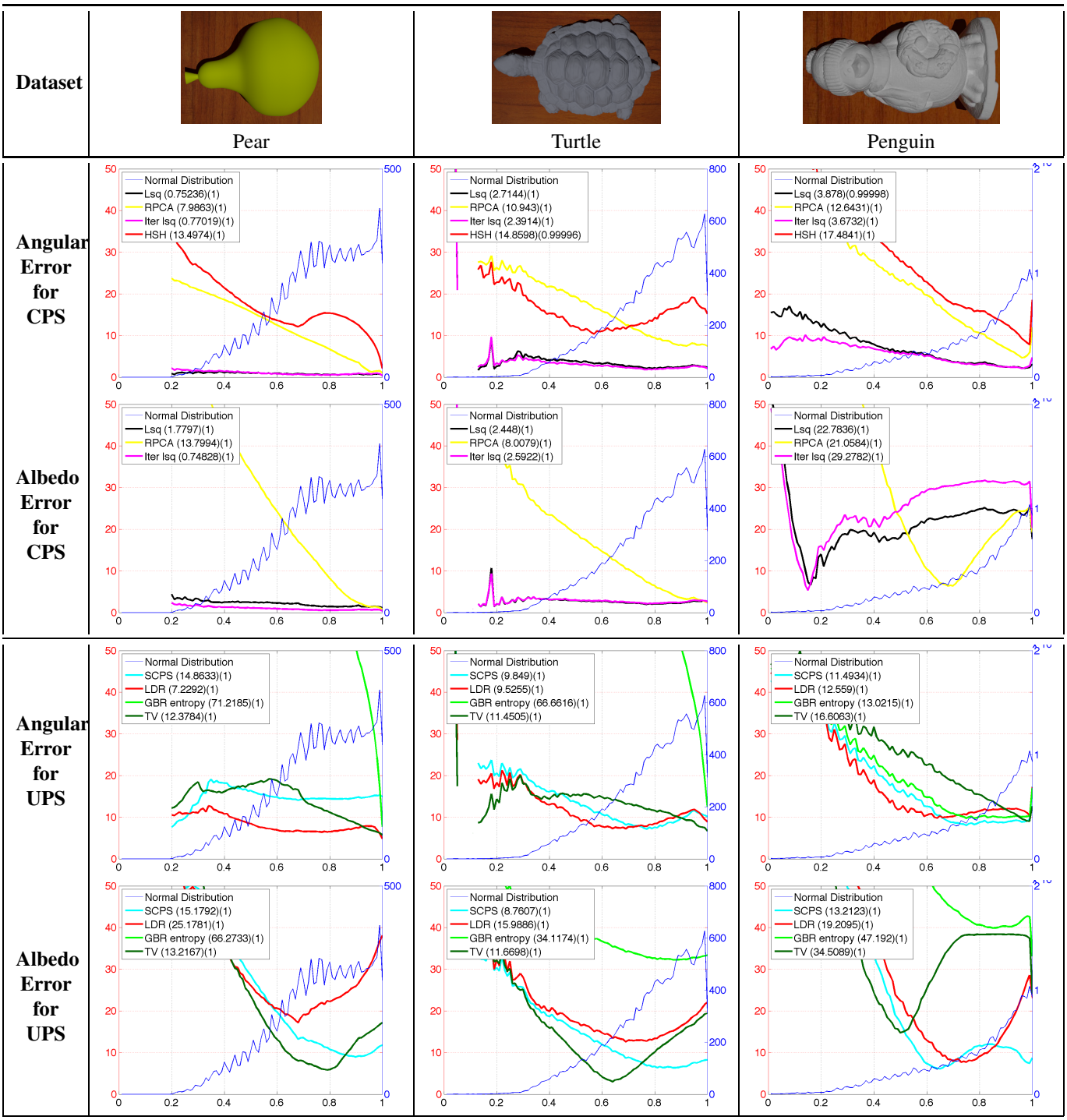


Figure 5. **Angular and albedo errors for selected objects (pear, turtle, penguin):** x-axis is the magnitude of the z component of normal vectors [0,1]; y-axis is the angular or albedo errors. The thin blue curve shows the histogram of the z- component of normal vectors in the ground truth normal map, and the second y-axis on the right side of each plot is the number of pixels for each histogram bin. In the legends, total average errors and ratio of the valid pixels are indicated from left to right respectively for each algorithm .

shape from sun position. In *Proceedings of the 12th European conference on Computer Vision-Volume Part II*, pages 357–370. Springer-Verlag, 2012. 4

[3] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele.

Photometric stereo for outdoor webcams. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 262–269. IEEE, 2012. 2, 5

[4] N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 2

[5] N. G. Alldrin, S. P. Mallick, and D. Kriegman. Resolving the Generalized Bas-Relief Ambiguity by Entropy Minimization. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–7. IEEE, 2007. 5

[6] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric Stereo with General, Unknown Lighting. *International journal of computer vision*, 72(3):239–257, June 2006. 5

[7] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003. 1, 5

[8] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Trans. Graph.*, 33(4):159:1–159:12, July 2014. 2

[9] M. Berger, J. A. Levine, L. G. Nonato, G. Taubin, and C. T. Silva. A benchmark for surface reconstruction. *ACM Transactions on Graphics (TOG)*, 32(2):20, 2013. 2

[10] S. Berkiten, X. Fan, and S. Rusinkiewicz. Merge2-3D: Combining multiple normal maps with 3D surfaces. *3DV 2014, International Conference on 3D Vision*, page 8, Dec. 2014. 4

[11] G. J. Brostow, C. Hernández, G. Vogiatzis, B. Stenger, and R. Cipolla. Video Normals from Colored Lights. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):2104–2114, Oct. 2011. 4

[12] M. Chandraker, S. Agarwal, and D. Kriegman. ShadowCuts: Photometric Stereo with Shadows. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8. IEEE, 2007. 5

[13] M. K. Chandraker, F. Kahl, and D. J. Kriegman. Reflections on the generalized bas-relief ambiguity. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 788–795. IEEE, 2005. 5

[14] X. Chen, A. Golovinskiy, and T. Funkhouser. A benchmark for 3d mesh segmentation. In *ACM Transactions on Graphics (TOG)*, volume 28, page 73. ACM, 2009. 2

[15] M. S. Drew, Y. Hel-Or, T. Malzbender, and N. Hajari. Robust estimation of surface properties and interpolation of shadow/specularity components. *Image and Vision Computing*, 30(4-5):317–331, May 2012. 5

[16] P. Favaro and T. Papadhimitri. A closed-form solution to uncalibrated photometric stereo via diffuse maxima. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 821–828. IEEE, 2012. 5

[17] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and Spatially-Varying BRDFs from Photometric Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1060–1071, June 2010. 1, 5

[18] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2335–2342. IEEE, 2009. 2

[19] C. Hernandez, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3):548–554, March 2008. 2

[20] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, Aug. 2005. 2, 5

[21] W. Jakob. Mitsuba renderer, 2010. http://www.mitsuba-renderer.org. 1, 2, 3, 7

[22] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, SGP '06, pages 61–70. Eurographics Association, 2006. 4

[23] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *EGSR'07: Proceedings of the 18th Eurographics conference on Rendering Techniques*. Eurographics Association, June 2007. 5

[24] T. Malzbender, T. Malzbender, D. Gelb, D. Gelb, H. Wolters, and H. Wolters. Polynomial texture maps. In *Computer Graphics, SIGGRAPH 2001 Proceedings*, 2001. 1, 4, 5

[25] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001. 2

[26] Y. Mukaigawa, Y. Ishii, and T. Shakunaga. Analysis of photometric factors based on photometric linearization. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 24(10):3326–3334, Oct. 2007. 5

[27] Y. Quéau, F. Lauze, and J.-D. Durou. Solving uncalibrated photometric stereo using total variation. *Journal of Mathematical Imaging and Vision*, pages 1–21, 2014. 5, 6

[28] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42, 2002. 2

[29] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1118–1125. IEEE, 2010. 5

[30] P. Tan, S. Lin, and L. Quan. Resolution-enhanced photometric stereo. In *Proceedings of the 9th European conference on Computer Vision-Volume Part III*, pages 58–71. Springer-Verlag, 2006. 5

[31] P. Tan, S. P. Mallick, L. Quan, D. Kriegman, and T. Zickler. Isotropy, Reciprocity and the Generalized Bas-Relief Ambiguity. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8. IEEE, 2007. 5

[32] C. Toler-Franklin, A. Finkelstein, and S. Rusinkiewicz. Illustration of complex real-world objects using images with normals. In *NPAR '07: Proceedings of the 5th international symposium on Non-photorealistic animation and rendering*, pages 111–119, New York, New York, USA, Aug. 2007. ACM Request Permissions. 2, 3, 4

[33] A. Wenger, A. Gardner, C. Tchou, J. Unger, T. Hawkins, and P. Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2005)*, 24(3):756–764, July 2005. 4

[34] R. J. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *22nd Annual Technical Symposium*, pages 136–143. International Society for Optics and Photonics, 1979. 1, 5

[35] C. Wu, Y. Liu, Q. Dai, and B. Wilburn. Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. *Visualization and Computer Graphics, IEEE Transactions on*, 17(8):1082–1095, 2011. 2

[36] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Computer Vision–ACCV 2010*, pages 703–717. Springer, 2011. 5

[37] Y. Xiong, A. Chakrabarti, R. Basri, S. J. Gortler, D. W. Jacobs, and T. Zickler. From Shading to Local Shape. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2014. 2

[38] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 158–164. IEEE, 1997. 5

[39] Z. Zhou and P. Tan. Ring-light photometric stereo. In *ECCV'10: Proceedings of the 11th European conference on Computer vision: Part II*. Springer-Verlag, Sept. 2010. 5