Photometric Segmentation: Simultaneous Photometric Stereo and Masking

Bjoern Haefner
TU Munich
Munich, Germany
bjoern.haefner@tum.de

Yvain Quéau
GREYC, UMR CNRS 6072
Caen, France
yvain.queau@ensicaen.fr

Daniel Cremers
TU Munich
Munich, Germany
cremers@tum.de

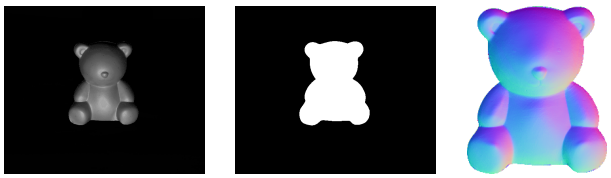
Abstract

This work is concerned with both the 3D-reconstruction of an object using photometric stereo, and its 2D-segmentation from the background. In contrast with previous works on photometric stereo which assume that a mask of the area of interest has been computed beforehand, we formulate 3D-reconstruction and 2D-segmentation as a joint problem. The proposed variational solution combines a differential formulation of photometric stereo with the classic Chan-Vese model for active contours. Given a set of photometric stereo images, this solution simultaneously infers a binary mask of the object of interest and a depth map representing its 3D-shape. Experiments on real-world datasets confirm the soundness of simultaneously solving both these classic computer vision problems, as the joint approach considerably simplifies the overall 3D-scanning process for the end-user.

1. Introduction

Photometric stereo [26] can be employed to estimate the 3D-shape of an object, given a set of images taken under the same viewing angle, but varying illumination. To this end, an image formation model describing the interactions between light and matter is inverted. Recent advances in the field have focused on relaxing several assumptions such as those of Lambertian reflectance and of calibrated lighting [23], in order to make the technique applicable in real-world scenarios and to simplify the 3D-scanning process.

However, all the approaches to photometric stereo which have been proposed so far assume that the area of interest is known in advance (see Figure 1). This means that the end-user is still required to perform a pre-segmentation of the object to reconstruct, before the 3D-reconstruction can be carried out. Such a pre-segmentation can be tedious and time-consuming: it would be way more convenient to achieve it automatically, while taking into account the information conveyed by the multi-light acquisition process of photometric stereo.



(a) (b) (c)

Figure 1. Given a set of input images such as (a), and a mask of the object to reconstruct (b), photometric stereo techniques infer 3D-geometry, represented in (c) under the form of surface normals. The present work aims at simplifying this process, by automatically achieving 2D-segmentation of the object in the same time as its 3D-reconstruction. That is to say, we aim at directly recovering the mask (b) and geometry (c) from the images (a).

In this work, we introduce a way to simultaneously achieve the 2D-segmentation of the object and its 3D-reconstruction, instead of first masking the object and then estimating its geometry. Building upon both the celebrated active contours model of Chan and Vese for two-region segmentation [2], and a recent PDE-based variational formulation of photometric stereo [20], we propose a joint variational approach to this problem. It comes down to estimating a minimal-length curve separating background from an area where the image formation model is satisfied and thus shape estimation is possible.

The proposed variational formulation for simultaneous 3D-reconstruction by photometric stereo and 2D-segmentation is detailed in Section 3, after discussing related variational approaches to photometric stereo and segmentation in Section 2. The resulting optimisation problem is numerically challenging, but recasting it as a level-set problem allows one to use classic convex optimisation techniques. Experimental results presented in Section 4 demonstrate the potential of this joint approach. Eventually, Section 5 concludes this study and suggests future research directions.

2. Variational Methods for Photometric Stereo and Segmentation

Assuming Lambertian reflectance with known directional lighting, neglecting shadows and assuming the object to reconstruct is pre-segmented, the classic formulation of photometric stereo [26] with m images consists in solving a set of equations such as

$$I_i(x) = \rho(x)\,\mathbf{n}(x)\cdot\mathbf{s}_i, \quad \forall x\in\Omega, i\in\{1,\dots,m\}, \quad (1)$$

with $\Omega\subset\mathbb{R}^2$ the mask of the object to reconstruct, $I_i : \Omega\to\mathbb{R}$ the i-th input graylevel image, ρ the reflectance (albedo) map, $\mathbf{n}$ the normal map (which encodes the 3D-geometry), and $\mathbf{s}_i\in\mathbb{R}^3$ a vector representing the incident lighting in the i-th image (in intensity and direction). Most of recent works on photometric stereo have focused on relaxing the assumptions of Lambertian reflectance (i.e., handling surfaces which exhibit a specular behavior) [4, 21, 11, 29] and calibrated directional lighting (i.e., handling unknown or non-uniform lighting) [5, 10, 13, 22], see for instance [23] for some discussion and [3] for a state-of-the-art joint solution to both issues using deep neural networks. However, in all of these recent works the object to reconstruct is assumed to be segmented a priori: the whole pipeline relies on the knowledge of the domain Ω.

In order to get rid of this assumption, we will make use of the recent variational formulation of photometric stereo exposed in [20], and also advocated in [6, 10, 12, 24]. Therein, 3D-reconstruction by photometric stereo is formulated as a variational problem aiming at directly reconstructing the underlying depth map $z:\Omega\to\mathbb{R}$, thus bypassing the need for normal estimation followed by normal integration. It is an optimisation-based approach of the form

$$\min_z \int_\Omega \mathcal{P}_{\text{PS}}(z(x))\,\mathrm{d}x, \quad (2)$$

with $\mathcal{P}_{\text{PS}}(z(x))$ a term evaluating the pixel-wise discrepency between the data and a differential formulation of the image formation model (1). Under orthographic projection, the normal is linked to the gradient $\nabla z:\Omega\to\mathbb{R}^2$ of the underlying depth map $z:\Omega\to\mathbb{R}$ according to $\mathbf{n}(x) = \frac{[\nabla z(x),-1]^\top}{\sqrt{|\nabla z(x)|^2+1}}$ [19], thus from a pair of equations such as (1), with $i\neq j$, one gets, for any $x\in\Omega$:

$$\frac{I_i(x)}{[\nabla z(x),-1]^\top\cdot\mathbf{s}_i} = \frac{\rho(x)}{\sqrt{|\nabla z(x)|^2+1}} = \frac{I_j(x)}{[\nabla z(x),-1]^\top\cdot\mathbf{s}_j}, \quad (3)$$

from which one can deduce:

$$\mathbf{a}_{ij}(x)\nabla z(x) = b_{ij}(x), \quad (4)$$

with $\mathbf{a}_{ij}(x) := \begin{bmatrix} I_i(x)s_j^1 - I_j(x)s_i^1 \\ I_i(x)s_j^2 - I_j(x)s_i^2 \end{bmatrix}^\top \in\mathbb{R}^{1\times 2}$, $b_{ij}(x) := I_i(x)s_j^3 - I_j(x)s_i^3\in\mathbb{R}$ and where $\mathbf{s}_i = \left[s_i^1, s_i^2, s_i^3\right]^\top$. This gives rise to $\binom{m}{2}$ different linear PDEs in z, which can be combined in a variational framework. Adding an arbitrary depth prior z_0 for numerical stability, depth estimation can then be formulated as (2), with

$$\mathcal{P}_{\text{PS}}(z(x)) := \frac{1}{\binom{m}{2}}\sum_{ij}\left(\mathbf{a}_{ij}(x)\nabla z(x) - b_{ij}(x)\right)^2 + \lambda(z(x)-z_0(x))^2, \quad (5)$$

where $\lambda>0$ is some hyper-parameter.

On the other hand, there is a large amount of literature on the image segmentation problem. Let us mention for instance early approaches based on region merging heuristics [17], active contours evolving towards edges in the images (aka snakes) [8], or recent deep learning frameworks [1]. Another class of methods is based on piecewise-smooth approximation of the input image [14], which comes down to image segmentation in the case of piewise-constant approximation. A classic example of such an approach is the Chan-Vese active contour model [2]. It aims at estimating a minimal-length curve C separating the image domain Ω between an area inside C where the graylevel image I is well-approximated by some value μ_1, and an area outside C where it is better-approximated by μ_2. This can be formulated as follows:

$$\min_{\mu_1,\mu_2,C}\int_{\text{inside}(C)}\mathcal{P}_1(\mu_1,I(x))\,\mathrm{d}x + \int_{\text{outside}(C)}\mathcal{P}_2(\mu_2,I(x))\,\mathrm{d}x + \nu\,\text{length}(C), \quad (6)$$

where $\nu\geq 0$ is a hyper-parameter controlling the length of the curve C, and $\mathcal{P}_j(\mu_j,I(x)) = (\mu_j - I(x))^2$, $j\in\{1,2\}$, such that the values μ_1,μ_2 resemble the mean intensity of the image I in the region inside and outside C, respectively.

Image segmentation and 3D-reconstruction may appear as two disconnected problems. Nevertheless, each task contributes information which may be interesting for the other, and the joint solving of these inverse problems has proven valuable, for instance, in the context of dense multi-view reconstruction [7], X-ray tomography [9], pose estimation [18], SLAM [25] or hyperspectral imaging [28]. Inspired by such joint approaches to simultaneous reconstruction and segmentation, in the rest of this work we revisit the photometric stereo problem in the case where no prior segmentation of the object has been performed i.e., domain Ω in (1) is unknown. This is detailed in the next section, where we propose a joint approach to photometric stereo and masking which combines the variational photometric stereo formulation (2) with the Chan-Vese variational segmentation approach (6).

3. Photometric Segmentation

The underlying assumption of the classic Chan-Vese segmentation model is that the brightness in the foreground largely differs from that of the background: (6) assumes that brightness in the foreground and background are well-approximated by two different constants μ_1 and μ_2. In the context of photometric stereo images, the objects are usually captured in the dark, thus it would make sense to assume that the background has a constant, low, brightness. Yet, the foreground may contain shadowed or low-albedo areas which might wrongly be classified as background.

The redundancy of information induced by the lighting variations contributes useful information to overcome shadowing issues. One could for instance apply a similar method as (6), but on the stack of m photometric stereo images. This can be simply achieved by setting $\mathcal{P}_j(\mu_j, I(x)) = \frac{1}{m}\sum_{i=1}^{m}(\mu_j - I_i(x))^2$ in (6). However, Figure 2 shows that this first approach remains unsatisfactory as it cannot distinguish between low albedo and background.

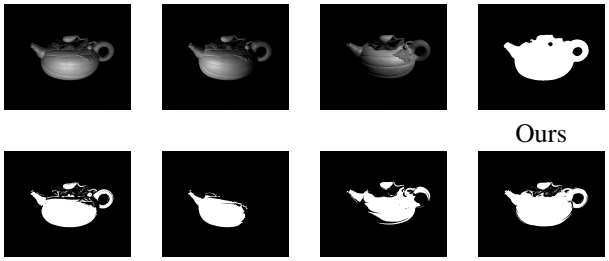


Figure 2. Top: three photometric stereo images (out of $m = 10$), and the segmentation obtained using the proposed joint reconstruction and segmentation method. Bottom: result of Chan-Vese segmentation applied to the single photometric stereo image shown above (CV-1), or to the whole set of 10 images (CV-10). It is difficult to segment a single image due to the ambiguity between background and shadows. Using multiple images improves results, but not as much as incorporating a photometric stereo model into segmentation, as we propose.

Instead of such a naive adaptation of the Chan-Vese model based on the average brightness, we suggest to drive segmentation by the image formation model. We define foreground as the set of pixels where the depth map z can be estimated from the differential photometric stereo model (4). On the contrary, we define background as the set of pixels where the photometric stereo model is not valid i.e., any uniform depth map z_0 can be set in (4). This way, the resulting depth map z will exhibit discontinuities along the separating curve C, which shall not fit to model (4). That is to say, it would "cost" more to wrongly include the boundaries of the object in the background or foreground, than to fit the curve separating foreground and background on the genuine object's boundaries.

In variational terms, this comes down to solving the following optimisation problem:

$$\begin{aligned}\min_{z,C} &\int_{\text{inside}(C)} \mathcal{P}_{\text{PS}}(z(x))\,\mathrm{d}x \\ &+ \int_{\text{outside}(C)} \mathcal{P}_{\text{PS}}(z_0(x))\,\mathrm{d}x \\ &+ \nu\,\text{length}(C),\end{aligned} \tag{7}$$

where $\mathcal{P}_{\text{PS}}$ is the differential photometric stereo fitting criterion defined in (5), $\nu \geq 0$ is a hyper-parameter controlling the length of the segmenting curve, and z_0 is an arbitrary depth prior which can be set using e.g., a prior on the camera-scene distance, or any arbitrary value if depth estimation up to an additive offset is acceptable (e.g., in our experiments we use $z_0 \equiv 1$).

Let us remark that the proposed variational model (7) differs from the Chan-Vese model (6) in two ways. First, the unknown depth z needs not being estimated in the background, which slightly simplifies the process. On the other hand the unknown depth is spatially-varying on the foreground, yet the estimation of such a spatially-varying quantity is made possible by the multiple image measurements under varying lighting.

Despite these differences, the optimisation problem (7) contains the same major difficulty as the origin problem of Chan and Vese: it involves both 2D (the depth map) and 1D (the curve) entities, which makes optimisation nontrivial. Besides, optimisation over the curve would require an appropriate parameterisation, which is known to yield nontrivial numerical issues such as setting the number of control points and uniformly sampling them over the curve. In the rest of this section we introduce an equivalent level-sets [16] formulation of the problem, which yields a simpler numerical solution. That is to say, we embed the problem in a higher-dimensional space where it is easier to solve numerically.

Let us define the curve C as the zero level-set of some function $\phi : \Omega \to \mathbb{R}$, such that $\phi \geq 0$ defines foreground and $\phi < 0$ defines background. Let us further denote by H the Heaviside step function ($H(x) = 1$ if $x \geq 0$ and $H(x) = 0$ elsewhere). Then, (7) is rewritten as follows:

$$\begin{aligned}\min_{z,\phi} &\int_{\Omega} H(\phi(x))\mathcal{P}_{\text{PS}}(z(x))\,\mathrm{d}x \\ &+ \int_{\Omega} (1 - H(\phi(x)))\mathcal{P}_{\text{PS}}(z_0(x))\,\mathrm{d}x \\ &+ \nu \int_{\Omega} |\nabla H(\phi(x))|\,\mathrm{d}x,\end{aligned} \tag{8}$$

where optimisation is now carried out over two real-valued 2D maps over Ω, and the segmenting curve C can be computed a posteriori from ϕ by thresholding.

To jointly solve the problems of photometric stereo (z-estimation) and segmentation (ϕ-estimation), we solve (8) alternatingly over each variable. At iteration (k), we solve:

$$z^{(k+1)} = \arg\min_{z} \int_{\Omega} H(\phi^{(k)}(x)) \mathcal{P}_{\text{PS}}(z(x))\, \mathrm{d}x, \tag{9}$$

$$\begin{aligned}\phi^{(k+1)} = \arg\min_{\phi} \quad & \int_{\Omega} H(\phi(x)) \mathcal{P}_{\text{PS}}(z^{(k+1)}(x))\, \mathrm{d}x \\ & + \int_{\Omega} (1 - H(\phi(x))) \mathcal{P}_{\text{PS}}(z_0(x))\, \mathrm{d}x \\ & + \nu \int_{\Omega} |\nabla H(\phi(x))|\, \mathrm{d}x. \end{aligned} \tag{10}$$

Problem (9) is a linear least squares problem, which can be solved using conjugate gradient iterations on the normal equations. Problem (10) is solved using gradient descent on the Euler-Lagrange equation

$$\delta(\phi(x)) \Bigg[\mathcal{P}_{\text{PS}}(z^{(k+1)}(x)) - \mathcal{P}_{\text{PS}}(z_0(x)) - \nu \operatorname{div}\left(\frac{\nabla \phi(x)}{|\nabla \phi(x)|} \right) \Bigg] = 0, \tag{11}$$

where $\delta(\phi(x))$ is a dirac delta, which can be considered as the derivative of the heaviside function $H(\phi(x))$.

Before the first iteration we initialise $z^{(0)}$ with the prior z_0, while the initial foreground is a circle of radius 10: $\phi^{(0)} = 10 - \sqrt{(x_1 - c_1)^2 + (x_2 - c_1)^2}$, where x_i are pixel coordinates and c_i corresponds to the center of the image, $i \in \{1, 2\}$. We stop iterations when the relative energy between two consecutive iterations falls under a threshold of 0.02, and we noticed that this was achieved in at most 20 iterations.

4. Experimental Validation

This section provides experimental results for the proposed variational photometric segmentation model. Real-world datasets of 10 photometric stereo images (see Figure 3) are extracted from a publicly available challenging photometric stereo benchmark [23] (images were preprocessed using [27] in order to fit the Lambertian assumption). The ten images are chosen such that the object is illuminated from every direction. All the experiments were conducted using Matlab on a standard laptop with 16GB of RAM and an Intel Core $i7$ with 2.2GHz. Convergence was always reached in at most 1 minute.

The impact of the hyper-parameter ν on the segmentation result will first be discussed. This is followed by a quantitative and qualitative evaluation of our segmentation results against other segmentation approaches. Eventually, we show that the estimated geometry of the scene is on average better, and en par compared to the result using no mask, and the ground-truth mask, respectively.

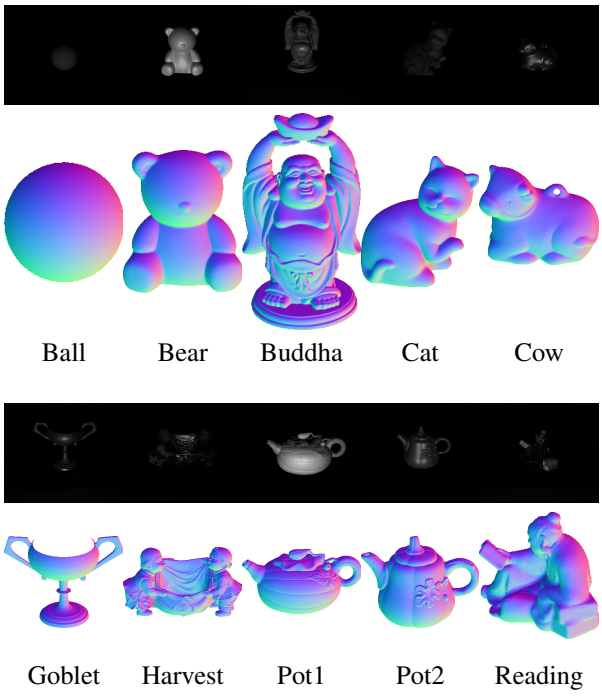
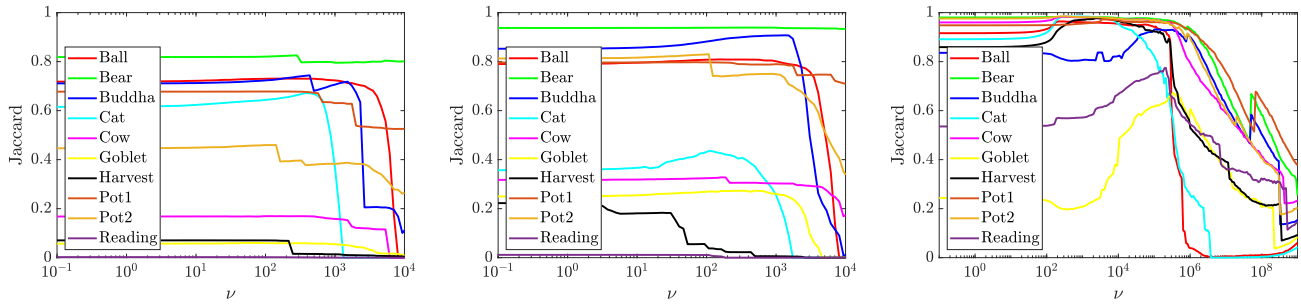


Figure 3. One out of ten grayscale images from [23] we used as input, along with the corresponding ground-truth normals.

4.1. Parameter Tuning

Our model comprises two tuning parameters λ and ν. In all our experiments we set λ to a very small value of 10^{-9}, since regularisation inside the mask is only intended to fix the translation ambiguity $z(x) := z(x) + \text{constant}$ in (4). Indeed, any small value of λ will solve this ambiguity and ensure convexity of the optimisation problem with respect to z, yet a high value of λ might bias the solution towards z_0. The tuning parameter ν is more crucial, hence we are going to evaluate it more thoroughly.

To this end we run our algorithm on the publicly available dataset [23] for values of $\nu \in [10^{-1}, 10^{9}]$ and evaluate the segmentation result using the Jaccard coefficient $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$, where A and B are two sets, in our case the ground truth mask and the estimated one. Our approach is compared against the classic Chan-Vese approach [2] presented in (6), which depends on the parameter ν as well. To show the impact of the choice of images in [2] we deploy two schemes. The first scheme uses a single image (randomly chosen from the data set) and we denote this approach with CV-1. The second scheme uses the same 10 images we use in the proposed method and we denote this approach with CV-10. The impact of the hyper-parameter ν can be evaluated in Figure 4. Not surprisingly, the Chan-Vese approach with ten images performs on average better than CV-1, as more data is used.



(a) CV-1 (b) CV-10 (c) Ours

Figure 4. Impact of the tuning parameter ν on the Jaccard coefficient (closer to one is better) for the classic Chan-Vese model (6) based on a single image (a) or multiple images (b), and for the proposed model (c). Our method systematically overcomes the Chan-Vese ones, provided that μ is not set too high in order to avoid over-segmentation.

Our model overcomes both Chan-Vese results throughout the evaluated range of ν, which shows that a tailored photometric segmentation cost function helps to find a better estimate of the mask of the object. In the rest of the evaluation we use the value of ν which provides the best results, for both Chan-Vese methods and for the proposed one.

4.2. Segmentation Accuracy

To quantitatively validate the accuracy of the estimated mask we compare the Jaccard coefficients against those obtained by three different segmentation approaches. The first one is GIMP's "Foreground Select Tool" which is based on statistical models of color variation [15] and can be considered as a standard way to generate a mask for an end-user of photometric stereo. It asks him to draw two scribbles in the image which provide the best statistical information in terms of color to separate background from foreground. The two Chan-Vese approaches (with a single or multiple images) already discussed in the previous paragraph are also considered, along with our approach with the best possible parameters for each dataset based on our evaluation in Figure 4. All quantitative and qualitative results can be seen in Table 1 and Figure 5. As already shown in Figure 4, the proposed approach overcomes both CV-1 and CV-10, which proves that the photometric term improves segmentation results. GIMP performs best on the Goblet dataset since it is able to separate the inner part of it, which no other method is able to do. Still, GIMP, CV-1 and CV-10 mainly suffer from oversegmentation, as they are not intended to distinguish background from shading. Only Harvest shows an undersegmented result, where GIMP considers too much background as foreground, due to too much dark shading variations in the object. Although the automated GIMP tool seems to overcome classic Chan-Vese segmentation, it can not keep up with our joint approach, which delivers the best segmentation results overall.

Dataset	GIMP	CV-1	CV-10	Proposed
Ball	0.6958	0.7307	0.8090	**0.9643**
Bear	0.8767	0.8254	0.9391	**0.9827**
Buddha	0.9112	0.7441	0.9074	**0.9320**
Cat	0.8567	0.6719	0.4352	**0.9842**
Cow	0.5536	0.1695	0.3277	**0.9829**
Goblet	**0.8706**	0.0601	0.2734	0.6727
Harvest	0.4830	0.0706	0.2227	**0.9773**
Pot1	0.7930	0.6781	0.7978	**0.9727**
Pot2	0.9145	0.4596	0.8305	**0.9851**
Reading	0.4970	0.0023	0.0114	**0.7748**

Table 1. Quantitative evaluation of the segmentation obtained using GIMP, CV-1, CV-10 and the proposed method, based on the Jaccard coefficient. The proposed approach overcomes the others in most cases.

4.3. Normal Reconstruction Accuracy

We also question whether an accurate segmentation may improve the quality of the estimated geometry. The best possible estimate is obtained using photometric stereo with ground truth mask i.e., solving (8) with fixed ϕ. Hence we consider the latter estimate as our "ground truth" geometry (as Figures 3 and 6 illustrate, such a baseline geometry may deviate from the real ground truth one, yet using the latter would bias the evaluation). For quantitative evaluation we calculate the mean angular error (MAE) in degrees between the baseline normals (resulting from the ground truth mask) and the estimated ones. As estimated normals we consider two approaches. The first one is an approach without any mask, that is every pixel is considered as valuable data point and used during optimisation. The second one is the proposed photometric segmentation approach. To make comparison fair, for both approaches we only evaluate MAE in the area corresponding to the intersection between the mask estimated by our approach and the ground truth one. Results can be seen in Table 2 and Figure 6.

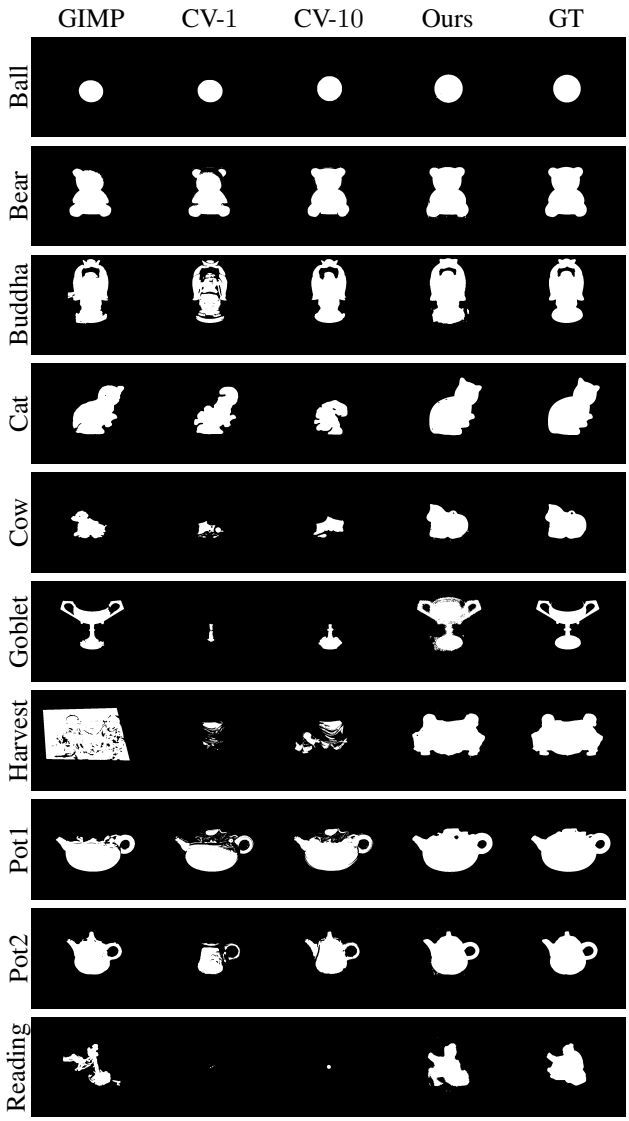


Figure 5. Qualitative comparison between the segmentation obtained using GIMP, CV-1, CV-10 and the proposed method which jointly estimates mask and geometry. In all the examples except Goblet, the estimated mask is the most accurate one.

These results show that geometry estimation indeed benefits from a joint 3D-reconstruction and segmentation approach: in most datasets our approach deviates much less from the baseline normals, compared to the approach without mask. Only the two data sets Harvest and Reading appear to perform better with no mask, but the loss in accuracy (0.02° and 0.15°, respectively) can be considered negligible. We believe that this gain comes from the fact that when using no mask, geometry is smoothed at the boundaries of the object, while when using a mask (or, when automatically finding this mask, as we propose) much sharper geometry can be recoverd near the boundaries.

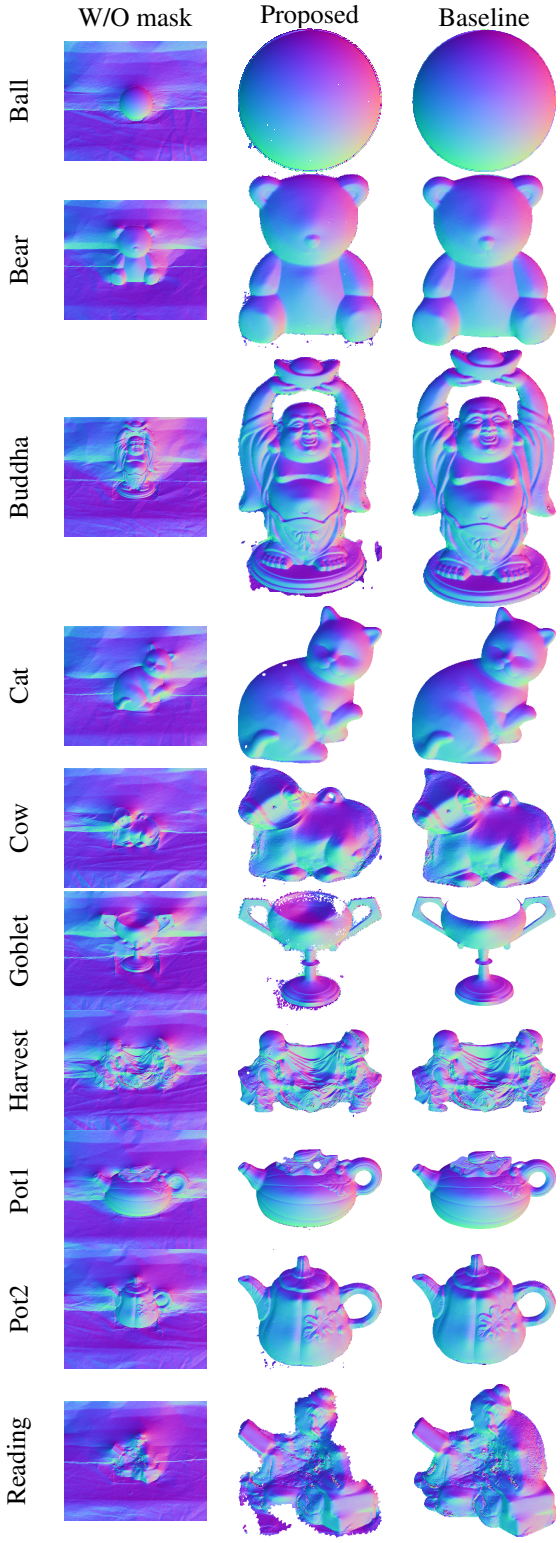


Figure 6. Qualititative results of the normal estimates using no mask and the proposed approach which jointly estimates mask and geometry. Estimated masks can be seen in Figure 5.

Dataset	W/O mask	Proposed
Ball	0.9290	**0.7522**
Bear	0.6211	**0.2974**
Buddha	0.7791	**0.5370**
Cat	0.2068	**0.0868**
Cow	0.9644	**0.6592**
Goblet	6.8144	**6.4709**
Harvest	**0.6204**	0.6816
Pot1	1.7623	**1.5196**
Pot2	0.8353	**0.3747**
Reading	**9.0507**	9.2291

Table 2. Comparison of the mean angular error (in degrees) on the estimated normals with respect to the baseline. The proposed approach slightly improves the geometry estimate in most cases.

Indeed, Figure 7 shows that the improvement becomes apparent at the boundaries of the estimates, where our approach has less error and larger deviation from the approach with no mask is visible. Especially in the case of the Buddha, where a large error appears in the hole of the arms and head, our approach has much less error, as it is able to detect this region as background. Only the results of Goblet and Reading largely deviate from the baseline normals inside the object. The difficulty with Goblet is the discontinuity in the upper part, which our approach is not able to detect. This results in geometry estimates across the discontinuity, inducing smoothing which deteriorates the overall geometry inside the object. Reading itself is very dark compared to the other objects, cf Figure 3. This causes the mask estimate to suffer from undersegmentation of the object.

5. Conclusion

We presented a joint variational approach to photometric stereo and segmentation. To the best of our knowledge this is the first methodology providing a scheme which is able to perform photometric stereo without the need of a mask. The proposed approach simplifies the photometric stereo process from the end-user perspective, by circumventing the need for tedious masking of the object and providing an end-to-end framework for object 3D-reconstruction, as shown in Figure 8. Experiments conducted on real-world benchmarks provided empirical evidence for the superiority of model-driven segmentation over naive segmentation based on brightness. Still, the proposed alternating optimisation strategy could be accelerated a lot by relying on parallel computing: in the future this will enable real-time results which will ease the setting of the hyper-parameter controlling the length of the boundary curve.

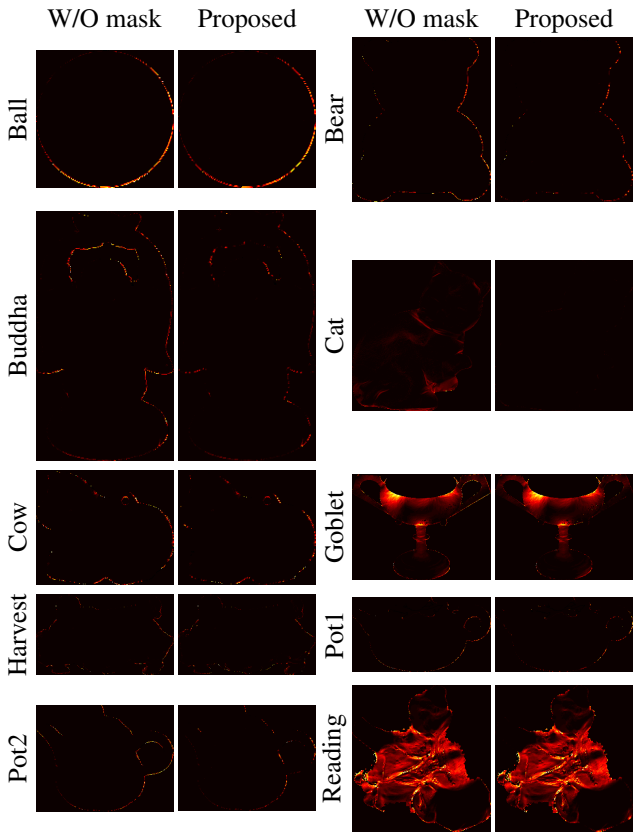
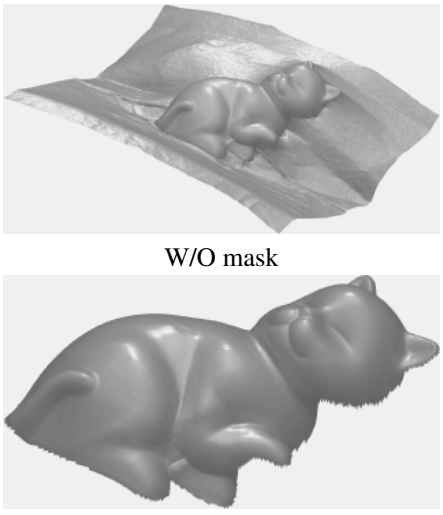


Figure 7. Spatial distribution of the angular error with respect to the baseline. Joint reconstruction and segmentation slightly reduces errors near the boundaries of the objects, since geometry estimate does not propagate over possibly discontinuous regions. Best seen in color on the Buddha and Cat examples.



W/O mask

Proposed

Figure 8. 3D-reconstruction of the Cat without masking, or using the proposed method. The latter makes possible an end-to-end 3D-scanning pipeline of objects.

References

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017. 2

[2] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001. 1, 2, 4

[3] G. Chen, K. Han, B. Shi, Y. Matsushita, and K.-Y. K. Wong. Self-calibrating Deep Photometric Stereo Networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[4] K. H. M. Cheng and A. Kumar. Revisiting Outlier Rejection Approach for Non-Lambertian Photometric Stereo. *IEEE Transactions on Image Processing*, 28(3):1544–1555, 2019. 2

[5] D. H. Cho, Y. Matsushita, Y. W. Tai, and I. S. Kweon. Semi-calibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2

[6] P. F. U. Gotardo, T. Simon, Y. Sheikh, and I. Matthews. Photogeometric scene flow for high-detail dynamic 3d reconstruction. In *IEEE International Conference on Computer Vision (ICCV)*, pages 846–854, 2015. 2

[7] C. Hane, C. Zach, A. Cohen, R. Angst, and M. Pollefeys. Joint 3D Scene Reconstruction and Class Segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013. 2

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988. 2

[9] F. Lauze, Y. Quéau, and E. Plenge. Simultaneous reconstruction and segmentation of CT scans with shadowed data. In *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 308–319, 2017. 2

[10] F. Logothetis, R. Mecca, and R. Cipolla. Semi-calibrated near field photometric stereo. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 941–950, 2017. 2

[11] F. Lu, X. Chen, I. Sato, and Y. Sato. SymPS: BRDF symmetry guided photometric stereo for shape and light source estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):221–234, 2017. 2

[12] R. Mecca, E. Rodolà, and D. Cremers. Realistic photometric stereo using partial differential irradiance equation ratios. *Computers & Graphics*, 51:8–16, 2015. 2

[13] Z. Mo, B. Shi, F. Lu, S.-K. Yeung, and Y. Matsushita. Uncalibrated photometric stereo under natural illumination. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2936–2945, 2018. 2

[14] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989. 2

[15] C. Nieuwenhuis and D. Cremers. Spatially varying color distributions for interactive multi-label segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1234–1247, 2013. 5

[16] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988. 3

[17] W. A. Perkins. Area segmentation of images using edge points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(1):8–15, 1980. 2

[18] V. A. Prisacariu, A. V. Segal, and I. Reid. Simultaneous monocular 2d segmentation, 3d pose recovery and 3d reconstruction. In *Asian Conference on Computer Vision*, pages 593–606, 2012. 2

[19] Y. Quéau, J.-D. Durou, and J.-F. Aujol. Normal integration: a survey. *Journal of Mathematical Imaging and Vision*, 60(4):576–593, 2018. 2

[20] Y. Quéau, R. Mecca, and J.-D. Durou. Unbiased photometric stereo for colored surfaces: A variational approach. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4359–4368, 2016. 1, 2

[21] Y. Quéau, T. Wu, F. Lauze, J.-D. Durou, and D. Cremers. A Non-Convex Variational Approach to Photometric Stereo under Inaccurate Lighting. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 350–359, 2017. 2

[22] S. Sengupta, H. Zhou, W. Forkel, R. Basri, T. Goldstein, and D. Jacobs. Solving uncalibrated photometric stereo using fewer images by jointly optimizing low-rank matrix completion and integrability. *Journal of Mathematical Imaging and Vision*, 60(4):563–575, 2018. 2

[23] B. Shi, Z. Mo, Z. Wu, D. Duan, S. Yeung, and P. Tan. A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):271–284, 2019. 1, 2, 4

[24] W. Smith and F. Fang. Height from photometric ratio with model-based light source selection. *Computer Vision and Image Understanding*, 145:128–138, 2016. 2

[25] K. Tateno, F. Tombari, and N. Navab. When 2.5 D is not enough: Simultaneous reconstruction, segmentation and recognition on dense SLAM. In *The IEEE International Conference on Robotics and Automation (ICRA)*, pages 2295–2302, 2016. 2

[26] R. J. Woodham. Photometric Method for Determining Surface Orientation from Multiple Images. *Optical Engineering*, 19(1):139–144, 1980. 1, 2

[27] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Asian Conference on Computer Vision (ACCV)*, pages 703–717, 2010. 4

[28] Q. Zhang, R. Plemmons, D. Kittle, D. Brady, and S. Prasad. Joint segmentation and reconstruction of hyperspectral data with compressed measurements. *Applied Optics*, 50(22):4417–4435, 2011. 2

[29] Q. Zheng, B. Kumar, A. Shi, and G. Pan. Numerical Reflectance Compensation for Non-Lambertian Photometric Stereo. *IEEE Transactions on Image Processing*, 28(7):3177–3191, 2019. 2