

Solving Uncalibrated Photometric Stereo using Total Variation

Yvain Quéau · François Lauze · Jean-Denis Durou

the date of receipt and acceptance should be inserted later

Abstract Estimating the shape and appearance of an object, given one or several images, is still an open and challenging research problem called 3D-reconstruction. Among the different techniques available, photometric stereo produces highly accurate results when the lighting conditions have been identified. When these conditions are unknown, the problem becomes the so-called uncalibrated photometric stereo problem, which is ill-posed. In this paper, we will show how total variation (TV) can be used to reduce the ambiguities of uncalibrated photometric stereo, and we will study two methods for estimating the parameters of the *generalized bas-relief* ambiguity. These methods will be evaluated through the 3D-reconstruction of real-world objects.

Keywords 3D-reconstruction, photometric stereo, total variation, generalized bas-relief ambiguity.

1 Introduction

Photometric stereo was introduced by Woodham in [47] in the early 80's, using the commonly adopted Lambertian model to recover both the surface shape and its albedo, given $m \geq 3$ pictures of a fixed scene taken from the same viewpoint but under different calibrated illumination conditions. However, calibrating the light

sources is not a trivial task, and it has been shown that slight errors in this calibration could result in major reconstruction errors [30]. Thus, uncalibrated photometric stereo seems to be a promising alternative, as the light vectors are estimated during the reconstruction. Unfortunately, this problem is ill-posed. If integrability is imposed, the ambiguity is reduced to a set of 3-parameter linear transformations called *generalized bas-relief* transformations [7].

The estimation of these parameters is the subject of this paper: we extend the preliminary results presented in [41] and show how TV-regularization of the M field, which encodes the albedo and the normals, or of the depth function u , considerably reduces the ambiguities: the *generalized bas-relief* ambiguity is reduced to the classical *bas-relief* ambiguity, which is easier to solve. We demonstrate that the proposed methods can be written as the minimization of simple convex functions, and are thus much faster than most previous work involving either pixel identification [19, 44] or non-convex minimization [3]. Despite the absence of statistical justification on the distributions of the albedo, the depth and the normal field, the results we obtain *a posteriori* justify the choice of using total variation.

The rest of the paper is organized as follows: subsequently to the recall of the equations of Lambertian photometric stereo in Section 2, we empirically justify in Section 3 the choice of using TV for solving the uncalibrated photometric stereo problem. In Section 4, we prove that minimizing the total variation of the M field or that of the depth u reduces the *generalized bas-relief* ambiguity to a simple *bas-relief* ambiguity. The methods are eventually evaluated in Section 5¹.

Y. Quéau
IRIT, UMR CNRS 5505, Toulouse, France
E-mail: yvain.queau@enseeiht.fr

F. Lauze
DIKU, Copenhagen, Denmark
E-mail: francois@diku.dk

J.-D. Durou
IRIT, UMR CNRS 5505, Toulouse, France
E-mail: durou@irit.fr

¹ The Matlab codes corresponding to the proposed methods are available online at:
<http://ubee.enseeiht.fr/photometricstereo/>

2 Background

2.1 Calibrated Photometric Stereo

Photometric Stereo (PS) was introduced in [47] to deal with 3D-reconstruction, allowing the user to recover both the normal field associated to a surface and the albedo. In this technique, m images of a scene are taken from the same viewpoint but under variable lighting conditions. An example of such images, with $m = 3$, is shown in Fig. 1.



Fig. 1: $m = 3$ images, of size 256×256 , of a Beethoven's bust, taken under 3 different illuminations².

PS is an extension of the shape-from-shading problem [29], which is known for being ill-posed [37]. The use of additional images with different lighting conditions allows to solving the ambiguities of shape-from-shading. Instead of a unique known light source, supposed to be located at infinite or finite distance (a recent shape-from-shading considered arbitrary location [21]), dealing with several known sources allows obtaining unambiguous, accurate and fast 3D-reconstructions. As the normal vectors and the albedo values are estimated at every pixel, *dense* depth and albedo maps are recovered, unlike using stereoscopy. To obtain a full 3D-model rather than a depth map, PS can be coupled with multi-view techniques, through the fusion of depth maps [13] or the combination with, e.g., structure-from-motion [50] or shape-from-silhouettes [26].

Assuming orthographic projection (the problem was also studied in the perspective case [36]), the surface to be reconstructed is represented by a Monge patch $(x, y, u(x, y))^T$ over a domain $\Omega \subset \mathbb{R}^2$, where the depth function $u(x, y)$ is assumed to be twice differentiable, at least piecewise. The vector field of unit outward normals is given by

$$N(x, y) = \frac{1}{\sqrt{\|\nabla u\|^2 + 1}} \begin{pmatrix} -\partial_x u \\ -\partial_y u \\ 1 \end{pmatrix} \quad (1)$$

Let $\rho(x, y) \in [0, 1]$ be the albedo map, and $M(x, y) = \rho(x, y)N(x, y)$, so that $\rho(x, y) = \|M(x, y)\|$. Given a

directional light source $S^i = (S_x^i, S_y^i, S_z^i)^T \in \mathbb{R}^3$, the graylevel $I^i(x, y)$ obtained at point (x, y) of the image plane for matte objects is given by Lambert's law

$$I^i(x, y) = \rho(x, y)N(x, y)^T S^i = M(x, y)^T S^i \quad (2)$$

(shadows are ignored).

In order to recover M (and thus ρ and N) without ambiguity, at least 3 images associated with 3 non-coplanar light vectors are mandatory. In the sequel the $m \geq 3$ images will be denoted I^1, \dots, I^m , and the corresponding light vectors S^1, \dots, S^m . It will also be assumed that $\forall(x, y), \rho(x, y) > 0$. Defining the observation vector $I(x, y) = (I^1(x, y), \dots, I^m(x, y))$, the light matrix $\mathbf{S} = (S^1, \dots, S^m)$, and omitting the dependencies in (x, y) , the following equality holds:

$$I = M^T \mathbf{S} \quad (3)$$

This is a system of linear equations which can be solved in a least-squares sense using the Moore-Penrose pseudo-inverse \mathbf{S}^+ of \mathbf{S} :

$$\widehat{M}^T = I\mathbf{S}^+, \quad \widehat{\rho} = \|\widehat{M}\|, \quad \widehat{N} = \frac{\widehat{M}}{\widehat{\rho}} \quad (4)$$

(all through this paper, a hat above a letter will indicate that this is an estimation).

The normal vectors being unitary, they have only two degrees of freedom, so the problem is often equivalently formulated as the estimation of ρ and $\nabla u = (p, q)^T$, with

$$\begin{cases} p = -\frac{N_x}{N_z} \\ q = -\frac{N_y}{N_z} \end{cases} \quad (5)$$

The first line of Fig. 2 shows the albedo and the depth gradient estimated from the images of Fig. 1.

Although direct depth estimation may be carried out [36], an additional step is usually needed to “integrate” the normal field into a depth map (see Fig. 2-d). This step, which can be problematic, will not be presented in this paper. It usually results in estimating the depth u as the least-squares solution of the equation $\nabla u = (\widehat{p}, \widehat{q})^T$ [15, 20, 24, 28, 45]. All the depth maps presented in this paper are obtained using the same solver [45].

2.2 About the Assumptions

The usual assumption about the number of images (*i.e.* the number of light sources) is that $m \geq 3$ and the light vectors should be non-coplanar. Some attempts have been made to solve the problem with $m = 2$ [32, 35, 38],

² <http://www.ece.ncsu.edu/imaging/Archives/ImageDatabase/>

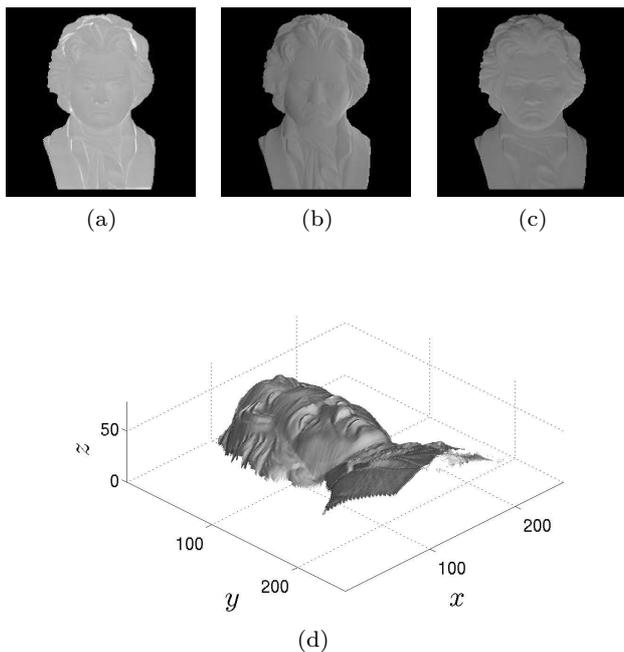


Fig. 2: Results of applying photometric stereo to the three images of Fig. 1: (a) $\hat{\rho}$; (b) \hat{p} ; (c) \hat{q} . (d) Relighting of the reconstructed surface. White pixels in (a)-(b)-(c) indicate high values.

and an interesting recent application of the coplanar configuration is 3D-reconstruction from an outdoor webcam (since the Sun moves within a plane) [1, 2]. In the present paper the traditional case of $m \geq 3$ non-coplanar light vectors is studied.

In addition, although the $m > 3$ case improves the robustness of PS thanks to the least-squares estimation (4), a lot of physical phenomena like specularities, self-shadows, cast-shadows or inter-reflections are not represented in Lambert's law (2): such phenomena can be treated as outliers in a robust estimation process like Expectation-Maximization [46]. Specular reflections have also been successfully handled by adapting the photometric stereo technique to non-purely diffuse models such as that of Torrance and Sparrow [22], considering that the luminance is obtained by summing a diffuse and a specular component.

The problem of self-shadows, specularities, and more generally arbitrary reflectances can also be avoided by using a reference object [27] whose geometry and reflectance are known. The case of cast-shadows is more complicated, since they are impossible to locally model, however they can be located and rejected from the estimation process [6]. This has been done efficiently using the popular Graph Cuts technique [12].

Another approach to deal with outliers is to preprocess the data: under the Lambertian assumption, any set of n pixels of the observed images forms a $n \times m$ matrix which should be of rank 3. However, this is not the case when dealing with real images, because of noise, quantization and outliers. Thus, Wu et al. propose in [48] an algorithm to correct the input, via low-rank approximation. It should also be noted that specularities can be removed from RGB images by an appropriate change of colour space [34]. In this work, the outliers are ignored, i.e. it is assumed that we are dealing either with synthetic data or with preprocessed real data.

2.3 Uncalibrated Photometric Stereo

In Section 4, we will focus on the problem with $m \geq 3$ unknown non-coplanar distant light sources, which is the so-called *uncalibrated photometric stereo* problem. In this case, it has been shown [17, 25] that the vector field M defined as the surface normal multiplied by the albedo can only be determined up to a 3×3 linear transformation $\mathbf{A} \in GL(3)$, since

$$M^\top \mathbf{S} = (\mathbf{A}^\top M)^\top \mathbf{A}^{-1} \mathbf{S} \quad (6)$$

However, since the depth function $u(x, y)$ is assumed to be twice differentiable, Schwarz' integrability constraint $\partial^2 u / \partial_x \partial_y = \partial^2 u / \partial_y \partial_x$ should be satisfied. This condition is equivalent to equating to zero the third component of $\text{curl}(N/N_z) = \text{curl}(M/M_z)$, denoted by $\overline{\text{curl}} M$:

$$\overline{\text{curl}} M = \frac{\partial}{\partial y} \left(\frac{M_x}{M_z} \right) - \frac{\partial}{\partial x} \left(\frac{M_y}{M_z} \right) = 0 \quad (7)$$

Imposing $\overline{\text{curl}} M = 0$ as a hard constraint is not possible, as Eq. (7) does not hold for points located at discontinuities, which are present in real-world objects. Thus, Yuille and Snow proposed in [49] to solve Eq. (7) in a least-squares sense, while respecting Lambert's law (3) (still in a least-squares sense). We reformulate this method as the estimation of a field M and a light matrix \mathbf{S} holding both constraints:

$$\begin{cases} \|M^\top \mathbf{S} - I\|^2 = \min \\ \|\overline{\text{curl}} M\|^2 = \min \end{cases} \quad (8)$$

where the notation \min signifies that the parameters minimize the left hand of the equation.

Adapting Hayakawa's method [25], Yuille and Snow give in [49] a closed-form solution holding these constraints, which will be referred to as (M^0, \mathbf{S}^0) in the following.

Not all transformations $\mathbf{A} \in GL(3)$ preserve the integrability of the field $\mathbf{A}^\top M$: it has been shown in [7, 49] that imposing the integrability constraint reduces the set of transformations \mathbf{A} to the set of 3-parameters transformations $\mathbf{A} = \mathbf{G}$ with \mathbf{G} of the form

$$\mathbf{G}(\mu, \nu, \lambda) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{pmatrix}, \quad \mathbf{G}(\mu, \nu, \lambda)^{-1} = \frac{1}{\lambda} \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ -\mu & -\nu & 1 \end{pmatrix} \quad (9)$$

where $(\mu, \nu, \lambda) \in \mathbb{R}^2 \times \mathbb{R} \setminus \{0\}$. This corresponds to the *generalized bas-relief* (GBR) ambiguity [7]. As the sign of λ is directly linked to the convexity of the object, and as the concave/convex ambiguity cannot be solved without additional information on the object, it is usually assumed that $\lambda > 0$.

Thus, problem (8) is still an ill-posed problem, as one can apply any GBR transformation without either changing the value of the data term (this comes from (6)), or breaking the integrability condition, since it is shown in [7, 49] that

$$\|\overline{\text{curl}} M\|^2 = \min \iff \|\overline{\text{curl}}(\mathbf{G}^\top M)\|^2 = \min \quad (10)$$

In terms of depth, the GBR transforms $u(x, y)$ into

$$\bar{u}(x, y) = \frac{u(x, y) - \mu x - \nu y}{\lambda} \quad (11)$$

Fig. 3 shows the effects of a GBR.

Research on the resolution of this GBR ambiguity is prolific. It can be assumed, quite reasonably, that real-world objects do not strictly follow the Lambertian model, so as to exploit the presence of outliers like specularities [14] or inter-reflections [12]. Local properties of the shape or the reflectance can also be used: for example, critical points in an image (the brightest pixels) can reveal the direction of the light vector [19], and intensity profiles allow identifying pixels which have the same normal but different albedos [44]. We argue that considering global properties is more general, and might be easier to exploit as it does not rely on the identification of critical points.

Quite surprisingly, very few attempts to exploit global properties have been made, apart from the use of prior on shapes [23] or on the albedo distribution: Alldrin *et al.* propose in [3] to estimate the GBR parameters which minimize the entropy of the albedo distribution, since the albedo is spread by the GBR. Their motivation is to favour materials which are ‘‘homogeneous’’, *i.e.* made up of a small value of components. Consistently with our formulation (8), we rewrite their method

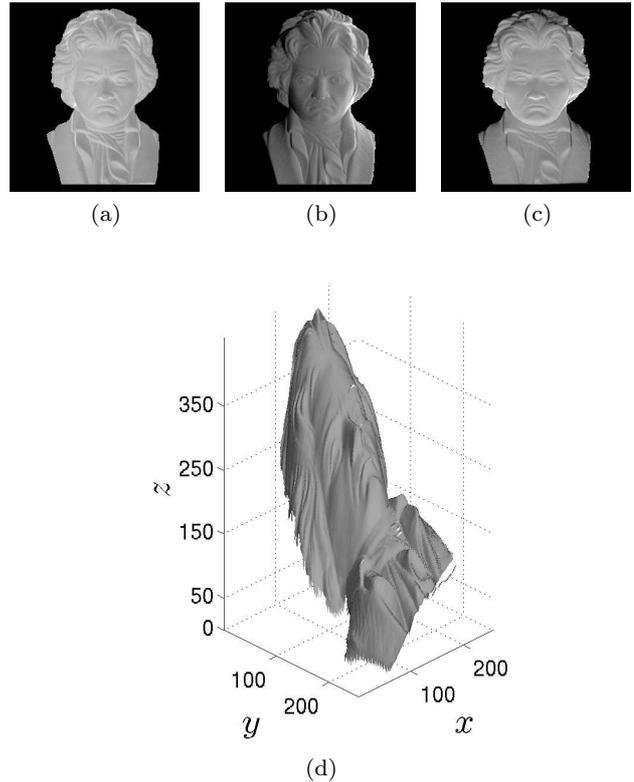


Fig. 3: Effects of a GBR of parameters $\mu = 1/2$, $\nu = 1$ and $\lambda = 1/3$ on: (a) the albedo, (b)-(c) the normals and (d) the shape of Fig. 2. The scales are the same as in Fig. 2, so as to illustrate the drifts in all the three directions. It is to be noted that the albedo (a) is spread, compared to that of Fig.2-a: this led to the Minimum-Entropy algorithm [3].

as the following pair of problems, which must be solved sequentially:

$$\begin{cases} \left\{ \begin{array}{l} \|M^{0\top} \mathbf{S}^0 - I\|^2 = \min \\ \|\overline{\text{curl}} M^0\|^2 = \min \end{array} \right. \\ (\hat{\mu}, \hat{\nu}, \hat{\lambda}) = \underset{\mu, \nu, \lambda}{\text{argmin}} \mathcal{E}(\|\mathbf{G}(\mu, \nu, \lambda)^\top M^0\|) \end{cases} \quad (12)$$

where $\mathcal{E}(\|\mathbf{G}^\top(\mu, \nu, \lambda) M^0\|)$ stands for the entropy of the transformed albedo distribution.

2.4 TV-Regularized Models in Computer Vision

The total variation of a function is a widely used measure for regularity. For an almost everywhere differentiable function $f : \Omega \subset \mathbb{R}^p \rightarrow \mathbb{R}$, it can be written

$$\text{TV}(f) = \int_{\Omega} \|\nabla f(x)\| dx \quad (13)$$

and extends to the class of so-called functions of bounded variations [5].

In the sequel we take $p = 2$ as we deal with images. Total variation was introduced to the imaging community by the work of Rudin, Osher and Fatemi (ROF) [43] for image denoising: given a (noisy) graylevel image $I_0 = I + \eta$, with η a Gaussian white noise of known variance σ^2 , the ROF model finds an image \hat{I} which solves $\|\hat{I} - I_0\|^2 \leq \sigma^2$ on Ω , and with minimal total variation $TV(\hat{I})$. The problem was shown in [10] to be equivalent to the minimization of the functional

$$E(I) = \|I - I_0\|^2 + \gamma_\sigma TV(I)$$

for a certain weight γ_σ . Efficient algorithms have been developed in the last decade for solving such problems, using for instance primal-dual schemes [9].

In this paper, since both the total variation of the depth u and that of the $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ vector field M will be considered, vectorial TV should also be defined. When f takes its values in \mathbb{R}^q with $q > 1$, several definitions can be considered (see [8] for some discussion). The most frequently used definition is probably

$$TV(f) = \int_{\Omega} \|J(f(x))\|_F dx \quad (14)$$

where $\|J(f(x))\|_F$ is the Frobenius norm of the Jacobian matrix of $f = (f_1 \dots f_q)^\top$ at point x . Another definition is:

$$TV(f) = \sum_{i=1}^q \int_{\Omega} \|\nabla f_i(x)\| dx = \sum_{i=1}^q TV(f_i) \quad (15)$$

The difference between both these definitions may be important in the denoising context, as the first introduces a coupling between the different channels, and is invariant by rotation of these channels. However, in our context, we are not concerned by this difference, as the ordering of the three components M_x , M_y and M_z is important for the integrability constraint (7), and as each parameter of a GBR affects a single component of M . Therefore, we will consider the definition (15), which is easier to manipulate and produces better results. In this paper, unlike the preliminary version [41], the total variation of the M field will thus refer to

$$TV(M) = TV(M_x) + TV(M_y) + TV(M_z) \quad (16)$$

In Section 4, we will show that minimizing the total variation of either the M field or the depth function u reduces the ambiguities of uncalibrated photometric stereo. As we will see, in this context, the minimization of TV is quite simple as it depends only on the GBR parameters. The resulting problems being convex, standard convex optimization methods such as quasi-Newton ones can be used: this offers considerably faster methods, compared to state-of-the-art approaches.

3 An Empirical Evidence for using TV

Let us now introduce our motivations for using TV in the context of uncalibrated photometric stereo. As it will be illustrated on some examples, a GBR affects the variations of both the albedo ρ , the depth u and its gradient ∇u , which is linked to the normals by (1). Moreover, the optimal values of μ and ν seem to correspond to a minimum of the total variation.

Firstly, let us state that this work is inspired by the method of Alldrin *et al.*, who show in [3] that the GBR parameters could be estimated by minimizing the entropy of the albedo distribution. However, the entropy being anything but convex, the minimization has to be global: Alldrin *et al.* use a brute-force discrete search on a user-defined interval, which results in a very long optimization process. Furthermore, this entropy does not consider the spatial variation of the albedo, as shown in Fig. 4.

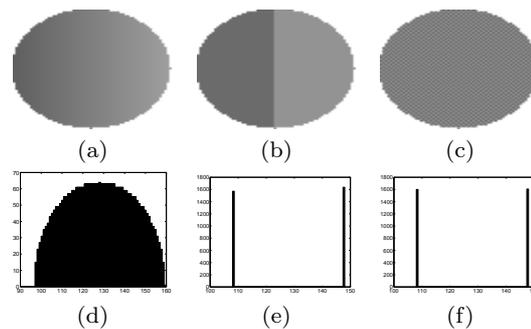


Fig. 4: Three different albedo configurations and the corresponding histograms: (a) has a significant entropy value, cf. histogram (d); (b) and (c) have the same slight entropy, cf. histograms (e) and (f), however the spatial distributions of the albedo are different. Total variation would tend to favour distributions such as (b).

When looking for “homogeneous” zones, one would expect that similar albedo pixels would be close to each other. This property can be obtained by using the total variation of the albedo, as one would expect that locally, the variations of albedo should be slight. Also, total variation having better differential properties than entropy, we expect considerably shorter computation times since convex optimization tools can be used instead of brute-force algorithms.

Since the GBR applies not only to the albedo but to the whole M field, not only the variations of the albedo ρ are affected but also those of the depth u and of its first order derivatives $\nabla u = (p, q)^\top$. We show in Fig. 5 the effect of a GBR on these variations.

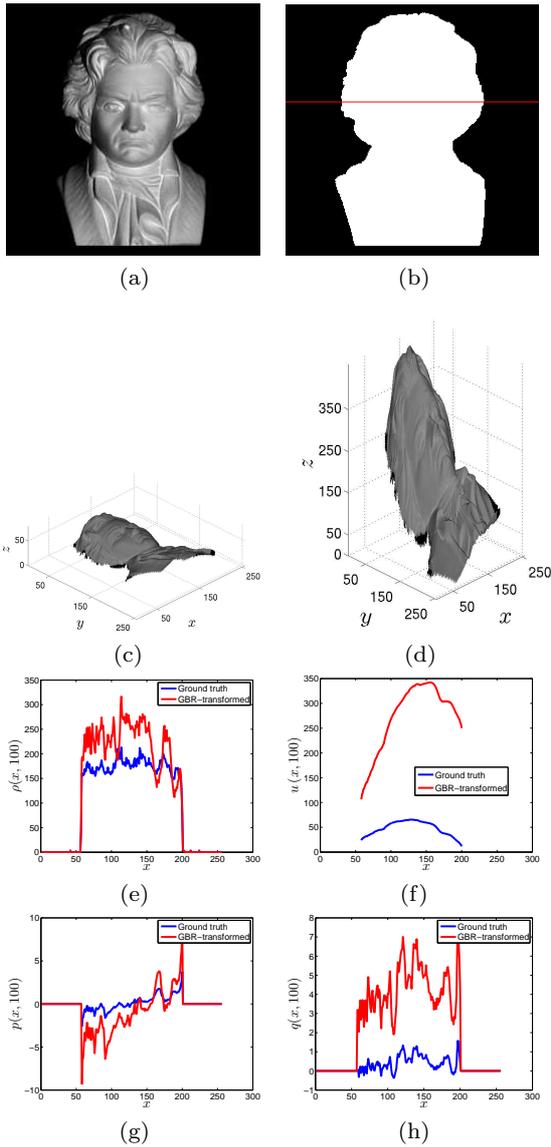


Fig. 5: The effects of a GBR transformation on the variations of the albedo, of the depth and of its derivatives. (a) One of the input images. (b) The $y = 100$ straight line. (c) Ground truth 3D-model. (d) GBR-transformed 3D-model. We plotted the values of the ground truth and the GBR-transformed values of (e) ρ , (f) u , (g) p and (h) q along the $y = 100$ straight line, with $\mu = 1/2$, $\nu = 1$ and $\lambda = 1/3$. Obviously, the GBR-transformed estimations have higher total variations.

Thus, to solve this GBR ambiguity, it seems reasonable to consider not only ρ , but also u and ∇u . As ρ , ∇u and M are linked by

$$M = \frac{\rho}{\sqrt{\|\nabla u\|^2 + 1}} \begin{pmatrix} -\nabla u \\ 1 \end{pmatrix} \quad (17)$$

we can also directly consider the M field.

In Fig. 6, we report the impact of μ and ν on the total variations of ρ , u and M estimated from the images of Beethoven's bust. To do so, we consider as the ground truth the solution of the calibrated photometric stereo problem and apply GBR transformations to this ground truth, with λ fixed to 1: the ground truth values thus correspond to $(\mu, \nu) = (0, 0)$. To independently study the impact of each parameter, one test per parameter is executed while fixing the other parameters to the ground truth value.

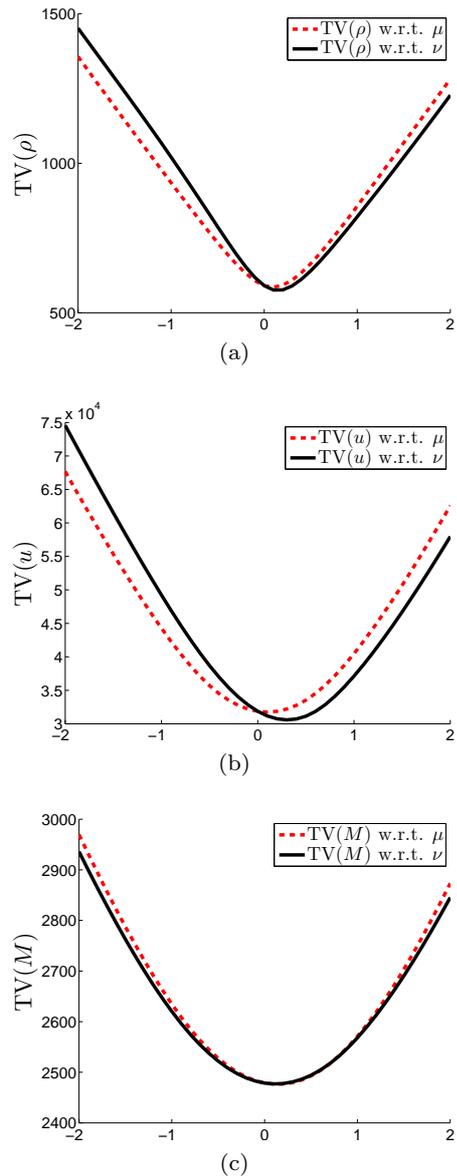


Fig. 6: The effects of μ and ν on the total variations of (a) the albedo ρ , (b) the depth function u and (c) the M field. The three total variations seem to be minimal for values of μ and ν close to the ground truth values.

This test tends to indicate that minimizing any of these total variations may solve the problem in μ and ν . The case of λ is different: the effect of λ on these TV, with μ and ν fixed to the ground truth values, is illustrated in Fig. 7. $\text{TV}(u)$ is minimized for $\lambda = +\infty$ and $\text{TV}(M)$ for $\lambda = 0$ (these observations are proven in Section 4), however these values correspond to degenerated depth maps (see Eq. (11)).

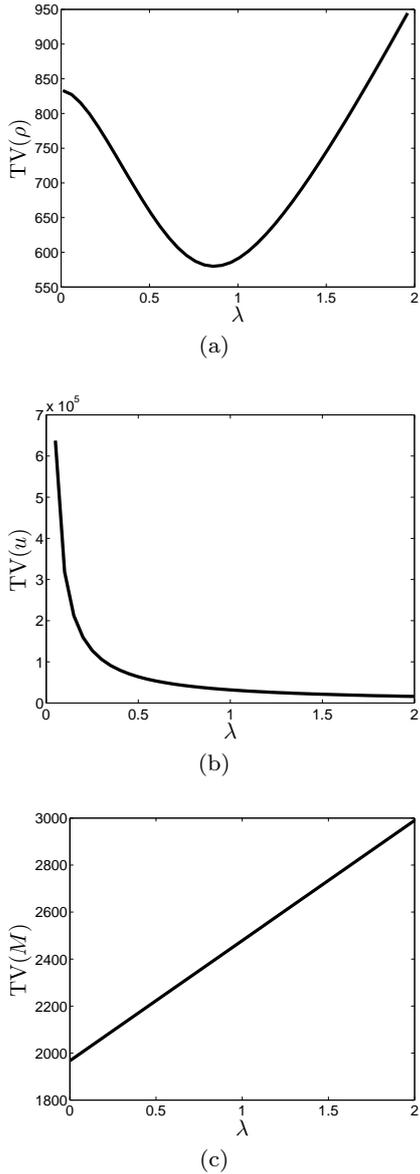


Fig. 7: The effects of λ on the total variations of (a) the albedo ρ , (b) the depth function u and (c) the M field. Only the total variation of the albedo has a minimum inside $]0; +\infty[$, whereas $\text{TV}(u)$ is inversely proportional to λ and $\text{TV}(M)$ is proportional to λ . Both these last observations are proven in Section 4.

Thus, only $\text{TV}(\rho)$ seems to have a non-degenerated minimum. However, in general, estimating λ by minimizing $\text{TV}(\rho)$ gives quite disappointing results: in fact the previous test is biased by the obviously uniform albedo of Beethoven's bust (which is made of plaster). Using the dataset Doll from [3], we obtain the plot of Fig. 8 for $\text{TV}(\rho)$ as a function of λ , with μ and ν fixed to the ground truth values.

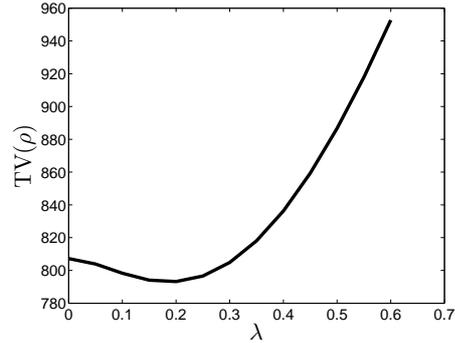


Fig. 8: The total variation of the albedo ρ , as a function of λ , on the Doll dataset. The value of λ minimizing $\text{TV}(\rho)$ is quite far away from the ground truth value 1.

Thus, to estimate λ , other methods should be considered. However, if μ and ν are already estimated, we will see in Section 4.3 that this estimation is equivalent to the resolution of the classical *bas-relief* ambiguity, which is considerably easier to solve than the GBR. As a consequence, very simple solutions do exist. For instance, we will show that the constant light magnitude constraint [25, 49] can be written as a simple linear system. Of course, any method of resolution of the GBR [3, 19, 44] would also be simplified by the knowledge of μ and ν (see for instance the TV- u / ME method in Section 5). Because of the link (17) between ρ , u and M , we restrict ourselves in this paper to the study of $\text{TV}(M)$ and $\text{TV}(u)$. Moreover, the following proposition shows that minimizing $\text{TV}(M)$ is linked (although not equivalent) to simultaneously minimizing $\text{TV}(\rho)$ and $\text{TV}(\nabla u)$:

Proposition 1 $\text{TV}(M) \leq 3\text{TV}(\rho) + 4\text{TV}(\nabla u)$

Proof Denoting $\nabla u = (p, q)^\top$ and $r = \|\nabla u\|$, we have

$$M_x = \frac{-\rho p}{\sqrt{r^2 + 1}}, \quad M_y = \frac{-\rho q}{\sqrt{r^2 + 1}}, \quad M_z = \frac{\rho}{\sqrt{r^2 + 1}} \quad (18)$$

The gradient of M_x is given by

$$\nabla M_x = -\frac{\rho \nabla p}{\sqrt{r^2 + 1}} - \frac{p \nabla \rho}{\sqrt{r^2 + 1}} + \frac{\rho p r \nabla r}{(r^2 + 1)^{3/2}} \quad (19)$$

$$(20)$$

Since $\frac{1}{\sqrt{r^2+1}} \leq 1$, $\frac{p}{\sqrt{r^2+1}} \leq 1$ and $\frac{r}{r^2+1} \leq 1$, we get

$$\|\nabla M_x\| \leq \rho \|\nabla p\| + \|\nabla \rho\| + \rho \|\nabla r\| \quad (21)$$

We obtain in the same way

$$\|\nabla M_y\| \leq \rho \|\nabla q\| + \|\nabla \rho\| + \rho \|\nabla r\| \quad (22)$$

The gradient of M_z is given by

$$\nabla M_z = \frac{\nabla \rho}{\sqrt{r^2+1}} - \frac{\rho r \nabla r}{(r^2+1)^{3/2}} \quad (23)$$

Thus we have

$$\|\nabla M_z\| \leq \|\nabla \rho\| + \rho \|\nabla r\| \quad (24)$$

Summing (21), (22) and (24) gives

$$\sum_{c=x,y,z} \|\nabla M_c\| \leq 3\|\nabla \rho\| + 3\rho \|\nabla r\| + \rho(\|\nabla p\| + \|\nabla q\|) \quad (25)$$

Finally, since $\nabla r = \frac{p\nabla p + q\nabla q}{r}$ and $\max(|p|, |q|) \leq r$,

$$\|\nabla r\| \leq \|\nabla p\| + \|\nabla q\| \quad (26)$$

and thus

$$\sum_{c=x,y,z} \|\nabla M_c\| \leq 3\|\nabla \rho\| + 4\rho(\|\nabla p\| + \|\nabla q\|) \quad (27)$$

Integrating this inequality over the reconstruction domain Ω leads to:

$$\text{TV}(M) \leq 3 \int_{\Omega} \|\nabla \rho\| dx + 4 \int_{\Omega} \rho(\|\nabla p\| + \|\nabla q\|) dx \quad (28)$$

Since $\rho \in [0, 1]$, using the definition (13) of $\text{TV}(\rho)$ and the definition (15) of $\text{TV}(\nabla u)$ leads to

$$\text{TV}(M) \leq 3 \text{TV}(\rho) + 4 \text{TV}(\nabla u) \quad (29)$$

which is the announced result. \square

Remark 1 In Eq. (28), $\int_{\Omega} \rho(\|\nabla p\| + \|\nabla q\|) dx$ is the ρ -weighted total variation of ∇u and when minimizing it, it allows for “relaxing” the minimization of $\text{TV}(\nabla u)$ where the albedo is low, i.e. where the material is dark. Since it is obvious that an albedo equal to zero induces ill-posedness, this “relaxation” allows us not to consider areas which would induce errors in the reconstruction.

Finally, let us state that, though the proposed limited empirical evidence should be completed by a statistical study of the involved distributions, the results shown in Section 5 will *a posteriori* justify the TV approach, as the accuracy of the estimations of μ and ν reach state-of-the-art results, and the computation times of the proposed methods are considerably shorter than those of other approaches, thanks to the convexity of the objective functions that are discussed in the next section.

4 Solving the GBR using TV

In this section, we show how TV-regularization can solve the uncalibrated photometric stereo problem. Our aim is to estimate a solution $(\widehat{M}, \widehat{\mathbf{S}})$ of (8) having minimal total variation. Given a solution (M^0, \mathbf{S}^0) of (8), this problem can be defined as the estimation of the GBR parameters $\widehat{\mu}$, $\widehat{\nu}$ and $\widehat{\lambda}$: the M field is eventually given by $\widehat{M} = \mathbf{G}(\widehat{\mu}, \widehat{\nu}, \widehat{\lambda})^\top M^0$.

4.1 The Proposed Model

In order to minimize the TV of the M field, different approaches can be considered. The first, inspired by denoising models, would be to replace (8) by:

$$\begin{cases} (\widehat{M}, \widehat{\mathbf{S}}) = \underset{M, \mathbf{S}}{\text{argmin}} \|M^\top \mathbf{S} - I\|^2 + \gamma \text{TV}(M) \\ \text{s.t.} \quad \|\overline{\text{curl}} M\|^2 = \min \end{cases} \quad (30)$$

where $\gamma > 0$ is a user-defined hyper-parameter. However, this is not the best solution: such a model would be useful in order to smooth some artifacts due to the presence of outliers, however, as already stated, in this paper we are not concerned about such outliers as we assume the data is preprocessed. Therefore, a weight between the data term $\|M^\top \mathbf{S} - I\|^2$ and the regularization term $\text{TV}(M)$ is not required, since we would like to keep the data term minimal³.

In the context of the resolution of the GBR ambiguity, we propose to solve the following problem, which seems more relevant:

$$\begin{cases} \widehat{M} = \underset{M}{\text{argmin}} \text{TV}(M) \\ \text{s.t.} \quad \begin{cases} \|M^\top \mathbf{S} - I\|^2 = \min \\ \|\overline{\text{curl}} M\|^2 = \min \end{cases} \end{cases} \quad (31)$$

To satisfy the first constraint in (31), orthogonal projection of a local least-squares estimation on the space of curl-free vector field might be considered, as such a projection can be achieved efficiently using, for instance, Fourier transform [20, 45]. Unfortunately, since the “curl” considered here (defined by Eq. (7)) is nonlinear, such an approach is not as simple to put in place as it may seem. Another possibility would be to introduce two Lagrange multipliers associated with the constraints. As the second constraint is nonlinear and as there are several unknowns, proximal methods [4, 11] could probably be used. Yet, this would require an iterative scheme, which would result in high computation times.

³ However, as mentioned in the perspectives, the model (30) could be useful to post-process the results.

Moreover, we did not use the fact that we do know a closed form solution of (8), and the analytical form of all acceptable transformations (the GBR matrices). We propose a simple solution to (31), based on the solution of Yuille and Snow [49] and the estimation of the GBR parameters. We use the following proposition:

Proposition 2 *Problem (31) is strictly equivalent to*

$$(\widehat{M}, \widehat{\mathbf{S}}) = \left(\mathbf{G}(\widehat{\mu}, \widehat{\nu}, \widehat{\lambda})^\top M^0, \mathbf{G}(\widehat{\mu}, \widehat{\nu}, \widehat{\lambda})^{-1} \mathbf{S}^0 \right) \quad (32)$$

where (M^0, \mathbf{S}^0) is the solution of Yuille and Snow, and

$$\begin{cases} (\widehat{\mu}, \widehat{\nu}, \widehat{\lambda}) = \underset{\mu, \nu, \lambda}{\operatorname{argmin}} \operatorname{TV}(\mathbf{G}(\mu, \nu, \lambda)^\top M^0) \\ \text{s.t. } \lambda > 0 \end{cases} \quad (33)$$

Proof Let $(\widehat{M}, \widehat{\mathbf{S}})$ be given by (32) and (33). A method for estimating $\widehat{\mu}$, $\widehat{\nu}$ and $\widehat{\lambda}$ is proposed in the following section. All the GBR matrices being invertible, we have $\|M^0{}^\top \mathbf{S}^0 - I\|^2 = \|(\mathbf{G}(\widehat{\mu}, \widehat{\nu}, \widehat{\lambda})^\top M^0)^\top (\mathbf{G}(\widehat{\mu}, \widehat{\nu}, \widehat{\lambda})^{-1} \mathbf{S}^0) - I\|^2$. As, by construction, $\|M^0{}^\top \mathbf{S}^0 - I\|^2 = \min$, it follows that $\|\widehat{M}^\top \widehat{\mathbf{S}} - I\|^2 = \min$. By construction also, M^0 is such that $\|\operatorname{curl} M^0\|^2 = \min$. From (10), we deduce $\|\operatorname{curl} \widehat{M}\|^2 = \min$. Both constraints are thus respected, thanks to the properties of the GBR. Finally, the GBR parameters being chosen so as to minimize $\operatorname{TV}(\mathbf{G}(\mu, \nu, \lambda)^\top M^0)$, \widehat{M} is a solution of (31).

Now, let \widehat{M} be a solution of (31), and $\widehat{\mathbf{S}}$ the associated light matrix. $(\widehat{M}, \widehat{\mathbf{S}})$ satisfies the two constraints, as does (M^0, \mathbf{S}^0) . As the only transformations which preserve these constraints are the GBR [7, 49], there necessarily exists a GBR matrix $\mathbf{G}(\widehat{\mu}, \widehat{\nu}, \widehat{\lambda})$ such that $\widehat{M} = \mathbf{G}(\widehat{\mu}, \widehat{\nu}, \widehat{\lambda})^\top M^0$ and $\widehat{\mathbf{S}} = \mathbf{G}(\widehat{\mu}, \widehat{\nu}, \widehat{\lambda})^{-1} \mathbf{S}^0$, i.e. $(\widehat{M}, \widehat{\mathbf{S}})$ can be expressed under the form (32). \square

Let us now explain how to solve (33).

4.2 Estimating μ and ν

Firstly, for computational issues, we propose to replace the constraint $\lambda > 0$ by $\lambda \geq \epsilon$, with $\epsilon > 0$. We will discuss the influence of ϵ in Section 4.3, and propose solutions for choosing it (in the preliminary results [41], we manually chose ϵ).

Now, remark that, from our definition (15) of TV and the form of a GBR, the objective function reads:

$$\begin{aligned} E(\mu, \nu, \lambda) &= \operatorname{TV}(\mathbf{G}(\mu, \nu, \lambda)^\top M^0) \\ &= \operatorname{TV}(g_1(\mu)M^0) + \operatorname{TV}(g_2(\nu)M^0) \\ &\quad + \operatorname{TV}(g_3(\lambda)M^0) \end{aligned} \quad (34)$$

where $g_1(\mu) = (1, 0, \mu)$, $g_2(\nu) = (0, 1, \nu)$ and $g_3(\lambda) = (0, 0, \lambda)$ denote the three lines of $\mathbf{G}(\mu, \nu, \lambda)^\top$.

Writing $\operatorname{TV}(g_3(\lambda)M^0) = \int_{\Omega} \|\nabla(\lambda M_z^0)\| dx$, one obtains $\partial_{\lambda} E = \lambda \operatorname{TV}(M_z^0)$, and thus $\operatorname{TV}(M)$ is linear in λ (cf. Fig. 7-c), and the solution in λ is given by:

$$\widehat{\lambda} = \epsilon \quad (35)$$

This explains why we replaced the constraint $\lambda > 0$ by $\lambda \geq \epsilon$, as $\lambda = 0$ would provide a degenerated depth function $\bar{u}(x, y)$ (see (11)). Thus, the choice of ϵ is important for the estimation of λ , but we can state the following:

Proposition 3 *The estimations of μ and ν are totally independent from that of λ , thus from ϵ .*

Proof Denoting by $E(\mu, \nu, \lambda) = \operatorname{TV}(\mathbf{G}(\mu, \nu, \lambda)^\top M^0) = \operatorname{TV}(M_x^0 + \mu M_z^0) + \operatorname{TV}(M_y^0 + \nu M_z^0) + \operatorname{TV}(\lambda M_z^0)$, its derivatives in μ and ν read:

$$\begin{cases} \partial_{\mu} E = \int_{\Omega} \frac{\mu \|\nabla M_z^0\|^2 + (\nabla M_x^0)^\top (\nabla M_z^0)}{\|\nabla(M_x^0 + \mu M_z^0)\|} dx \\ \partial_{\nu} E = \int_{\Omega} \frac{\nu \|\nabla M_z^0\|^2 + (\nabla M_y^0)^\top (\nabla M_z^0)}{\|\nabla(M_y^0 + \nu M_z^0)\|} dx \end{cases} \quad (36)$$

Since λ does not appear in Eq. (36), this proves the proposition. It should be stated that this comes from the choice of the definition (15) of TV: λ would remain in the equations if the definition (14) were used. \square

Remark 2 Differentiating Eq. (36) w.r.t. μ and ν proves the convexity of $\operatorname{TV}(M)$ w.r.t. μ and ν .

As will be shown in Section 4.3, the choice of ϵ actually corresponds to the resolution of the classical *bas-relief* ambiguity, which can be solved quite simply. Hence, the proposed TV formulation reduces the GBR to the *bas-relief* ambiguity. As already stated by Belhumeur *et al.* in [7], this is consistent with the work of Fan and Wolff [18], who show that the Hessian of u can only be recovered up to a scaling factor (which here corresponds to λ). In their work, two other parameters are required (the boundary conditions of a differential equation): these parameters correspond to the remaining parameters μ and ν .

At this point, let us focus on the estimations of μ and ν . They involve the minimization of a convex function which depends only on two parameters μ and ν . Given the expression (36) of the gradient ∇E , we solve for μ and ν in problem (33) using a quasi-Newton method. In our implementation, we use the `fminunc` Matlab function. We can accelerate the optimization by choosing an adequate initialization, such as the solution of the Tikhonov regularization, given by

$$\begin{cases} \widehat{\mu}_0 = \underset{\mu}{\operatorname{argmin}} \int_{\Omega} \|\nabla(M_x^0 + \mu M_z^0)\|^2 dx \\ \widehat{\nu}_0 = \underset{\nu}{\operatorname{argmin}} \int_{\Omega} \|\nabla(M_y^0 + \nu M_z^0)\|^2 dx \end{cases} \quad (37)$$

which admits the closed-form solution

$$\begin{cases} \hat{\mu}_0 = -\frac{\int_{\Omega} (\nabla M_x^0)^\top (\nabla M_z^0) dx}{\int_{\Omega} \|\nabla M_z^0\|^2 dx} \\ \hat{\nu}_0 = -\frac{\int_{\Omega} (\nabla M_y^0)^\top (\nabla M_z^0) dx}{\int_{\Omega} \|\nabla M_z^0\|^2 dx} \end{cases} \quad (38)$$

4.3 Estimating λ

According to (35), the solution in λ of problem (33) is $\hat{\lambda} = \epsilon$. Similarly to a hyper-parameter, ϵ can be manually tuned by the user so as to obtain visually acceptable shapes. However, instead of manually choosing this parameter, we can use additional information, which is greatly facilitated by the following observation:

Corollary 1 *The estimation of λ can be performed after estimating μ and ν .*

Proof This is a direct consequence of Prop. 3. \square

This corollary justifies that μ and ν can be estimated before fixing ϵ : we can replace (33) by the following pair of problems, which must be solved sequentially:

$$\begin{cases} (\hat{\mu}, \hat{\nu}) = \underset{\mu, \nu}{\operatorname{argmin}} \operatorname{TV}(\mathbf{G}(\mu, \nu, 1)^\top M^0) \\ \left\{ \begin{array}{l} \hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \operatorname{TV}(\mathbf{G}(\hat{\mu}, \hat{\nu}, \lambda)^\top M^0) \\ \text{s.t. } \lambda \geq \epsilon \end{array} \right. \end{cases} \quad (39)$$

where ϵ can then be chosen after estimating μ and ν . It can be stated that:

Lemma 1 *The estimation of λ corresponds to the resolution of the classical bas-relief ambiguity.*

Proof As we have

$$\mathbf{G}(\hat{\mu}, \hat{\nu}, \lambda)^\top M^0 = \mathbf{G}(0, 0, \lambda)^\top (\mathbf{G}(\hat{\mu}, \hat{\nu}, 1)^\top M^0) \quad (40)$$

we are now dealing with GBR matrices $\mathbf{G}(0, 0, \lambda)$ which will transform the field $M^1 = \mathbf{G}(\hat{\mu}, \hat{\nu}, 1)^\top M^0$: this particular GBR ambiguity with $\mu = \nu = 0$ corresponds to a *bas-relief* ambiguity on the field M^1 [7]. \square

The influence of λ is clear: $\lambda \rightarrow \infty$ will “flatten” the depth function u while increasing the values of the albedo ρ , whereas $\lambda \rightarrow 0$ will give “peaky” shapes with lower albedo. Indeed, according to (11), the transformed depth \bar{u} writes

$$\bar{u}(x, y) = \frac{u(x, y)}{\lambda} \quad (41)$$

and, as shown in [7], the transformed albedo writes

$$\bar{\rho}(x, y) = \rho(x, y) \left(\frac{\lambda^2 + \|\nabla u\|^2}{1 + \|\nabla u\|^2} \right)^{1/2} \quad (42)$$

Further discussion on the link between the geometry of the object and its albedo under *bas-relief* transformations can be found in [31]. We show in Fig. 9 the effects of such a transformation on the shape and albedo of Beethoven’s bust, estimated from the images of Fig. 1.

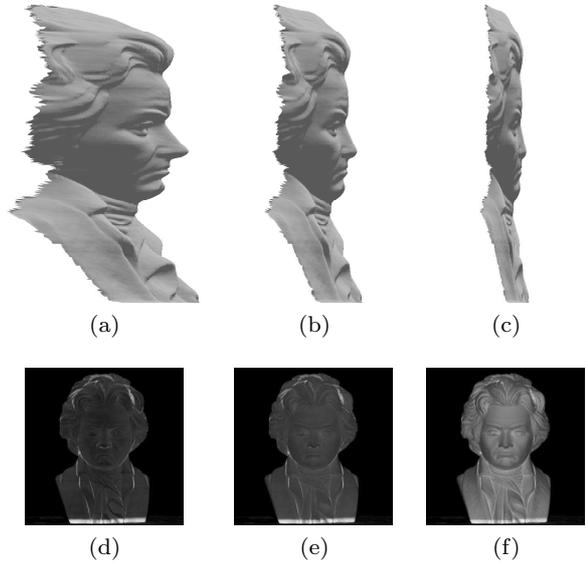


Fig. 9: Effects of a *bas-relief* transformation. First line: reconstructed shapes with (a) $\lambda = 1/2$, (b) $\lambda = 1$ (original shape), and (c) $\lambda = 2$. Second line: albedo with (d) $\lambda = 1/2$, (e) $\lambda = 1$, and (f) $\lambda = 2$.

As stated in Section 3, one possible solution involving total variation to solve for this last parameter is to minimize the total variation of the albedo:

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \operatorname{TV}(\|\mathbf{G}(0, 0, \lambda)^\top M^1\|) \quad (43)$$

This approach is very similar to that of Alldrin *et al.* [3] who minimize the entropy of the albedo distribution, nevertheless here only λ has to be estimated. However, we already stated in Section 3 that in general this approach gives poor results, as real-world albedos are usually not sufficiently regular, and moreover the cumbersome expression of the total variation of the albedo leads to numerical difficulties.

Thus, it is better to use another method for estimating λ . In the experiments, we solve this residual ambiguity by adapting the constant light magnitude constraint [25, 49] to the case of the *bas-relief* ambiguity, i.e. we assume that all the light vectors $S^i, i \in [1, m]$, have the same norm S_0 . When in addition μ and ν are unknown, it is shown in [49] that this constraint requires solving a constrained least-squares problem which is not trivial.

Here, μ and ν being already estimated, the problem is much easier. Let us denote $\mathbf{S}^1 = \mathbf{G}(\hat{\mu}, \hat{\nu}, 1)^{-1} \mathbf{S}^0$. The constant light magnitude constraint reads:

$$\forall i \in [1, m], (\mathbf{G}(0, 0, \lambda)^{-1} S^{1,i})^\top (\mathbf{G}(0, 0, \lambda)^{-1} S^{1,i}) = S_0^2 \quad (44)$$

where $S^{1,i}$ refers to the i -th column of \mathbf{S}^1 , S_0 is the unknown magnitude and λ is the *bas-relief* parameter.

This is nothing more than a linear system with two unknowns ($1/\lambda^2, S_0^2$) and m equations, which can be solved using Moore-Penrose pseudo-inverse. We can thus solve for λ^2 using this method, then deduce λ , since we assume $\lambda > 0$ for convexity.

Note that this system might not admit a real solution (as the estimation of λ^2 might be negative), which is not stated in [25, 49]. Such problems might occur for example if the distribution of the light magnitudes is far away from a uniform distribution: in such cases, other constraints should be considered. It is noticeable that the proposed method for estimating μ and ν is general, and can be used to simplify any method of resolution of the GBR. In Section 5, we adapt the Minimum-Entropy method [3] when the constant light magnitude constraint is not valid: the discrete search on all the possible values of μ, ν and λ proposed in [3] is reduced to a discrete search on the possible values of λ , which considerably accelerates the method.

4.4 An Alternative Approach: Depth Regularization

As will be shown in Section 5, the estimation of μ and ν proposed in Section 4.2, coupled with an efficient resolution of the *bas-relief* ambiguity, reaches state-of-the-art results, and is considerably faster compared to other methods. It has however, in practice, the drawback of involving numerical approximations of the partial derivatives of M by finite differences. We can avoid computing these derivatives by regularizing the depth map u . Indeed, let us state that:

1. Formulation (36) requires the knowledge of $J(M^0)$. This involves necessary numerical approximations of $\nabla M_x^0, \nabla M_y^0$ and ∇M_z^0 in the implementation, which might occasionally be a source of problems, depending on the presence of discontinuities in the depth function. As such discontinuities are present in real-world objects, it might be necessary to apply Gaussian filtering on M^0 so as not to bias the results (the same “trick” is used during the integrability enforcement, as advised in [3]).
2. We showed in Section 3 that the GBR has an impact on the variations of the albedo, the depth and

its first order derivatives, however we did not explicitly use information about the depth.

Thus, we propose an alternative approach involving total variation of the depth u , which, as we will show, does not involve numerical approximations of ∇u by finite differences, as ∇u depends only on the known quantities $p^0 = -M_x^0/M_z^0, q^0 = -M_y^0/M_z^0$ and on the GBR parameters.

Let us recall that the depth function u is assumed to be piecewise twice differentiable, so the integrability condition $\partial^2 u / \partial x \partial y = \partial^2 u / \partial y \partial x$ is implicitly satisfied almost everywhere, and problem (31) can be reformulated in terms of depth as the following problem:

$$\begin{cases} \hat{u} = \operatorname{argmin} \operatorname{TV}(u) \\ \text{s.t. } \|\rho^u \frac{(-\partial_x u, -\partial_y u, 1)}{\sqrt{\|\nabla u\|^2 + 1}} \mathbf{S} - I\|^2 = \min \end{cases} \quad (45)$$

Since this is as complicated to directly solve as (31), we adopt the same parametric formulation. Let us denote by $u^0(x, y)$ the depth map obtained by integrating a normal field N^0 estimated by the method of Yuille and Snow [49] (we do not need to explicitly perform this integration), by $U^0 = (x, y, u^0(x, y))^\top$ the corresponding Monge patch and by $g(\mu, \nu, \lambda) = \frac{1}{\lambda}(-\mu, -\nu, 1)$ the vector depending on the GBR parameters, which corresponds to the third line of $\mathbf{G}(\mu, \nu, \lambda)^{-1}$. Similarly to Prop. 2, we have:

Proposition 4 *A solution of (45) is given by*

$$\hat{u} = g(\hat{\mu}, \hat{\nu}, \hat{\lambda}) U^0 \quad (46)$$

where

$$\begin{cases} (\hat{\mu}, \hat{\nu}, \hat{\lambda}) = \operatorname{argmin}_{\mu, \nu, \lambda} \operatorname{TV}(g(\mu, \nu, \lambda) U^0) \\ \text{s.t. } \lambda > 0 \end{cases} \quad (47)$$

Proof Simply remark that, according to (11), a GBR-modified depth function reads:

$$\bar{u} = g(\mu, \nu, \lambda) U^0 \quad (48)$$

The rest of the proof is similar to that of Prop. 2. \square

We consider u^0 and U^0 are implicitly obtained by integrating the normal field estimated by using [49]. We focus on:

$$(\hat{\mu}, \hat{\nu}, \hat{\lambda}) = \operatorname{argmin}_{\mu, \nu, \lambda} \operatorname{TV}(g(\mu, \nu, \lambda) U^0) \quad (49)$$

As the gradient of the modified depth function $\bar{u} = g(\mu, \nu, \lambda) U^0$ is equal to:

$$\begin{aligned} \nabla \bar{u} &= -\frac{1}{\lambda} (M_x^0/M_z^0 + \mu, M_y^0/M_z^0 + \nu)^\top \\ &= -\frac{1}{\lambda} \begin{pmatrix} (M_x^0 + \mu M_z^0)/M_z^0 \\ (M_y^0 + \nu M_z^0)/M_z^0 \end{pmatrix} \end{aligned} \quad (50)$$

the estimation of λ is problematic, as it leads to $\lambda \rightarrow \infty$. In practice, similarly to the previous case, we have to impose a constraint $\lambda \leq \epsilon'$, $\epsilon' > 0$ to avoid this degenerated solution. As previously, ϵ' can be manually tuned so as to obtain satisfactory 3D-reconstructions, or λ can be estimated by using another constraint, as already discussed in Section 4.3, which is justified by the following lemma:

Lemma 2 *The estimations of μ and ν do not depend on that of λ .*

Proof Eq. (50) shows that

$$\nabla(g(\mu, \nu, \lambda)U^0) = \frac{1}{\lambda} \nabla(g(\mu, \nu, 1)U^0) \quad (51)$$

and thus:

$$\operatorname{argmin}_{\mu, \nu} \operatorname{TV}(g(\mu, \nu, \lambda)U^0) = \operatorname{argmin}_{\mu, \nu} \operatorname{TV}(g(\mu, \nu, \lambda')U^0) \quad (52)$$

for any $\lambda' > 0$. Note also that Eq. (51) shows that $\operatorname{TV}(u)$ is inversely proportional to λ (cf. Fig. 7-b). \square

This lemma justifies the choice of separating the estimations of (μ, ν) and that of λ : we can first fix $\lambda = 1$ (for example) and estimate μ and ν . Once again, this reduces the *generalized bas-relief* ambiguity to the *bas-relief* ambiguity.

Thus, we estimate μ and ν by solving:

$$(\hat{\mu}, \hat{\nu}) = \operatorname{argmin}_{\mu, \nu} \operatorname{TV}(g(\mu, \nu, 1)U^0) \quad (53)$$

The solution of (53) can be obtained once again using quasi-Newton methods, the derivatives of $E'(\mu, \nu) = \operatorname{TV}(g(\mu, \nu, 1)U^0) = \operatorname{TV}(-\mu x - \nu y + u^0(x, y))$ reading:

$$\begin{cases} \partial_{\mu} E' = \int_{\Omega} \frac{M_x^0 + \mu M_z^0}{\sqrt{(M_x^0 + \mu M_z^0)^2 + (M_y^0 + \nu M_z^0)^2}} dx \\ \partial_{\nu} E' = \int_{\Omega} \frac{M_y^0 + \nu M_z^0}{\sqrt{(M_x^0 + \mu M_z^0)^2 + (M_y^0 + \nu M_z^0)^2}} dx \end{cases} \quad (54)$$

To accelerate the estimation, we choose to start from the solution $(\hat{\mu}_0, \hat{\nu}_0)$ of the Tikhonov regularization:

$$(\hat{\mu}_0, \hat{\nu}_0) = \operatorname{argmin}_{\mu, \nu} \int_{\Omega} \|\nabla(g(\mu, \nu, 1)U^0)\|^2 dx \quad (55)$$

which easily gives

$$\begin{cases} \hat{\mu}_0 = \bar{p}^0 \\ \hat{\nu}_0 = \bar{q}^0 \end{cases} \quad (56)$$

where \bar{p}^0 and \bar{q}^0 stand for the means of $p^0 = -M_x^0/M_z^0$ and of $q^0 = -M_y^0/M_z^0$. In fact, (55) results in forcing the mean value of N to be equal to $[0, 0, 1]^{\top}$. This simply signifies that, on average, the normals should be oriented toward the camera, which is quite reasonable. This has also been used as a prior in an Expectation-Maximization algorithm for *calibrated* photometric stereo in presence of outliers [46].

Moreover, note that in (54), only known terms M_x^0 , M_y^0 and M_z^0 are involved: no numerical approximation of $J(M^0)$ is needed.

4.5 GBR-consistency

It was recently emphasized in [39] that an appropriate method for estimating the GBR parameters should be GBR-consistent, i.e. it should return the same estimates \widehat{M} and $\widehat{\mathbf{S}}$, whatever the initial solution (M^0, \mathbf{S}^0) of (8). Papadimitri and Favaro prove in [39] that this property holds for the ‘‘Diffuse Maxima’’ method (DM) initially described in [19]. However, this is not the case for the ‘‘Minimum Entropy’’ method [3], neither for the ‘‘Self calibrating’’ method [44]. The estimations of μ and ν previously discussed actually hold this property, and we can state the following result:

Proposition 5 *Both the TV methods described in Sections 4.2 and 4.4 are GBR-consistent.*

Proof We present the proof for the TV-regularization of M , the other proof is similar.

Let $(M^{01}, \mathbf{S}^{01})$ and $(M^{02}, \mathbf{S}^{02})$ be two different solutions of (8). The problem of estimating a field \widehat{M} with minimal TV, given the initial field M^{01} , is synonymous to solving:

$$\begin{cases} (\hat{\mu}, \hat{\nu}, \hat{\lambda}) = \operatorname{argmin}_{\mu, \nu, \lambda} \operatorname{TV}(\mathbf{G}(\mu, \nu, \lambda)^{\top} M^{01}) \\ \text{s.t. } \lambda > 0 \end{cases} \quad (57)$$

On the other hand, there are necessarily three parameters $(\mu^{01}, \nu^{01}, \lambda^{01}) \in \mathbb{R}^2 \times \mathbb{R}_+^*$ such that the equality $M^{01} = \mathbf{G}(\mu^{01}, \nu^{01}, \lambda^{01})^{\top} M^{02}$ holds. From the form of a GBR, we have $\mathbf{G}(\mu, \nu, \lambda)^{\top} \mathbf{G}(\mu^{01}, \nu^{01}, \lambda^{01})^{\top} = \mathbf{G}(\mu^{01} + \mu\lambda^{01}, \nu^{01} + \nu\lambda^{01}, \lambda\lambda^{01})^{\top}$. Writing $\mu' = \mu^{01} + \mu\lambda^{01}$, $\nu' = \nu^{01} + \nu\lambda^{01}$ and $\lambda' = \lambda\lambda^{01}$, the problem becomes:

$$\begin{cases} (\hat{\mu}, \hat{\nu}, \hat{\lambda}) = \operatorname{argmin}_{\mu', \nu', \lambda'} \operatorname{TV}(\mathbf{G}(\mu', \nu', \lambda')^{\top} M^{02}) \\ \text{s.t. } \lambda' > 0 \end{cases} \quad (58)$$

which is exactly the problem of estimating a field \widehat{M} with minimal TV, given an initial field M^{02} . Thus, the solution in M is the same, whether we initialize the algorithm with M^{01} or with M^{02} . \square

5 Results

In this section we evaluate the accuracy of the two proposed reconstruction methods. Some authors [19, 44] evaluate this accuracy by calculating the mean angular error between the estimated normal field and that estimated by calibrated photometric stereo (Eq. (4)), although it is advocated in [3] that uncalibrated photometric stereo might give overall “better” results than calibrated photometric stereo, as the use of prior information might compensate for some inaccuracy in the light parameters. Thus, we evaluate our methods firstly on synthetic datasets, for which ground truth normals are known.

5.1 Results on Synthetic Images

To evaluate the accuracy of the reconstructions obtained by our methods, referred to as TV- M and TV- u , we use synthetic images of a vase. The images were created in the following way:

- The depth equation can be written as the equation of a polynom (this equation can be found for example in [16]), thus the analytical expression of the normals is known. The depth function u is shown in Fig. 10-a, and its derivatives p and q in Figs. 10-b and 10-c.
- We created a $3 \times m$ light matrix \mathbf{S} , choosing $m = 22$ light vectors having the same intensity (so as to solve the residual *bas-relief* ambiguity).
- We manually created an albedo map, thus obtaining a ground truth M field (see below).
- Eventually, the 256×256 images were generated using the Lambertian model: $I = \max(M^T \mathbf{S}, 0)$ (the max operator models the *self-shadows*).

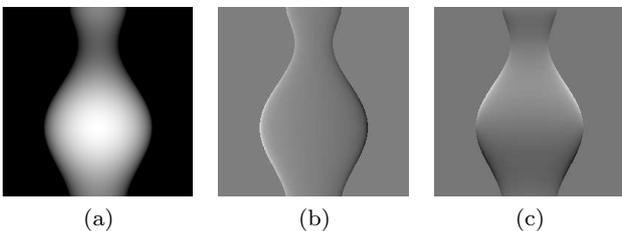


Fig. 10: Ground truth functions (a) u , (b) p and (c) q for the synthetic vase dataset.

We compared our reconstructions to those of [3] and [19], for which Matlab codes are freely available on the authors’ websites. All our codes were developed

using Matlab too so as to provide a fair comparison. CPU times were measured on a personal I7 processor at 2.9GHz. Those methods solve the GBR in this way:

- The Minimum-Entropy method (ME) [3] estimates the GBR by minimizing the entropy of the albedo.
- The Diffuse-Maxima method (DM) [19] considers the brightest pixels as locally oriented towards the light source and thus deduces the GBR.

As these methods have different features, we are studying three cases here, depending on the way we create the albedo:

- With a uniform albedo (Fig. 11-a), the albedo has a very low entropy, and the maxima of intensity give the orientation of the light, so both [3] and [19] should work perfectly.
- With a piecewise uniform albedo (Fig. 11-b), the albedo still has a low entropy, but the maxima of intensity might correspond to high albedo values, so this case favours [3].
- With a radial gradient albedo (Fig. 11-c), the entropy is important, and maxima of intensity might again correspond to high albedo values.

One image of each dataset is shown in the second line of Fig. 11. To evaluate the accuracy of the different methods, we can calculate the relative errors between the estimated GBR parameters and the ground truth parameters, as well as the angular errors between the estimated normals and the ground truth normals. All these results are reported in Table 1.

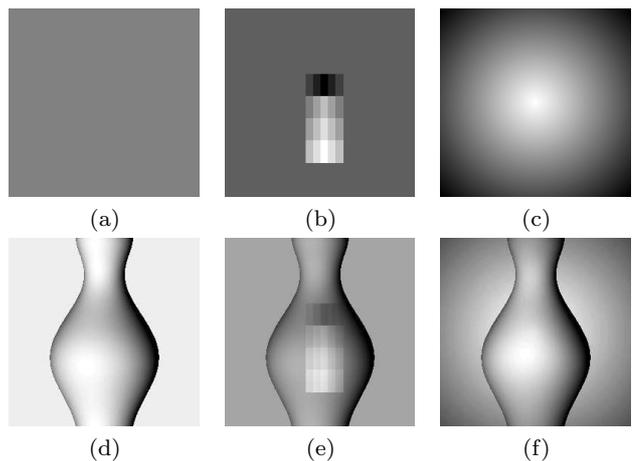


Fig. 11: First line: ground truth albedo maps: (a) uniform; (b) piecewise uniform; (c) radial gradient. Second line: one image of each dataset, corresponding to each different albedo (the same lighting was used to generate these three images).

	$\frac{ \mu - \hat{\mu} }{ \mu }$	$\frac{ \nu - \hat{\nu} }{ \nu }$	$\frac{ \lambda - \hat{\lambda} }{ \lambda }$	MAE (deg.)	Max-AE (deg.)	Min-AE (deg.)	σ_{AE} (deg.)	CPU (s)
<i>Uniform</i>								
ME	1.54	0.00	0.05	0.63	16.41	0.03	1.10	26.41
DM	5.36	0.28	0.01	0.98	16.04	0.01	0.95	2.79
TV- <i>M</i>	93.72	0.27	0.57	10.09	24.77	0.03	3.14	0.13
TV- <i>M</i> (G.f)	0.61	0.02	0.07	0.57	16.75	0.00	1.28	0.26
TV- <i>u</i>	0.61	0.02	0.07	0.57	16.75	0.00	1.28	0.57
<i>P/w unif.</i>								
ME	1.19	0.02	0.04	0.59	16.35	0.00	1.09	28.54
DM	24.11	1.59	0.11	4.82	14.97	0.06	1.25	3.34
TV- <i>M</i>	53.38	0.11	0.36	7.42	21.53	0.11	2.11	0.07
TV- <i>M</i> (G.f)	0.50	0.02	0.07	0.56	16.77	0.00	1.27	0.26
TV- <i>u</i>	0.50	0.02	0.07	0.56	16.77	0.00	1.27	0.50
<i>Rad. grad.</i>								
ME	8.01	18.22	0.21	15.59	18.89	2.03	3.64	30.82
DM	8.77	20.74	0.10	15.53	18.40	3.03	3.45	0.56
TV- <i>M</i>	102.61	0.89	0.20	4.43	18.76	0.07	1.35	0.06
TV- <i>M</i> (G.f)	2.70	0.03	0.07	0.75	16.53	0.01	1.26	0.15
TV- <i>u</i>	2.41	0.02	0.07	0.75	16.53	0.01	1.26	0.23

Table 1: Comparison of the results of the Minimum-Entropy method (ME), the Diffuse-Maxima method (DM) and the proposed methods TV-*M* (without or with applying Gaussian filtering (G.f) before calculating $J(M^0)$) and TV-*u*. We show the relative errors on the GBR parameters μ , ν and λ , as well as the mean angular error (MAE), the maximal angular error (Max-AE), the minimal angular error (Min-AE), the standard deviation of the angular error (σ_{AE}) on the normals, and the CPU time. *Uniform* refers to the synthetic vase with uniform albedo, *P/w unif.* to the vase with piecewise uniform albedo, and *Rad. grad.* to the vase with a radial gradient albedo. The proposed methods overcome state-of-the-art results, and are considerably faster. They can also deal with situations state-of-the-art methods usually fail on (smoothly varying albedo).

As expected, the ME and DM methods give satisfactory results in the first case, but fail if the albedo varies smoothly. On the contrary, our methods produce comparable results in all three cases, as they do not rely on a hypothesis on the albedo, nor on pixel identification. They are thus more general methods, at least for estimating μ and ν .

The ME method uses a global optimization to minimize the (non-convex) entropy of the albedo distribution, resulting in very high CPU times. The CPU time for the DM method is also totally unpredictable, since it depends on the number of local maxima found in the images. On the contrary, the TV methods solely involve simple convex optimization tools and are thus faster.

Finally, let us explain the differences between the results of TV-*M* and TV-*u*. As mentioned, TV-*M* involves numerical approximations, which is totally biased by the non-integrability of the vase on the boundaries. It is for that reason that TV-*u* performs better. It is possible to apply Gaussian filtering before calculating the numerical derivatives in (36) and (38), which reduces the impact of these approximations and improves the results of the TV-*M* method. This modified method is referred to as TV-*M* (G.f).

5.2 Evaluation on Real-World Datasets

5.2.1 Datasets

To evaluate the accuracy of the reconstructions on real-world datasets, since no ground truth is available, we can only compare the results to those of calibrated photometric stereo. It should be recalled that this comparison is purely informative, since estimating the normal field by calibrated photometric stereo is subject to inaccuracy in the given light parameters. We used:

- Ten datasets of almost Lambertian objects. The Buddha, Cat, Horse, Owl and Rock datasets, courtesy of Dan Goldman and Steven Seitz⁴, are composed of twelve 320×300 images (except Horse, which is composed of twelve 270×350 images). The Octopus, Redfish and Korean Doll datasets can be found on Neil Alldrin’s homepage⁵: they are composed of five images, respectively of size 321×281 , 351×301 and 1321×521 . The Doll dataset was presented in [19], and is composed of fifteen 405×250 images. Finally, Beethoven dataset is composed of three 256×256 images.

⁴ <http://courses.cs.washington.edu/courses/cse455/10wi/projects/>

⁵ <http://vision.ucsd.edu/~nalldrin/research/>

- The first twenty sets of images from the Extended Yale DataFace B [23, 33]. These datasets originally contain 64 different illumination conditions: we selected only the 27 with lowest elevation angles, so as to discard the images the most corrupted by shadows. Each image of each dataset is 192×168 .

Because real-world images in reality contain outliers such as shadows or specularities, we apply the preprocessing from [48], which involves a hyper-parameter C that we fixed, as advised in [19], to 1.7 for the datasets with at least 12 images, and to 3 for the other datasets. One preprocessed image of each dataset is shown in Fig. 12. However such a preprocessing cannot remove all the outliers: specular highlights are still visible in, for instance, the Owl dataset (see the eyes of the owl in Fig. 12).

5.2.2 Results

The relative errors on the GBR parameters, the mean angular errors on the normal fields and the CPU times can be found in Tables 2 and 3. As can be seen, the proposed methods reach, in general, state-of-the-art results on those datasets, and considerably overcome the two others in terms of computation times.

For some datasets, no error is given: this corresponds to the cases where the resolution of (44) results in negative values of λ^2 . The constraint of uniform light magnitude is not adapted to such datasets. However, as the estimated values of μ and ν do not rely on this constraint, the estimation of λ can be achieved for example by adapting another method of resolution of the GBR to the resolution of the *bas-relief* ambiguity. Thus, we also report the results of estimating λ by ME, with μ and ν estimated by TV- u (this method is referred to as TV- u / ME). Unifying models (12) and (53), this method can be written:

$$\begin{cases} (\hat{\mu}, \hat{\nu}) = \underset{\mu, \nu}{\operatorname{argmin}} \operatorname{TV}(g(\mu, \nu, 1)U^0) \\ \hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \mathcal{E}(\|\mathbf{G}(\hat{\mu}, \hat{\nu}, \lambda)^\top M^0\|) \end{cases} \quad (59)$$

The discrete search for the minimization of the entropy is thus performed only with respect to λ , which results in considerably shorter computation times, compared to ME. The example of the Korean Doll dataset (cf. Table 2) is striking: the computation time is reduced from more than two minutes to a couple of seconds using TV- u / ME, while obtaining a very acceptable level of accuracy. This confirms that the proposed method for the estimation of μ and ν can be used to accelerate any method of GBR resolution.

Moreover, these tests empirically prove the validity of the ‘‘Minimum-TV’’ constraint for a wide variety of objects, since the obtained errors are comparable to state-of-the-art methods. However, as the proposed methods do not rely either on pixel identification or on minimizing non-convex functions, they are faster than other methods.

Finally, we show in Fig. 13 a side-view of a relighting of each 3D-model obtained after integrating the normals estimated with the TV method which gives the lowest MAE and warping the estimated albedo on the surface. We use three light sources (one in front of the object, one on the left side and one on the right side) and the Phong model [40] for rendering.

5.2.3 Guidelines

Finally, we provide some guidelines to help the user choose the right method. To use the proposed methods, two choices have to be made:

- TV-regularization of M or TV-regularization of u : this is linked to the estimation of μ and ν .
- How to solve the residual *bas-relief* ambiguity, which is linked to the estimation of λ .

The TV-regularization of u should, in the most general case, give better results, as it does not rely on numerical differentiations of the initial field M^0 , which are biased in the presence of depth discontinuities (e.g., the Owl dataset). However, if both the surface and the albedo are smooth enough (e.g., the Horse and the B13 datasets), the TV-regularization of M will offer comparable or better results. However, it is hard to *a priori* ensure the absence of depth discontinuities.

Thus, to choose the appropriate method, we advise considering the following strategy:

- If no prior information is available, the best choice is probably to firstly try the TV- u method: the estimation of μ and ν does not involve any numerical approximation, and the estimation of λ using the constant light magnitude constraint is quite fast.
- If the surface and the albedo seem smooth enough, the TV- M method may give overall better estimations of μ and ν .
- Finally, if the constant light magnitude constraint is not satisfied, in both cases (TV- u or TV- M) negative values of λ^2 will be obtained, or the surface will look either too flat or too peaky. In that case, other methods of estimation of λ should be considered. If the albedo is regular enough, one can easily adapt the Min. Entropy method to the estimation of λ only (we provide such an implementation in Matlab): this corresponds to the TV- u / ME method. Other methods [19, 44] would also be easy to adapt.



Fig. 12: One preprocessed image of each real-world dataset used in the experiments.

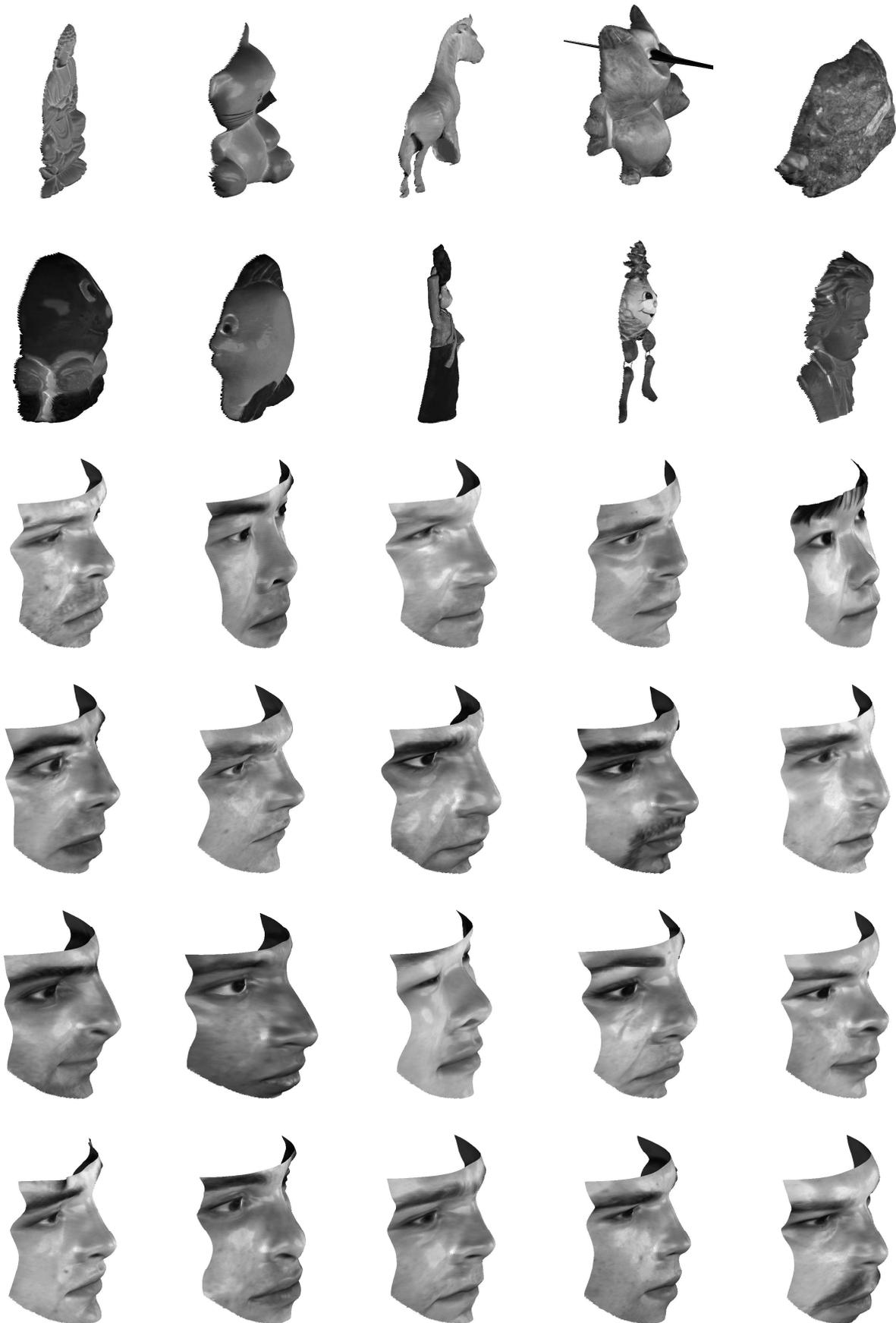


Fig. 13: Relighting of each 3D-model obtained by the TV method which gives the lowest MAE.

	$\frac{ \mu-\hat{\mu} }{ \mu }$	$\frac{ \nu-\hat{\nu} }{ \nu }$	$\frac{ \lambda-\hat{\lambda} }{ \lambda }$	MAE	CPU		$\frac{ \mu-\hat{\mu} }{ \mu }$	$\frac{ \nu-\hat{\nu} }{ \nu }$	$\frac{ \lambda-\hat{\lambda} }{ \lambda }$	MAE	CPU
<i>B01</i>						<i>B11</i>					
ME	18.83	5.26	0.26	14.78	17.03	ME	2.29	4.62	0.13	19.10	17.99
DM	33.92	5.17	0.06	14.78	9.05	DM	2.17	2.32	0.16	12.06	1.67
TV- <i>M</i>	35.45	3.27	0.75	16.23	0.03	TV- <i>M</i>	0.14	1.96	0.68	16.49	0.03
TV- <i>u</i>	24.45	4.23	0.75	16.22	0.03	TV- <i>u</i>	2.02	1.52	0.51	14.36	0.03
TV- <i>u</i> / ME	24.45	4.23	0.22	12.69	0.26	TV- <i>u</i> / ME	2.02	1.52	0.33	12.79	0.43
<i>B02</i>						<i>B12</i>					
ME	2.27	13.24	0.30	17.08	16.92	ME	4.33	22.20	0.46	25.51	21.27
DM	3.80	15.51	0.16	18.56	4.34	DM	0.83	10.92	0.36	14.92	1.06
TV- <i>M</i>	3.95	3.45	1.02	15.41	0.14	TV- <i>M</i>	5.52	5.30	0.04	7.92	0.04
TV- <i>u</i>	1.12	8.86	1.15	16.66	0.12	TV- <i>u</i>	6.70	1.34	0.03	6.63	0.03
TV- <i>u</i> / ME	1.12	8.86	0.11	9.37	0.26	TV- <i>u</i> / ME	6.70	1.34	0.36	11.19	0.38
<i>B03</i>						<i>B13</i>					
ME	7.70	3.25	0.23	11.56	23.57	ME	0.84	1.03	0.08	5.21	22.26
DM	10.79	2.86	0.10	9.23	4.67	DM	2.21	1.89	0.10	8.36	11.63
TV- <i>M</i>	22.32	0.64	0.51	12.95	0.04	TV- <i>M</i>	0.39	0.43	0.41	7.13	0.03
TV- <i>u</i>	10.18	0.61	0.33	9.01	0.03	TV- <i>u</i>	2.99	3.29	3.53	26.68	0.03
TV- <i>u</i> / ME	10.18	0.61	0.16	5.41	0.28	TV- <i>u</i> / ME	2.99	3.29	0.09	14.82	0.52
<i>B04</i>						<i>B15</i>					
ME	77.65	2.90	0.18	10.41	20.85	ME	22.59	3.06	0.16	11.47	19.11
DM	73.36	4.69	0.01	14.23	4.44	DM	38.75	4.45	0.09	16.38	2.62
TV- <i>M</i>	121.13	2.11	0.63	15.69	0.03	TV- <i>M</i>	35.21	2.06	0.63	16.25	0.03
TV- <i>u</i>	49.40	4.22	0.94	19.15	0.03	TV- <i>u</i>	30.25	4.07	1.17	22.55	0.04
TV- <i>u</i> / ME	49.40	4.22	0.14	12.84	0.39	TV- <i>u</i> / ME	30.25	4.07	0.15	14.80	0.35
<i>B05</i>						<i>B16</i>					
ME	0.40	0.12	0.26	7.50	18.54	ME	4.89	2.16	0.26	8.55	21.61
DM	0.04	0.39	0.13	3.38	5.65	DM	8.42	4.12	0.07	9.96	6.34
TV- <i>M</i>	2.38	2.91	0.59	13.93	0.03	TV- <i>M</i>	7.53	0.68	0.34	8.71	0.04
TV- <i>u</i>	0.86	2.51	0.58	13.08	0.03	TV- <i>u</i>	7.42	1.57	0.38	9.65	0.03
TV- <i>u</i> / ME	0.86	2.51	0.02	8.87	0.28	TV- <i>u</i> / ME	7.42	1.57	0.21	7.25	0.33
<i>B06</i>						<i>B17</i>					
ME	3.32	3.73	0.19	12.89	19.23	ME	3.35	2.72	0.19	10.15	21.45
DM	4.53	2.93	0.19	11.35	3.11	DM	4.86	3.25	0.02	10.84	3.68
TV- <i>M</i>	1.28	1.58	0.30	9.05	0.03	TV- <i>M</i>	5.06	2.00	0.53	12.99	0.03
TV- <i>u</i>	4.96	2.02	0.55	13.43	0.03	TV- <i>u</i>	2.25	2.59	0.64	14.61	0.03
TV- <i>u</i> / ME	4.96	2.02	0.21	9.85	0.32	TV- <i>u</i> / ME	2.25	2.59	0.18	9.56	0.32
<i>B07</i>						<i>B18</i>					
ME	1.94	3.54	0.24	11.21	18.53	ME	2.01	1.98	0.18	10.12	20.21
DM	2.29	2.62	0.15	7.98	11.48	DM	5.14	1.89	0.12	11.19	1.39
TV- <i>M</i>	2.28	2.00	0.53	11.97	0.03	TV- <i>M</i>	4.53	0.18	0.27	9.38	0.03
TV- <i>u</i>	3.63	3.19	0.83	16.41	0.03	TV- <i>u</i>	4.16	1.52	0.61	14.64	0.03
TV- <i>u</i> / ME	3.63	3.19	0.21	10.73	0.35	TV- <i>u</i> / ME	4.16	1.52	0.13	9.33	0.43
<i>B08</i>						<i>B19</i>					
ME	157.37	4.66	0.34	19.70	18.31	ME	24.39	5.47	0.18	16.24	21.73
DM	127.10	3.69	0.25	14.75	2.90	DM	32.28	5.71	0.24	17.80	3.18
TV- <i>M</i>	136.65	0.87	0.33	10.01	0.03	TV- <i>M</i>	23.29	2.45	0.45	13.26	0.10
TV- <i>u</i>	154.10	1.13	0.38	10.96	0.03	TV- <i>u</i>	37.93	3.61	0.73	17.61	0.08
TV- <i>u</i> / ME	154.10	1.13	0.22	7.90	0.40	TV- <i>u</i> / ME	37.93	3.61	0.18	12.50	0.82
<i>B09</i>						<i>B20</i>					
ME	5.50	6.99	0.30	17.98	23.36	ME	1.83	6.90	0.33	17.20	18.05
DM	7.24	7.22	0.14	15.98	1.10	DM	3.89	5.17	0.21	11.86	5.69
TV- <i>M</i>	7.76	2.11	0.30	10.95	0.03	TV- <i>M</i>	2.76	1.90	0.30	8.38	0.03
TV- <i>u</i>	8.24	0.88	0.21	9.36	0.03	TV- <i>u</i>	5.91	1.40	0.31	8.71	0.03
TV- <i>u</i> / ME	8.24	0.88	0.30	9.38	0.32	TV- <i>u</i> / ME	5.91	1.40	0.22	7.10	0.32
<i>B10</i>						<i>B21</i>					
ME	12.34	217.43	0.15	6.92	19.22	ME	43.56	1.80	0.27	9.98	16.76
DM	16.48	459.18	0.14	11.34	5.32	DM	90.43	2.64	0.18	11.93	2.35
TV- <i>M</i>	6.22	340.09	0.22	9.10	0.03	TV- <i>M</i>	112.68	1.05	0.55	14.06	0.03
TV- <i>u</i>	25.54	87.17	0.17	7.83	0.03	TV- <i>u</i>	103.96	1.82	0.67	15.67	0.03
TV- <i>u</i> / ME	25.54	87.17	0.13	6.26	0.33	TV- <i>u</i> / ME	103.96	1.82	0.16	10.38	0.48

Table 2: Comparison of the results of the ME, DM and TV methods, on the first twenty datasets of the extended Yale DataFace B. Note that B14 does not exist in this database. Units are the same as in Table 1.

	$\frac{ \mu - \hat{\mu} }{ \mu }$	$\frac{ \nu - \hat{\nu} }{ \nu }$	$\frac{ \lambda - \hat{\lambda} }{ \lambda }$	MAE	CPU
<i>Buddha</i>					
ME	0.81	0.22	0.06	6.38	13.87
DM	0.07	0.06	0.12	4.99	1.91
TV- M	0.91	0.35	0.21	10.92	0.06
TV- u	1.03	0.51	1.34	18.85	0.03
TV- u / ME	1.03	0.51	0.00	14.79	0.35
<i>Cat</i>					
ME	0.48	0.49	0.17	14.37	17.82
DM	0.95	0.05	0.08	5.37	0.57
TV- M	1.00	0.10	0.27	6.16	0.06
TV- u	0.13	0.54	5.94	35.57	0.04
TV- u / ME	0.13	0.54	0.18	16.04	0.45
<i>Horse</i>					
ME	11.33	0.28	0.09	10.88	14.21
DM	14.25	0.20	0.03	4.80	1.25
TV- M	11.64	0.11	0.21	6.68	0.07
TV- u	34.83	0.34	0.64	18.25	0.03
TV- u / ME	34.83	0.34	0.13	15.59	0.34
<i>Owl</i>					
ME	0.41	0.63	0.25	15.67	22.71
DM	0.35	0.00	0.07	6.63	0.43
TV- M	0.50	0.23	0.13	8.42	0.08
TV- u	0.15	0.13	0.08	5.54	0.05
TV- u / ME	0.15	0.13	0.15	7.54	0.54
<i>Rock</i>					
ME	14.77	0.75	0.23	18.37	32.30
DM	5.35	0.39	0.05	11.61	0.61
TV- M	3.13	0.88	-	-	0.08
TV- u	9.43	1.19	-	-	0.07
TV- u / ME	9.43	1.19	0.19	27.56	0.76
<i>Octopus</i>					
ME	0.06	0.07	0.07	3.06	22.49
DM	0.10	0.16	0.19	6.64	0.26
TV- M	0.03	0.14	-	-	0.07
TV- u	0.11	0.29	-	-	0.07
TV- u / ME	0.11	0.29	0.28	9.41	0.67
<i>Redfish</i>					
ME	7.06	0.06	0.03	4.38	21.52
DM	8.85	0.18	0.05	5.60	0.79
TV- M	20.83	0.06	3.48	25.92	0.10
TV- u	3.28	3.75	2.93	26.07	0.05
TV- u / ME	3.28	3.75	0.17	14.17	0.61
<i>Korean Doll</i>					
ME	0.04	0.19	0.08	3.39	151.99
DM	0.20	0.80	0.27	17.74	2.28
TV- M	0.11	0.46	-	-	0.49
TV- u	0.09	0.50	-	-	0.34
TV- u / ME	0.09	0.50	0.15	10.36	3.87
<i>Doll</i>					
ME	0.05	2.64	0.47	26.24	17.13
DM	0.03	0.81	0.34	12.15	0.59
TV- M	0.05	1.81	-	-	0.06
TV- u	0.04	1.82	-	-	0.03
TV- u / ME	0.04	1.82	0.44	19.84	0.35
<i>Beethoven</i>					
ME	10.50	6.69	0.57	63.83	16.39
DM	0.02	0.25	0.41	10.01	1.31
TV- M	0.15	0.34	0.65	16.44	0.05
TV- u	0.06	0.31	0.47	13.70	0.03
TV- u / ME	0.06	0.31	0.07	13.30	0.32

Table 3: Comparison of the results of the ME, DM and TV methods, on ten almost Lambertian objects. Units are the same as in Table 1.

6 Conclusion and Future Work

Contributions. In this paper, we showed that the properties of total variation were highly useful in the resolution of the uncalibrated photometric stereo problem, since estimating the parameters of the GBR transformation as the minimizers of the total variation of either the M field or the depth function u reduces the GBR ambiguity to the classical *bas-relief* ambiguity, which is considerably easier to solve. We experimentally proved that the proposed methods reached state-of-the-art results in terms of accuracy, with much shorter computation times, thanks to the parametric formulation of the original problem (31), which we converted into the problem of minimizing a convex function depending only on 3 parameters.

Future prospects. Despite the fact that very satisfactory results were obtained using the proposed methods, their justification remains empirical: a statistical study of the distributions of ρ , u and ∇u over a wide variety of objects would probably help understanding why total variation offers such satisfactory results.

All the work presented here assumes the light sources considered are directional, so as to obtain a linear formulation of the problem. In practice, directional lighting is difficult to ensure, and point-light sources should be considered. The problem becoming nonlinear, the GBR ambiguity would disappear. However, the problem would become numerically much more difficult, so efficient strategies should be considered.

Finally, for a robust estimation, it would be necessary to consider outliers such as noise, specularities and shadows, as they induce a bias in the estimation. Indeed, as the preprocessing [48] is purely driven by the rank of the matrix, it does not take spatial considerations into account and thus gives poor results in presence of noise or strong outliers, therefore some artifacts in the reconstructions are still visible (see the eye of the owl in Fig. 13). However, such artifacts could easily be removed by “smoothing” the normal field: the model (30) could be used as a post-processing method after the estimation of M and before the integration step. To ensure that the normal fields are “almost everywhere” integrable, we could switch the \mathcal{L}^2 minimization $\|\overline{\text{curl}} M\|^2 = \min$ in (30) to a \mathcal{L}^1 minimization $\|\overline{\text{curl}} M\|_1 = \min$ (this could be efficiently achieved by using iteratively reweighted least-squares). Proceeding so would actually result in an improvement of the method of Reddy *et al.* [42], who are concerned in enforcing integrability on the normal field, but not in the particular case of photometric stereo, for which not only the normal field, but also the albedo has to be corrected.

References

1. Abrams A, Hawley C, Pless R (2012) Heliometric Stereo: Shape from Sun Position. In: Proceedings of the 12th European Conference on Computer Vision, Lecture Notes in Computer Science, vol 7573, pp 357–370
2. Ackermann J, Langguth F, Fuhrmann S, Goesele M (2012) Photometric Stereo for Outdoor Webcams. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 262–269
3. Alldrin NG, Mallick SP, Kriegman DJ (2007) Resolving the Generalized Bas-relief Ambiguity by Entropy Minimization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
4. Attouch H, Bolte J, Redont P, Soubeyran A (2008) Alternating Proximal Algorithms for Weakly Coupled Convex Minimization Problems. Applications to Dynamical Games and PDE's. *Journal of Convex Analysis* 15(3):485–506
5. Aubert G, Kornprobst P (2006) *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations* (second edition), Applied Mathematical Sciences, vol 147. Springer-Verlag
6. Barsky S, Petrou M (2003) The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(10):1239–1252
7. Belhumeur PN, Kriegman DJ, Yuille AL (1999) The Bas-Relief Ambiguity. *International Journal of Computer Vision* 35(1):33–44
8. Bresson X, Chan T (2008) Fast Dual Minimization of the Vectorial Total Variation Norm and Applications to Color Image Processing. *Inverse Problems and Imaging* 2(4):455–484
9. Chambolle A (2004) An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision* 20(1):89–97
10. Chambolle A, Lions PL (1997) Image recovery via total variation minimization and related problems. *Numerische Mathematik* 76(2):167–188
11. Chambolle A, Pock T (2011) A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision* 40(1):120–145
12. Chandraker MK, Kahl F, Kriegman DJ (2005) Reflections on the Generalized Bas-Relief Ambiguity. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (volume I), pp 788–795
13. Chang JY, Lee KM, Lee SU (2007) Multiview normal field integration using level set methods. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Beyond Multiview Geometry: Robust Estimation and Organization of Shapes from Multiple Cues
14. Drbohlav O, Chantler M (2005) Can Two Specular Pixels Calibrate Photometric Stereo? In: Proceedings of the 10th IEEE International Conference on Computer Vision (volume II), pp 1850–1857
15. Durou JD, Courteille F (2007) Integration of a normal field without boundary condition. In: Proceedings of the First International Workshop on Photometric Analysis For Computer Vision
16. Durou JD, Falcone M, Sagona M (2008) Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding* 109(1):22–43
17. Epstein R, Yuille A, Belhumeur PN (1996) Learning and Recognizing Objects Using Illumination Subspaces. In: Proceedings of the 4th European Conference on Computer Vision, International Workshop on Object Representation for Computer Vision
18. Fan J, Wolff LB (1997) Surface curvature and shape reconstruction from unknown multiple illumination and integrability. *Computer Vision and Image Understanding* 65(2):347–359
19. Favaro P, Papadimitri T (2012) A Closed-Form Solution to Uncalibrated Photometric Stereo via Diffuse Maxima. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 821–828
20. Frankot RT, Chellappa R (1988) A Method for Enforcing Integrability in Shape from Shading Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10(4):439–451
21. Galliani S, Ju YC, Breuß M, Bruhn A (2013) Generalised Perspective Shape from Shading in Spherical Coordinates. In: Proceedings of the 4th International Conference on Scale Space and Variational Methods in Computer Vision, pp 222–233
22. Georgiades AS (2003) Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In: Proceedings of the 9th IEEE International Conference on Computer Vision (volume II), pp 816–823
23. Georgiades AS, Kriegman DJ, Belhumeur PN (2001) From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6):643–660

24. Harker M, O’Leary P (2008) Least squares surface reconstruction from measured gradient fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
25. Hayakawa H (1994) Photometric stereo under a light-source with arbitrary motion. *Journal of the Optical Society of America - Part A: Optics, Image Science, and Vision* 11(11):3079–3089
26. Hernández C, Vogiatzis G, Brostow GJ, Stenger B, Cipolla R (2008) Multiview Photometric Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(3):548–554
27. Hertzmann A, Seitz SM (2005) Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(8):1254–1264
28. Horn BKP, Brooks MJ (1986) The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing* 33(2):174–208
29. Horn BKP, Brooks MJ (1989) *Shape from Shading*. MIT Press
30. Horowitz I, Kiryati N (2004) Depth from gradient fields and control points: Bias correction in photometric stereo. *Image and Vision Computing* 22(9):681–694
31. Koenderink JJ, Van Doorn AJ (1997) The generic bilinear calibration-estimation problem. *International Journal of Computer Vision* 23(3):217–234
32. Kozera R (1991) Existence and uniqueness in photometric stereo. *Applied Mathematics and Computation* 44(1):1–103
33. Lee K, Ho J, Kriegman D (2005) Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(5):684–698
34. Mallick S, Zickler T, Kriegman D, Belhumeur P (2005) Beyond lambert: Reconstructing specular surfaces using color. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
35. Mecca R, Durou JD (2011) Unambiguous photometric stereo using two images. In: Proceedings of the 16th International Conference on Image Analysis and Processing, pp 286–295
36. Mecca R, Rosman G, Kimmel R, Bruckstein AM (2013) Perspective Photometric Stereo with Shadows. In: Proceedings of the 4th International Conference on Scale Space and Variational Methods in Computer Vision, pp 258–269
37. Oliensis J (1991) Uniqueness in shape from shading. *International Journal of Computer Vision* 6(2):75–104
38. Onn R, Bruckstein AM (1990) Integrability Disambiguates Surface Recovery in Two-Image Photometric Stereo. *International Journal of Computer Vision* 5(1):105–113
39. Papadimitri T, Favaro P (2013) A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *International Journal of Computer Vision* (to appear)
40. Phong BT (1975) Illumination for computer generated pictures. *Communications of the ACM* 18(6):311–317
41. Quéau Y, Lauze F, Durou JD (2013) Solving the Uncalibrated Photometric Stereo Problem Using Total Variation. In: Proceedings of the 4th International Conference on Scale Space and Variational Methods in Computer Vision, pp 270–281
42. Reddy D, Agrawal A, Chellappa R (2009) Enforcing integrability by error correction using L1-minimization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
43. Rudin LI, Osher S, Fatemi E (1992) Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D: Nonlinear Phenomena* 60(1–4):259–268
44. Shi B, Matsushita Y, Wei Y, Xu C, Tan P (2010) Self-calibrating Photometric Stereo. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
45. Simchony T, Chellappa R, Shao M (1990) Direct analytical methods for solving Poisson equations in computer vision problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(5):435–446
46. Verbiest F, Van Gool L (2008) Photometric stereo with coherent outlier handling and confidence estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
47. Woodham RJ (1980) Photometric Method for Determining Surface Orientation from Multiple Images. *Optical Engineering* 19(1):139–144
48. Wu L, Ganesh A, Shi B, Matsushita Y, Wang Y, Ma Y (2010) Robust Photometric Stereo via Low-Rank Matrix Completion and Recovery. In: Proceedings of the 10th Asian Conference on Computer Vision
49. Yuille AL, Snow D (1997) Shape and Albedo from Multiple Images using Integrability. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 158–164
50. Zhou Z, Wu Z, Tan P (2013) Multi-view Photometric Stereo with Spatially Varying Isotropic Materials. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition