A L^1-TV Algorithm for Robust Perspective Photometric Stereo with Spatially-Varying Lightings

Yvain Quéau[1]([✉]), François Lauze[2], and Jean-Denis Durou[1]

[1] IRIT, UMR CNRS 5505, Toulouse, France
yvain.queau@enseeiht.fr, durou@irit.fr
[2] Department of Computer Science, University of Copenhagen,
Copenhagen, Denmark
francois@diku.dk

Abstract. We tackle the problem of perspective 3D-reconstruction of Lambertian surfaces through photometric stereo, in the presence of outliers to Lambert's law, depth discontinuities, and unknown spatially-varying lightings. To this purpose, we introduce a robust L^1-TV variational formulation of the recovery problem where the shape itself is the main unknown, which naturally enforces integrability and permits to avoid integrating the normal field.

Keywords: Uncalibrated photometric stereo · Spatially-varying lightings · Perspective projection · Total variation · Proximal methods

1 Introduction

Photometric stereo refers to the problem of inferring the shape of an object, given a set of m images of this object acquired under different known (*calibrated* photometric stereo) or unknown (*uncalibrated* photometric stereo) illuminations but from the same point of view, by inverting a generative image model accounting for the projection model (orthographic or perspective), the object reflectance (Lambert, Phong, etc.) and the lighting model (directional, punctual, etc.). Early photometric stereo considered orthographic projection, Lambertian reflectance and directional lightings [17], which is the simplest case. Extending photometric stereo to more general models has become an important research direction, along with the natural need for fast and robust solutions. For example, perspective projection can replace the orthographic projection in both calibrated [11] and uncalibrated [14] cases, the Lambertian assumption can be relaxed so as to take specular highlights into account [7] and robustness to outliers like shadows can be obtained through numerous techniques such as sparse regression [10]. However, extending photometric stereo to non-directional lightings remains an open and challenging problem. We focus on this aspect of the problem and propose a robust variational formulation of perspective photometric stereo with spatially-varying lightings, considered as additional unknowns so as not to impose a parametric lighting model. The different hypotheses our approach relies on are described hereafter.

1.1   Image Formation Model

Radiance Model. Lambert's law assumes that the surface $\mathcal{S}$ reflects light diffusively, the reflectance at $\mathbf{x} \in \mathcal{S}$ being characterised by the albedo $\rho(\mathbf{x}) \in [0, 1]$. Given a parallel uniform light beam $\mathbf{s}_0 \in \mathbb{R}^3$, oriented towards the source, and the unit outward normal $\mathbf{n}(\mathbf{x}) \in \mathbb{R}^3$ to the surface at $\mathbf{x}$, the *emitted* Lambertian radiance $r_e(\mathbf{x})$ at $\mathbf{x}$ is proportional to the irradiance $\max\{0, \mathbf{s}_0 \cdot \mathbf{n}(\mathbf{x})\}$, where the max operator aims at modeling the self-shadows. Forgetting this operator (shadows will be dealt with by robust estimators), the Lambertian emitted radiance reads:

$$r_e(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\pi} \, \mathbf{s}_0 \cdot \mathbf{n}(\mathbf{x}) \tag{1}$$

In this work, we do not suppose that the lightings consist in parallel uniform beams. Even more, we do not use any explicit lighting model, allowing us to cope with, for instance, near point light sources, large sources, self- and cast-shadows and even secondary reflections. In the most general case, the irradiance at $\mathbf{x}$ is equal to the sum of elementary contributions coming from all the unit directions $\mathbf{v}(\theta, \phi)$ characterised by the spherical angles (θ, ϕ) defined with respect to $\mathbf{n}(\mathbf{x})$:

$$r_e(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\pi} \iint_{\theta \in [0, \pi/2], \, \phi \in [0, 2\pi]} r_i(\mathbf{x}, \theta, \phi) \mathbf{v}(\theta, \phi) \cdot \mathbf{n}(\mathbf{x}) \sin\theta \, d\theta \, d\phi \tag{2}$$

where $r_i(\mathbf{x}, \theta, \phi)$ is the *incident* radiance at $\mathbf{x}$ in the direction $\mathbf{v}(\theta, \phi)$. Let $\mathbf{s}(\mathbf{x}) = \iint_{\theta, \phi} r_i(\mathbf{x}, \theta, \phi) \mathbf{v}(\theta, \phi) \sin\theta \, d\theta \, d\phi$. We call the vector field $\mathbf{s}$ the *light field*. Eq. (2) then provides a more general Lambertian emitted radiance model than (1):

$$r_e(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\pi} \, \mathbf{s}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) \tag{3}$$

In the photometric stereo framework, m different light fields $\mathbf{s}^i$, $i \in [1, m]$, successively illuminate the scene, providing m images I^i, $i \in [1, m]$, assumed to be graylevel images. The graylevel is assumed to be proportional to the emitted radiance $r_e(\mathbf{x})$ as expressed in (3). To our knowledge, this very general radiance model has not been used in the context of photometric stereo so far. In most applications, a single point light source at infinity is assumed, so that each lighting is represented by a constant vector in $\mathbb{R}^3$ [1,7,14,15,17]. Closer to this general expression is the important work of Basri *et al.* in [2], where a distribution of point light sources at infinity is assumed, allowing the authors to develop shape recovery algorithms based on low order spherical harmonic decompositions.

Camera Model.   We assume the camera is a calibrated pinhole camera, represented by its focal length f and its principal point, considered as the origin of the image domain $\Omega \subset \mathbb{R}^2$. Each pixel $\mathbf{x}_p \in \Omega$ is then uniquely associated to a 3D point $\mathbf{x}(\mathbf{x}_p) = [x(\mathbf{x}_p), y(\mathbf{x}_p), z(\mathbf{x}_p)]^\top \in \mathbb{R}^3$, where $z > 0$ is called the *depth*, so that the surface of the object to be reconstructed can be represented as the set of 3D points $\mathcal{S} = \{\mathbf{x}(\mathbf{x}_p), \mathbf{x}_p \in \Omega\}$, with:

$$\mathbf{x}(\mathbf{x}_p) = \exp(u(\mathbf{x}_p)/f) \begin{bmatrix} \frac{1}{f}\mathbf{x}_p \\ 1 \end{bmatrix}, \quad u(\mathbf{x}_p) = f \log z(\mathbf{x}_p) \tag{4}$$

It follows from (4) that the unit outward normal $\mathbf{n}(\mathbf{x})$ to the surface at $\mathbf{x}$ can be equivalently defined on the image domain Ω, through (see e.g. [11]):

$$\mathbf{n}_u\left(\mathbf{x}_p\right) = \frac{1}{\sqrt{\left\|\nabla u\left(\mathbf{x}_p\right)\right\|^2 + \left(1 + \frac{1}{f}\mathbf{x}_p \cdot \nabla u(\mathbf{x}_p)\right)^2}} \begin{bmatrix} -\nabla u\left(\mathbf{x}_p\right) \\ 1 + \frac{1}{f}\mathbf{x}_p \cdot \nabla u\left(\mathbf{x}_p\right) \end{bmatrix} \qquad (5)$$

In the radiance model (3), the albedo ρ and the light field $\mathbf{s}$ are defined on the surface $\mathcal{S}$, which is a subset of $\mathbb{R}^3$. Yet, since they are independent from the projection model, they can be equivalently defined onto the image domain Ω through the mapping (4). We will thus refer to their values as $\rho(\mathbf{x}_p)$ and $\mathbf{s}(\mathbf{x}_p)$. Forgetting the normalisation factor $\frac{1}{\pi}$, the image formation model finally reads:

$$I^i(\mathbf{x}_p) = \rho(\mathbf{x}_p)\,\mathbf{s}^i(\mathbf{x}_p) \cdot \mathbf{n}_u(\mathbf{x}_p), \qquad \mathbf{x}_p \in \Omega, \ i \in [1, m] \qquad (6)$$

knowing that this equality is in fact a relation of proportionality.

1.2   Overview of Our Contributions

In most of literature on photometric stereo, the main unknown is the couple (albedo, normal field) which is estimated from the set of linear equations (6) without any prior on the camera model: knowledge of the camera parameters only has importance when *integrating* this estimated normal field into a depth map [6]. Yet, recent photometric approaches explicitly benefit from the perspective camera model: it is shown in [14] that the directional uncalibrated photometric stereo problem is better constrained under perspective projection, provided the normal field is considered integrable. Also, Mecca *et al.* show in [11] that the system of nonlinear PDEs in $z = \exp(u/f)$ resulting from (5) and (6) can be solved through a semi-Lagrangian scheme, avoiding the classical two-steps approach. The path we follow here combines the benefits from both these recent works, and improve their robustness through the introduction of a variational formalism.

To emphasise the advantages of our approach over prior work, let us state that: 1) it is the first variational approach to uncalibrated photometric stereo; 2) it enhances robustness through L^1-TV optimisation (only least-squares approaches were considered in the uncalibrated case so far [7,14,15]); 3) depth discontinuities can be recovered without the need for *a posteriori* robust normal field integration through e.g., techniques like those described in [6]; and 4) unknown, spatially-varying, lightings and albedo are considered (the one-step approach from [11] assumes calibrated directional lightings and known albedo).

The rest of this paper is organised as follows. Using a Bayesian rationale, we use the hypotheses above to derive in Section 2 a generic maximum a posteriori formulation of the recovery problem, and a variational formulation where regularisations use total variation semi-norms. Then, we propose in Section 3 a proximal algorithm for minimising the associated energy, before demonstrating in Section 4 the benefit of using spatially-varying lightings and the proposed robust recovery.

2   From Bayesian Inference to Variational Formulation

We derive in this section a variational approach for inverting the image formation model (6), via ideas from Bayesian inference for factoring the posterior probability, in the spirit of Mumford's work [12]. For the sake of compactness, we will denote the set of images $\mathbf{i} = [I^1, \dots, I^m]^\top$ (vector field $\Omega \to \mathbb{R}^m$) and that of lightings $\mathbf{S} = [\mathbf{s}^1, \dots, \mathbf{s}^m]$ (matrix field $\Omega \to \mathbb{R}^{3 \times m}$).

2.1   MAP Estimation

We first propose to recover $(u, \rho, \mathbf{S})$ as the maximum a posteriori (MAP) of the distribution

$$\mathcal{P}(u, \rho, \mathbf{S}|\mathbf{i}) = \frac{\mathcal{P}(\mathbf{i}|u, \rho, \mathbf{S}) \, \mathcal{P}(u, \rho, \mathbf{S})}{\mathcal{P}(\mathbf{i})} \qquad (7)$$

where $\mathcal{P}(\mathbf{i})$ is the evidence, which can be discarded since it is constant and plays no role in MAP, $\mathcal{P}(\mathbf{i}|u, \rho, \mathbf{S})$ is the likelihood, and $\mathcal{P}(u, \rho, \mathbf{S})$ is the prior, which is factored as $\mathcal{P}(u, \rho, \mathbf{S}) = \mathcal{P}(\mathbf{S}|u, \rho)\mathcal{P}(u, \rho)$. In all photometric stereo works, u and ρ are expressed as independent functions on Ω, which is also the point of view used here, thus we factor $\mathcal{P}(u, \rho) = \mathcal{P}(u)\mathcal{P}(\rho)$.

For sake of simplicity, we assume that the lightings are independent from the surface and the albedo as well, so that the conditional prior simplifies: $\mathcal{P}(\mathbf{S}|u, \rho) = \mathcal{P}(\mathbf{S})$. Of course, this simplification is abusive in the presence of secundary reflections. Yet, such reflections are in general sufficiently sparse to be neglected almost everywhere, and their effect on depth recovery can be limited by an appropriate choice of regularisation, as the one proposed hereafter. Studying the benefit of using a joint conditional prior is left as a future research direction.

With these assumptions, we eventually obtain $\mathcal{P}(u, \rho, \mathbf{S}) = \mathcal{P}(u)\mathcal{P}(\rho)\mathcal{P}(\mathbf{S})$. Maximisation of the probability (7) is then equivalent to minimisation of the neg-log-posterior $\mathcal{E}(u, \rho, \mathbf{S}) = -\log \mathcal{P}(u, \rho, \mathbf{S}|\mathbf{i}) - \log \mathcal{P}(\mathbf{i})$ which is given by:

$$\mathcal{E}(u, \rho, \mathbf{S}) = -\log \mathcal{P}(\mathbf{i}|u, \rho, \mathbf{S}) - \log \mathcal{P}(u) - \log \mathcal{P}(\rho) - \log \mathcal{P}(\mathbf{S}) \qquad (8)$$

where the first term is a data term, and the others are regularisation measures.

2.2   Continuous Variational Energy

Data Term. Lambert's law (6) does generally not hold perfectly: noise in the measurements will prevent the equality. We assume that Lambert's law residuals are iid Laplacian, with some outliers that will account for the shadows (unexplained low graylevels), the highlights (unexplained high graylevels) and the depth discontinuities (∇u not defined). We thus use a L^1-norm based data term that accounts for the sparsity of such effects:

$$\mathcal{E}_D(u, \rho, \mathbf{S}) = \sum_{i=1}^m \int_\Omega \left| I^i(\mathbf{x}_p) - \rho(\mathbf{x}_p)\mathbf{s}^i(\mathbf{x}_p) \cdot \mathbf{n}_u(\mathbf{x}_p) \right| \mathrm{d}\mathbf{x}_p = \sum_{i=1}^m \left\| I^i - \rho\mathbf{s}^i \cdot \mathbf{n}_u \right\|_{L^1(\Omega)} \quad (9)$$

Regularisations. In the 3D-reconstruction framework, surfaces are usually assumed to be differentiable, but this hypothesis is rarely realistic because of the presence of edges and depth discontinuities. At a macroscopic scale, surfaces should, more reasonably, be supposed to be smooth almost everywhere, with continuous, non-differentiable edges and some depth discontinuities. A regulariser that naturally allows for the corresponding class of depth functions is the total variation (TV) semi-norm:

$$J(u) = \int_\Omega \|\nabla u(\mathbf{x}_p)\| \, d\mathbf{x}_p = \|\nabla u\|_{L^1(\Omega)} \tag{10}$$

As suggested in literature [1,15], we restrict our study to the case of piecewise-smooth ("Arlequin-like") albedos, which can be enforced by a TV semi-norm as well:

$$J(\rho) = \int_\Omega \|\nabla \rho(\mathbf{x}_p)\| d\mathbf{x}_p = \|\nabla \rho\|_{L^1(\Omega)} \tag{11}$$

In the directional orthographic case, the problem is inherently ill-posed, but it is shown in [15] that TV-regularisation of both the depth and the albedo reduces the ambiguities to a simple translational ambiguity. The perspective case is better constrained (u can be recovered up to a scale ambiguity on $\mathcal{S}$ without regularisation [14]), so these regularisations would essentially enforce smoothness. Well-posedness is less clear in perspective case with spatially-varying lightings. Yet, it seems rather intuitive that, by limiting the variations of the lightings, the problem gets closer to the directional one, at least piecewise. This general discussion is left as future work, but it invites us to introduce a regularisation measure for the lightings as well. Since the depth discontinuities of both u and $\mathbf{s}^i$ will, in general, coincide, it is reasonable to assume that $\mathbf{s}^i$ should have bounded variations too, and to introduce its TV-regularisation:

$$J(\mathbf{s}^i) = \int_\Omega \|\mathbf{J}(\mathbf{s}^i)(\mathbf{x}_p)\|_F \, d\mathbf{x}_p = \|\mathbf{J}(\mathbf{s}^i)\|_{L^1(\Omega)}, \qquad i \in [1, m] \tag{12}$$

where $\mathbf{J}(\mathbf{s}^i)$ is the Jacobian matrix of $\mathbf{s}^i$ and $\|\cdot\|_F$ denotes the Frobenius norm. For the sake of compactness, we will denote $J(\mathbf{S}) = \sum_{i=1}^m J(\mathbf{s}^i)$. As stated earlier, the independence of $\mathbf{S}$ and u serves as a simplifying hypothesis, but it has no physical motivation: in future work, a joint regularisation of both these fields shall be introduced, using for instance a coupled L^1 semi-norm [8].

Energy. A continuous analogue of the MAP problem of minimising the neg-log posterior (8) can thus be written as the minimisation of $\mathcal{E}(u, \rho, \mathbf{S}) = \mathcal{E}_D(u, \rho, \mathbf{S}) + \alpha J(u) + \beta J(\rho) + \gamma J(\mathbf{S})$, with α, β, γ some positive weights. This functional is however not coercive in u, because of the scale ambiguity on $\mathcal{S}$ inherent to monocular perspective 3D-reconstruction. Coercivity can be forced by setting arbitrarily the depth in one point, or by adding a quadratic prior u_0:

$$\mathcal{E}(u, \rho, \mathbf{S}) = \sum_{i=1}^m \left\| I^i - \rho \, \mathbf{s}^i \cdot \mathbf{n}_u \right\|_{L^1} + \frac{\epsilon}{2} \|u - u_0\|_{L^2}^2 + \alpha J(u) + \beta J(\rho) + \gamma J(\mathbf{S}) \tag{13}$$

with $\epsilon > 0$ (u_0 can for instance be a prior on the mean camera-object distance).

3   Proximal Recovery Using Split-Bregman Iterations

We propose a proximal algorithm [5] for dealing with the TV semi-norms in (13), and introduce Split-Bregman iterations for decoupling the L^1 data term. Adopting the same notations as in [9] for the Bregman variables (d, b), whose components will be denoted (d^i, b^i), the augmented energy is then defined as:

$$\bar{\mathcal{E}}(u, \bar{u}, \rho, \bar{\rho}, \mathbf{S}, \bar{\mathbf{S}}, d, b) = \sum_{i=1}^{m} \left\| d^i \right\|_{L^1} + \frac{1}{2\theta_d} \sum_{i=1}^{m} \left\| d^i - \left(I^i - \bar{\rho}\, \bar{\mathbf{s}}^i \cdot \mathbf{n}_{\bar{u}} \right) - b^i \right\|_{L^2}^2 + \frac{\epsilon}{2} \left\| \bar{u} - u_0 \right\|_{L^2}^2$$

$$+ \frac{1}{2\theta_u} \|\bar{u} - u\|_{L^2}^2 + \alpha J(u) + \frac{1}{2\theta_\rho} \|\bar{\rho} - \rho\|_{L^2}^2 + \beta J(\rho) + \frac{1}{2\theta_{\mathbf{S}}} \left\| \bar{\mathbf{S}} - \mathbf{S} \right\|_{L^2}^2 + \gamma J(\mathbf{S}) \quad (22)$$

Algorithm 1 L^1-TV Perspective Photometric Stereo

Input: Images I^i, $i \in [1, m]$; camera parameters; initial light fields $\mathbf{s}^{i,0}$, $i \in [1, m]$; shape prior u_0; model parameters α, β, γ, ϵ; proximal parameters θ_d, θ_ρ, θ_u and $\theta_{\mathbf{S}}$; maximum number of iterations N.

Output: Point cloud $\mathbf{x}$, albedo ρ and updated light fields $\mathbf{s}^i$, $i \in [1, m]$.

Initialisation: Compute the normals and initialise the albedo ρ^0 by per-pixel classical photometric stereo; integrate the normals to initialise u^0; $\bar{\rho}^0 = \rho^0$, $\bar{u}^0 = u^0$; $d^{i,0} = b^{i,0} = 0$, $\bar{\mathbf{s}}^{i,0} = \mathbf{s}^{i,0}$, $i \in [1, m]$.

    for $k = 1 \dots N$ **do**

$$\bar{u}^{k+1} = \underset{\bar{u}}{\operatorname{argmin}} \frac{1}{2\theta_d} \sum_{i=1}^{m} \left\| d^{i,k} - \left(I^i - \bar{\rho}^k \bar{\mathbf{s}}^{i,k} \cdot \mathbf{n}_{\bar{u}} \right) - b^{i,k} \right\|_{L^2}^2 + \frac{\epsilon}{2} \|\bar{u} - u_0\|_{L^2}^2 + \frac{1}{2\theta_u} \left\| \bar{u} - u^k \right\|_{L^2}^2$$
$$(14)$$

$$\bar{\rho}^{k+1} = \underset{\bar{\rho}}{\operatorname{argmin}} \frac{1}{2\theta_d} \sum_{i=1}^{m} \left\| d^{i,k} - \left(I^i - \bar{\rho}\, \bar{\mathbf{s}}^{i,k} \cdot \mathbf{n}_{\bar{u}^{k+1}} \right) - b^{i,k} \right\|_{L^2}^2 + \frac{1}{2\theta_\rho} \left\| \bar{\rho} - \rho^k \right\|_{L^2}^2 \quad (15)$$

$$\bar{\mathbf{s}}^{i,k+1} = \underset{\bar{\mathbf{s}}^i}{\operatorname{argmin}} \frac{1}{2\theta_d} \left\| d^{i,k} - \left(I^i - \bar{\rho}^{k+1} \bar{\mathbf{s}}^i \cdot \mathbf{n}_{\bar{u}^{k+1}} \right) - b^{i,k} \right\|_{L^2}^2 + \frac{1}{2\theta_{\mathbf{S}}} \left\| \bar{\mathbf{s}}^i - \mathbf{s}^{i,k} \right\|_{L^2}^2, i \in [1, m]$$
$$(16)$$

$$d^{i,k+1} = \underset{d^i}{\operatorname{argmin}} \left\| d^i \right\|_{L^1} + \frac{1}{2\theta_d} \left\| d^i - \left(I^i - \bar{\rho}^{k+1} \bar{\mathbf{s}}^{i,k+1} \cdot \mathbf{n}_{\bar{u}^{k+1}} \right) - b^{i,k} \right\|_{L^2}^2, i \in [1, m]$$
$$(17)$$

$$b^{i,k+1} = b^{i,k} + \left(I^i - \bar{\rho}^{k+1} \bar{\mathbf{s}}^{i,k+1} \cdot \mathbf{n}_{\bar{u}^{k+1}} - d^{i,k+1} \right), i \in [1, m] \quad (18)$$

$$u^{k+1} = \underset{u}{\operatorname{argmin}} \frac{1}{2\theta_u} \left\| \bar{u}^{k+1} - u \right\|_{L^2}^2 + \alpha J(u) \quad (19)$$

$$\rho^{k+1} = \underset{\rho}{\operatorname{argmin}} \frac{1}{2\theta_\rho} \left\| \bar{\rho}^{k+1} - \rho \right\|_{L^2}^2 + \beta J(\rho) \quad (20)$$

$$\mathbf{s}^{i,k+1} = \underset{\mathbf{s}^i}{\operatorname{argmin}} \frac{1}{2\theta_{\mathbf{S}}} \left\| \bar{\mathbf{s}}^{i,k+1} - \mathbf{s}^i \right\|_{L^2}^2 + \gamma J(\mathbf{s}^i), i \in [1, m] \quad (21)$$

    end for
    $\mathbf{x} = \exp(u/f) \left[\mathbf{x}_p^\top / f, 1 \right]^\top$

---

for some small positive $\theta_d, \theta_u, \theta_\rho, \theta_{\mathbf{S}}$. We minimise the augmented energy (22) iteratively w.r.t. each variable, according to Algorithm 1.

3.1   Solutions of Convex Subproblems in Algorithm 1

The update (17) in d is a basis pursuit problem, which is solvable explicitely by shrinkage:

$$d^{i,k+1} = \text{shrink}(I^i - \bar{\rho}^{k+1}\bar{\mathbf{s}}^{i,k+1} \cdot \mathbf{n}_{\bar{u}^{k+1}} + b^{i,k}, 1/\theta_d), \; i \in [1, m] \qquad (23)$$

with $\text{shrink}(x, \gamma) = \text{sign}(x) \max\{|x| - \gamma, 0\}$. The update (18) in b is the Bregman update. Justification for this update can be found in [9]. It is the equivalent of "adding back the noise" in iterative image denoising methods [13].

   Regarding the updates in u, ρ and $\mathbf{S}$ (Eqs (19), (20), and (21)), they are instances of the L^2-TV problem which we solve using Chambolle's dual projection algorithm [4], with the natural vectorial extension presented in [3] in the case of $\mathbf{S}$. We experimentally noticed that the convergence rate is better when only performing a few projection iterations (typically 5) at each global iteration k, rather than solving these problems to full convergence (the same behaviour is observed in [9]). As for the update (15) in $\bar{\rho}$, writing the normal equations leads to the explicit update:

$$\bar{\rho}^{k+1} = \frac{\theta_d \rho^k + \theta_\rho \sum_{i=1}^{m}(I^i - d^{i,k} + b^{i,k})(\bar{\mathbf{s}}^{i,k} \cdot \mathbf{n}_{\bar{u}^{k+1}})}{\theta_d + \theta_\rho \sum_{i=1}^{m}(\bar{\mathbf{s}}^{i,k} \cdot \mathbf{n}_{\bar{u}^{k+1}})^2} \qquad (24)$$

   In a similar way, from the vectorial normal equations derived from (16), one obtains the explicit solution in $\bar{\mathbf{s}}^i$, $i \in [1, m]$:

$$\bar{\mathbf{s}}^{i,k+1} = \left(\theta_d \mathbf{I}_3 + \theta_{\mathbf{S}} \left(\bar{\rho}^{k+1}\right)^2 \mathbf{n}_{\bar{u}^{k+1}} \mathbf{n}_{\bar{u}^{k+1}}^\top\right)^{-1} \left(\theta_d \mathbf{s}^{i,k} + \theta_{\mathbf{S}} \bar{\rho}^{k+1}(I^i - d^{i,k} + b^{i,k})\mathbf{n}_{\bar{u}^{k+1}}\right) \; (25)$$

which can be computed e.g., through Cholesky factorisation.

3.2   Solution in $\bar{u}$

We now describe the minimisation in $\bar{u}$ with greater care. Let us denote:

$$\mathcal{E}_{\bar{u}}^k(\bar{u}) = \frac{1}{2\theta_d}\sum_{i=1}^{m} \left\| d^{i,k} - \left(I^i - \bar{\rho}^k\,\bar{\mathbf{s}}^{i,k} \cdot \mathbf{n}_{\bar{u}}\right) - b^{i,k} \right\|_{L^2}^2 + \frac{\epsilon}{2} \left\| \bar{u} - u_0 \right\|_{L^2}^2 + \frac{1}{2\theta_u} \left\| \bar{u} - u^k \right\|_{L^2}^2 \; (26)$$

Proposition 1. *The necessary optimality condition for $\bar{u}$ reads as the vanishing of the gradient of the energy* (26) *w.r.t. $\bar{u}$, which is given by:*

$$\nabla_{\bar{u}}\mathcal{E}_{\bar{u}}^k(\bar{u}) = \text{div}\left(\frac{f_{\bar{u}}^k}{\theta_d}\left(\nabla\bar{u} + (1 + \nabla\bar{u} \cdot \mathbf{u})\mathbf{u}\right)\right) + \text{div}\left(\frac{1}{\theta_d}\pi_{1,2}(\mathbf{g}_{\bar{u}}^k)\right)$$

$$- \text{div}\left(\frac{\pi_3\left(\mathbf{g}_{\bar{u}}^k\right)}{\theta_d}\mathbf{u}\right) + \left(\epsilon + \frac{1}{\theta_u}\right)\bar{u} - \epsilon u_0 - \frac{1}{\theta_u}u^k \qquad (27)$$

where $\pi_{1,2}$ and π_3 are the projections $\pi_{1,2}(a,b,c) = (a,b)$, $\pi_3(a,b,c) = c$, and

$$f_{\bar{u}}^k = \sum_{i=1}^{m} \frac{\tilde{I}^{i,k}\left(\tilde{I}^{i,k} + d^{i,k} - I^i - b^{i,k}\right)}{\|\nabla\bar{u}\|^2 + (1 + \nabla\bar{u} \cdot \mathbf{u})^2}, \mathbf{g}_{\bar{u}}^k = \sum_{i=1}^{m} \frac{\bar{\rho}^k\left(\tilde{I}^{i,k} + d^{i,k} - I^i - b^{i,k}\right)}{\sqrt{\|\nabla\bar{u}\|^2 + (1 + \nabla\bar{u} \cdot \mathbf{u})^2}}\bar{\mathbf{s}}^{i,k}$$

$$\mathbf{u} = \mathbf{x}_p/f, \qquad \tilde{I}^{i,k} = \bar{\rho}^k \bar{\mathbf{s}}^{i,k} \cdot \mathbf{n}_{\bar{u}} \qquad (28)$$

Proof. The directional derivative of $\mathbf{n}_{\bar{u}}$ at $\bar{u}$ in the direction w is given by:

$$d_w\mathbf{n}_{\bar{u}}(\bar{u}) = -\frac{\nabla w \cdot (\nabla\bar{u} + (1 + \nabla\bar{u} \cdot \mathbf{u})\mathbf{u})}{\|\nabla\bar{u}\|^2 + (1 + \nabla\bar{u} \cdot \mathbf{u})^2}\mathbf{n}_{\bar{u}} - \frac{1}{\sqrt{\|\nabla\bar{u}\|^2 + (1 + \nabla\bar{u} \cdot \mathbf{u})^2}}\begin{bmatrix} \nabla w \\ -\nabla w \cdot \mathbf{u} \end{bmatrix} \quad (29)$$

Let $D^i(\bar{u}) = \dfrac{1}{2\theta_d}\left\|d^{i,k} - (I^i - \bar{\rho}^k\,\bar{\mathbf{s}}^{i,k}\cdot\mathbf{n}_{\bar{u}}) - b^{i,k}\right\|_{L^2}^2$. Using the chain rule, its first variation $L^i(w)$ is given by:

$$L^i(w) = -\frac{1}{\theta_d}\int_{\Omega} f_{\bar{u}}^{i,k}\nabla w \cdot (\nabla\bar{u} + (1 + \nabla\bar{u} \cdot \mathbf{u})\mathbf{u}) - \frac{1}{\theta_d}\int_{\Omega}\mathbf{g}_{\bar{u}}^{i,k} \cdot \begin{bmatrix} \nabla w \\ -\nabla w \cdot \mathbf{u} \end{bmatrix} \quad (30)$$

where $f_{\bar{u}}^{i,k}$ and $\mathbf{g}_{\bar{u}}^{i,k}$ represent the terms inside the sums in (28). The first variation of (26) is thus given by:

$$L(w) = -\frac{1}{\theta_d}\int_{\Omega}\left(\sum_{i=1}^{m} f_{\bar{u}}^{i,k}\right)(\nabla\bar{u} + (1 + \nabla\bar{u} \cdot \mathbf{u})\mathbf{u}) \cdot \nabla w + \int_{\Omega}\left(\left(\epsilon + \frac{1}{\theta_u}\right)\bar{u} - \epsilon u_0 - \frac{1}{\theta_u}u^k\right)w$$

$$- \frac{1}{\theta_d}\int_{\Omega}\left[\pi_{1,2}\left(\sum_{i=1}^{m}\mathbf{g}_{\bar{u}}^{i,k}\right) - \pi_3\left(\sum_{i=1}^{m}\mathbf{g}_{\bar{u}}^{i,k}\right)\mathbf{u}\right] \cdot \nabla w \quad (31)$$

Using Dirichlet boundary conditions on $\bar{u}$, application of the Green formula eventually provides the result announced in (27). □

This ressembles the steady state of a reaction-diffusion equation. The data term of (26) being nonlinear and non-convex in $\bar{u}$, there is no guarantee for the negativity of $f_{\bar{u}}^k$. For this reason, we choose a descent solver with a quasi-Newton step (BFGS) in the implementation. We now evaluate experimentally the benefit of using the proposed approach.

4   Experiments

Robustness to Noise and Outliers. We first evaluate the ability of the proposed proximal algorithm to enhance the results of classical photometric stereo. We create 20 images, of size 128×128, of a "Canadian tent" shape with a "pears" albedo, illuminated from $m = 20$ known lightings, and corrupt them simultaneously by an additive zero-mean Gaussian noise with standard deviation equal to $\sigma\%$ of the maximum graylevel, and a salt-and-pepper noise affecting $p\%$ of the pixels. We use the method from [10] to estimate the initial albedo ρ^0 and the normals which minimise the L^1 data term (9). Up to this point,
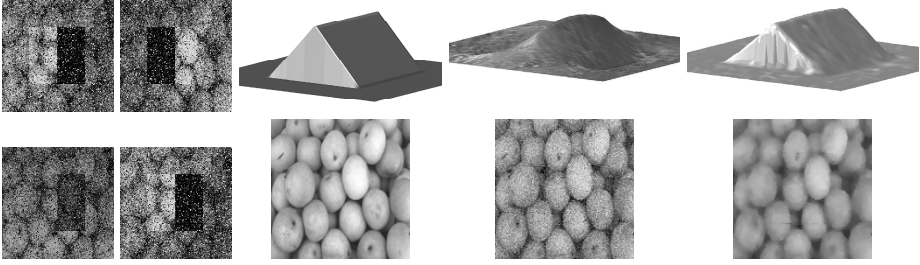
Fig. 1. Robustness to noise and outliers. First and second columns: 4 of the $m = 20$ images, with $\sigma = 20\%$ and $p = 10\%$; Third column: ground-truth depth map and albedo; Fourth column: results from [10], using a least squares integrator with homogeneous Dirichlet BC; Fifth column: results after L^1-TV refinement. State-of-the-art methods provide a rather noisy albedo, and the quality of the final depth map is widely dependent from the choice of a specific integrator. Rather than choosing a robust integrator, as advised e.g. in [6], we believe that it makes more sense to use the proposed scheme to directly refine the depth map and the albedo from the images, so as to avoid the propagation of any bias due to normal estimation. Indeed, rather than seeking a piecewise-smooth shape explaining a possibly biased normal field, it makes more sense to look for a piecewise-smooth shape explaining the images themselves.

in view of the piecewise smooth nature of the shape, robust integration should be considered to obtain a depth map [6]. To emphasise the ability of our scheme to recover such shapes as well, we integrate the normals using the DCT solver from [16] instead, with homogeneous Dirichlet BC, before iteratively refining the depth and the albedo through the proposed L^1-TV scheme (Figure 1): this allows recovering the depth discontinuities, while simultaneously denoising the albedo. We used $\theta_d = 0.1$, $\theta_u = 0.1\theta_d$, $\theta_\rho = 0.001\theta_u$, $\theta_{\mathbf{S}} = \theta_\rho$, $\epsilon = 10^{-8}$, $u_0 = 0$, $\alpha = 2/\theta_u$, $\beta = 5\alpha$, and $\gamma = \alpha$.

Influence of the Parameters. Since the proposed proximal algorithm involves numerous parameters, it is natural to question their influence. Let us first make a distinction between the model parameters α, β, γ and ϵ, and the optimisation parameters $\theta_\star$. Those latter will mainly affect the convergence rate (though, since we consider nonconvex optimisation, they may also affect the results), while the choice of the former corresponds to some *a priori* knowledge about the regularity of the unknowns. If we choose high values for α, β, γ, we will get oversmoothed results, while the 3D-reconstruction may not be robust to noise, or not well-conditioned, if we choose low values. The parameter ϵ only enforces coercivity, thus any low value should be satisfactory (we set $\epsilon = 10^{-8}$ in all the experiments). As for the optimisation parameters, setting high $\theta_\star$ values will result in slow convergence, while the optimisation may become unstable with low values. In our tests, we experimented no major difficulty in finding a set of these numerous parameters offering "reasonable" results. Tuning more precisely
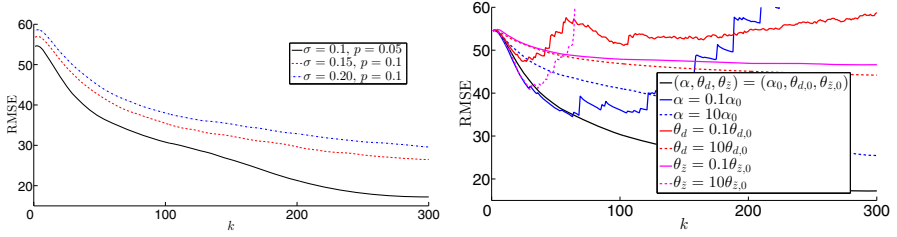
Fig. 2. Quantitative evaluation of the diffusion process. Left: RMSE in u as a function of the number k of iterations, for several levels of noise (σ, p), with the same reference values of the parameters as in the first test, denoted $(\alpha_0, \theta_{d,0}, \theta_{u,0})$ in the case of $(\alpha, \theta_d, \theta_u)$. Right: influence of the model parameter α and of the optimisation parameters θ_d and θ_u on the diffusion process. Convergence is stable but slow if α or θ_d is too high (oversmoothed depth map), or if θ_u is too high (small descent steps). On the contrary, the descent is too fast to be stable if we set a too low value for α or θ_d (no regularisation, inducing numerical difficulties), or if θ_u is too low (large descent steps).

one or the other parameter can be a little more tricky, as illustrated in Figure 2. We used the same data as in the previous test, with $\sigma = 0.1$ and $p = 0.05$, for studying the evolution of the RMSE in u (the scale ambiguity on $\mathcal{S}$ is a posteriori solved to minimise this RMSE) with respect to a specific choice of α, θ_d or θ_u (the
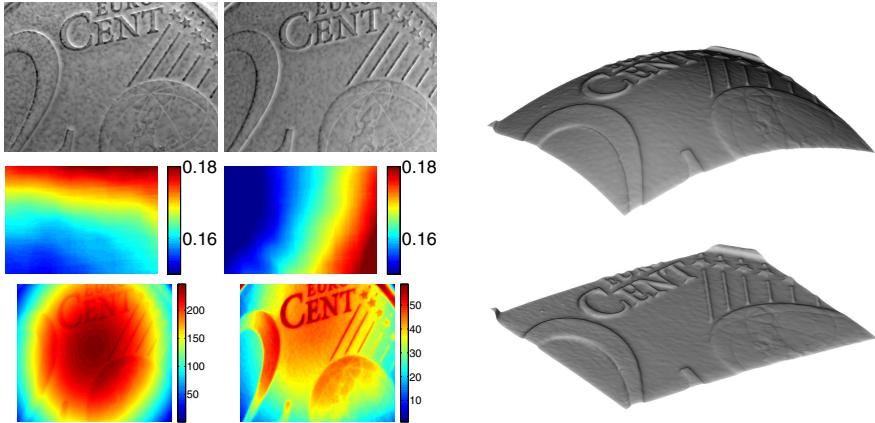

Fig. 3. Bias correction by using spatially-varying lightings. Left, from top to bottom: two of the $m = 15$ input images; associated calibrated lighting intensities; estimated log depth maps using least-squares photometric stereo with a directional model and with the calibrated spatially-varying lightings. Right: relighted views of the estimated sets of 3D-points, using directional (top) or spatially-varying lightings (bottom). Neglecting the spatial variations of the lightings creates a global drift in the reconstruction, which is drastically reduced by considering spatially-varying lightings. This 3D-reconstruction could then be further improved using the proposed L^1-TV scheme.
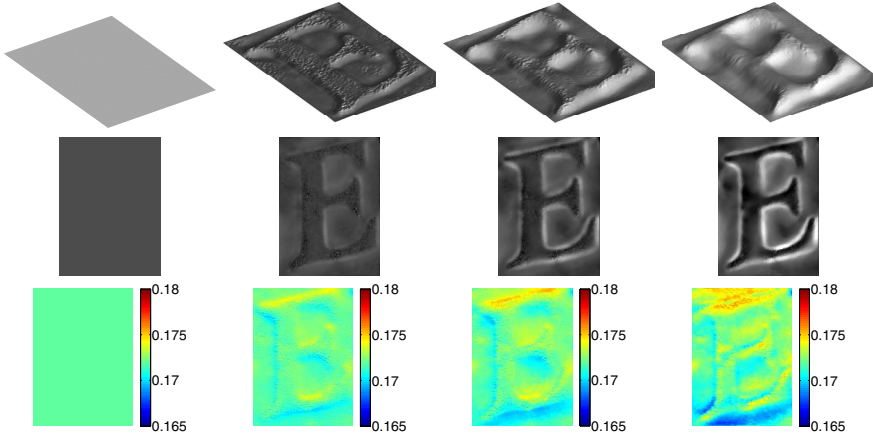
Fig. 4. L^1-TV optimisation with rough initialisation. We show the diffusion process of u (top), ρ (middle) and $\|\mathbf{s}^1\|$ (bottom), at $k = 0$, $k = 10$, $k = 20$ and $k = 40$. The sharp features are first captured by the L^1 data term, and then the TV proximal operators smooth out the residual noise. Realistic estimation can be provided, even with a very rough initialisation.

other parameters behave similarly). Finding heuristics for automatic selection of the parameters, based on these observations, is left as future work.

Experimental Setup and Lighting Initialisation. For tests on real-world data, we used a calibrated camera surrounded by twenty LEDs oriented towards a white-painted two cents coin. The LEDs are only controllable by groups of 5, providing 15 different lighting configurations with either 5, 10, 15 or 20 LEDs turned on together. Thanks to the proposed spatially-varying lightings, we do not need model-driven calibration of the lightings: to initialise the light fields, we used a planar calibration grid composed of regularly located white truncated pyramids, before interpolating them so as to obtain dense light fields.

   We first evaluate the benefit of using spatially-varying lightings, by comparing the results of the classical two-steps approach, using either directional lightings or the calibrated spatially-varying ones (least-squares estimations are used). Results shown in Figure 3 prove that the widely-spread idea according to which photometric stereo can recover high-frequency details, but not low-frequency ones is wrong: it is only a matter of correctly estimating the lightings.

L^1-TV Refinement from Rough Initialisation. We now question the importance of the initialisation for the full (shape, reflectance and lightings) proximal recovery, regarding the non-convexity of the energy. Due to the computational cost required to perform one Cholesky factorisation per pixel at each iteration in the lightings estimation, we consider a small close-up on the "E" letter of the coin,

with size 201×146 pixels. The evolution of the anisotropic diffusion process, starting from a flat shape with uniform albedo and directional lightings, is illustrated in Figure 4: it proves that the proposed proximal algorithm converges towards a realistic solution even if a rough initialisation is considered. We used $\theta_d = 0.05$, $\theta_u = 100\theta_d$, $\theta_\rho = \theta_{\tilde{d}}$, $\theta_{\mathbf{S}} = 0.1\theta_d$, $\epsilon = 10^{-8}$, $\alpha = 0.05/\theta_u$, $\beta = 0.05/\theta_\rho$, $u_0 = 0$ and $\gamma = 0.0001/\theta_S$.

5   Conclusion and Perspectives

We proposed a variational formulation for the joint recovery of shape, reflectance and spatially-varying lightings from observed images, derived from a formal MAP approach. Despite the apparent difficulty of tackling a non-convex problem, robust recovery of the depth can be obtained without integrating the normals. By considering the lightings as unknown vector fields, situations usually considered as "hard", like nearby or extended sources, are easily handled. In future work, we will explore joint regularisation of the lightings, the depth and the albedo and study with more care the well-posedness of the inverse problem, the robustness with respect to the numerous parameters, and more efficient numerics.

References

1. Alldrin, N.G., Mallick, S.P., Kriegman, D.J.: Resolving the generalized bas-relief ambiguity by entropy minimization. In: CVPR (2007)
2. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. IJCV **72**(3), 239–257 (2007)
3. Bresson, X., Chan, T.: Fast dual minimization of the vectorial total variation norm and applications to color image processing. IPI **2**(4), 455–484 (2008)
4. Chambolle, A.: An algorithm for total variation minimization and applications. JMIV **20**(1–2), 89–97 (2004)
5. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. JMIV **40**(1), 120–145 (2011)
6. Durou, J.-D., Aujol, J.-F., Courteille, F.: Integrating the normal field of a surface in the presence of discontinuities. In: Cremers, D., Boykov, Y., Blake, A., Schmidt, F.R. (eds.) EMMCVPR 2009. LNCS, vol. 5681, pp. 261–273. Springer, Heidelberg (2009)
7. Georghiades, A.: Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In: ICCV (2003)
8. Goldluecke, B., Strekalovskiy, E., Cremers, D.: The natural vectorial total variation which arises from geometric measure theory. SIIMS **5**(2), 537–563 (2012)
9. Goldstein, T., Osher, S.: The Split Bregman method for L1-regularized problems. SIIMS **2**(2), 323–343 (2009)
10. Ikehata, S., Wipf, D., Matsushita, Y., Aizawa, K.: Robust photometric stereo using sparse regression. In: CVPR (2012)

11. Mecca, R., Tankus, A., Wetzler, A., Bruckstein, A.M.: A direct differential app-
    roach to photometric stereo with perspective wiewing. SIIMS **7**(2), 579–612 (2014)
12. Mumford, D.: Bayesian rationale for the variational formulation. In: Geometry-
    Driven Diffusion in Computer Vision, pp. 135–146 (1994)
13. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization
    method for total variation-based image restoration. MMS **4**(2), 460–489 (2005)
14. Papadhimitri, T., Favaro, P.: A new perspective on uncalibrated photometric
    stereo. In: CVPR (2013)
15. Quéau, Y., Lauze, F., Durou, J.D.: Solving uncalibrated photometric stereo using
    total variation. To appear in JMIV (2015)
16. Simchony, T., Chellappa, R., Shao, M.: Direct analytical methods for solving Pois-
    son equations in computer vision problems. PAMI **12**(5), 435–446 (1990)
17. Woodham, R.J.: Photometric method for determining surface orientation from
    multiple images. Optical Engineering **19**(1), 139–144 (1980)