JMIV manuscript No.
(will be inserted by the editor)

Variational Reflectance Estimation from Multi-view Images

Jean Mélou[1,2] · **Yvain Quéau**[3] · **Jean-Denis Durou**[1] · **Fabien Castan**[2] · **Daniel Cremers**[3]

Received: date / Accepted: date

Abstract We tackle the problem of reflectance estimation from a set of multi-view images, assuming known geometry. The approach we put forward turns the input images into reflectance maps, through a robust variational method. The variational model comprises an image-driven fidelity term and a term which enforces consistency of the reflectance estimates with respect to each view. If illumination is fixed across the views, then reflectance estimation remains under-constrained: a regularization term, which ensures piecewise-smoothness of the reflectance, is thus used. Reflectance is parameterized in the image domain, rather than on the surface, which makes the numerical solution much easier, by resorting to an alternating majorization-minimization approach. Experiments on both synthetic and real datasets are carried out to validate the proposed strategy.

Keywords Reflectance · Multi-view · Shading · Variational Methods.

Jean Mélou
E-mail: jeme@mikrosimage.eu

Yvain Quéau
E-mail: yvain.queau@tum.de

Jean-Denis Durou
E-mail: durou@irit.fr

Fabien Castan
E-mail: faca@mikrosimage.eu

Daniel Cremers
E-mail: cremers@tum.de

[1]IRIT, UMR CNRS 5505, Université de Toulouse, France
[2]Mikros Image, Levallois-Perret, France
[3]Department of Computer Science, Technical University of Munich, Garching, Germany

1 Introduction

Acquiring the shape and the reflectance of a scene is a key issue, *e.g.*, for the movie industry, as it allows proper relighting. Well-established shape acquisition techniques such as multi-view stereo exist for accurate 3D-reconstruction. Nevertheless, they do not aim at recovering the surface reflectance. Hence, the original input images are usually mapped onto the 3D-reconstruction as texture. Since the images mix shading information (induced by lighting and geometry) and reflectance (which is characteristic of the surface), relighting based on this approach usually lacks realism. To improve the results, reflectance needs to be separated from shading.

In order to more precisely illustrate our purpose, let us take the example of a Lambertian surface. In a 2D-point (pixel) $\mathbf{p}$ conjugate to a 3D-point $\mathbf{x}$ of a Lambertian surface, the graylevel $I(\mathbf{p})$ is usually written

$$I(\mathbf{p}) = \rho(\mathbf{x})\,\mathbf{s}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}). \tag{1}$$

In the right-hand side of (1), $\rho(\mathbf{x}) \in \mathbb{R}$ is the albedo[1], $\mathbf{s}(\mathbf{x}) \in \mathbb{R}^3$ is the lighting vector, and $\mathbf{n}(\mathbf{x}) \in \mathbb{S}^2 \subset \mathbb{R}^3$ is the outer unit-length normal to the surface. All these elements a priori depend on $\mathbf{x}$ *i.e.*, they are defined locally. Whereas I is always supposed to be given, different situations can occur, according to which are also known, among ρ, $\mathbf{s}$ and $\mathbf{n}$.

One equation (1) per pixel is not enough to simultaneously estimating the reflectance ρ, the lighting $\mathbf{s}$ and the geometry, represented here by $\mathbf{n}$, because there are much more unknowns than equations. Figure 1 illustrates this source of ill-posedness through the so-called "workshop metaphor" introduced by Adelson and Pentland in [1].

[1] Since the albedo suffices to characterize the reflectance of a Lambertian surface, we will name it "reflectance" as well.

(a) (b) (c) (d)

Fig. 1 The "workshop metaphor" (extracted from a paper by Adelson and Pentland [1]). Image (a) may be interpreted either by: (b) incorporating all the brightness variations inside the reflectance; (c) modulating the lighting of a white planar surface; (d) designing a uniformly white 3D-shape. This last interpretation is a solution of the shape-from-shading problem.

Photometric 3D-reconstruction usually assumes that the lighting $\mathbf{s}(\mathbf{x})$ is known. Still, there remains three scalar unknowns per equation (1): $\rho(\mathbf{x})$ and $\mathbf{n}(\mathbf{x})$, which has two degrees of freedom. Shape-from-shading [16] uses the shading $\mathbf{s}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})$ as unique clue to recover the shape $\mathbf{n}$, assuming known reflectance ρ, but the problem is still ill-posed. A way to make photometric 3D-reconstruction well-posed is to use $m > 1$ images taken using a single camera pose, but under varying lighting:

$$I^i(\mathbf{p}) = \rho(\mathbf{x})\, \mathbf{s}^i(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}), \quad i \in \{1, \dots, m\} \tag{2}$$

In this variant of shape-from-shading called "photometric stereo" [37], the reflectance $\rho(\mathbf{x})$ and the normal $\mathbf{n}(\mathbf{x})$ can be estimated without any ambiguity, as soon as $m \geq 3$ non-coplanar lighting vectors $\mathbf{s}^i(\mathbf{x})$ are used.

Symmetrically to (2), solving the problem:

$$I^i(\mathbf{p}) = \rho(\mathbf{x})\, \mathbf{s}(\mathbf{x}) \cdot \mathbf{n}^i(\mathbf{x}), \quad i \in \{1, \dots, m\} \tag{3}$$

allows to estimate the lighting $\mathbf{s}(\mathbf{x})$, as soon as the reflectance $\rho(\mathbf{x})$ and $m \geq 3$ non-coplanar normals $\mathbf{n}^i(\mathbf{x})$ are known. This can be carried out, for instance, by placing a small calibration pattern with known color and known shape near each 3D-point $\mathbf{x}$ [31].

The problem we aim at solving in this paper is slightly different. Suppose we are given a series of m images of a scene taken using a single lighting, but m camera poses. According to Lambert's law, this ensures that a 3D-point looks equally bright in all the images where it is visible. Such invariance is the basic clue of multi-view stereo (MVS), which has become a very popular technique for 3D-reconstruction [12]. Therefore, since an estimate of the surface 3D-shape is available, $\mathbf{n}$ is known. Now, we have to index the pixels by the image number i. Fortunately, additional data provided by MVS are the correspondences between the different views, taking the form of m-tuples of pixels $(\mathbf{p}^i)_{i \in \{1,\dots,m\}}$ which are conjugate to a common 3D-point $\mathbf{x}$.

Our problem is written[2]:

$$I^i(\mathbf{p}^i) = \rho(\mathbf{x})\, \mathbf{s}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}), \quad i \in \{1, \dots, m\} \tag{4}$$

where $\mathbf{p}^i$ is the projection of $\mathbf{x}$ in the i-th image, and $\rho(\mathbf{x})$ and $\mathbf{s}(\mathbf{x})$ are unknown. Obviously, this system reduces to Equation (1), since its m equations are the same one: the right-hand side of (4) does not depend on i, not more than the left-hand side $I^i(\mathbf{p}^i)$ since, as already noticed, the lighting $\mathbf{s}$ does not vary from one image to another, and the surface is Lambertian.

Multi-view helps estimating the reflectance, because it provides the 3D-shape via MVS. However, even if $\mathbf{n}(\mathbf{x})$ is known, Equation (1) remains ill-posed. This is illustrated, in Figure 1, by the solutions (b) and (c), which correspond to the same image (a) and to the same planar surface. In the absence of any prior, Equation (1) has an infinity of solutions in $\rho(\mathbf{x})\, \mathbf{s}(\mathbf{x})$. In addition, determining $\rho(\mathbf{x})$ from each of these solutions gives rise to another ambiguity, since $\mathbf{s}(\mathbf{x})$ is not forced to be unit-length.

Such a double source of ill-posedness probably explains why various methods for reflectance estimation have been designed, introducing a variety of prior in order to disambiguate the problem. Most of them assume that brightness variations induced by reflectance changes are likely to be strong, but sparsely distributed, while the lighting is likely to induce smoother changes. This suggests to separate a single image into a piecewise smooth layer and a more oscillating one. In the computer vision literature, this is often referred to as "intrinsic image decomposition", while the terminology "cartoon + texture decomposition" is more frequently used by the mathematical imaging community.

[2] Even if they look very similar, Problems (2), (3) and (4) have completely different peculiarities.

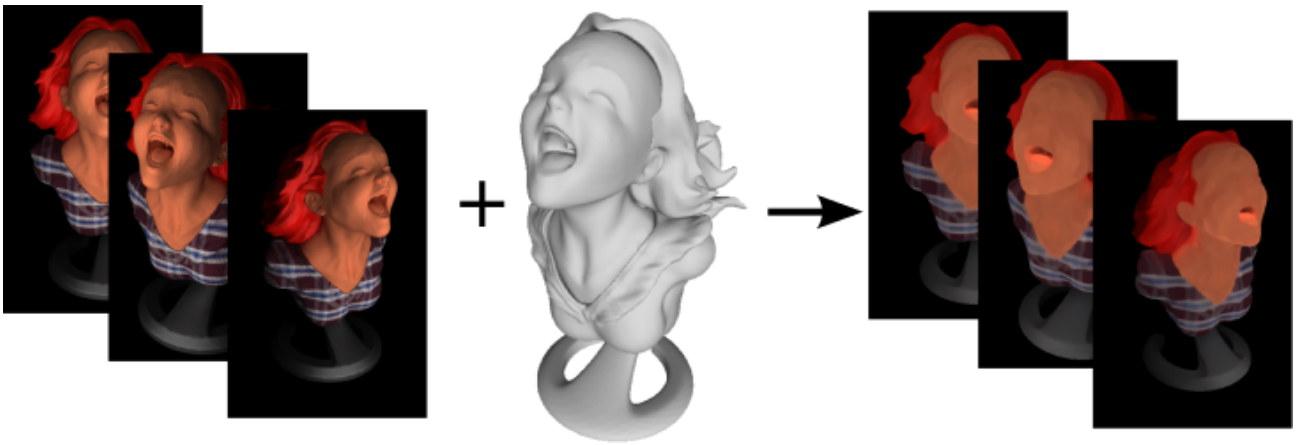


Fig. 2 Overview of our contribution. From a set of n images of a surface acquired under different angles, and a coarse geometry obtained for instance using multi-view stereo, we estimate a shading-free reflectance map per view.

Contributions. In this work, we show the relevance of using multi-view images for reflectance estimation. Indeed, this enables a prior shape estimation using MVS, which essentially reduces the decomposition problem to the joint estimation of a set of reflectance maps, as illustrated in Figure 2. We elaborate on the variational approach to multi-view shading-reflectance decomposition which we initially presented in [26]. The latter introduced a robust l^1-TV framework for the joint estimation of piecewise-smooth reflectance maps and of spherical harmonics lighting, with an additional term ensuring the consistency of the reflectance maps. The present paper extends this approach by developing the theoretical foundations of this variational model. In this view, our parameterization choices are further discussed and the underlying ambiguities are exhibited. The variational model is motivated by a Bayesian rationale, and the proposed numerical scheme is interpreted in terms of a majorization-minimization algorithm. Finally, we conclude that varying the lighting along with the viewing angle, in the spirit of photometric stereo, is the only way to estimate the reflectance without resorting to any prior.

Organization of the Paper. After reviewing related approaches in Section 2, we formalize in Section 3 the problem of multi-view reflectance estimation. Section 4 then introduces a Bayesian-to-variational approach to this problem. A simple numerical strategy for solving the resulting variational problem, which is based on alternating majorization-minimization, is presented in Section 5. Experiments on both synthetic and real data are then conducted in Section 6, before summarizing our achievements and suggesting future research directions in Section 7.

2 Related Works

Studied since the 1970s [21], the problem of decomposing an image (or a set of images) into a piecewise-smooth component (cartoon-like reflectance) and an oscillatory one (texture, or shading) is a fundamental computer vision problem, which has been addressed in numerous ways.

Cartoon + Texture Decomposition. Researchers in the field of mathematical imaging have suggested various variational models for this task, using for instance non-smooth regularization and Fourier-based frequency analysis [3], or l^1-TV variational models [23]. However, such techniques do not use an explicit photometric model for justifying the decomposition. On the other hand, photometric analysis is another important branch of computer vision, which may be a source of inspiration for motivating new variational models.

Photometric Stereo. As discussed in the introduction, photometric stereo techniques [37] are able to unambiguously estimate the reflectance and the geometry, by considering several images obtained from the same viewing angle but under calibrated, varying lighting. Photometric stereo has even been extended to the case of uncalibrated, varying lighting [5]. As in uncalibrated photometric stereo, our goal is to estimate reflectance under unknown lighting. However, in our case, we cannot ensure that lighting is varying, so the problem is less constrained. Our hope is that this can be somewhat compensated by the prior knowledge of geometry, and by the resort to appropriate priors. Various priors for reflectance have been discussed in the context of intrinsic image decomposition.

Intrinsic Image Decomposition. Separating reflectance from shading in a single image is a challenging problem, often referred to as intrinsic image decomposition. Given the ill-posed nature of this problem, prior information on shape, reflectance and lighting must be introduced. Most of the existing works are based on the Retinex theory [21], which states that most of the slight brightness variations in an image are due to lighting, while the reflectance is piecewise-constant (*e.g.*, a Mondrian image). A variety of clustering-based [13,34] or sparsity-enhancing methods [4,14,28,34,35] have been developed based on this theory. However, most of them suffer from the fundamental ambiguity of shape-from-shading, since the geometry of the scene is unknown. Some other methods disambiguate the problem by requiring the user to "brush" uniform reflectance parts [8,28], or by resorting to a crowdsourced database [7]. Still, these works require user interactions, which may not be desirable in certain cases.

Multi-view 3D-reconstruction. Instead of introducing possibly unreliable priors on shape, or relying on user interactions, ambiguities can be reduced by assuming that the geometry of the scene is known. Intrinsic image decomposition has for instance been addressed using an RGB-D camera [9] or, closer to our proposal, multiple views of the same scene under different angles [19,20]. In the latter works, the geometry is first extracted from the multi-view images, before the problem of reflectance estimation is addressed. Geometry computation can be achieved using multi-view stereo [33]. MVS techniques have seen significant growth over the last decade, an expansion which goes hand in hand with the development of structure-from-motion (SfM) solutions [27]. Indeed, MVS requires the parameters of the cameras, outputs of the SfM algorithm. Nowadays, these mature methods are commonly used in uncontrolled environments, or even with large scale Internet data [2]. For the sake of completeness, let us also mention that some efforts in the direction of multi-view and photometrically consistent 3D-reconstruction have been devoted recently [17, 18,22,24,25]. Similar to these methods, we will resort to a compact representation of lighting, namely the spherical harmonics model.

Spherical Harmonics Lighting Model. Let us consider a point $\mathbf{x}$ lying on the surface $\mathcal{S} \subset \mathbb{R}^3$ of the observed scene, and let $\mathbf{n}(\mathbf{x})$ be the unit-length outward normal vector to $\mathcal{S}$ in $\mathbf{x}$. Let $\mathcal{H}(\mathbf{x})$ be the hemisphere centered in $\mathbf{x}$, having as basis plane the tangent plane to $\mathcal{S}$ in $\mathbf{x}$. Each light source visible from $\mathbf{x}$ can be associated to a point ω on $\mathcal{H}(\mathbf{x})$. If we describe by the vector $\mathbf{s}(\mathbf{x},\omega)$ the corresponding elementary light beam (oriented towards the source), then the luminance of $\mathbf{x}$ in the direction $\mathbf{v}$ is given by

$$L(\mathbf{x},\mathbf{v}) = \frac{1}{\pi}\int_{\mathcal{H}(\mathbf{x})} r(\mathbf{x},\mathbf{n}(\mathbf{x}),\mathbf{s}(\mathbf{x},\omega),\mathbf{v}) \max\{0,\mathbf{s}(\mathbf{x},\omega)\cdot\mathbf{n}(\mathbf{x})\}\,\mathrm{d}\omega, \quad (5)$$

where r is the "reflectance" (BRDF) of the surface which, in general, depends on the viewing direction $\mathbf{v}$, and the rightest factor in the integral is the surface illuminance, or shading (the max operator encodes self-shadows).

This expression of the luminance is intractable in the general case. However, if we restrict our attention to Lambertian surfaces, then the reflectance reduces to the albedo $\rho(\mathbf{x})$, which is independent of the viewing direction $\mathbf{v}$. If the light sources are further assumed to be distant enough from the object, then $\mathbf{s}(\mathbf{x},\omega)$ is independent of $\mathbf{x}$ *i.e.*, the light beams are the same for the whole (supposedly convex) object, and thus the lighting is completely defined on the unit sphere. Therefore, the integral (5) acts as a convolution on $\mathcal{H}(\mathbf{x})$, having as kernel $\max\{0,\mathbf{s}(\omega)\cdot\mathbf{n}\}$. Spherical harmonics, which can be considered as the analogue to the Fourier series on the unit sphere, have been shown to be an efficient low-dimensional representation of this convolution [6, 32]. Many vision applications [18,38] use second order spherical harmonics, which can captured over 99% of the natural lighting [11] using only nine coefficients. This yields an approximation of the luminance of the form

$$L = \rho\,\boldsymbol{\sigma}\cdot\boldsymbol{\nu}, \quad (6)$$

where $\rho \in \mathbb{R}$ is the reflectance, $\boldsymbol{\sigma} \in \mathbb{R}^9$ is a compact lighting representation, and $\boldsymbol{\nu} \in \mathbb{R}^9$ stores the local geometric information. The latter is deduced from the normal according to:

$$\boldsymbol{\nu} = \begin{bmatrix} \mathbf{n} \\ 1 \\ n_1\,n_2 \\ n_1\,n_3 \\ n_2\,n_3 \\ n_1^2 - n_2^2 \\ 3n_3^2 - 1 \end{bmatrix}. \quad (7)$$

In (6), the lighting vector $\boldsymbol{\sigma}$ is the same in all the points of the surface, but the reflectance ρ and the geometric vector $\boldsymbol{\nu}$ vary along the surface $\mathcal{S}$ of the observed scene. Hence we will write (6) as:

$$L(\mathbf{x}) = \rho(\mathbf{x})\,\boldsymbol{\sigma}\cdot\boldsymbol{\nu}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{S}. \quad (8)$$

Our aim in this paper is to estimate the reflectance $\rho(\mathbf{x})$ in each point $\mathbf{x} \in \mathcal{S}$, as well as the lighting vector $\boldsymbol{\sigma}$, given a set of multi-view images and the geometry $\boldsymbol{\nu}$. We formalize this problem in the next section.

3 Multi-view Reflectance Estimation

In this section, we describe with more care the problem of reflectance estimation from a set of multi-view images. First, we need to explicit the relationship between graylevel, reflectance, lighting and geometry.

3.1 Image Formation Model

Let $\mathbf{x} \in \mathcal{S}$ be a point on the surface of the scene. Assume that it is observed by a graylevel camera with linear response function and let $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ be the image, where Ω is the projection of $\mathcal{S}$ onto the image plane. Then, the graylevel in the pixel $\mathbf{p} \in \Omega$ conjugate to $\mathbf{x}$ is proportional to the luminance of $\mathbf{x}$ in the direction of observation $\mathbf{v}$:

$$I(\mathbf{p}) = \gamma\, L(\mathbf{v}, \mathbf{x}), \tag{9}$$

where the coefficient $\gamma > 0$, referred to in the following as the "camera coefficient", is unknown[3]. By assuming Lambertian reflectance and the light sources distant enough from the object, Equations (8) and (9) yield:

$$I(\mathbf{p}) = \gamma\, \rho(\mathbf{x})\, \boldsymbol{\sigma} \cdot \boldsymbol{\nu}(\mathbf{x}). \tag{10}$$

Now, let us assume that m images I^i, $i \in \{1, \dots, m\}$ of the surface, obtained while moving a single camera, are available, and discuss how to adapt (10).

Case 1: unknown, yet fixed lighting and camera coefficient. If all the automatic settings of the camera are disabled, then the camera coefficient is independent from the view *i.e.*, $\gamma^i = \gamma$. If the illumination is also fixed, then the lighting vectors are independent from the view *i.e.*, $\boldsymbol{\sigma}^i = \boldsymbol{\sigma}$. We can thus incorporate the camera coefficient into the lighting vector: $\boldsymbol{\sigma} := \gamma\, \boldsymbol{\sigma}$. If the point $\mathbf{x}$ is visible in the i-th view, Equation (10) becomes:

$$I^i(\pi^i(\mathbf{x})) = \rho(\mathbf{x})\, \boldsymbol{\sigma} \cdot \boldsymbol{\nu}(\mathbf{x}), \tag{11}$$

where we denote by π^i the 3D-to-2D projection associated to the i-th view. In (11), the unknowns are the reflectance $\rho(\mathbf{x})$ and the lighting vector $\boldsymbol{\sigma}$. Equations (11), $i \in \{1, \dots, m\}$, constitute a generalization of (4) to more complex illumination scenarios. For the whole scene, this is a problem with $n + 9$ unknowns and up to mn equations, where n is the number of 3D-points $\mathbf{x}$ which have been estimated by multi-view stereo. However, as for System (4), only n equations are linearly independent, hence the problem of reflectance and lighting estimation is under-constrained.

[3] This coefficient depends on several factors such as the lens aperture, the magnification, the exposure time, etc.

Case 2: unknown and varying lighting and camera coefficient. If lighting is varying, then we have to make the lighting vector view-dependent. The camera coefficient can also be assumed to vary and be integrated to the lighting vector *i.e.*, $\boldsymbol{\sigma}^i := \gamma^i\, \boldsymbol{\sigma}^i$, since the estimation of each $\boldsymbol{\sigma}^i$ will include that of each γ^i. Equation (10) becomes:

$$I^i(\pi^i(\mathbf{x})) = \rho(\mathbf{x})\, \boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}(\mathbf{x}), \tag{12}$$

There are even more unknowns $(n + 9m)$, but this time the mn equations are not linearly dependent, at least as long as the $\boldsymbol{\sigma}^i$ are not proportional *i.e.*, if not only the camera coefficient or the lighting intensity vary across the views, but also the lighting direction[4]. Typically, n is of the order of $[10^3, 10^6]$, hence the problem is over-constrained as soon as at least two out of the m lighting vectors are non-collinear. This is a situation similar to uncalibrated photometric stereo [5], but much more favorable: the geometry is known, hence the ambiguities arising in uncalibrated photometric stereo are likely to be reduced. However, contrarily to uncalibrated photometric stereo, lighting is not actively controlled in our case. Lighting variations are likely to happen *e.g.*, in outdoor scenarios, yet they will be limited. The m lighting vectors $\boldsymbol{\sigma}^i$, $i \in \{1, \dots, m\}$, will thus be close to each other: lighting variations will not be sufficient in practice for disambiguation (ill-conditioning).

Since (11) is under-constrained and (12) is ill-conditioned, additional information will have to be introduced either ways, and we can restrict our attention to the varying lighting case (12).

3.2 Extension to RGB Images

So far, we have assumed that graylevel images were available. To extend our study to RGB images, we must replace the camera coefficient γ in (9) by the transmission spectrum $c_\star(\lambda)$ of the camera, which depends both on the channel $\star \in \{R, G, B\}$ and on the wavelength λ. Besides, the spectral dependency of the reflectance and of the lighting intensity must be considered in the definition (5) of the luminance, which must then be integrated over the visible spectrum. Altogether, assuming Lambertian reflectance and distant sources, we get the following image formation model in channel $\star$, for a

[4] Another case, which we do not study here, is when the lighting and camera coefficient are both varying, yet only lighting is calibrated. This is known as "semi-calibrated" photometric stereo [10].

pixel $\mathbf{p}$ conjugate to a 3D-point $\mathbf{x}$:

$$I_\star(\mathbf{p}) = \int_0^{+\infty} c_\star(\lambda)\,\rho(\mathbf{x},\lambda) \underbrace{\left[\int_{\mathcal{H}(\mathbf{x})} \max\{0, \mathbf{s}(\mathbf{x},\omega,\lambda)\cdot\mathbf{n}(\mathbf{x})\}\,\mathrm{d}\omega\right]}_{\boldsymbol{\sigma}(\lambda)\cdot\boldsymbol{\nu}(\mathbf{x})} \mathrm{d}\lambda. \quad (13)$$

This integral expression of the color level $I_\star$ can be simplified in two cases:

- When the surface is non-colored ($\rho(\mathbf{x},\lambda) = \rho(\mathbf{x})$), then (13) is rewritten:

$$I_\star(\mathbf{p}) = \rho(\mathbf{x})\,\boldsymbol{\sigma}_\star\cdot\boldsymbol{\nu}(\mathbf{x}), \quad (14)$$

where

$$\boldsymbol{\sigma}_\star = \int_0^{+\infty} c_\star(\lambda)\,\boldsymbol{\sigma}(\lambda)\,\mathrm{d}\lambda \quad (15)$$

is the colored lighting vector, relatively to the response of the camera in channel $\star$. In the multi-view case, (14) becomes the following extension of (12):

$$I^i_\star(\pi^i(\mathbf{x})) = \rho(\mathbf{x})\,\boldsymbol{\sigma}^i_\star\cdot\boldsymbol{\nu}(\mathbf{x}). \quad (16)$$

- When the lighting is non-colored ($\boldsymbol{\sigma}(\lambda) = \boldsymbol{\sigma}$), then (13) is rewritten:

$$I_\star(\mathbf{p}) = \rho_\star(\mathbf{x})\,\boldsymbol{\sigma}\cdot\boldsymbol{\nu}(\mathbf{x}), \quad (17)$$

where

$$\rho_\star(\mathbf{x}) = \int_0^{+\infty} c_\star(\lambda)\,\rho(\mathbf{x},\lambda)\,\mathrm{d}\lambda \quad (18)$$

is the colored reflectance, relatively to the response of the camera in channel $\star$. In the multi-view case, (17) becomes the following extension of (12):

$$I^i_\star(\pi^i(\mathbf{x})) = \rho_\star(\mathbf{x})\,\boldsymbol{\sigma}^i\cdot\boldsymbol{\nu}(\mathbf{x}). \quad (19)$$

However, it is not possible to find a tractable expression such as

$$I^i_\star(\pi^i(\mathbf{x})) = \rho_\star(\mathbf{x})\,\boldsymbol{\sigma}^i_\star\cdot\boldsymbol{\nu}(\mathbf{x}) \quad (20)$$

to extend model (12) to color, in the case where both the illumination and the reflectance are colored. Still, Model (20) is convenient to use in practice, because it makes the extension to RGB images of the proposed approach straightforward: it is indeed enough to apply the same framework independently in each color channel. Hence, we consider hereafter the graylevel case only *i.e.*, we consider the image formation model (12).

The question which arises now is how to estimate the reflectance ρ from a set of equations such as (12), when the geometry is known but the lighting and the camera coefficient are unknown.

3.3 Reflectance Estimation on the Surface

We place ourselves at the end of the multi-view 3D-reconstruction pipeline. Thus, the projections π^i are known (in practice, they are estimated using structure-from-motion techniques), as well as the geometry, represented by a set of n 3D-points $\mathbf{x}_j \in \mathbb{R}^3$, $j \in \{1,\dots,n\}$ and the corresponding normals $\mathbf{n}(\mathbf{x}_j)$ (obtained for instance using multi-view stereo techniques), from which the n geometric vectors $\boldsymbol{\nu}_j := \boldsymbol{\nu}(\mathbf{x}_j)$ are easily deduced according to (7).

The unknowns are then the n reflectance values $\rho_j := \rho(\mathbf{x}_j) \in \mathbb{R}$ and the m lighting vectors $\boldsymbol{\sigma}^i \in \mathbb{R}^9$, which are independent from the 3D-point number j due to the distant light assumption. One may naively think that their estimation can be carried out in a purely data-driven manner, using some fitting function $F : \mathbb{R} \to \mathbb{R}$:

$$\min_{\substack{\{\rho_j\in\mathbb{R}\}_j \\ \{\boldsymbol{\sigma}^i\in\mathbb{R}^9\}_i}} \sum_{i=1}^{m}\sum_{j=1}^{n} v^i_j\,F\left(\rho_j\,\boldsymbol{\sigma}^i\cdot\boldsymbol{\nu}_j - I^i_j\right), \quad (21)$$

where we denote $I^i_j = I^i(\pi^i(\mathbf{x}_j))$, and v is a visibility boolean such that $v^i_j = 1$ if the 3D-point $\mathbf{x}_j$ is visible in the i-th image, and $v^i_j = 0$ otherwise.

Let us consider, for the sake of pedagogy, the simplest case of least-squares fitting ($F(x) = x^2$) and perfect visibility ($v^i_j \equiv 1$). Then, Problem (21) is rewritten in matrix form:

$$\min_{\substack{\boldsymbol{\rho}\in\mathbb{R}^{n\times 1} \\ \mathbf{S}\in\mathbb{R}^{9\times m}}} \|\mathbf{N}\,(\boldsymbol{\rho}\otimes\mathbf{S}) - \mathbf{I}\|^2_F, \quad (22)$$

where vector $\boldsymbol{\rho} \in \mathbb{R}^{n\times 1}$ stores the reflectance values, matrix $\mathbf{S} \in \mathbb{R}^{9\times m}$ stores the lighting vectors, column-wise, matrix $\mathbf{I} \in \mathbb{R}^{n\times m}$ stores the graylevels, $\otimes$ is the Kronecker product, $\|\cdot\|_F$ is the Frobenius norm, and $\mathbf{N} \in \mathbb{R}^{n\times 9n}$ is a block-diagonal matrix where the j-th block, $j \in \{1,\dots,n\}$, is the row vector $\boldsymbol{\nu}_j^\top$.

Using the pseudo-inverse $\mathbf{N}^\dagger$ of $\mathbf{N}$, (22) is rewritten:

$$\min_{\substack{\boldsymbol{\rho}\in\mathbb{R}^{n\times 1} \\ \mathbf{S}\in\mathbb{R}^{9\times m}}} \left\|\boldsymbol{\rho}\otimes\mathbf{S} - \mathbf{N}^\dagger\,\mathbf{I}\right\|^2_F. \quad (23)$$

Problem (23) is a nearest Kronecker product problem, which can be solved by singular value decomposition (SVD) [15, Theorem 12.3.1].

However, this matrix factorization approach suffers from three shortcomings:

1) It is valid only if all 3D-points are visible under all the viewing angles, which is rather unrealistic. In practice, (22) should be replaced by

$$\min_{\substack{\boldsymbol{\rho}\in\mathbb{R}^{n\times 1} \\ \mathbf{S}\in\mathbb{R}^{9\times m}}} \|\mathbf{V}\circ[\mathbf{N}\,(\boldsymbol{\rho}\otimes\mathbf{S}) - \mathbf{I}]\|^2_F, \quad (24)$$

where $\mathbf{V} \in \mathbb{R}^{n\times m}$ is a visibility matrix containing the values v_j^i, and $\circ$ is the Hadamard product. This yields a Kronecker product problem with missing data, which is much more arduous to solve.

2) It is adapted only to least-squares estimation. Considering a more robust fitting function would prevent a direct SVD solution.
3) If lighting is not varying ($\boldsymbol{\sigma}^i = \boldsymbol{\sigma}, \forall i \in \{1,\dots,m\}$), then it can be verified that (22) is ill-posed. Among its many solutions, the following trivial one can be exhibited:

$$\mathbf{S}_{\text{trivial}} = \boldsymbol{\sigma}_{\text{diffuse}}\, \mathbf{1}_{1\times m}, \tag{25}$$

$$\boldsymbol{\rho}_{\text{trivial}} = [\rho_{\text{trivial},j}]_j\,, \tag{26}$$

with:

$$\boldsymbol{\sigma}_{\text{diffuse}} = [0,0,0,1,0,0,0,0,0]^\top \tag{27}$$

$$\rho_{\text{trivial},j} = E_i[I_j^i], \qquad \forall j \in \{1,\dots,n\}, \tag{28}$$

where E_i is the mean over the view indices i. This trivial solution means that the lighting is assumed to be completely diffuse[5], and that the reflectance is equal to the image graylevel, up to noise only. Obviously, this is not an acceptable interpretation. As discussed in the previous subsection, in real-world scenarios we will be very close to this degenerate case, hence additional regularization will have to be introduced, which makes things even harder.

Overall, the optimization problem which needs to be addressed is not as easy as (23). It is a non-quadratic regularized problem of the form:

$$\min_{\substack{\{\rho_j\in\mathbb{R}\}_j\\ \{\boldsymbol{\sigma}^i\in\mathbb{R}^9\}_i}} \sum_{i=1}^{m}\sum_{j=1}^{n} v_j^i\, F\big(\rho_j\, \boldsymbol{\sigma}^i\cdot\boldsymbol{\nu}_j - I_j^i\big) + \sum_{j=1}^{n}\sum_{k|\mathbf{x}_k\in\mathcal{V}(\mathbf{x}_j)} R(\rho_j,\rho_k), \tag{29}$$

where $\mathcal{V}(\mathbf{x}_j)$ is a set of neigbors of $\mathbf{x}_j$ on the surface $\mathcal{S}$, and the regularization function R needs to be chosen appropriately to ensure piecewise-smoothness.

However, the sampling of the points $\mathbf{x}_j$ on the surface $\mathcal{S}$ is usually non-uniform, because the shape of $\mathcal{S}$ is potentially complex. It may thus be difficult to design appropriate fidelity and regularization functions F and R, and to design an appropriate numerical solving. In addition, some thin brightness variations may be missed if the sampling is not dense enough. Overall, direct estimation of reflectance on the surface looks promising at first sight, but rather tricky in practice. Therefore, we leave this as an interesting future research direction and follow in this paper a simpler approach, which consists in estimating reflectance in the image domain.

[5] In the computer graphics community, this is referred to as "ambient lighting".

3.4 Reflectance Estimation in the Image Domain

Instead of trying to colorize the n 3D-points estimated by multi-view stereo *i.e.*, of parameterizing the reflectance over the surface $\mathcal{S}$, we can also formulate the reflectance estimation problem in the image (2D) domain.

Equation (12) is equivalently written, for all pixels $\mathbf{p} \in \Omega^i := \pi^i(\mathcal{S})$:

$$I^i(\mathbf{p}) = \rho^i(\mathbf{p})\, \boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i(\mathbf{p}), \tag{30}$$

where we denote $\rho^i(\mathbf{p}) := \rho({\pi^i}^{-1}(\mathbf{p}))$ and $\boldsymbol{\nu}^i(\mathbf{p}) := \boldsymbol{\nu}({\pi^i}^{-1}(\mathbf{p}))$. Instead of estimating one reflectance value per estimated 3D-points, the reflectance estimation problem is thus turned into the estimation of m "reflectance maps"

$$\rho^i : \Omega^i \subset \mathbb{R}^2 \to \mathbb{R}. \tag{31}$$

On the one hand, the 2D parameterization (31) does not explicitly enforce the consistency of the reflectance maps. This will have to be explicitly enforced later on. Besides, the surface will not be directly colorized, but the estimated reflectance maps could be back-projected and fused over the surface in a final step.

On the other hand, the question of occlusions (visibility) does not arise, and the domains Ω^i are subsets of a uniform square 2D-grid. Therefore, it will be much easier to design appropriate fidelity and regularization terms. Besides, there will be as many reflectance estimates as pixels in those sets: with modern HD cameras, this number is much larger than the number of 3D-points estimated by multi-view stereo. Estimation is thus much denser.

With such a parameterization choice, the regularized problem (29) will be turned into:

$$\begin{aligned}\min_{\substack{\{\rho^i:\,\Omega^i\to\mathbb{R}\}_i\\ \{\boldsymbol{\sigma}^i\in\mathbb{R}^9\}_i}} &\sum_{i=1}^{m}\sum_{\mathbf{p}\in\Omega^i} F\big(\rho^i(\mathbf{p})\, \boldsymbol{\sigma}^i\cdot\boldsymbol{\nu}^i(\mathbf{p}) - I^i(\mathbf{p})\big)\\ &+\sum_{i=1}^{m}\sum_{\mathbf{p}\in\Omega^i}\sum_{\mathbf{q}\in\mathcal{V}^i(\mathbf{p})} R(\rho^i(\mathbf{p}),\rho^i(\mathbf{q}))\\ &\text{s.t. } C(\{\rho^i\}_i) = 0,\end{aligned} \tag{32}$$

with C some function to ensure multi-view consistency, and where $\mathcal{V}^i(\mathbf{p})$ is the set of neighbors of pixel $\mathbf{p}$ which lie in Ω^i. Note that, since Ω^i is a subset of a square, regular 2D-grid, this neighborhood is much easier to handle than that appearing in (29).

In the next section, we discuss appropriate choices for F, R and C in (32), by resorting to a Bayesian rationale.

4 A Bayesian-to-variational Framework for Multi-view Reflectance Estimation

Let us now introduce a Bayesian-to-variational framework for estimating reflectance and lighting from multi-view images.

4.1 Bayesian Inference

Our problem consists in estimating the m reflectance maps $\rho^i : \Omega^i \to \mathbb{R}$ and the m lighting vectors $\boldsymbol{\sigma}^i \in \mathbb{R}^9$, given the m images $I^i : \Omega^i \to \mathbb{R}$, $i \in \{1, \dots, m\}$. As we already stated, a maximum likelihood approach is hopeless, because a trivial solution arises. We rather resort to Bayesian inference, estimating $(\{\rho^i\}_i, \{\boldsymbol{\sigma}^i\}_i)$ as the maximum a posteriori (MAP) of the distribution

$$\mathcal{P}(\{\rho^i\}_i, \{\boldsymbol{\sigma}^i\}_i | \{I^i\}_i) = \frac{\mathcal{P}(\{I^i\}_i | \{\rho^i\}_i, \{\boldsymbol{\sigma}^i\}_i)\, \mathcal{P}(\{\rho^i\}_i, \{\boldsymbol{\sigma}^i\}_i)}{\mathcal{P}(\{I^i\}_i)}, \tag{33}$$

where the denominator is the evidence, which can be discarded since it depends neither on the reflectance nor on the lighting, and the factors in the numerator are the likelihood and the prior, respectively.

Likelihood. The image formation model (30) is never strictly satisfied in practice, due to noise, cast-shadows and possibly slightly specular surfaces. We assume that such deviations from the model can be represented as independent (with respect to pixels and views) Laplace laws with zero mean and scale parameter α:

$$\begin{aligned}\mathcal{P}(\{I^i\}_i | \{\rho^i\}_i, \{\boldsymbol{\sigma}^i\}_i) &= \prod_{i=1}^{m} \left(\frac{1}{2\alpha}\right)^{|\Omega^i|} \exp\left\{-\frac{1}{\alpha} \left\| \rho^i\, \boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i - I^i \right\|_{i,1}\right\} \\ &= \left(\frac{1}{2\alpha}\right)^{\sum_{i=1}^m |\Omega^i|} \exp\left\{-\frac{1}{\alpha} \sum_{i=1}^{m} \left\| \rho^i\, \boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i - I^i \right\|_{i,1}\right\}\end{aligned} \tag{34}$$

where $\|\cdot\|_{i,p}$, $p \geq 0$, is the ℓ^p norm over Ω^i and $|\Omega^i|$ is the cardinal of Ω^i.

Prior. Since the reflectance maps and the lighting vectors are independent, the prior can be factorized to $\mathcal{P}(\{\rho^i\}_i, \{\boldsymbol{\sigma}^i\}_i) = \mathcal{P}(\{\rho^i\}_i)\mathcal{P}(\{\boldsymbol{\sigma}^i\}_i)$. Since lighting vectors are independent, the prior distribution of the lighting vectors factorizes to $\mathcal{P}(\{\boldsymbol{\sigma}^i\}_i) = \prod_{i=1}^m \mathcal{P}(\boldsymbol{\sigma}^i)$. As each lighting vector is unconstrained, we can consider the same uniform distribution *i.e.*, $\mathcal{P}(\boldsymbol{\sigma}^i) = \tau$, independently from the view index i. This distribution being independent from the unknowns, we can discard the lighting prior from the inference process. Regarding the reflectance maps, we follow the Retinex theory and consider each of them as piecewise-constant. The natural prior for each map is thus the Potts model:

$$\mathcal{P}(\rho^i) = K^i \exp\left\{-\frac{1}{\beta^i} \left\| \nabla \rho^i \right\|_{i,0}\right\} \tag{35}$$

where $\nabla \rho^i(\mathbf{p}) = \left[\partial_x \rho^i(\mathbf{p}), \partial_y \rho^i(\mathbf{p})\right]^\top$ represents the gradient of ρ^i at pixel $\mathbf{p}$ (approximated, in practice, using first-order forward stencils with a Neumann boundary condition), and with K^i a normalization coefficient and β^i a scale parameter. Note that we use the abusive ℓ^0 norm notation $\|\nabla \rho^i\|_{i,0}$ to denote:

$$\left\| \nabla \rho^i \right\|_{i,0} = \sum_{\mathbf{p} \in \Omega^i} \sum_{\mathbf{q} \in \mathcal{V}^i(\mathbf{p})} f\left(\rho^i(\mathbf{p}) - \rho^i(\mathbf{q})\right) \tag{36}$$

with $f(x) = 1$ if $x \neq 0$, and $f(x) = 0$ otherwise.

The m reflectance maps are obviously not independent: the reflectance characterizes the surface and it is thus independent from the view. It follows that the parameters (K^i, β^i) are the same for each Potts model (35), and that the reflectance prior $\mathcal{P}(\{\rho^i\}_i)$ can be taken as the product of m independent distributions with the same parameters (K, β):

$$\mathcal{P}(\{\rho^i\}_i) = K^m \exp\left\{-\frac{1}{\beta} \sum_{i=1}^{m} \left\| \nabla \rho^i \right\|_{i,0}\right\} \tag{37}$$

but only if the coupling between the reflectance maps is enforced by the following linear constraint:

$$C^{i,j}(\rho^i - \rho^j) = 0, \ \forall (i,j) \in \{1, \dots, m\}^2, \tag{38}$$

where $C^{i,j}$ is a $\Omega^i \times \Omega^j \to \{0, 1\}$ "correspondence function", which is easily created from the (known) projection functions $\{\pi^i\}_i$ and the geometry, and which is defined as follows:

$$C^{i,j}(\mathbf{p}^i, \mathbf{p}^j) = \begin{cases} 1 & \text{if pixels } \mathbf{p}^i \text{ and } \mathbf{p}^j \text{ correspond} \\ & \quad \text{to the same surface point,} \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

Since maximizing the MAP probability (33) is equivalent to minimizing its negative logarithm, we eventually obtain the following constrained variational problem, which explicits the functions F, R and C in (32):

$$\begin{aligned}&\min_{\substack{\{\rho^i : \Omega^i \to \mathbb{R}\}_i \\ \{\boldsymbol{\sigma}^i \in \mathbb{R}^9\}_i}} \sum_{i=1}^{m} \left\| \rho^i\, \boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i - I^i \right\|_{i,1} + \lambda \sum_{i=1}^{m} \left\| \nabla \rho^i \right\|_{i,0} \\ &\text{s.t.} \quad C^{i,j}(\rho^i - \rho^j) = 0, \ \forall (i,j) \in \{1, \dots, m\}^2,\end{aligned} \tag{40}$$

where $\lambda = \alpha/\beta$ and where we neglect all the normalization coefficients.

4.2 Relationship with Cartoon + Texture Decomposition

Applying a logarithm transformation to both sides of (30), we obtain:

$$\tilde{I}^i(\mathbf{p}) = \tilde{\rho}^i(\mathbf{p}) + \log\left(\boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i(\mathbf{p})\right), \tag{41}$$

where the tilde notation is used as a shortcut for the logarithm.

By applying the exact same Bayesian-to-variational rationale, we would end up with the following variational problem:

$$\min_{\substack{\{\tilde{\rho}^i:\,\Omega^i\to\mathbb{R}\}_i \\ \{\boldsymbol{\sigma}^i\in\mathbb{R}^9\}_i}} \sum_{i=1}^m \left\| \tilde{\rho}^i + \log\left(\boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i\right) - \tilde{I}^i \right\|_{i,1} + \lambda \sum_{i=1}^m \left\|\nabla \tilde{\rho}^i\right\|_{i,0}$$
$$\text{s.t.} \quad C^{i,j}(\tilde{\rho}^i - \tilde{\rho}^j) = 0, \ \forall (i,j) \in \{1,\dots,m\}^2, \tag{42}$$

Variational problem (42) can be interpreted as a multi-view cartoon + texture decomposition problem, where each image $\tilde{I}$ is decomposed into a component $C^i := \tilde{\rho}^i$ which is piecewise-smooth ("cartoon", here the log-reflectance), and a component $T^i := \log\left(\boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i\right)$ which contains higher-frequency details ("texture", here the log-shading). In contrast with conventional methods for such a task, the present one uses an explicit shading model for the texture term, here the spherical harmonics model.

Note however that such a decomposition is justified only if the log-images $\tilde{I}^i$ are considered. If using the original images I^i, our framework should rather be considered as a multi-view cartoon "×" texture decomposition framework.

4.3 Simplification of the Variational Model (40)

Problem (40) is a non-convex, non-smooth variational problem, due to the ℓ^0 regularizers. Although some efforts have recently been devoted to the resolution of such uneasy optimization problems [36], we prefer to keep the optimization simple, and approximate these regularizers by (convex, but non-smooth) anisotropic total variation terms:

$$\sum_{i=1}^m \left\|\nabla \rho^i\right\|_{i,0} \approx \sum_{i=1}^m \left\|\nabla \rho^i\right\|_{i,1}. \tag{43}$$

Besides, the correspondence function may be slightly inaccurate in practice, due to errors in the prior geometry estimation obtained via multi-view stereo. Therefore, we turn the linear constraint in (40) into an additional term. Eventually, we replace the non-differentiable absolute values arising from the ℓ^1-norms by the (differentiable) Moreau envelope *i.e.*, the Huber loss[6]:

$$|x| \approx \phi_\delta(x) := \begin{cases} \dfrac{x^2}{2\,\delta}, & |x| \le \delta \\ |x| - \dfrac{\delta}{2}, & |x| > \delta \end{cases} \tag{44}$$

Altogether, this yields the following variational problem:

$$\min_{\substack{\rho:=\{\rho^i:\,\Omega^i\to\mathbb{R}\}_i \\ \boldsymbol{\sigma}:=\{\boldsymbol{\sigma}^i\in\mathbb{R}^9\}_i}} \varepsilon(\rho,\boldsymbol{\sigma}) := \sum_{i=1}^m \varepsilon_{\text{Photo}}(\rho^i,\boldsymbol{\sigma}^i)$$
$$+ \lambda \sum_{i=1}^m \varepsilon_{\text{Smooth}}(\rho^i) + \mu \sum_{1\le i<j\le m} \varepsilon_{\text{MV}}(\rho^i,\rho^j). \tag{45}$$

In Equation (45), the first term ensures photometric consistency (in the sense of the Huber loss function):

$$\varepsilon_{\text{Photo}}(\rho^i,\boldsymbol{\sigma}^i) = \sum_{\mathbf{p}\in\Omega^i} \phi_\delta\left(\rho^i(\mathbf{p})\,\boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i(\mathbf{p}) - I^i(\mathbf{p})\right), \tag{46}$$

the second one ensures reflectance smoothness (smoothed anisotropic total variation):

$$\varepsilon_{\text{Smooth}}(\rho^i) = \sum_{\mathbf{p}\in\Omega^i} \left[\phi_\delta\left(\partial_x \rho^i(\mathbf{p})\right) + \phi_\delta\left(\partial_y \rho^i(\mathbf{p})\right)\right], \tag{47}$$

and the third term ensures multi-view consistency of the reflectance estimates (again, in the sense of the Huber loss function):

$$\varepsilon_{\text{MV}}(\rho^i,\rho^j) = \sum_{\mathbf{p}^i\in\Omega^i} \sum_{\mathbf{p}^j\in\Omega^j} C_{i,j}(\mathbf{p}^i,\mathbf{p}^j)\,\phi_\delta\left(\rho^i(\mathbf{p}^i) - \rho^j(\mathbf{p}^j)\right). \tag{48}$$

At last, λ and μ are tunable hyper-parameters controlling the reflectance smoothness and the multi-view consistency, respectively.

5 Alternating Majorization-Minimization for Solving (45)

To solve (45), we propose an alternating majorization-minimization method, which combines alternating and majorization-minimization optimization techniques. As sketched in Figure 3, this algorithm works as follows. Given an estimate $(\rho^{(k)},\boldsymbol{\sigma}^{(k)})$ of the solution at iteration (k), the lighting vectors and the reflectance maps are successively updated according to:

$$\rho^{(k+1)} = \operatorname*{argmin}_{\rho}\ \varepsilon_\rho^{(k)}(\rho), \tag{49}$$

$$\boldsymbol{\sigma}^{(k+1)} = \operatorname*{argmin}_{\boldsymbol{\sigma}}\ \varepsilon_{\boldsymbol{\sigma}}^{(k)}(\boldsymbol{\sigma}), \tag{50}$$

where $\varepsilon_\rho^{(k)}$ and $\varepsilon_{\boldsymbol{\sigma}}^{(k)}$ are local quadratic majorants of $\varepsilon(\cdot,\boldsymbol{\sigma}^{(k)})$ and $\varepsilon(\rho^{(k+1)},\cdot)$ around, respectively, $\rho^{(k)}$ and $\boldsymbol{\sigma}^{(k)}$. Then, the process is repeated until convergence.

[6] We use $\delta = 10^{-4}$, in the experiments.

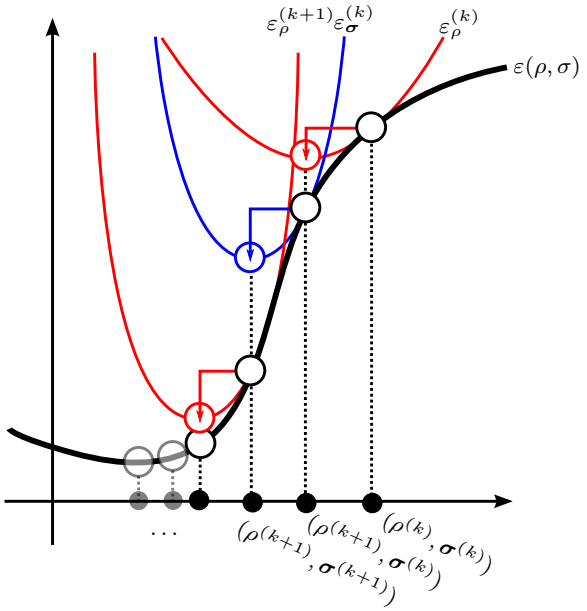


Fig. 3 Sketch of the proposed alternating majorization-minimization solution. The partially freezed energies $\varepsilon(\cdot,\boldsymbol{\sigma})$ and $\varepsilon(\rho,\cdot)$ are locally majorized by the quadratic functions ε_ρ (in red) and $\varepsilon_{\boldsymbol{\sigma}}$ (in blue). Then, these quadratic majorants are (globally) minimized and the process is repeated until convergence is reached.

5.1 Majorants

Let us first remark that the function

$$\psi_\delta(x;\,x_0)=\begin{cases}\dfrac{x^2}{2\,\delta}, & |x_0|\le\delta,\\[2mm] \dfrac{x^2}{2\,|x_0|}+\dfrac{|x_0|}{2}-\dfrac{\delta}{2}, & |x_0|>\delta,\end{cases}\tag{51}$$

is such that $\psi_\delta(x_0;\,x_0)=\phi_\delta(x_0)$, and is a proper local quadratic majorant of ϕ_δ around x_0, $\forall x_0\in\mathbb{R}$. This is easily verified if $|x_0|\le\delta$, from the definition (44) of ϕ_δ. If $|x_0|>\delta$, the difference $\psi_\delta(x;\,x_0)-\phi_\delta(x)$ writes:

$$\begin{cases}\dfrac{(|x_0|-\delta)\,(|x_0|\,\delta-x^2)}{2\,|x_0|\,\delta}, & |x|\le\delta,\\[2mm] \dfrac{(|x|-|x_0|)^2}{2\,|x_0|}, & |x|>\delta,\end{cases}\tag{52}$$

which is positive in any case. Therefore, the function

$$\begin{aligned}\varepsilon_\rho^{(k)}(\rho):=&\sum_{i=1}^m\sum_{\mathbf{p}\in\Omega^i}\psi_\delta\Big(\rho^i(\mathbf{p})\,\boldsymbol{\sigma}^{i,(k)}\cdot\boldsymbol{\nu}^i(\mathbf{p})-I^i(\mathbf{p});\;r^{i,(k),(k)}\Big)\\&+\lambda\sum_{i=1}^m\sum_{\mathbf{p}\in\Omega^i}\Big[\psi_\delta\Big(\partial_x\rho^i(\mathbf{p});\;\partial_x\rho^{i,(k)}(\mathbf{p})\Big)\\&\qquad+\psi_\delta\Big(\partial_y\rho^i(\mathbf{p});\;\partial_y\rho^{i,(k)}(\mathbf{p})\Big)\Big]\\&+\mu\sum_{1\le i<j\le m}\sum_{\mathbf{p}^i\in\Omega^i}\sum_{\mathbf{p}^j\in\Omega^j}C_{i,j}(\mathbf{p}^i,\mathbf{p}^j)\\&\psi_\delta\Big(\rho^i(\mathbf{p}^i)-\rho^j(\mathbf{p}^j);\;\rho^{i,(k)}(\mathbf{p}^i)-\rho^{j,(k)}(\mathbf{p}^j)\Big),\end{aligned}\tag{53}$$

with

$$r^{i,(k_1),(k_2)}=\rho^{i,(k_1)}(\mathbf{p})\,\boldsymbol{\sigma}^{i,(k_2)}\cdot\boldsymbol{\nu}^i(\mathbf{p})-I^i(\mathbf{p}),\tag{54}$$

is a local quadratic majorant of $\varepsilon(\cdot,\boldsymbol{\sigma}^{(k)})$ around $\rho^{(k)}$ which is suitable for the update (49).

Similarly, the function

$$\begin{aligned}\varepsilon_\sigma^{(k)}(\boldsymbol{\sigma}):=&\sum_{i=1}^m\sum_{\mathbf{p}\in\Omega^i}\psi_\delta\Big(\rho^{i,(k+1)}(\mathbf{p})\boldsymbol{\sigma}^i\cdot\boldsymbol{\nu}^i(\mathbf{p})-I^i(\mathbf{p});\;r^{i,(k+1),(k)}\Big)\\&+\lambda\sum_{i=1}^m\varepsilon_{\text{Smooth}}(\rho^{i,(k+1)})+\mu\sum_{1\le i<j\le m}\varepsilon_{\text{MV}}(\rho^{i,(k+1)},\rho^{j,(k+1)})\end{aligned}\tag{55}$$

is a local quadratic majorant of $\varepsilon(\rho^{(k+1)},\cdot)$ around $\boldsymbol{\sigma}^{(k)}$ which is suitable for the update (50).

5.2 Numerical Solutions of (49) and (50)

Let us first consider the update (50), which is easier. Using (51) and (55), and neglecting the constants which play no role in the optimization, (50) is rewritten as the following m independent weighted least-squares problems, $i\in\{1,\ldots,m\}$:

$$\begin{aligned}\boldsymbol{\sigma}^{i,(k+1)}=\operatorname*{argmin}_{\boldsymbol{\sigma}^i\in\mathbb{R}^9}&\sum_{\mathbf{p}\in\Omega^i}w_i^{(k+1),(k)}(\mathbf{p})\\&\Big(\rho^{i,(k+1)}(\mathbf{p})\,\boldsymbol{\sigma}^i\cdot\boldsymbol{\nu}^i(\mathbf{p})-I^i(\mathbf{p})\Big)^2,\end{aligned}\tag{56}$$

with

$$w_i^{(k_1),(k_2)}(\mathbf{p})=\frac{1}{|I^i(\mathbf{p})-\rho^{i,(k_1)}(\mathbf{p})\,\boldsymbol{\sigma}^{i,(k_2)}\cdot\boldsymbol{\nu}^i(\mathbf{p})|_\delta}.\tag{57}$$

and

$$|x|_\delta=\max\{|x|,\delta\}.\tag{58}$$

Each problem (56) is a small-scale linear least-squares problem which can be solved, for instance, by resorting to the pseudo-inverse.

Similarly, using (51) and (53), and neglecting the constants, the update (49) is rewritten as the following weighted least-squares problem:

$$\begin{aligned}\rho^{(k+1)}=\operatorname*{argmin}_{\{\rho^i:\,\Omega^i\to\mathbb{R}\}_i}&\sum_{i=1}^m\sum_{\mathbf{p}\in\Omega^i}\varepsilon^{\text{W}}_{\text{Photo}}(\rho^i,\boldsymbol{\sigma}^{i,(k)},\mathbf{p})\\&+\lambda\sum_{i=1}^m\sum_{\mathbf{p}\in\Omega^i}\varepsilon^{\text{W}}_{\text{Smooth}}(\rho^i,\mathbf{p})\\&+\mu\sum_{1\le i<j\le m}\sum_{\mathbf{p}^i\in\Omega^i}\sum_{\mathbf{p}^j\in\Omega^j}\varepsilon^{\text{W}}_{\text{MV}}(\rho^i,\rho^j,\mathbf{p}^i,\mathbf{p}^j),\end{aligned}\tag{59}$$

where we introduce the following weighted local energies:

$$\varepsilon^{\mathrm{W}}_{\mathrm{Photo}}(\rho^i, \boldsymbol{\sigma}^{i,(k)}, \mathbf{p}) := w_i^{(k),(k)}(\mathbf{p}) \left(\rho^i(\mathbf{p})\, \boldsymbol{\sigma}^{i,(k)} \cdot \boldsymbol{\nu}^i(\mathbf{p}) - I^i(\mathbf{p})\right)^2, \quad (60)$$

$$\varepsilon^{\mathrm{W}}_{\mathrm{Smooth}}(\rho^i, \mathbf{p}) := \left\| \begin{bmatrix} \sqrt{w^{(k)}_{\partial_x \rho^i}(\mathbf{p})} & \\ & \sqrt{w^{(k)}_{\partial_y \rho^i}(\mathbf{p})} \end{bmatrix} \nabla \rho^i(\mathbf{p}) \right\|^2, \quad (61)$$

$$\varepsilon^{\mathrm{W}}_{\mathrm{MV}}(\rho^i, \rho^j, \mathbf{p}^i, \mathbf{p}^j) := C_{i,j}(\mathbf{p}^i, \mathbf{p}^j) w^{(k)}_{i,j}(\mathbf{p}^i, \mathbf{p}^j)\, (\rho^i(\mathbf{p}^i) - \rho^j(\mathbf{p}^j))^2, \quad (62)$$

which involve the weights $w_i^{(k),(k)}$ defined in (57), as well as the following new ones:

$$w^{(k)}_{\partial_x \rho^i}(\mathbf{p}) = \frac{1}{|\partial_x \rho^{i,(k)}(\mathbf{p})|_\delta}, \quad (63)$$

$$w^{(k)}_{\partial_y \rho^i}(\mathbf{p}) = \frac{1}{|\partial_y \rho^{i,(k)}(\mathbf{p})|_\delta}, \quad (64)$$

$$w^{(k)}_{i,j}(\mathbf{p}^i, \mathbf{p}^j) = \frac{1}{|\rho^{i,(k)}(\mathbf{p}^i) - \rho^{j,(k)}(\mathbf{p}^j)|_\delta}. \quad (65)$$

Contrarily to the lighting vectors updates, the reflectance maps updates are not independent, because of the multi-view consistency prior. All estimations must thus be carried out simultaneously. Stacking all the reflectance values in a large vector $\boldsymbol{\rho} \in \mathbb{R}^N$, with $N = \sum_i |\Omega^i|$, the optimisation problem (59) can be turned into a linear least-squares problem having the following form:

$$\boldsymbol{\rho}^{(k+1)} = \underset{\boldsymbol{\rho} \in \mathbb{R}^N}{\operatorname{argmin}} \left\| \mathbf{A}^{(k)} \boldsymbol{\rho} - \mathbf{b}^{(k)} \right\|^2, \quad (66)$$

with $\mathbf{A}^{(k)}$ a large, sparse matrix. Problem (66) can be solved by applying conjugate gradient iterations to the associated normal equations

$$\mathbf{A}^{(k)\top} \mathbf{A}^{(k)} \boldsymbol{\rho} = \mathbf{A}^{(k)\top} \mathbf{b}^{(k)}. \quad (67)$$

5.3 Implementation Details

We iterate optimisation steps (49) and (50) until convergence or a maximum iteration number is reached, starting from the trivial solution of the non-regularized ($\lambda = \mu = 0$) problem. This non-regularized solution is attained by considering diffuse lighting (see (27)) and using the input images as reflectance maps. In our experiments, we found 50 iterations were always sufficient to reach a stable solution (10^{-3} relative residual between two consecutive energy values $\varepsilon(\rho^{(k)}, \boldsymbol{\sigma}^{(k)})$ and $\varepsilon(\rho^{(k+1)}, \boldsymbol{\sigma}^{(k+1)})$.

Proving convergence of our scheme is beyond the scope of this paper, but the proof could certainly be derived from that in [30], where a similar alternating majorization-minimization called "alternating reweighted least-squares" is used. Note, however, that the convergence rate seems to be sublinear (see Figure 4), hence possibly faster numerical strategies could be explored in the future.

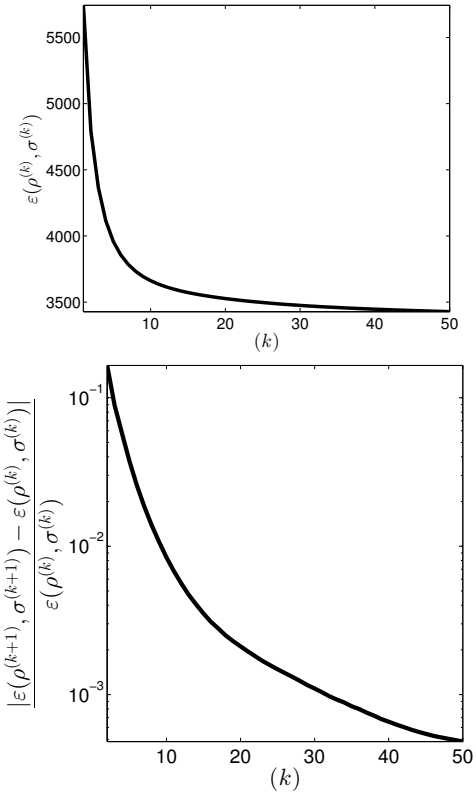


Fig. 4 Top: evolution of the energy $\varepsilon(\rho^{(k)}, \boldsymbol{\sigma}^{(k)})$ defined in (45), in function of iterations (k), concerning the test presented in Figure 8. Bottom: absolute value of the relative variation between two successive energy values. Our algorithm stops when this value is less than 10^{-3}, which happens in less than 50 iterations and takes around 3 minutes on a recent i7 processor, with non-optimized Matlab codes for $m = 13$ images of size 540×960.

6 Results

In this section, we evaluate the proposed variational method for multi-view reflectance estimation, on a variety of synthetic and real-world datasets. We start by a quantitative comparison of our results with two single-view methods, namely, the cartoon + texture decomposition method from [23] and the intrinsic image decomposition method from [14].

6.1 Quantitative Evaluation on a Synthetic Dataset

We first test our reflectance estimation method using $m = 13$ images, of size 540×960, of a synthetic object whose geometry is perfectly known (see Figure 5-a). Two scenarios are considered:

- In Figure 6, a purely-Lambertian, piecewise-constant reflectance is mapped onto the surface of the object, which is then lit under a "sky-dome" *i.e.*, almost diffuse, lighting. Shading effects are thus rather limited, hence applying to each image an estimation method which does not use an explicit reflectance model *e.g.*, the cartoon + texture decomposition method from [23], should already provide satisfactory results. The reflectance being perfectly piecewise constant, applying sparsity-based intrinsic image decomposition methods such as [14] to each image should also work well.
- In Figure 7, a more complicated (non-uniform) reflectance is mapped onto the shirt, the hair is made partly specular, and the diffuse lighting is replaced by a single extended light source, which induces much stronger shading effects. It will thus be much harder to remove shading without an explicit reflectance model (cartoon + texture approach), while the single-view image decomposition approach should be non-robust to specularities.

In both cases, the competing methods [23] and [14] are applied independently to each of the $m = 13$ images. The estimates are thus not expected to be consistent, which may be problematic if the reflectance maps should be further mapped onto the surface for, *e.g.*, relighting applications. On the contrary, our approach simultaneously, and consistently, estimates the m reflectance maps.

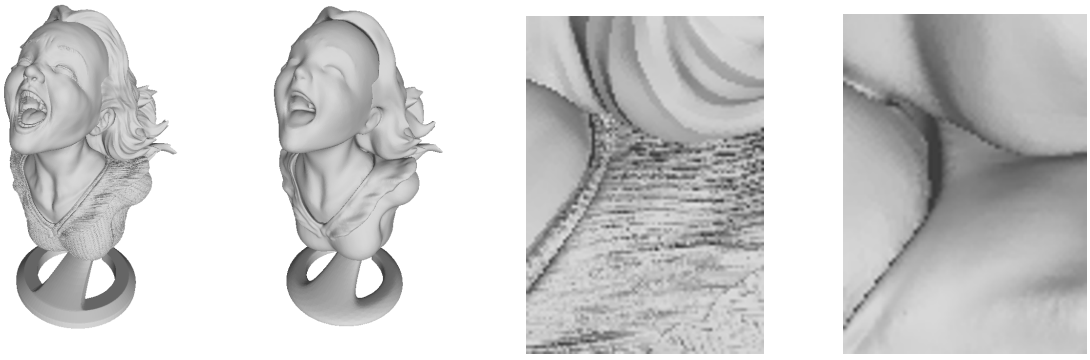
As we dispose of the reflectance ground truth, we can numerically evaluate these results by estimating the root mean square error (RMSE) for each method, over the whole set of $m = 13$ images. The values are presented in Table 1. In order to compare comparable things, the reflectance estimated by each method is scaled, in each channel, by a factor common to the $m = 13$ reflectance maps, so as to minimize the RMSE. This should thus highlight inconsistencies between the reflectance maps.

Based on the qualitative results from Figures 6 and 7, and the quantitative ones shown in Table 1, we can make the following three observations:

1) Considering an explicit image formation model improves cartoon + texture decomposition. Indeed, the "cartoon" part from the cartoon + texture decomposition is far less uniform than the reflectance estimated using both other methods. Shading is only blurred, and not really removed. This could be improved by augmenting the regularization weight, but the price to pay would be a loss of detail in the parts containing thinner details (as the shirt, in the example of Figure 7).

2) Simultaneously estimating the multi-view reflectance maps makes them consistent and improves robustness to specularities. When estimating each reflectance map individually, inconsistencies arise, which is obvious for the hair in the third line of Figure 6, and explains the RMSE values in Table 1. In contrast, our results confirm our basic idea *i.e.*, that reflectance estimation benefits in two ways from the multi-view framework: this allows us not only to estimate the 3D-shape, but also to constrain the reflectance of each surface point to be the same in all the pictures where it is visible. In addition, since the location of bright spots due to specularity depends on the viewing angle, they usually occur in some places on the surface only under certain viewing angles. Considering multi-view data should thus improve robustness to specularities. This is confirmed in Figure 7 by the reflectance estimates in the hair, where the specularities are slightly better removed than with single-view methods.

3) A sparsity-based prior for the reflectance should be preferred over total variation. As we use a TV-smoothing term, which favors piecewise-smooth reflectance, the satisfactory results of Figure 6 were predictable. However, some penumbra remains visible around the neck. Since we also know the object geometry, it seems that we could compensate for penumbra. However, this would require that the lighting is known as well, which is not the case in the framework of the targeted usecase, since an outdoors lighting is uncontrolled. Moreover, we would have to consider not only the primary lighting, but also the successive bounces of light on the different parts of the scene (these were taken into account by the ray-tracing algorithm). In contrast, the sparsity-based approach [14] is able to eliminate penumbra rather well, without modeling secundary reflections. It is also able to more appropriately remove shading on the face in the example of Figure 7, while not degrading as much as total variation the thin structures of the shirt. Hence, the relative simplicity of the numerical solution, which is a consequence of the choice of replacing the Potts prior by a total variation one (see Section 4.3), comes with a price. In future works, it may be important to design a numerical strategy handling the original non-smooth, non-convex problem (40).



(a) (b) (c) (d)

Fig. 5 (a) 3D-shape used in the tests (the well-known "Joyful Yell" 3D-model), which will be imaged under two scenarios (see Figures 6 and 7). (b) Same, after smoothing, thus less accurate. (c)-(d) Zooms of (a) and (b), respectively, near the neck.

Table 1 RMSE on the reflectance estimates (the estimated and ground truth reflectance maps are scaled to $[0, 1]$), with respect to each channel and to the whole set of images, for our method and two single-view approaches. Our method overcomes the latter on the two considered datasets. See text for details.

Test	Channel	Cartoon + texture [23]	Intrinsic decomposition [14]	Ours
Purely-Lambertian surface + Piecewise-constant reflectance + Skydome lighting (see Figure 6)	R	0.62	0.26	**0.07**
	G	0.23	0.14	**0.04**
	B	0.38	0.24	**0.07**
Non-uniform shirt reflectance + Partly specular hair reflectance + Single extended light source (see Figure 7)	R	0.60	0.29	**0.22**
	G	0.32	0.22	**0.13**
	B	0.24	0.21	**0.12**

6.2 Handling Inaccurate Geometry

In the previous experiments, the geometry was perfectly known. In real-world scenarios, errors in the 3D-shape estimation using SfM and MVS are unavoidable. Therefore, it is necessary to evaluate the ability of our method to handle inaccurate geometry.

Thus, we use for the next experiment the surface shown in Figure 5-b (zoomed in Figure 5-d), which is obtained by smoothing the original 3D-shape of Figure 5-a (zoomed in Figure 5-c), using a tool from the `meshlab` software. The results provided in Figure 8 show that our method seems robust to such small inaccuracies in the object geometry, and is thus relevant for the intended application.

In Figure 9, we qualitatively evaluate our method on the outputs of an SfM/MVS pipeline, which provides a rough geometry and camera parameters estimates. These experiments confirm that small inaccuracies in the geometry input can be handled. The specularities are also appropriately removed, and the reflectance maps present the expected "cartoon"-like aspect. However, the reflectance is under-estimated in the sides of the nose and around the chin. Indeed, since lighting is fixed, these areas are self-shadowed in all the images. Two workarounds could be used: forcing the regularization term (and, possibly, losing fine-scale details), or actively controlling the lighting in order to be sure that no point on the surface is shadowed in all the views. This is further discussed in the next subsection.

6.3 Tuning the Hyper-parameters λ and μ

In the previous experiments, we arbitrarily chose the values of parameters λ and μ which provided the "best" results. Of course, such a tuning may be tedious and must be discussed.

In order to highlight the influence of these parameters, let us first question what would happen without neither regularization nor multi-view consistency *i.e.*, when $\lambda = \mu = 0$. In that case, only the photometric term (46) would be optimised, which corresponds to the maximum likelihood case. If lighting is not varying, then we are in a degenerate case which may result in estimating diffuse lighting (see Equation (27)) and replacing the reflectance maps by the images. Lighting will thus be "baked in" the reflectance maps, which is precisely what we pretend not to do.

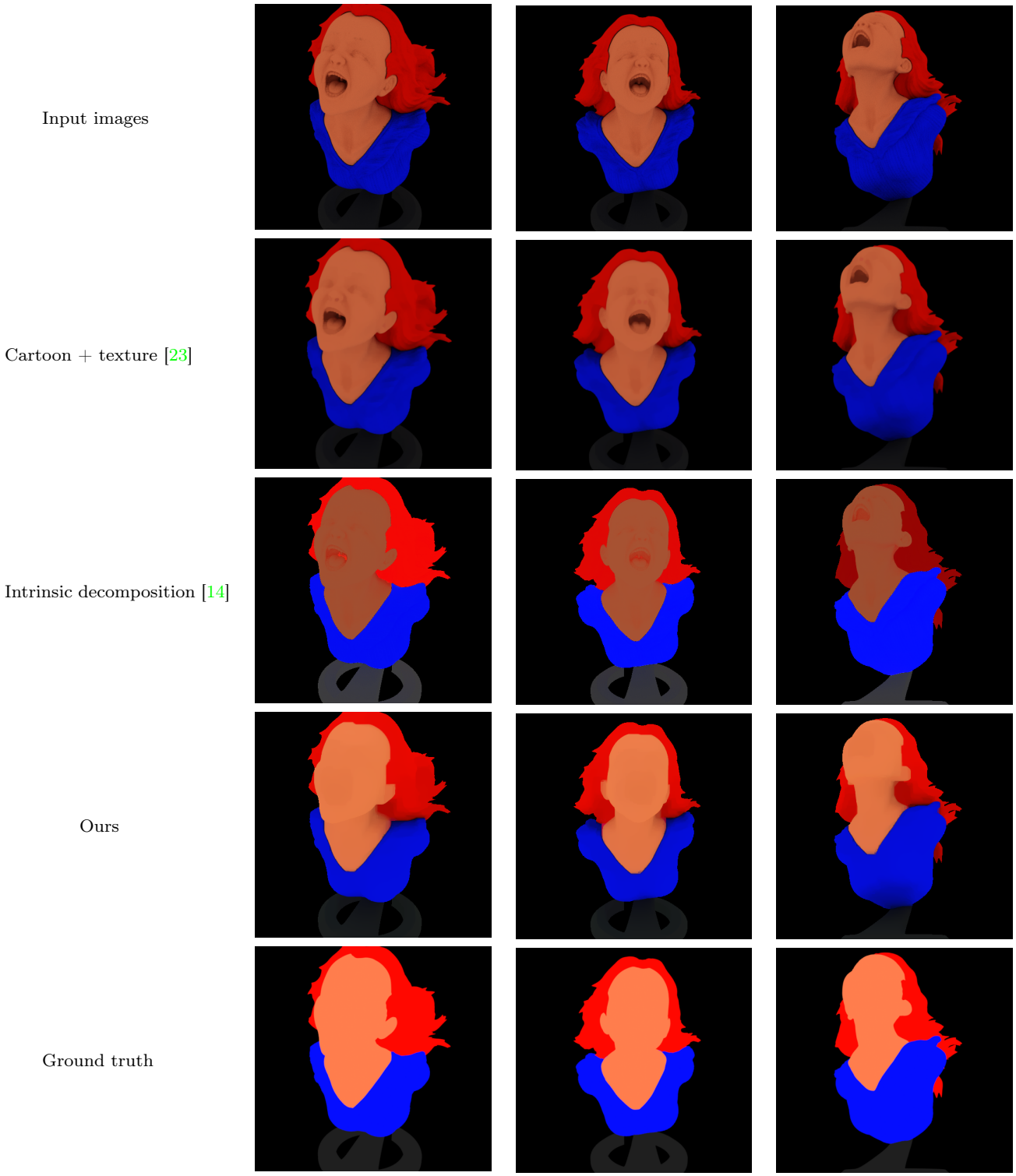


Fig. 6 First row: three (out of $m = 13$) synthetic views of the object of Figure 5-a, computed with a purely-Lambertian reflectance taking only four different values (hair, face, shirt and plinth), under "sky-dome" lighting. Second row: estimation of the reflectance using the cartoon + texture decomposition described in [23] (with its parameter fixed to 0.4). Third row: estimation of the reflectance using the method proposed in [14] (with 4 clusters). Forth row: estimation of the reflectance using the proposed approach (with $\lambda = 8$ and $\mu = 1000$). Fifth row: ground truth.

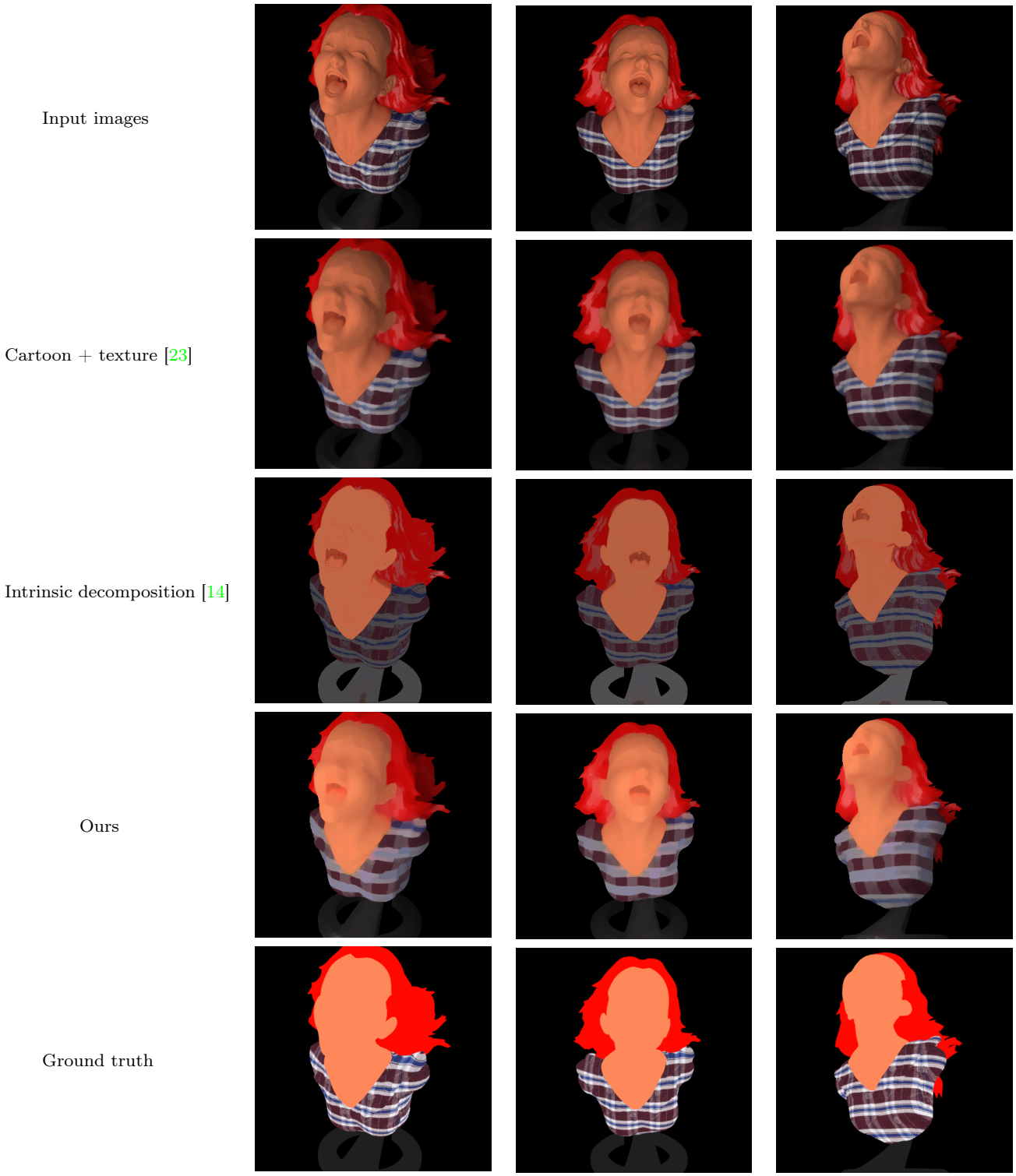


Fig. 7 First row: three (out of $m = 13$) synthetic views of the object of Figure 5-a, computed with a non-uniform shirt reflectance, a uniform, but partly specular hair reflectance, illuminated by a single extended light source. Second row: estimation of the reflectance using the cartoon + texture decomposition described in [23] (with its parameter fixed to 0.4). Third row: estimation of the reflectance using the method proposed in [14] (with 6 clusters). Forth row: estimation of the reflectance using the proposed approach (with $\lambda = 2.5$ and $\mu = 1000$). Fifth row: ground truth.

Fig. 8 Same test as in Figure 7, using a coarse version of the 3D-shape (see Figures 5-b and 5-d), with $\lambda = 2.5$ and $\mu = 1000$. Results are qualitatively similar to those shown in Figure 7, obtained with perfect geometry. The RMSE in the RGB channels are, respectively: 0.24, 0.14 and 0.13, which are only slightly higher than those attained with perfect geometry (see Table 1).

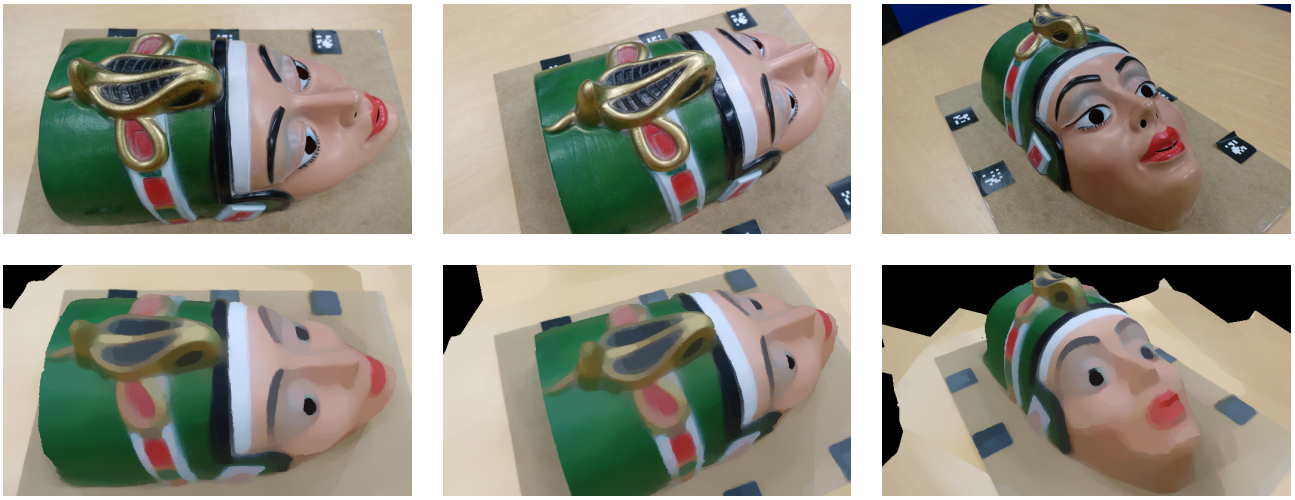


Fig. 9 Test on a real-world dataset. First row: three (out of $m = 8$) views of the scene. Second row: estimated reflectance maps using the proposed approach (with $\lambda = 2$ and $\mu = 1000$). Geometry and camera parameters were estimated using an SfM/MVS pipeline.

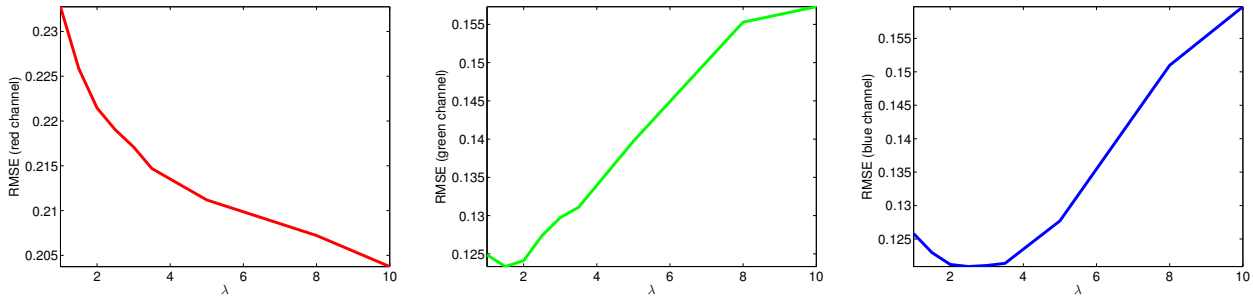


Fig. 10 Quantitative influence of parameter λ, using images from the same dataset as that of Figure 7, with $\mu = 1000$.

To avoid this effect, the smoothness term (47) must be activated by setting $\lambda > 0$. If we still consider $\mu = 0$, then the variational problem (45) comes down to m independent image restoration problems. In fact, these problems are similar to ℓ^1-TV denoising problems, except that a physically plausible fidelity term is used to help removing the illumination artifacts not only from the total variation regularization, but also by incorporating prior knowledge of the surface geometry. However, because the photometric term (46) is invariant by the transformation $(\rho^i, \boldsymbol{\sigma}^i) := (\kappa^i \rho^i, \boldsymbol{\sigma}^i / \kappa^i)$, $\kappa^i > 0$, each reflectance map ρ^i is estimated only up to a scale factor, hence the m maps will not be consistent, as is the case for the competing single-view methods.

The latter issue is solved by activating the multi-view consistency term (48) *i.e.*, by setting $\mu > 0$. In that case, there is still an ambiguity $\{\rho^i, \boldsymbol{\sigma}^i\}_i := \{\kappa\rho^i, \boldsymbol{\sigma}^i/\kappa\}$, $\kappa > 0$, but this ambiguity is now global *i.e.*, independent from i. It is enough in practice to set one reflectance value arbitrarily, or to normalize the reflectance values, to solve the ambiguity.

Overall, it is necessary to ensure that both λ and μ are strictly positive. The choice of μ is not really critical. Indeed, the multi-view consistency regularizer which is controlled by μ arises from relaxing a hard constraint (see (40) and (45)). Hence, μ only needs to be chosen "high enough" so that the regularizer approximates fairly well a hard constraint. In all the experiments, we used $\mu = 1000$ and did not face any particular problem. Obviously, if the correspondences were not appropriately computed by SfM, then this value should be reduced, but SfM solutions such as [27] are now mature enough to provide accurate correspondences.

The choice of λ is much more critical. This is illustrated in Figure 10, which shows the RMSE in each channel at convergence of our algorithm, as a function of λ. This graph shows that the "optimal" value of λ is very hard to find: in this example, a high value of λ would diminish the RMSE in the face and the hair (which are mostly red), because this would make them uniform as expected (see Figure 11, last rows). However, a much lower value of λ is required in order to preserve the thin shirt details, which mostly contain green and blue components (see Figure 11, first rows).

There is one situation where this tuning is much easier. It is when the lighting is not fixed, but strongly varying. As discussed in Section 3, the problem of jointly estimating reflectance and lighting is then over-determined, which theoretically makes the regularization unnecessary. In Figure 12, we show the results obtained in the case where each image is obtained under a different lighting. In that case, the thin structures of the shirt are preserved, while shading on the face is largely reduced, despite the choice of a very low regularization weight $\lambda = 1$. Note that we cannot use the limit case $\lambda = 0$ because not all pixels have correspondences in all images: there may thus be a few pixels for which the problem remains under-determined, and for which diffusion is required. Overall, this experiment shows that the only way to avoid introducing an empirical prior on the reflectance, and thus its tuning, is to actively control lighting during the acquisition process. This means, combining multi-view and photometric stereo.

It happens that this problem is actively being addressed by the computer vision community [29]. Interestingly, in this research the focus is put on highly accurate geometry estimation, and not so much on reflectance estimation (no reflectance estimation result is shown). Therefore, it may be an interesting future research direction to incorporate our reflectance estimation framework in such multi-view, multi-lighting approaches. Both highly accurate geometry and reflectance could indeed be expected.

7 Conclusion and Perspectives

We have proposed a variational framework for estimating the reflectance of a scene from a series of multi-view images. We advocate a 2D parameterization of reflectance, turning the problem into that of converting the input images into reflectance maps. Invoking a Bayesian rationale leads to a variational model comprising a ℓ^1-norm-based photometric data term, a Potts regularizer and a multi-view consistency constraint. For simplicity, both the latter are relaxed into a total variation term and a ℓ^1-norm term, respectively. Numerical solving is carried out using an alternating majorization-minimization algorithm. Empirical results on both synthetic and real-world datasets demonstrate the interest of considering multi-view images for reflectance estimation, as it allows to benefit from prior knowledge of the geometry, to improve robustness to specularities and to guarantee consistency of the reflectance estimates.

However, the critical analysis of our results also highlighted some limitations and possible future research directions. For instance, avoiding the relaxation of the non-smooth, non-convex regularization, seems to be necessary in order to really ensure that the estimated reflectance maps are piecewise-constant. In addition, the choice of parameterizing reflectance in the image (2D) domain is advocated for reasons of numerical simplicity, yet it seems somewhat more natural to work directly on the surface (this would avoid the multi-view consistency constraint). However, this would require turning our simple variational framework into a more arduous optimization problem over a manifold. Finally, it seems that the only way to avoid resorting to an arbitrary prior for limiting the arising ambiguities consists in actively controlling the lighting (this would avoid resorting to spatial regularization). Therefore, another extension of our work consists in estimating reflectance from multi-view, multi-lighting data, in the spririt of multi-view photometric stereo techniques. However, this would require appropriately modifying the SfM/MVS pipeline, which relies on the constant brightness assumption.

References

1. Adelson, E.H., Pentland, A.P.: Perception as Bayesian inference, chap. The perception of shading and reflectance,

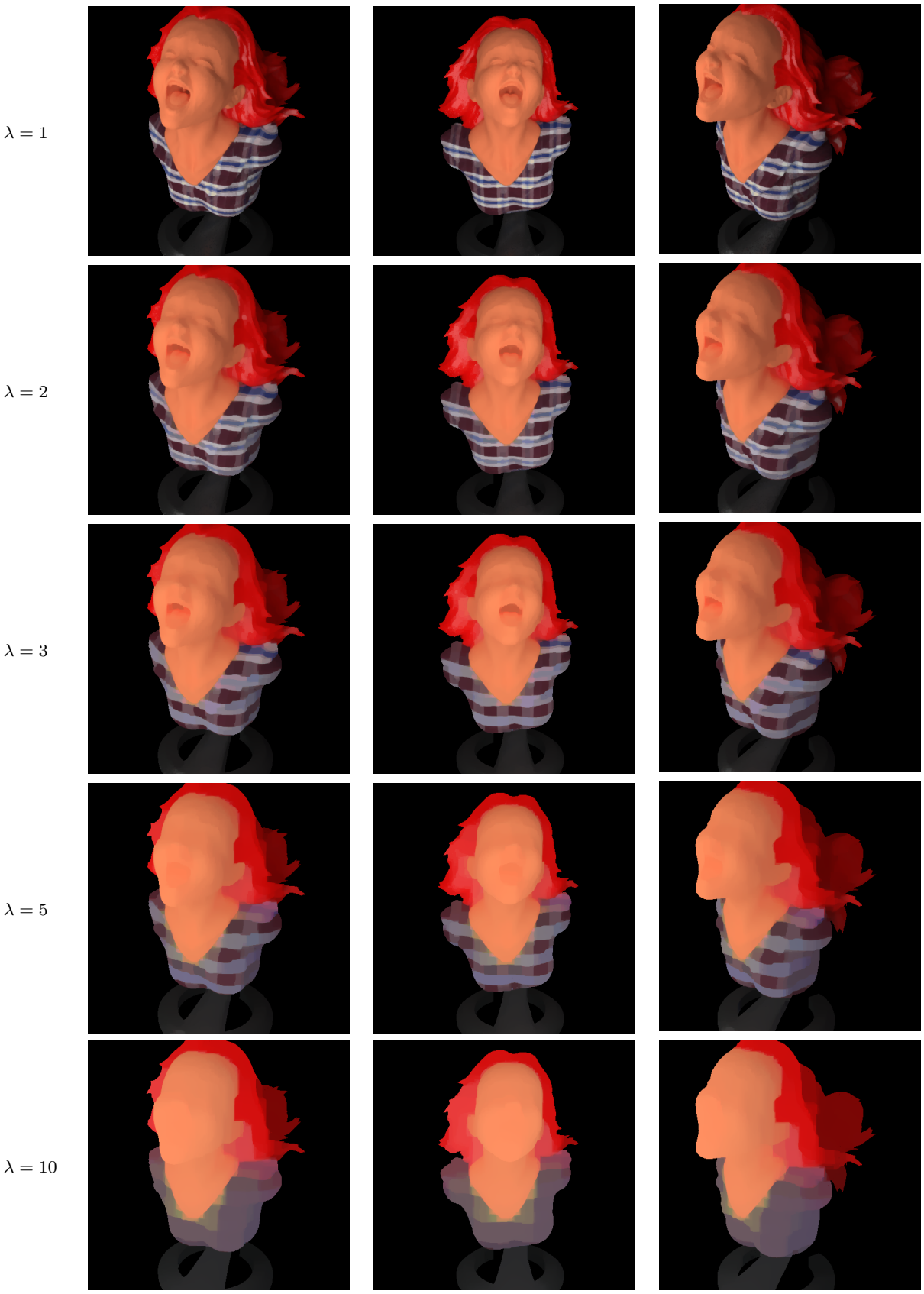


Fig. 11 Qualitative influence of parameter λ, using images from the same dataset as that of Figure 7, with $\mu = 1000$.

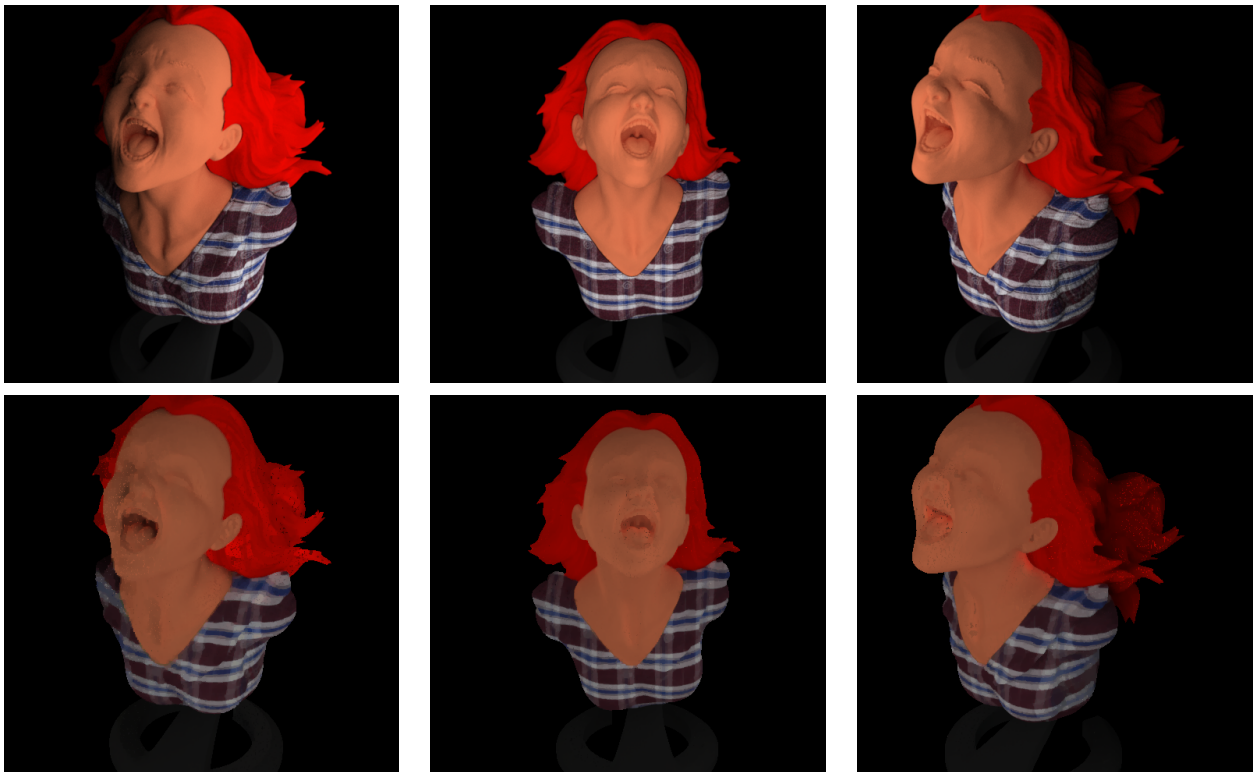


Fig. 12 First row: three (out of $m = 13$) synthetic images computed under varying lighting (which comes here from the right, from the front and from the left, respectively). Second row: estimated reflectance maps using the proposed approach (with $\lambda = 1$ and $\mu = 1000$). The thin structures of the shirt are preserved, while shading on the face is largely reduced. These results must be compared with those of the first row in Figure 11, obtained with the same value of λ but under fixed lighting.

pp. 409–423. Cambridge University Press (1996) 1, 2
2. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building Rome in a Day. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 72–79 (2009) 4
3. Aujol, J.F., Gilboa, G., Chan, T., Osher, S.: Structure-Texture Image Decomposition – Modeling, Algorithms, and Parameter Selection. International Journal of Computer Vision 67(1), 111–136 (2006) 3
4. Barron, J., Malik, J.: Shape, illumination, and reflectance from shading. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(8), 1670–1687 (2015) 4
5. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric Stereo with General, Unknown Lighting. International Journal of Computer Vision 72(3), 239–257 (2007) 3, 5
6. Basri, R., Jacobs, D.P.: Lambertian reflectances and linear subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(2), 218–233 (2003) 4
7. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. ACM Transactions on Graphics 33(4), 159:1–159:12 (2014) 4
8. Bousseau, A., Paris, S., Durand, F.: User assisted intrinsic images. ACM Transactions on Graphics 28(5), 130:1–130:10 (2009) 4
9. Chen, Q., Koltun, V.: A simple model for intrinsic image decomposition with depth cues. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 241–248 (2013) 4
10. Cho, D., Matsushita, Y., Tai, Y.W., Kweon, I.S.: Photometric stereo under non-uniform light intensities and exposures. In: Proceedings of the European Conference on Computer Vision. pp. 170–186 (2016) 5
11. Frolova, D., Simakov, D., Basri, R.: Accuracy of spherical harmonic approximations for images of lambertian objects under far and near lighting. In: Proceedings of the European Conference on Computer Vision. pp. 574–587 (2004) 4
12. Furukawa, Y., Hernández, C., et al.: Multi-view stereo: A tutorial. Foundations and Trends® in Computer Graphics and Vision 9(1-2), 1–148 (2015) 2
13. Garces, E., Munoz, A., Lopez-Moreno, J., Gutierrez, D.: Intrinsic Images by Clustering. Computer Graphics Forum 31(4), 1415–1424 (2012) 4
14. Gehler, P., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering Intrinsic Images with a Global Sparsity Prior on Reflectance. In: Advances in Neural Information Processing Systems. pp. 765–773 (2011) 4, 11, 12, 13, 14, 15
15. Golub, G.H., Van Loan, C.F.: Matrix Computations (4th Ed.). Johns Hopkins University Press (2013) 6
16. Horn, B.K.P.: Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View. Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (1970) 2
17. Jin, H., Cremers, D., Wang, D., Yezzi, A., Prados, E., Soatto, S.: 3-D Reconstruction of Shaded Objects from Multiple Images Under Unknown Illumination. International Journal of Computer Vision 76(3), 245–256 (2008) 4

18. Kim, K., Torii, A., Okutomi, M.: Multi-view Inverse Rendering Under Arbitrary Illumination and Albedo. In: Proceedings of the European Conference on Computer Vision. pp. 750–767 (2016) 4
19. Laffont, P.Y., Bousseau, A., Drettakis, G.: Rich Intrinsic Image Decomposition of Outdoor Scenes from Multiple Views. IEEE Transactions on Visualization and Computer Graphics 19(2), 210–224 (2013) 4
20. Laffont, P.Y., Bousseau, A., Paris, S., Durand, F., Drettakis, G.: Coherent intrinsic images from photo collections. ACM Transactions on Graphics 31, 202:1–202:11 (2012) 4
21. Land, E.H., J., M.J.: Lightness and retinex theory. Journal of the Optical Society of America 61, 1–11 (1971) 3, 4
22. Langguth, F., Sunkavalli, K., Hadap, S., Goesele, M.: Shading-aware Multi-view Stereo. In: Proceedings of the European Conference on Computer Vision. pp. 469–485 (2016) 4
23. Le Guen, V.: Cartoon + Texture Image Decomposition by the TV-L1 Model. Image Processing On Line 4, 204–219 (2014) 3, 11, 12, 13, 14, 15
24. Maier, R., Kim, K., Cremers, D., Kautz, J., Nießner, M.: Intrinsic3d: High-quality 3D reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In: Proceedings of the IEEE International Conference on Computer Vision (2017) 4
25. Maurer, D., Ju, Y.C., Breuß, M., Bruhn, A.: Combining Shape from Shading and Stereo: A Variational Approach for the Joint Estimation of Depth, Illumination and Albedo. In: Proceedings of the British Machine Vision Conference (2016) 4
26. Mélou, J., Quéau, Y., Durou, J.D., Castan, F., Cremers, D.: Beyond Multi-view Stereo: Shading-Reflectance Decomposition. In: Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision. pp. 694–705 (2017) 3
27. Moulon, P., Monasse, P., Marlet, R.: openMVG: An open multiple view geometry library. https://github.com/openMVG/openMVG (2014) 4, 17
28. Nadian-Ghomsheh, A., Hassanian, Y., Navi, K.: Intrinsic Image Decomposition via Structure-Preserving Image Smoothing and Material Recognition. PLoS ONE 11(12), 1–22 (2016) 4
29. Park, J., Sinha, S.N., Matsushita, Y., Tai, Y.W., Kweon, I.S.: Robust multiview photometric stereo using planar mesh parameterization. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(8), 1591–1604 (2017) 17
30. Quéau, Y., Durix, B., Wu, T., Cremers, D., Lauze, F., Durou, J.D.: LED-based Photometric Stereo: Modeling, Calibration and Numerical Solution. Journal of Mathematical Imaging and Vision (2017), (to appear) 11
31. Quéau, Y., Pizenberg, M., Durou, J.D., Cremers, D.: Microgeometry capture and RGB albedo estimation by photometric stereo without demosaicing. In: Proceedings of QCAV. Tokyo, Japon (2017) 2
32. Ramamoorthi, R., Hanrahan, P.: An Efficient Representation for Irradiance Environment Maps. In: Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques. pp. 497–500 (2001) 4
33. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. vol. 1, pp. 519–528 (2006) 4
34. Shen, L., Yeo, C.: Intrinsic images decomposition using a local and global sparse representation of reflectance. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 697–704 (2011) 4
35. Song, J., Cho, H., Yoon, J., Yoon, S.M.: Structure adaptive total variation minimization-based image decomposition. IEEE Transactions on Circuits and Systems for Video Technology (2017), (to appear) 4
36. Storath, M., Weinmann, A.: Fast partitioning of vector-valued images. SIAM Journal on Imaging Sciences 7(3), 1826–1852 (2014) 9
37. Woodham, R.J.: Photometric Method for Determining Surface Orientation from Multiple Images. Optical Engineering 19(1), 139–144 (1980) 2, 3
38. Wu, C., Wilburn, B., Matsushita, Y., Theobalt, C.: High-quality shape from multiview stereo and shading under general illumination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 969–976 (2011) 4