HAL
open science

Beyond Multi-view Stereo: Shading-Reflectance Decomposition

Jean Mélou, Yvain Quéau, Jean-Denis Durou, Fabien Castan, Daniel Cremers

To cite this version:

Jean Mélou, Yvain Quéau, Jean-Denis Durou, Fabien Castan, Daniel Cremers. Beyond Multi-view Stereo: Shading-Reflectance Decomposition. 6th International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2017), Jun 2017, Kolding, Denmark. pp. 694-705, 10.1007/978-3-319-58771-4_55 . hal-01787432

HAL Id: hal-01787432
https://hal.science/hal-01787432

Submitted on 7 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 18993

The contribution was presented at SSVM 2017 : http://ssvm2017.compute.dtu.dk/
To link to this article URL :
http://dx.doi.org/10.1007/978-3-319-58771-4_55

To cite this version : Mélou, Jean and Quéau, Yvain and Durou, Jean-Denis and Castan, Fabien and Cremers, Daniel *Beyond Multi-view Stereo: Shading-Reflectance Decomposition.* (2017) In: 6th International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2017), 4 June 2017 - 8 June 2017 (Kolding, Denmark).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Beyond Multi-view Stereo: Shading-Reflectance Decomposition

Jean Mélou[1,3(✉)], Yvain Quéau[2], Jean-Denis Durou[1], Fabien Castan[3], and Daniel Cremers[2]

[1] IRIT, UMR CNRS 5505, Université de Toulouse, Toulouse, France
[2] Department of Informatics, Technical University Munich, Munich, Germany
[3] Mikros Image, Levallois-Perret, France
jeme@mikrosimage.eu

Abstract. We introduce a variational framework for separating shading and reflectance from a series of images acquired under different angles, when the geometry has already been estimated by multi-view stereo. Our formulation uses an l^1-TV variational framework, where a robust photometric-based data term enforces adequation to the images, total variation ensures piecewise-smoothness of the reflectance, and an additional multi-view consistency term is introduced for resolving the arising ambiguities. Optimisation is carried out using an alternating optimisation strategy building upon iteratively reweighted least-squares. Preliminary results on both a synthetic dataset, using various lighting and reflectance scenarios, and a real dataset, confirm the potential of the proposed approach.

Keywords: Reflectance · Multi-view · Shading · Variational methods

1 Introduction

Acquiring the shape and the reflectance of a scene is a key issue for the movie industry, as it allows proper relighting. Well-established shape acquisition techniques such as multi-view stereo can provide accurate 3D-reconstructions in a robust manner. 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 characterises the surface), relighting based on this approach is usually unsatisfactory. To improve results, the reflectance needs to be further extracted. As shown in Fig. 1, our aim is to separate shading and reflectance using several images of a surface taken under different angles, assuming known (but possibly gross) geometry.

We formulate this task as a variational problem, introducing the knowledge that the albedo (we assume Lambertian surface, hence albedo characterizes reflectance) is independent from the viewing angle as a prior within the variational model. We further tackle the problem of robustness to specularities and imperfect alignment, using a robust l^1-norm-based photometric

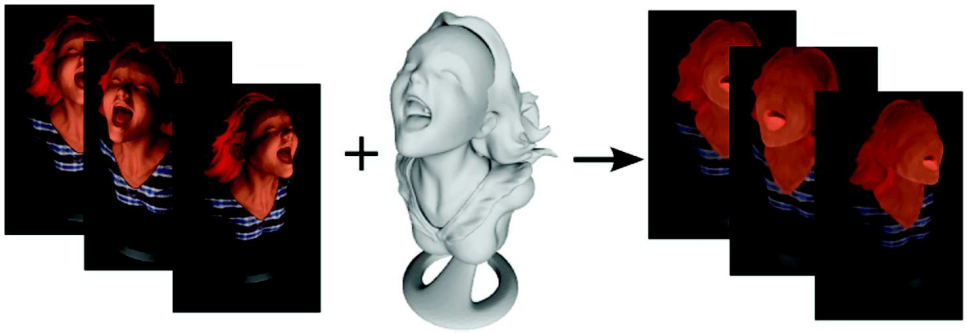

Fig. 1. Overview of our contribution. From a set of n images of a surface acquired under different angles, and a coarse geometry obtained for instance using multi-view stereo, we estimate a shading-free reflectance map per view. (Color figure online)

data term. Assuming that the albedo is piecewise-smooth, we further involve a TV-regularizer.

After reviewing related approaches in Sect. 2, we present our variational solution in Sect. 3. Our numerical strategy for solving this variational problem, which is presented in Sect. 4, is based on alternating optimisation of reflectance and lighting, and on an iteratively reweighted least-squares strategy for handling the l^1-terms. Preliminary experiments on both synthetic and real data are conducted in Sect. 5, which confirm the interest of multi-view shading-reflectance decomposition strategy. Our work is eventually summarised in Sect. 6.

2 Related Work

The problem of decomposing an image into a low-frequency component (representing reflectance) and a higher-frequency one has been tackled in various ways. One famous example is the so-called "cartoon-texture" image decomposition [2], which can be achieved efficiently using an l^1-TV approach [8]. Yet, such methods may fail in the presence of smooth brightness variations due to shading.

In this view, a photometric model would be helpful, as it would explicitely describe the interactions between the geometry, the potentially complex lighting and the reflectance. Yet, it is not possible to estimate all these parameters from a single image. In fact, estimating only the geometry, with known reflectance and lighting, is already an ill-posed problem, known as shape-from-shading [4]. It can be disambiguated by using several images obtained from the same angle, but under varying lighting, a variant known as photometric stereo [14] which can simultaneously recover shape, reflectance and lighting [3].

Still, photometric stereo requires very controlled acquisition environments, which limit potential applications. Multi-view stereo (MVS) methods [12], which focus on estimating the shape, are less restrictive. These 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 [10]. Indeed, MVS requires the cameras' parameters, outputs of the SfM algorithm. Nowadays,

these mature methods are commonly used in uncontrolled environments, or even with large scale Internet data [1].

Considering strengths and weaknesses of both photometric and multi-view methods, a joint approach may help recovering a precise 3D-reconstruction as well as albedo and lighting information. For instance, the method in [5] iteratively refines the object surface using shading information, assuming that the albedo is constant or piecewise-constant [5]. This assumption was recently relaxed in [6,7], by resorting to properly designed priors on shape, shading and reflectance.

Still, these methods aim at refining a mesh by adding fine-scaled surface details, along with the associated reflectance information. As a result, the complexity of the mesh may be a source of concern when it comes to real-time manipulation by graphic artists. We argue that it may be more meaningful to keep the geometry simple, yet providing the artists with a series of 2D-maps representing reflectance and fine-scale geometry, under the form of albedo and depth maps. A first step in this direction has recently been achieved in [9], where a variational framework for the joint recovery of depth, reflectance and lighting maps is introduced. Yet, this approach relies on the choice of a single reference view, the other views being used only for the sake of stereo matching.

On the contrary, we target a symmetric approach, where all the reflectance maps corresponding to the different views are simultaneously recovered, thus avoiding the problem of selecting the main image. We focus in this exploratory work on separating shading and reflectance, using the MVS results, and not yet on refining fine-scale geometric surface details. Indeed, we will see that even with a smoothed geometry, reasonable reflectance estimation can be carried out. In this view, the next section introduces a simple and effective l^1-TV variational model for estimating a set of reflectance maps, assuming known geometry.

3 Joint Variational Estimation of the Albedo and of Spherical Harmonics Lighting

We consider a set of n pictures of an object, obtained under different angles (and, possibly, different lighting) $\{I^i : \Omega^i \subset \mathbb{R}^2 \to \mathbb{R}\}_{i=1\ldots n}$, where Ω^i represents the mask of the object of interest in image I^i. These masks are assumed to be known, as well as a (possibly inaccurate) representation of the geometry under the form of normal maps $\mathbf{n}^i : \Omega^i \to \mathbb{R}$ (which can be obtained by using SfM and MVS).

Our aim is to extract, from each image I^i, a representation of the surface reflectance through an albedo map $\rho^i : \Omega^i \to \mathbb{R}$, and an estimate of the lighting in this image. Assuming Lambertian reflectance and general lighting, a second-order spherical harmonics lighting model can be used [11], and thus the i-th lighting can be represented by a vector $\boldsymbol{\sigma}^i \in \mathbb{R}^9$. Our problem then comes down to solving the following set of equations[1]:

$$I^i(p) = \rho^i(p)\boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i(p), \quad \forall p \in \Omega^i,\ \forall i = 1 \ldots n, \tag{1}$$

where vectors $\boldsymbol{\nu}^i(p) \in \mathbb{R}^9$ can be deduced from the normals coordinates [3].

[1] This model is valid for greyscale images. To handle RGB images, our approach can be applied independently to each color channel.

Obviously, it is not possible to solve the set of Eq. (1) without introducing additional priors. Although a *global* (*i.e.*, same for each image and each pixel) scale ambiguity on the albedo is acceptable (because these values can always be normalized), the set of Eq. (1) exhibits such an ambiguity for each image. We tackle this issue by proposing a multi-view consistency prior on the albedo. Indeed, the albedo characterizes the surface, and is thus independent from the view. Besides, to ensure spatial consistency of the albedo estimate, a total variation prior is also introduced. Eventually, to ensure robustness to specularities, we solve the set of Eq. (1) in the l^1-norm sense. Overall, this leads us to estimate $\{\boldsymbol{\sigma^i} \in \mathbb{R}^9, \rho^i : \Omega^i \to \mathbb{R}\}_{i=1\ldots n}$ as minimisers of the following energy:

$$\varepsilon(\{\boldsymbol{\sigma^i}, \rho^i\}_i) = \sum_{i=1}^{n} \varepsilon_{\text{Photo}}(\boldsymbol{\sigma^i}, \rho^i) + \lambda \sum_{i=1}^{n} \varepsilon_{\text{Smooth}}(\rho^i) + \mu \sum_{i<j}\sum \varepsilon_{\text{MV}}(\rho^i, \rho^j). \quad (2)$$

In Eq. (2), the first component ensures photometric consistency:

$$\varepsilon_{\text{Photo}}(\boldsymbol{\sigma^i}, \rho^i) = \sum_{p \in \Omega^i} |I^i(p) - \rho^i(p)\boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i(p)|, \quad (3)$$

the second one ensures albedo smoothness:

$$\varepsilon_{\text{Smooth}}(\rho^i) = \sum_{p \in \Omega^i} |\partial_x \rho^i(p)| + |\partial_y \rho^i(p)|, \quad (4)$$

where $\nabla \rho^i(p) = \left[\partial_x \rho^i(p), \partial_y \rho^i(p)\right]^\top$ represents the gradient of ρ^i at pixel p (approximated, in practice, using first-order forward stencils), and the third component ensures multi-view consistency of the albedo estimate:

$$\varepsilon_{\text{MV}}(\rho^i, \rho^j) = \sum_{p^i \in \Omega^i} \sum_{p^j \in \Omega^j} C_{i,j}(p^i, p^j) |\rho^i(p^i) - \rho^j(p^j)|, \quad (5)$$

where $C_{i,j}$ is a "correspondence function" defined as follows:

$$C_{i,j}(p^i, p^j) = \begin{cases} 1 \text{ if pixels } p^i \text{ and } p^j \text{ correspond to the same surface point;} \\ 0 \text{ otherwise.} \end{cases} \quad (6)$$

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

The values of function $C_{i,j}$ are easily deduced from an initial geometry estimate, as proposed for instance by Langguth *et al.* in [7] for their evaluation of geometric error.

Applying an SfM algorithm to the images $\{I^i\}_{i=1\ldots n}$, we obtain the cameras intrinsics $\mathbf{K}^i \in \mathbb{R}^{3\times 3}$, and the poses of the cameras, represented by rotation matrices $\mathbf{R}^i \in \mathbb{R}^{3\times 3}$ and translation vectors $\boldsymbol{t}^i \in \mathbb{R}^3$. Then, a point $X \in \mathbb{R}^3$ on the surface is projected to a pixel $p^i \in \mathbb{R}^2$ in the i-th image according to

$$[{p^i}^\top, 1]^\top = \pi_i(X) = \mathbf{K}^i(\mathbf{R}^i X + \boldsymbol{t}^i), \quad (7)$$

where π_i stands for the projection from surface to image I^i.

Hence, the correspondence function $C_{i,j}$ defined in Eq. (6) can be redefined as follows, introducing some threshold $\epsilon > 0$ (we use $\epsilon = 3$, in the experiments) and using the known depth (obtained by MVS) to compute the inverse projections:

$$C_{i,j}(p^i, p^j) = \begin{cases} 1 \text{ if } \left\| {\pi_i}^{-1}([{p^i}^\top, 1]^\top) - {\pi_j}^{-1}([{p^j}^\top, 1]^\top) \right\| \leq \epsilon, \\ 0 \text{ otherwise.} \end{cases} \tag{8}$$

4 Resolution

Let us now introduce our numerical strategy for minimising the energy (2). This problem being bi-convex, we opt for an alternating estimation strategy: at iteration (k), we successively update the lighting and the albedo as:

$$\{\boldsymbol{\sigma}^{i,(k+1)}\}_i = \underset{\{\boldsymbol{\sigma}^i \in \mathbb{R}^9\}_i}{\operatorname{argmin}} \, \varepsilon(\{\boldsymbol{\sigma^i}, \rho^{i,(k)}\}_i), \tag{9}$$

$$\{\rho^{i,(k+1)}\}_i = \underset{\{\rho^i \colon \Omega^i \to \mathbb{R}\}_i}{\operatorname{argmin}} \, \varepsilon(\{\boldsymbol{\sigma}^{i,(k+1)}, \rho\}_i), \tag{10}$$

taking as initial guess $\rho^{i,(0)} \equiv I^i$ and $\boldsymbol{\sigma}^{i,(0)} = \mathbf{1}_{\mathbb{R}^9}$.

To handle the non-differentiable l^1-norm terms, we opt for an iteratively reweighted least-squares approach. Since the n lighting vectors $\boldsymbol{\sigma}^i$, $i = 1 \ldots n$, are independent, Eq. (9) is then replaced by the following n independent reweighted least-squares updates, which can be solved by resorting to the pseudo-inverse:

$$\boldsymbol{\sigma}^{i,(k+1)} = \underset{\boldsymbol{\sigma}^i \in \mathbb{R}^9}{\operatorname{argmin}} \sum_{p \in \Omega^i} w_i^{(k)}(p) |I^i(p) - \rho^{i,(k)}(p) \, \boldsymbol{\sigma}^i \cdot \boldsymbol{\nu}^i(p)|^2, \quad \forall i = 1 \ldots n, \tag{11}$$

with

$$w_i^{(k)}(p) = \frac{1}{|I^i(p) - \rho^{i,(k)}(p) \, \boldsymbol{\sigma}^{i,(k)} \cdot \boldsymbol{\nu}^i(p)|_\delta}, \tag{12}$$

where we denote $|\cdot|_\delta = \max\{\delta, |\cdot|\}$ (we use $\delta = 10^{-4}$, in the experiments).

Similarly, albedo update (10) is approximated as follows:

$$\begin{aligned} \{\rho^{i,(k+1)}\}_i = \underset{\{\rho^i \colon \Omega^i \to \mathbb{R}\}_i}{\operatorname{argmin}} & \sum_{i=1}^{n} \sum_{p \in \Omega^i} w_i^{(k)}(p) |I^i(p) - \rho^i(p) \, \boldsymbol{\sigma}^{i,(k)} \cdot \boldsymbol{\nu}^i(p)|^2 \\ & + \lambda \sum_{i=1}^{n} \sum_{p \in \Omega^i} w_{\partial_x \rho^i}^{(k)}(p) |\partial_x \rho^i(p)|^2 + w_{\partial_y \rho^i}^{(k)}(p) |\partial_y \rho^i(p)|^2 \\ & + \mu \sum_{i<j} \sum \sum_{p^i \in \Omega^i} \sum_{p^j \in \Omega^j} C_{i,j}(p^i, p^j) w_{i,j}^{(k)}(p^i, p^j) |\rho^i(p^i) - \rho^j(p^j)|^2, \end{aligned} \tag{13}$$

with

$$w^{(k)}_{\partial_x \rho^i}(p) = \frac{1}{|\partial_x \rho^{i,(k)}(p)|_\delta}, \tag{14}$$

$$w^{(k)}_{\partial_y \rho^i}(p) = \frac{1}{|\partial_y \rho^{i,(k)}(p)|_\delta}, \tag{15}$$

$$w^{(k)}_{i,j}(p^i, p^j) = \frac{1}{|\rho^{i,(k)}(p^i) - \rho^{j,(k)}(p^j)|_\delta}. \tag{16}$$

This time, due to the multi-view consistency prior, the albedo estimates w.r.t. the different images are not independent: all estimations must be carried out simultaneously. Stacking all the albedo values in a large vector $\boldsymbol{\rho} \in \mathbb{R}^N$, with $N = \sum_i |\Omega^i|$, the optimisation problem (13) can be turned into the following large, sparse, linear least-squares problem:

$$\boldsymbol{\rho}^{(k+1)} = \underset{\boldsymbol{\rho} \in \mathbb{R}^N}{\operatorname{argmin}} \left\| \begin{bmatrix} \operatorname{Diag}(\{\sqrt{w^{(k)}_i(p)}\, \boldsymbol{\sigma}^{i,(k+1)} \cdot \boldsymbol{\nu}^i(p)\}_{i,p}) \\ \sqrt{\lambda}\, \operatorname{Diag}(\{\sqrt{w^{(k)}_{\partial_x \rho^i}(p)}\}_{i,p})\ \mathbf{D}_x \\ \sqrt{\lambda}\, \operatorname{Diag}(\{\sqrt{w^{(k)}_{\partial_y \rho^i}(p)}\}_{i,p})\ \mathbf{D}_y \\ \sqrt{\mu}\, \operatorname{Diag}(\{\sqrt{w^{(k)}_{i,j}(p^i, p^j)}\}_{i,j,p^i,p^j})\ \mathbf{C} \end{bmatrix} \boldsymbol{\rho} - \begin{bmatrix} \operatorname{Diag}(\{\sqrt{w^{(k)}_i(p)}\}_{i,p}) \boldsymbol{I} \\ \mathbf{0}_{N \times 1} \\ \mathbf{0}_{N \times 1} \\ \mathbf{0}_{\sum_{i=1}^{n-1}(n-i)|\Omega^i| \times 1} \end{bmatrix} \right\|_2^2. \tag{17}$$

In the least-squares problem (17), the first matrix block stacks the weighted shading values, $\mathbf{D}_x$ and $\mathbf{D}_y$ are large sparse matrices obtained by concatenating the n finite differences matrices relative to each domain Ω^i, $i = 1 \dots n$, and the last one is a large $\sum_{i=1}^{n-1}(n-i)|\Omega^i| \times N$ matrix used to represent the correspondence functions defined in Eq. (6). As for the non-null vector in the second row of Eq. (17), it stacks all the intensity values in $\boldsymbol{I}$, and weights them. Each function $C_{i,j}$ is easily represented as an $|\Omega^i| \times |\Omega^j|$ matrix $\mathbf{C}_{i,j}$ with at most one nonzero element per row. By arranging these matrices by block in a matrix $\mathbf{C}$, all values $C_{i,j}(p^i, p^j)\left(\rho^i(p^i) - \rho^j(p^j)\right)$ can be compactly represented in matrix form as $\mathbf{C}\boldsymbol{\rho}$. For instance, considering an $n = 4$ pictures set, we get the following matrix:

$$\mathbf{C} = \begin{pmatrix} \mathbf{B}_{1,2} & \mathbf{C}_{1,2} & & \\ \mathbf{B}_{1,3} & & \mathbf{C}_{1,3} & \\ \mathbf{B}_{1,4} & & & \mathbf{C}_{1,4} \\ & \mathbf{B}_{2,3} & \mathbf{C}_{2,3} & \\ & \mathbf{B}_{2,4} & & \mathbf{C}_{2,4} \\ & & \mathbf{B}_{3,4} & \mathbf{C}_{3,4} \end{pmatrix}, \tag{18}$$

where $\mathbf{B}_{i,j}$ is a diagonal $|\Omega^i| \times |\Omega^i|$ matrix with entries equal to -1 on the lines where $\mathbf{C}_{i,j}$ is non-null.

Problem (17) is a linear least-squares problem where the matrix is very sparse. For its resolution, we apply a conjugate gradient algorithm to the associated normal equations. We iterate optimisation steps (9) and (10) until convergence or a maximum iteration number is reached. In our experiments, we found 50 iterations were always sufficient to reach a stable solution (10^{-3} relative residual between two consecutive energy values). Proving convergence of our scheme is beyond the scope of this paper, but it was empirically observed in all experiments, although the convergence rate seems to be sublinear.

5 Results

We first test our shading-reflectance decomposition method in a very simple situation. Let us simulate $n = 13$ pictures of the object in Fig. 2-a, supposed purely-Lambertian, with a camera whose intrinsic and extrinsic parameters are known, under a "sky-dome" lighting. Since the object geometry is known, the problem unknows are reflectance and lighting in each surface point.

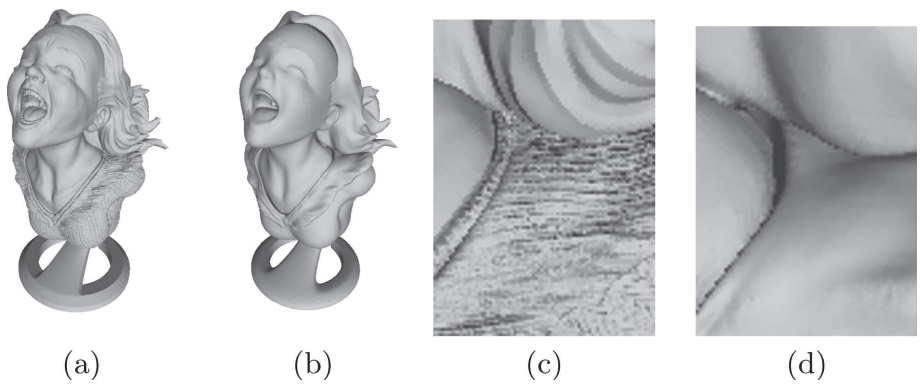

Fig. 2. (a) 3D-shape used in the tests (the well-known "Joyful Yell" 3D-model), which will be covered with two different albedos. (b) Same 3D-shape after smoothing, thus less accurate. (c)–(d) Zooms of (a) and (b), respectively, near the neck.

Figure 3 shows three of these pictures, of size 540×960, generated using a renderer, and the estimated colored albedo using our method.

As expected, since the lighting used in this simulation (sky-dome) is the most appropriate to a spherical harmonics modelisation, these first results are very satisfactory. As a comparison, the third row of Fig. 3 shows the results of the cartoon-texture decomposition method described in [8]. This method needs only one image. The "cartoon" part, which is more or less equivalent to an albedo, is far less uniform than the albedo estimated using shading-reflectance decomposition, in the four parts (hair, face, shirt and plinth) which have received a uniform albedo (compare the second and third lines of Fig. 3).

This comparison a posteriori confirms our basic idea i.e., that reflectance estimation benefits in two ways from the multi-view framework: indeed, this

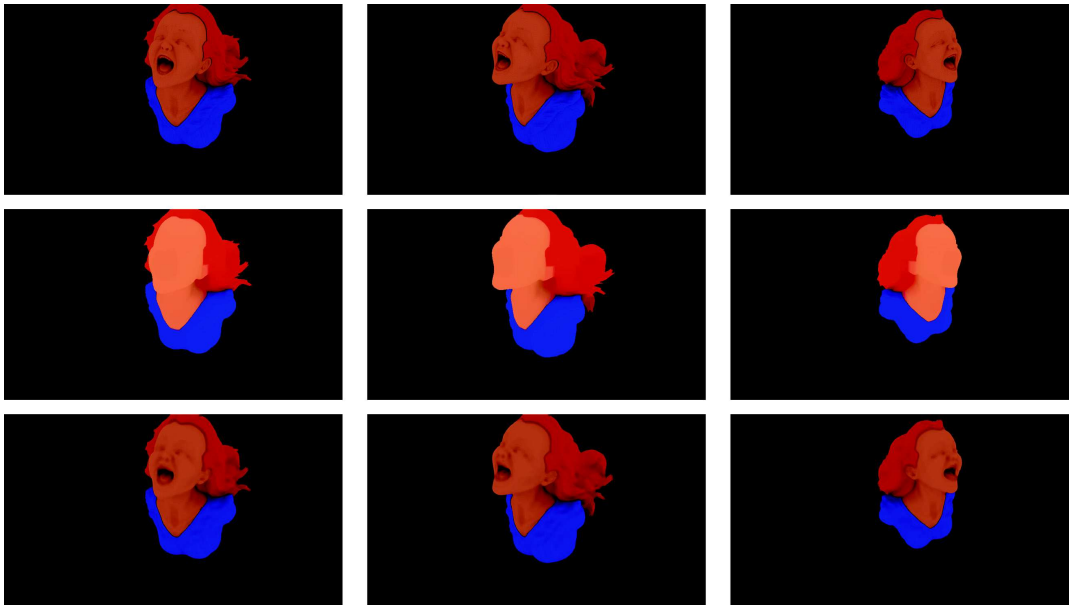

Fig. 3. First line: Three (out of $n = 13$) synthetic views of the object of Fig. 2-a, computed with a purely-Lambertian reflectance divided into four parts (hair, face, shirt and plinth) which receive a uniform albedo, under "sky-dome" lighting. Second line: Colored estimated albedos, using the proposed approach. Geometry and camera parameters are supposed to be known, but the lighting is unknown. Third line: Empirical estimation of the albedo using the cartoon-texture decomposition described in [8]. (Color figure online)

allows us not only to estimate the 3D-shape, but also to constrain the albedo of each surface point to be the same in all the pictures where it is visible. In contrast, the cartoon-texture decomposition cannot correct the shading effects, which explains, for instance, why the cartoon is so dark inside the mouth.

As we dispose of the albedo ground truth, we can numerically evaluate these results by estimating the albedo variance in the $n = 13$ pictures, in each part of the object where the albedo is uniform[2]. The values presented in Table 1 confirm that our estimation is more accurate. As well, we observe that the albedo variance is higher for both zones which have concave parts, namely hair and face (in this last case, the albedo is largely under-estimated in the mouth). Indeed, such points only partly see the sky-dome, which causes a penumbra effect.

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). Actually, one of the main difficulties of our problem is to consider unknown lighting.

[2] In order to compare comparable things, we scale the estimated albedos in each part, so that its median is equal to the associated ground truth value.

Table 1. Variances of the estimated albedos inside each of the four homogeneous parts of the colored 3D-model used for the tests of Fig. 3, computed after renormalization, in the three channels. In each box: the real value of the albedo is given on the left; on the right, the variances computed from our shading-reflectance decomposition, and from the cartoon-texture decomposition are given, respectively, above and below.

Channel	Hair		Face		Shirt		Plinth	
Red	1.0000	0.0135	1.0000	0.0015	0.0196	0.0002	0.1216	0.0004
		0.0274		0.0106		0.0006		0.0005
Green	0.0314	0.0007	0.5333	0.0016	0.0549	0.0000	0.1216	0.0003
		0.0007		0.0065		0.0001		0.0005
Blue	0.0000	0.0000	0.3608	0.0006	1.0000	0.0104	0.1216	0.0001
		0.0000		0.0031		0.0217		0.0003

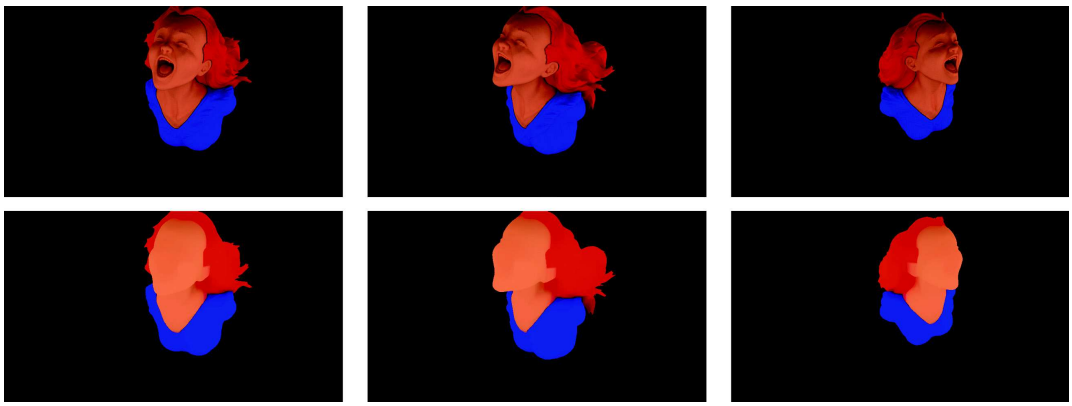

Fig. 4. Same test as in Fig. 3, but the scene is illuminated by four extended light sources. Obviously, the results are not much affected by the light configuration, since the shading-reflectance decomposition is still effective. (Color figure online)

Another test, where the sky-dome lighting is replaced with four extended light sources, is thus appropriate. Figure 4 shows that, under the same assumptions as for Fig. 3, the results are really close. The proposed method seems little sensitive to the lighting configuration, which is a significant advantage.

As we use a TV-smoothing term, which favors piecewise-constant albedos, the satisfactory results of Figs. 3 and 4 were predictable. Let us now modify the shirt albedo in order to simulate thin stripes. Figure 5 shows that the proposed method still works well if the smoothing weight λ is well-tuned (λ is 12 times smaller for Fig. 5 than for Figs. 3 and 4).

The use of an l^1-term for photometric consistency offers a competitive advantage to our method: it is robust to outliers such as deviations from the Lambertian model (1), which are unavoidable in practice. We generated a new set of $n = 13$ pictures, considering now the hair and plinth reflectances as partly specular. Indeed, the results presented in Fig. 6 are similar to those of Fig. 5.

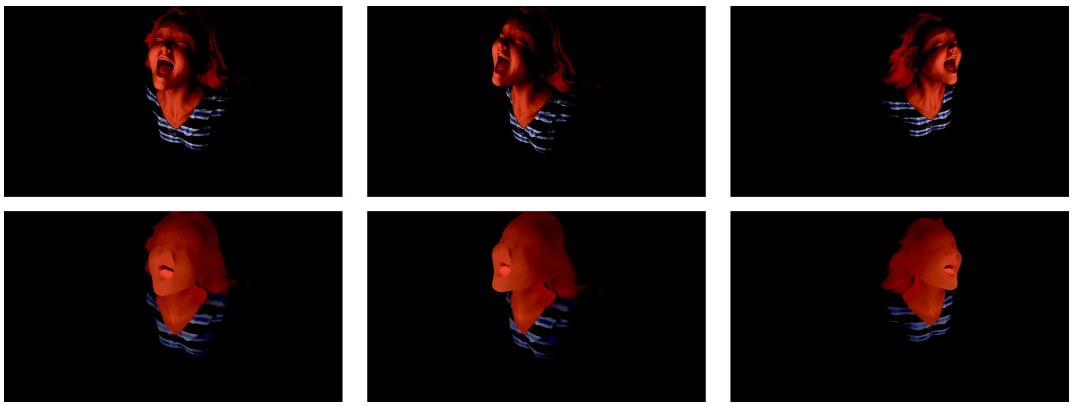

Fig. 5. Same tests as in Fig. 4, with a single extended light source, but the shirt has now a non-uniform albedo. Our method still works well, although the new configuration of the shirt albedo with fine stripes is a priori less adapted to a TV-smoothing term. (Color figure online)

Fig. 6. Same test as in Fig. 5, but the hair and plinth reflectances are now partly specular. Our method seems to be robust against outliers, due to the l^1-data term. (Color figure online)

For the next step, we suppose moreover that the scene geometry is inaccurately known. This will necessarily be the case with real data. The surface shown in Fig. 2-b (zoomed in Fig. 2-d) is obtained by smoothing the original 3D-shape of Fig. 2-a (zoomed in Fig. 2-c), using a tool from the `meshlab` software. The results provided in Fig. 7 show that our method is robust as well to small inaccuracies in the object geometry, and is thus relevant for the intended application.

As a digest of all these tests, Table 2 gives the variance of the albedo estimated inside the hair part in the red channel, which is the most significant.

Finally, we put this work in real context. The proposed algorithm is applied to the outputs of an SfM/MVS pipeline, which provides a rough geometry and camera parameters estimates. Figure 8 confirms that small inaccuracies in the geometry input do not degrade significantly the results of our method.

Fig. 7. Same test as in Fig. 6, using a coarse version of the 3D-shape (cf. Fig. 2-b and d). Our method is robust as well to small inaccuracies in the object geometry. (Color figure online)

Table 2. Variance of the estimated albedo inside the hair part in the red channel, for all tests on the synthetic dataset, except that of Fig. 5 (which would give the same value as that of Fig. 4).

Channel	Figure 3	Figure 4	Figure 6	Figure 7
Red	0.0135	0.0367	0.0475	0.0626

Fig. 8. Left: One real view of the object 'fountain-P11' [13]. Right: Colored estimated albedo, using the proposed approach. Geometry and camera parameters estimates are outputs of an SfM/MVS pipeline using 25 input images. Only 8 images have been used as inputs of our algorithm. (Color figure online)

6 Conclusion and Perspectives

We have proposed a variational framework for separating shading and reflectance from images based on an initial 3D-reconstruction obtained by SfM and MVS techniques. We have shown that the ambiguities can be raised by introducing a multi-view consistency prior on the reflectance. Robustness is further enforced by considering an l^1-norm-based photometric data term, and a piecewise-smoothness constraint on the albedo is introduced under the form of total

variation regularization. Preliminary results on a synthetic dataset covered with two different albedos, and lit in various manners, as well as on a real dataset, demonstrate the potential of the approach.

We now plan to estimate not only the reflectance, but also fine-scale geometric details. In this view, our alternating scheme will be modified in order to include an additional step aiming at estimating the normals, in the spirit of the recent multi-view shape-from-shading approach presented in [9].

References

1. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building Rome in a day. In: Proceedings of ICCV (2009)
2. Aujol, J.F., Gilboa, G., Chan, T., Osher, S.: Structure-texture image decomposition - modeling, algorithms, and parameter selection. Int. J. Comput. Vis. **67**(1), 111–136 (2006)
3. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. Int. J. Comput. Vis. **72**(3), 239–257 (2007)
4. 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)
5. Jin, H., Cremers, D., Wang, D., Yezzi, A., Prados, E., Soatto, S.: 3-D reconstruction of shaded objects from multiple images under unknown illumination. Int. J. Comput. Vis. **76**(3), 245–256 (2008)
6. Kim, K., Torii, A., Okutomi, M.: Multi-view inverse rendering under arbitrary illumination and albedo. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 750–767. Springer, Cham (2016). doi:10.1007/978-3-319-46487-9_46
7. Langguth, F., Sunkavalli, K., Hadap, S., Goesele, M.: Shading-aware multi-view stereo. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 469–485. Springer, Cham (2016). doi:10.1007/978-3-319-46487-9_29
8. Le Guen, V.: Cartoon + Texture image decomposition by the TV-L1 model. Image Process. On Line **4**, 204–219 (2014). https://doi.org/10.5201/ipol.2014.103
9. 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 BMVC (2016)
10. Moulon, P., Monasse, P., Marlet, R.: openMVG: an open multiple view geometry library. https://github.com/openMVG/openMVG
11. Ramamoorthi, R., Hanrahan, P.: An efficient representation for irradiance environment maps. In: Proceedings of SIGGRAPH (2001)
12. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proceedings of CVPR (2006)
13. Strecha, C., Von Hansen, W., Van Gool, L.J., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: Proceedings of CVPR (2008)
14. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Opt. Eng. **19**(1), 139–144 (1980)