

# Estimation de la réflectance à partir de données multi-vues

Jean MÉLOU<sup>1,3,\*</sup> Yvain QUÉAU<sup>2</sup> Jean-Denis DUROU<sup>1</sup> Fabien CASTAN<sup>3</sup> Daniel CREMERS<sup>2</sup>

<sup>1</sup> IRIT, UMR CNRS 5505, Université de Toulouse

<sup>2</sup> TUM, Munich, Allemagne

<sup>3</sup> Mikros Image, Levallois-Perret

\* jeme@mikrosimage.eu

## Résumé

Nous présentons une méthode variationnelle permettant de séparer l'ombrage de la réflectance, à partir d'un ensemble d'images d'une même scène acquises sous différents points de vue, sous l'hypothèse que la géométrie de la scène a été préalablement estimée par une méthode de reconstruction 3D multi-vues. Le modèle proposé est de type  $l^1$ -TV : un terme d'attache aux données en norme  $l^1$  est couplé à un terme de cohérence inter-vues, qui permet de lever certaines ambiguïtés, tandis que la variation totale force la réflectance à être uniforme par morceaux. Pour résoudre ce problème variationnel, nous utilisons une stratégie d'optimisation alternée de problèmes en moindres carrés repondérés. Le potentiel de notre approche est validé par plusieurs résultats sur données de synthèse, où nous testons différents cas de figure pour l'éclairage et la réflectance, puis sur données réelles.

## Mots Clef

Vision 3D, réflectance, multi-vues, ombrage, méthodes variationnelles.

## 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 solving 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

3D-vision, Reflectance, Multi-view, Shading, Variational Methods.

## 1 Introduction

L'acquisition du relief et de la réflectance d'un décor est devenue indispensable à l'industrie du cinéma. Cela est nécessaire, en particulier, pour simuler correctement les rééclairages. Si plusieurs techniques bien rodées permettent d'obtenir une estimation précise du relief, qui plus est suffisamment robuste, il n'en va pas de même de l'estimation de la réflectance. Faute de mieux, les images originales sont généralement offertes de texture, c'est-à-dire qu'elles sont plaquées telles quelles sur le relief. Or, comme ces images combinent l'éclairage, la géométrie de la scène et sa réflectance, les rééclairages produits sont alors, dans l'ensemble, très peu satisfaisants. Pour améliorer la procédure, il est indispensable d'estimer la réflectance de façon plus précise. Comme le montre la figure 1, notre but est de séparer l'ombrage (qui dépend de l'éclairage et de la géométrie de la scène) de la réflectance, en utilisant plusieurs images de la surface acquises sous différents angles, sous l'hypothèse qu'une estimation préalable du relief (pas forcément très précise) est disponible.

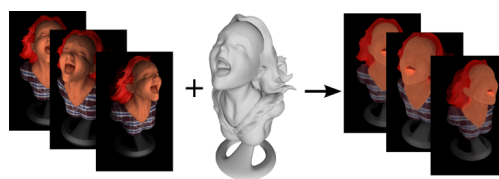


FIGURE 1 – À partir d'un ensemble d'images acquises sous différents angles, et d'une estimation grossière du relief de la scène obtenue par reconstruction 3D multi-vues, nous estimons la réflectance dans chaque vue.

En supposant la surface lambertienne, ce qui signifie que sa réflectance est caractérisée par l'albédo, nous formulons ce problème sous forme variationnelle. Nous utilisons un terme d'attache aux données en norme  $l^1$ , afin de rendre le modèle robuste aux spécularités et aux alignements imparfaits, ainsi que deux autres termes concernant l'albédo : un terme qui le force à être indépendant du point de vue, et un terme de régularisation de type « variation totale », qui vise à le rendre uniforme par morceaux.

Après avoir présenté plusieurs approches voisines de la nôtre dans le paragraphe 2, nous proposons un nouveau modèle variationnel dans le paragraphe 3. Notre stratégie de résolution itérative, présentée dans le paragraphe 4, estime alternativement la réflectance et l'éclairage. Les difficultés liées à l'utilisation de la norme  $l^1$  sont surmontées en utilisant des moindres carrés repondérés. Les tests décrits dans le paragraphe 5, effectués sur données de synthèse et sur données réelles, confirment l'intérêt de notre approche. Enfin, notre travail est résumé dans le paragraphe 6.

## 2 État de l'art

Le problème de la décomposition d'une image en basses fréquences (composante qui caractérise la réflectance) et hautes fréquences a suscité de nombreuses approches. L'un des exemples les plus connus est celui de la décomposition « *cartoon-texture* » [2], qui peut être mise en œuvre efficacement par une approche  $l^1$ -TV [9]. Cependant, ces approches peuvent échouer en présence de légères variations de luminance dues à l'ombrage (*shading*).

L'utilisation d'un modèle photométrique permet d'explicitier les interactions entre l'éclairage (potentiellement complexe), la géométrie de la scène et sa réflectance. Il est impossible d'estimer tous ces paramètres à partir d'une seule image. L'estimation du relief à partir de l'ombrage, en supposant l'éclairage et la réflectance connus, est un problème mal posé appelé *shape-from-shading* [4], qui devient bien posé si l'on utilise plusieurs images prises sous le même angle, mais sous différents éclairages. Cette variante, connue sous le nom de *stéréophotométrie* [15], permet effectivement d'estimer simultanément le relief, la réflectance et l'éclairage [3].

Toutefois, la mise en œuvre de la stéréophotométrie nécessite des conditions opératoires très contrôlées, qui rendent son champ d'application très limité. Les méthodes de stéréo multi-vues (MVS) [13], qui visent à estimer uniquement le relief, sont moins restrictives. Elles ont de plus connu des progrès considérables ces dernières années, grâce à l'essor des méthodes de *structure-from-motion* (SfM) [11]. Elles sont en outre complémentaires, puisque la MVS nécessite de connaître les paramètres des caméras, qui peuvent être estimés par SfM. De nos jours, ces méthodes ont mûri et sont communément utilisées dans des environnements peu contrôlés, voire sur de grands jeux de données récupérées sur Internet [1].

Considérant les forces et faiblesses des méthodes de reconstruction 3D photométrique et des méthodes multi-vues, une approche mixte semble pouvoir fournir simultanément une reconstruction 3D précise de la scène et des informations sur sa réflectance et sur l'éclairage. Par exemple, la méthode proposée dans [5] utilise l'information d'ombrage pour affiner itérativement le relief, en supposant l'albédo uniforme, au moins par morceaux. Kim *et al.* [6] et Langguth *et al.* [8] se sont récemment affranchis de cette hypothèse en proposant de nouveaux termes d'a priori sur la géométrie, l'ombrage et la réflectance.

Cependant, ces méthodes visent à affiner, d'une part, la réflectance, d'autre part, le relief décrit par un maillage, auquel sont ajoutés des détails de faible échelle. Or, le maillage final peut devenir tellement complexe qu'il sera difficile pour un graphiste de le manipuler. Il nous paraît plus avisé de limiter la complexité de la géométrie et de fournir au graphiste une série de données 2D caractérisant la réflectance et la géométrie, sous la forme de « cartes » d'albédo et de « cartes » de profondeur. Un premier pas dans cette direction a été effectué dans [10], où une approche variationnelle permet de retrouver conjointement la profondeur, la réflectance et l'éclairage. Toutefois, cette approche choisit une vue de référence, et n'utilise les autres images que pour la mise en correspondance.

Nous préconisons au contraire de traiter les images de façon uniforme. L'ensemble des cartes de réflectance, qui correspondent aux différentes vues, sont évaluées simultanément, ce qui nous évite de devoir sélectionner une image de référence. Contrairement aux travaux précédents, nous nous contentons de séparer l'ombrage de la réflectance en utilisant, sans l'affiner, le relief obtenu par MVS. En effet, nous verrons que l'estimation de la réflectance reste satisfaisante, même à partir d'un relief imprécis. Dans le prochain paragraphe, nous introduisons un modèle variationnel  $l^1$ -TV simple et efficace, qui permet d'estimer un ensemble de cartes de réflectance lorsque le relief de la scène est connu.

## 3 Estimation simultanée de l'albédo et de l'éclairage

Soit  $\{I^i : \Omega^i \subset \mathbb{R}^2 \rightarrow \mathbb{R}\}_{i=1\dots n}$  un ensemble de  $n$  images d'un objet vu sous différents angles (et, éventuellement, sous différents éclairages), où  $\Omega^i$  désigne le masque de l'objet dans l'image  $I^i$ . Nous supposons ces masques connus, tout comme le relief de l'objet (sous une forme éventuellement imprécise), décrit par  $n$  cartes de normales  $\mathbf{n}^i : \Omega^i \rightarrow \mathbb{R}$  préalablement estimées par l'utilisation conjointe de méthodes de SfM et de MVS.

Nous cherchons à extraire de chaque image  $I^i$  une représentation de la réflectance de la surface sous la forme d'une carte d'albédo  $\rho^i : \Omega^i \rightarrow \mathbb{R}$ , ainsi qu'une estimation de l'éclairage. Nous supposons la surface lambertienne. Sans hypothèse particulière sur l'éclairage, nous pouvons utiliser des harmoniques sphériques [12] pour le modéliser. Avec des harmoniques sphériques d'ordre 2, le  $i^{\text{ème}}$  éclairage est décrit par un vecteur  $\vec{\sigma}^i \in \mathbb{R}^9$ . Notre problème revient donc à résoudre le système d'équations suivant<sup>1</sup> :

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

où les coordonnées du vecteur  $\vec{\nu}^i(p) \in \mathbb{R}^9$  se déduisent de celles de la normale  $\mathbf{n}^i(p)$  [3].

1. Ce modèle concerne uniquement les images en niveaux de gris. Il suffit d'appliquer notre méthode à chaque canal pour pouvoir traiter les images en couleur.

Bien entendu, il est impossible de résoudre le système d'équations (1) sans y ajouter d'a priori. L'ambiguïté *globale* (c'est-à-dire commune à tous les pixels de toutes les images) sur l'échelle de l'albédo n'est pas gênante en soi, puisque les albédos peuvent être normalisés in fine, mais chacune des équations (1) comporte une ambiguïté de ce type, ce qui est plus gênant. Plutôt que de lever l'ambiguïté de façon arbitraire, par exemple en imposant aux vecteurs d'éclairage d'être unitaires, nous préférons forcer l'albédo de chaque point de la surface à être indépendant de la vue, ce qui se justifie puisque l'albédo est une caractéristique physique de la surface. En outre, afin de garantir la cohérence spatiale de l'albédo, un autre a priori est utilisé, de type « variation totale ». Enfin, pour rendre la méthode robuste aux spécularités et aux alignements imparfaits, nous résolvons le système d'équations (1) en norme  $l^1$ . Tout cela nous amène à estimer les vecteurs  $\{\vec{\sigma}^i \in \mathbb{R}^9\}_{i=1\dots n}$  et les fonctions  $\{\rho^i : \Omega^i \rightarrow \mathbb{R}\}_{i=1\dots n}$  qui minimisent l'énergie :

$$\begin{aligned} \varepsilon(\{\vec{\sigma}^i, \rho^i\}_i) &= \sum_{i=1}^n \varepsilon_{\text{Photo}}(\vec{\sigma}^i, \rho^i) \\ &+ \lambda \sum_{i=1}^n \varepsilon_{\text{Lissage}}(\rho^i) \\ &+ \mu \sum_{1 \leq i < j \leq n} \varepsilon_{\text{MV}}(\rho^i, \rho^j). \end{aligned} \quad (2)$$

Dans cette expression de  $\varepsilon(\{\vec{\sigma}^i, \rho^i\}_i)$ , le premier terme impose la cohérence photométrique avec les images :

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

le second terme force l'albédo à être lisse :

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

où  $\nabla \rho^i(p) = [\partial_x \rho^i(p), \partial_y \rho^i(p)]^\top$  représente le gradient de  $\rho^i$  au pixel  $p$  (nous utilisons des différences finies avant du premier ordre), et le dernier terme vise à garantir la cohérence inter-vues de l'albédo :

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

où  $C_{i,j}$  est une « fonction de correspondance » définie comme suit :

$$C_{i,j}(p^i, p^j) = \begin{cases} 1 & \text{si } p^i \text{ et } p^j \text{ correspondent à un} \\ & \text{même point de la surface,} \\ 0 & \text{sinon.} \end{cases} \quad (6)$$

Enfin,  $\lambda$  et  $\mu$  sont deux hyper-paramètres du modèle, qui permettent de moduler, respectivement, le lissage de l'albédo et son niveau de cohérence inter-images.

Une méthode de SfM comporte une étape de mise en correspondance entre images utilisant des descripteurs, par

exemple de type SIFT. Les valeurs de la fonction  $C_{i,j}$  définie en (6) peuvent donc être calculées à partir des résultats de ce traitement préalable.

Elles peuvent également se déduire de l'estimation initiale du relief obtenue par SfM/MVS, comme l'ont proposé Langguth *et al.* dans [8] pour leur évaluation de l'erreur géométrique. En effet, si une méthode de SfM est appliquée aux images  $\{I^i\}_{i=1\dots n}$ , nous pouvons connaître les paramètres intrinsèques  $\mathbf{K}^i \in \mathbb{R}^{3 \times 3}$  de chaque caméra, ainsi que sa pose, qui est caractérisée par une matrice de rotation  $\mathbf{R}^i \in \mathbb{R}^{3 \times 3}$  et un vecteur de translation  $\vec{t}^i \in \mathbb{R}^3$ . Un point  $X \in \mathbb{R}^3$  de la surface, visible dans la  $i^{\text{ème}}$  image, est donc « conjugué » du pixel  $p^i \in \mathbb{R}^2$  par la fonction de projection  $\pi_i$  définie par :

$$\begin{bmatrix} p^i \\ 1 \end{bmatrix} = \pi_i(X) = \mathbf{K}^i(\mathbf{R}^i X + \vec{t}^i). \quad (7)$$

Par conséquent, en introduisant un seuil  $\epsilon > 0$  ( $\epsilon = 3$ , dans nos tests) et en utilisant le relief obtenu par MVS pour inverser les projections  $\pi_i$ , la fonction de correspondance  $C_{i,j}$  définie en (6) peut être calculée de la façon suivante :

$$C_{i,j}(p^i, p^j) = \begin{cases} 1 & \text{si } \|\pi_i^{-1}([p^i, 1]^\top) - \pi_j^{-1}([p^j, 1]^\top)\| \leq \epsilon, \\ 0 & \text{sinon.} \end{cases} \quad (8)$$

## 4 Résolution numérique

Nous présentons dans ce paragraphe notre méthode de minimisation de l'énergie (2). Le problème étant bi-convexe, nous optons pour une stratégie d'optimisation alternée : à l'itération ( $k$ ), nous mettons successivement à jour l'éclairage et l'albédo de la façon suivante :

$$\{\vec{\sigma}^{i,(k+1)}\}_i = \underset{\{\vec{\sigma}^i \in \mathbb{R}^9\}_i}{\text{argmin}} \varepsilon(\{\vec{\sigma}^i, \rho^{i,(k)}\}_i), \quad (9)$$

$$\{\rho^{i,(k+1)}\}_i = \underset{\{\rho^i : \Omega^i \rightarrow \mathbb{R}\}_i}{\text{argmin}} \varepsilon(\{\vec{\sigma}^{i,(k+1)}, \rho^i\}_i), \quad (10)$$

en utilisant les initialisations  $\rho^{i,(0)} \equiv I^i$  et  $\vec{\sigma}^{i,(0)} \equiv \vec{1}_{\mathbb{R}^9}$ . Afin de pallier la non différentiabilité des termes d'énergie en norme  $l^1$ , nous appliquons une approche de type moindres carrés repondérés. Puisque les  $n$  vecteurs d'éclairage  $\vec{\sigma}^i$  sont indépendants, nous remplaçons le problème (9) par  $n$  problèmes en moindres carrés repondérés indépendants, à savoir, pour  $i = 1 \dots n$  :

$$\vec{\sigma}^{i,(k+1)} = \underset{\vec{\sigma}^i \in \mathbb{R}^9}{\text{argmin}} \sum_{p \in \Omega^i} w_i^{(k)}(p) (I^i(p) - \rho^{i,(k)}(p) \vec{\sigma}^i \cdot \vec{\nu}^i(p))^2, \quad (11)$$

où le poids  $w_i^{(k)}(p)$  s'écrit :

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

et où la notation  $|\cdot|_\delta = \max\{\delta, |\cdot|\}$  est utilisée ( $\delta = 10^{-4}$ , dans nos tests). Chaque problème (11) peut être résolu en calculant une matrice pseudo-inverse.

Par un raisonnement analogue, l'estimation de l'albédo est mise à jour de la façon suivante :

$$\begin{aligned} \{\rho^{i,(k+1)}\}_i = & \operatorname{argmin}_{\{\rho^i: \Omega^i \rightarrow \mathbb{R}\}_i} \sum_{i=1}^n \sum_{p \in \Omega^i} \varepsilon_{\text{Photo}}^W(\vec{\sigma}^{i,(k+1)}, \rho^i, p) \\ & + \lambda \sum_{i=1}^n \sum_{p \in \Omega^i} \varepsilon_{\text{Lissage}}^W(\rho^i, p) \\ & + \mu \sum_{1 \leq i < j \leq n} \sum_{p^i \in \Omega^i} \sum_{p^j \in \Omega^j} \varepsilon_{\text{MV}}^W(\rho^i, \rho^j, p^i, p^j), \end{aligned} \quad (13)$$

où nous introduisons les notations suivantes :

$$\begin{aligned} \varepsilon_{\text{Photo}}^W(p, \rho^i, \vec{\sigma}^{i,(k+1)}) = & w_i^{(k)}(p) \\ & (I^i(p) - \rho^i(p) \vec{\sigma}^{i,(k+1)} \cdot \vec{v}^i(p))^2, \end{aligned} \quad (14)$$

$$\begin{aligned} \varepsilon_{\text{Lissage}}^W(\rho^i, p) = & w_{\partial_x \rho^i}^{(k)}(p) (\partial_x \rho^i(p))^2 \\ & + w_{\partial_y \rho^i}^{(k)}(p) (\partial_y \rho^i(p))^2, \end{aligned} \quad (15)$$

$$\begin{aligned} \varepsilon_{\text{MV}}^W(\rho^i, \rho^j, p^i, p^j) = & C_{i,j}(p^i, p^j) \\ & w_{i,j}^{(k)}(p^i, p^j) (\rho^i(p^i) - \rho^j(p^j))^2, \end{aligned} \quad (16)$$

qui font intervenir les poids :

$$w_{\partial_x \rho^i}^{(k)}(p) = \frac{1}{|\partial_x \rho^{i,(k)}(p)|_\delta}, \quad (17)$$

$$w_{\partial_y \rho^i}^{(k)}(p) = \frac{1}{|\partial_y \rho^{i,(k)}(p)|_\delta}, \quad (18)$$

$$w_{i,j}^{(k)}(p^i, p^j) = \frac{1}{|\rho^{i,(k)}(p^i) - \rho^{j,(k)}(p^j)|_\delta}. \quad (19)$$

Grâce à l'a priori sur la cohérence inter-vues de l'albédo, les cartes d'albédo correspondant aux différentes images ne sont pas estimées indépendamment mais, au contraire, simultanément. En concaténant toutes les valeurs de l'albédo dans un vecteur  $\vec{\rho} \in \mathbb{R}^N$ , où  $N = \sum_i |\Omega^i|$ , le problème d'optimisation (13) peut être reformulé sous la forme d'un problème en moindres carrés linéaires de grande dimension :

$$\begin{aligned} \vec{\rho}^{(k+1)} = & \operatorname{argmin}_{\vec{\rho} \in \mathbb{R}^N} \left\| \begin{bmatrix} \operatorname{Diag}(\{\sqrt{w_i^{(k)}}(p) \vec{\sigma}^{i,(k+1)} \cdot \vec{v}^i(p)\}_{i,p}) \\ \sqrt{\lambda} \operatorname{Diag}(\{\sqrt{w_{\partial_x \rho^i}^{(k)}}(p)\}_{i,p}) \mathbf{D}_x \\ \sqrt{\lambda} \operatorname{Diag}(\{\sqrt{w_{\partial_y \rho^i}^{(k)}}(p)\}_{i,p}) \mathbf{D}_y \\ \sqrt{\mu} \operatorname{Diag}(\{\sqrt{w_{i,j}^{(k)}}(p^i, p^j)\}_{i,j,p^i,p^j}) \mathbf{C} \end{bmatrix} \vec{\rho} \right. \\ & \left. - \begin{bmatrix} \operatorname{Diag}(\{\sqrt{w_i^{(k)}}(p)\}_{i,p}) \vec{I} \\ \vec{0}_{N \times 1} \\ \vec{0}_{N \times 1} \\ \vec{0}_{\sum_{i=1}^{n-1} (n-i) |\Omega^i| \times 1} \end{bmatrix} \right\|_2^2. \end{aligned} \quad (20)$$

Dans le problème en moindres carrés (20), le premier bloc matriciel contient les valeurs (pondérées) de l'ombrage, les matrices creuses  $\mathbf{D}_x$  et  $\mathbf{D}_y$ , qui concatènent les différences finies relatives aux  $n$  domaines  $\Omega^i$ ,  $i = 1 \dots n$ , tandis que la matrice  $\mathbf{C}$ , de taille  $(\sum_{i=1}^{n-1} (n-i) |\Omega^i|) \times N$ , est construite à partir des fonctions de correspondance  $C_{i,j}$  définies en (6). La partie non nulle du vecteur situé sur la deuxième ligne de (20) pondère les niveaux de gris concaténés dans le vecteur  $\vec{I}$ . Chaque fonction  $C_{i,j}$  est facile à représenter par une matrice de taille  $|\Omega^i| \times |\Omega^j|$ , notée  $\mathbf{C}_{i,j}$ , dont un élément par ligne au plus est non nul. En arrangeant ces matrices par blocs dans la matrice  $\mathbf{C}$ , l'ensemble des expressions  $C_{i,j}(p^i, p^j)$  ( $\rho^i(p^i) - \rho^j(p^j)$ ) peut être représenté par le produit matriciel  $\mathbf{C} \vec{\rho}$ . Par exemple, si nous disposons de  $n = 4$  images, la matrice  $\mathbf{C}$  s'écrit :

$$\mathbf{C} = \begin{bmatrix} \mathbf{B}_{1,2} & \mathbf{C}_{1,2} & & \\ \mathbf{B}_{1,3} & & \mathbf{C}_{1,3} & \\ \mathbf{B}_{1,4} & & & \mathbf{C}_{1,4} \\ & \mathbf{B}_{2,3} & \mathbf{C}_{2,3} & \\ & \mathbf{B}_{2,4} & & \mathbf{C}_{2,4} \\ & & \mathbf{B}_{3,4} & \mathbf{C}_{3,4} \end{bmatrix}, \quad (21)$$

où  $\mathbf{B}_{i,j}$  est une matrice carrée de taille  $|\Omega^i| \times |\Omega^i|$ , diagonale, dont les éléments diagonaux valent  $-1$  sur les lignes où la matrice  $\mathbf{C}_{i,j}$  n'est pas identiquement nulle, et 0 sinon.

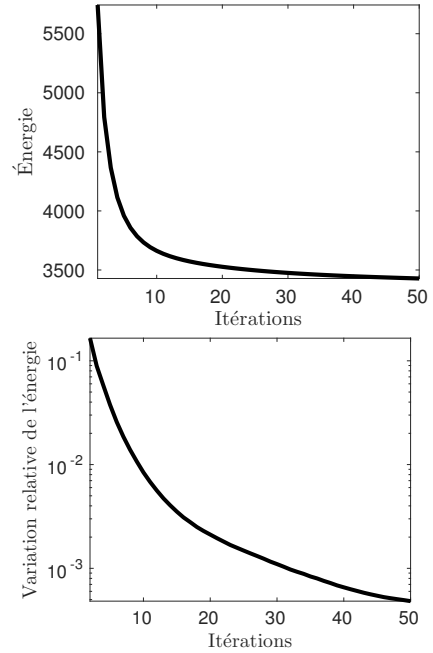


FIGURE 2 – En haut : évolution de l'énergie définie en (2), en fonction des itérations, sur les données de la figure 8. En bas : valeur absolue de la variation relative entre deux valeurs successives de l'énergie. Notre algorithme s'arrête lorsque cette dernière passe en-dessous de  $10^{-3}$ , ce qui se produit en moins de 50 itérations et prend 10 minutes sur un processeur I7 récent, avec un code Matlab non optimisé, pour des images de taille  $540 \times 960$ .

Le problème (20) est un problème en moindres carrés linéaires, de matrice très creuse. Pour le résoudre, nous appliquons un algorithme de gradient conjugué aux équations normales associées. Nous itérons l'ensemble de ces étapes de résolution jusqu'à atteindre la convergence, ou jusqu'à ce qu'un nombre maximal d'itérations soit atteint. Il ressort de nos tests (cf. figure 2) que 50 itérations suffisent à atteindre une solution stable (définie par une variation relative de  $10^{-3}$  entre deux valeurs successives de l'énergie définie en (2)). La preuve de convergence de cet algorithme dépasse l'objectif de cet article, mais la convergence a été observée dans tous les tests, bien que sa vitesse semble être sous-linéaire. Notons enfin que d'autres approches auraient pu être envisagées, comme par exemple les algorithmes primaux-duaux.

## 5 Résultats

Commençons par tester notre méthode de décomposition ombrage-réflectance dans un cas très simple. L'objet de la figure 3-a, supposé parfaitement lambertien, est vu par  $n = 13$  caméras virtuelles de paramètres internes et de poses connus, sous un éclairage de type « dôme éclairant » (*sky-dome*). Nous supposons dans un premier temps que le relief de cet objet est parfaitement connu. Les inconnues du problème sont donc la réflectance en chaque point de la surface, et l'éclairage. La première ligne de la figure 4 montre deux de ces images, qui ont été simulées par lancer de rayons à l'aide d'un moteur de rendu. La ligne suivante montre l'albédo coloré estimé par notre méthode.

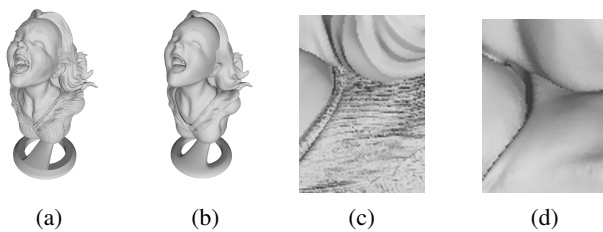


FIGURE 3 – (a) Le modèle 3D bien connu *Joyful Yell*, que nous utilisons pour les tests, sera recouvert de deux albédos différents. (b) Ce même modèle lissé, donc moins précis. (c)-(d) Zooms de (a) et (b) sur une zone proche du cou.

Vu que le type d'éclairage utilisé est celui qui se prête le mieux à une modélisation par harmoniques sphériques, nous pouvons effectivement nous attendre à ce que ce premier résultat soit très satisfaisant. À titre de comparaison, la troisième ligne de la figure 4 montre le résultat obtenu par décomposition *cartoon-texture* [9], qui ne nécessite l'utilisation que d'une seule image : la composante *cartoon*, qui est un équivalent plus ou moins empirique de la réflectance, est manifestement bien moins uniforme que l'albédo estimé par décomposition ombrage-réflectance. Cela valide a posteriori l'idée fondatrice de notre travail, à savoir que l'estimation de la réflectance bénéficie doublement de l'uti-

lisation de plusieurs vues : cela permet non seulement de connaître le relief, mais également de contraindre l'albédo, en chaque point de la surface, à être le même dans toutes les images où ce point est visible. A contrario, la décomposition *cartoon-texture* ne peut pas réellement corriger l'ombrage, ce qui explique, par exemple, que l'« albédo » estimé à l'intérieur de la bouche soit très sombre.

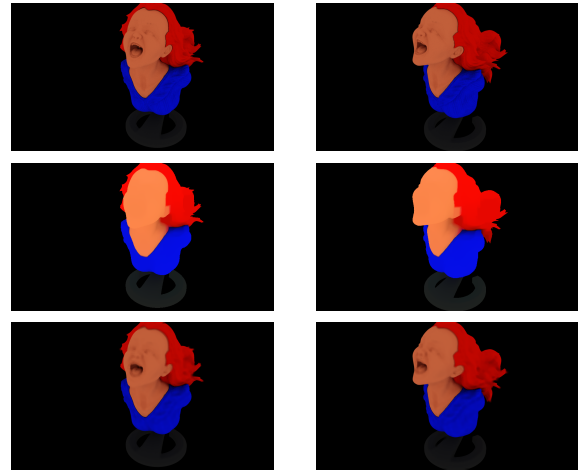


FIGURE 4 – Première ligne : deux vues de synthèse (parmi  $n = 13$ ) de l'objet de la figure 3-a, de réflectance purement lambertienne divisée en quatre parties d'albédo uniforme (cheveux, visage, chemise et socle), placé sous un éclairage de type *sky-dome*. Deuxième ligne : albédos colorés estimés par la méthode proposée. Troisième ligne : estimation empirique de l'albédo par la méthode de décomposition *cartoon-texture* décrite dans [9].

Comme nous disposons des valeurs réelles de l'albédo, nous pouvons évaluer numériquement ces résultats, en calculant la variance de l'albédo estimé pour l'ensemble des pixels des  $n = 13$  images correspondant aux cheveux, au visage, à la chemise ou au socle. Les variances ainsi calculées devraient être nulles puisque, lors de la simulation, chacune de ces parties a reçu un albédo uniforme, y compris à l'intérieur de la bouche dans le cas du visage<sup>2</sup>. Les valeurs compilées dans la table 1 confirment que notre estimation est plus précise que par décomposition *cartoon-texture*. Nous observons également que la variance de l'albédo est plus élevée pour les deux parties comportant des zones concaves, à savoir les cheveux et le visage (dans ce dernier cas, l'albédo est nettement sous-estimé à l'arrière du cou). En effet, comme les zones concaves ne voient qu'une partie du dôme éclairant, cela se traduit par un phénomène d'assombrissement de type « pénombre ».

Puisque nous connaissons le relief, nous pourrions essayer de compenser ce phénomène de pénombre, mais il nous faudrait également connaître l'éclairage, ce qui est difficilement envisageable dans les conditions opératoires de

2. Afin de comparer des choses comparables, nous fixons l'échelle de l'albédo estimé de manière à ce que sa médiane soit égale à l'albédo réel.

TABLE 1 – Évaluation numérique des résultats de la figure 4 : variance de l’albédo estimée dans chaque partie du modèle 3D où l’albédo est uniforme (cheveux, visage, chemise et socle), calculée après renormalisation, dans chaque canal coloré. Pour chaque partie et chaque canal : la valeur réelle de l’albédo est indiquée à gauche ; les variances calculées pour la décomposition ombrage-réflectance (en haut) et pour la décomposition *cartoon-texture* (en bas) sont indiquées à droite.

Canal	Cheveux		Visage		Chemise		Socle	
Rouge	1.0000	0.0135 0.0274	1.0000	0.0015 0.0106	0.0196	0.0002 0.0006	0.1216	0.0004 0.0005
Vert	0.0314	0.0007 0.0007	0.5333	0.0016 0.0065	0.0549	0.0000 0.0001	0.1216	0.0003 0.0005
Bleu	0.0000	0.0000 0.0000	0.3608	0.0006 0.0031	1.0000	0.0104 0.0217	0.1216	0.0001 0.0003

l’application visée : en effet, les scènes peuvent être filmées en extérieur, donc sous un éclairage non contrôlé ; en outre, il nous faudrait non seulement connaître les éclairages primaires, mais également tenir compte des rebonds de la lumière sur les différentes parties de la scène (ce phénomène a été pris en compte, lors du calcul des images, par lancer de rayons). Effectivement, une des difficultés du problème qui nous intéresse vient de ce que l’éclairage est inconnu.

Il paraît donc opportun d’effectuer un nouveau test, dans lequel l’éclairage de type « dôme éclairant » est remplacé par quatre sources lumineuses étendues. La figure 5 montre que, dans des conditions opératoires similaires à celles de la figure 4, les résultats sont quasiment identiques. La méthode proposée semble donc peu sensible à l’éclairage, ce qui lui confère un avantage très important.

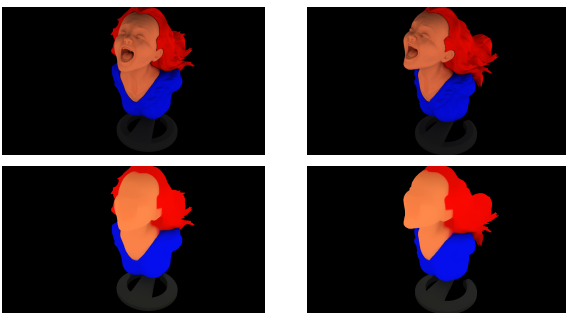


FIGURE 5 – Même test que celui de la figure 4, à ceci près que la scène est éclairée par quatre sources étendues. Les résultats ne sont que très peu affectés par cette nouvelle configuration de l’éclairage.

Comme nous utilisons un terme de lissage de type « variation totale », qui favorise les zones homogènes, il est normal que les résultats des figures 4 et 5 soient satisfaisants, puisque l’albédo y est uniforme par morceaux. Or, si nous modifions l’albédo de la chemise de manière à simuler des rayures, la figure 6 montre que la méthode proposée

donne encore satisfaction<sup>3</sup>, si le paramètre  $\lambda$  est correctement ajusté (en l’occurrence,  $\lambda$  est 12 fois plus faible pour le test de la figure 6 que pour les tests des figures 4 et 5).

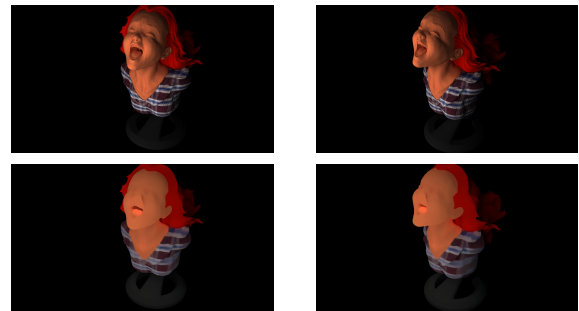


FIGURE 6 – Même test que celui de la figure 5, avec une seule source lumineuse étendue, lorsque l’albédo de la chemise n’est plus uniforme. Notre méthode fonctionne encore, bien que le nouvel albédo soit moins adapté à un terme de régularisation de type « variation totale ».

Par ailleurs, l’utilisation d’un terme d’attache aux données en norme  $l^1$  confère à notre méthode un autre avantage, certes prévisible, à savoir d’être robuste aux données aberrantes, en particulier aux écarts au modèle lambertien (1), qui sont inévitables. Nous avons donc calculé un nouveau jeu de  $n = 13$  images de l’objet de la figure 3-a, en supposant que la réflectance des cheveux et du socle comportait une partie spéculaire. Or, les résultats de la figure 7 sont très similaires à ceux de la figure 6.

Nous effectuons un nouveau test, où nous supposons que la connaissance du relief de la scène est imparfaite, ce qui sera forcément le cas avec des données réelles. Le relief de la figure 3-b a été obtenu en lissant celui de la figure 3-a (à l’aide d’un outil du logiciel *meshlab*). Ce relief comporte donc des écarts au modèle 3D réel. Les résultats de la figure 8 montrent que notre méthode est à nouveau robuste, lorsque la connaissance du relief est imprécise.

3. Les albédos étant estimés à un facteur près, cela explique que leurs apparences puissent fortement différer entre les figures 5 et 6.

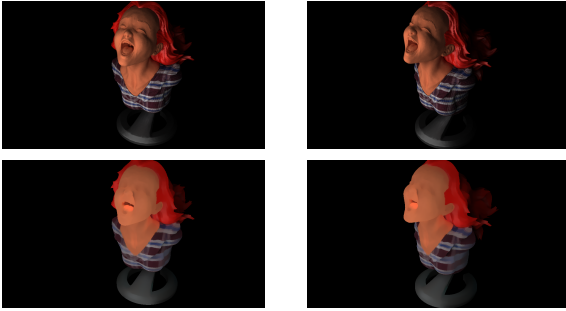


FIGURE 7 – Même test que celui de la figure 6, lorsque la réflectance des cheveux et du socle est partiellement spéculaire. Notre méthode semble robuste aux données aberrantes, grâce au terme d’attache aux données en norme  $l^1$ .

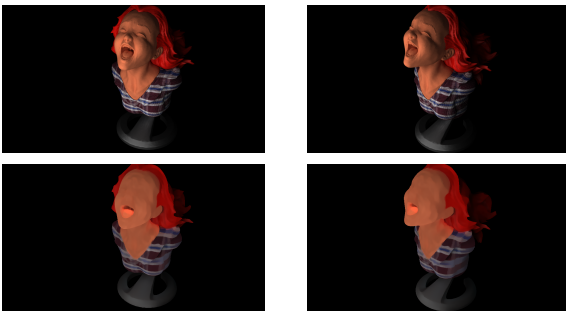


FIGURE 8 – Même test que celui de la figure 7, lorsque le relief utilisé est imprécis (cf. figures 3-b et 3-d). Notre méthode semble effectivement robuste à une certaine imprécision sur la connaissance du relief.

La table 2 synthétise l’ensemble des tests effectués sur images de synthèse, en indiquant la variance de l’albédo estimé sur les cheveux, dans le canal rouge, qui est le canal le plus significatif puisqu’il s’agit de la couleur dominante de cette partie.

Pour finir, nous replaçons ces travaux dans un contexte de prise de vues réelles. La méthode proposée est appliquée en sortie d’un algorithme de SfM/MVS, qui fournit une estimation grossière du relief, ainsi que l’estimation des paramètres des caméras. Le résultat de la figure 9 confirme que l’estimation de l’albédo fournie par notre méthode n’est pas trop durement affectée par une connaissance imparfaite de la géométrie de la scène, ce qui est inévitable dans les conditions opératoires qui ont motivé cette étude.

## 6 Conclusion et perspectives

Nous avons proposé un modèle variationnel permettant de séparer l’ombrage de la réflectance à partir d’un ensemble d’images. Ce modèle utilise une estimation initiale du relief obtenue par SfM et MVS. Nous avons montré que les ambiguïtés du problème pouvaient être éliminées en introduisant un a priori sur la cohérence inter-images de la ré-

TABLE 2 – Variance de l’albédo estimé sur les cheveux, dans le canal rouge, pour l’ensemble des tests effectués sur images de synthèse, excepté celui de la figure 6 (qui donnerait le même résultat que le test de la figure 5).

Canal	Fig. 4	Fig. 5	Fig. 7	Fig. 8
Rouge	0.0135	0.0367	0.0475	0.0626



FIGURE 9 – Première ligne : une vue réelle (parmi 25) de l’objet *fountain-P11* [14]. Deuxième ligne : albédo coloré estimé par l’approche proposée. La géométrie et les paramètres des caméras ont été estimés par SfM/MVS à partir des 25 images. Seules 8 images ont été fournies à notre algorithme. Troisième ligne : zooms de ces deux images, sur une même partie du mur de brique à l’arrière de la fontaine.

flectance. La robustesse est garantie par l’utilisation de la norme  $l^1$  pour le terme d’attache aux données. D’autre part, la réflectance est contrainte à être uniforme par morceaux, grâce à un terme de régularisation de type « variation totale ». Les premiers tests, effectués non seulement sur données de synthèse, à partir d’un modèle 3D sur lequel sont plaqués deux albédos différents, et qui est soumis à différents éclairages, mais également sur données réelles, démontrent le potentiel de notre approche.

Nous prévoyons à présent de ne pas seulement estimer la réflectance, mais également d'affiner l'estimation du relief. Notre schéma de résolution sera alors modifié, afin de lui adjoindre une étape d'estimation du relief, dans l'esprit de ce qui a été récemment proposé dans [10].

## Références

- [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. *International Journal of Computer Vision* 67(1), 111–136 (2006)
- [3] Basri, R., Jacobs, D., Kemelmacher, I. : Photometric Stereo with General, Unknown Lighting. *International Journal of Computer Vision* 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. *International Journal of Computer Vision* 76(3), 245–256 (2008)
- [6] Kim, K., Torii, A., Okutomi, M. : Multi-view Inverse Rendering Under Arbitrary Illumination and Albedo. In : Proceedings of ECCV (2016)
- [7] 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)
- [8] Langguth, F., Sunkavalli, K., Hadap, S., Goesele, M. : Shading-aware Multi-view Stereo. In : Proceedings of ECCV (2016)
- [9] Le Guen, V. : Cartoon + Texture Image Decomposition by the TV-L1 Model. *Image Processing On Line* 4, 204–219 (2014), <https://doi.org/10.5201/ipol.2014.103>
- [10] 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)
- [11] Moulon, P., Monasse, P., Marlet, R. : openMVG : An open multiple view geometry library. <https://github.com/openMVG/openMVG>
- [12] Ramamoorthi, R., Hanrahan, P. : An Efficient Representation for Irradiance Environment Maps. In : Proceedings of SIGGRAPH (2001)
- [13] 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)
- [14] 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)
- [15] Woodham, R.J. : Photometric Method for Determining Surface Orientation from Multiple Images. *Optical Engineering* 19(1), 139–144 (1980)