

Que peut-on apprendre d'une scène vue par une webcam à partir d'images prises au cours d'une journée ensoleillée ?

What can we learn from a scene seen by a webcam from images taken within a sunny day?

Yvain Quéau¹

Jean-Denis Durou¹

Xavier Descombes²

¹ Université de Toulouse, IRIT, UMR CNRS 5505

² INRIA Sophia-Antipolis Méditerranée

yvain.queau@enseeiht.fr durou@irit.fr xavier.descombes@inria.fr

Résumé

Nous étudions le problème de la reconstruction 3D à partir d'images prises par une webcam au cours d'une journée ensoleillée. Sachant que le flux lumineux provenant du soleil est parallèle et uniforme, et que sa direction varie au cours de la journée, il semble opportun d'utiliser la stéréophotométrie. Toutefois, le mouvement du soleil étant circonscrit à un plan, le problème est dégénéré. Nous étudions dans cet article les ambiguïtés apparaissant dans ce cas, et montrons qu'en introduisant une hypothèse sur la différentiabilité de la surface à reconstruire, le problème peut être reformulé comme un simple problème d'étiquetage binaire, puis résolu très efficacement par l'algorithme de coupure de graphe.

Mots Clef

Reconstruction 3D, stéréophotométrie, webcam, coupure de graphe.

Abstract

We tackle the 3D-reconstruction problem using a timelapse sequence of images captured during a sunny day by a webcam. Considering the sun as an infinitely distant punctual light source, images acquired at different times will reveal the scene as lit under different directions, allowing the use of the photometric stereo technique. Yet, since the sun moves within a plane, the 3D-reconstruction problem is under-constrained: we study in this paper the arising ambiguities, and show that by introducing a hypothesis on the differentiability of the surface, this 3D-reconstruction problem can be reformulated as a binary labelling problem, and solved by graph cuts.

Keywords

3D-reconstruction, photometric stereo, outdoor webcam, graph cuts.

1 Introduction

Les données de nombreuses webcams sont aujourd'hui disponibles à la consultation sur Internet, quasiment en temps réel. Citons par exemple le projet AMOS (*the Archive of Many Outdoor Scenes*) [?], qui totalisait, au moment où cet article a été écrit, plus de 650 millions d'images acquises par près de 25000 caméras réparties en divers endroits du globe, par exemple au centre Pompidou à Paris, sous la Statue de la Liberté à New York (cf. figure ??) ou sur les étendues gelées de l'Antarctique. Chaque image est accompagnée des coordonnées GPS de la webcam, ainsi que de la date et de l'heure de prise de vue.

Parmi les applications envisageables, il a été proposé dans [?, ?] d'utiliser ces immenses bases de données pour calculer le relief et la réflectance des scènes observées. Le présent travail reprend cette idée. Nous tentons de formaliser le problème pour en étudier les ambiguïtés et tirer de cette étude un algorithme efficace permettant de reconstruire le relief de scènes à partir d'images prises au cours d'une même journée.

Après avoir décrit dans le paragraphe ?? les travaux présentés dans [?, ?], nous analysons dans le paragraphe ?? les ambiguïtés inhérentes au problème de la stéréophotométrie dans le cas où les éclairages sont coplanaires. Nous montrons ensuite, dans le paragraphe ??, comment une hypothèse sur le caractère différentiable de la surface à reconstruire permet de lever ces ambiguïtés, et comment le problème traité se prête particulièrement bien à une résolution par l'algorithme de coupure de graphe. Des extensions à divers cas d'usage plus réalistes (intensités des éclairages inconnues et images en couleur) sont enfin présentées dans le paragraphe ??.

2 Travaux antérieurs

Les auteurs de [?, ?] utilisent les variations d'éclairage induites par la course du soleil sur de très longues périodes (plusieurs saisons, voire plusieurs années) afin de mettre en œuvre la technique dite de « stéréophotométrie » [?] sur un

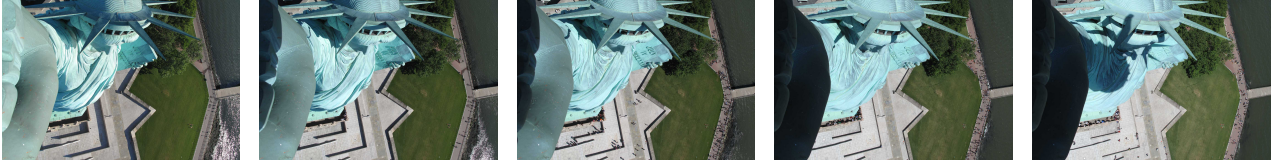


FIGURE 1 – Images de la Statue de la Liberté prises le 5 juillet 2014 entre 11h30 et 17h30 (source : projet AMOS). La course du soleil permet de fournir des images de la scène sous différentes directions incidentes. Le but de ce travail est d'utiliser ces images comme données de stéréophotométrie, afin de reconstruire un modèle 3D de la scène.

ensemble de m images I^i , $i \in [1, m]$.

À partir des coordonnées GPS de la webcam, de la date et de l'heure de prise de vue, il est possible d'estimer précisément la position du soleil relativement au référentiel terrestre [?]. La pose de la webcam peut être estimée, par exemple, en ajustant aux images les paramètres d'un modèle statistique décrivant le ciel [?]. Une fois ces étapes franchies, les directions d'éclairage, relativement à la caméra, sont connues. Nous notons $\mathbf{s}^i \in \mathbb{R}^3$ le vecteur *unitaire* décrivant la direction d'éclairage pour l'image numéro i . Bien que ces m directions d'éclairage soient parfaitement connues, cela n'est pas le cas des intensités $\phi^i \in \mathbb{R}$, qui dépendent de nombreux phénomènes météorologiques difficiles à prédire (cf. paragraphe ??).

En supposant que la clarté d'un point de la scène observée est indépendante de la position de l'observateur (hypothèse « lambertienne »), la réflectance est caractérisée par un scalaire ρ , appelé *albédo*. En négligeant les phénomènes d'ombres (propres ou portées), le niveau de gris I^i est alors donné par la *loi de Lambert* :

$$I^i = \rho \phi^i \mathbf{s}^i \cdot \mathbf{n} \quad (1)$$

où \mathbf{n} désigne la normale unitaire sortante à la surface. Une extension de ce modèle aux images couleur est proposée dans le paragraphe ??.

L'estimation des paramètres inconnus \mathbf{n} , ρ et ϕ^i est reformulée dans [?, ?] comme un problème d'optimisation aux moindres carrés pondérés :

$$\min_{\mathbf{n}, \rho, \phi^i} \sum_{i=1}^m w^i (I^i - \rho \phi^i \mathbf{s}^i \cdot \mathbf{n})^2 \quad (2)$$

où les w^i sont des poids permettant de filtrer les images « corrompues » par des conditions météorologiques non favorables, déterminés de façon heuristique en analysant par exemple la couleur du ciel. La minimisation est réalisée au moyen d'algorithmes génériques d'optimisation aux moindres carrés non linéaires.

L'idée sous-jacente à ces méthodes est que les directions d'éclairage \mathbf{s}^i présentent une variabilité suffisante pour rendre inutile l'analyse du conditionnement du problème. En pratique, un grand nombre d'images sont utilisées (plusieurs milliers), et la qualité des résultats obtenus force à reconnaître que cette approche « big data » est pertinente. Nous pensons cependant qu'elle pourrait tirer bénéfice d'une étude plus approfondie du cas qui nous intéresse.

Dans cet article, nous nous concentrons sur le cas dégénéré où la position du soleil est circonscrite à un plan. Autrement dit, quel que soit $m \geq 2$, la *matrice d'éclairage* $\mathbf{S} = [\mathbf{s}^1, \dots, \mathbf{s}^m]$ est de rang 2, ce qui rend le problème inverse (??) mal posé, même lorsque seules les normales sont inconnues. En théorie, comme le plan contenant la course du soleil varie d'un jour à l'autre, le rang de \mathbf{S} est égal à 3 dès que l'on utilise des images prises sur une période de plusieurs jours, mais la matrice \mathbf{S} est très mal conditionnée s'il s'agit d'une période trop courte. On comprend alors mieux pourquoi seuls des jeux d'images prises en différentes saisons sont utilisés dans [?, ?] : cela permet de rendre le problème (??) à la fois bien posé et bien conditionné sans introduire d'hypothèse supplémentaire sur la scène.

Cependant, il est parfois impossible d'utiliser des images prises en différentes saisons, par exemple si le relief de la scène est fixe à l'échelle d'une journée, mais est susceptible d'évoluer notablement à l'échelle de plusieurs mois. C'est le cas par exemple d'une webcam observant un glacier. Réaliser la reconstruction 3D à partir de séquences prises au cours d'une seule journée permettrait donc de relever quotidiennement le relief des glaciers, afin d'affiner les relevés stéréoscopiques [?]. Nous présentons dans cet article un premier pas vers l'utilisation de la stéréophotométrie pour résoudre de tels problèmes, en explicitant les ambiguïtés du problème et en proposant une approche numérique efficace pour réaliser la reconstruction 3D en supposant l'albédo ρ connu, ou du moins homogène, une hypothèse qui justement semble réaliste dans le cas des glaciers.

3 Mise en évidence des ambiguïtés

Nous supposons dans un premier temps que nous disposons de $m \geq 2$ images en niveaux de gris I^i , $i \in [1, m]$, acquises au cours d'une même journée, et que nous connaissons les directions d'éclairage \mathbf{s}^i , leurs intensités ϕ^i et l'albédo ρ . Nous utilisons comme modèle la loi de Lambert (??), et négligeons les ombres et les reflets spéculaires, considérés comme des données aberrantes (en pratique, nous disposons de suffisamment d'images pour garantir une certaine robustesse vis-à-vis de ces phénomènes). Même dans un tel cas, apparemment très simple, le problème est mal posé : il a été montré dans [?] que la stéréophotométrie à $m \geq 2$ éclairages coplanaires admet deux

solutions \mathbf{n}^+ et \mathbf{n}^- , éventuellement égales, en chaque point (cf. figure ??). Nous proposons ici une méthode pour les expliciter. En écrivant la loi de Lambert (??) pour chaque image I^i , $i \in [1, m]$, on obtient en chaque point du domaine de reconstruction Ω un système linéaire de m équations :

$$\underbrace{\begin{bmatrix} \frac{I^1}{\rho\phi^1} \\ \frac{I^2}{\rho\phi^2} \\ \vdots \\ \frac{I^m}{\rho\phi^m} \end{bmatrix}}_{\mathbf{i} \in \mathbb{R}^m} = \underbrace{\begin{bmatrix} \mathbf{s}^{1\top} \\ \mathbf{s}^{2\top} \\ \vdots \\ \mathbf{s}^{m\top} \end{bmatrix}}_{\mathbf{S}^\top \in \mathbb{R}^{m \times 3}} \mathbf{n} \quad (3)$$

où le vecteur inconnu $\mathbf{n} \in \mathbb{R}^3$ est unitaire. Le problème aux moindres carrés s'écrit, en chaque point :

$$\begin{cases} \min_{\mathbf{n}} \|\mathbf{i} - \mathbf{S}^\top \mathbf{n}\|^2 \\ \text{s.c. } \|\mathbf{n}\| = 1 \end{cases} \quad (4)$$

où $\|\cdot\|$ désigne la norme euclidienne. Par décomposition en valeurs singulières (SVD), la matrice \mathbf{S}^\top s'écrit :

$$\mathbf{S}^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^\top \quad (5)$$

où $\sigma_1 \geq \sigma_2 > 0$ sont les deux valeurs singulières strictement positives de \mathbf{S} , \mathbf{u}_1 et \mathbf{u}_2 sont des vecteurs de \mathbb{R}^m , et \mathbf{v}_1 et \mathbf{v}_2 des vecteurs de \mathbb{R}^3 . Le vecteur

$$\mathbf{n}_0 = \frac{\mathbf{u}_1^\top \mathbf{i}}{\sigma_1} \mathbf{v}_1 + \frac{\mathbf{u}_2^\top \mathbf{i}}{\sigma_2} \mathbf{v}_2 \quad (6)$$

est l'unique solution de :

$$\begin{cases} \min_{\mathbf{n}} \|\mathbf{i} - \mathbf{S}^\top \mathbf{n}\|^2 \\ \text{s.c. } \|\mathbf{n}\| = \min \end{cases} \quad (7)$$

Si $\|\mathbf{n}_0\| = 1$, alors \mathbf{n}_0 est l'unique solution exacte du problème (??) :

$$\mathbf{n}^+ = \mathbf{n}^- = \mathbf{n}_0 \quad (8)$$

Si $\|\mathbf{n}_0\| < 1$, l'ensemble des solutions du problème non contraint est donné par :

$$\mathbf{n}_0 + \mathbb{R}\mathbf{v}_3 \quad (9)$$

où \mathbf{v}_3 est un vecteur unitaire du noyau de \mathbf{S} . Parmi ces solutions, deux sont unitaires :

$$\begin{cases} \mathbf{n}^+ = \mathbf{n}_0 + \sqrt{1 - \|\mathbf{n}_0\|^2} \mathbf{v}_3 \\ \mathbf{n}^- = \mathbf{n}_0 - \sqrt{1 - \|\mathbf{n}_0\|^2} \mathbf{v}_3 \end{cases} \quad (10)$$

Ce sont donc les deux seules solutions de (??).

Enfin, si $\|\mathbf{n}_0\| > 1$, le problème (??) n'admet aucune solution exacte. On pourra choisir la solution approchée :

$$\mathbf{n}^+ = \mathbf{n}^- = \frac{\mathbf{n}_0}{\|\mathbf{n}_0\|} \quad (11)$$

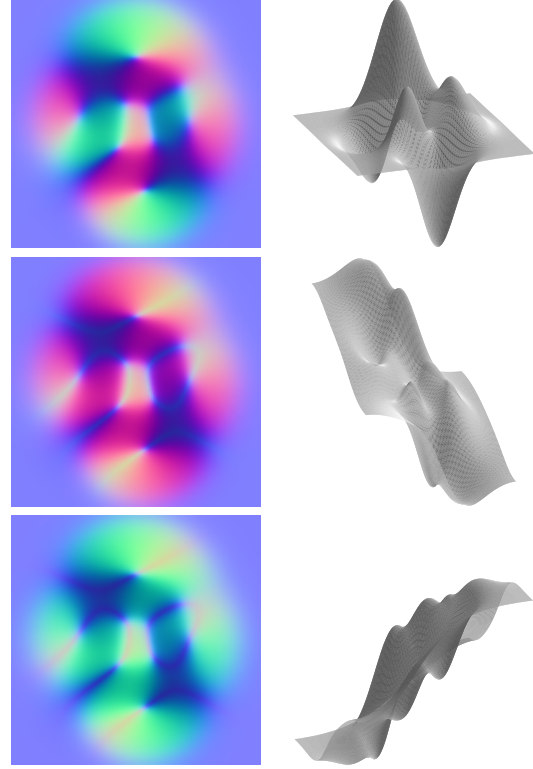


FIGURE 2 – Illustration des ambiguïtés inhérentes au problème. À gauche : champ de normales. À droite : relief obtenu en intégrant les normales par transformation en cosinus discrète [?]. De haut en bas : vérité terrain, solution \mathbf{n}^+ et solution \mathbf{n}^- . Les reliefs correspondant aux champs \mathbf{n}^+ et \mathbf{n}^- ne sont que deux reliefs parmi une infinité, puisque toute combinaison de ces deux champs est également plausible.

Dans les faits, le cas $\|\mathbf{n}_0\| < 1$, qui correspond à une ambiguïté d'ordre 2, est le plus courant. On peut donc définir $2^{|\Omega|}$ champs de normales aussi plausibles les uns que les autres, où $|\Omega|$ désigne le nombre de pixels contenus dans le domaine de reconstruction Ω . Il a été récemment proposé de contraindre davantage le problème en supposant que les normales sont distribuées selon une loi de Laplace [?]. Ceci constitue une connaissance a priori très forte sur la surface, qui n'est réaliste que pour des reliefs plus ou moins symétriques (les auteurs de [?] s'intéressent effectivement à la reconstruction 3D de visages). Inspirés par les résultats théoriques sur le caractère bien posé ou non du problème de la reconstruction 3D par stéréophotométrie à deux images [?, ?], nous proposons plutôt d'utiliser l'intégrabilité du champ de normales comme critère, ce qui est beaucoup moins restrictif (la surface est simplement supposée deux fois différentiable). Nous allons montrer dans le prochain paragraphe comment utiliser ce critère et l'algorithme de coupure de graphe pour déterminer efficacement une solution globale.

4 Désambiguisation du problème

Supposons maintenant que la surface à reconstruire soit représentable par une fonction de profondeur $u : \Omega \rightarrow \mathbb{R}$, et que u soit différentiable sur le domaine Ω , qui est entendu ici comme un sous-ensemble de la grille régulière des pixels. Sous l'hypothèse d'une caméra orthographique :

$$\mathbf{n} = \frac{1}{\sqrt{\|\nabla u\|^2 + 1}} \begin{bmatrix} -\nabla u \\ 1 \end{bmatrix} \quad (12)$$

où $\nabla u = [u_x, u_y]$ est le gradient de u . En notant $\mathbf{n} = [n_1, n_2, n_3]^T$ les composantes de \mathbf{n} , on en déduit :

$$\nabla u = \begin{bmatrix} -n_1/n_3 \\ -n_2/n_3 \end{bmatrix} \quad (13)$$

Nous supposons dorénavant que u est deux fois différentiable, et notons $p = -n_1/n_3$ et $q = -n_2/n_3$. Alors, en vertu du théorème de Schwarz, $(u_x)_y = (u_y)_x$, ce qui se traduit par la *condition d'intégrabilité* sur les normales :

$$p_y = q_x \quad (14)$$

Nous proposons donc de choisir, parmi les $2^{|\Omega|}$ champs de normales possibles, celui qui est « le plus intégrable ». Autrement dit, il s'agit d'attribuer à chaque pixel $(x, y) \in \Omega$ une étiquette $e \in \{+, -\}$ correspondant au choix de la normale \mathbf{n}^+ ou \mathbf{n}^- (cf. figure ??) :

$$\min_e \sum_{\Omega} (p_y^e - q_x^e)^2 \quad (15)$$

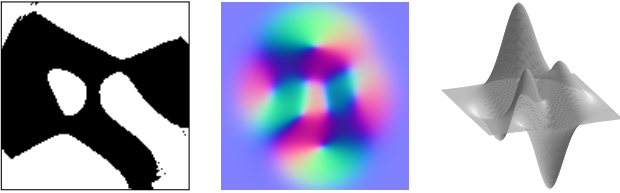


FIGURE 3 – Désambiguisation par coupure de graphe. La minimisation de l'intégrabilité à partir des champs \mathbf{n}^+ et \mathbf{n}^- de la figure ?? fournit l'étiquetage de gauche (couleur noire pour \mathbf{n}^+ , blanche pour \mathbf{n}^-). En choisissant cette combinaison, on obtient le champ de normales de la figure centrale, qui correspond au relief de droite. Ce champ de normales et ce relief sont similaires à ceux de la vérité terrain présentés sur la figure ?? . Le temps de calcul sur cet exemple de taille 128×128 est inférieur à 1 ms.

Plutôt que de procéder à une évaluation exhaustive de tous les champs possibles, nous allons résoudre ce problème d'optimisation combinatoire par coupure de graphe [?]. Pour discrétiser les dérivées spatiales de l'équation (??), nous utilisons des différences finies d'ordre 1. Ceci nécessite de considérer les quatre familles de cliques suivantes :

$$\begin{cases} \mathcal{C}_1 = \{ \{(x, y), (x-1, y), (x, y-1)\} \in \Omega^3 \} \\ \mathcal{C}_2 = \{ \{(x, y), (x+1, y), (x, y-1)\} \in \Omega^3 \} \\ \mathcal{C}_3 = \{ \{(x, y), (x-1, y), (x, y+1)\} \in \Omega^3 \} \\ \mathcal{C}_4 = \{ \{(x, y), (x+1, y), (x, y+1)\} \in \Omega^3 \} \end{cases} \quad (16)$$

Dans la suite, nous notons (e_1, e_2, e_3) les étiquettes associées aux trois pixels des cliques \mathcal{C}_j , $j \in [1, 4]$, dans l'ordre où ils sont définis en (??). Le problème (??) se réécrit alors :

$$\min_e \sum_{c \in \mathcal{C}} V_c^{\text{int}}(e_1, e_2, e_3) \quad (17)$$

où $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4$ et le *potentiel* V_c^{int} représente l'intégrabilité locale $(p_y - q_x)^2$ pour la clique et l'étiquetage courants. Par exemple, pour une clique c_1 de \mathcal{C}_1 :

$$V_{c_1}^{\text{int}}(e_1, e_2, e_3) = \left[(p^{e_1}(x, y) - p^{e_3}(x, y-1)) - (q^{e_1}(x, y) - q^{e_2}(x-1, y)) \right]^2 \quad (18)$$

Le problème (??) est un problème d'étiquetage où le potentiel local dépend du pixel courant et de deux de ses voisins. De tels problèmes d'optimisation combinatoire ont été étudiés dans [?], où il est notamment prouvé que l'algorithme de coupure de graphe peut être utilisé pour minimiser l'énergie, à condition de s'assurer de sa *régularité*, ce qui suppose ici :

$$V_c^{\text{int}}(+, +, e_3) + V_c^{\text{int}}(-, -, e_3) \leq V_c^{\text{int}}(+, -, e_3) + V_c^{\text{int}}(-, +, e_3) \quad (19)$$

$$V_c^{\text{int}}(+, e_2, +) + V_c^{\text{int}}(-, e_2, -) \leq V_c^{\text{int}}(+, e_2, -) + V_c^{\text{int}}(-, e_2, +) \quad (20)$$

$$V_c^{\text{int}}(e_1, +, +) + V_c^{\text{int}}(e_1, -, -) \leq V_c^{\text{int}}(e_1, +, -) + V_c^{\text{int}}(e_1, -, +) \quad (21)$$

Ces inégalités n'ayant a priori aucune raison d'être vérifiées, nous forçons la régularité en ajoutant à V_c^{int} trois termes de régularisation sous la forme de potentiels d'Ising :

$$V_c(e_1, e_2, e_3) = V_c^{\text{int}}(e_1, e_2, e_3) + \beta_c^{1,2} \delta(e_1 \neq e_2) + \beta_c^{1,3} \delta(e_1 \neq e_3) + \beta_c^{2,3} \delta(e_2 \neq e_3) \quad (22)$$

où $\beta_c^{1,2}$, $\beta_c^{1,3}$ et $\beta_c^{2,3}$ sont trois coefficients positifs ou nuls à définir, de telle sorte que le nouveau problème :

$$\min_e \sum_{c \in \mathcal{C}} V_c(e_1, e_2, e_3) \quad (23)$$

puisse être résolu par coupure de graphe. Notons que le problème (??) est différent du problème initial (??), puisqu'un a priori sur le caractère lisse par morceaux de l'étiquetage optimal est introduit. Il s'avère que cet a priori est justifié dans le cas où les solutions \mathbf{n}^+ et \mathbf{n}^- sont calculées comme indiqué dans le paragraphe ?? . Cependant, il est bien sûr préférable de choisir des valeurs aussi faibles que possible pour les coefficients β_c . En appliquant les conditions de régularité (??), (??) et (??) à l'énergie ainsi modifiée, on obtient les minoration suivantes de ces coefficients :

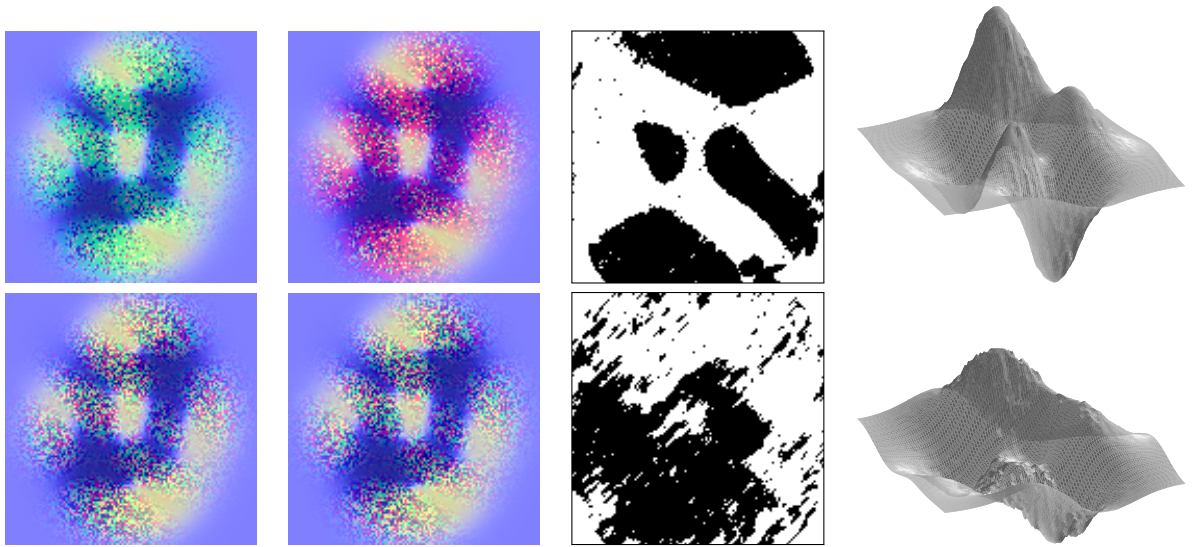


FIGURE 4 – En échangeant aléatoirement $r = 20\%$ (première ligne) et $r = 50\%$ (deuxième ligne) des valeurs de \mathbf{n}^+ (première colonne) et \mathbf{n}^- (deuxième colonne), les résultats se dégradent (voir également la figure ??). En effet, même si l'influence du terme de régularisation est limitée par le choix de paramètres β_c minimaux, l'a priori sous la forme de potentiels d'Ising s'oppose à un étiquetage non uniforme par morceaux : les étiquetages optimaux sont beaucoup moins lisses que ceux estimés ici (troisième colonne), ce qui induit un biais important dans la reconstruction 3D (quatrième colonne). Il est donc important de favoriser le caractère lisse des champs \mathbf{n}^+ et \mathbf{n}^- , ce qui est naturellement le cas en choisissant la solution algébrique proposée dans le paragraphe ??.

$$\beta_c^{1,2} \geq \frac{1}{2} \max \left\{ \begin{aligned} &V_c(+, +, +) + V_c(-, -, +) \\ &- V_c(+, -, +) - V_c(-, +, +), \\ &V_c(+, +, -) + V_c(-, -, -) \\ &- V_c(+, -, -) - V_c(-, +, -), 0 \end{aligned} \right\} \quad (24)$$

$$\beta_c^{1,3} \geq \frac{1}{2} \max \left\{ \begin{aligned} &V_c(+, +, +) + V_c(-, +, -) \\ &- V_c(+, +, -) - V_c(-, +, +), \\ &V_c(+, -, +) + V_c(-, -, -) \\ &- V_c(+, -, -) - V_c(-, -, +), 0 \end{aligned} \right\} \quad (25)$$

$$\beta_c^{2,3} \geq \frac{1}{2} \max \left\{ \begin{aligned} &V_c(+, +, +) + V_c(+, -, -) \\ &- V_c(+, +, -) - V_c(+, -, +), \\ &V_c(-, +, +) + V_c(-, -, -) \\ &- V_c(-, +, -) - V_c(-, -, +), 0 \end{aligned} \right\} \quad (26)$$

On peut donc déterminer a priori, pour chaque clique c , la plus petite valeur de chacun de ces coefficients permettant de rendre l'énergie régulière, ce qui permet d'éviter tout sur-lissage (cf. figure ??). Les coefficients $\beta_c^{1,2}$, $\beta_c^{1,3}$ et $\beta_c^{2,3}$ ne doivent donc pas être vus comme des paramètres, puisqu'ils n'ont pas à être réglés manuellement et ne servent qu'à garantir la régularité de l'énergie.

Néanmoins, l'introduction d'un terme de régularisation correspond à un certain a priori sur le caractère lisse (au

moins par morceaux) de l'étiquetage, et donc sur celui des champs \mathbf{n}^+ et \mathbf{n}^- . Heureusement, s'ils sont estimés comme indiqué dans le paragraphe ??, ces champs présentent ces caractéristiques. Mais si l'on simule une répartition moins lisse des normales entre \mathbf{n}^+ et \mathbf{n}^- , en échangeant aléatoirement une proportion r de valeurs de \mathbf{n}^+ et \mathbf{n}^- , les résultats se dégradent très nettement (cf. figures ?? et ??).

La méthode que nous préconisons comporte donc deux étapes :

1. Calcul des deux solutions \mathbf{n}^+ et \mathbf{n}^- en chaque pixel, comme indiqué dans le paragraphe ??.
2. Désambiguïsation du problème par l'intégrabilité, en utilisant la coupure de graphe.

Ces étapes peuvent toutes deux être menées très efficacement : le calcul des deux solutions initiales est explicite, et l'optimisation par coupure de graphe garantit un temps de calcul minimal.

À ce stade, il semblerait opportun de valider notre approche sur des images réelles. Cela est malheureusement impossible, car il n'est pas réaliste de supposer, pour des images acquises par une webcam, que les intensités lumineuses et l'albédo sont connus. Dans le prochain paragraphe, nous montrons comment étendre cette approche au cas où les intensités ϕ^i sont inconnues et colorées.

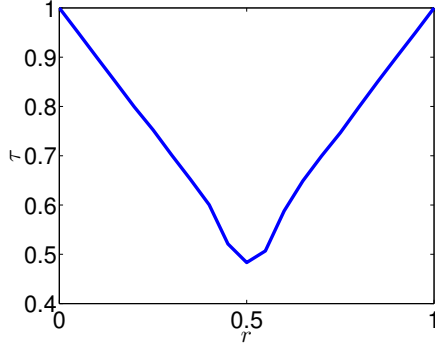


FIGURE 5 – Proportion τ de bonnes étiquettes en fonction de la proportion r d'échanges entre \mathbf{n}^+ et \mathbf{n}^- (cf. figure ??). Les résultats sont excellents lorsque les deux champs \mathbf{n}^+ et \mathbf{n}^- sont lisses : dans les cas $r = 0$, qui correspond aux données de la figure ??, et $r = 1$, qui correspond à l'échange global entre \mathbf{n}^+ et \mathbf{n}^- , on obtient $\tau = 1$. En échangeant aléatoirement \mathbf{n}^+ et \mathbf{n}^- , le taux de bons étiquetages décroît linéairement jusqu'à $\tau = 0,5$ pour $r = 0,5$ (cette valeur de τ correspond à un choix totalement aléatoire des étiquettes).

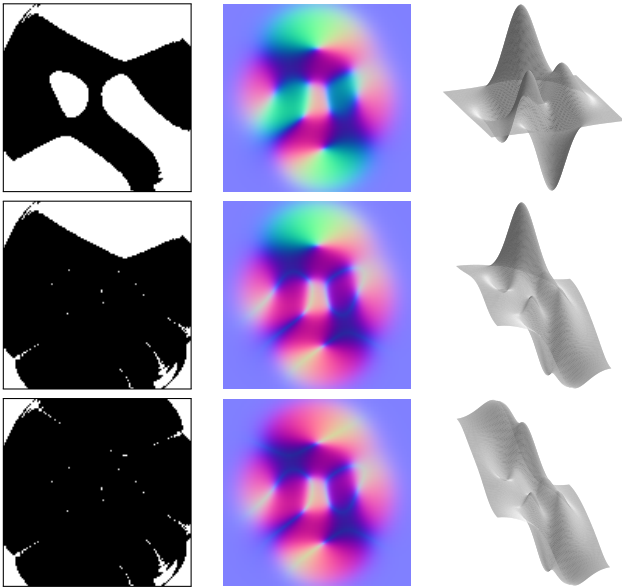


FIGURE 6 – Sur-lissage de l'étiquetage pour des valeurs élevées de β_c . Nous montrons, de gauche à droite, l'étiquetage, le champ de normales estimé, et le relief correspondant, en fixant les paramètres, de haut en bas, à 5, 10 et 100 fois leur valeur minimale. Même s'il est nécessaire d'introduire une régularisation de l'étiquetage pour obtenir une énergie régulière au sens de [?], on a tout intérêt à minimiser cette régularisation afin de ne pas biaiser la solution.

5 Cas des images réelles

5.1 Intensités inconnues

Nous supposons maintenant que les intensités lumineuses ϕ^i sont inconnues. Cela est notamment le cas des images prises en extérieur, puisque ces intensités dépendent de nombreux phénomènes météorologiques. Nous pouvons modéliser ce nouveau problème de la façon suivante :

$$\min_{\mathbf{n}, \phi^i} \sum_{i=1}^m (I^i - \rho \phi^i \mathbf{s}^i \cdot \mathbf{n})^2$$

$$\text{s.c.} \begin{cases} \|\mathbf{n}\| = 1 \\ \sum_{\Omega} [(-n_1/n_3)_y - (-n_2/n_3)_x]^2 = \min \end{cases} \quad (27)$$

Il s'agit alors de déterminer la normale \mathbf{n} en chaque point, et l'intensité lumineuse ϕ^i pour chaque image, de façon à minimiser l'écart quadratique entre les données et les images « reprojetées ». Parmi l'infinité de champs de normales permettant de minimiser ce critère, nous retenons le plus intégrable.

L'optimisation peut être réalisée en alternant les estimations en \mathbf{n} et en ϕ^i , à partir d'une solution initiale $\mathbf{n}^0 = [0, 0, 1]^T$ (relief plan) et $\phi^{i,0} = 1, i \in [1, m]$ (éclairages de même intensité). La mise à jour des normales s'effectue comme indiqué précédemment, en estimant dans un premier temps tous les champs unitaires acceptables (paragraphe ??), puis en sélectionnant le plus intégrable par coupure de graphe (paragraphe ??). La condition d'optimalité vis-à-vis des variables ϕ^i donne, pour $i \in [1, m]$:

$$\phi^{i,k+1} = \frac{\sum_{\Omega} I^i \rho \mathbf{s}^i \cdot \mathbf{n}^k}{\sum_{\Omega} (\rho \mathbf{s}^i \cdot \mathbf{n}^k)^2} \quad (28)$$

Afin de vérifier expérimentalement si ce schéma converge, et si oui, à quelle vitesse, nous simulons les images du relief synthétique de la figure ?? sous $m = 100$ éclairages dont les directions sont contenues dans un plan vertical. L'angle des directions d'éclairage avec la verticale varie uniformément entre $-\pi/6$ et $\pi/6$, c'est-à-dire que le soleil est au zénith à midi, afin de limiter les ombres, qui sont toutefois simulées en seillant à zéro les valeurs négatives du niveau de gris. Les intensités lumineuses ϕ^i sont tirées aléatoirement selon une loi uniforme sur $[0, 1]$.

Nous évaluons l'évolution de l'erreur de projection

$$E_r = \sqrt{\frac{1}{m|\Omega|} \sum_{\Omega} \sum_{i=1}^m (I^i - \rho \phi^i \mathbf{s}^i \cdot \mathbf{n})^2} \quad (29)$$

et celle de l'erreur d'intégrabilité

$$E_i = \sqrt{\frac{1}{|\Omega|} \sum_{\Omega} (p_y - q_x)^2} \quad (30)$$

en fonction du nombre d'itérations k . Les résultats de la figure ?? montrent une convergence très rapide de ces deux critères.

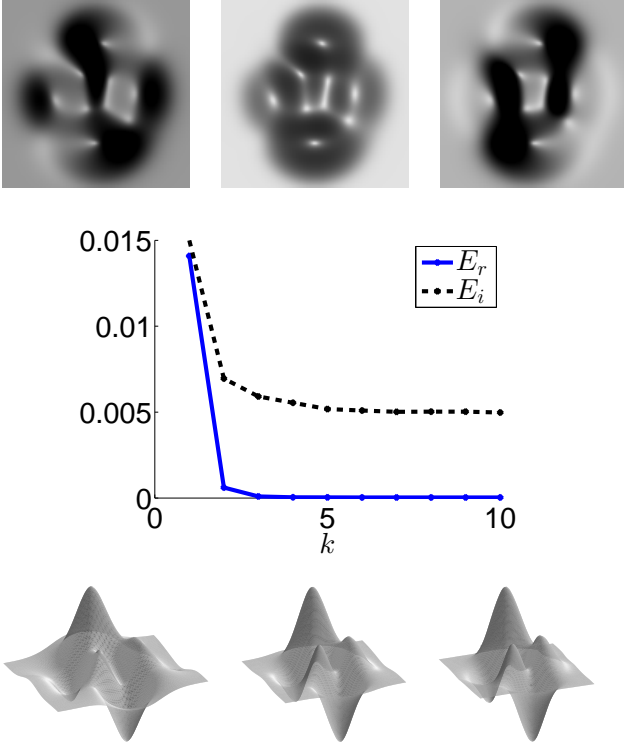


FIGURE 7 – Estimation simultanée du relief et des intensités lumineuses. Première ligne : trois des $m = 100$ images simulées. Deuxième ligne : évolution de E_r et E_i en fonction du nombre d’itérations k ($k = 1, 2, \dots, 10$). Troisième ligne : relief estimé pour $k = 1$, $k = 2$ et $k = 10$.

5.2 Éclairages colorés

Nous nous sommes restreints jusqu’ici à des images en niveaux de gris. Il est également possible d’utiliser des images RVB, en supposant que la couleur observée dépend uniquement du spectre incident, et non de la texture de l’objet. Toute notion de couleur est alors entendue relativement à celle de la scène, qui est supposée homogène et définit le blanc de référence. Ceci restreint toutefois notre étude à des objets non texturés, le cas d’objets texturés étant laissé en perspective. Sous cette hypothèse, le nouveau modèle lambertien s’écrit :

$$[I_r^i, I_v^i, I_b^i] = \rho \mathbf{s}^i \cdot \mathbf{n} [\phi_r^i, \phi_v^i, \phi_b^i] \quad (31)$$

où ϕ_r^i , ϕ_v^i et ϕ_b^i représentent les intensités lumineuses dans les canaux R, V et B.

On peut ainsi retrouver l’évolution du spectre au cours de la journée, en même temps que la géométrie de la scène, car l’algorithme précédent s’adapte sans difficulté : l’estimation des intensités est réalisée canal par canal, et le système linéaire (??) devient maintenant un système de $3m$ équations correspondant aux trois canaux (l’estimation de \mathbf{n} n’en est que plus robuste). Ceci est particulièrement utile dans le cas des webcams, car en général la balance des blancs s’adapte automatiquement à la couleur de l’éclairage. Estimer des intensités lumineuses vectorielles permet

donc de compenser l’absence de calibrage photométrique de la webcam.

Un exemple de reconstruction 3D d’un morceau de la Statue de la Liberté est montré sur la figure ?? . Les intensités colorées estimées, relativement à la couleur de la statue, sont représentées sur la figure ??, et correspondent qualitativement aux variations de couleur perceptibles sur les images.

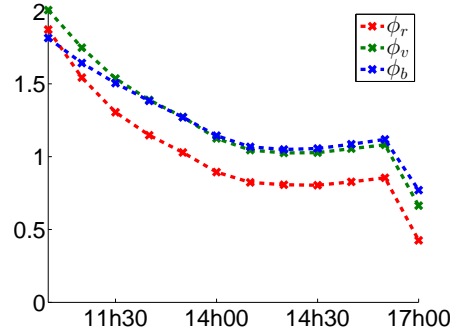


FIGURE 9 – Estimation des intensités lumineuses pour les $m = 12$ images de l’expérience de la figure ?? . Les variations d’intensité correspondent à des changements perceptibles sur les images (l’image 12 comporte une ombre, ce qui provoque une chute brutale des intensités).

6 Conclusion et perspectives

Dans cet article, nous avons présenté une étude théorique du problème de la reconstruction 3D par stéréophotométrie dans le cas dégénéré où les éclairages sont coplanaires, et nous avons proposé une résolution efficace utilisant l’intégrabilité du champ de normales et l’algorithme de coupe de graphe. Nous avons montré que cette approche permettait d’estimer les intensités colorées des éclairages et la géométrie d’une scène à partir d’images en couleur prises par une webcam au cours d’une même journée ensoleillée. Toutefois, nous avons supposé l’albédo connu, ou au moins uniforme, ce qui limite en pratique notre méthode à des données synthétiques ou à des objets réels non texturés (exemple : les glaciers). Une extension naturelle de ce travail consistera à traiter le cas d’un albédo inconnu et non uniforme. Il faudra probablement utiliser un a priori sur la distribution de l’albédo, à la manière de ce qui est fait dans [?] pour résoudre les ambiguïtés de la stéréophotométrie non calibrée dans le cas d’éclairages non coplanaires.

Nous avons également fait l’hypothèse d’un modèle de caméra orthographique. Une autre piste consiste à étendre ce travail à une caméra perspective, ce qui implique une relation légèrement différente entre la profondeur u et la normale \mathbf{n} : il faudra alors adapter la définition du critère d’intégrabilité, comme cela a déjà été fait dans le cas de la stéréophotométrie non calibrée mais non dégénérée [?].

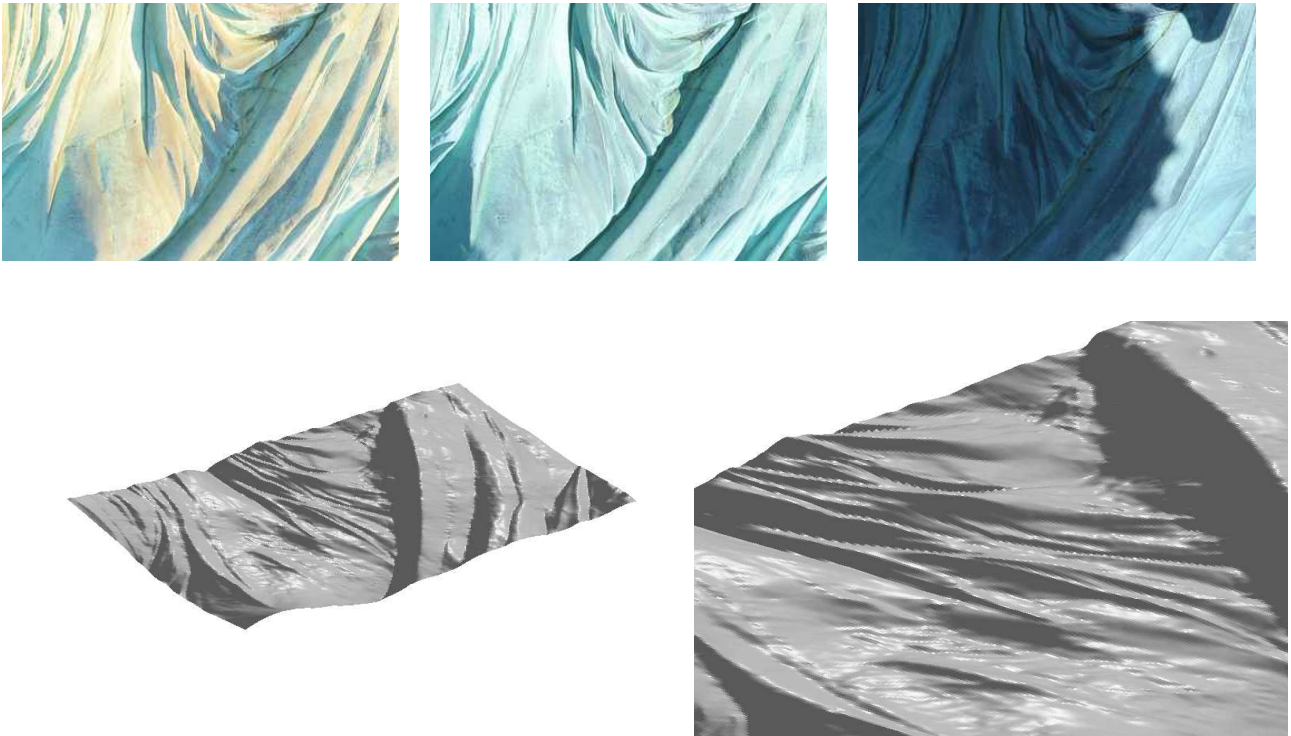


FIGURE 8 – Reconstruction 3D d’une partie de la Statue de la Liberté, de couleur uniforme, à partir de 12 images, de taille 231×356 , prises entre 10h30 et 17h00. Première ligne : I^1 , I^6 et I^{12} . La radiométrie varie beaucoup d’une image à l’autre, à cause de l’effet combiné du changement de spectre lumineux au cours de la journée, de la balance automatique des blancs de la webcam, et du passage de nuages. Ces effets peuvent être pris en compte en utilisant les trois canaux RVB. Deuxième ligne : reconstruction 3D obtenue, et gros plan sur les « plis » de la Statue de la Liberté (temps de calcul pour 5 itérations : moins d’une seconde sur un processeur I7 récent). Une telle reconstruction 3D pourrait servir à enrichir des modèles 3D grossiers : par exemple, le modèle 3D de la Statue de la Liberté disponible dans Google Earth ne comporte aucun des détails très fins retrouvés ici.