HAL open science

Calibrage Radiométrique en Utilisant des Collections de Photos

Mauricio Díaz, Peter Sturm

To cite this version:

Mauricio Díaz, Peter Sturm. Calibrage Radiométrique en Utilisant des Collections de Photos. ORASIS 2011 - Congrès des jeunes chercheurs en vision par ordinateur, INRIA Grenoble Rhône-Alpes, Jun 2011, Praz-sur-Arly, France. inria-00595363

HAL Id: inria-00595363

https://inria.hal.science/inria-00595363v1

Submitted on 24 May 2011

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.

Calibrage Radiométrique en Utilisant des Collections de Photos
Radiometric Calibration using Photo Collections

Mauricio Díaz Peter Sturm

INRIA Grenoble Rhône-Alpes – Laboratoire Jean Kuntzmann
38330 Montbonnot, France
Mauricio.Diaz@inrialpes.fr, Peter.Sturm@inrialpes.fr

Résumé

Mesurer l'irradiance émise par une scène est une caractéristique hautement souhaitable dans de nombreux algorithmes de vision par ordinateur. Des diverses applications comme, par exemple, l'estimation de la BRDF ou la réalité augmentée nécessitent des estimations des propriétés photométriques fiables. La méthode la plus simple pour les obtenir est avec l'aide d'un appareil photo. Toutefois, un premier pas est nécessaire pour atteindre cet objectif. Les valeurs d'intensités obtenues par un appareil photo doivent être liées aux valeurs d'irradiance émises par l'objet. Ce phénomène est modelé par une fonction non linéaire appelée la réponse radiométrique de la camera (CRF ou camera response function en anglais). Dans cet article, nous proposons d'exploiter la grande diversité des apparences d'un objet dans des collections de photos pour récupérer cette fonction non linéaire pour chacune des caméras en utilisant les images disponibles. Ce processus, également connu sous le nom de calibrage radiométrique, utilise un ensemble non structuré d'images pour récupérer la position des cameras et un modèle 3D de la scène. À partir de cette entrée, la fonction de réponse de la caméra est estimée pour chaque image. Ce problème très mal posé est uniquement traitable en se servant des à prioris appropriés. L'approche proposée est basée sur un modèle empirique pour paramétrer la réponse radiométrique des caméras introduit par Grossberg et Nayar. Des méthodes linéaires sont proposées. Ces approches permettent de calculer des solutions approximatives. Ensuite, des solutions plus exactes sont obtenues par des algorithmes d'optimisation non linéaire.

Mots Clef

Photométrie, Modélisation 3D, Calibrage de Caméra.

Abstract

Access to the scene irradiance is a desirable feature in many computer vision algorithms. Applications like BRDF estimation, relighting or augmented reality need measurements of the object's photometric properties and the simplest method to get them is using a camera. However, the first step necessary to achieve this goal is the computation of the function that relates scene irradiance to image intensities. In this paper we propose to exploit the large variety of an object's appearances in photo collections to recover this non linear function for each of the cameras that acquired the available images. This process, also known as radiometric calibration, uses an unstructured set of images, to recover the camera's geometric calibration and a 3D scene model, using available methods. From this input, the camera response function is estimated for each image. This highly ill-posed problem is made tractable by using appropriate priors. The proposed approach is based on the empirical prior on camera response functions introduced by Grossberg and Nayar. Linear methods are proposed that allow to compute approximate solutions, which are then refined by non–linear least squares optimization.

Keywords

Photometry, 3D Modeling, Camera Calibration.

1 Introduction

Le but du travail présenté dans ce document est de récupérer le calibrage radiométrique pour plusieurs appareils photo en utilisant une collection d'images acquises dans des conditions non contrôlées. De nos jours, des grandes collections d'images montrant différentes vues et angles d'une scène, comme par exemple, une structure touristique, son disponibles sur Internet. Ces bases de données, construites de façon collaborative par la communauté, présentent une grande richesse en variations de l'apparence. Notre motivation est d'exploiter tout cette gamme exubérante de fluctuations et d'extraire des informations uniquement à partir des images. Depuis quelques années, l'utilisation de grandes bases de données est un sujet d'intérêt en vision par ordinateur. Parmi les pionniers, le travail de Snavely *et al.* [23] a proposé l'utilisation d'une technique de *Structure From Motion* appliquée à de grandes collections d'images pour récupérer la position et l'orientation des appareils photo impliqués. Cet algorithme constitue l'épine dorsale du système PhotoSytnh, une application qui offre une façon intuitive, agréable et efficace de naviguer à travers de grandes collections d'images. Entre temps, d'autres projets ont centré leurs efforts en résolvant le problème de reconstruction 3D robuste dans des environnements non contrôlés, utilisant une mesure photométrique consistante (*e.g.* [6]). Les résultats obtenus par ces travaux en particu-

FIGURE 1 – Cathédrale de Rouen de Claude Monet. Trois tableaux de la même scène peints dans différentes conditions d'illumination.

lier permettent aujourd'hui de créer un modèle 3D dense d'un objet. Par exemple, il est possible de reconstruire un monument célèbre en utilisant des images capturées dans des conditions totalement décontractées et naturelles. Ce modèle géométrique est un point de départ pour déduire des propriétés photométriques de la scène.

Si on choisit une cible déterminée, il est possible de trouver dans de nombreux sites web des milliards d'images de l'objet en question. Même si ces photos représentent le même objet, une grande gamme d'apparences est disponible dans toutes les images. Cette richesse de couleurs s'explique car la scène est capturée dans des conditions d'éclairage différentes, par exemple, des images prises à différents moments de la journée (ou de l'année) avec des conditions météorologiques distinctes. Depuis le XIXme siècle, le célèbre peintre impressionniste Claude Monet avait déjà montré une fascination particulière pour ce phénomène comme on peut voir dans sa série intitulée *La Cathédrale de Rouen* (voir figure 1). C'est cette opulence en apparence le facteur que nous voulons exploiter. De plus, les appareils photo ont tous des capteurs différents, qui fonctionnent, la plupart des cas, d'une façon très hétérogène. On n'obtient pas la même image avec deux appareils photo différents, même si les conditions restent inchangées. En conséquence, modéliser l'aspect photométrique des objets ne peut pas être réalisé directement à partir des images. Les images sont le résultat de l'interaction de différents facteurs : la géométrie de la scène, la réflectance de la surface, les conditions d'éclairage et les capteurs utilisés. Par ce dernier, nous nous référons dans le présent document à la fonction de réponse photométrique de la caméra (CRF), mais d'autres aspects existent également, telles que la profondeur de champ ou des aberrations optiques. Afin d'accéder aux informations de la réflectance sur la surface, il semble essentiel de connaître la CRF. Cependant, ce n'est généralement pas le cas pour les images dans des collections de photos. De plus, les conditions d'éclairage sont en général inconnues.

Nous pouvons formuler notre problème comme suit : pour un certain groupe d'images non calibrées qui correspondent à une scène particulière, il est possible de trouver un modèle 3D de la scène (dans notre travail, ce modèle est obtenu en utilisant Bundler [24] et PMVS [7]), nous souhaitons déterminer les CRF pour toutes les images, même si elles sont toutes différentes. Ce problème est en général très mal posé : même si nous connaissions parfaitement les conditions d'illumination pour chaque image, nous pourrions éventuellement, pour chaque ensemble hypothétique de CRF, définir des propriétés photométriques pour chaque point de la surface qui sont compatibles avec les observations dans toutes les images. De toute évidence, afin de résoudre le problème, une connaissance à priori est nécessaire. Cette information pourrait concerner l'éclairage, la réflectance de la surface ainsi que l'espace de CRFs possibles. Dans cet article, nous utilisons une restriction forte mais réaliste sur le type de CRF qui peuvent se former, ainsi que des modèles simples pour la réflectance et l'éclairage.

2 Travaux Précédents

Ces dernières années, une partie de la communauté de vision par ordinateur a concentré son attention sur l'utilisation des collections de photos pour obtenir des informations sur la géométrie de la scène. Après les résultats prometteurs de [24], où une géométrie éparse de la scène et les caméras est récupéré, d'autres travaux ont exploré différents aspects. Haber *et al.* [10] ont développé une approche pour récupérer les propriétés photométriques et l'éclairage en utilisant une représentation de l'interaction entre la lumière et la surface dans le cadre des ondelettes. Dans ce travail, les auteurs supposent que les images extraites des collections de photos peuvent être corrigées, en termes photométriques, par une courbe de gamma. Il a été largement démontré que les fonctions de réponse des appareils photo numériques sont souvent des fonctions plus complexes.

Il existe différentes approches pour le calibrage radiométrique d'un appareil photo. Deux catégories peuvent être distinguées : les méthodes actives et passives. La récupération active de la CRF rassemble toutes les méthodes qui nécessitent une présence physique au cours du processus d'acquisition. Des approches communes, notamment en utilisant soit des objets d'étalonnage spécifiques (chartes de couleurs tels que le populaire ColorChecker Gretag-Macbeth, voir `http://www.xrite.com`) [3, 11], soit plusieurs images acquises par un appareil photo statique, mais avec différents durées d'exposition [4, 18]. Naturellement, ces approches ne sont pas applicables si le but est le calibrage automatique des images dans une collection partagée par la communauté. Une deuxième catégorie comprend des approches où la présence physique sur la scène n'est pas nécessaire. L'idée de ce type d'algorithmes est de chercher des caractéristiques dans l'image qui reflètent la non linéarité produite par la fonction de réponse de la caméra. Dans ce cas, les régions d'intérêt à chercher sont, par exemple, des représentations de bas niveau comme les bords où les changements de couleur se produisent de façon drastique [14], [15] ou des invariants géométriques [19]. Ces régions permettent d'extrapoler une estimation de la CRF. D'autres artefacts peuvent être aussi utilisés, comme par exemple le bruit présent dans les images numériques [25]. Par exemple dans le cas de[14], Lin *et al.* utilisent qu'une seule image pour calculer la CRF à partir des bords où deux régions de couleur constant peuvent être clairement différenciées. Dans [13], Kuthirummal *et al.* ont

trouvé que les grandes collections des photos révèlent des formes statistiques clairement définies. A partir de ces à prioris, ils estiment les fonctions de réponse des modèles génériques d'appareils photo, en supposant que toutes les instances d'un modèle d'appareil photo ont les mêmes propriétés et que de nombreuses images prises par la caméra à calibrer, sont disponibles. Des dispositifs personnalisés, *e.g.* , les appareils photo à objectifs interchangeables, ne peuvent pas être modélisés en utilisant cette méthode. La plupart des méthodes mentionnées ci-dessus utilisent le modèle empirique proposé par Grossberg et Nayar [9], un modèle basé sur une analyse en composantes principales (voir la section 3).

Luong *et al.* [16] utilisent, comme notre approche, un modèle 3D de la scène imagée, pour effectuer le calibrage radiométrique et l'estimation d'éclairage. Toutefois, ils ont utilisé un modèle linéaire pour les CRF. En plus leur approche exige que plusieurs images soient prises avec le même appareil photo, dans des conditions contrôlées. Dans le cas des images téléchargées sur Internet, un modèle linéaire pour la CRF n'est pas exact et les conditions d'acquisition sont totalement incontrôlées.

Dans notre travail, nous estimons une CRF non linéaire à partir d'images acquises dans des conditions non contrôlées. De plus, il n'est pas nécessaire d'avoir plus d'une image par caméra. Les à prioris que nous utilisons concernent les propriétés de réflectance de la scène (surfaces Lambertiennes), l'éclairage (la combinaison d'une source lumineuse directionnelle et un terme ambiant) et les CRF à calibrer.

3 Énoncé du Problème

Comme expliqué précédemment, l'objectif des méthodes présentées dans ce document est le calibrage radiométrique des caméras à travers une collection d'images non structurée de la même scène. Concrètement, nous estimons la fonction non linéaire qui représente la réponse photométrique de la caméra (CRF) par image et par canal, ainsi que les conditions d'éclairage (également par image et par canal). Cette illumination comporte une source de lumière directionnelle et un facteur d'éclairage ambiant par rapport à la longueur d'onde du canal considéré. (La direction de la source lumineuse est la même pour tous les canaux, mais sa "force" peut varier, ainsi que le terme ambiant). Le cas où il est connu que deux images ont la même CRF et/ou les mêmes conditions d'éclairage, est une extension directe de notre algorithme.

Entrées de l'algorithme. La seule entrée principale de notre algorithme, sont les images elles-mêmes et les à prioris que nous utilisons (voir ci-dessous). Comme un prétraitement de nos méthodes, nous allons d'abord obtenir le calibrage géométrique de tous les appareils photo et un modèle de la scène 3D. Le calibrage géométrique est effectué en utilisant le logiciel BUNDLER par Snavely *et al.* [24] et le modèle 3D est calculé en utilisant le logiciel PMVS développé par Furukawa *et al.* [7]. Le modèle 3D est formé à partir d'éléments de la surface (points dans l'espace 3D et les normales associées). Des maillages ou d'autres représentations de modèles 3D peuvent aussi être utilisés de façon analogue. L'entrée complète pour notre algorithme est donc un ensemble de M images et leur calibrage géométrique, ainsi qu'une collection d'éléments de surface et des informations sur leur visibilité dans les images.

Formation de l'image. Le modèle de formation de l'image que nous utilisons est composé par une réflectance lambertienne qui caractérise les éléments de la surface et par un éclairage formé par une source de lumière directionnelle et par l'éclairage ambiant. Dans ces termes, l'irradiance E d'un élément j qui appartient à la surface, dans les conditions d'éclairage de l'image i, est donc exprimée par :

$$E_{ij} = \rho_j(\boldsymbol{n}_j^{\mathsf{T}}\boldsymbol{l}_i + \mu_i) \ , \tag{1}$$

où,

$$\boldsymbol{l_i} = \begin{bmatrix}\beta_i \cos\theta_i \cos\phi_i & \beta_i \sin\theta_i \cos\phi_i & \beta_i \sin\phi_i\end{bmatrix}^{\mathsf{T}}$$

est un vecteur qui décrit l'illumination pour l'image i. Ce vecteur est exprimé en coordonnées sphériques par rapport à un repère global et il est dépendant de la direction de la source de lumière directionnelle (θ_i,ϕ_i) et son intensité (β_i). L'albedo du $j^{\text{ème}}$ élément de la surface est dénoté par ρ_j, sa normale (un vecteur unitaire) est $\boldsymbol{n}_j$ et μ_i modélise le terme ambiant de l'éclairage. Suivant la notation décrite par Luong *et al.* [16], nous pouvons exprimer l'équation (1) par le produit de vecteurs $\boldsymbol{L}_i = \begin{bmatrix}\boldsymbol{l_i} & \mu_i\end{bmatrix}^{\mathsf{T}}$ et $\boldsymbol{N}_j = \begin{bmatrix}n_{j1} & n_{j2} & n_{j3} & 1\end{bmatrix}^{\mathsf{T}}$. C'est ainsi que la valeur de la luminosité d'un élément de surface j, observé dans l'image i avec une CRF f_i, est donné par :

$$B_{ij} = f_i(\rho_j \boldsymbol{N}_j^{\mathsf{T}}\boldsymbol{L}_i) \ , \tag{2}$$

où β_i, ρ_j, μ_i et f_i sont définis par canal de couleur.

À Priori sur les CRF. Selon l'étude de Grossberg et Nayar [9] les CRF qui appartiennent au monde réel sont enfermées dans un espace bien défini qui, peut être modélisé en utilisant une base empirique. Ce résultat nous permet d'exprimer les CRF en termes de quelques paramètres, trois semblent être suffisants [9]. Cette affirmation repose sur deux hypothèses importantes : la fonction de réponse f est la même pour tous les pixels qui appartiennent à une image particulière et la fonction est monotone. Cette dernière restriction est plausible dans la pratique, alors que la première peut être adressée dans les travaux futurs, afin de gérer des effets potentiels sur le vignettage des images. En plus, nous supposons que la sensibilité spectrale du capteur est similaire pour toutes les gammes d'intensité.

D'après les résultats de Grossberg et Nayar, la relation entre la luminosité de l'image B et l'irradiance E peut être exprimée comme une combinaison linéaire d'une valeur moyenne h_0 et de composantes principales h_n :

$$f_i(E) = h_0(E) + \sum_{n=1}^{N} w_{in} h_n(E) \ . \tag{3}$$

La CRF moyenne, et les composantes principales peuvent être représentées à l'aide des tableaux précédemment al-

loués ou en utilisant des polynômes. Sans perte de généralité, nous utilisons la deuxième possibilité dans notre travail. La principale différence entre les deux représentations, est que pour imposer la monotonie à la CRF estimée f_i, moins de contraintes d'inégalité sont nécessaires dans le cas des polynômes. Outre cette question, les formules qui décrivent nos méthodes seraient strictement analogues dans le cas des tableaux précédemment alloués.
Donc, la base pour l'espace des CRF est représentée par des polynômes de degré D :

$$h_n(E) = \sum_{d=0}^{D} c_{nd} E^d \quad . \tag{4}$$

D'après les auteurs de ce modèle empirique [9], les fonctions h_n sont exprimées par rapport à des valeurs normalisées de l'irradiance et de l'intensité, telles que $c_{n0} = 0$ et $\sum c_{nd} = 1$ pour tout $n = 1 \cdots N$. Le degré des polynômes a été choisi pour une représentation adéquate des courbes qui forment la base, ce qui a été obtenu avec $D = 9$. Ces polynômes sont connus, ce qui est inconnu est l'ensemble des coefficients w_{in} de leur combinaison linéaire (3).

Formulation complète du problème. Pour un canal donné, les valeurs non connues sont $\boldsymbol{\Theta} = \{\boldsymbol{L}_i, \rho_j, w_{in}\}$, avec un groupe de coefficients d'illumination $\boldsymbol{L}_i$ par image (les inconnues β_i et μ_i son calculées par canal de l'image, tandis que la direction de la source de lumière (θ_i et ϕ_i) est partagée par tous les canaux de l'image), un albedo ρ_j par point 3D sur la surface et par un groupe de coefficients w_{in} pour décrire les CRF par image. Avec les modèles exposés ci-dessus, les valeurs de luminosité prédites, étant donné les inconnues estimées sont :

$$\begin{aligned} \hat{B}_{ij} &= f_i(\rho_j \boldsymbol{N}_j^\top \boldsymbol{L}_i) \\ &= h_0(\rho_j \boldsymbol{N}_j^\top \boldsymbol{L}_i) + \sum_{n=1}^{N} w_{in} h_n(\rho_j \boldsymbol{N}_j^\top \boldsymbol{L}_i) \quad . \end{aligned} \tag{5}$$

Nous cherchons à estimer les inconnues en minimisant la différence entre les échantillons observés et les valeurs de luminosité estimées :

$$\underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i=1}^{M} \sum_{j=1}^{J} (B_{ij} - v_{ij} \hat{B}_{ij})^2 \quad . \tag{6}$$

Les scalaires v_{ij} sont des booléens, qui prennent une valeur de 1 si le point j est visible dans l'image i, sinon la valeur est 0. Notez que les inconnues ne peuvent être estimées sans ambiguïté : les albédos ρ_j et les coefficients d'éclairage $\boldsymbol{L}_i$ ne peuvent être estimés qu'à un facteur d'échelle global près. En plus de la fonction de coût (6), on peut également imposer des contraintes, par exemple sur la monotonie de la CRF estimée.

4 Solutions Proposées

Le type de problème posé dans l'équation (6) est connu dans le contexte de l'optimisation comme une estimation non linéaire aux moindres carrés. Le processus complet est décrit dans la section 4.3. Dans les sections suivantes, nous proposons une solution à partir des moindres carrés linéaires pour une fonction de coût modifiée, qui peut être utilisée pour initialiser l'optimisation non linéaire. Nous examinons d'abord le cas minimal de deux images. Cet exemple simple nous permet de montrer que, en théorie, deux images sont suffisantes pour résoudre le problème. Ensuite, la formulation est généralisée à plusieurs images.

4.1 Le Cas de Deux Images

Pour transformer notre problème au cas linéaire, nous considérons l'inverse de la CRF : $g_i(B) = f_i^{-1}(E)$. Selon Grossberg et Nayar [8], la CRF inverse peut aussi être représentée comme une somme de composantes principales de la forme que montre l'équation (3) :

$$g_i(B) = p_0(B) + \sum_{n=1}^{N} x_{in} p_n(B) \quad . \tag{7}$$

Nous allons estimer les coefficients x_{in} définissant les CRF inverses. Ensuite, les CRF directes peuvent être estimées par l'ajustement des coefficients à la courbe, voir ci-dessous.

À partir de l'équation (2), nous avons :

$$\rho_j = \frac{g_i(B_{ij})}{\boldsymbol{N}_j^\top \boldsymbol{L}_i} \quad . \tag{8}$$

Si un point 3D j est visible dans deux images, nous pouvons formuler la égalité suivante :

$$\frac{g_1(B_{1j})}{\boldsymbol{N}_j^\top \boldsymbol{L}_1} = \frac{g_2(B_{2j})}{\boldsymbol{N}_j^\top \boldsymbol{L}_2} \quad . \tag{9}$$

Donc,

$$g_1(B_{1j}) \boldsymbol{N}_j^\top \boldsymbol{L}_2 = g_2(B_{2j}) \boldsymbol{N}_j^\top \boldsymbol{L}_1 \quad .$$

En détail :

$$\begin{aligned} &\left(p_0(B_{1j}) \boldsymbol{N}_j^\top\right) \boldsymbol{L}_2 + \sum_{n=1}^{N} \left(p_n(B_{1j}) \boldsymbol{N}_j^\top\right) (x_{1n} \boldsymbol{L}_2) = \\ &\left(p_0(B_{2j}) \boldsymbol{N}_j^\top\right) \boldsymbol{L}_1 + \sum_{n=1}^{N} \left(p_n(B_{2j}) \boldsymbol{N}_j^\top\right) (x_{2n} \boldsymbol{L}_1) \quad , \end{aligned} \tag{10}$$

où nous avons regroupé les entités connues et inconnues, respectivement. Cette équation est linéaire par rapport aux inconnues suivantes : $\boldsymbol{L}_1, \boldsymbol{L}_2, x_{1n}\boldsymbol{L}_2, x_{2n}\boldsymbol{L}_1$, pout $n = 1 \cdots N$. Soit $\boldsymbol{q}_{nij} = p_n(B_{ij}) \boldsymbol{N}_j$. Puis, considérant J points 3D, les équations correspondant à (10) sont regrou-

pées et exprimées en notation matricielle de cette manière :

$$\begin{bmatrix} \boldsymbol{q}_{021}^{\mathsf{T}} & \cdots & \boldsymbol{q}_{02J}^{\mathsf{T}} \\ \boldsymbol{q}_{121}^{\mathsf{T}} & \cdots & \boldsymbol{q}_{12J}^{\mathsf{T}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{q}_{N21}^{\mathsf{T}} & \cdots & \boldsymbol{q}_{N2J}^{\mathsf{T}} \\ -\boldsymbol{q}_{011}^{\mathsf{T}} & \cdots & -\boldsymbol{q}_{01J}^{\mathsf{T}} \\ -\boldsymbol{q}_{111}^{\mathsf{T}} & \cdots & -\boldsymbol{q}_{11J}^{\mathsf{T}} \\ -\boldsymbol{q}_{N11}^{\mathsf{T}} & \cdots & -\boldsymbol{q}_{N1J}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \boldsymbol{L}_1 \\ x_{21}\boldsymbol{L}_1 \\ \vdots \\ x_{2N}\boldsymbol{L}_1 \\ \boldsymbol{L}_2 \\ x_{11}\boldsymbol{L}_2 \\ \vdots \\ x_{1N}\boldsymbol{L}_2 \end{bmatrix} = \boldsymbol{0} \ . \qquad (11)$$

Ce produit matriciel est un système linéaire et homogène de $8(N+1)$ inconnues (pour rappel, N est le nombre de composantes principales pour le modèle empirique des CRF, dans notre travail, il est égal à 3). Ce système d'équations peut être résolu en utilisant, par exemple, la méthode de décomposition en valeurs singulières appliquée sur la matrice du modèle. Les inconnues sont, comme expliqué ci-dessus, estimées à l'échelle près seulement. Pour que la solution soit unique (à l'échelle près), les normales des points 3D doivent être suffisamment diversifiées, par exemple, si toutes les normales sont identiques, c'est–à–dire si la scène est plane, alors le système d'équations est mal contraint. Additionnellement, les valeurs de luminosité observées doivent être suffisamment diversifiées. Il est possible d'établir les conditions nécessaires pour le bien-posé du problème, mais cela est laissé pour les travaux à venir.

Une fois le système d'équations est résolu, les inconnues individuelles sont estimées comme suit. Les coefficients de l'éclairage $\boldsymbol{L}_i$ sont un résultat direct de la solution du système (à l'échelle). Les coefficients de la CRF x_{in} peuvent être facilement calculés : soit $\boldsymbol{X}_{in}$ la solution estimée pour $x_{in}\boldsymbol{L}_{3-i}$ $(i = 1, 2)$. Donc, la solution en termes des moindres carrés de x_{in} est :

$$x_{in} = \frac{\boldsymbol{L}_{3-i}^{\mathsf{T}}\boldsymbol{X}_{in}}{\boldsymbol{L}_{3-i}^{\mathsf{T}}\boldsymbol{L}_{3-i}} \ .$$

Même si les vecteurs $\boldsymbol{L}_i$ sont estimés à une échelle près, il n'existe aucune ambiguïté sur l'échelle des coefficients de la CRF x_{in}. Enfin, l'albédo ρ_j peut être calculé en utilisant l'équation (8) par exemple, en faisant la moyenne des estimations provenant des deux images.

4.2 Le Cas Vues Multiples

Dans le cas de vues multiples, si M images sont considérées, alors une solution linéaire pour toutes les inconnues est possible. Pour toutes les paires d'images on peut empiler les équations correspondantes (11) en formant ainsi un système fortement épars. Si on considère le cas le plus général, où toutes les paires d'images ont des points 3D visibles en commun, il existe en total $4M(1 + N(M-1))$ inconnues $\boldsymbol{L}_i$ et $x_{in}\boldsymbol{L}_i$. Si les J points 3D sont visibles dans toutes les images, il y a $JM(M-1)/2$ équations. Un certain nombre de $J \geq 8(N+1)$ points 3D est en général suffisant pour une solution unique à l'échelle près.

4.3 Optimisation non linéaire

Les méthodes linéaires exposées précédemment sont efficaces, dans le cas de plusieurs images uniquement si on n'utilise pas toutes les paires correspondantes. En présence de bruit, l'algorithme décrit n'est pas optimal, puisque la fonction de coût qui est minimisée (la norme de l'expression dans l'équation (6)) est algébrique et les inconnues estimées constituent une paramétrisation redondante des vraies inconnues du problème. Pour obtenir une solution optimale il faut appliquer une optimisation non linéale complète de la fonction de coût (6) (la fonction qui donne l'estimation du maximum de vraisemblance si les intensités recueillies sont affectées par un bruit gaussien).

Pour implémenter notre algorithme de façon efficace, il est nécessaire de calculer le Jacobien et le gradient de notre problème des moindres carrés. Ce calcul est trivial (on l'omet en ce document en raison du manque d'espace). Le Jacobien est une matrice hautement éparse. Étant donné que les CRF plausibles sont monotones, nous imposons également des contraintes d'inégalité dans le calcul des coefficients de la CRF. Pour cela, nous forçons à chaque valeur entre 1024 échantillons équidistants de l'irradiance à être plus grand que la valeur précédente. Selon l'utilisation ou non de ces contraintes, nous utilisons soit la fonction d'optimisation `fmincon` soit la fonction `lsqnonlin` disponibles en MATLAB. Ces fonctions profitent pleinement des avantages de l'utilisation d'un Jacobien épars.

Dans nos expériences, nous avons utilisé différentes méthodes pour initialiser les inconnues. Additionellement à l'utilisation de la méthode linéaire exposée dans la section précédente, nous avons également initialisé l'algorithme en utilisant la moyenne des fonctions linéaires du modèle empirique proposé par Grossberg et Nayar ($f_i = h_0$).

5 Résultats

Pour évaluer la précision du modèle proposé dans la section précédente, nous avons suivi des expériences sur des données synthétisées et aussi sur des images réelles. Pour le second cas, nous avons créé une base de données simulant les conditions d'images trouvées dans les collections provenant d'Internet. De plus, nous avons testé l'algorithme sur un ensemble d'images téléchargées à partir du dépôt d'images Flickr.

5.1 Données de Synthèse

Nous avons généré un ensemble de 250.000 "points virtuels 3D" en utilisant une large gamme d'albédos et de normales. Chaque élément de la surface est composé par les coordonnées du point 3D plus sa normale respective. 50 images (500x500 pixels) ont été produites. Pour chaque image, les points virtuels ont été exposés à différentes conditions d'illumination, cet–à–dire, différentes sources de lumière directionnelle et d'éclairage ambiant. Pour produire les valeurs d'intensité respectives, les CRF on été choisi au hasard entre les 201 CRF recueillies dans [8]. Une fonction différente a été utilisée pour chaque image. Sans bruit, la CRF originale et la CRF estimée sont identiques, confirmant la validité de nos méthodes. Ensuite, nous avons

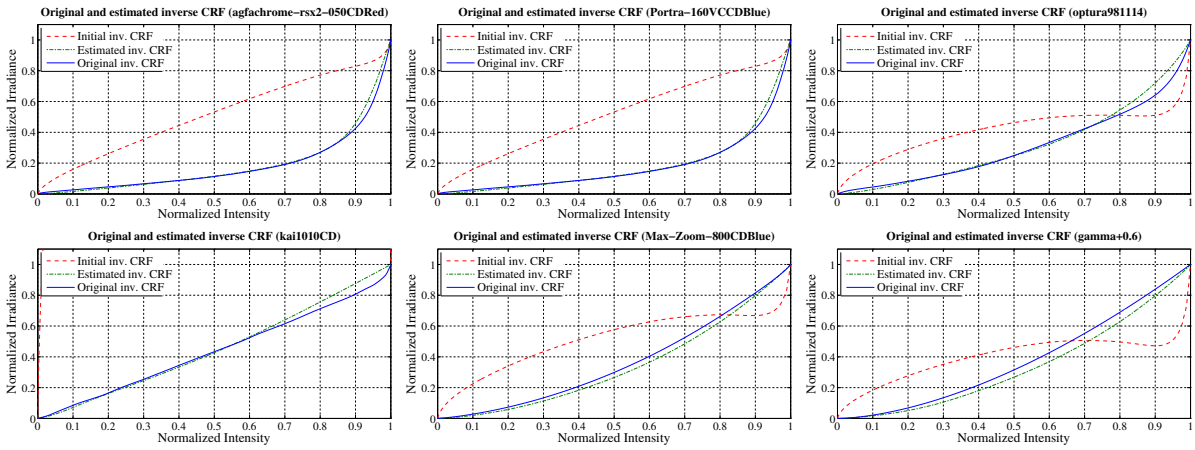

FIGURE 2 – Six inverse CRF estimées en utilisant 50 images synthétisées. Les pixels des images ont été perturbés par un bruit gaussien (*cf.* text). L'initialisation pour l'algorithme non–linéal est donné par la méthode décrite en section 4.2.

Bruit (Éc. Type. [%])	0.001	0.01	0.1	1
Erreur RMS pour la CRF	0.032	0.040	0.044	0.043
Erreru RMS pour l'albedo	0.073	0.098	0.112	0.138

TABLE 1 – Médiane des erreurs pour 50 images de synthèse.

ajouté un bruit gaussien sur les valeurs d'intensité produites, avec des écarts types entre $0,001$ et 1 pour cent et une valeur moyenne de 1 pour cent de la valeur maximum d'intensité. Nous avons mené 50 essais pour chaque niveau de bruit. La figure 2 montre les CRF originales et estimées pour 6 parmi les 50 images utilisées pour un essai typique. Les performances de l'approche linéaire sont relativement médiocres. Cependant, les résultats de l'optimisation non linéaire sont très satisfaisants, même lorsque l'algorithme est initialisé avec la solution de l'approche linéaire. Une évaluation quantitative est présentée dans le tableau 1, en utilisant différents valeurs de bruit. La qualité de la CRF estimée est mesurée en calculant la racine de l'erreur quadratique moyenne (RMS) entre la fonction originale et la fonction calculée : $\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(f(E_k)-\hat{f}(E_k)\right)^2}$ où les E_k correspondent à 1024 valeurs équidistantes entre 0 et 1.

5.2 Images réelles

Rarement les fabricants pourvurent des informations techniques sur le comportement des intensités fournies par le capteur et comment ces variables sont liées à l'irradiance de la scène. De plus, suite à notre définition de la CRF (une fonction dépendant de l'appareil photo et de la scène en même temps), la CRF comprend également des modifications induites par l'environnement, l'optique, la configuration de la caméra (balance des blancs, exposition, etc.), et d'autres processus internes de la caméra. En conséquence, une vérité terrain pour tester nos algorithmes n'est pas facilement disponible. Même s'il est difficile d'accéder à la vérité terrain, il est possible d'obtenir des approximations. Une méthode consiste à utiliser une charte de couleurs avec des échantillons d'irradiances connues sous certaines circonstances. Cette méthode est la plus employée dans l'étalonnage radiométrique. Cependant, il existe une controverse sur la bonne utilisation de cet outil, étant donné que différents auteurs réfutent les valeurs fournies par le fabricant de la charte (*cf.* [20]). En plus, afin d'obtenir les valeurs correctes dans l'espace couleur souhaité, un illuminant spécifique doit être établi. Dans notre cas, nous utilisons la charte *ColorChecker GretagMacbeth* et les valeurs CIE $L^*a^*b^*$ fournies par le fabricant. Nous réalisons une transformation entre l'espace $L^*a^*b^*$ et l'espace sRGB en utilisant la balance des blancs enregistrée dans les informations EXIF de chaque image, si disponible, sinon nous utilisons l'illuminant D65.

Pour valider notre algorithme dans des conditions réelles, nous avons compilé deux bases de données et implémenté d'autres méthodes pour la comparaison. La première base de données est composée par des images recueillies à partir d'une collection de photos sur Internet ; une seconde a été construit pour nos soins, et comprends des images acquises par nous afin d'avoir accès à la scène pour obtenir une approximation à la vérité terrain. Cette collection d'images est composée de 113 images d'une scène d'extérieur. Cinq appareils photo différents ont été utilisés. Différentes conditions d'illumination sont reflétées dans les photos. Ces images ont été prises en utilisant une configuration complètement automatique dans toutes les caméras (balance de blancs, temps d'exposition, ouverture, etc.). En même temps, pour certaines positions particulières, nous avons également pris plusieurs images avec des expositions différentes, en tenant l'appareil photo statique. Dans ces scénarios, nous avons aussi inclu la charte de calibrage, juste quelques secondes après les premières prises. Entre la prise de l'image originale et l'image contenant la charte de couleurs, nous avons conservé, dans la mesure du possible, les mêmes réglages pour l'appareil photo. Nous sommes conscients que l'inclusion de la charte de couleurs peut changer les paramètres de la caméra lors du tournage en mode automatique (par exemple, la balance des blancs), alors, lorsque possible, on a réglé l'appareil photo afin d'utiliser les mêmes paramètres que lors de la prise initiale.

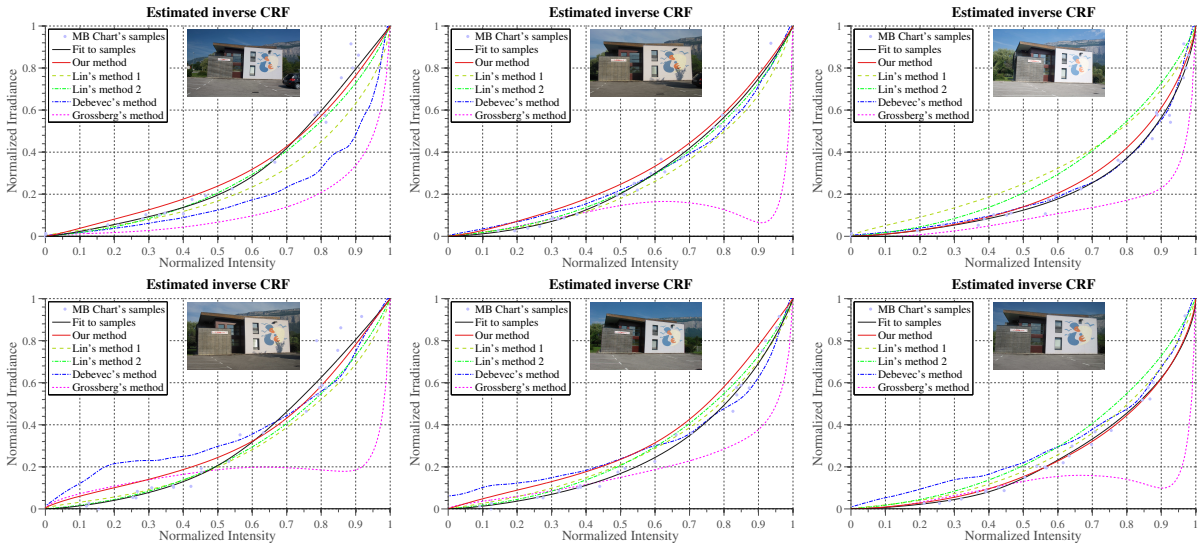

FIGURE 3 – CRF estimées par l'algorithme décrit dans la section 4.3. Ces résultats son comparés avec : la CRF mesurée à partir des échantillons de la charte de couleurs. La méthode de Lin *et al.* avec deux valeurs différents pour le paramètre $\lambda = \{10^{-1}, 10^{-4}\}$ (*cf.* [14]). Les résultats du logiciel HDRShop [4]. L'algorithme d'exposition multiples proposé par [8]. Des contraintes sur la monotonicité de la CRF estimée sont imposées.

Les 24 échantillons prélevés de la charte de couleurs ont été interpolés en utilisant la technique décrite dans [8]. Ces résultats sont alors considérés comme notre vérité terrain. En utilisant les images acquises avec plusieurs expositions, nous sommes également en mesure de calculer une estimation de la CRF en utilisant le logiciel *HDRShop* [4] et la méthode d'expositions multiples décrite par Grossberg [8]. Nous avons aussi implémenté notre propre version de l'algorithme proposé par Lin *et al.* [14] (voir section 2).

Nous avons testé notre algorithme avec 22.021 points 3D obtenus après le processus de reconstruction. Les résultats pour 6 images sont présentés dans la figure 3. Notre estimation est la plupart du temps très proche de la courbe ajustée aux échantillons de la charte de couleurs. La méthode de Lin [14] présente également de bonnes estimations, mais les meilleurs résultats de cet algorithme ne sont pas toujours obtenus avec la valeur recommandée par les auteurs pour le paramètre de régularisation impliqué ($\lambda = 10^{-4}$) , voir par exemple, le graphique en bas à droite de la figure 3. Le graphique en haut à droite de la même figure montre un exemple où les résultats de Lin *et al.* sont relativement loin de la vérité terrain. Les méthodes d'expositions multiples présentent des courbes irrégulières dans certaines estimations. Cela peut s'expliquer par le fait qu'il est presque impossible d'obtenir des images parfaitement alignées lors des prises d'images à l'extérieur, même si le délai entre chaque prise est moins d'une seconde. Les ombres, les reflets et les conditions peuvent changer rapidement. Le succès de ces méthodes dépend fortement d'un alignement parfait entre les pixels de toutes les images.

Pour valider notre approche nous avons utilisé une deuxième base de données composée de 958 images récoltées d'Internet. Une fois que nous avons reconstitué le modèle 3D de la scène, 21.612 points 3D sont disponibles. Nous avons lancé notre algorithme sur une sélection aléatoire de 103 caméras (nous n'utilisons pas toutes les images afin de réduire le temps de traitement). Dans ce cas, il n'est pas possible d'insérer la charte de couleurs ni d'obtenir des expositions multiples avec une caméra fixe, alors nous comparons nos résultats avec les estimations calculées par l'algorithme de Lin. Dans la figure 4, quatre CRF inverses estimées sont affichées. Même si la méthode de Lin est fortement dépendante du terme de régularisation déjà mentionné, cette approche est à ce jour l'une des méthodes les plus fiables pour estimer la CRF en utilisant une seule image. La solution calculée par notre algorithme est la plupart des fois très proche des résultats de l'algorithme de Lin.

Erreur RMS/Base de donnés	Canal R	Canal G	Canal B
1. Base de donnés propre	0.0431	0.0528	0.0463
2. Base de donnés Internet	0.0627	0.0716	0.0679

TABLE 2 – Moyenne des erreurs RMS pour le deux bases de donnés décrites dans la section 5.2. Dans le deux cas, l'erreur RMS est calculé par rapport a l'estimation de Lin *et al.*

6 Discussion et Conclusions

Nous avons présenté une approche destinée à récupérer les fonctions de réponse radiométrique des caméras, en utilisant un ensemble d'images extraites d'une collection de photos. L'approche exploite le modèle 3D généré par des méthodes disponibles et un modèle empirique pour paramétrer les CRF, sans lequel le problème serait mal posé.

Par rapport à d'autres méthodes, notre approche ne nécessite pas que plusieurs images soient alignées ou prises sous les mêmes conditions d'illumination. Notre méthode est envisagée comme une première étape dans le processus de récupération des propriétés photométriques pour une

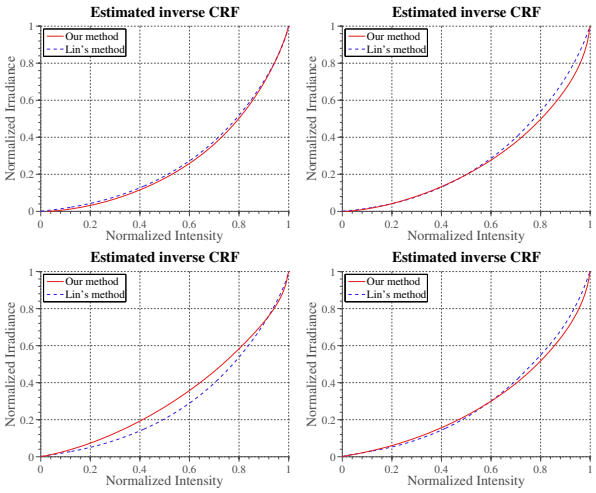

FIGURE 4 – Estimation de la CRF inverse en utilisant notre méthode et une collection d'images extraites d'Internet. Ces résultats sont comparés avec notre propre implémentation de la méthode de Lin *et al.* [14].

scène donnée, y compris, les propriétés de reflectance de la surface et l'illumination de la scène. Dans ce contexte, il n'est pas convenable de comparer nos résultats avec ceux d'autres méthodes qui n'utilisent qu'une seule image ([14, 15, 17, 26, 12, 25]). Néanmoins, nos résultats sont aussi bons (ou meilleurs) que ceux estimés par ces algorithmes. De plus, notre méthode permet d'estimer toutes les CRF pour un groupe d'images en même temps, de façon automatique, tandis que, selon notre expérience, les méthodes d'image unique nécessitent la plupart du temps un certain contrôle ou un réglage manuel des paramètres, en fonction de l'image d'entrée. Pour le cas des collections d'images, l'obtention des CRF avec ces méthodes pourrait être une tâche extrêmement lente et fastidieuse. Dans ce contexte, nos résultats sont prometteurs.

Les limitations de notre approche sont les suivantes. Le modèle d'éclairage, constitué d'une seule source de lumière directionnelle et l'éclairage ambiant, n'est certainement pas tout à fait réaliste. Des modèles plus généraux peuvent en principe être utilisés. Par exemple un modèle basé en harmoniques sphériques [1, 21] représente d'une façon plus réaliste l'illumination, mais il reste à voir si son estimation est bien posée. Une autre limitation est l'hypothèse de la réflectance lambertienne, bien que cela est relativement facile de contourner en utilisant une fonction de pondération robuste dans l'optimisation non linéaire. Cette fonction est appelée de façon itérative afin d'exclure les grossières erreurs pendant l'optimisation aux moindres carrés, par exemple. Actuellement, nous ne gérons pas les ombres associées à la source de lumière. Cela pourrait également être pris en compte dans une certaine mesure par une fonction de pondération robuste. Il est aussi envisageable de gérer les ombres explicitement depuis la géométrie de la scène et l'estimation actuelle de la position de la source de lumière. Les ombres peuvent être testées et utilisées pour prédire correctement les valeurs de l'intensité.

Références

[1] Ronen Basri, David Jacobs, *Lambertian reflectance and linear subspaces*, IEEE Trans. on PAMI, 2003.

[2] A Chakrabarti, D Scharstein, Todd Zickler, *An empirical camera model for internet color vision*, BMVC, 2009.

[3] Y.-C. Chang, J.F. Reid, *RGB calibration for color image analysis in machine vision*, IEEE Trans. on Image Processing (1996), 1414–1422.

[4] Paul Debevec, Jitendra Malik, *Recovering high dynamic range radiance maps from photographs*, SIGGRAPH, 1997.

[5] G.D. Finlayson, S.D. Hordley, C. Lu, M.S. Drew, *On the removal of shadows from images*, IEEE Trans. on PAMI **28**, 2006, 59–68.

[6] Yasutaka Furukawa, Jean Ponce, *Accurate, dense, and robust multi-view stereopsis*, IEEE Trans. on PAMI, 2009.

[7] _____, *Patch-based multi-view stereo software*, Web, 2009, http ://grail.cs.washington.edu/software/pmvs.

[8] Michael D Grossberg, Shree K Nayar, *What is the space of camera response functions ?*, CVPR,2003, 602.

[9] _____, *Modeling the space of camera response functions*, IEEE Trans. on PAMI **26**, 2004, 1272 – 1282.

[10] Tom Haber, Christian Fuchs, Philippe Bekaer, Hans-Peter Seidel, Michael Goesele, Hendrik Lensch, *Relighting objects from image collections*, CVPR, 2009, 627 – 634.

[11] A. Ilie, G. Welch, *Ensuring color consistency across multiple cameras*, ICCV,2005, 1268–1275.

[12] Seon Joo Kim, J Frahm, M Pollefeys, *Radiometric calibration with illumination change for outdoor scene analysis*, CVPR, 2008.

[13] Sujit Kuthirummal, Aseem Agarwala, Dab Goldman, Shree K Nayar, *Priors for large photo collections and what they reveal about cameras*, ECCV (2008), 74–87.

[14] S Lin, Jinwei Gu, S Yamazaki, Heung-Yeung Shum ;, *Radiometric calibration from a single image*, CVPR,2004, 938 – 945.

[15] S Lin, L Zhang, *Determining the radiometric response function from a single grayscale image*, CVPR, 2005, 66 – 73.

[16] Q Luong, P Fua, Y Leclerc, *The radiometry of multiple images*, IEEE Trans. on PAMI **24** ,2002, 19 – 33.

[17] Y Matsushita, S Lin, *Radiometric calibration from noise distributions*, CVPR, 2007.

[18] Toomo Mitsunaga, Shree K Nayar, *Radiometric self calibration*, CVPR 1999.

[19] Tian-Tsong Ng, Shih-Fu Chang, Mao-Pei Tsui ;, *Using geometry invariants for camera response function estimation*, CVPR,2007.

[20] Danny Pascale, *Rgb coordinates of the macbeth colorchecker*, Tech. report, The BabelColor Company, Jun 2006.

[21] R. Ramamoorthi, P. Hanrahan, *A signal-processing framework for inverse rendering*, SIGGRAPH, 2001, 117–128.

[22] Fabiano Romeiro, Todd Zickler, *Blind reflectometry*, ECCV, 2010, 45–58.

[23] Noah Snavely, Steven M Seitz, Richard Szeliski, *Photo tourism : exploring photo collections in 3D*, SIGGRAPH, 2006.

[24] _____, *Modeling the world from internet photo collections*, IJCV, 2008.

[25] J Takamatsu, Y Matsushita, K Ikeuchi, *Estimating camera response functions using probabilistic intensity similarity*, CVPR, 2008.

[26] B Wilburn, Hui Xu, Y Matsushita, *Radiometric calibration using temporal irradiance mixtures*, CVPR, 2008.