HAL open science

Radiometric Calibration using Photo Collections

Mauricio Díaz, Peter Sturm

To cite this version:

Mauricio Díaz, Peter Sturm. Radiometric Calibration using Photo Collections. ICCP 2011 - IEEE International Conference on Computational Photography, IEEE, Carnegie Mellon University, Apr 2011, Pittsburgh, United States. pp.1-8, 10.1109/ICCPHOT.2011.5753117 . inria-00590275

HAL Id: inria-00590275

https://inria.hal.science/inria-00590275v1

Submitted on 9 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.

Radiometric Calibration using Photo Collections

Mauricio Díaz
Laboratoire Jean Kuntzmann
38400 Saint Martin d'Hères, France
Mauricio.Diaz@inrialpes.fr

Peter Sturm
INRIA Grenoble Rhône-Alpes
38330 Montbonnot, France
Peter.Sturm@inrialpes.fr

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.

1. Introduction

The work presented in this paper aims at recovering radiometric camera calibration from multiple images acquired in uncontrolled conditions. It is motivated by the goal of exploiting the richness of appearance variations captured by images in large community image collections. During the last years, the use of these large databases in computer vision has been a widely disseminated research subject and important advances have been obtained. For example, geometric reconstruction using an unconstrained and unstructured set of images gathered from the Internet is in these days a doable task. In a seminal work, Snavely *et al*. [23] proposed to use structure from motion (SFM) techniques to recover positions of cameras having acquired images of some object of interest. This algorithm constitutes the backbone of the PhotoSynth system[1] whose goal is to offer intuitive, pleasant and efficient means of browsing image collections. In the meanwhile, other projects have centered their efforts in solving the problem of robust 3D reconstruction for uncontrolled environments, using multiple views, see *e.g*. [6]. Combining these results, *i.e*. robust SFM for camera position estimation and a solid multiview reconstruction, nowadays it is possible to generate dense 3D models of an object, *e.g*. a major monument, using images captured under completely casual and natural conditions. Having this geometric model is a starting point to deduce photometric properties of the scene.

Images uploaded by users to large photo collections exhibit a wide range of appearances. The objects that compose the scene are captured under different illumination conditions, for example, images taken at different times of the day (or the year) with different weather conditions. This opulence in appearance information is the factor we want to exploit. Modeling the appearance of the objects can not be achieved directly from the images. Images are the result of an interplay of different factors, scene geometry, surface reflectance, illumination conditions and the way cameras acquire and process images. By the latter, we refer in this paper to the camera response function (CRF), although other aspects also exist, such as depth of field or lens aberrations. In order to access surface reflectance information, it seems essential to know the CRF. However, this is usually not the case for images in community image collections. Further, the illumination conditions are in general unknown.

We can formulate our problem as follows: for some given uncalibrated images that correspond to a particular scene and a 3D model of the scene (in our work, obtained using Bundler [24] and PMVS [7]), we wish to determine the CRF's of all images, even if they are all different. This problem is in general highly ill-posed: even if we perfectly knew the illumination conditions for every image, one could, for every hypothetical set of CRF's, define a BRDF (bidirectional reflectance distribution function) per surface point that is consistent with its observa-

[1] http://photosynth.net/ Consultation date:15/02/2010

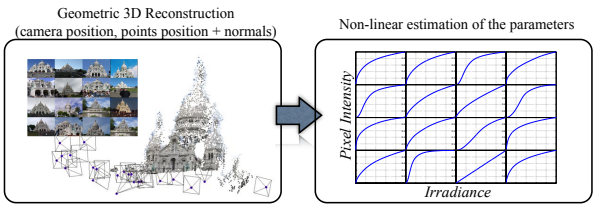

Figure 1. From an unordered set of images collected from the Internet and a 3D model from the scene, radiometric properties of the surface and camera response functions are estimated.

tions in all images. Clearly, in order to solve the problem, priors are needed. They could concern the illumination, surface reflectance as well as the CRF's. In this paper, we use a realistic strong prior on CRF's, together with priors on reflectance and illumination. Grossberg and Nayar have radiometrically calibrated 201 different camera-lens-combinations and performed a principal component analysis over the obtained CRF's [8]; this revealed that all these functions can be very well modeled as linear combinations of a few basis function (3 are often sufficient). We use this as prior on the CRF's to be estimated in our problem formulation. However, without additional priors on reflectance and illumination, the problem would still be intractable. In this paper, we use simple priors for these: we assume Lambertian surfaces (but with spatially varying albedo) and illumination conditions modeled by the combination of a single directional light source and ambient lighting.

2. Related Work

In recent years, the computer vision community has focused its attention on the use of public photo collections to get information about the geometry of the scene. After the promising results of [24], where a sparse geometry of the scene and the cameras is recovered, other works have explored different aspects. Haber *et al*. [10] have developed an approach to recover reflectance properties and illumination using a wavelet framework. In this work, authors assume that images extracted from photo collections can be photometrically corrected by mapping with a gamma curve. It has been widely proven that camera response functions of digital cameras are often more complex function.

There exist different approaches for radiometric camera calibration. Two categories can be distinguished: active and passive methods. The active recovery of CRF's gathers all the methods that require physical presence during the acquisition process. Common approaches include using either specific calibration objects (color charts such as the popular GretagMacbeth ColorChecker, see http://www.xrite.com) [3, 11] or images acquired by a static camera but with varying exposure times [4, 18]. Naturally, these approaches are not applicable if the goal is the radiometric calibration of images in a community image collection. A second category includes approaches where the physical presence on the scene is not necessary. This kind of algorithm seeks image characteristics that reflect the non-linearity produced by the camera response function. Low level representations as edges in regions where color changes drastically [14], [15] or geometric invariants [19] allow to extrapolate an estimation of the CRF. Other artifacts common in digital images as noise [25] can also be used to estimate the radiometric response of a camera. In [13], Kuthirummal *et al*. found priors for statistics on large photo collections. These priors allow them to estimate the response functions of generic camera models, assuming that all instances of one camera model have the same properties and that many images taken by the camera to be calibrated, are available. Personalized devices, *e.g*., cameras with interchangeable lenses, can not be modeled using this method.

Most of the methods above mentioned uses the empirical model proposed by Grossberg and Nayar [9], based on the principal component analysis mentioned in section 1. Recently, a different camera model is proposed in [2]. In this work authors propose to model the CRF as the product of a white-balance transform matrix and a polynomial of fifth degree. This approach provides a model using mixed color information but increases the number of parameters for the CRF model (24-parameter according the authors). The space of possible CRFs spanned is dramatically large, and there is no guarantee that the estimated CRF correspond to a real one.

Luong *et al*. [16] use, like our approach, a 3D model of the imaged scene, to perform radiometric calibration and illumination estimation. However, they used a linear model for the CRF's and required that several images be taken with the same camera, in controlled conditions. In the case of images downloaded from the Internet, a linear model for the CRF's is not accurate and acquisition conditions are completely uncontrolled.

In our work, we estimate non-linear CRF's from images acquired in uncontrolled conditions; further, it is not required that more than one image be available per camera. As explained in the introduction, radiometric calibration is ill-posed in this case. The priors we use concern the reflectance properties of the scene (Lambertian), the illumination (combination of a directional light source and ambient lighting), and the CRF's to be calibrated.

3. Problem statement

As explained above, the goal of the methods presented in this paper is the radiometric calibration of images in an unstructured image collection of the same scene. Concretely, we estimate one non-linear camera response function (CRF) per image and channel, as well as lighting conditions (also per image and channel), consisting of a directional light

source and ambient lighting relative to the wavelength of the considered channel. (Direction of the light source is the same for all channels, but its "strength" can vary, as well as the ambient term). The case where it is known that two images have the same CRF and/or the same lighting conditions, is a straightforward extension.

Input. The only primary input are the images themselves and the priors we use (see below). As a pre-processing for our methods, we first obtain the geometric calibration of all images and a 3D scene model. Geometric calibration is performed using the BUNDLER software by Snavely *et al*. [24] and the 3D model is computed using the PMVS software by Furukawa *et al*. [7]. As 3D model, we use a collection of surface elements (3D points plus normals) – meshes or other dense models could be used in the same manner. It is necessary to handle visibility of the surface elements; this is also estimated by the PMVS software. The complete input to our methods are thus a set of M images and their geometric calibration, as well as a collection of surface elements and information on their visibility in the images.

Image formation. The image formation model we use consists of the assumption of Lambertian reflectance of the surface elements and an illumination consisting of one directional light source and ambient lighting. The irradiance E of a surface element j under the lighting conditions of image i is thus expressed by:

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

where,

$$\mathbf{l_i} = \begin{bmatrix}\beta_i \cos\theta_i \cos\phi_i & \beta_i \sin\theta_i \cos\phi_i & \beta_i \sin\phi_i\end{bmatrix}^{\mathrm{T}} \ ,$$

corresponds to the direction (θ_i,ϕ_i) and strength (β_i) of the directional source light for image i expressed in global spherical coordinates (not a unit vector), ρ_j is the albedo of that surface element, $\mathbf{n}_j$ its normal (a unit vector) and μ_i is an ambient lighting term. Following the notations by Luong *et al*. [16], we can express equation (1) using a dot product of the vectors $\mathbf{L}_i = \begin{bmatrix}\mathbf{l_i} & \mu_i\end{bmatrix}^{\mathrm{T}}$ and $\mathbf{N}_j = \begin{bmatrix}n_{j1} & n_{j2} & n_{j3} & 1\end{bmatrix}^{\mathrm{T}}$. Then, the brightness value of a surface element j, observed in image i with CRF f_i, is given by:

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

where β_i, ρ_j, μ_i and f_i are defined by channel.

Prior on CRF's. Grossberg and Nayar [9] have found that CRF's belonging to real world cameras lie in a small part of a theoretical function space that can be modeled using a small empirical basis. This result allows to express the CRF's in terms of few parameters, three seeming to be sufficient [9]. This claim relies on two important assumptions: the response function f is the same for all the pixels that belong to a particular image and the function is monotonic. The latter is plausible in practice whereas the former may have to be addressed in future work, in order to handle potential residual vignetting effects in images. Additionally, we assume that the spectral sensitivity has similar ranges for all cameras.

Using Grossberg and Nayar's results, the relation between image brightness B and irradiance E can be expressed as a linear combination of an average CRF h_0 and principal components h_n:

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

The average CRF and the principal components can for example be represented as lookup tables or polynomials fitted to these. Without loss of generality, we use the latter in our work. The main difference between the two representations is that in order to impose monotonicity on the estimated CRF's f_i, fewer inequality constraints are required in the case of polynomials, compared to lookup tables. Besides this issue, the formulas underlying our methods would be strictly analogous in the case of lookup tables.

The basis CRF's are thus represented as polynomials of degree D:

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

Note that according to [9], the h_n are expressed relative to normalized brightness and irradiance values, such that $c_{n0} = 0$ and $\sum c_{nd} = 0$ for all $n = 1 \cdots N$. The degree of the polynomials was chosen for an adequate representation of the curves that form the basis, which was obtained with $D = 9$. These polynomials are known; what is unknown is the coefficients w_{in} of their linear combination (3).

Complete problem formulation. For a given channel, the unknowns are $\mathbf{\Theta} = \{\mathbf{L}_i, \rho_j, w_{in}\}$, with one set of illumination coefficients $\mathbf{L}_i$ per image (the unknowns β_i and μ_i composing the vector $\mathbf{L}_i$ are computed per image channel, while the direction of the light source, θ_i and ϕ_i, is shared by all the image's channels), one albedo ρ_j per 3D point and one set of CRF coefficients w_{in} per image. With the above models, the predicted brightness values, given estimates of the unknowns, are:

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

We aim at estimating the unknowns by minimizing the difference between observed and predicted brightness values:

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

The scalars v_{ij} are booleans, a value of 1 indicating that surface element j is visible in image i, otherwise the value being 0. Note that the unknowns can not be estimated without ambiguity: albedos ρ_j and lighting coefficients $\mathbf{L}_i$ can only be estimated up to one global scale factor. In addition to the cost function (6), we may also impose constraints, such as on the monotonicity of the estimated CRF's.

4. Proposed Solutions

The problem stated in equation (6) is of the non-linear least squares type. The full non-linear optimization is described in section 4.3. In the following sections, we propose linear least squares solutions for a modified cost function, which may be used to initialize the non-linear optimization. We first address the minimal case of two images, showing that this is in theory already sufficient to solve the problem. This is then generalized to multiples images.

4.1. Two Images

To turn our problem into a linear one, we consider the inverse CRF's: $g_i(B) = f_i^{-1}(E)$. According to Grossberg and Nayar [8], the inverse CRF's can also be represented as a sum of principal components of the form (3):

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

We will estimate the coefficients x_{in} defining the inverse CRF's, after which the direct CRF's can be estimated by fitting to sample values, see below.

From equation (2), we have:

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

If surface element j is visible in both images, then we have:

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

Thus:

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

In full detail:

$$\left(p_0(B_{1j}) \mathbf{N}_j^{\mathrm{T}}\right) \mathbf{L}_2 + \sum_{n=1}^{N} \left(p_n(B_{1j}) \mathbf{N}_j^{\mathrm{T}}\right) (x_{1n} \mathbf{L}_2) =$$
$$\left(p_0(B_{2j}) \mathbf{N}_j^{\mathrm{T}}\right) \mathbf{L}_1 + \sum_{n=1}^{N} \left(p_n(B_{2j}) \mathbf{N}_j^{\mathrm{T}}\right) (x_{2n} \mathbf{L}_1) \quad , \tag{10}$$

where we grouped known and unknown entities respectively. This equation is linear in the following unknowns: $\mathbf{L}_1, \mathbf{L}_2, x_{1n}\mathbf{L}_2, x_{2n}\mathbf{L}_1$, for $n = 1 \cdots N$. Let $\mathbf{q}_{nij} = p_n(B_{ij}) \mathbf{N}_j$. Then, if we consider J surface elements, the associated equations (10) can be grouped together in the following matrix equation:

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

This is a linear homogeneous equation system in $8(N+1)$ unknowns (as a reminder, N is the number of principal components in the prior on CRF's, in our work it is equal to 3). It can be solved to least squares using *e.g.* a singular value decomposition of the design matrix. The unknowns are, as explained above, estimated up to scale only. In order for the solution to be unique (up to scale), the normals of the available surface elements need to be sufficiently diverse; for example, if all normals are identical, *i.e.* if the scene is planar, then the equation system is underconstrained. Also, the observed brightness values need to be sufficiently diverse. It may be possible to establish necessary conditions for the well-posedness of the problem, but this is left for future work.

Once the equation system is solved, the individual unknowns are estimated as follows. The lighting coefficients $\mathbf{L}_i$ are directly given (up to scale) by the solution of the system. The CRF coefficients x_{in} can be easily computed: let $\mathbf{X}_{in}$ be the estimated solution of $x_{in}\mathbf{L}_{3-i}$ ($i = 1, 2$). Then, the least squares solution of x_{in} is

$$x_{in} = \frac{\mathbf{L}_{3-i}^{\mathrm{T}} \mathbf{X}_{in}}{\mathbf{L}_{3-i}^{\mathrm{T}} \mathbf{L}_{3-i}} \quad .$$

Note that although the $\mathbf{L}_i$ are estimated up to scale only, there is no scale ambiguity on the CRF coefficients x_{in}. Finally, the albedos ρ_j can be estimated using equation (8), *e.g.* by averaging the estimates coming from the two images.

4.2. Multiple Images

If M images are considered, a linear solution of all unknowns is possible by simply stacking equations (11) for all pairs of images on top of each other, giving a large but highly sparse equation system. In the most general case, where all pairs of images have visible scene elements in common, there are a total of $4M(1+N(M-1))$ unknowns $\mathbf{L}_i$ and $x_{in}\mathbf{L}_i$. If all J surface elements are visible in all images, then there are $JM(M-1)/2$ equations. A number of $J \geq 8(N+1)$ surface elements is in general sufficient for a unique solution up to scale (a tighter bound on J is possible but complicated).

This approach becomes intractable if M becomes of the order of several hundreds or larger. In that case, we only use equations linking one reference image with all other images.

4.3. Non-linear Optimization

The above linear methods are efficient (in the case of many images only if not all equations are used, however). In the presence of noise they are suboptimal, since the cost function that is minimized (the norm of the expression in equation (11)) is algebraic and since the estimated unknowns constitute a redundant parameterization of the actual unknowns of the problem. An optimal solution thus requires the full non-linear optimization of the meaningful cost function (6) (the function gives the maximum likelihood estimate if observed brightness values are affected by i.i.d. Gaussian noise).

The Jacobian and gradient of the least squares problem (6) are trivial to compute analytically (not given due to lack of space) and the Jacobian is highly sparse. Since plausible CRF's are monotonic, we also impose inequality constraints on the CRF coefficients, as follows: for 1024 equidistant irradiance values between 0 and 1, we enforce that the CRF is larger for each sampled irradiance value than for the next smaller one. Depending on the use or not of these constraints, we use either the `fmincon` or the `lsqnonlin` optimization functions of MATLAB, which make full use of the sparsity of the Jacobian.

In our experiments, we used different methods to initialize the unknowns. Besides using the linear method of the previous section, we also initialized the CRF's as linear functions $f_i(E) = E$ or as the average CRF of Grossberg and Nayar's analysis: $f_i = h_0$.

If the linear method of the previous section is used for initialization, we first have to estimate the coefficients of the direct CRF's, from those of the estimated inverse CRF's. To do so, we simply generate sample pairs of brightness and irradiance values from the computed inverse CRF's g_i, for a dense equidistant sampling of the range of brightness values. For each image, we then fit a direct CRF to these samples, by estimating the coefficients w_{in} of (3).

Noise (St. Dev. [%])	0.001	0.01	0.1	1
RMS Error inverse CRF	0.032	0.040	0.044	0.043
RMS Error for albedo	0.073	0.098	0.112	0.138

Table 1. Median RMS error for 50 synthesized images

5. Experiments

To evaluate the accuracy of the model proposed in the previous section, we ran experiments on synthesized data and real images. For the second case, we created a database simulating conditions of images found in Internet collections. Also, we tested the algorithm on a set of images downloaded from the image repository Flickr[2].

5.1. Synthetic data

We generated a set of 250,000 "virtual" 3D points using a large range of albedos and normals. Each surface element is composed of the position of the 3D point plus its normal. 50 images (500x500 pixels) were produced. For each image, surface elements were exposed to different illuminations, *i.e.* different directional light sources and ambient lighting of different "strengths". To produce the respective brightness values, a randomly selected CRF was used for every image. They were selected among CRF's following equation (3), fitted to the data of 201 real CRF's, collected in [8]. Without noise, the recovered CRF's were identical to the original functions, confirming the validity of our methods. We then added Gaussian noise to the generated brightness values, with standard deviations between 0.001 and 1 percent and a mean value of 1 percent of the range of brightness values. We ran 50 trials for each noise level. Figure 2 shows estimated and original inverse CRF's for 6 of the 50 cameras in a typical trial. The linear approach performs relatively poorly. The non-linear optimization however, even when initialized from the solution of the linear approach, leads to good results. A quantitative assessment is presented in table 1, for a varying noise level. The quality of an estimated CRF is measured by computing the following root mean square error (RMS) between the true and estimated inverse CRF: $\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(f(E_k)-\hat{f}(E_k)\right)^2}$ where the E_k are taken as the $K = 1024$ equidistant values between 0 and 1.

5.2. Real images

Manufacturers rarely provide technical information about the behavior of the intensities captured and how they are related to the irradiances of the scene. Moreover, following our definition of CRF (a camera-scene dependent function), it also includes changes induced by the environment, the optics, the camera set-up (white balance, exposure

[2] www.flicker.com

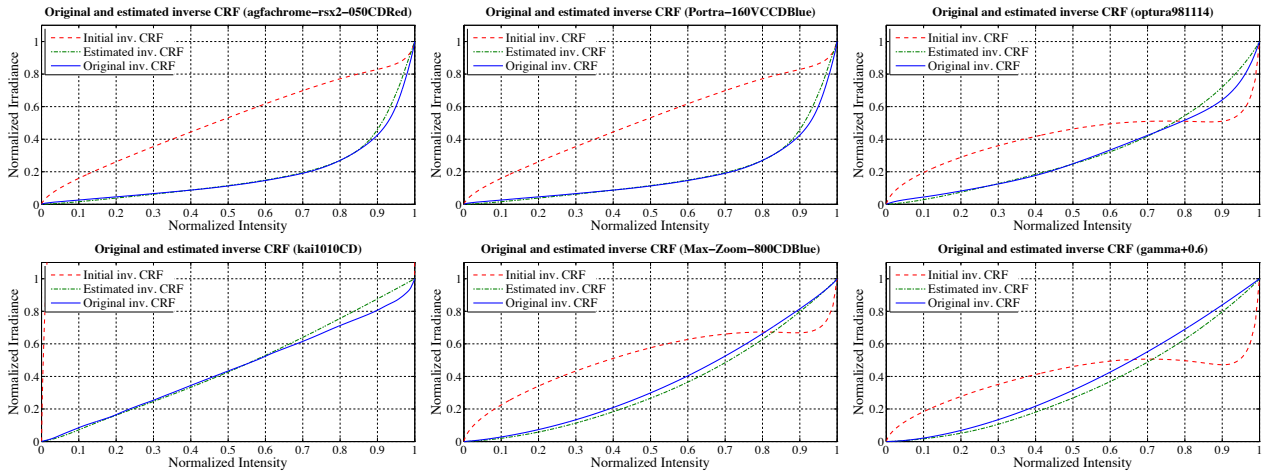

Figure 2. Six estimated inverse CRF's using 50 synthesized images perturbed by Gaussian noise of standard deviation 0.01 and mean 0.1 ($c.f.$ text). The initialization for the non-linear algorithm is given by the method of section 4.2. Non-linear estimation was taking into account monotonicity constraints on the CRF's.

value, etc.), and internal processing of the data. As a consequence, a ground truth to test our algorithms is not readily available. Approximations to the ground truth can be obtained using samples of a color calibration chart, however controversy and discussion around the right RGB values for the chart's patches have been raised (see [20]). Also, in order to get the correct values in the desired color space, a specified illuminant should be established. In our case, we use the GretagMacbeth ColorChecker and the CIE $L^*a^*b^*$ values provided by the manufacturer transformed to an sRGB space using the white balance point stored in the EXIF information of the cameras, when available, otherwise we use the illuminant D65.

To validate our algorithm using real world conditions we compiled two databases and ran several other methods for comparison. One database was gathered from an Internet photo collection; a second one was built by images acquired by us in order to allow for a comparison to a ground truth. As for the latter database, we took 113 photos of an outdoor scene with a set of 5 different cameras and changing illumination conditions (to see some samples of these images we invite the reader to explore the supplemental material). These images were taken using an automatic setup for the cameras (white balance, aperture, exposition time, etc.). At the same time, for some particular positions, we also took multiple images with different exposures, without moving the camera. In those cases, just some seconds after the shots, the GretagMacbeth ColorChecker was also included in the scene, keeping, when possible, the camera settings locked between the original shot and the image containing the color chart. We are aware that the inclusion of the color chart could change camera parameters when shooting in automatic mode ($e.g.$ white balance) then, when possible we change the mode to manual and we use the settings determined at the initial shoot.

The 24 samples extracted from the ColorChecker were interpolated using the technique reported in [8]. These results are then considered as our ground truth for the first experiment. Using the multiple exposures images, we are also able to compute the estimated CRF using the *HDRShop* software [4] and the multiple exposures method described by Grossberg [8]. CRF estimation using our own implementation (with help by the authors) of Lin's *et al.* algorithm [14] is also calculated. We ran our algorithm on a set of 21,021 surface elements given by the 3D reconstruction process. Results for 6 image-camera pairs are shown in figure 3. Our estimation lies most of the time very close to the curve fitted to the color checker samples. Lin's method also exhibits good results, however best results are not always obtained with the recommended value of 10^{-4} for the involved regularization parameter λ [14], see $e.g.$ the lower right graph of figure 3. The upper right graph shows an example where Lin *et al.*'s approach gives a result relatively far from the ground truth. Multiple exposure methods present irregular curves in some estimations. This can be explained because it is almost impossible to get perfectly aligned images when shooting outdoor scenes, even if the time between each shot is shorter than one second. Shadows, reflections and other artifacts can change rapidly, and success of these methods depends on perfect pixel alignment.

We tested the algorithm using a second database composed of 958 images harvested from the Internet. Once we reconstructed the 3D model of the scene, 21,612 surface elements were available. We extracted a random selection of 103 cameras and ran our algorithm (we do not use all im-

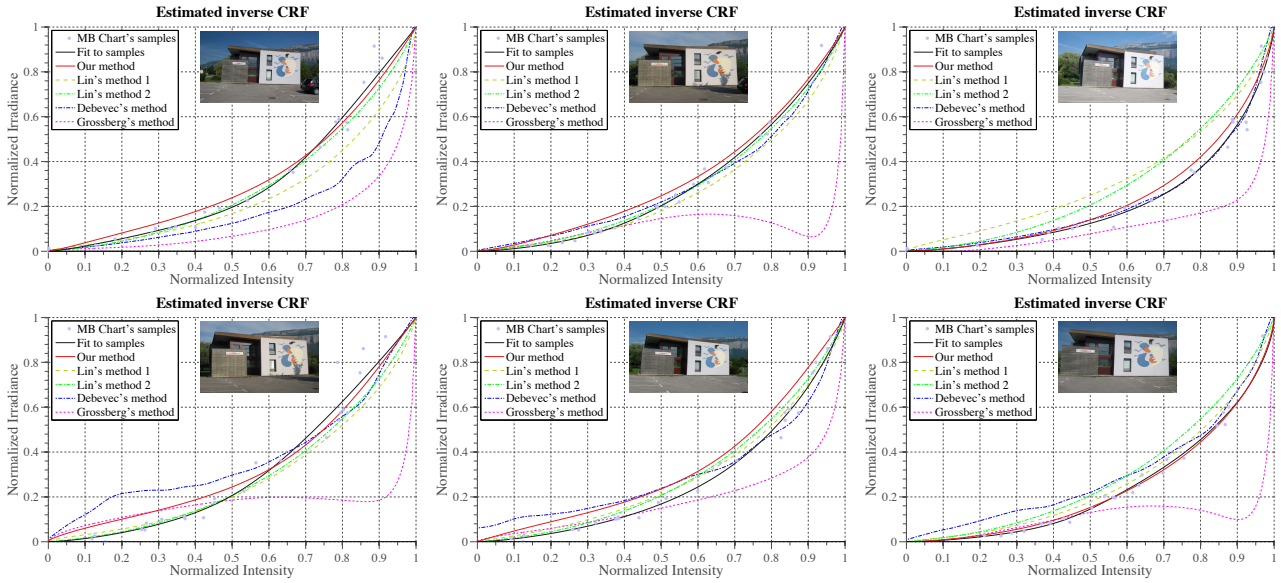

Figure 3. Estimated inverse CRF's found with the algorithm described in section 4.3 and compared to: Measured inverse CRF's with the color checker samples interpolated. Lin's *et al*. algorithm using two different values for parameter $\lambda = \{10^{-1}, 10^{-4}\}$ (*c.f.* [14]). HDRShop results [4]. Grossberg's *et al*. multiple exposures approach [8]. Constraints in the monotonicity of the estimated CRF's are imposed.

RMS error/Database	Channel R	Channel G	Channel B
1. Own database	0.0431	0.0528	0.0463
2. Internet Database	0.0627	0.0716	0.0679

Table 2. Average RMS error for the two databases described in section 5.2. RMS error is calculated relative to Lin's estimation (with fixed parameter λ) in both cases.

ages in order to reduce processing time and complexity). Because in this case it is impossible to collect samples using a color calibration chart neither to obtain multiple exposures with a static camera, we can only compare our results with the estimations computed by Lin's algorithm. In figure 4, six recovered inverse CRF's are shown. Despite the dependence of Lin's method to the aforementioned regularization term, this approach has been to date one of the most reliable methods for estimating the CRF of single images. Our proposed solution is very close in most of the cases to the results of Lin's algorithm. We are also aware that Lin's method could present problems when it handles preprocessed images (the non-linearity of the CRF is altered along edges in this case).

6. Discussion and conclusions

We have presented an approach to recover the camera response functions of a set of images in a photo collection, acquired with different cameras. The approach exploits a 3D model generated using available methods and a powerful empirical prior on CRF's, without which the problem would be ill-posed. Compared to other methods, our approach does not require multiple aligned images with the same CRF or the same lighting conditions. Our method is motivated as a first step in the pipeline of recovering surface properties and illumination conditions from multiple images and, in this context it can not be fairly compared with methods for single image CRF recovering [14, 15, 17, 26, 12, 25] (even so, results are as good as those estimated by these algorithms). Moreover, our method estimates all the CRFs for a group of images at the same time, in an automatic way, while, according to our experience, single image methods require most of the time some supervision or a manual tuning of the parameters, dependent on the input image. For the case of image collections obtaining the CRF with these methods could be an extremely slow and tedious task. In this context, our results are promising.

Limitations of our approach are as follows. The illumination model, consisting of a single directional light source and ambient lighting, is certainly not entirely realistic. More general models can in principle be used, *e.g.* ones based on spherical harmonics [1, 21], but it remains to be seen if their estimation is well-posed. Finally, empirical priors on outdoor lighting conditions seem to be a good alternative. This "spherical harmonics" framework opens the door to the formulation of priors based on complexes BRDFs as illustrated in [22]. The empirical camera model proposed by Grossberg provides a good approximation to CRF's of cameras, where there is no evidence that different

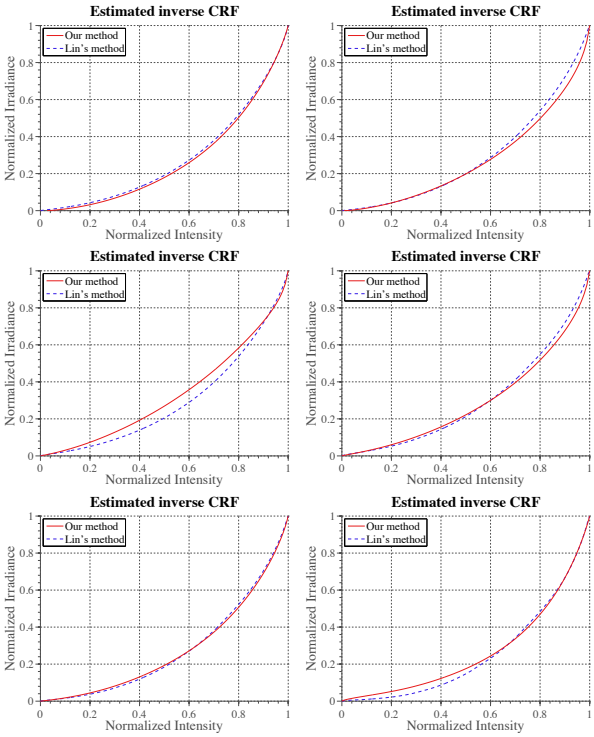

Figure 4. Estimated Inverse CRF's using our algorithm on a 3D model reconstructed from Internet images. Results of our own implementation of Lin's *et al*. [14] work are also shown.

color channels' CRF's are correlated. However, the joint estimation of CRFs over all channels generates consistent CRFs. Color information could be used in the future, adding a camera sensitivity factor in order to model the influence of the light wavelength. Another limitation is the assumption of Lambertian reflectance, although this is relatively easy to circumvent by using a robust weighting function in the non-linear optimization and having recourse to iteratively reweighted least squares optimization for example. We currently do not handle shadows associated with the directional light source. This could also be taken care of to some degree by a robust weighting function. One may also handle shadows explicitly since from the known scene geometry and the current estimate of the light source, shadows can be predicted and used to correctly predict brightness values. However, the resulting cost function would become discontinuous and harder to optimize. An alternative would be to directly detect shadows in the image with dedicated methods such as [5] and use the result as input to our methods.

References

[1] R. Basri, D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. on PAMI*, 2003. 7

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

[3] Y.-C. Chang, J. Reid. RGB calibration for color image analysis in machine vision. *IEEE Trans. on Image Processing*, 1414–1422, 1996. 2

[4] P. Debevec, J. Malik. Recovering high dynamic range radiance maps from photographs. *SIGGRAPH*, 1997. 2, 6, 7

[5] G. Finlayson, S. Hordley, C. Lu, M. Drew. On the removal of shadows from images. *PAMI*, 59–68, 2006. 8

[6] Y. Furukawa, J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on PAMI*, 2009. 1

[7] Y. Furukawa, J. Ponce. Patch-based multi-view stereo software. 2009. http://grail.cs.washington.edu/software/pmvs. 1, 3

[8] M. D. Grossberg, S. K. Nayar. What is the space of camera response functions? *CVPR*, 2:602, 2003. 2, 4, 5, 6, 7

[9] M. D. Grossberg, S. K. Nayar. Modeling the space of camera response functions. *PAMI*, 1272 – 1282, 2004. 2, 3

[10] T. Haber, C. Fuchs, P. Bekaer, H.-P. Seidel, M. Goesele, H. Lensch. Relighting objects from image collections. *CVPR*, 627–634, 2009. 2

[11] A. Ilie, G. Welch. Ensuring color consistency across multiple cameras. In *ICCV*, 1268–1275, 2005. 2

[12] S. J. Kim, J. Frahm, M. Pollefeys. Radiometric calibration with illumination change for outdoor scene analysis. *CVPR*, 2008. 7

[13] S. Kuthirummal, A. Agarwala, D. Goldman, S. K. Nayar. Priors for large photo collections and what they reveal about cameras. In *ECCV*, 74–87, 2008. 2

[14] S. Lin, J. Gu, S. Yamazaki, H.-Y. Shum;. Radiometric calibration from a single image. *CVPR*, 2938–945, 2004. 2, 6, 7, 8

[15] S. Lin, L. Zhang. Determining the radiometric response function from a single grayscale image. *CVPR*, 2005. 2, 7

[16] Q. Luong, P. Fua, Y. Leclerc. The radiometry of multiple images. *IEEE Trans. on PAMI*, 19–33, Jan 2002. 2, 3

[17] Y. Matsushita, S. Lin. Radiometric calibration from noise distributions. *CVPR*, 2007. 7

[18] T. Mitsunaga, S. K. Nayar. Radiometric self calibration. *CVPR*, 1, 1999. 2

[19] T.-T. Ng, S.-F. Chang, M.-P. Tsui. Using geometry invariants for camera response function estimation. *CVPR*, 2007. 2

[20] D. Pascale. Rgb coordinates of the macbeth colorchecker. Technical report, The BabelColor Company, Jun 2006. 6

[21] R. Ramamoorthi, P. Hanrahan. A signal-processing framework for inverse rendering. *SIGGRAPH*, 117–128, 2001. 7

[22] F. Romeiro, T. Zickler. Blind reflectometry. In *ECCV*, 45–58, 2010. 7

[23] N. Snavely, S. M. Seitz, R. Szeliski. Photo tourism: exploring photo collections in 3D. *SIGGRAPH*, 2006. 1

[24] N. Snavely, S. M. Seitz, R. Szeliski. Modeling the world from internet photo collections. *IJCV*, 2008. 1, 2, 3

[25] J. Takamatsu, Y. Matsushita, K. Ikeuchi. Estimating camera response functions using probabilistic intensity similarity. *CVPR*, 2008. 2, 7

[26] B. Wilburn, H. Xu, Y. Matsushita. Radiometric calibration using temporal irradiance mixtures. *CVPR*, 2008. 7