Single-Image Estimation of the Camera Response Function in Near-Lighting*

Pedro Rodrigues[1] and João P. Barreto[1,2]
[1]Institute of Systems and Robotics, University of Coimbra
Coimbra, Portugal
{prodrigues,jpbar}@isr.uc.pt

[2]Department of Electrical and Computer Engineering, Faculty of Sciences and Technology
University of Coimbra, Coimbra, Portugal

Abstract

The camera response function (CRF) relates quantised image pixel values with physical incoming light. This paper describes a method to estimate the CRF from a single image of a general two-coloured surface for which the albedo ratio between the coloured regions is known a priori. *While other radiometric calibration methods either use multiple frames or require the light to be infinitely distant, the algorithm herein proposed makes no assumptions about lighting conditions and can handle cameras with strong vignetting. Although the approach is generic, in the sense that can be applied to any camera system, the method is particularly well suited for determining the CRF of near-lighting endoscopes in the operating room. This is a very pertinent problem for which no practical, effective solutions have been proposed. The robustness, repeatability, and accuracy of the algorithm is experimentally validated in real images acquired with different endoscopic set-ups.*

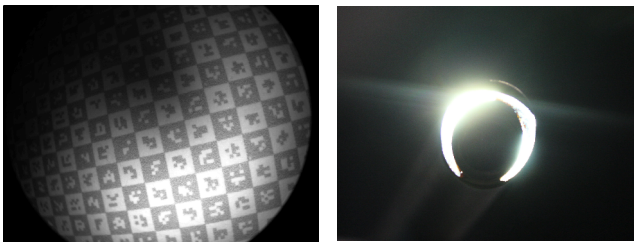
1. Introduction

In minimally invasive procedures the doctor executes the clinical action of surgery or diagnosis based exclusively in the video acquired by a small endoscopic camera that is inserted in the targeted anatomical cavity. Such procedures are prone to errors and very difficulty to execute, with the surgeons having to undergo a long training period till mastering the surgical technique. In this context, improving visualization conditions and developing systems for assisting the doctor during the procedures is of paramount importance to decrease clinical errors and to reduce the surgeon learning curve.

It is well known that, in most cameras, the relation between incident light and the image digitized values is non-linear. Endoscopic cameras are no exception with such non-linearity being either due to limitations in camera/optics manufacturing, or intentionally used for compressing the dynamic range of the sensors or adapting the visualization to a particular display technology. The camera response function (CRF) models this non-linear relation by describing how (physically meaningful) incoming light is mapped to quantised image brightness values. Calibration of the CRF can be useful in many ways in the context of clinical endoscopy. Since the color of organs and tissues is a primary cue in many diagnosis procedures, fidelity in color visualization is of key importance. Thus, the CRF can be used in an image post-processing stage for the purpose of color constancy and white-balance [9], *i.e.*, invariance of color to different illumination, lenses, and cameras. The CRF estimation is a crucial step for deblurring as described in [18] and is an important first step for the application of photometric vision algorithms for the purpose of developing systems of computer-aided surgery (CAS). Reconstruction methods like shape-from-shading [20, 6] or photometric stereo [3] assume a linear (or affine) mapping between physical scene radiance and image brightness, and such techniques are becoming increasingly popular in CAS as a way of performing 3D modelling of deformable surfaces [12].

Radiometric calibration is a classical problem in computer vision with several different methods described in the literature [4, 15, 5, 8, 10, 16, 13, 19]. However, and as discussed in the subsequent section, the estimation of the CRF of medical endoscopes poses specific challenges that are not properly addressed by the current state-of-the-art. First, the anatomical cavity is illuminated via a light guide that runs along the endoscope, which means that we can neither consider a far light source, nor a light source coincident with the optical center [17]. It is a situation of near-light source [20] for which the vast majority of radiometric calibration methods are not applicable. Moreover, and as pointed out

*The method and apparatus described in this work are object of the US Provisional Patent Application No. 62/146,643.



(a) (b)

Figure 1: The calibration grid imaged with an endoscopic camera (a) and photograph of the endoscope tip with the light on (b).

by Collins and Bartoli [3], the light source can hardly be considered punctual or isotropic (see figure 1). As a bottom line, and since the shape and format of the distal tip of the light guide varies across endoscopes, it is highly desirable to carry the CRF estimation without making assumptions about illumination conditions. Second, since in endoscopy the lens scope is mounted on the camera-head immediately before starting the intervention, the calibration procedure must be fast, robust, and require minimal user intervention in order to be carried by the surgeon in the operating room (OR) without disturbing the existing clinical routine. Melo *et al*. [14] have recently proposed a method for geometric camera calibration in endoscopy that requires a single image of a checkerboard pattern acquired from a generic pose [2]. To the best of our knowledge, there is no similar solution for radiometric calibration that addresses this important usability requirement (as discussed later, the single-image methods of [10, 16, 19, 13] do not apply to medical endoscopes).

In this article, we describe a method for estimating the CRF from a single image of a surface with regions of two colors for which the albedo ratio is known in advance. The algorithm makes no assumptions about the illumination conditions or the image vignetting. Although the shape of the calibration surface is irrelevant for the CRF estimation, all the experiments are carried using a planar checkerboard target similar to the one described in [1]. The reasons are twofold: the automatic segmentation and identification of the two color regions is straightforward, and by combining our radiometric calibration method with the geometric calibration approach proposed in [14] it is possible to fully model the camera from a single calibration frame.

As a final remark, please note that knowing the two albedos is only required to avoid the exponential ambiguity on the CRF estimation.

1.1. Related work

CRF estimation is a mature topic with several methods described in the literature. However, the need for a single-image approach allied to several technicalities to account for (near lightning, a strong vignetting effect, blurring, chromatic aberrations) render current approaches inapplicable to our set-up.

- Standard CRF estimation is done with a planar color chart (*e.g*., classic 24-patch ColorChecker) of which the albedos are known. This is not applicable in endoscopy set-ups, as it is only applicable when the light is uniformly distributed in the chart and the vignetting effect can be overlooked.

- Another class of approaches involves the use of multiple registered images of the same scene, either photo collections [5, 8] or images with different exposures [4, 15]. The CRF estimation is carried from pixel correspondences across multiple frames of the same scene. This is a plausible approach for our application, but a controlled multiple-image calibration procedure cannot be imposed on a surgeon in the OR.

- A few single-image CRF estimation have been proposed.

  Lin and Yamazaki [10] have proposed a method where they explore the non-linearity of the transitions between two albedos of a generic image to estimate the CRF. The method lies on the assumption that vignetting and light do not change across these edges and that demosaicing, blur, and chromatic aberrations do not occur. Most of these assumptions do not hold in endoscopic rigs.

  Ng *et al*. [16] identify points where the irradiance can be locally approximated by a plane. The authors show that on such points there is a relationship between the derivatives of the measured pixel values and the CRF with no unknowns other than the CRF itself. This relationship is used to estimate the CRF. The method tends to use edges that suffer from chromatic aberrations, demosaicing inaccuracies, and blurring.

  Matsushita and Lin [13] propose the use of noise distributions to find the CRF. However it showed to not be applicable in low-light conditions. It is highly dependent on having multiple noise readings of the full spectrum of intensities, and on being able to detect uniform regions (which is not doable in our case due to strong vignetting and the near lighting conditions).

  Wilburn *et al*. [19] propose two approaches for single-image CRF estimation: one that uses an uncalibrated monitor and a controled camera shutter speed (the two must be synchronized); the second uses motion blur in a similar way to [10]. The first approach is not applicable with strong vignetting and the white balance could

not be performed on the same image. The second approach suffers from the same problems as [10] with the additional issue of the need of acquiring an image with motion blur.

- On applications to medical endoscopy we have the work of Collins and Bartoli [3], where the radiometric calibration is said to be performed *a priori*, and thus we can only assume it is done off-line using conventional techniques and, as such, it does not meet the usability constraints. Wu *et al*. [20] propose a method valid for near-lighting and strong vignetting conditions. The calibration involves the acquisition of an image for each color patch (of the classic 24-patch ColorChecker) with the same relative camera pose. The idea is that a pixel across different patch images is affected by the same amount of vignetting and light, the only change the albedo. This allows a system of equations to be built where the only unknown is the CRF. The calibration is troublesome and time consuming, not being suited to be carried by a surgeon in the OR. Nevertheless, it is the closest related work and we will use it as benchmark.

Our approach follows a similar reasoning as the one proposed by Wu *et al*. in [20]. But, instead of acquiring multiple images to be able to construct equations invariant to lighting and vignetting, we will find points where this invariance holds on a single image.

1.2. Notation

Matrices are represented by symbols in a sans serif font, *e.g*., A. Vectors and vector functions are represented by bold symbols, and scalars are denoted by plain font letters, *e.g*., $\mathbf{x} = (x, y)^\mathsf{T}$ and $\mathbf{f}(\mathbf{x}) = (f_x(\mathbf{x}), f_y(\mathbf{x}))^\mathsf{T}$. Sets are denoted by upper-case calligraphic letters, *e.g*., $\mathcal{L}$.

2. Formulation and Overview

The radiometric image formation model can be written as

$$d(\mathbf{x}) = f(\alpha e(\mathbf{x})) \tag{1}$$

where d is the map of imaged/stored values (*i.e*., the actual image), $\mathbf{x}$ is a scene point, and e is the sensor/image irradiance. f is the CRF, a monotonically increasing curve, and α is the exposure, a combination of sensor gains, shutter time, and aperture. The sensor irradiance is given by

$$e(\mathbf{x}) = m(\mathbf{x})l(\mathbf{x}) \tag{2}$$

with m being the vignetting (natural, optical, mechanical, and pixel vignetting) and other spatial attenuation effects, and l the scene radiance.

A generic model of the scene radiance can be written as

$$l(\mathbf{x}) = \rho(\mathbf{x})q(\mathbf{n}(\mathbf{x}), \mathbf{i}(\mathbf{x})) \tag{3}$$

where ρ is the albedo and q is the light intensity on the scene as a function of the scene normals $\mathbf{n}$ and the vector $\mathbf{i}$ from the scene point $\mathbf{x}$ to a light. For multiple/extended lights there will be multiple light vectors, and function q is, in general, a sum/integral of the light functions. This is not relevant for our approach, as it is independent of the light effect on the scene as well as of the vignetting.

Since f is invertible, to facilitate the estimation, we rewrite the above equations as

$$f^{-1}(d(\mathbf{x})) = \alpha\rho(\mathbf{x})m(\mathbf{x})q(\mathbf{n}(\mathbf{x}), \mathbf{i}(\mathbf{x})) \tag{4}$$

giving the irradiance as a function of the image.

As mentioned in the previous section, we will look for pixels, on a single image, where equations can be written with invariance to the vignetting and the light behaviour. As in [20], we will look for pairs of pixels with different albedos for which we know that both the vignetting and the light effect can be cancelled. On their approach, both the vignetting and the light effect remain constant between albedos (different ColorChecker patch images). On a single-image approach, one cannot expect to find regions where both are constant without modelling them. However, in fact, we do not need to be invariant to both effects, only to their joint effect. In this way, we are able to build a system of equations where the only unknowns are the CRF and the albedos, without making assumptions about the vignetting or the light behaviour. This is of crucial importance for our set-up, since the vignetting is not always central (as with most set-ups) and the lights are at close range, are typically not punctual, and may not be isotropic [3].

More specifically, let the sensor irradiance be written as

$$f^{-1}(d(\mathbf{x})) = \rho(\mathbf{x})u(\mathbf{x}) \tag{5}$$

where $u(\mathbf{x})$ is composed of exposure, vignetting, and the light effect on the scene. Throughout the paper we will call this function the irradiance normalized by albedo (INA). On the ith isovalue line of the INA

$$u(\mathbf{x}_j) = \kappa_i = \frac{f^{-1}(d(\mathbf{x}_j))}{\rho(\mathbf{x}_j)} \tag{6}$$

$\forall j \in \mathcal{L}_i$, where $\mathcal{L}_i$ is the set of pixels crossed by isovalue line i and κ_i is a constant. Thus, if a line i passes through multiple albedos (ρ_1 and ρ_2 as an example), one will have

$$\frac{f^{-1}(d(\mathbf{x}_j))}{\rho_1} = \frac{f^{-1}(d(\mathbf{x}_k))}{\rho_2} \tag{7}$$

$\forall j \in \mathcal{L}_i \cap \mathcal{A}_{\rho_1}, \forall k \in \mathcal{L}_i \cap \mathcal{A}_{\rho_2}$, with an equation for each pair of albedos on each isoline. $\mathcal{A}_{\rho_n}$ is the set of points with

a specific albedo $\rho(\mathbf{x}) = \rho_n$. This last equation will then be used for the single-image CRF estimation.

Please note that this approach is scalable. Additional images with the same set-up would provide additional equations that can be grouped together to improve the estimation. Even with different exposures, different poses, and changes in the vignetting (due to changes of zoom and/or aperture).

2.1. Assumptions and Ambiguities

The proposed approach lies on the following assumptions: that we have a scene of a two-color (at least) lambertian surface; and that we know the albedo ratio between the regions, which can be estimated off-line with a radiometrically calibrated camera. In addition, to perform white balance with the same frame one of the colors must be white/gray.

In the present work, we have used a planar two-color CALTag grid [1] (see figure 2) geometrically calibrated with the method proposed in [14]. In fact, we do not need it to be planar nor a grid for the framework to succeed. However, we have chosen this grid to be able to perform both the geometric and radiometric calibrations with a single image.

We have chosen to use a white and light gray grid instead of the usual black, so that we have an higher signal-to-noise ratio on the darker regions. This is specially important for poorly lit scenes and for the calibration of the lower brightness spectrum of the CRF.

The fact that we determine the albedo ratio ρ_1/ρ_2 beforehand removes problems of exponential ambiguities present otherwise. This happens because, if the albedo ratio is not known and we include its estimation on the solver, from equation 7,

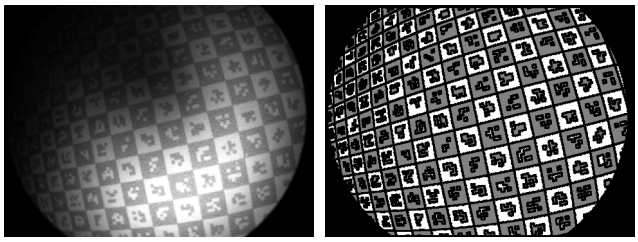
$$\left(f^{-1}(d(\mathbf{x}_j))\right)^\gamma = \left(\frac{\rho_1}{\rho_2} f^{-1}(d(\mathbf{x}_k))\right)^\gamma, \tag{8}$$

$\forall j \in \mathcal{L}_i \cap \mathcal{A}_{\rho_1}, \forall k \in \mathcal{L}_i \cap \mathcal{A}_{\rho_2}$, which gives valid solutions independently of γ. Instead of being determining the CRF up-to-scale, we would also be doing it up to an exponential ambiguity. Similarly to what is described in [11].

The scale ambiguity is not a problem because, for a color image, a white balance method can be used to establish the relative scale of the channels. The scale ambiguity of a grayscale image, or the global scale of color image, will simply give more or less brightness and more or less saturation to the image. It is the equivalent to the exposure constant α and can be set to match the brightness of the original image and/or to minimize the saturated pixels.

3. Detailed Algorithm

In this section we will describe in detail the algorithm used for the single-image CRF estimation. This framework can be conceptually divided into three steps. The segmentation of the two albedos, the detection of INA isovalue lines, and the actual CRF estimation and parametrization.



(a) (b)

Figure 2: The calibration grid (a) and its segmentation into the two albedos (b): in white are the patches with lighter albedo and in gray are the patches with darker albedo.

3.1. Segmentation

Having used the method proposed in [14] using CALTag grids, the segmentation is straightforward. Since the positions on the scene plane of every fiducial marker and grid corner are known or can be easily estimated, using the geometric calibration information, we can warp the grid to the image plane. This warped image is itself our segmentation. To avoid regions of blurring, a morphological erosion is performed to both regions. Figure 2 shows a CALTag grid and its segmentation.

3.2. Isoline detection

Since $u(\mathbf{x})$ is not known, we need to evaluate how its isolines behave on the image $d(\mathbf{x})$. By rewriting (6) as

$$d(\mathbf{x}_j) = f(\rho(\mathbf{x}_j)\kappa_i) \tag{9}$$

$\forall j \in \mathcal{L}_i$, it is shown that, for a given albedo, an isoline in the sensor irradiance is also an isoline in the image $d(\mathbf{x})$. In addition, along an isoline of $u(\mathbf{x})$, $\mathcal{L}_i$, the image values form a piecewise constant function (with a different constant value for each albedo). Figure 3 shows each image component individually. Specifically, this behaviour can be seen in figure 3b, where we show each image component on an isoline of $u(\mathbf{x})$.

In the image space we have a set of isolines for each albedo. However, the isolines of $d(\mathbf{x})$ for each albedo are the same, and equal to the isolines of $u(\mathbf{x})$, except for its value (figure 3b).

To find the isolines of $u(\mathbf{x})$, let us model the image along one albedo ρ_n as a generic model h_n where the isolines are known. We can write for two albedos on the image space

$$d(\mathbf{x}_j) = h_1(\mathbf{x}_j) \tag{10a}$$
$$d(\mathbf{x}_k) = h_2(\mathbf{x}_k) \tag{10b}$$

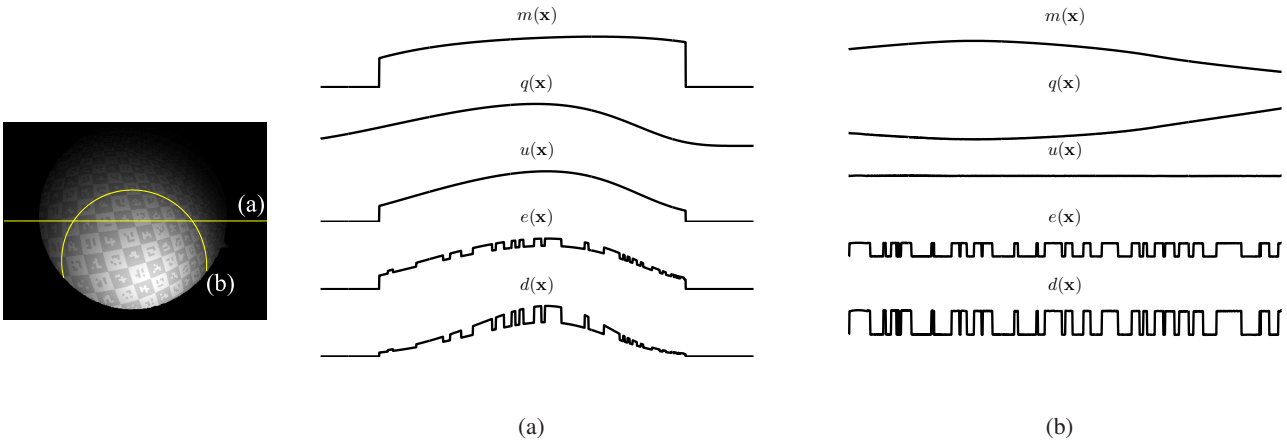


Figure 3: Values (arbitrary units) of each image component showing the formation process of an image of a checkerboard along (a) a horizontal line and (b) an isoline of the irradiance normalized by albedo.

$\forall j \in \mathcal{A}_{\rho_1}, \forall k \in \mathcal{A}_{\rho_2}$. From before, we know that the isolines of $h_1(\mathbf{x})$ and $h_2(\mathbf{x})$ will have the same shape as the ones in $u(\mathbf{x})$ but with different values. The shape of the surfaces represented by the models are different, since the step between isolines varies from one formulation to the other, but the isolines are the same. If we consider one of the albedos, let us say ρ_1, as a reference we can show that the two models are related by

$$h_1(\mathbf{x}_k) = f\left(\frac{\rho_1}{\rho_2} f^{-1}(h_2(\mathbf{x}_k))\right) = g(h_2(\mathbf{x}_k)) \tag{11}$$

$\forall k \in \mathcal{A}_{\rho_2}$, where g is a positive and monotonically increasing function that is used to transform the model h_2 into the model h_1. This function g is the equivalent of having a gain for each isoline for the points of the albedo ρ_2, to be able to use only the model h_1 (relative to the albedo ρ_1) for both albedos.

More specifically, we have used a polynomial model for $h_1(\mathbf{x}) = \mathsf{B}\mathbf{p}$, where $\mathbf{p}$ are the polynomial coefficients and $B = (\mathbf{b}_1, \mathbf{b}_2, \cdots)^\mathsf{T}$, with $\mathbf{b}_n = \left(1, x_n, y_n, x_n^2, y_n^2, x_n y_n, \cdots\right)^\mathsf{T}$. x and y are the image coordinates, *i.e.*, $\mathbf{x} = (x, y)^\mathsf{T}$. This way we get

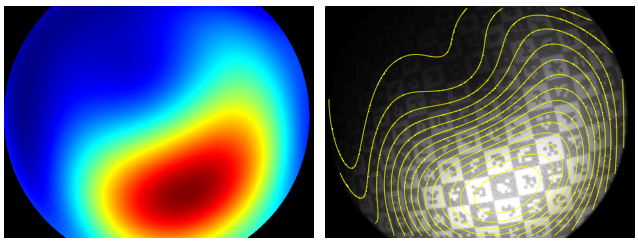
$$\mathbf{b}_j^\mathsf{T}\mathbf{p} = d(\mathbf{x}_j) \tag{12a}$$

$$\mathbf{b}_k^\mathsf{T}\mathbf{p} = g(d(\mathbf{x}_k)). \tag{12b}$$

The isovalue lines will then be extracted as the level sets of the polynomial. As for the linear system of equation to be solved, it is defined as

$$\begin{bmatrix} \mathbf{b}_j^\mathsf{T} & \mathbf{0}^\mathsf{T} \\ \mathbf{b}_k^\mathsf{T} & -\mathbf{s}^\mathsf{T}(d(\mathbf{x}_k)) \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} d(\mathbf{x}_j) \\ 0 \end{bmatrix} \tag{13}$$

$\forall j \in \mathcal{A}_{\rho_1}, \forall k \in \mathcal{A}_{\rho_2}$, where $\mathbf{g}$ is a discrete array version of the function g and $\mathbf{s}(n)$ is sparse vector with a single value of 1 on the element n.



(a) (b)

Figure 4: A depiction of (a) the estimated h_1 (a surface with the same isolines as the irradiance normalized by albedo) and (b) the calibration grid image superposed with some of the isolines.

An example of the determined isolines can be observed in figure 4.

3.3. Estimation

Having the isolines, we estimate the CRF using (7). For a given isoline i and albedo ρ_n we compute

$$\tilde{d}_{i\rho_n} = \underset{j \in \mathcal{L}_i \cap \mathcal{A}_{\rho_n}}{\text{median}}\ d(\mathbf{x}_j) \tag{14}$$

and rewrite (7) as

$$f^{-1}(\tilde{d}_{i\rho_1}) - \frac{\rho_1}{\rho_2} f^{-1}(\tilde{d}_{i\rho_2}) = 0. \tag{15}$$

To find the CRF the following linear system is built

$$\mathbf{a}_i^\mathsf{T}\mathbf{v}_{f^{-1}} = 0 \tag{16}$$

$\forall i$, where

$$\mathbf{a}_i = \mathbf{s}(\tilde{d}_{i\rho_1}) - \frac{\rho_1}{\rho_2}\mathbf{s}(\tilde{d}_{i\rho_2}) \tag{17}$$

and we will solve for $\mathbf{v}_{f^{-1}}$, the discrete array version of the function f^{-1}.

We also enforce monotonicity constraints to $\mathbf{v}_{f^{-1}}$ and a Tikhonov regularization of its second derivatives.

3.3.1 Parametrization

At this point, from the previous optimization step, we have a sparse representation of the CRF that is defined by $\mathbf{d}^*$, the brightness values described in the equations, and $\mathbf{v}^*_{f^{-1},n}$, the corresponding optimized irradiance values (up-to-scale). To interpolate and extrapolate to other values of image brightness that were not covered by the equations in the previous estimation step, we will parametrize our CRF by fitting a well known model.

Among many possible parametric models that could have been used, we have the EMoR [7] and the generalized gamma curve model (GGCM) [16]. The following description will use the GGCM, but this parameterization could be done with other models as well. For the inverse CRF, the GGCM is defined as

$$f^{-1}_{\text{GGCM}}(d,\beta) = d^{1/\sum_{n=0}^{N}\beta_n d^n} \tag{18}$$

where N is the order of the polynomial that defines the model. As in Ng *et al*. [16], we have used $N = 1$.

The reason for not parametrizing directly in the previous estimation step is simply the nonlinearity of the GGCM. To have faster and more robust results we have decided to parametrize the model in a post-processing stage.

We have added a shift variable δ_0 and a scaling variable δ_1 because of sensor gains and shifts, thus

$$f^{-1}_{\text{sGGCM}}(d,\delta,\beta) = f^{-1}_{\text{GGCM}}(\delta_1\,(d+\delta_0)\,,\beta). \tag{19}$$

Let us define a vector $\mathbf{f}(\mathbf{d},\delta,\beta) = \left(f^{-1}_{\text{sGGCM}}(d_1,\delta,\beta), f^{-1}_{\text{sGGCM}}(d_2,\delta,\beta),\cdots\right)^{\mathsf{T}}$. The GGCM is then fitted to $\mathbf{v}^*_{f^{-1}}$ using least squares and subject to monotonicity constraints. The parameters $\delta = (\delta_0,\delta_1)^{\mathsf{T}}$ are initialized with a fitting of a linear model (and thus $\beta = (\beta_0,\beta_1)^{\mathsf{T}}$ is initialized as $\beta = (0,1)^{\mathsf{T}}$).

3.4. RGB Images and White Balance

For grayscale images, the above estimation procedure is sufficient. Nevertheless, for displaying RGB images after CRF correction there is still the need to perform white balance.

White balancing is the process through which three gains are estimated, one for each channel, to make actual white/gray points appear white/gray on the image. The CRF estimation can be done independently for each channel (as explained before) followed by a standard white balancing step using white/gray points. However, using a few constraints on the solver, this can be performed directly in the linear estimation.

Table 1: Description of the camera set-ups used on the acquired datasets. Please see text for more details on each component.

	Camera	Mount	Lens	Light
R1	Sentech	no zoom	arthroscope	lens tip
R2	Sentech	no zoom	arthroscope	lamp
R3	Flea	zoom	laparoscope	lens tip
R4	Grasshopper	-	Stryker	lamp
R5	Flea	zoom	arthroscope	sunlight
R6	Flea	zoom	arthroscope	lens tip

For the joint CRF estimation and white balance, we estimate the 3 CRFs simultaneously and enforce white balance constraints on the solver. To estimate the 3 CRFs, the linear system becomes

$$\begin{bmatrix} \mathsf{A}_{\mathsf{R}} & 0 & 0 \\ 0 & \mathsf{A}_{\mathsf{G}} & 0 \\ 0 & 0 & \mathsf{A}_{\mathsf{B}} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{f_{\mathsf{R}}^{-1}} \\ \mathbf{v}_{f_{\mathsf{G}}^{-1}} \\ \mathbf{v}_{f_{\mathsf{B}}^{-1}} \end{bmatrix} = \mathbf{0}. \tag{20}$$

Let $\mathbf{p}_c$ be an array containing the histogram of $d_c(\mathbf{x}_{j\in\mathcal{W}})$, the measured values (of channel $c \in \{\mathsf{R},\mathsf{G},\mathsf{B}\}$) of the pixels chosen for the white balance procedure. The selected pixels follow 3 criteria: they belong to the white patches, are above a certain threshold to avoid noise, and are not saturated. The white balance equations are defined as

$$\begin{bmatrix} \mathbf{p}_{\mathsf{R}}{}^{\mathsf{T}} & -\mathbf{p}_{\mathsf{G}}{}^{\mathsf{T}} & \mathbf{0}^{\mathsf{T}} \\ \mathbf{0}^{\mathsf{T}} & \mathbf{p}_{\mathsf{G}}{}^{\mathsf{T}} & -\mathbf{p}_{\mathsf{B}}{}^{\mathsf{T}} \\ \mathbf{0}^{\mathsf{T}} & \mathbf{p}_{\mathsf{G}}{}^{\mathsf{T}} & \mathbf{0}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{f_{\mathsf{R}}^{-1}} \\ \mathbf{v}_{f_{\mathsf{G}}^{-1}} \\ \mathbf{v}_{f_{\mathsf{B}}^{-1}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sum_{j\in\mathcal{W}} d_{\mathsf{G}}(\mathbf{x}_j) \end{bmatrix}. \tag{21}$$

The first two equations perform the white balance, the third equation constrains the overall brightness of the CRF-corrected image to be similar to the original image.

The general shape of the CRF can be assumed to be preserved between channels, thus the CRFs will be computed as

$$f^{-1}_{\text{sGGCM},c}(d_c,\delta_0,\delta_c,\beta) = f^{-1}_{\text{GGCM}}(\delta_c\,(d_c+\delta_0)\,,\beta) \tag{22}$$

to account for the in-camera white balance parameters (δ_c and δ_0).

The sum along the three channels of the mean square error is minimized. As before, the parameters $\beta = (\beta_0,\beta_1)^{\mathsf{T}}$ and $\delta = (\delta_0,\delta_{\mathsf{R}},\delta_{\mathsf{G}},\delta_{\mathsf{B}})^{\mathsf{T}}$ are initialized with a fitting of a linear model.

4. Experimental Validation

To validate the proposed method, we have compared it to Wu *et al*. approach [20]. Six datasets were acquired with

different set-ups, described in table 1. The cameras used were a Point Grey (Point Grey Research inc, BC, Canada) Flea3 CMOS USB3 camera, a Point Grey Grasshopper2 CCD FireWire camera, and a Sentech (Sentech co. ltd, Kanagawa Prefecture, Japan) HD203DV CMOS HDMI camera. As for the lens, we have used a Smith & Nephew (Smith & Nephew plc, London, UK) Dyonics laparoscopic lens, a Smith & Nephew Dyonics arthroscopic lens, and a Stryker (Stryker Corporation, MI, USA) arthroscopic lens. We have used three different mount adaptors to couple the lens to the cameras and three different light sources. The gamma parameter intrinsic to the cameras was also altered between datasets to cover a wider range of scenarios. The albedo ratio necessary for our approach was previously determined using a radiometrically calibrated Canon 600D (Canon inc, Tokyo, Japan).

Each dataset is composed of five CALTag grid images with different poses and one image of each of the 24 patches of the ColorChecker, all with the same pose and exposure. All CRF estimations, including the estimation from Wu *et al*. approach, are defined up-to-scale and, therefore, are not directly comparable. To be able to compare the CRFs we perform a linear optimization for each individual scaling factor (one for each estimation). Figure 5 shows the resulting CRFs for each dataset.

Our approach showed high repeatability and invariance to the light source. On some cases, the extrapolation gives erroneous estimations, which is an expected behaviour when the range of values on the image is small. For these cases, since the proposed method is scalable (as discussed in section 2), a multiple image approach can be used.

Note that this single-image approach successfully calibrated the sunlight dataset (R5), showing that it does not require both strong vignetting and near-lighting conditions.

In a few datasets, mainly R3 and R4, the estimation tends to deviate from the Wu *et al*. approach. However it still shows high repeatability.

Figure 6 demonstrates the color constancy properties of the CRF correction and white balance procedures. Please note that, since there is a scaling factor that is not being taken into account and there is still the presence of vignetting and light components, the insets only need to match the ground truth on the background in terms of color and not overall brightness.

5. Conclusions

With this paper we have proposed a single-image approach for estimating the CRF for near-light and/or strong vignetting conditions, for which a general image of a two color albedo surface can be used.

Unlike other single-image calibration algorithms that require the detection of regions where light is constant on the scene and no vignetting is present, our approach benefits from the effects of near-lighting and vignetting to perform the estimation.

Our approach is specially suited for calibration of endoscopic rigs which suffer in general from radial distortion and, thus, we have proposed the use of CALTag grids so that the geometric caliration, the CRF estimation, and the white balance can be performed with a single frame.

Other applications include indoor calibration (when the light incident on the scene cannot be considered as a distant light), the calibration of smartphone cameras with flash at close range, and the calibration of generic cameras with strong vignetting.

The method showed high repeatability and good results when compared to the 24 same-pose image calibration procedure proposed in [20]. A practical limitation is that the scene needs to be well lit for the equations to be able to describe the full color spectrum, *i.e*., to reduce extrapolation. However, this is an issue common to all CRF estimation techniques. This can be overcome using multiple images of the calibration target, with the advantage that registration is not required nor the images need to be at the same pose.

As future work, we want to eliminate the need for accurately knowing the albedo ratio in advance, similarly to what has been done in [15].

Acknowledgements

The authors want to thank QREN-Mais Centro for funding under the research project "Novas Tecnologias para Apoio à Saúde e Qualidade de Vida, Projecto Cirurgia e Diagnóstico Assistido por Computador Usando Imagem".

References

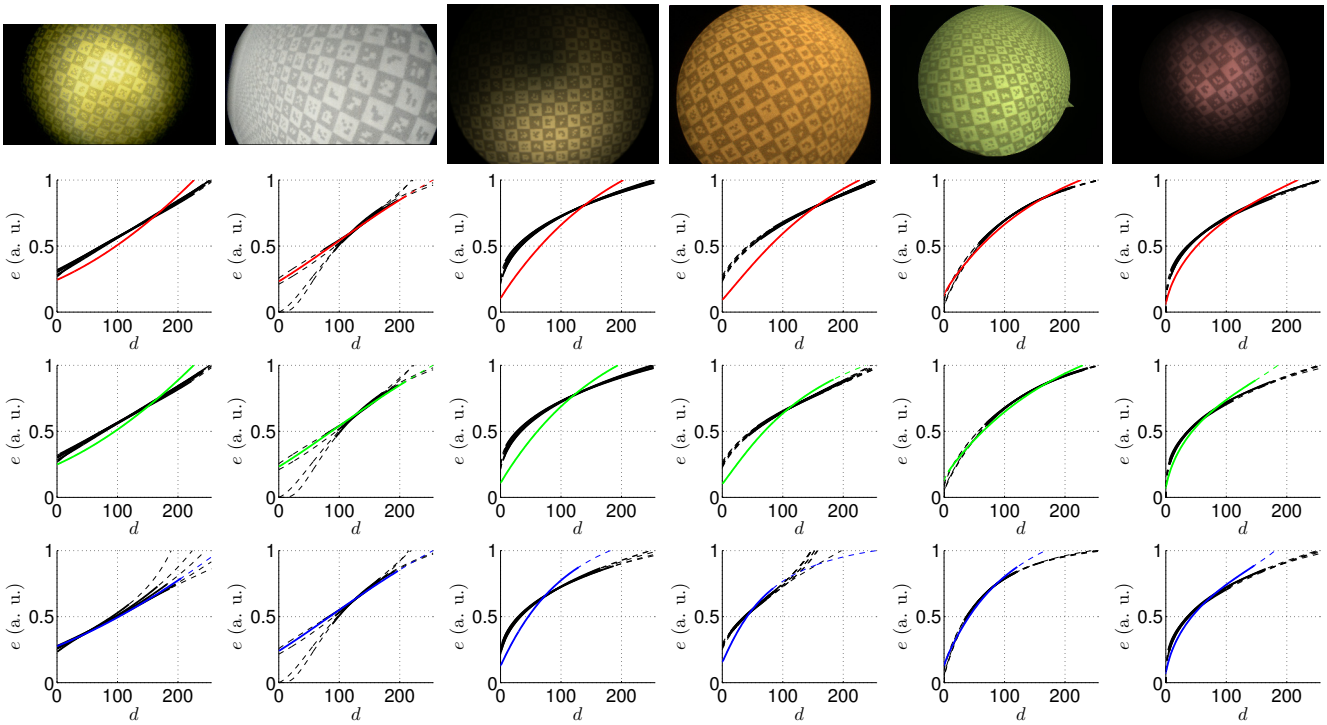
[1] B. Atcheson, F. Heide, and W. Heidrich. CALTag: High Precision Fiducial Markers for Camera Calibration. *Vision, Modeling, and Visualization*, pages 41–48, 2010. 2, 4

[2] J. a. P. Barreto, J. Roquette, P. Sturm, and F. Fonseca. Automatic Camera Calibration Applied to Medical Endoscopy. In *Procedings of the British Machine Vision Conference 2009*, number 20091000069650, pages 52.1–52.10. British Machine Vision Association, 2009. 2

[3] T. Collins and A. Bartoli. 3D reconstruction in laparoscopy with close-range photometric stereo. In N. Ayache, H. Delingette, P. Golland, and K. Mori, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, volume 15 of *Lecture Notes in Computer Science*, pages 634–42. Springer Berlin Heidelberg, Jan. 2012. 1, 2, 3

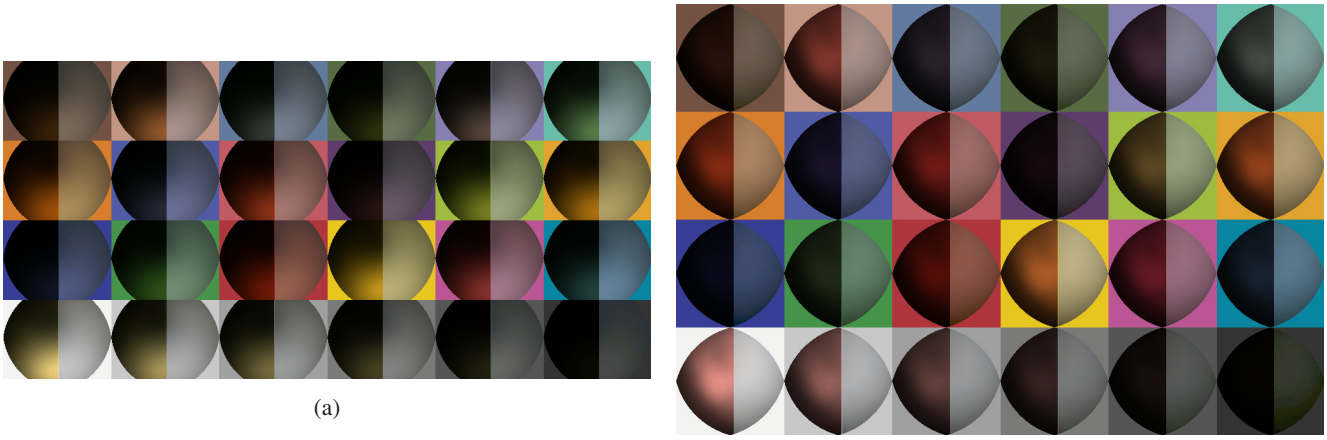
[4] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 1997*, page 1, New York, New York, USA, 1997. ACM Press. 1, 2

[5] M. Díaz and P. Sturm. Radiometric calibration using photo collections. In *2011 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, Apr. 2011. 1, 2



(a) R1 (b) R2 (c) R3 (d) R4 (e) R5 (f) R6

Figure 5: Each column shows an example of a CALTag grid image acquired with the respective rig, and the resulting camera response functions for the red, green, and blue channels (from top to bottom). Each plot shows the result for the Wu *et al.* approach in color, and the results for five CALTag images using our approach. The dashed segment of the curves are the extrapolated regions.



(a) (b)

Figure 6: Examples of the correction of the camera response function estimated with a single CALTag grid using rigs (a) R3 and (b) R6. Each rectangle has the background with the ground truth color from the ColorChecker and two insets with the original image of the color patch (left) and its corrected version (right).

[6] N. Gonçalves, D. Roxo, J. a. P. Barreto, and P. Rodrigues. Perspective shape from shading for wide-FOV near-lighting endoscopes. *Neurocomputing*, 150:136–146, Oct. 2015. 1

[7] M. D. Grossberg and S. K. Nayar. Modeling the space of camera response functions. *IEEE transactions on pattern analysis and machine intelligence*, 26(10):1272–82, Oct.

2004. 6

[8] S. J. Kim and M. Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE transactions on pattern analysis and machine intelligence*, 30(4):562–76, Apr. 2008. 1, 2

[9] E. Y. Lam and G. S. K. Fung. Automatic White Balancing in Digital Photography. In R. Lukac, editor, *Single-Sensor Imaging: Methods and Applications for Digital Cameras*, chapter 10, pages 267–94. CRC Press, 2009. 1

[10] S. Lin and S. Yamazaki. Radiometric calibration from a single image. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages 938–945. IEEE, 2004. 1, 2, 3

[11] A. Litvinov and Y. Y. Schechner. Addressing Radiometric Nonidealities: A Unified Framework. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 52–59. IEEE, 2005. 4

[12] A. Malti and A. Bartoli. Combining conformal deformation and Cook-Torrance shading for 3-D reconstruction in laparoscopy. *IEEE transactions on bio-medical engineering*, 61(6):1684–92, June 2014. 1

[13] Y. Matsushita and S. Lin. Radiometric Calibration from Noise Distributions. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. Ieee, June 2007. 1, 2

[14] R. Melo, J. a. P. Barreto, and G. Falcão. A new solution for camera calibration and real-time image distortion correction in medical endoscopy-initial technical evaluation. *IEEE transactions on bio-medical engineering*, 59(3):634–44, Mar. 2012. 2, 4

[15] T. Mitsunaga and S. Nayar. Radiometric self calibration. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, pages 374–380. IEEE Comput. Soc, 1999. 1, 2, 7

[16] T.-T. Ng, S.-F. Chang, and M.-P. Tsui. Using Geometry Invariants for Camera Response Function Estimation. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, Minneapolis, MN, June 2007. IEEE. 1, 2, 6

[17] E. Prados and O. Faugeras. Shape from shading. In N. Paragios, Y. Chen, and O. Faugeras, editors, *Handbook of Mathematical Models in Computer Vision*, pages 375–88. Springer, 2006. 1

[18] Y.-W. Tai, X. Chen, S. Kim, S. J. Kim, F. Li, J. Yang, J. Yu, Y. Matsushita, and M. S. Brown. Nonlinear camera response functions and image deblurring: theoretical analysis and practice. *IEEE transactions on pattern analysis and machine intelligence*, 35(10):2498–512, Oct. 2013. 1

[19] B. Wilburn, H. Xu, and Y. Matsushita. Radiometric calibration using temporal irradiance mixtures. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7. Ieee, June 2008. 1, 2

[20] C. Wu, S. G. Narasimhan, and B. Jaramaz. A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes. *International Journal of Computer Vision*, 86(2-3):211–228, Feb. 2010. 1, 3, 6, 7