Computer Vision and Image Understanding
journal homepage: www.elsevier.com

Photometric camera characterization from a single image with invariance to light intensity and vignetting

Pedro M. C. Rodrigues[a,**], João P. Barreto[a,b], Michel Antunes[b]

[a] *Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal*
[b] *Perceive 3D, Coimbra, Portugal*

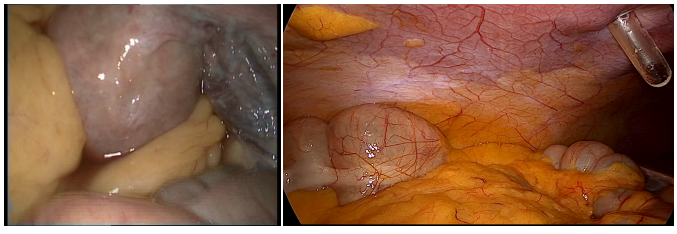
ABSTRACT

Photometric characterization of a camera entails describing how the camera transforms the light reaching its sensors into an image and how this image can be defined in a standard color space. Although the research in this area has been extensive, the current literature lacks practical methods designed for cameras operating under near light. There are two major application scenarios considered in this paper that would benefit from this type of approaches. Camera rigs for minimally-invasive-procedures cannot be calibrated in the operating room with the current methods. This comes from the fact that existing approaches need multiple images, assume uniform lighting, and/or use over-simplistic camera models, which does not allow for the calibration of near light setups in a fast and reliable way. The second scenario refers to the calibration of cellphone cameras, which currently cannot be calibrated at close range with a single image, specially if the flash is used, as there would be non-uniform lighting on the scene. In this work, we describe a method to characterize cameras from a single image of a known target. This enables both geometric and photometric calibrations to be performed on-the-fly without making assumptions on the vignetting nor on the spatial properties of the light. The presented method showed good repeatability and color accuracy even when compared to multiple-image approaches. Applications to laparoscopic cameras, generic cameras (such as cellphone cameras), and cameras other than trichromatic are shown to be viable.

1. Introduction

Camera response characterization determines how a camera transforms incoming light to create pixel values on a digital image. It has been a research topic for a few decades and its relevance is increasing because of the proliferation of computer vision algorithms in this domain. Applications include augmented reality, 3D reconstruction by shape-from-shading (Gonçalves et al. (2015)), and color normalization across camera vendors (Seon Joo Kim et al. (2012)).

The most simple form of camera response characterization is the estimation of what is usually known as the camera response function (CRF). This function is a camera



(a) (b)

Fig. 1: Images of similar anatomical cavities taken from two endoscopes attached to camera control units from different vendors. Note the difference in the color response of the two cameras on tissue with the similar color properties.

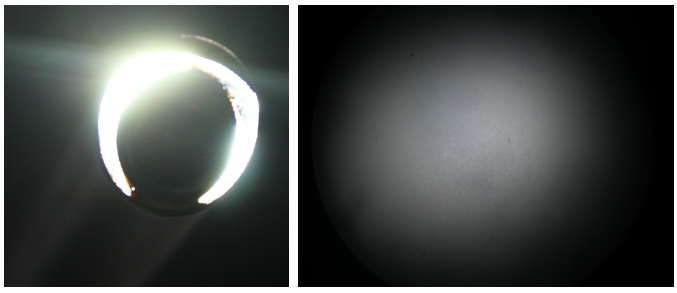
**Corresponding author:
e-mail: prodrigues@isr.uc.pt (Pedro M. C. Rodrigues)

model that, for each image channel independently, maps the value of the incoming light to pixel values. Currently, most methods in the literature only address this issue. The CRF may be adequate in some cases, such as for RAW images, where the only mapping is the light sensors response. However, for most cameras, the CRF does not fully describe the transformation (Seon Joo Kim et al. (2012)). It aims to linearize the response of each color channel independently, but does not describe the color transformations performed in-camera nor the color space of the camera. To fully describe a camera one must estimate a more complex model. Each vendor implements in-camera its own closed-source algorithm. See figure 1 for two images of similar scenes taken from two cameras with different responses. Even if two vendors follow the same color standard (*e.g.*, sRGB) the resulting images can be quite different. Thus, an universal camera model can be hard to establish. Several models have been proposed in the literature. Recently, an insightful work by Seon Joo Kim et al. (2012) aimed to establish a more accurate camera model. Although it requires a large amount of images with the same camera, it was able to establish a better model for the in-camera processing, *i.e.*, from RAW values to RGB pixel values. The proposed model entails a CRF, a gamut mapping function, and a color transformation matrix. It also proved that the in-camera processing is not scene dependent as had been suggested in Chakrabarti et al. (2009). Missing from the Kim *et al.* model is the mapping from the actual incoming light on the sensor to the RAW camera-specific space. Describing the incoming light is usually done in a standard color space and the mapping between a camera-specific color space and a standard color space can be referred to as photometric or colorimetric calibration.

With the present work and in contrast with the work of Seon Joo Kim et al. (2012), we aim to achieve single-frame characterization, including photometric calibration, of digital cameras with generic vignetting and acquired under a generic illuminant. Currently, this is an unsolved problem in the literature. Papers addressing calibration in these conditions usually employ simple camera models with assumptions that do not hold in practice. Moreover, works using more accurate models, assume light is uniform on the scene and that vignetting either can be ignored or has a radial behavior.

The set-up we address in this work is of particular interest in minimally-invasive procedures (MIP), where the previous assumptions do not hold. Moreover, cameras designed for MIP are composed of inter-changeable components (camera, optics, and light). Thus, a medical professional must be able to make the calibration as the parts are mounted, on-the-fly, in the operating room (OR). Hence, the desire to develop a single-image method. Note that indoor calibration of generic cameras, such as cellphone cameras, also tend to violate the uniform illuminant assumption.

This work is an extension of what is described in Rodrigues and Barreto (2015). As in that work, we propose a method that uses a surface with multiple regions of different colors as a calibration target, for which the albedo values are known in advance. Although the shape of the calibration surface is irrelevant for the estimation, all the experiments are carried using a planar checkerboard target similar to the one described in Atcheson et al. (2010). The reasons are twofold: the automatic segmentation and identification of the albedo regions is straightforward, and by combining our method with the geometric calibration approach proposed in Melo et al. (2012) it is possible to fully model the camera from a single calibration frame. In addition to the work presented in Rodrigues and Barreto (2015), this document provides a deeper analysis and theoretical insight about the image formation physics, the camera models, and the necessary conditions for the method to work. It also extends the method that only was able to estimate the CRF, to also perform color response characterization. This extension brings two advantages: (a) it relaxes the assumption that the camera color response is in sRGB with positive impact in accuracy/repeatability of the CRF, and (b) it models the color response of the camera allowing for its calibration. Relaxing this assumption and characterizing the camera color response entails estimating a color transformation matrix in addition to the CRF. As a consequence of estimating a more accurate model, the theory behind the approach for dealing with generic lighting setups must be revisited. In addition, the number of albedos needed in the calibration target changes from two albedos to at least three albedos for the usual trichromatic cameras.



(a) (b)

Fig. 2: A photograph of the lens tip of an endoscopic camera (a) and a white planar scene imaged with an endoscopic camera (b).

The current manuscript will present a thorough experimental validation for different set-ups, a study on the influence of the number of images and colors involved in the calibration, and examples of practical applications using real endoscopic imagery. Our method will also be evaluated for non-endoscopic cameras and we will demonstrate its applicability to such cameras. The method of Wu et al. (2010a,b) was identified as being the closest to our work and will be used for comparison purposes. Further information about their approach is given in section 2.2.

1.1. 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 $\mathbf{f}(\mathbf{x}) = (f_x(\mathbf{x}), f_y(\mathbf{x}))^\mathsf{T}$. $f_x(\mathbf{x})$ denotes a scalar function. Sets are denoted by upper-case calligraphic letters, *e.g.*, $\mathcal{L}$.

Indices are denoted as underscripts and descriptive information is provided as superscripts. For example, as one can see in section 3, $\rho_\mathrm{R}^\mathrm{sRGB}$ would denote the albedo of index R (red channel) in the sRGB color space.

1.2. Overview

Section 2 discusses the relevant literature. Section 3 discusses the physics inherent to the image formation along with the camera model that will be used in this work. Section 4 presents the theory and the implementation details of the algorithm proposed to estimate the camera models. Section 5 and 6 provide the experimental results and discussion about this work.

2. Related Work

Literature in classical CRF estimation is abundant and there are some well established methods. However, to the best of our knowledge, there is not a practical method that can cope with set-ups with near light from an illuminant with generic shape and/or generic vignetting (see figure 2). Morevover, most methods work with RAW images, without in-camera software being applied to the acquired pixel values, and thus can use simple camera models that will not generalize to RGB images. Most off-the-shelf cameras, however, cannot export RAW images, making this a very important limitation of these methods. The proposed approach relaxes this assumptions allowing for a more complete camera model.

In this section, we will provide an overview of the classical calibration (CRF estimation) and of the relevant works in the state of the art, and discuss why these methods cannot be applied in the scenarios tackled in this paper. Table 1 provides a summary of the methods and their requirements.

2.1. Conventional Calibration

Typically, CRF estimation is performed using multiple registered images of the same static scene. For this, photo collections, *e.g.*, a panoramic shot, can be used to infer the CRF from the overlapping regions (Díaz and Sturm (2011); Kim and Pollefeys (2008)). In Díaz and Sturm (2011), authors assume a distant light model for the parameterization. As for Kim and Pollefeys (2008), the work assumes that the vignetting is static and has a circular attenuation effect from the lens center. Moreover, the light is assumed to be static relative to the scene, whereas in our case the light moves with the camera.

Another possibility, and the one most often adopted, is using a perfectly static camera and a static scene to acquire images at different exposure values (Debevec and Malik (1997); Mitsunaga and Nayar (1999)), usually by varying the shutter speed. While this is a widely used approach, many off-the-shelf cameras do not allow to manually set the shutter speed and, in MIP, it can be difficult to acquire multiple images in the same pose as these cameras are designed to be handheld and must be sterilized. Seon Joo Kim et al. (2012) estimate a more complex camera model, but their method uses approximately 30 images, including same-pose images with different exposures and camera settings.

Some single-image approaches for CRF estimation for uniform light have also been proposed (Lin and Yamazaki (2004); Ng et al. (2007); Matsushita and Lin (2007); Wilburn et al. (2008)). These methods all have in common the assumption that the light is uniform on the regions being analyzed in each method, and this is hardly the case when dealing with near lighting (see figure 2).

As mentioned before, none of these methods work with more complete camera models. Instead they limit the camera model to the CRF, which either means that the authors are dealing with RAW data or that the cameras are not being fully characterized.

2.2. Near-Light Set-ups

Only a small number of works explicitly deal with near light set-ups. In Wu et al. (2010b), the authors propose a method to estimate the CRF using 24 same-pose same-exposure images each of a single-color patch of known albedo. However, this approach cannot be done on-the-fly in the OR, which is a major disadvantage for application in MIP. It is also not applicable in generic cameras for which the shutter speed cannot be manually set. Moreover, like the conventional calibration methods mentioned above, it has the drawback of using a camera model with just the CRF. The methods proposed in Haneishi et al. (1995); Funt and Bastani (2014); Finlayson et al. (2015) can also cope with near-light set-ups. However, the authors assume the camera response is already in a linear space. Therefore, instead of performing CRF estimation, the authors use a 3 by 3 matrix to remap the RGB values of the camera.

2.3. Other Related Works

Works in the color stabilization field aim to match the color properties of two images with shared content (Vazquez-Corral and Bertalmio (2015); Frigo et al. (2016); Vazquez-Corral and Bertalmío (2014); Vazquez-Corral and Bertalmio (2016)). While this type of approach could potentially be used in our scenario, these works do not aim to characterize the camera. For example, Vazquez-Corral and Bertalmio (2016) estimate what could be perceived as a CRF and Vazquez-Corral and Bertalmío (2014) estimate a 3 by 3 matrix. However, the estimated models do not describe the camera in a canonical color space. Instead, these works aim to match the tonal properties of two or more images, so that the images could be perceived to be

Table 1: State-of-the-art regarding camera response characterization. From left to right: the methods described in the introduction; the number of images necessary; whether the method needs to change or fix specific camera settings; whether it can cope with near lighting setups; the camera model used (CRF – camera response function; M – color mapping; LUT – 3D look-up table); whether it performs photometric calibration. Note that photometric calibration requires the use of a calibration target and that single-image based methods do not require variable camera settings. *Díaz and Sturm (2011); Kim and Pollefeys (2008). †Debevec and Malik (1997); Mitsunaga and Nayar (1999). ‡Lin and Yamazaki (2004); Ng et al. (2007); Matsushita and Lin (2007); Wilburn et al. (2008). §Haneishi et al. (1995); Funt and Bastani (2014); Finlayson et al. (2015).

method	# images	rig settings	near light	model	photometry
photo collections*	multiple	no	no	CRF	no
variable exposure†	multiple (same-pose)	yes	yes	CRF	no
single-image methods‡	single	-	no	CRF	no
photometry§	single	-	yes	M	yes
Wu et al. (2010b)	multiple (same-pose)	yes	yes	CRF	yes
Seon Joo Kim et al. (2012)	multiple (same-pose)	yes	yes	CRF+M+LUT	no
Rodrigues and Barreto (2015)	single	-	yes	CRF	yes
ours	single	-	yes	CRF+M	yes

taken under similar circunstances, *i.e.*, similar lighting and camera settings.

3. Image Formation: Theory and Assumptions

In this section we will discuss the physics of image formation and the theory behind its modeling, as well as the assumptions that are made in this work. For a summary of the assumptions related to the image formation please see section 3.1.

The image formation model can be conceptually divided in two parts: the physics of the incoming light and the camera model. First, let us look at the physics part of the image formation. The sensor irradiance adopted in this work is

$$e(\mathbf{x},\lambda) = v(\mathbf{x})q(\mathbf{x},\lambda)\rho(\mathbf{x},\lambda) \tag{1}$$

as in Forsyth (1990); Goldman (2010), where $\mathbf{x}$ are the pixel coordinates that correspond to the projection of a scene point, λ is the wavelength, v is the vignetting (a combination of natural, optical, mechanical, and pixel vignetting), q is the light component reflected from the scene, and ρ is the albedo (color). For a trichromatic camera, the generic image acquisition process is

$$\mathbf{d}\left(\mathbf{x}\right) = \mathbf{f}\left(\alpha \begin{bmatrix} \int e\left(\mathbf{x},\lambda\right) s_{\mathrm{R}}\left(\lambda\right)\, d\lambda \\ \int e\left(\mathbf{x},\lambda\right) s_{\mathrm{G}}\left(\lambda\right)\, d\lambda \\ \int e\left(\mathbf{x},\lambda\right) s_{\mathrm{B}}\left(\lambda\right)\, d\lambda \end{bmatrix}\right) \tag{2}$$

as in Wu and Allebach (2000); Forsyth (1990), where $\mathbf{f} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is a generic camera model that comprises any non-linearities on the sensors and camera intrinsic processing, $\mathbf{d} = \left(d_{\mathrm{R}}, d_{\mathrm{G}}, d_{\mathrm{B}}\right)^{\mathsf{T}}$ is the acquired image, α is the exposure (a combination of sensor gains, shutter time, and aperture), $s_c\left(\lambda\right)$ is the spectral sensitivity of the sensors on channel $c \in \{\mathrm{R},\mathrm{G},\mathrm{B}\}$. The equations present in this work generalize for other cameras such as grayscale cameras and hyperspectral/multispectral cameras, but for an easier read we will use RGB cameras in this description.

Since we are interested in calibrating the camera and, therefore, transform image $\mathbf{d}$ to a new image in a canonical space, we are actually interested in modeling the inverse of the camera model $\mathbf{f}^{-1}$. Note that $\mathbf{f}$ is not guaranteed to have an inverse. In fact, it does not usually have one. However, $\mathbf{f}^{-1}$ is denoted here as the best possible approximation of the inverse of $\mathbf{f}$.

Furthermore, if one can assume that light component has the same spectrum across all $\mathbf{x}$, then $q\left(\mathbf{x},\lambda\right) = q_{\mathbf{x}}\left(\mathbf{x}\right) q_\lambda\left(\lambda\right)$. In practice, this is the case for a single illuminant with constant spectrum, or for multiple illuminants all with the same spectra. This can also be assumed in a near light scenario where the intensity of the ambient light is negligible when compared to the near lights. Another requirement for this assumption to be case is that the bidirectional reflectance distribution function (BDRF) is the same for all $\mathbf{x}$. In that case, one can write

$$\mathbf{f}^{-1}\left(\mathbf{d}\left(\mathbf{x}\right)\right) = \alpha v\left(\mathbf{x}\right) q_{\mathbf{x}}\left(\mathbf{x}\right) \begin{bmatrix} \int q_\lambda\left(\lambda\right) \rho\left(\mathbf{x},\lambda\right) s_{\mathrm{R}}\left(\lambda\right)\, d\lambda \\ \int q_\lambda\left(\lambda\right) \rho\left(\mathbf{x},\lambda\right) s_{\mathrm{G}}\left(\lambda\right)\, d\lambda \\ \int q_\lambda\left(\lambda\right) \rho\left(\mathbf{x},\lambda\right) s_{\mathrm{B}}\left(\lambda\right)\, d\lambda \end{bmatrix}$$

or

$$\mathbf{f}^{-1}\left(\mathbf{d}\left(\mathbf{x}\right)\right) = \alpha v\left(\mathbf{x}\right) q_{\mathbf{x}}\left(\mathbf{x}\right) \boldsymbol{\rho}^{\mathrm{RAW}}\left(\mathbf{x}\right) \tag{3}$$

where $\boldsymbol{\rho}^{\mathrm{RAW}} = \left(\rho_{\mathrm{R}}^{\mathrm{RAW}}, \rho_{\mathrm{G}}^{\mathrm{RAW}}, \rho_{\mathrm{B}}^{\mathrm{RAW}}\right)^{\mathsf{T}}$ is the albedo of the scene as imaged under the light of spectrum q_λ in the RAW color space, *i.e.*, the color space of the camera sensors or, in other words, the color space before any in-camera processing.

As for the camera model, the adopted $\mathbf{f}^{-1}$ entails a strictly increasing function for each channel $g_c : \mathbb{R} \rightarrow \mathbb{R}$, the inverse CRF, followed by an invertible square matrix M^{m}, the color transformation matrix. Formally,

$$\mathbf{f}^{-1}\left(\mathbf{d}\right) = \mathsf{M}^{\mathrm{m}} \mathbf{g}\left(\mathbf{d}\right) \tag{4}$$

where $\mathbf{g}\left(\mathbf{d}\right) = \left(g_{\mathrm{R}}\left(d_{\mathrm{R}}\right), g_{\mathrm{G}}\left(d_{\mathrm{G}}\right), g_{\mathrm{B}}\left(d_{\mathrm{B}}\right)\right)^{\mathsf{T}}$. While there has been some discussion on the validity of this model (Grundmann et al. (2013); Chakrabarti et al. (2009)), it has been used with good results and has been proven to be valid for most of the range of possible image values (Seon Joo Kim et al. (2012)).

In this work, we are interested in performing photometric calibration which entails the mapping of the color space of the camera with a 3 by 3 matrix M^{s} that converts the values in a standard space related to the standard human observer, such as the sRGB color space, to the camera space. This matrix cannot fully describe this transformation (Luther-Ives condition – Vazquez-Corral et al. (2014)). However, it is a widely made approximation that is enough for practical purposes. Such matrix could also have other dimensions and, in fact may not be a square matrix, *e.g.*, the case where one wants to map a trichromatic camera to a single-channel space or map a hyperspectral camera to a trichromatic space. However, throughout this work only square matrices are considered. Let us rewrite equation (3) as

$$\mathbf{f}^{-1}\left(\mathbf{d}\left(\mathbf{x}\right)\right) = u\left(\mathbf{x}\right)\boldsymbol{\rho}^{\mathrm{RAW}}\left(\mathbf{x}\right) \tag{5}$$

where $u\left(\mathbf{x}\right)$, what we will refer to as the albedo-normalized irradiance, is composed of exposure, vignetting, and the light effect on the scene: $u\left(\mathbf{x}\right) = \alpha v\left(\mathbf{x}\right) q_{\mathbf{x}}\left(\mathbf{x}\right)$. Using the camera model from equation (4), the color standardization matrix as

$$\boldsymbol{\rho}^{\mathrm{sRGB}} = \mathsf{M}^{\mathrm{s}}\boldsymbol{\rho}^{\mathrm{RAW}}, \tag{6}$$

and defining the color transformation matrix as $\mathsf{M} = \mathsf{M}^{\mathrm{s}}\mathsf{M}^{\mathrm{m}} = \left(\mathbf{m}_{\mathrm{R}}, \mathbf{m}_{\mathrm{G}}, \mathbf{m}_{\mathrm{B}}\right)^{\mathsf{T}}$ one can finally write the image formation model as

$$\mathsf{M}\mathbf{g}\left(\mathbf{d}\left(\mathbf{x}\right)\right) = u\left(\mathbf{x}\right)\boldsymbol{\rho}^{\mathrm{sRGB}}\left(\mathbf{x}\right) \tag{7}$$

in a standard color space, in this case sRGB.

Figure 3 provides a schematic overview of the image formation and of the calibration processes. In this figure, we show the light of an illuminant being reflected from the scene and entering the camera to form an image. We show a schematic representation of how the equations change at each step and how $\boldsymbol{\rho}^{\mathrm{RAW}}$, $\boldsymbol{\rho}^{\mathrm{CAM}}$, $\boldsymbol{\rho}^{\mathrm{sRGB}}$ relate to each other. We also show how the calibration is used to create a new corrected image $\mathbf{d}'$ (dashed arrow).

Figure 4 provides a brief qualitative analysis to show how well the camera model that we will be using (*i.e.*, $\mathsf{M}\mathbf{g}(.)$) describes an actual camera. In this figure, we depict a comparison between two images of planar patches of a single albedo, where the images were taken in the same camera pose. Since the albedo is constant on each scene, the variations throughout the image $\mathbf{d}$ are due to the vignetting and the amount of reflected light from the scene, *i.e.*, the albedo-normalized irradiance $u(\mathbf{x})$. Also, since the camera pose and the illuminants are the same for the two images being compared, the scalar function $u(\mathbf{x})$ should be the same for the two images. Therefore, as long as the camera model is valid, isocurves on the two images being compared should be the coincident. They would have different values, but the same position. This is expected because the adopted camera model $\mathbf{f}^{-1}$ does not allow for constant values of $u\left(\mathbf{x}\right)$ to be mapped to non constant values on the image side $\mathbf{d}\left(\mathbf{x}\right)$. To test this hypothesis we represent isocurves of different color patches in the same plot for comparison. If the isocurves are coincident for all pairs of image channels and albedos, the adopted $\mathbf{f}^{-1}$ should be valid. The isocurves represented in figure 4 were defined by searching for pixels of same value on a specific image channel. The left column compares two low-saturation color-patch images to show that they are coincident, even across channels. On the right column we compare a low- with a high-saturation patch image. This shows that the isocurves of the two images are similar but not coincident. This is because, in this case, the camera model starts to be less accurate as the camera software tends to be more complex on the saturated values. Color saturation did not seam to be the only factor, as suggested by Seon Joo Kim et al. (2012), since we were able to find two low-saturation patches with deviations in the isocurves and we could also find a pair of a low- and a high- saturation patches with coincident isocurves.

According to Seon Joo Kim et al. (2012) a more accurate model for the camera must also include a gamut mapping function $\mathbb{R}^3 \rightarrow \mathbb{R}^3$, in the form of a 3D lookup table (LUT). However, we will not explore this model, since our goal is to perform the camera characterization with a single frame and to estimate a full 3D LUT we would need a wide range of data points. For comparison, to model a camera Seon Joo Kim et al. (2012) use 30 images, including same-pose images with different exposures and camera settings.

3.1. Assumptions

In summary, the light component q is assumed to have the same spectrum for all $\mathbf{x}$, and the ambient light is assumed to negligible. This can be experimentally guaranteed by inserting the calibration target within a black box and performing the calibration there.

Regarding the assumptions about the calibration target, the surfaces do not have to be Lambertian. Yet, the BRDF must not change across albedos, which could happen if the surface properties change. In other words, we assume that the material has the same properties throughout the calibration target.

The exact number of colors that are needed for the calibration target varies with the model that is used. If only the CRF needs to be estimated or, in other words, the color transformation matrix M can be assumed to be diagonal, only two colors are needed. Three colors must be used if a full $\mathsf{M}_{3\times3}$ is to be estimated. For a matrix of generic size the number of colors needed is equal to the number of columns of the matrix.

4. Camera Characterization

In this section we will discuss a preliminary work (Rodrigues and Barreto (2015)) and how that work can be extended to estimate both the CRF $\mathbf{g}$ and the color transformation matrix M.

For an easier reformulation for other types of cameras (*e.g.*, different number of channels) and other color spaces, the channels of the acquired image will be indexed by the

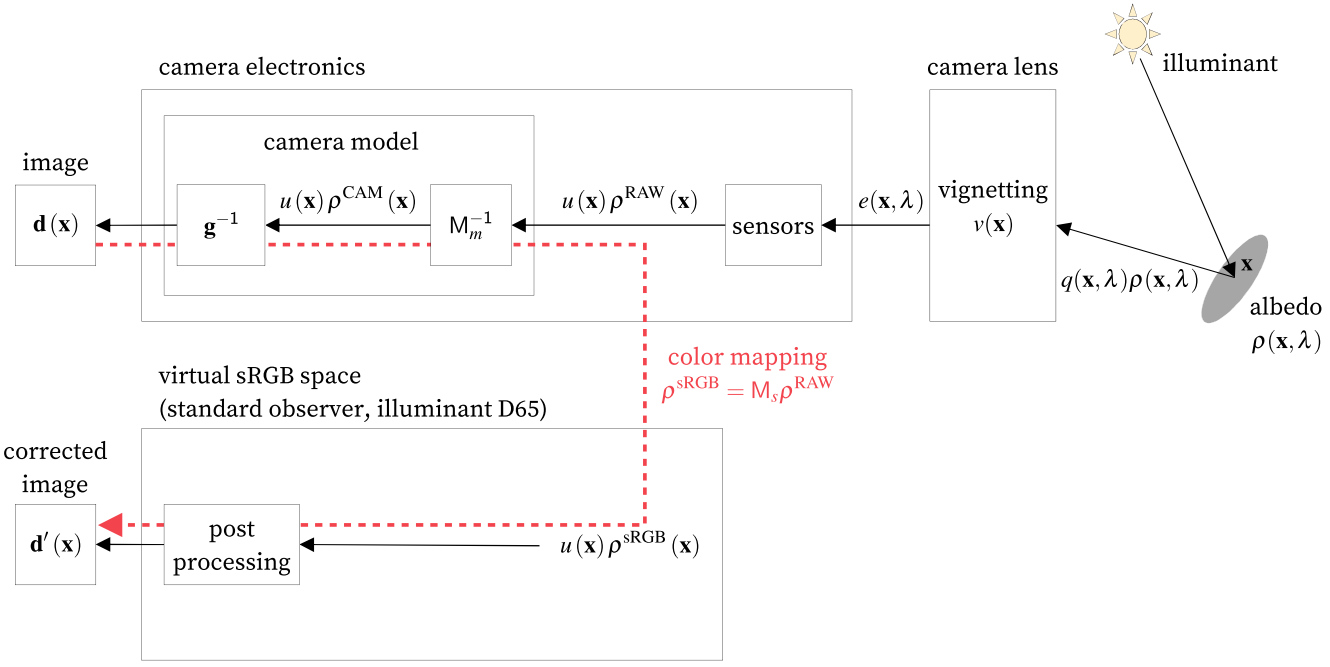


Fig. 3: Schematic representation of the image formation process along with how the calibration procedure goes from the acquired image to the sRGB color space in order to create a corrected image. The calibration is represented by the dashed red arrow.

letter c and the elements of $\boldsymbol{\rho}^{\mathrm{sRGB}}$ (the channels of the corrected image) will be indexed by the letter s. But note that usually $s \equiv c$, because, for most cases, one wants to map from an RGB space to sRGB, which have equivalent number of channels and channel names.

In Rodrigues and Barreto (2015), we propose the use of a single-image of a two-albedo (white and gray) target with known albedos to estimate the CRF. In both the preliminary work and what we propose with this work, we must perform estimation of isocurves, which will be explained in section 4.2. While the preliminary work only uses two sets albedo patches, the method proposed in this work was generalized to include more albedos so that the isocurves of u can be determined more accurately.

4.1. Segmentation

The proposed approach lies on the assumption that we have a scene of a multiple-color surface. In the present work, we have used a planar multiple-color CALTag grid (Atcheson et al. (2010)) geometrically calibrated with the method proposed in Melo et al. (2012). We do not need it to be planar nor a grid for the framework to succeed, but we do need to segment the scene into regions of constant albedo. We have chosen this grid to be able to perform both the geometric and photometric calibrations with a single image. The fact that this type of grid has tags in each square also allows for easier localization of each specific square and to ensure that a specific tag corresponds to a specific albedo.

Having used the method proposed in Melo et al. (2012) to obtain the geometric calibration of the scene and the camera, 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, we can warp the grid to the image plane. This warped image is itself our segmentation.

A morphological erosion is then performed to the segmented regions of each albedo to avoid problematic albedo boundary regions. On these regions, it is a possibility that the image values are influenced by optical and motion blur, chromatic aberrations, and demosaicing inaccuracies. Segmentation inaccuracies could also be another problematic factor on boundary regions.

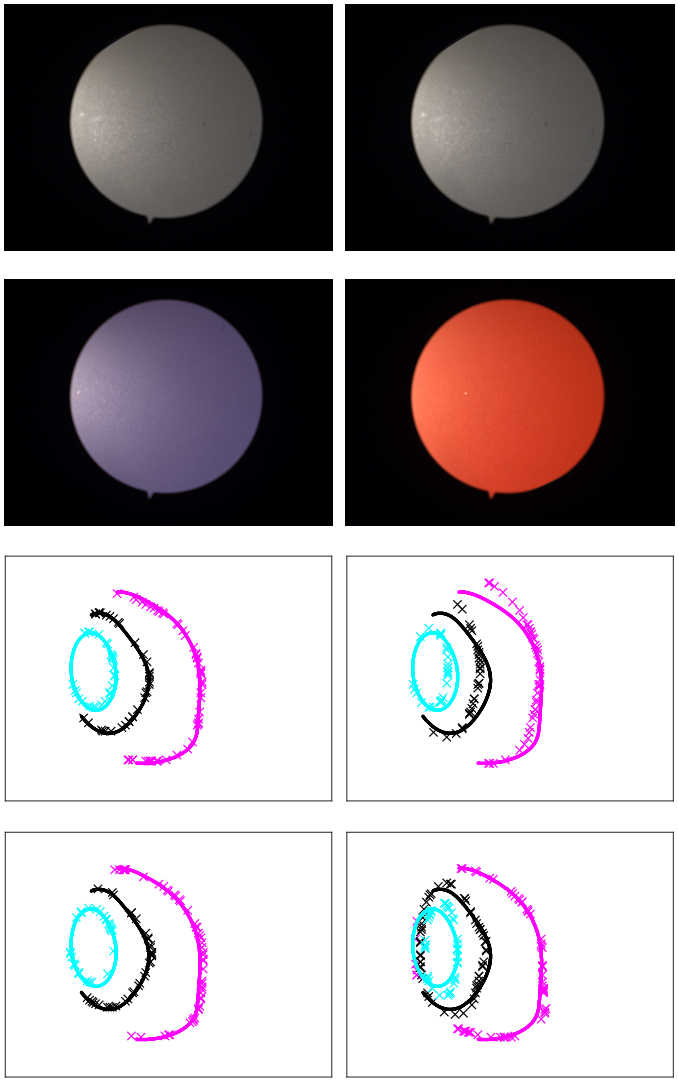
Figure 5 shows a CALTag grid and its segmentation.

4.2. Isocurves Estimation

To use a single image for camera response characterization, we seek to find, within the same image, pairs of pixels where one could write equations invariant to the vignetting and the light effect. 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 $u(\mathbf{x})$. 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 on the scene. 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 (see figure 2), and may not be isotropic (Collins and Bartoli (2012)).

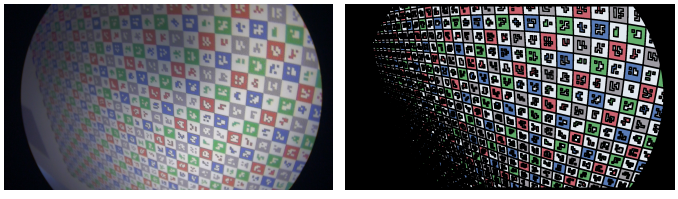
Let us define the albedos in yet another color space, the color space of the image $\mathbf{d}$, as $\boldsymbol{\rho}^{\mathrm{CAM}}$ such that

$$\boldsymbol{\rho}^{\mathrm{RAW}} = \mathsf{M}^{\mathrm{m}}\boldsymbol{\rho}^{\mathrm{CAM}}\,. \tag{8}$$



(a) low S (b) low S *vs.* high S

Fig. 4: Comparison of the isocurves of $d_c(\mathbf{x})$ on real images acquired using endoscopic lens. The columns show the comparison between: (a) two low-saturation color patches, and; (b) a low-saturation and a high-saturation patches. From top to bottom: the first and second color patches being compared, the comparison between isocurves of the green channels of the two patches (lines — first image; crosses — second image), and the comparison between the isocurves of the green channel of the first patch with the red channel of the second patch (lines — first image; crosses — second image). See text for additional insight.



(a) $\mathbf{d}(\mathbf{x})$ (b) $\boldsymbol{\rho}^{\mathrm{sRGB}}(\mathbf{x})$

Fig. 5: An image of a 5-color calibration grid (a) and the corresponding segmentation into regions of constant albedos (b) color-coded with the true sRGB values. Black pixels occur when albedo information is not available.

This color space differs from the RAW color space due to in-camera processing (see figure 3).

The particular case where the color transfomation matrix M^{m} is approximated by a diagonal matrix has been addressed in Rodrigues and Barreto (2015), and we will build on this approach to eliminate this assumption.

For an easier understanding of the equations that follow, it is helpful to think of $u(\mathbf{x})$ as being the image that reaches the camera sensors but without any color information (whether from light sources or from the actual scene). By definition, an isocurve of $u(\mathbf{x})$ will contain points $\mathbf{x}$ with the same value of $u(\mathbf{x})$. Note that the points in an isocurve of $u(\mathbf{x})$ may correspond to pixels with different values in the actual image $d(\mathbf{x})$ since albedo information is present there. With this in mind and using equations (6), (7) and (8), if M is invertible, one can write

$$g_c(d_c(\mathbf{x})) = u(\mathbf{x})\,\rho_c^{\mathrm{CAM}}(\mathbf{x}) \tag{9}$$

and, on the ith isocurve of u,

$$u(\mathbf{x}_j) = \kappa_i = \frac{g_c(d_c(\mathbf{x}_j))}{\rho_c^{\mathrm{CAM}}(\mathbf{x}_j)}, \quad j \in \mathcal{L}_i \tag{10}$$

where $\mathcal{L}_i$ is the set of pixels crossed by isocurve i and κ_i is a constant (the value of the isocurve). Such is true for all c and for whatever albedo is crossed by the isocurve. If a curve i passes through multiple albedos one will have, for an albedo pair, $\boldsymbol{\rho}_n$ and $\boldsymbol{\rho}_{n'}$,

$$\rho_{c',n'}^{\mathrm{CAM}} g_c(d_c(\mathbf{x}_j)) = \rho_{c,n}^{\mathrm{CAM}} g_{c'}(d_{c'}(\mathbf{x}_k)), \\ j \in \mathcal{L}_i \cap \mathcal{A}_{\boldsymbol{\rho}_n}, k \in \mathcal{L}_i \cap \mathcal{A}_{\boldsymbol{\rho}_{n'}} \tag{11}$$

where $\mathcal{A}_{\boldsymbol{\rho}_n}$ is the set of points with a specific albedo $\boldsymbol{\rho}(\mathbf{x}) = \boldsymbol{\rho}_n$. Equation (11) will then be used for the single-image CRF estimation. Note that $c, c' \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$ and thus c and c' are both channel indices. c' is used to show that the equations can be written by combining different channels or different albedos.

See figure 6 for more intuition on the image formation process and how it relates to the isocurves of $u(\mathbf{x})$. This figure shows a synthetic grayscale image and each individual component that form the image. Figure 6(a) shows how the vignetting $v(\mathbf{x})$ and the light component $q(\mathbf{x})$ combine to form the albedo-normalized irradiance $u(\mathbf{x})$. It then shows what $u(\mathbf{x})$ combined with the albedo looks like and, finally, what a CRF does to that combination to create the final image. Figure 6(b) shows the same process for an isocurve of $u(\mathbf{x})$.

Since $u(\mathbf{x})$ is not known, we need to evaluate how its isocurves behave on the image $\mathbf{d}(\mathbf{x})$. From (10), it is clear that, for a given albedo, an isocurve in the sensor irradiance is also an isocurve in the image $d(\mathbf{x})$. In addition, along an isocurve of $u(\mathbf{x})$, $\mathcal{L}_i$, the image values form a piecewise constant function (with a different constant value for each albedo).

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

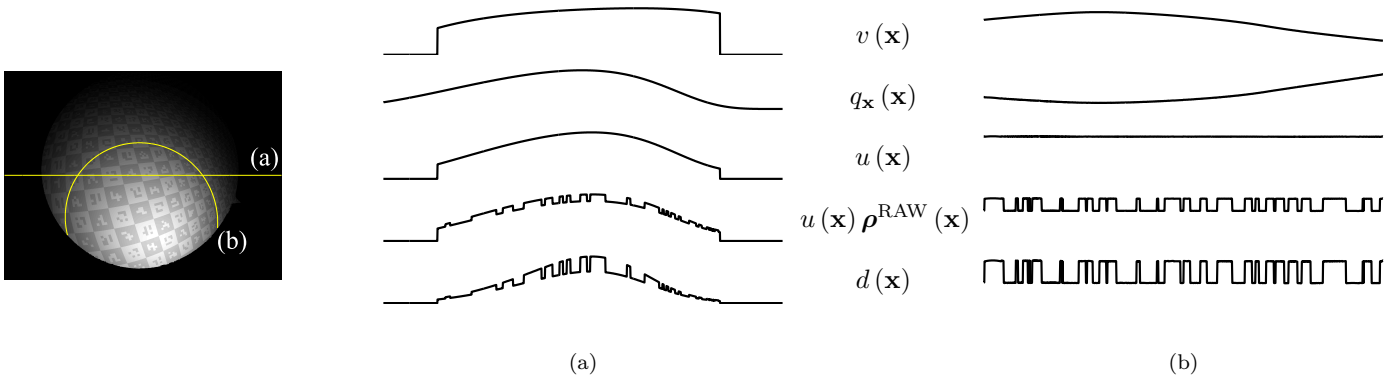


Fig. 6: Values (arbitrary units) of each image component showing the formation process of an image of a checkerboard along (a) a horizontal line and (b) an isocurve of the albedo-normalized irradiance $u(\mathbf{x})$. For this example we used a synthetic grayscale image of a two-albedo checkerboard. Please refer to the text, specifically equations (3) and (5), for details on each image component.

value. The same is true for the one albedo and different image color channels. Thus, to find the isocurves of $u(\mathbf{x})$, it is reasonable to fit a single surface $\mu_{c,n}$ to the image along one albedo $\boldsymbol{\rho}_n$ and channel c. If there is a single color suitable to fit μ across the whole calibration target, then a single color may be used. However, for better results, all albedos and all channels can be used.

Let us approximate the image along one of the albedos n and channel c by a generic model $\mu_{c,n}$ where the isocurves are known or can easily be extracted. We can write for two albedos on the image space

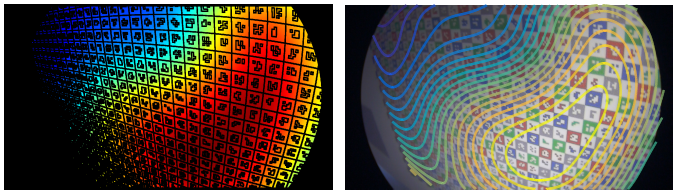
$$d_c(\mathbf{x}_j) \sim \mu_{c,n}(\mathbf{x}_j), \qquad j \in \mathcal{A}_{\boldsymbol{\rho}_n} \tag{12}$$
$$d_{c'}(\mathbf{x}_k) \sim \mu_{c',n'}(\mathbf{x}_k), \qquad k \in \mathcal{A}_{\boldsymbol{\rho}_{n'}}\,. \tag{13}$$

From before, we know that the isocurves of $\mu_{c,n}(\mathbf{x})$ and $\mu_{c',n'}(\mathbf{x})$ 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 isocurves varies from one formulation to the other, but the isocurves are the same. One can show that the two models are related by

$$d_{c'}(\mathbf{x}_k) \sim \mu_{c',n'}(\mathbf{x}_k) \tag{14}$$
$$g_c^{-1}\left(\frac{\rho_{c,n}^{\mathrm{CAM}}}{\rho_{c',n'}^{\mathrm{CAM}}} g_{c'}(d_{c'}(\mathbf{x}_k))\right) \sim \mu_{c,n}(\mathbf{x}_k) \tag{15}$$
$$h_{c,c',n,n'}(d_{c'}(\mathbf{x}_k)) \sim \mu_{c,n}(\mathbf{x}_k), \quad k \in \mathcal{A}_{\boldsymbol{\rho}_{n'}} \tag{16}$$

where $h_{c,c',n,n'}$ is a positive and monotonically increasing function that is used to transform the model $\mu_{c',n'}$ into the model $\mu_{c,n}$. This function h is the equivalent of having a gain for each isocurve value for the points of the albedo $\boldsymbol{\rho}_{n'}$ on channel c', to be able to use only the model $\mu_{c,n}$ for both albedos using equations (12) and (16).

We have used a polynomial model as $\mu_{c,n}$, *i.e.*, $\mu_{c,n}(\mathbf{x};\mathbf{p}) = \hat{\mathbf{x}}^{\mathsf{T}}\mathbf{p}$, where $\hat{\mathbf{x}} = \left(1, x_1, x_2, x_1^2, x_1x_2, x_2^2, \dots\right)^{\mathsf{T}}$ and $\mathbf{p}$ are the polynomial coefficients. The isocurves can then be extracted as the level sets of the polynomial. This is done by searching the polynomial function for points $\mathbf{x}$ where the polynomial is constant.



(a) (b)

Fig. 7: A depiction of (a) the estimated $\mu_{c,n}(\mathbf{x})$ (a surface with the same isocurves as the albedo-normalized irradiance $u(\mathbf{x})$) and (b) the calibration grid image superposed with some of the isocurves.

As for the linear system of equations to be solved, for predefined c and n (*e.g.*, green channel and white albedo), it is defined as

$$\forall j \in \mathcal{A}_{\boldsymbol{\rho}_n}, \forall c', \forall n', \forall k \in \mathcal{A}_{\boldsymbol{\rho}_{n'}}:$$
$$\begin{bmatrix} \hat{\mathbf{x}}_j^{\mathsf{T}} & \mathbf{0}^{\mathsf{T}} \\ \hat{\mathbf{x}}_k^{\mathsf{T}} & -\mathbf{s}^{\mathsf{T}}(d_{c'}(\mathbf{x}_k)) \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \mathbf{h}_{c,c',n,n'} \end{bmatrix} = \begin{bmatrix} d_c(\mathbf{x}_j) \\ 0 \end{bmatrix} \tag{17}$$

where $\mathbf{h}$ is a discrete array version of the function h and $\mathbf{s}(n)$ is sparse vector with a single value of 1 on the element n. This is solved by minimization of the least squares with quadratic programming. We also constrain the various instances of $\mathbf{h}$ to be monotonically increasing by introducing additional equations that force the local derivatives to be positive. The polynomial used was of the 6th order (28 parameters). However, this can be altered to meet the requirements of other calibration scenes.

An example of the determined isocurves can be observed in figure 7.

4.3. Model Estimation

Having determined the isocurves, to estimate the camera model one would want to minimize the residuals derived from equation (11), and equations (6) and (8), as

$$\forall i, \forall s, \forall s', \forall n, \forall n': e_{i,n,n',s,s'} =$$
$$\rho_{s',n'}^{\mathrm{sRGB}} \mathbf{m}_s^{\mathsf{T}} \mathbf{g}\left(\mathbf{d}_{i,\boldsymbol{\rho}_n}\right) - \rho_{s,n}^{\mathrm{sRGB}} \mathbf{m}_{s'}^{\mathsf{T}} \mathbf{g}\left(\mathbf{d}_{i,\boldsymbol{\rho}_{n'}}\right) \tag{18}$$

where $\mathbf{d}_{i,\boldsymbol{\rho}}$ is the median value of the image values on isocurve i and albedo $\boldsymbol{\rho}$. However, this a nonlinear programming problem that can be hard to solve and time-consuming.

In Rodrigues and Barreto (2015), M is assumed to be diagonal and thus the problem can be estimated per channel using convex optimization by minimizing residuals of the form

$$\forall i, \forall s, \forall n, \forall n' : e_{i,n,n',s} = \rho^{\mathrm{sRGB}}_{s,n'} g_c\left(d_{c,i,\boldsymbol{\rho}_n}\right) - \rho^{\mathrm{sRGB}}_{s,n} g_c\left(d_{c,i,\boldsymbol{\rho}_{n'}}\right) \quad (19)$$

with $s = c$ as $\boldsymbol{\rho}^{\mathrm{sRGB}} = \boldsymbol{\rho}^{\mathrm{CAM}}$.

In this work, we propose the estimation of the CRF and M in an iterative process of two convex-optimization steps. The CRF is initially assumed to be linear, *i.e.*, $\forall c : g_c(d) = d$. Then, in each iteration, M is computed in a convex optimization framework using equations derived from equation (7) as

$$\forall x : \left[\boldsymbol{\rho}^{\mathrm{sRGB}}(x)\right]_{\times} \mathsf{M}\mathbf{g}\left(\mathbf{d}(x)\right) = \mathbf{0} \quad (20)$$

where $[\cdot]_{\times}$ denotes the skew-symmetric matrix. This is followed by the estimation of $\mathbf{g}$, which can be done with equations very similar way to equation (19). However, for robustness we have decided to introduce inter-channel equations. In other words, instead of only combining equations for different albedos (where $u(\mathbf{x})$ is constant) to exclude $u(\mathbf{x})$ from the equations, we combined equations from different channels as well.

The albedos are updated using equation (8) and the values used for M are the ones computed on the previous step. The residuals to minimize then become

$$\forall i, \forall c, \forall c', \forall n, \forall n' : e_{i,n,n',c,c'} = \rho^{\mathrm{CAM}}_{c',n'} g_c\left(d_{c,i,\boldsymbol{\rho}_n}\right) - \rho^{\mathrm{CAM}}_{c,n} g'_c\left(d_{c',i,\boldsymbol{\rho}_{n'}}\right). \quad (21)$$

At the end of each iteration, a metric is computed to ensure the method is converging to a solution. The metric used was the average angle between the albedo vectors $\boldsymbol{\rho}^{\mathrm{sRGB}}$ and their respective corrected image pixel $\mathsf{M}\mathbf{g}(\mathbf{d})$, which should be co-linear. The angle θ, as computed between the vectors $\boldsymbol{\rho}^{\mathrm{sRGB}}$ and $\mathsf{M}\mathbf{g}(\mathbf{d})$, was used instead of an euclidean distance or a traditional color metric because the estimation is performed up-to-scale due to $u(x)$. This iterative process continues until there is no longer improvement on the metric or a fixed number of iterations is reached (50 iterations).

This approach is scalable with additional images. Even images with different exposures, different poses, and changes in the vignetting (due to changes of zoom and/or aperture) can be used to augment the number of equations. Additional images would provide additional equations that can be grouped together to potentially achieve a more robust estimation.

There is no need for both the light to be non-uniform on the scene and the vignetting to be strong. Either one is sufficient and, while the method benefits from these effects, neither is mandatory. If the calibration albedos cover the range of possible values, good results can be obtained as long as there is a slight variability enough for isocurves to be obtainable (for instance, a scene under sunlight with a faint shadow).

4.4. Parametrization

Since no assumption is made on the form of the CRF $\mathbf{g}$ with the isocurves approach, one is not bound to a specific CRF model. In fact, equation (21) can rewritten using specific parametrization for each g_c (*e.g.*, polynomial) or in a non-parametric way. In this work we use a non-parametric formulation for the CRF $\mathbf{g}$

The isocurves may not define equations for all values in the 0-255 range of image pixel values for the CRFs $g_c(d_c)$. Therefore, after estimating the camera model with non-parametric CRFs, these must be handled to fill values that might be missing, especially in the larger and lower values of this range. With this in mind, the CRF is parametrized with the Empiric Model of Response (EMoR) (Grossberg and Nayar (2004)), a linear basis model obtained by using principal component analysis in real CRFs.

Please check algorithm 1 for a summary of the steps of the camera characterization procedure.

Algorithm 1 Camera characterization algorithm.

1: image acquisition
2: geometric calibration
3: segmentation
4: estimation of $h(\mathbf{x})$ (equation (17))
5: level sets
6: 2-step estimation of M and $\mathbf{g}$, the camera model (equations (20) and (21))
7: parametrization of $\mathbf{g}$

5. Experimental Validation

For the experimental validation we evaluated the color accuracy, CRF accuracy, and CRF repeatablity using endoscopic and non-endoscopic set-ups.

Three endoscopic datasets were acquired along with a cellphone camera and a monochromatic camera to showcase the applicability to generic cameras and single-channel cameras:

1. PointGrey Flea3 (FLIR Systems, Wilsonville, OR) CMOS camera with a 30° Stryker (Stryker Corporation, Kalamazoo, MI) endocopic lens and a Smith and Nephew (Smith and Nephew plc, United Kingdom) light source connected through the built-in light guide
2. PontGrey Grasshopper2 CCD camera with a laparoscopic lens and an external lamp light
3. Sentech (Sentech co. ltd., Kanagawa Prefecture, Japan) CMOS camera with a 30° Dyonics (Smith and Nephew plc) endocopic lens and a light source connected trough the built-in light guide

4. Cellphone (Samsung S6 Edge - Samsung Electronics co. ltd., South Korea) CMOS camera with the built-in LED flash turned on
5. PointGrey Dragonfly monochromatic CCD camera with a lamp as the light source

The CALTag checkerboard albedos were obtained with a iWave WR10 (iWave Systems Technologies, Bangalore, India) colorimeter.

Throughout the experimental results we compare the method proposed in this work (referred to as **isocurves-MCRF**) with the method presented in Rodrigues and Barreto (2015) (**isocurves-CRF**) and a fourth method (**direct-M**) where only M is estimated according to equation (20). The latter is essentially the first step of the iterative 2-step optimization proposed before. In it the CRF is assumed to be linear and thus no isocurves are necessary.

We also compare our methods to our implementation of the Wu et al. (2010b) approach, which requires the acquisition of same-pose same-exposure images of multiple single-albedo patches with known albedo values. Since the camera pose is static and the exposure is fixed, the only quantity that changes between images is the albedo. The authors assume that M is diagonal and they estimate each g_c function independently. To allow the comparison to this work, for each dataset that we acquired, in addition to the CALTag chart images, a sequence of images of the X-Rite (X-Rite, Inc., MI) ColorChart was acquired under the circumstances required by the method.

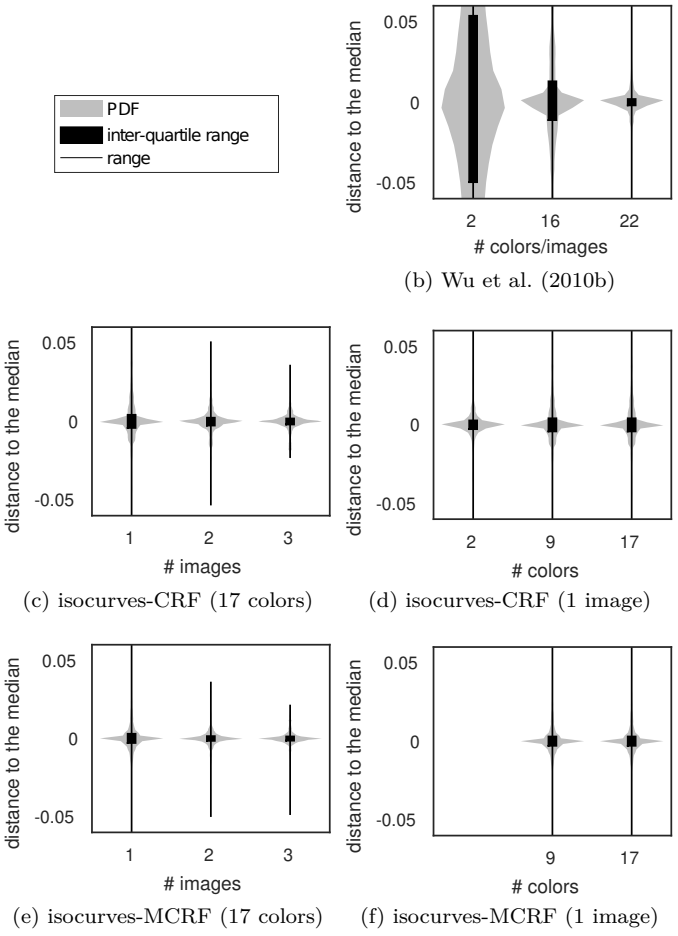
For all methods, the acquisitions were performed with ambient light present. There were no other illuminants close to the calibration scene, but there was indirect sunlight present in the scene.

The computational time of the isocurves-MCRF method was of approximately 12 seconds for 17-albedo 1600x1200 image using a MATLAB/C++ code develop for research porposes only and that has not been optimized for speed.

5.1. Color Accuracy and CRF Repeatability

To assess the repeatability of the CRF estimation we present the histograms of the distance to the median CRF. For such metric, we performed the camera model estimation for all cameras and all poses. Since the camera models are estimated up to a global scale factor, the relative factors must be estimated for a fair comparison. Thus, for each camera, the CRFs of the red, green, and blue channels are aligned across the estimations on different camera poses, by estimating a single global scale factor for each pose. Then, the median CRF across poses is computed and subtracted from the estimated CRFs. The violin plot of figure 8, shows the distributions and boxplots of these differences from all datasets, all poses, all three channels, and all values in the pixel value range. Figure 9 shows the median and the inter-quartile range across different poses for the CRF estimation.

Using less calibration images reduces drastically the repeatability of the Wu *et al.* method, as shown by the dis-



(b) Wu et al. (2010b)

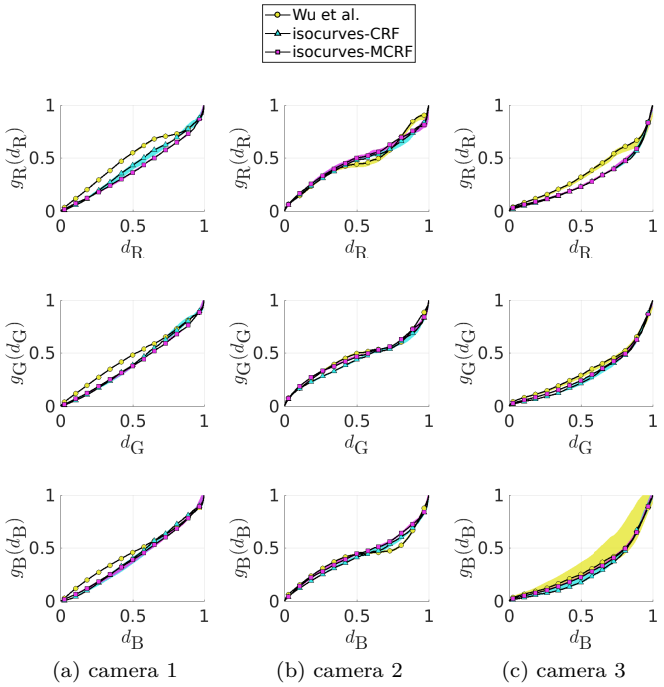
(c) isocurves-CRF (17 colors)

(d) isocurves-CRF (1 image)

(e) isocurves-MCRF (17 colors)

(f) isocurves-MCRF (1 image)

Fig. 8: Repeatability of the CRF estimation for the different approaches. The plots show the distribution and the boxplot of the distance to the median CRF. See text for more details on the metric.



(a) camera 1 (b) camera 2 (c) camera 3

Fig. 9: Median and inter-quartile range for the inverse CRFs for the (top to bottom) red, green, and blue channels of estimated for each dataset. Wu *et al.* approach was computed with 22 colors/images and the single-image approaches were computed with a single image of a 17-color target.

tributions of figures 8 and 9. For our method, these results show that additional images and additional albedo regions only marginally improves repeatability. However, good repeatability is already obtained with a single image and fewer colors.

To evaluate color accuracy we use two metrics: θ, computed as the angle between the albedo vectors $\boldsymbol{\rho}^{\text{sRGB}}$ and their respective corrected image pixel $\mathsf{M}\mathbf{g}(\mathbf{d})$; and $\Delta u'v'$, usually known as chromaticity distance. The latter is the euclidean distance in the $u'v'$-space, which can be obtained directly from the sRGB. Note that more standard color metrics like ΔE^*_{00} (Sharma et al. (2005)) cannot be used because the colors are being compared up-to-scale, due mainly to the near light on the scene. Since the light is not constant on the scene, as it would be in a direct sunlight set-up, the pixel values cannot be compared without estimating a scale factor for each pixel. Another reason is that we are estimating the camera model up-to-scale, therefore at least one global scale factor would need to be further estimated for color comparison. Both the θ and the $\Delta u'v'$ provide comparison between colors with invariance to their "lightness".

Color accuracy was evaluated using the X-Rite ColorChart. The single-albedo images were used for testing the camera calibration performed with the isocurves-MCRF method, the isocurves-CRF method, and the Wu et al. (2010b) method. Wu et al. (2010b) was evaluated using a cross-validation approach. Pixels with values in the lower 10% of the 0-255 range have been discarded as these will be greatly affected by noise and thus will generate metrics that do not necessarily represent the data. Moreover, perceptually, the color mapping of the lower values is less important. For qualitative evaluation of the color accuracy, figure 10 presents the test pixel values after being corrected with the estimated model against the ground truth chromaticity from the X-Rite ColorChart. For sake of comparison, we also show the values of the original image without calibration, *i.e.*, simply assuming that camera is already in sRGB.

Table 2 shows a summary of both the CRF repeatability and the color metrics.

As already shown in figure 8 and confirmed in table 2, in terms of CRF repeatability, the isocurves-MCRF method presented here is able to achieve similar results to the 22-image approach.

Although Wu *et al.* approach linearizes the color space by estimating a CRF, it performs worse, in terms of color accuracy, than the other methods evaluated here. In fact, like the isocurves-CRF method, it even shows to have worse color accuracy than the original image. This happens because, in both methods, M is assumed to be diagonal.

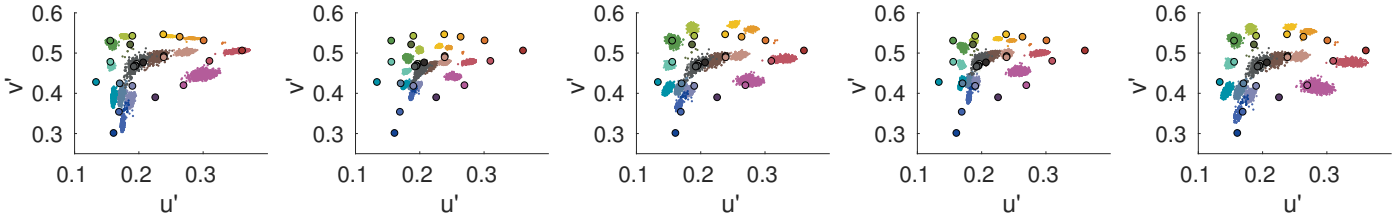
The estimation using the direct-M approach, *i.e.* the estimation of M assuming that $\forall c : g_c$ are linear, is purely color driven. This explains the fact that the method is able to achieve a better color metric. The cost function used for estimation of M comes from equation (20), which allows for the estimation of M to be done by writing equations that only take into consideration color accuracy and do not compare colors at different brightness values. On the one hand, this method is able to benefit from having a cost function that does not favor camera models that better linearize the color space. On the other hand, the method does not linearize the space and, thus, it does not fully describe the camera. Our method does linearize the color space while still obtaining an 20% improvement of chromaticity distance.

To evaluate the repeatability of the method when using general purpose cameras, further tests were performed. Regarding single-channel cameras, to use our method one must assume that the values of ρ^{RAW} are known. While we cannot know in advance the spectrum of the camera and the light, mapping the sRGB albedos to grayscale achieves good repeatability, suggesting that the spectra is not as relevant as in trichromatic cameras with built-in software that perform color mapping. Figure 11 shows the median and the inter-quartile range across different poses for the CRF estimation using generic cameras.

As in the previous repeatability tests, the method is able to achieve good repeatability as the inter-quartile range for the CRF estimation is barely noticeable.

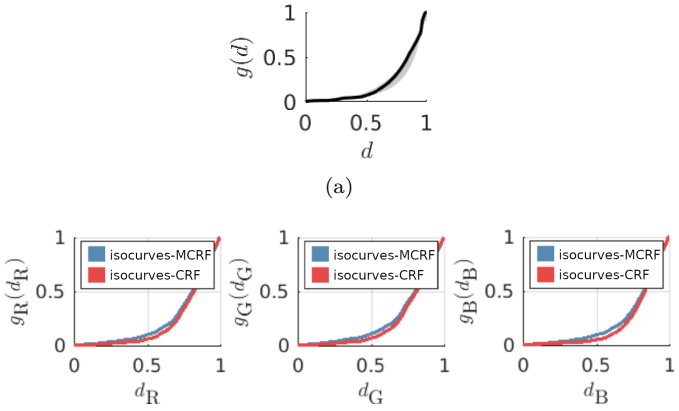
5.2. Brightness Transfer Functions

For this test, we aim to compare our CRFs estimations to the actual CRF the camera. However, theoretically valid evaluation of the CRF against ground truth is non-trivial under near light, since a direct comparison to



(a) original (b) Wu et al. (2010b) (22 patches) (c) direct-M (17 colors) (d) isocurces-CRF (17 colors) (e) isocurves-MCRF (17 colors)

Fig. 10: Ground truth color points and respective corrected image pixel values in the $u'v'$ space.



(a) (b) (c) (d)

Fig. 11: Median and inter-quartile range for the inverse CRFs for generic cameras: (a) monochromatic camera; (b-d) red, green, and blue channels of a cellphone camera. In red is isocurves-CRF method and in blue is the isocurves-MCRF method.

ground truth albedos is not possible. For a fair comparison we used brightness transfer functions (BTFs) and a single-channel camera so the color mapping and the albedo values do not influence the results. From equation (3) we can derive for a monochromatic camera its image formation model as

$$f^{-1}\left(d\left(\mathbf{x}\right)\right) = \alpha v\left(\mathbf{x}\right) q_{\mathbf{x}}\left(\mathbf{x}\right) \rho^{\mathrm{RAW}}\left(\mathbf{x}\right) \tag{22}$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is the single channel CRF. The BTF b between two same-pose single-channel images of different exposures, d_1 and d_2, can be described as

$$d_2\left(\mathbf{x}\right) = f^{-1}\left(\frac{\alpha_2}{\alpha_1} f\left(d_1\left(\mathbf{x}\right)\right)\right) = b\left(d_1\left(\mathbf{x}\right)\right) . \tag{23}$$

By plotting the pixel values of one image against the other one can compare to our estimation of the BTF only by estimating the constant k that minimizes

$$e_{\mathbf{x}} = g\left(d_2\left(\mathbf{x}\right)\right) - k g\left(d_1\left(\mathbf{x}\right)\right) . \tag{24}$$

Figure 12 shows the BTF estimation results for a generic scene imaged with a monochromatic camera with different poses and exposures.

The results show that the estimated BTFs follow the actual BTFs obtained directly from the camera. This gives a good indication that the estimated CRF is close to the true camera CRF.

Table 2: Average metrics across the three datasets. The table shows the average (Avg), standard deviation (SD) and root mean square (RMS) for both the $\Delta u'v'$ and the θ metrics. The average inter-quartile range (IQR) metric used to assess repeatability is also presented.

	# albedos	# images	$\Delta u'v'$		θ		IQR
			Avg±SD	RMS	Avg±SD	RMS	
original	-	-	0.171±0.100	0.199	0.029±0.022	0.037	-
Wu et al. (2010b)	2	2	0.394±0.154	0.426	0.064±0.020	0.068	0.104
	22	22	0.193±0.097	0.217	0.029±0.014	0.032	0.005
direct-M	17	1	0.131±0.077	0.152	0.021±0.014	0.025	-
isocurves-CRF	2	1	0.182±0.118	0.217	0.027±0.019	0.033	0.008
	17	1	0.177±0.101	0.204	0.026±0.017	0.031	0.009
isocurves-MCRF	17	1	0.139±0.073	0.157	0.022±0.013	0.026	0.006

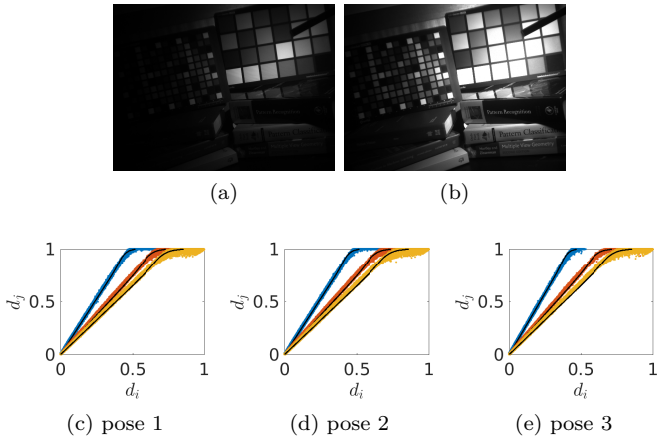


Fig. 12: Brightness transfer functions of a monochromatic camera for same-pose images at three different exposures. The top row shows two of the images taken at different exposures. The bottom row presents the ground truth and the estimated brightness transfer functions. Each color represents a pair of exposures, and in black is our estimation.

5.3. Applications

Three applications were chosen to showcase the potential of the presented method: shape-from-shading (SfS) using generic cameras, color-based segmentation, and color standardization application for MIP cameras.

5.3.1. Shape-from-shading

Shape-from-shading algorithms require linearization of the camera space. To show the applicability of our method, we have performed shape-from-shading, using a cellphone camera with its built-in flash light on. We calibrated the camera device using our method and, then, assuming a point light source (not centered in the lens) and the inverse square law for the light falloff, we have used the method made available in Visentini-Scarzanella et al. (2012). Figure 13 shows the results of the reconstruction.

5.3.2. Color-based Segmentation

Color segmentation is used to perform image segmentation based on the color information of the pixel, typically disregarding the pixel brightness/luminosity/"lightness". Having calibrated the camera response, and consequently having a more accurate color representation on the camera, it is expected some improvements on this type of application.

To segment an image based on the chromaticity of a region defined by a user, we implemented a method that uses an hysteresis threshold on the euclidean distance between the chromaticity of the user-defined region and the rest of the image. This method is simple and does aim to compete with the state-of-the-art on color-based segmentation. In contrast, this test aims to show that color-based segmentation can be improved by calibrating the camera. Figure 14 shows the chromaticity-based segmentation both before and after calibration of the camera response with the proposed method. It shows a clear improvement in descerning

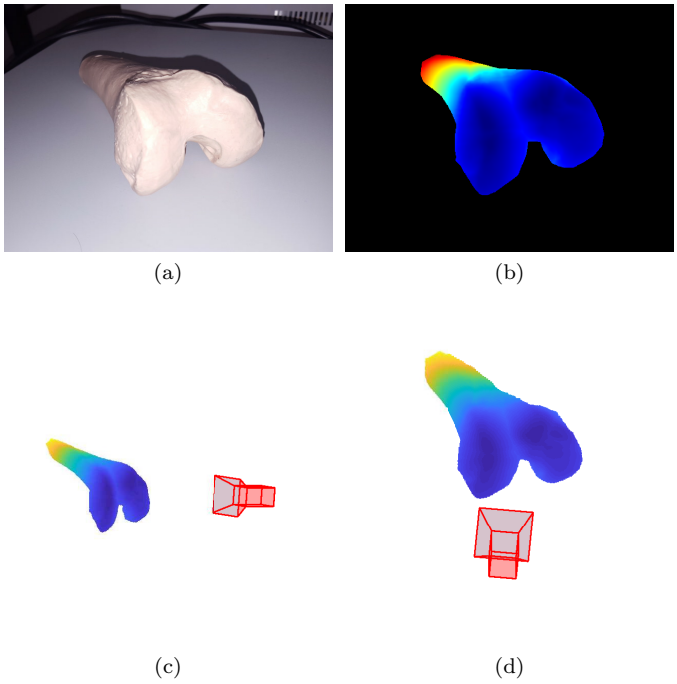


Fig. 13: Shape from shading using a cellphone camera: (a) the original frame, (b) the reconstructed depths, (c-d) 3D reconstructions from different point-of-views.

between certain colors, leading to a more accurate segmentation of that particular color defined by the user. These results imply that if this simple method can be improved by calibrating the camera, other methods, such as unsupervised segmentation, can also be improved as long as it uses color as a base for segmentation.

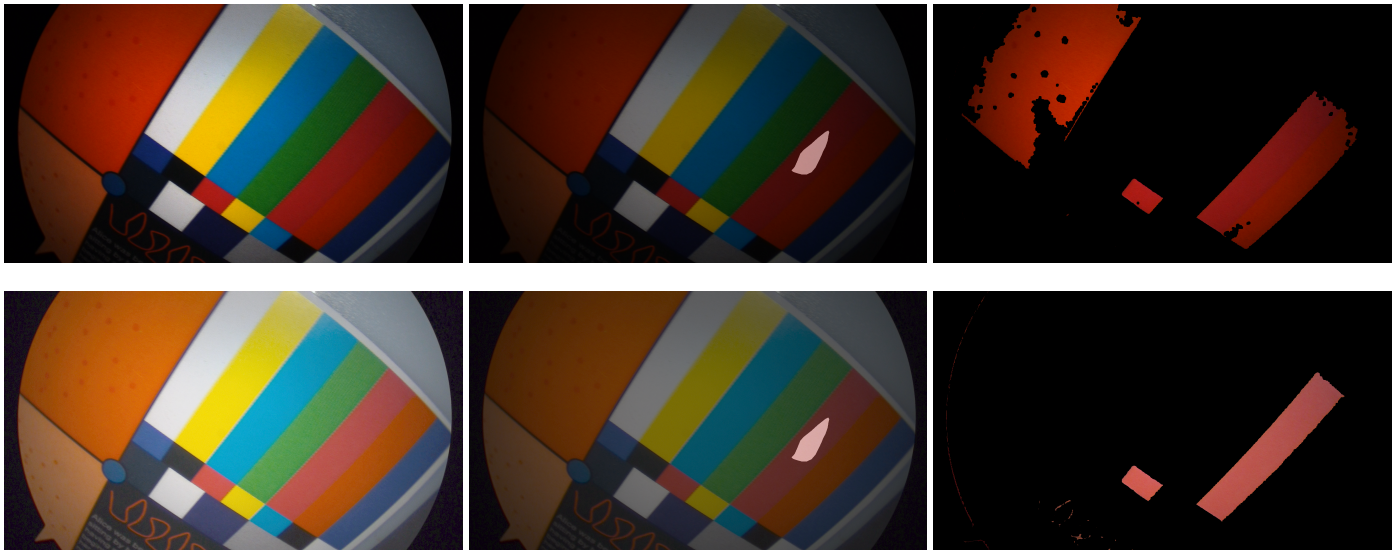
5.3.3. Color Standardization

Regarding MIP, one important application of the camera characterization is the standardization between instruments from different vendors, which inevitably have different camera response properties. Camera characterization is particularly important when post-processing is applied that relies on color information. This is very common in MIP systems. For instance, the visualization toolboxes i-Scan (Pentax Medical, HOYA Corporation, Tokyo, Japan) and SPIES (KARL STORZ GmbH & Co. KG, Tuttlingen, Germany) both use post-processing that manipulates color. Figure 15 shows how our method can improve the color standardization between equipments of two different vendors (Storz and Covidien – Medtronic, MN, USA) in abdominal laparoscopy.

The standardization shows noticeable color convergence after the camera response correction, which consequently translates in color convergence in the visualization mode.

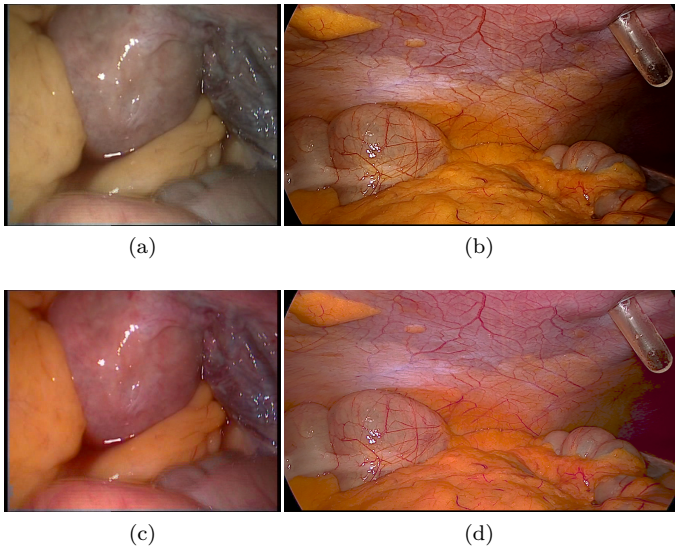
6. Conclusions

In this work we proposed a camera response characterization method than can be used in cameras with vi-



(a) image (b) hand-drawn segmentation (c) segmented regions

Fig. 14: Color-based segmentation. The top row is the original image as acquired by the camera and the bottom row is image after correction to sRGB with the proposed calibration method. Note that the segmentation in the top row struggles to differentiate between similar colors, while the image in the second row provides better results.



(a) (b)

(c) (d)

Fig. 15: Laparoscopic video frames from two MIP towers from different vendors: original frames (top), frame after calibration and correction (bottom). Note that the corresponding colors in the bottom row are perceived to be more similar.

gnetting and/or operating under near light. This single-image method achieved good repeatability results, even better than with Wu et al. (2010b) method that requires 16 images. It was also possible to obtain a 20% improvement on the color accuracy tests. The method is generic since no assumptions are made about the vignetting, and is able to cope with a variety of illuminant types and shapes. Another advantage of our method is that it is easily scalable as more images lead to a new small set of equations that can be stacked if needed. Moreover, if one needs better representation of the sRGB colorspace, adding colors to the targets is the obvious approach. While, for Wu et al. (2010b) approach, defining a new color implies a new image, adding colors in our method requires simply adding new regions to the same target.

The proposed method extends a CRF-estimation method to a more complete camera model, while maintaining the single-image requirement. This is crucial as it allows the employment of the method in the OR without disturbing the clinical workflow, or for an easy use in consumer electronics such as cellphone cameras.

As for limitations, our work depends on some assumptions that may not hold in some scenarios. Specifically, the assumption that the illuminants must have the same spectrum and that the ambient light is negligible. While these are limitations to method presented here, they can be guaranteed in the calibration set-up by using a controlled scene where the calibration target is inserted.

Acknowledgments

The authors thank the Portuguese Science Foundation and COMPETE2020 program for generous funding through the project VisArthro (ref.:

PTDC/EEIAUT/3024/2014) and the PhD scholarship SFRH/BD/113315/2015.

References

Atcheson, B., Heide, F., Heidrich, W., 2010. CALTag: High Precision Fiducial Markers for Camera Calibration. Vision, Modeling, and Visualization , 41–48doi:10.2312/PE/VMV/VMV10/041-048.

Chakrabarti, A., Scharstein, D., Zickler, T., 2009. An Empirical Camera Model for Internet Color Vision. Proceedings of the British Machine Vision Conference 2009 , 51.1–51.11URL: http://www.bmva.org/bmvc/2009/Papers/Paper364/Paper364.html, doi:10.5244/C.23.51.

Collins, T., Bartoli, A., 2012. 3D reconstruction in laparoscopy with close-range photometric stereo, in: Ayache, N., Delingette, H., Golland, P., Mori, K. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. Springer Berlin Heidelberg. volume 15 of *Lecture Notes in Computer Science*, pp. 634–42. doi:10.1007/978-3-642-33418-4_78.

Debevec, P.E., Malik, J., 1997. Recovering high dynamic range radiance maps from photographs, in: ACM SIGGRAPH 1997, ACM Press, New York, New York, USA. p. 1. doi:10.1145/1401132.1401174.

Díaz, M., Sturm, P., 2011. Radiometric calibration using photo collections, in: 2011 IEEE International Conference on Computational Photography (ICCP), IEEE. pp. 1–8. doi:10.1109/ICCPHOT.2011.5753117.

Finlayson, G.D., Mohammadzadeh Darrodi, M., Mackiewicz, M., 2015. The alternating least squares technique for nonuniform intensity color correction. Color Research and Application 40, 232–242. URL: http://doi.wiley.com/10.1002/col.21889, doi:10.1002/col.21889.

Forsyth, D.A., 1990. A novel algorithm for color constancy. International Journal of Computer Vision 5, 5–35. URL: http://link.springer.com/article/10.1007/BF00056770http://link.springer.com/10.1007/BF00056770, doi:10.1007/BF00056770.

Frigo, O., Sabater, N., Delon, J., Hellier, P., 2016. Motion Driven Tonal Stabilization. IEEE Transactions on Image Processing 25, 5455–5468. doi:10.1109/TIP.2016.2601267.

Funt, B., Bastani, P., 2014. Irradiance-independent camera color calibration. Color Research and Application 39, 540–548. URL: http://doi.wiley.com/10.1002/col.21849, doi:10.1002/col.21849.

Goldman, D.B., 2010. Vignette and exposure calibration and compensation. IEEE transactions on pattern analysis and machine intelligence 32, 2276–88. URL: http://www.ncbi.nlm.nih.gov/pubmed/20975123, doi:10.1109/TPAMI.2010.55.

Gonçalves, N., Roxo, D., Barreto, J.P., Rodrigues, P., Gonçalves, N., Barreto, J.P., 2015. Perspective shape from shading for wide-FOV near-lighting endoscopes. Neurocomputing 150, 21–30. doi:10.1016/j.neucom.2014.08.077.

Grossberg, M.D., Nayar, S.K., 2004. Modeling the space of camera response functions. IEEE transactions on pattern analysis and machine intelligence 26, 1272–82. doi:10.1109/TPAMI.2004.88.

Grundmann, M., McClanahan, C., Kang, S.B., Essa, I., 2013. Post-processing approach for radiometric self-calibration of video. IEEE International Conference on Computational Photography (ICCP) , 1–9URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6528307, doi:10.1109/ICCPhot.2013.6528307.

Haneishi, H., Shiobara, T., Miyake, Y., 1995. Color correction for colorimetric color reproduction in an electronic endoscope. Optics Communications 114, 57–63. doi:10.1016/0030-4018(94)00531-X.

Kim, S.J., Pollefeys, M., 2008. Robust radiometric calibration and vignetting correction. IEEE transactions on pattern analysis and machine intelligence 30, 562–76. doi:10.1109/TPAMI.2007.70732.

Lin, S., Yamazaki, S., 2004. Radiometric calibration from a single image, in: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., IEEE. pp. 938–945. doi:10.1109/CVPR.2004.1315266.

Matsushita, Y., Lin, S., 2007. Radiometric Calibration from Noise Distributions, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, Ieee. pp. 1–8. doi:10.1109/CVPR.2007.383213.

Melo, R., Barreto, J.P., Falcão, G., 2012. A new solution for camera calibration and real-time image distortion correction in medical endoscopy-initial technical evaluation. IEEE transactions on biomedical engineering 59, 634–44. doi:10.1109/TBME.2011.2177268.

Mitsunaga, T., Nayar, S., 1999. Radiometric self calibration, in: Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), IEEE Comput. Soc. pp. 374–380. doi:10.1109/CVPR.1999.786966.

Ng, T.T., Chang, S.F., Tsui, M.P., 2007. Using Geometry Invariants for Camera Response Function Estimation, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Minneapolis, MN. pp. 1–8. doi:10.1109/CVPR.2007.383000.

Rodrigues, P., Barreto, J.P., 2015. Single-image estimation of the camera response function in near-lighting, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. pp. 1028–1036. URL: http://www.cv-foundation.org/openaccess/content{_}cvpr{_}2015/html/Rodrigues{_}Single-Image{_}Estimation{_}of{_}2015{_}CVPR{_}paper.html, doi:10.1109/CVPR.2015.7298705.

Seon Joo Kim, Hai Ting Lin, Zheng Lu, Susstrunk, S., Lin, S., Brown, M.S., 2012. A New In-Camera Imaging Model for Color Computer Vision and Its Application. IEEE Transactions on Pattern Analysis and Machine Intelligence 34, 2289–2302. doi:10.1109/TPAMI.2012.58.

Sharma, G., Wu, W., Dalal, E.N., 2005. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. Color Research & Application 30, 21–30. URL: http://doi.wiley.com/10.1002/col.20070, doi:10.1002/col.20070.

Vazquez-Corral, J., Bertalmío, M., 2014. Color stabilization along time and across shots of the same scene, for one or several cameras of unknown specifications. IEEE Transactions on Image Processing 23, 4564–4575. doi:10.1109/TIP.2014.2344312.

Vazquez-Corral, J., Bertalmio, M., 2015. Simultaneous blind gamma estimation. IEEE Signal Processing Letters 22, 1316–1320. doi:10.1109/LSP.2015.2396299.

Vazquez-Corral, J., Bertalmio, M., 2016. Log-encoding estimation for color stabilization of cinematic footage, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 3349–3353. URL: http://ieeexplore.ieee.org/document/7532980/, doi:10.1109/ICIP.2016.7532980.

Vazquez-Corral, J., Connah, D., Bertalmío, M., 2014. Perceptual Color Characterization of Cameras. Sensors 14, 23205–23229. doi:10.3390/s141223205.

Visentini-Scarzanella, M., Stoyanov, D., Yang, G.Z., 2012. Metric depth recovery from monocular images using Shape-from-Shading and specularities, in: 2012 19th IEEE International Conference on Image Processing, IEEE. pp. 25–28. URL: http://ieeexplore.ieee.org/document/6466786/, doi:10.1109/ICIP.2012.6466786.

Wilburn, B., Xu, H., Matsushita, Y., 2008. Radiometric calibration using temporal irradiance mixtures, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, Ieee. pp. 1–7. doi:10.1109/CVPR.2008.4587650.

Wu, C., Jaramaz, B., Narasimhan, S.G., 2010a. A full geometric and photometric calibration method for oblique-viewing endoscopes. Computer Aided Surgery doi:10.3109/10929081003718758.

Wu, C., Narasimhan, S.G., Jaramaz, B., 2010b. A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes. International Journal of Computer Vision 86, 211–228. doi:10.1007/s11263-009-0207-3.

Wu, W., Allebach, J.P., 2000. Imaging Colorimetry Using a Digital Camera. Journal of Imaging Science and Technology 44, 267–279.