Photometric Calibration using Multiples Images and a prior on Camera Response Functions

Mauricio Díaz
Laboratoire Jean Kuntzmann
Institut National Polytechnique de Grenoble
Saint Martin d'Hères, France
Email: Mauricio.Diaz@inrialpes.fr

Peter Sturm
Perception Team
INRIA Grenoble Rhône-Alpes
Montbonnot, France
Email: Peter.Sturm@inrialpes.fr

***Abstract*—Measurement of the photometric properties from a particular scene is an essential task for many machine vision applications. Sophisticated instruments are used in high level applications (e.g. gonioreflectometers) but, in some way, the operating principle of a simple camera resides on the same basis and, under specific circumstances, a commercial camera could accomplish the same task. One of the problems is that sensors packed in digital cameras map intentionally the scene radiance into brightness values through non linear functions. If we are able to find this relation, known as the camera response function (CRF), we may get important information on the scene. Our work proposes a method to recover CRF's for different cameras using multiples images. We exploit the fact that the scene radiance is shared by all images. A particular case is studied: a scene consisting of a Lambertian plane.**

***Index Terms*—Photometry, Calibration, Machine vision, Color measurement, Image analysis.**

I. Introduction

Measure accurately the scene radiance is a primordial task in many computer vision algorithms. Realistic rendering, photometric stereo or augmented reality are but a few examples. Such applications require an explicit or implicit modeling of the appearance of objects, i.e. their "colors" or more generally, their reflectance properties. Formally, the function that characterize, in a photometry sense, the reflectance properties from a given material is called the Bidirectional Reflectance-Distribution Function (BRDF). The BRDF is the ratio between the reflected radiance by the object and the irradiance that origins that reflection. Under laboratory conditions, the acquisition of the BRDF can be done using a gonioreflectometer, a device that positions a source and a photo-sensor at various positions above the surface to be measured. Other methods uses previously acquired images to infer object's BDRFs using a passive approach, e.g. [1], (for a complete review of these methods see [2]). The latter mentioned methods are known to be less precise due to a large quantity of source errors, but their benefits are more than evident. One of the mentioned error source is associated to the camera response function (CRF). This function is defined as the map between scene radiance and image intensities and generally is modeled as a non linear curve. Manufactures used to design specific CRF in order to improve the visual quality of the photographs and, in general, they do not publish information about how the camera sensor modifies the values captured towards the final output.

Several methods have been published in order to estimate the CRF of a camera using multiples exposures of a static scene [3], [4], or the nonlinearities present as artifacts in a single image [5], [6], [7], [8]. Unlike those methods our motivation is to use a complete image formation model and scene geometry to estimate multiples CRF at the same time.

The work described in this paper is inspired by the goal of exploiting the richness in appearance available in large unstructured photo collections. In a seminal work, Snavely *et al.* [9] proposed to use online photo collections and structure-from-motion techniques to recover geometric information of the scene (e.g. coarse 3D model of that object and camera positions respect to that object). This constitutes the backbone of the PhotoSynth system[1], whose goal is not the 3D modeling *per se*, but rather to use the recovered camera positions and coarse 3D model to offer intuitive, pleasant, and efficient means of browsing image collections. Indeed, for major monuments, community image collections contain images showing a wide range of appearances of objects, for various lighting conditions: images taken at different times of day and in different weather conditions. Recently, some other works have exploited these sources of data to infer generic properties from a set of cameras produced by the same manufacturer [10], or even to synthesize new illumination environments [11]. The closest approach to the method that we propose is the work presented in [12], however in this work authors limit the CRF to a linear function, which in most of the cases can not model precisely real behaviour of camera sensors.

The problem we address in this paper is as follows. Given a set of geometrically calibrated images of the same scene and a 3D scene model, we wish to determine the CRF's of all images, even if they are all different. This problem is in general highly ill-posed: even if we perfectly knew the illumination conditions for every image, one could, for every hypothetical set of CRF's, define a BRDF per surface point that is consistent with its observations in all images. Clearly, in order to solve the problem, priors are needed. They could concern the illumination, surface reflectance as well as the CRF's. In this paper, we use a realistic strong prior on CRF's, together with

[1] *http://photosynth.net/*, Feb 2010.

priors on reflectance and illumination. Grossberg and Nayar have radiometrically calibrated 201 different camera-lens-combinations and performed a principal component analysis over the obtained CRF's [13]; they revealed that all these functions can be very well modeled as linear combinations of a few basis function (3 are often sufficient). We use this as prior on the CRF's to be estimated in our problem formulation. Without priors on reflectance and illumination, the problem would still be intractable however. In this paper, we use the simplest priors for these: we assume a Lambertian surface (but with spatially varying albedo) and a single directional light source. These priors may be generalized, e.g. towards simple BRDF's including specular terms and illumination consisting of a directional light source as well as ambient illumination.

Further, in this exploratory study, we assume that the object is planar. This is not a major restriction though: it simplifies the problem formulation and essentially reduces the number of unknowns per light source direction from 2 (direction) to 1 (angle between direction and normal of object plane), but besides this, there is no major difference here between planar and 3D scenes.

II. Problem Statement

In a digital image, a pixel measurement is the result of the interaction of several factors, including surface BRDF, geometry of the object, illumination conditions and sensor properties. If one ignores the nonlinear mapping induced by the camera sensor, the radiance E emitted by a point j on the surface of the object can be modeled as show in equation (1).

$$E_j = \int_\Omega L_j(\omega) F_j(\omega, n_j) d\omega \tag{1}$$

Where L_j represents the illumination and F_j the BRDF at the particular point. Under the assumption of a lambertian but spatially varying BRDF, this expression could be reduced to a vectorial point product operation between the light and the normal on the surface, scaled by a factor called albedo ρ (cf. equation (3)).

The success of our algorithm is based on the modeling of the CRF as a parametric function. The CRF allow us to express the relation between image brightness (B) and image radiance (E) as the function $f(E) = B$. In [13], the authors claim that it is possible to find a basis for the theoretical space of all CRF's. In consequence, any CRF belonging to this space, can be modeled using this basis and as few as 3 coefficients. This affirmation is made under two important assumptions: the response f is the same for all pixels on the sensor and the response function f is monotonic. According to these results, the function $f(E)$ is modeled in the following way, where h_0 is the mean and h_n are the principal components of the CRF's:

$$f_i(E) = h_0(E) + \sum_{n=1}^{N} c_{in} h_n(E) \ . \tag{2}$$

We suppose to have M geometrically calibrated images and a 3D model of the scene, i.e. we know, for any point on the scene surface, its image observations. Let B_{ij} the brightness of surface point j measured in image i. Our unknowns are now: the CRF for each image (coefficients c_{in}, cf. equation (2)), parameters of each surface point's reflectance model (see below), and lighting parameters for each image (see below). Since we assume Lambertian reflectance, a planar scene and a directional light source, image irradiance is expressed as the product of the albedo and the cosine of the angle formed by the plane normal and the light source [14], as mentioned earlier:

$$\hat{B}_{ij} = f_i(\hat{\rho}_j \cos \hat{\alpha}_i) \ . \tag{3}$$

Our general problem formulation is then to solve the following minimization problem:

$$\underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{M} \sum_{j=1}^{J} \| B_{ij} - f_i(\rho_j \cos \alpha_i) \|^2 \tag{4}$$

where θ is the set of all unknowns: the c_{in}, ρ_j, α_i.

III. Solution Approaches

Our problem comes down to solving the minimization shown in equation (4). Since this is a non-linear problem, we need to initialize the unknowns. We propose here two approaches for doing so, one for the minimal case of two images, the other for multiple images. The initialization is performed using the inverse CRF's, $g(B) = f^{-1}(E)$, which can also be expressed via principal components like in equation (2) [13]. Once the inverse CRF's are estimated, we estimate the direct CRF's from them and then minimize equation (4).

A. Two images

Consider the observed brightness values of point j in the two images: $B_{ij} = f_i(\rho_j \cos_i)$. Hence:

$$\rho_j = \frac{f_i^{-1}(B_{ij})}{\cos \alpha_i} \ , \tag{5}$$

and thus,

$$\frac{g_1(B_1)}{\cos \alpha_1} = \frac{g_2(B_2)}{\cos \alpha_2} \ , \tag{6}$$

where the function g_j is the inverse CRF for each camera. Like in equation 2, the inverse CRF can also be modeled as a weighted sum of a basis $g_i(B) = p_0(B) + \sum_{n=1}^{N} w_{in} p_n(B)$. Lets denote the cosines of the angles between the plane and the light sources $a_1 = 1/\cos \alpha_1$ and $a_2 = 1/\cos \alpha_2$. We can express equation (6) as a linear equation system:

$$\begin{gathered} (\mathbf{p}_{01} + \mathsf{P}_1 \mathbf{w}_1)\, a_2 - (\mathbf{p}_{02} + \mathsf{P}_2 \mathbf{w}_2)\, a_1 = \mathbf{0} \ , \\ \begin{bmatrix} \mathbf{p}_{01} & | & -\mathbf{p}_{02} & | & \mathsf{P}_1 & | & -\mathsf{P}_2 \end{bmatrix} \begin{bmatrix} a_2 \\ a_1 \\ \mathbf{w}_1 a_2 \\ \mathbf{w}_2 a_1 \end{bmatrix} = \mathbf{0} \ . \end{gathered} \tag{7}$$

The system in equation (7) can be solved by any least squares method. We are able to find the estimations for all a_j up to scale, but the coefficients for the inverse CRF ($\mathbf{w}_j$) are found without ambiguity.

B. Multiple images, multiple cameras

For the multiple cameras case, one alternative consists in creating a sparse linear system computing all pairs of images respect to one previously selected. For example, if the selected camera generates image 1, the linear system to solve can be expressed as follows:

$$\mathsf{A}\mathbf{x} = \begin{bmatrix}\mathbf{p}_{01} & \mathbf{p}_{01} & \dots & \mathbf{p}_{01}\end{bmatrix}^{\mathsf{T}}, \text{ where} \tag{8}$$

$$\mathsf{A} = \begin{bmatrix} \mathsf{P}_1 & \begin{bmatrix}\mathbf{p}_{02} & \mathsf{P}_2\end{bmatrix} & \mathbf{0} & \dots & \mathbf{0} \\ \mathsf{P}_1 & \mathbf{0} & \begin{bmatrix}\mathbf{p}_{03} & \mathsf{P}_3\end{bmatrix} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \mathbf{0} \\ \mathsf{P}_1 & \mathbf{0} & \mathbf{0} & \dots & \begin{bmatrix}\mathbf{p}_{0n} & \mathsf{P}_n\end{bmatrix} \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix}-\mathbf{w}_1 & \frac{a_2}{a_1} & \frac{a_2}{a_1}\mathbf{w}_2 & \frac{a_3}{a_1} & \frac{a_3}{a_1}\mathbf{w}_3 & \dots & \frac{a_n}{a_1} & \frac{a_n}{a_1}\mathbf{w}_n\end{bmatrix}^{\mathsf{T}}.$$

Solving this system (8), we can find the coefficients to model all CRF's and the ratios between the cosines of the angles. However, if one (or all) the images are perturbed with noise, this method can yield to a bad approximation. In that case, we propose a final stage of optimization, using as initialization points the results of equation (8) (that allow us to speed up the convergence and constrain the problem). Optimization is applied directly over the function in equation 4. The parameters to minimize (θ) are the albedo, the cosines of the angles and the coefficients that model the CRF's.

IV. Results

To test our approaches, we created three different classes of synthesized images: for the first one, we produced images from perfectly synthesized CRF (these functions correspond to linear combinations of the basis depicted in [13]). A second set was generated taking images created from real CRF's (using again data provided by Grossberg's database). Finally a third group which includes synthesized images with different levels of Gaussian noise added. All synthesized images simulate a plane with Lambertian properties whose normal forms an angle α with the light source over a large range of possible values (between 0 and 80 degrees). This set up allow us to check out in an experimental environment our previous analysis taking into account different variables.

For the first dataset, we selected images by pairs and we solved equation (7). Estimated CRF's and angles ratio are identical to the values used in the generation of the synthesized images. Parameter recovery was also possible when applying the multiple images algorithm (equation 8) to this dataset. In this hypothetical and ideal case, the problem can be solved without ambiguities. In real world cases, the solution depends on the number of albedo present on the scene.

Figure 1 shows the real inverse CRF's and the estimated functions for a couple of randomly selected images from the second dataset, using equation (7). First, we assume α_1 and w_1 known values and we find the inverse CRF and the angle for image 2. Then, we compute the parameters ($\cos\alpha_j$ and $\mathbf{w}_j$) for both images at same time. The legend in the figure shows plots

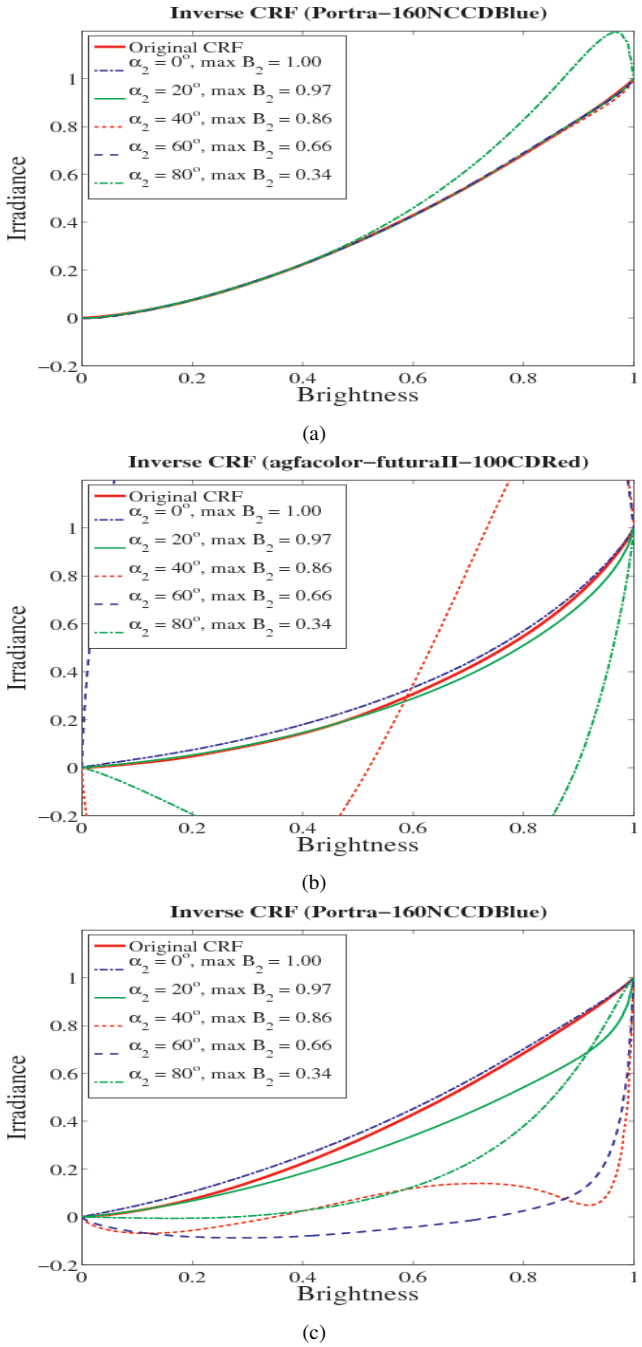


Fig. 1. (a) Estimated and real inverse CRF's for camera 2 when g_1 and angle α_1 are known. Figures (b) and (c) are estimated and original CRF's for camera 1 (g_1) and camera 2 (g_2) respectively, using different values of α_2 (α_1 is set to zero). Evident out-of-range values arise when α_2 increases and deserved brightness values are not sufficient to model the curve. All curves estimated use 3 coefficients to model the inverse CRF.

for different angles α_2 and the maximum brightness value for image 2, while keeping the first image static. Angle α_1 is intentionally fixed to zero and, in image 1, all the range of possible brightness values is available. This figure shows acceptable results when the α_2 values are lower. On the other hand, when the angle α_2 increases, the estimation is logically far from the ground truth because the fully range of normalized

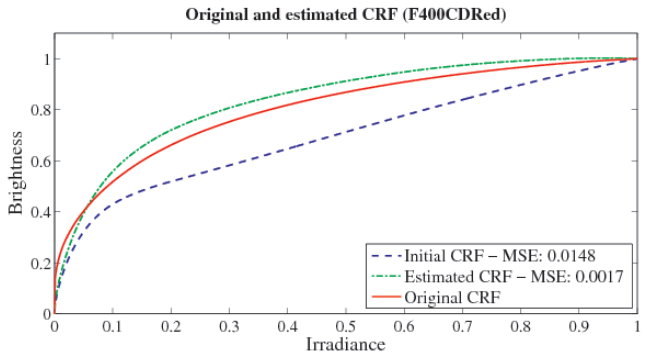
brightness is not available. In this case the CRF can not be fully recovered. It is important to mention that angles greater that 40 degrees are not usual in real world environments. Moreover, the two images case is the more simple scenario and using multiple images improves this results adding data redundancy.

In the multiple image case, the results show the benefits of the data redundancy. We solve equation (8) using 20 randomly selected images from the third dataset described above. Figure 2 shows the real and the estimated CRF for one of the cameras. We also define a quantitative assessment for a varying noise level. The quality of an estimated CRF is measured by computing the following root mean square error (RMS) between the true and estimated inverse CRF:

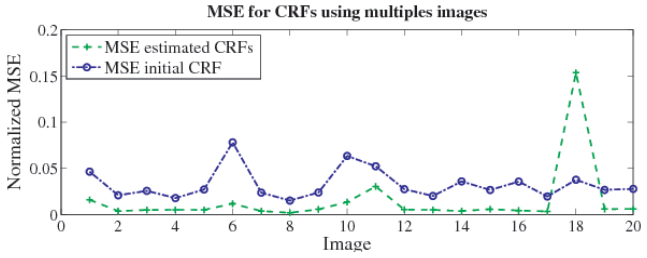
$$\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(f(E_k)-\hat{f}(E_k)\right)^2}\ . \tag{9}$$

In figure 2 we also show the RMS for the 20 estimated CRF's using the approach described in section III-B. The RMS is calculated by taking the square difference between the real measured CRF and its estimation over 1024 points. In general, when dealing with noisy images, optimization improves the results of our algorithm. Finally, we created six groups with different number of images, using those that belong to the third dataset. Images were disturbed with different levels of noise. Figure 4 shows the RMS average for 50 iterations of our algorithm. In this case all the groups of images varying the level of noise. The robustness of the results increases when the number of images is high. This behavior is explained because the increasing number of images then the number of brightness samples is higher and therefore the approximation is better.

Figure 3 shows same analysis this time using images that belong to third set (images corrupted by Gaussian noise). It is clear that the system is seriously affected by the noise even when a small perturbation disturbed it. In the case of multiple images this problem is mitigated with a latter optimization stage.



(a)



(b)

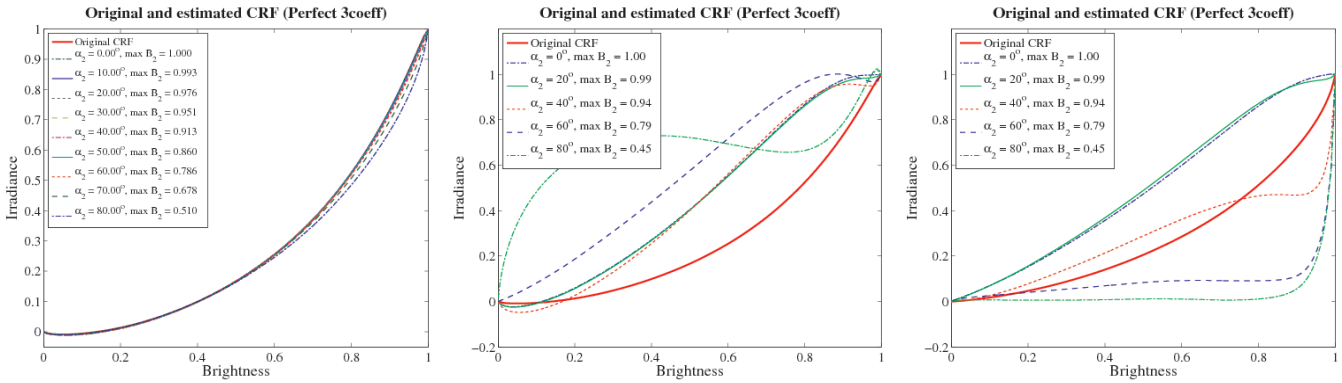
Fig. 2. (a) Estimated CRF for image 8, using a set of 20 images, before and after optimization. Figure (b) shows errors between estimated and real CRF's for all images (before and after optimization).

V. Conclusions

This work illustrates a theoretical framework to recover the photometric calibration using multiple images in the particular case of planar surfaces with lambertian BRDF. We exploit the fact that the scene radiance in a scene is shared by all observers (cameras) and it could be estimated as a function of the illumination and the BRDF (albedo). Although our results are achieved in fully synthetic environments, we are able to determine the conditions under which the algorithm might work in real world cases. These conditions can be summarized as follows: we assume a common CRF for all the pixels in each camera, the CRF is monotonic and we omit variations on the image associated to lens phenomena. The intrinsic condition of our analysis is that to obtain a good estimation of the CRF's, a large variety of albedo is necessary, otherwise the estimated function will be only valid in the available range.. Our results are promising and they might open the door for a complete photometric autocalibration using multiple unaligned images. Limitations of our approach are as follows. The illumination model, consisting of a simple directional light source is certainly not entirely realistic. More general models can in principle be used, e.g. including and ambient illumination term or methods based on spherical harmonics [15], [16], but it remains to be seen if their estimation is well-posed. The empirical camera model proposed by Grossberg [13] provides a good approximation to CRF's of real world cameras. However, if we want to completely characterize a camera, wavelength information must be take it into account. An alternative solution could be to use the model proposed by [17]. This model incorporates additional parameters to manage camera sensitivity and wavelength. The use of the model proposed by these authors in our algorithm could generate consistent CRF's over all channels. Another limitation is the assumption of Lambertian reflectance, although this is relatively easy to circumvent by using a robust weighting function in the non-linear optimization and having recourse to iteratively reweighted least squares optimization for example. Our promising results encourages us to continue this research avenue using more realistic reflectance models and real world environments.

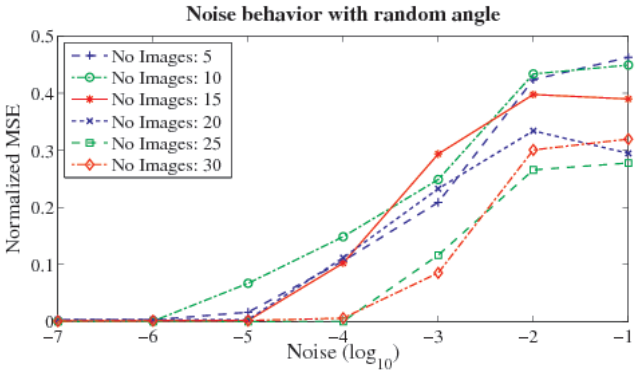
Acknowledgment

Mauricio Díaz was supported by the Programme Alβan, the European Union Programme of High Level Scholarships for Latin America, scholarship No. E07D402742CO.

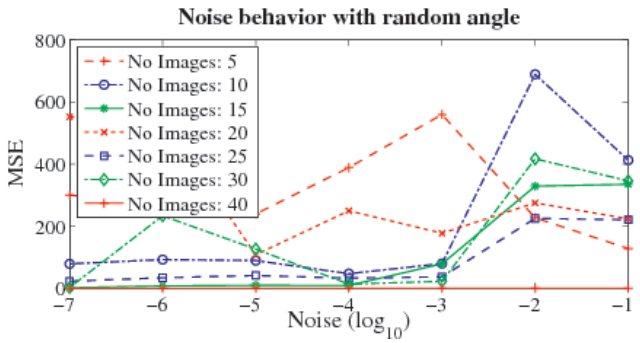


(a) (b) (c)

Fig. 3. (a) Estimated and original CRF for camera 2 when g_1 and angle α_1 are knowns. Estimated and original CRF for camera 1 (g_1) (b) and camera 2 (g_2) (c) are shown, when using different values of α_2. Same conditions that in figure 1 but in this case images are corrupted with Gaussian noise. Estimated responses use 3 coefficients to model the CRF.



(a)



(b)

Fig. 4. Average RMS for the estimated CRFs using multiple images algorithm. Synthesized images have different levels of noise, and different angle. The images were synthesized with real measured CRFs and 3 coefficients. In most cases the RMS is low when the number of images used increases.

REFERENCES

[1] F. Romeiro, Y. Vasilyev, and T. Zickler, "Passive reflectometry," in *Computer Vision ECCV*, Springer, Ed., 2008, pp. 859–872.

[2] T. Weyrich, J. Lawrence, H. Lensch, S. Rusinkiewicz, and T. Zickler, "Principles of appearance acquisition and representation," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 2, pp. 75–191, 2008.

[3] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," *SIGGRAPH*, Aug 1997.

[4] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," *CVPR*, vol. 1, Jul 1999.

[5] S. Lin and L. Zhang, "Determining the radiometric response function from a single grayscale image," *CVPR 2005.*, vol. 2, pp. 66 – 73 vol. 2, 2005.

[6] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum;, "Radiometric calibration from a single image," *CVPR 2004.*, vol. 2, pp. II–938 – II–945 Vol.2, 2004.

[7] J. Takamatsu, Y. Matsushita, and K. Ikeuchi, "Estimating camera response functions using probabilistic intensity similarity," *CVPR 2008.*, pp. 1 – 8, 2008.

[8] T.-T. Ng, S.-F. Chang, and M.-P. Tsui;, "Using geometry invariants for camera response function estimation," *CVPR '07.*, pp. 1 – 8, 2007.

[9] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d," *SIGGRAPH*, 2006.

[10] S. Kuthirummal, A. Agarwala, D. Goldman, and S. K. Nayar, "Priors for large photo collections and what they reveal about cameras," in *ECCV*, Oct 2008, pp. 74–87.

[11] T. Haber, C. Fuchs, P. Bekaer, H.-P. Seidel, M. Goesele, and H. Lensch, "Relighting objects from image collections," *CVPR*, 2009.

[12] Q. Luong, P. Fua, and Y. Leclerc, "The radiometry of multiple images," *IEEE Transactions on PAMI*, vol. 24, no. 1, pp. 19 – 33, Jan 2002.

[13] M. D. Grossberg and S. K. Nayar, "What is the space of camera response functions?" *CVPR*, 2003.

[14] B. Horn, *Robot Vision*. McGraw-Hill Higher Education, Jan 1986.

[15] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on PAMI*, Jan 2003.

[16] R. Ramamoorthi and P. Hanrahan, "A signal-processing framework for inverse rendering," in *SIGGRAPH*, 2001, pp. 117–128.

[17] A. Chakrabarti, D. Scharstein, and T. Zickler, "An empirical camera model for internet color vision," in *BMVC*, 2009.