Exploiting Image Collections for Recovering Photometric Properties

Mauricio Díaz, Peter Sturm

Exploiting Image Collections for Recovering Photometric Properties

Mauricio Diaz[1] and Peter Sturm[1]

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

Abstract. We address the problem of jointly estimating the scene illumination, the radiometric camera calibration and the reflectance properties of an object using a set of images from a community photo collection. The highly ill-posed nature of this problem is circumvented by using appropriate representations of illumination, an empirical model for the nonlinear function that relates image irradiance with intensity values and additional assumptions on the surface reflectance properties. Using a 3D model recovered from an unstructured set of images, we estimate the coefficients that represent the illumination for each image using a frequency framework. For each image, we also compute the corresponding camera response function. Additionally, we calculate a simple model for the reflectance properties of the 3D model. A robust non-linear optimization is proposed exploiting the high sparsity present in the problem.

Keywords: illumination conditions, reflectance estimation, radiometric calibration, photo collections.

1 Introduction

Capturing the photometric properties of a scene is a complex process and requires exhaustive work. A scene captured in a digital image is completely described, in a photometrical sense, if we are able to represent the objects by an appearance model, if we estimate the illumination conditions during the acquisition time and finally if we know the radiometric calibration of the camera. Once the parameters for the models that describe these processes are estimated, they may be used, for example, in augmented reality, relighting or realistic rendering applications. Several issues arise in the formulation and computation of models for object appearance, lighting and camera radiometric response using only images. The main obstacle is that intensity values registered by the sensor are the result of the interaction between surface geometry, object reflectance, scene illumination and camera properties. In order to estimate one or several of these, it is thus important to take into account or be robust to the respective other factors.

In this work we aim at estimating the photometric description for a particular scene using an unstructured and heterogeneous set of images. The use of photo collections is motivated by the goal of exploiting the richness of appearance variations present in these repositories. Recent works have shown the potential of using large databases in computer vision applications. For example, nowadays it is possible to create an acceptable 3D geometrical model, using images taken under completely casual conditions. But estimating the photometric properties from these datasets is still one of the most difficult conundrums for the experts.

Our problem can be formulated as follows: for a 3D geometry obtained from a set of M unstructured images and that is composed by J surface elements, and given the camera's geometric calibration, we wish to determine the camera response function (CRF) for every image, along with the image illumination conditions and the surface reflectance properties for every surface element. We start from scratch, recovering 3D structure of the scene and the camera pose from an image collection. The points belonging to the recovered structure are called surface elements. Depending on the representation, these points are vertices of a mesh or 3D coordinates of a cloud of points. The 3D reconstruction is done using publicly available tools [17, 5]. To represent the image intensities we model the scene radiance as the result of a linear combination of the radiometric camera model, the illumination for each acquisition time and the material reflectance. We estimate the parameters by grouping the unknowns in two subsets: the camera–illumination variables (CRF coefficients and spherical harmonic coefficients) and point variables (albedos). This notation allows us to exploit the high sparsity present in the problem using a robust estimation algorithm.

2 Related Work

Radiometric Calibration. A common strategy used to estimate the CRF consists on posing specific calibration objects (*e.g.* color charts) into the scene at the acquisition time [2, 9]. Other methods require multiples images taken under variable exposure times [3, 13]. The main drawback of these algorithms is that physical access to the scene during the acquisition must be guaranteed. On the other hand, researchers have proposed approaches exploiting image characteristics that reflect the non-linearity produced by the camera response function. For example, [10] and [11] use low level representations such as edges in regions with constant color to extrapolate the CRF. [14] uses geometric invariants looking for the same goal. Noise present in a simple digital image is used in [18] to infer the CRF. A common point in most of the methods above mentioned is the use of a simple, but realistic model for the CRF. In [7], Grossberg and Nayar proposed an empirical model based on the principal component analysis of real world CRF's. The non-linear radiometric response of the camera is composed by a few coefficients multiplying a precomputed basis. A different camera model is also introduced in [1]. In this work, authors model the CRF as the product of a white-balance transform matrix and a polynomial of fifth degree. In this approach the space of possible CRFs is dramatically large and there is no guarantee that the estimated CRFs correspond to the real ones. In the work [4], authors estimate CRFs using a photo collection, but without inferring information about the illumination or the surface reflectance.

Illumination and reflectance estimation. On the side of illumination estimation, we found also different approaches. The most common method, known also as inverse lighting in the computer graphics community, uses a reflective sphere inserted in the scene to recover natural illumination [19]. Non invasive methods have been possible from the formulation of the "signal processing framework" introduced in [16]. This work has allowed to simplify the integration over the hemisphere of the BRDF and the light source as the multiplication of some coefficients in the appropriate space. For the case of Lambertian surfaces, illumination estimation becomes a simple operation, at least in analytical terms. Other methods estimate at the same time different unknowns. For example, Luong *et al.* [12] use, like our approach, a 3D model to perform radiometric calibration and illumination estimation. They used a linear model for the CRF's and required that several images be taken with the same camera, under controlled conditions. Their model for the illumination consists of a point light source. However, most of the time a linear model for the CRF's is not accurate enough. Haber *et al.* [8] have developed an approach to recover reflectance properties and illumination using a wavelet framework. In this work, authors assume that images extracted from photo collections can be photometrically corrected by mapping with a traditional "gamma correction" curve.

3 Image Formation and Estimation Problem

Image irradiance coming from a Lambertian surface under distant illumination, is a magnitude dependent on the surface orientation and the incoming incident illumination. In this work we assume that illumination sources are distant and can be modeled via an environment map. Also, we assume that studied surfaces are characterized by a Lambertian reflectance with spatially varying albedo. Cast shadowing and interreflections are ignored (we explain later how to alleviate this restriction when applying our algorithm to real world cases). Under these considerations, image irradiance E for a surface element j with albedo ρ_j and normal $\mathbf{n}_j$ under an illumination L is computed by:

$$E(\rho_j, \mathbf{n}_j, L) = \rho_j \int_{\Omega} L(\theta_i, \phi_j) \cos\theta_j d\Omega \ , \tag{1}$$

where θ_i and ϕ_j are the inclination and azimuth angles respectively of the light directions, represented in a local coordinate system around the surface normal $\mathbf{n}_j$ and Ω denotes the hemisphere of all possible incoming light directions.

Camera Response Function. The mapping between image irradiance and intensity values is determined by the CRF. A simple but efficient model for the CRF is proposed in [7]. Authors found that CRF's belonging to real world cameras lie in a small part of a theoretical function space that can be spanned using a small basis. This result allows to express the CRF's in terms of N coefficients. For an image i, irradiance E is related to image intensities B by a linear combination of an average CRF h_0 and N principal components h_n:

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

The basis CRF's are thus represented as polynomials of degree D: $h_n(E) = \sum_{d=0}^{D} c_{nd}E^d$. Note that according to [7], the h_n are expressed relative to normalized brightness and irradiance, such that $c_{n0} = 0$ and $\sum c_{nd} = 1$ for all $n = 1 \cdots N$. The degree of the polynomials was chosen for an adequate representation of the curves forming the basis, which was obtained with $D = 9$. These polynomials are known; unknown are the coefficients w_{in} of their linear combination (2).

Global Illumination Model. Representing illumination is a key factor to obtaining 3D models from real world images. An approach that has recently gained importance is to analyze illumination conditions in a frequency framework. We represent the lighting falling on a surface element by a hemisphere centered in the normal position, using spherical harmonics. In this context, Ramamoorthi [15] has shown that for convex Lambertian surfaces, the image irradiance is well represented by a linear combination of coefficients and an orthonormal basis set. This basis is composed by the spherical harmonics $Y_{lm}(\mathbf{n_j})$ rotated around the plane defined locally by the surface normal. The indices obey $l \geq 0$ and $-l \leq m \leq l$ (there are $2l+1$ basis functions for a given order l). Equation (1) is expressed in terms of this basis as a combination of L coefficients as follows:

$$E(\rho_j, \mathbf{n}_j, E_{lm}) = \rho_j \sum_{l=0}^{L} \sum_{m=-l}^{l} E_{lm} Y_{lm}(\mathbf{n_j}) \ . \tag{3}$$

Given a geometric reconstruction of a 3D surface, the normal $\mathbf{n}_j$ becomes a known value and the inverse lighting estimation problem is reduced to the computation of the coefficients E_{lm} that best fit the basis of spherical harmonics rotated at the point normal. Thus, image irradiance per channel is eventually parametrized as a function of the material albedo and the illumination spherical coefficients: $E(\rho_j, \mathbf{n}_j, E_{lm})$.

Estimation Problem. Having defined a linear representation for the CRF's and for the illumination, the intensity value for a surface element j is calculated using the normal at the point $\mathbf{n}_j$ and equations (2) and (3). Since we are dealing with a set of images taken with different illumination conditions but keeping static surface properties, the image irradiance emitted by a surface element depends on the lighting and the material reflectance properties. Additionally, each camera has a different CRF. Therefore, if we denote B_{ij} as the intensity value for a particular color channel, describing the surface element j and the image i, the normalized intensity is estimated by:

$$\tilde{B}_{ij} = f_i(E(p_j, \mathbf{n}_j, E^i_{lm})) = f_i(\rho_j \sum_{l=0}^{L} \sum_{m=-l}^{l} E^i_{lm} Y_{lm}(\mathbf{n_j})) \ . \tag{4}$$

To simplify notations we express equation (4) as a vector multiplication, where the vector $\mathbf{E_i}$ is the set of 9 spherical coefficients that describe illumination in camera i (E^i_{lm} with $L = 2$ and $-l \leq m \leq l$) and $\mathbf{Y_j}$ is the spherical harmonics basis expressed in terms of the coordinate plane around the normal in point j. Combining this vector multiplication with the equation (2), intensity is:

$$\tilde{B}_{ij,\mathrm{ch}} = h_0(\rho_{j,\mathrm{ch}}\mathbf{E}_{i,\mathrm{ch}}^{\mathrm{T}}\mathbf{Y}_j) + \sum_{n=1}^{N} w_{in,\mathrm{ch}} h_n(\rho_{j,\mathrm{ch}}\mathbf{E}_{i,\mathrm{ch}}^{\mathrm{T}}\mathbf{Y}_j) \ , \quad (5)$$

where ch is a suffix indicating the color channel to evaluate (red, green, blue).

Let us denote the vector $\mathbf{a}_i = \left[\mathbf{E}_{iR}^{\mathrm{T}}\ \mathbf{E}_{iG}^{\mathrm{T}}\ \mathbf{E}_{iB}^{\mathrm{T}}\ \mathbf{w}_{iR}^{\mathrm{T}}\ \mathbf{w}_{iG}^{\mathrm{T}}\ \mathbf{w}_{iB}^{\mathrm{T}}\right]^{\mathrm{T}}$ describing illumination and CRF per channel. The vector $\mathbf{E}_{i,\mathrm{ch}}$ has dimension O while the vector $\mathbf{w}_{i,\mathrm{ch}}$ has N components. Then, the dimension of $\mathbf{a}_i$ is $3\times O+3\times N$. The vector $\mathbf{b}_j$ of dimension 3 represents the surface material albedo: $\mathbf{b}_j = \left[\rho_{jR}\ \rho_{jG}\ \rho_{jB}\right]$. We define our estimation function $B(\mathbf{a}_i, \mathbf{b}_j)$ as a function from $\Re^{3\times(O+N+1)} \to \Re^3$. To estimate the unknowns $\mathbf{a}_i$, $\mathbf{b}_j$, we minimize the difference between the observed and predicted intensity values. An optimal solution to calculate the unknowns requires a full non-linear optimization of the cost function, defined as the squared difference between the measured intensity and its correspondent estimation. Given a set of J surface elements projected in M images, the optimization problem is formulated as follows:

$$\min_{\mathbf{a}_i,\mathbf{b}_j} \sum_{i=1}^{M} \sum_{j=1}^{J} \left(B_{ij} - v_{ij}\hat{B}(\mathbf{a}_i, \mathbf{b}_j)\right)^2 \ . \quad (6)$$

The scalars v_{ij} are booleans, a value of 1 indicating that surface element j is visible in image i, otherwise the value being 0. Note that the unknowns can not be estimated without ambiguity: albedos ρ_j and lighting coefficients $\mathbf{E}_i$ can only be estimated up to one global scale factor. Additionally, we impose a constraint on the monotonicity of the estimated CRF's (plausible CRF's are monotonic)

In our experiments, we initialized the optimization algorithm using a vector $\mathbf{b}_j$ containing the mean values of all observed intensities of surface element j. In the case of the vector $\mathbf{a}_i$, the spherical harmonic coefficients ($\mathbf{E}_i$) are initialized with ones while the CRF coefficients ($\mathbf{w}_i$) correspond to a vector of zeros (the initial CRF's are $f_i = h_0$). To avoid the problems related with outliers (*i.e.* intensity samples not included in the 3D model, cast-shadowing, imperfections on the camera pose estimation, surface materials with specular reflection properties, interreflexions, *cf.* section 3), the least squares minimization presented in equation (6) is transformed to a robust estimation problem using the Iterative Reweighted Least Squares (IRLS) algorithm [20].

4 Results

We evaluate the performance of our algorithm in real world conditions using two databases. Both collections target architectural structures in outdoor environments. Images were taken during different periods of the day with natural illumination and different cameras. For the first database (DB1) we had access to the scene and the acquisition equipments. This database is composed by 120

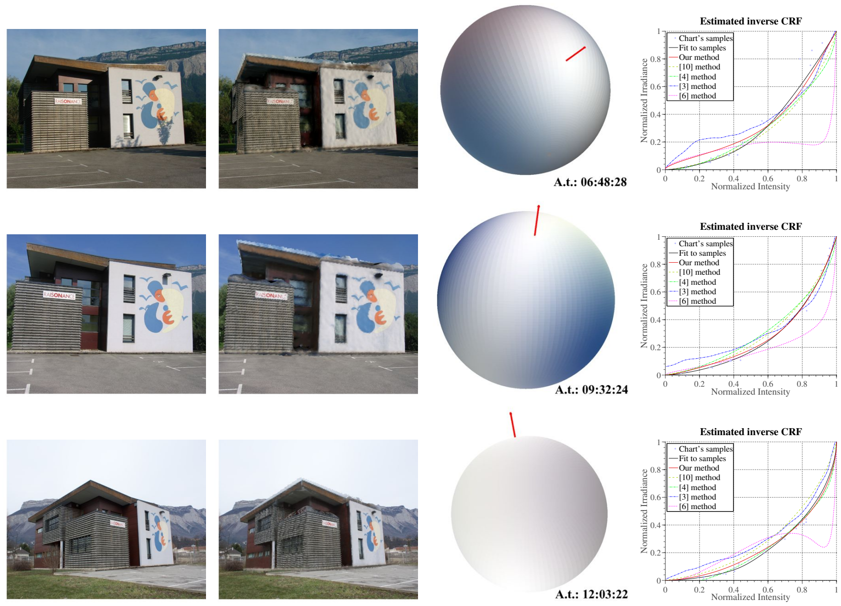

Fig. 1. The 1st column shows some image samples of DB1 used in the reconstruction. On the 2nd column we present corresponding rendered images using the estimated CRFs, illumination conditions and albedos. 3rd and 4th column show representations of the estimated parameters (*cf.* text).

images used to reconstruct a mesh with 112,504 vertices. We got 10 extra images containing a color checker board inside the scene, also we took multiple exposure images for these extra samples, just seconds after the image used for 3D reconstruction was taken. The second database (DB2) was collected from an internet repository and 928 images were used to reconstruct a mesh with 80,444 vertices. In our implementation, we modeled the CRF with 3 coefficients ($N = 3$) and we used 9 spherical harmonics coefficients to model the illumination ($O = 9$). The number of parameters to estimate is $3 \times (J + M \times (O + N))$.

CRF Estimation. To validate our results, we compared the estimated CRF with the ground truth, obtained by placing a color chart in the scene depicted when usin DB1. Four column of figure 1 shows our results with the CRF computed using the *HDRShop* software [3] and the technique described in [6]. These algorithms present poor estimations due to the difficulty of having perfectly aligned images when shooting outdoor scenes (shadows, reflections may change rapidly). CRF estimation using the single image method described in [10] is included. We also show the CRF obtained with the algorithm presented in [4]. When using DB2, we compare our estimated CRF with results of algorithms that do not require physical access to the scene [10, 4].

Illumination Estimation. The performance of our technique when estimating the illumination is evaluated by rendering a synthesized image using the calcu-

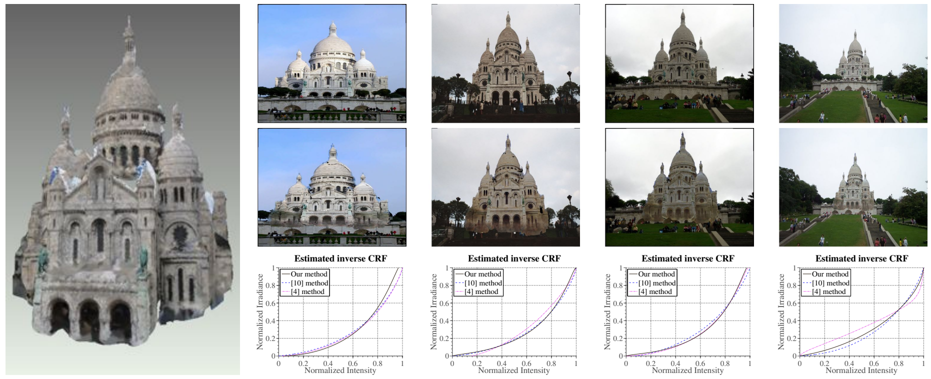

Fig. 2. At the left, the 3D model rendered with the average of the pixel intensities. At the right, four sample images of DB2 and the projection of 3D model with estimated parameter over the original background. Estimated CRF is shown in the third row.

lated lighting. We evaluated the root mean square difference (RMS) between the rendered and original images. These values are calculated for intensities scaled between zero and one. Using DB1 the median RMS difference is 1.2% of the full pixel intensity while using DB2 is around 1.8%. We performed a cross-validation test, using one subset of the database and rendering the synthesized images with the illumination and CRF calculated in a different subset (see figure 1, columns 1-2). For this case the median RMS difference was 7.7% for DB1 and around 23% for DB2. When using DB2, RMS error increases, since the original images contain sometimes pedestrians or objects not taken into account in the 3D model. Third column of figure 1 represents the computed spherical harmonics projected on a sphere viewed from the same point of view as the original images. An arrow indicates the maximum point, the direction where the illumination is strongest. It was mentioned in section 3 that albedos and illumination coefficients can only be estimated up to a global scale factor. This is the case for all three color channels. Hence, in order to display RGB illumination models and surface colors, we first have to estimate the ratios of these scales, between color channels. These scales are calculated by selecting a portion of the sky and projecting its pixels on the sphere. We found the right scale by fitting the spherical harmonics coefficients to the color of some manually selected pixels projected on the surface of the sphere. Images where the presence of a directional light source can be deduced form shadows show a correct estimation of the illumination direction. For cloudy skies, illumination is more uniform (*cf.* third image) and the maximum is less pronounced.

5 Discussion and Conclusion

We have presented a method to estimate jointly photometric properties for a scene. The computed CRFs show good performance, similar to state–of–the–art methods, with the added value that our method provides illumination and reflectance information. Illumination estimation presents a suitable environmental lighting to render new views for the captured scene. One limitation remains on the use of Lambertian reflectance. Although this constraint is contoured using a robust optimization, if we wish to calculate accurately the surface reflectance properties, a more complete model must be used. In that case, the number of parameters to compute may increase dramatically because the illumination and the reflectance interact over all directions of the upper hemisphere centered at the normal of a surface point. Other frameworks may be explored.

References

1. Chakrabarti, A., Scharstein, D., Zickler, T.: An empirical camera model for internet color vision. In: BMVC (2009).
2. Chang, Y.C., Reid, J.: RGB calibration for color image analysis in machine vision. IEEE Trans. on Image Processing 5(10), 1414–1422 (1996)
3. Debevec, P., Malik, J.: Recovering high dynamic range radiance maps from photographs. SIGGRAPH (Aug 1997).
4. Diaz, M., Sturm, P.: Radiometric calibration using photo collections. ICCP (2011).
5. Furukawa, Y., Ponce, J.: Patch-based multi-view stereo software. Web (2009), http://grail.cs.washington.edu/software/pmvs
6. Grossberg, M.D., Nayar, S.K.: What is the space of camera response functions? CVPR 2, 602 (2003).
7. Grossberg, M.D., Nayar, S.K.: Modeling the space of camera response functions. IEEE Trans. on PAMI 26(10), 1272 – 1282 (Oct 2004).
8. Haber, T., Fuchs, C., Bekaer, P., Seidel, H.P., Goesele, M., Lensch, H.: Relighting objects from image collections. CVPR. pp. 627 – 634 (Jun 2009)
9. Ilie, A., Welch, G.: Ensuring color consistency across multiple cameras. In: ICCV. pp. 1268–1275 (2005)
10. Lin, S., Gu, J., Yamazaki, S., Shum;, H.Y.: Radiometric calibration from a single image. CVPR 2004. 2, II–938 – II–945 Vol.2 (2004).
11. Lin, S., Zhang, L.: Determining the radiometric response function from a single grayscale image. CVPR 2005. 2, 66 – 73 vol. 2 (2005).
12. Luong, Q., Fua, P., Leclerc, Y.: The radiometry of multiple images. IEEE Trans. on PAMI 24(1), 19 – 33 (Jan 2002).
13. Mitsunaga, T., Nayar, S.K.: Radiometric self calibration. CVPR. (Jul 1999).
14. Ng, T.T., Chang, S.F., Tsui, M.P.: Using geometry invariants for camera response function estimation. CVPR. pp. 1 – 8 (2007).
15. Ramamoorthi, R.: Modeling illumination variation with spherical harmonics. Face Processing: Advanced Modeling and Methods (2002).
16. Ramamoorthi, R., Hanrahan, P.: A signal-processing framework for inverse rendering. In: SIGGRAPH. pp. 117–128 (2001).
17. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from internet photo collections. IJCV (Jan 2008).
18. Takamatsu, J., Matsushita, Y., Ikeuchi, K.: Estimating camera response functions using probabilistic intensity similarity. CVPR. pp. 1 – 8 (2008).
19. Yu, Y., Debevec, P., Malik, J., Hawkins, T.: Inverse global illumination: recovering reflectance models of real scenes from photographs. SIGGRAPH (1999).
20. Zhang, Z.: Parameter estimation techniques: A tutorial with application to conic fitting. Image and vision Computing (Jan 1997).