# Priors for Large Photo Collections and What They Reveal about Cameras

Sujit Kuthirummal[1], Aseem Agarwala[2],
Dan B Goldman[2], and Shree K. Nayar[1]

[1] Columbia University
[2] Adobe Systems, Inc.

**Abstract.** A large photo collection downloaded from the internet spans a wide range of scenes, cameras, and photographers. In this paper we introduce several novel priors for statistics of such large photo collections that are independent of these factors. We then propose that properties of these factors can be recovered by examining the deviation between these statistical priors and the statistics of a slice of the overall photo collection that holds one factor constant. Specifically, we recover the radiometric properties of a particular camera model by collecting numerous images captured by it, and examining the deviation of this collection's statistics from that of a broader photo collection whose camera-specific effects have been removed. We show that using this approach we can recover both a camera model's non-linear response function and the spatially-varying vignetting of the camera's different lens settings. All this is achieved using publicly available photographs, without requiring images captured under controlled conditions or physical access to the cameras. We also apply this concept to identify bad pixels on the detectors of specific camera instances. We conclude with a discussion of future applications of this general approach to other common computer vision problems.

## 1 Introduction

Large publicly-available photo collections such as Flickr have recently spawned new applications such as Photo Tourism [1] and Internet Stereo [2]. They have also been exploited for filling in holes in images [3], inserting objects into scenes [4], and object recognition [5]. These research efforts have demonstrated the power of using large photo collections to develop novel applications as well as to solve hard computer vision problems.

In this paper, we examine the statistics of such a large photo collection and develop priors that are independent of the factors that influence any one photograph: the scene, the camera, and the photographer. Statistical priors for single images have already been used for a wide range of computer vision tasks [6,7,8,9,10,11]. We argue that priors on the statistics of photo collections have the potential to be similarly powerful, since the statistics of a slice of the photo collection that holds one factor constant should yield information as to how that factor distorts the priors. We investigate this approach to recover

camera properties. We first compute statistical priors from a photo collection with camera-specific effects removed; that is, we use known camera calibration profiles to remove radiometric distortion from a photo collection. As a result, that collection becomes camera-independent. Then, we describe and experimentally validate priors for (a) the spatial distribution of average image luminances and (b) the joint histogram of irradiances at neighboring pixels. Next, we compute these same statistics for a camera-model-specific photo collection whose images have *not* had their distortion removed. We can then recover that camera model's radiometric properties – its non-linear response function and the spatially-varying vignetting for different lens settings – by minimizing the deviation of these statistics from the camera-independent priors. We also show how the same concept can be used to identify bad pixels on the detectors of specific camera instances.

Our approach to recovering properties of specific camera models assumes that all instances of a model have the same properties. This is a reasonable assumption to make for point and shoot cameras [12] since they do not have the variability that arises from attaching different lenses to SLR camera bodies. Hence, in this paper, we restrict ourselves to only point-and-shoot cameras. Also, the camera model properties we recover are aggregate estimates over many instances of the model; for most applications, these estimates are more than adequate. Thus, our approach provides an attractive alternative to traditional camera calibration methods which are typically tedious. Also, since our approach can be used to recover a camera's properties using existing pictures, it provides a convenient means to create a database of camera properties. Such a database would be similar in spirit to the databases available with commercial products like DxO [13] and PTLens [14], but with the important advantage that the cost of creating it would be effectively zero — there would be no need to buy the cameras and manually calibrate them. A photo-sharing website could use our approach to leverage its growing image collection to continually update and add to its database of profiles, and allow users to either undistort their images or make photometrically-correct edits. More importantly, our results demonstrate that the statistics of large photo collections contain significant information about scenes, cameras, and photographers, and our work represents a first step towards extracting and exploiting that information.

## 2   Related Work

A number of image priors have been proposed to describe the statistics of individual photographs, such as the sparsity of outputs of band-pass filters (e.g. derivative filters) [6,7], biases in the distribution of gradient orientations [8,9], and $1/f$ fall-off of the amplitude spectrum [10,11]. These priors have been exploited for applications such as deriving intrinsic images from image sequences [15], super-resolution and image demosaicing [16], removing effects of camera shake [17], and classifying images as belonging to different scene categories [9,18]. We focus on the aggregate statistics of large photo collections, which tend to have

less variability than the statistics of a single image. We thus propose two new priors for aggregate statistics of large photo collections and describe how they can be exploited to recover radiometric properties of cameras.

The most popular method for estimating the camera response function involves taking multiple registered images of a static scene with varying camera exposures [19,20]. Grossberg and Nayar [21] relax the need for spatial correspondences by using histograms of images at different exposures. If the exposure cannot be varied, but can be locked, the response can be estimated by capturing multiple registered images of a static scene illuminated by different combinations of light sources [22]. All these methods require significant user effort and physical access to the camera. Farid [23] assumes that the response function has the form of a gamma curve and estimates it from a single image. However, in practice response functions can differ significantly from gamma curves. Lin et al. [24] also estimate the response from a single image by exploiting intensity statistics at edges. Their results depend on the kinds of edges detected, and their method employs a non-linear optimization which needs multiple initial guesses for robustness. In contrast, we automatically and robustly estimate the response function using numerous existing photographs.

Vignetting can be estimated by imaging a uniformly illuminated flat textureless Lambertian surface, and comparing the intensity of every pixel with that of the center pixel (which is assumed to have no vignetting) [25,26]. Unfortunately, realizing such capture conditions is difficult. One approach is to use a device called an "integrating sphere," but this specialized hardware is expensive. Stumpfel et al. [27] capture many images of a known illuminant at different locations in the image and fit a polynomial to the measured irradiances. The same principle has been used to estimate vignetting from overlapping images of an arbitrary scene [28,29,30] using measured irradiances of the same scene point at different image locations. All these methods require the user to acquire new images under controlled conditions. Some of the above approaches [28,29] can be used to simultaneously estimate the vignetting and the response function of a camera, but there are typically ambiguities in recovering this information. Since we recover both properties independently, we do not have any ambiguities. Recently, Zheng et al. [31] have proposed estimating vignetting from a single image by assuming that a vignette-corrected image will yield an image segmentation with larger segments. Their optimization algorithm, which consists of many alternating image segmentation and vignetting estimation steps, is highly non-linear and hence is likely to have local minima issues. In contrast, we estimate vignetting linearly and efficiently.

During manufacturing, bad pixels are typically identified by exposing image detectors to uniform illuminations. However, some pixels develop defects later and it is difficult for consumers to create uniform environments to detect them. Dudas et al. [32] detect such pixels by analyzing a set of images in a Bayesian framework. However, they only show simulation results. We propose a simple technique that is able to detect bad pixels, albeit using many images.

# 3 Aggregate Statistics of Photo Collections

We now describe how we collect various internet photo collections and how we use them to form and experimentally validate two statistical priors that are independent of specific scenes, cameras, and photographers. We form image collections by downloading images from Flickr. Flickr supports searching for images from a particular camera model; we chose five popular models and downloaded thousands of images for each. We also manually calibrated these cameras using HDRShop [19] for response functions and an integrating sphere for vignetting (across different lens settings). To validate our approach, we then used the collection of one camera model – Canon S1IS – as a training set to undistort its corresponding downloaded images and form camera-independent priors;[1] the other camera models and their downloaded images were used to test our hypotheses.

Internet photo collections can contain outliers that corrupt our aggregate statistics. For example, images captured with flash, edited in Photoshop, or cropped would add distortion beyond the radiometric properties that we are recovering. Fortunately, EXIF tags allow us to cull most outliers; we remove flash images, images with certain Software fields, portrait-mode images, and images that are not full resolution. Our resultant collections contain about 40,000 images per camera model, which we then group using lens settings since camera properties vary with aperture and focal length. We would like to point out that since Flickr does not support searching for all images with a particular camera-lens setting, there are configurations for which we could not collect enough photographs to compute robust statistics. However, as we will show, for configurations with sufficient photographs, our approach gives uniformly good results.

## 3.1 Spatial Distribution of Average Image Luminances

Torralba and Oliva [18] and artist Jason Salavon (salavon.com) have made an interesting observation: the average of a set of photographs of similar scenes is not spatially stationary, but has a certain structure to it. So we ask: does the average photograph obtained by pixel-by-pixel averaging of many photographs captured with the same lens setting have a particular structure? To investigate this question we computed the average of the log-luminance of the photographs in the undistorted training set photo collection with the same lens setting. Figures 1 (a,b) show the average log-luminance of two groups of photographs captured with the same focal length, but different f-number. One can see that we have averaged out particular scenes, but the average image is not uniform. This is illustrated in Figures 1 (c) and (d) which show the contrast enhanced versions of the images in Figures 1 (a) and (b), respectively. We can immediately make two interesting observations. (i) The average images have a vertical gradient as can also be seen in Figure 1 (e) which shows log-luminances along a column of the

---

[1] We assume that undistortion is enough to make an image collection camera-independent for the purpose of training priors. While this may not be true in all cases, we have experimentally verified that our priors are accurate across all five camera models (from four different manufacturers) that we used in our experiments.
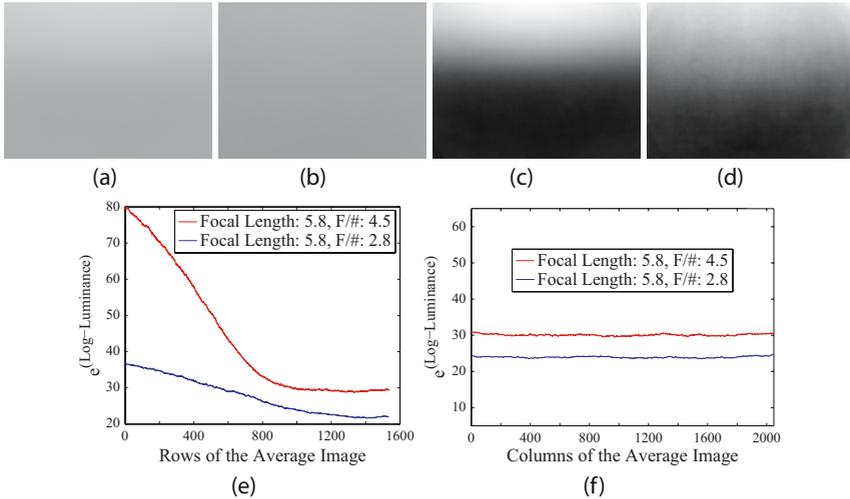
**Fig. 1.** (a) Average log-luminance of 15,550 photographs captured by Canon S1IS cameras with focal length 5.8 mm and f-number 4.5. The photographs, all $2048 \times 1536$, were linearized and vignette corrected before averaging. (b) Average log-luminance of 13,874 photographs captured by Canon S1IS cameras with focal length 5.8 mm and f-number 2.8. (c,d) Contrast-enhanced versions (for illustration only) of the images in (a) and (b), respectively. (e,f) Plots of the average log-luminances of respectively the $1000^{th}$ column and $1000^{th}$ row for the two settings in (a) and (b). Response functions were normalized so that luminance values were in the range $(0,255)$ prior to averaging.

average images. This is possibly because illumination sources are typically above – outdoors, from the sun and sky, while indoors, from ceiling-mounted light fixtures. (ii) The average images do not have a horizontal gradient, illustrated by Figure 1 (f) which shows log-luminances along a row. We have found that these two observations are general and they hold true for all camera models and lens settings. In summary, in the absence of vignetting, average log-luminance images have a vertical gradient, but no horizontal gradient. This observation serves as the prior, which we exploit to recover vignetting in Section 4.2.

### 3.2  Joint Histogram of Irradiances at Neighboring Pixels

A prior on the distribution of gradients in a single image is commonly used in computer vision estimation tasks [16,17]. However, the larger data set of a photo collection allows us to measure how this gradient distribution varies as a function of irradiance values. Therefore, we compute the joint histogram of irradiances at neighboring pixels (where neighborhood is defined as 4-connected). Note that we characterize the joint histogram only for a small block of pixels, since we know from Section 3.1 that this statistic would also vary spatially.

   We now describe how we compute the joint histogram of irradiances for a color channel of a camera model. We assume that we know the inverse response
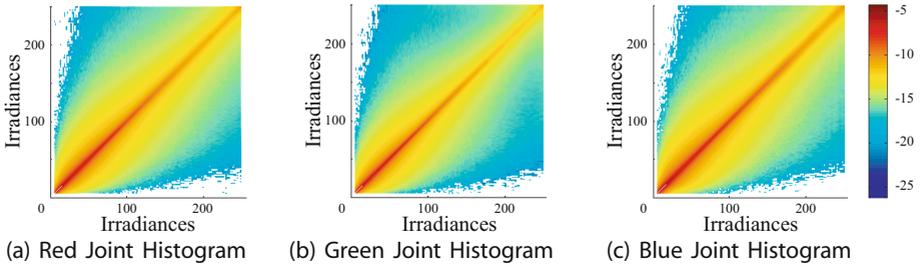
**Fig. 2.** Log of the joint histograms of (a) red, (b) green, and (c) blue irradiances computed from 15,550 photographs captured by Canon S1IS cameras with the extreme lens setting – smallest focal length (5.8 mm) and largest f-number (4.5). The inverse camera response functions used were normalized so that irradiance values were in the range (0,255). When computing the histograms we ignored irradiances less than 5 and greater than 250 to avoid the effects of under-exposure and saturation, respectively.

function, $R$, for that channel, where $R(i)$ is the irradiance value corresponding to intensity $i$. Using $R$ we linearize that channel in photographs from that model and compute a joint histogram, $JH$, where $JH(i,j)$, gives the number of times irradiances $R(i)$ and $R(j)$ occur in neighboring pixels in a desired pixel block. We interpret the joint histogram as the joint probability distribution of irradiances by assuming that the distribution is piecewise uniform within each bin. However, since the values of $R$ are typically non-uniformly spaced, the bins have different areas. Therefore, to convert the joint histogram to a probability distribution, we divide the value of each bin by its area. Note that the values of $R$ determine the sampling lattice, so to enable comparisons between joint histograms for different response functions we resample the histogram on a regular grid in irradiance space. Finally, we normalize the resampled distribution so that it sums to one.

We computed joint histograms of red, green, and blue irradiances for several camera models using $31 \times 31$ pixel blocks at the center of photographs. Figure 3 shows the joint histograms for the Canon S1IS camera model computed from photographs with the smallest focal length and largest f-number. These histograms show that the probability of any two irradiances being incident on neighboring pixels varies depending on the values of the irradiances. Also, the probability of the same irradiance occuring at neighboring pixels is greater for low irradiance values and decreases slowly as the irradiance value increases. Finally, note that the histograms for different color channels differ slightly, illustrating that the visual world has different distributions for different colors. We have empirically observed that for any particular color channel, the joint histogram looks very similar across camera models, especially when computed for the extreme lens setting – smallest focal length and largest f-number. This is not surprising, because the extreme setting is chosen by different camera models for similar types of scenes. We quantified this similarity using the symmetric Kullback-Leibler (KL) divergence between corresponding histograms. The symmetric KL divergence between distributions $p$ and $q$ is defined as

$$KLDiv_{Sym}(p, q) = \Sigma_i q(i) \log(\frac{q(i)}{p(i)}) + \Sigma_i p(i) \log(\frac{p(i)}{q(i)}), \tag{1}$$

where $p(i)$ and $q(i)$ are the samples. For the Canon S1IS and Sony W1 camera models, the symmetric KL divergence between corresponding joint histograms for the extreme lens setting were 0.059 (red channel), 0.081 (green channel), and 0.068 (blue channel). *These small numbers illustrate that the histograms are very similar across camera models.* Therefore, we can use the joint histograms computed for any one camera model as non-parametric priors on these statistics.
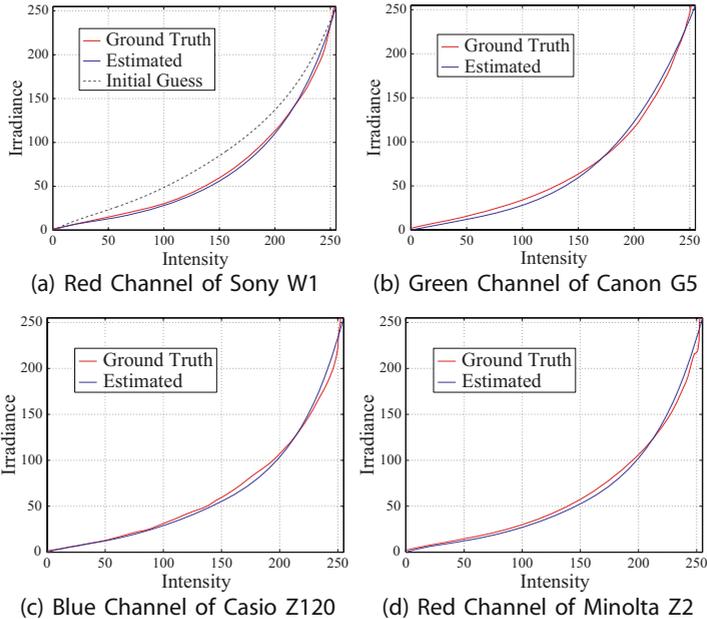
## 4   Using the Priors for Radiometric Calibration

In this section we use these camera-independent statistical priors to recover the response function of a camera model, the vignetting of a camera model for different lens settings, and the bad pixels on the detector of a specific camera. We use the same basic approach for all three applications; given a photo collection, we estimate camera properties that minimize the difference between the statistics of the photo collection and the priors defined in the previous section.

### 4.1   Estimating Camera Response Function

We estimate a camera model response function by minimizing the difference between the joint histogram of irradiances (Section 3.2) for the camera model and the camera-independent prior joint histogram. To estimate the response for a color channel of a camera model, we first compute the joint histogram, $JH$, of intensities in a $31 \times 31$ pixel block at the center of photographs from a collection with the smallest focal length and largest f-number. Say $R$ is an estimate of the inverse response function. Since $R$ is a one-to-one mapping from image intensities to irradiances, $JH$ can be used to compute the joint histogram of irradiances, as described in Section 3.2. We can then determine the 'goodness' of the estimate $R$ by computing the symmetric KL Divergence (Equation 1) between this histogram and the prior histogram for that color channel. Therefore, we can estimate the response function using an optimization over $R$ that minimizes this divergence. We use a simple polynomial [20] as the parametric representation of $R$, and optimize over its coefficients. We define $R(i) = 255 * \sum_{k=1}^{N} \alpha_k (\frac{i}{255})^k$, where $R(i)$ is the irradiance corresponding to intensity $i$, $\alpha_k$ are the coefficients, and $N$ is the degree of the polynomial. We normalize $R(.)$ such that $R(255) = 255$. We have used $N = 5$ in our experiments, since we found it to be a good fit for all inverse response functions in our data set; the mean RMS fitting error was 0.41%. We use the Nelder-Mead Simplex method [33] for the optimization. Note that the joint histogram of image intensities has to be computed only once, though a resampling and normalization step must be performed at each iteration.

We used the priors obtained from the Canon S1IS model to estimate the inverse response functions of Sony W1, Canon G5, Casio Z120, and Minolta Z2

Fig. 3. Estimated and ground truth inverse response functions of one channel for four camera models – (a) Sony W1, (b) Canon G5, (c) Casio Z120, and (d) Minolta Z2. For these estimates we used 17,819, 9,529, 1,315, and 3,600 photographs, respectively. (a) also shows the initial guess used by our optimization. (e) RMS percentage errors of the estimated inverse response functions for camera models from four different manufacturers obtained using our proposed method and the method of [24].

camera models. Due to space constraints, we only show the inverse responses of one of their channels in Figures 3(a-d). For comparison we also show the ground truth inverse response functions obtained using HDRShop [19][2]. As we can see, the estimated curves are very close to the ground truth curves. The difference between the two sets of curves is greater at higher image intensities, for which HDRShop typically provides very noisy estimates.

The RMS estimation errors are shown in Figure 3(e). Even though our estimation process uses a non-linear optimization, we have found it to be robust to

---

[2] Inverse response functions can only be estimated up to scale. To compare the inverse responses produced by our technique and HDRShop, we scaled the results from HDRShop by a factor that minimizes the RMS error between the two curves.
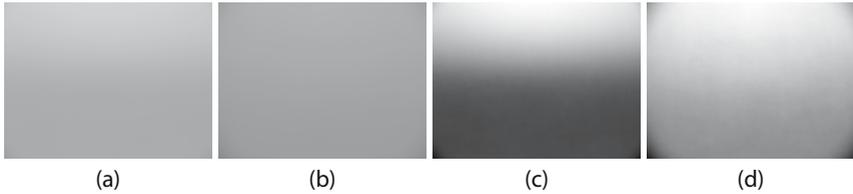
**Fig. 4.** (a) Average log-luminance of 15,550 photographs captured by Canon S1IS cameras with focal length 5.8 mm and f-number 4.5. (b) Average log-luminance of 13,874 photographs captured by Canon S1IS cameras with focal length 5.8 mm and f-number 2.8. (c,d) Contrast enhanced versions of the images in (a) and (b), respectively.

choices of the initial guess. For all our results we used the mean inverse response from the EMoR database [34], shown in Figure 3(a), as the initial guess. For comparison, Figure 3(e) also shows the estimation errors obtained when using the method of Lin et al. [24] on large image sets (the same ones used by our method) for robustness; the overall mean RMS error of their estimates is 28% greater than ours. An interesting question to ask is: How many photographs does our technique need to get a good estimate? We have found that only around 200 photographs are required to get an estimate with RMS error of about 2%. In some cases, as few as 25 photographs are required. (See [35] for details.)

## 4.2   Determining Vignetting for a Lens Setting

Vignetting in a photograph depends on the lens setting – the focal length and the f-number – used to capture it. In Section 3.1, we have seen that the average log-luminance of a group of linearized and vignette-corrected photographs captured with the same lens setting has a vertical gradient but no horizontal gradient. Using the technique in Section 4.1, we can recover response functions, linearize photographs and compute average log-luminance images. Figures 4 (a, b) show the average log-luminances for two groups of linearized photographs captured by Canon S1IS cameras with the same focal length, but different f-number. The photographs used were not vignette-corrected. The contrast-enhanced versions of these images are shown in Figures 4 (c) and (d), respectively. Note the darkening of the corners, which suggests that vignetting information is embedded in the average images. The average images now have a horizontal gradient in addition to a vertical gradient. This observation coupled with our prior model (Section 3.1) leads to a simple vignetting estimation algorithm: find a vignetting function that yields a corrected average log-luminance image with no horizontal gradient.

Since vignetting affects all color channels equally, we only need to analyze its effect on luminance. The measured luminance $m$ at pixel $(x, y)$ in photograph $i$ can be written as:

$$m_i(x, y) = v(x, y) * l_i(x, y), \tag{2}$$

where $v(x, y)$ is the vignetting at that pixel and $l_i(x, y)$ is the luminance that would have been measured in the absence of vignetting. Taking the log on both

sides of Equation 2 and computing the average log-luminance in $N$ photographs with the same lens setting, we get

$$\frac{1}{N}\Sigma_i \log(m_i(x,y)) = \log(v(x,y)) + \frac{1}{N}\Sigma_i \log(l_i(x,y)). \tag{3}$$

Writing the measured average log-luminance, $\frac{1}{N}\Sigma_i \log(m_i(x,y))$, as $M(x,y)$, $\log(v(x,y))$ as $V(x,y)$, and the average log-luminance in the absence of vignetting, $\frac{1}{N}\Sigma_i \log(l_i(x,y))$, as $L(x,y)$, Equation 3 becomes

$$M(x,y) = V(x,y) + L(x,y). \tag{4}$$

According to our prior model, in the absence of vignetting an average log-luminance image does not have a horizontal gradient, i.e., all values in a row are equal. This implies that Equation 4 can be rewritten as

$$M(x,y) = V(x,y) + L(y). \tag{5}$$

Note that $M$ is known, while $V$ and $L$ are unknown. We assume that vignetting is radially symmetric about the center of the image. Therefore, vignetting at pixel $(x,y)$ can be expressed as a function of the distance, $r$, of the pixel from the image center. We model the log of the vignetting as a polynomial in $r$: $V(x,y) = \sum_{k=1}^{N}\beta_k r^k$, where $\beta_k$ are the coefficients and $N$ is the degree of the polynomial. In our experiments we have used $N = 9$. Note that the value of $V$ is zero at the center of the image, modeling the fact that there is no vignetting there. This model reduces Equation 5 to a set of linear equations in the unknowns $L(y)$ and the vignetting coefficients $\beta_k$, which we can solve for efficiently.

The average log-luminance images in Figures 4 (a) and (b) can be used to estimate vignetting. However, we have observed that the top halves of photographs contain many saturated pixels, especially photographs taken with small focal lengths and large f-numbers (typically used for outdoor scenes with lots of light). For instance, photographs in our data set captured by Canon S1IS cameras with such a setting had pixels in the top half that were saturated approximately 30% of the time. This means that we significantly underestimate the average value for pixels in the top half. Since statistics of the top half of the average images are unreliable, we have used the bottom half to recover vignetting. Figures 5 (a-f) show the estimated vignetting curves obtained using this approach for two lens settings each of three camera models – Canon S1IS, Sony W1, and Canon G5. For comparison, ground truth vignetting curves obtained from photographs captured in an integrating sphere are also shown. As one can see, the estimated vignetting and ground truth curves are very close to each other. Figure 5(g) shows the RMS and mean estimation errors. We have found that our technique needs around 3000 photographs to get an estimate with RMS error of about 2%. (See [35] for details.)
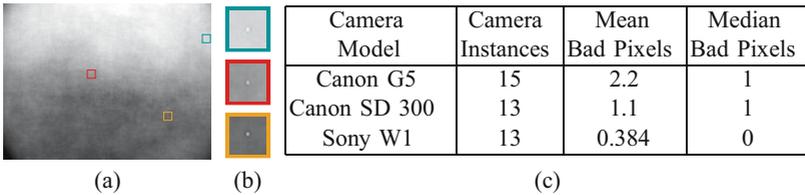
We have observed that statistics at the center of photographs differ slightly from those of other portions of the image. We believe that this is due to a compositional bias – faces are usually captured in the center region. This deviation
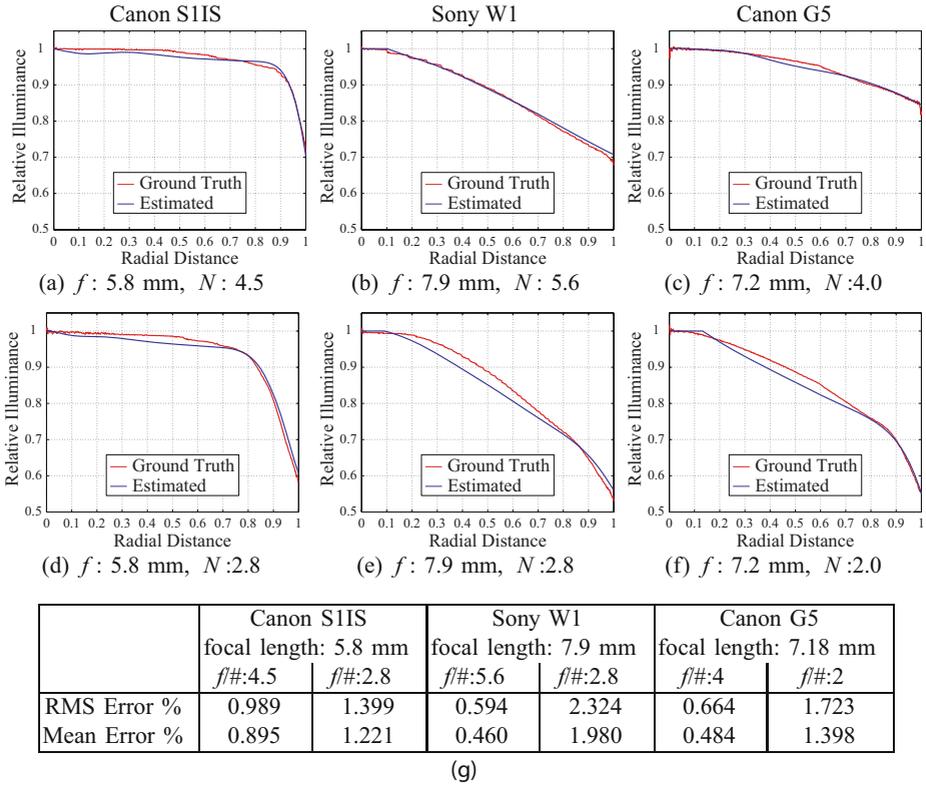
**Fig. 5.** (a-f) Vignetting estimated for two lens settings each of Canon S1IS, Sony W1, and Canon G5 cameras, using the bottom half of their respective average log-luminance images. 15,550, 13,874, 17,819, 15,434, 12,153, and 6,324 photographs, respectively were used for these estimates. ($f$ and $N$ stand for focal length and f-number respectively.) (g) RMS and mean percentage errors of the estimated vignetting for two lens settings each of three camera models; estimation errors are typically less than 2%.

in statistics sometimes causes relative illuminance near the image center to be incorrectly estimated as greater than one. We have handled this by clamping the curves to have a maximum value of one. Note that for lens settings with smaller f-numbers, the estimation is slightly poorer for a larger region near the image center. Such a setting is usually chosen for indoor scenes, where people are typically closer to the camera and their faces occupy a larger region near the image center, thus accentuating this compositional bias.

It is interesting to note from Figure 5 that for these camera models, at lens settings with small f-numbers (large apertures), the corners of the photograph get about 40% less light than the center! This large difference becomes very noticeable if overlapping photographs are stitched together without vignette correction. If photographs are corrected for vignetting, then the overlap seams become barely visible as was shown by [28,29].

| Camera Model | Camera Instances | Mean Bad Pixels | Median Bad Pixels |
|---|---|---|---|
| Canon G5 | 15 | 2.2 | 1 |
| Canon SD 300 | 13 | 1.1 | 1 |
| Sony W1 | 13 | 0.384 | 0 |

(a)        (b)        (c)

**Fig. 6.** (a) Contrast enhanced luminance of the average of 1,186 photographs from a particular Canon S1IS camera. (b) Zoomed in portions of the image in (a) in which we can clearly see bad pixels that have very different intensities from their neighbors. (c) A comparative study of the number of bad detector pixels in a particular camera instance for three different camera models.

### 4.3 Identifying Bad Pixels on a Camera Detector

During manufacturing, camera detectors are exposed to uniform illuminations so that bad pixels – pixels with abnormal sensitivities and biases – stand out and can be easily identified. However, some pixels develop defects later and it is difficult for consumers to create uniform environments to detect them. In Section 3.1 we saw that by averaging a large number of photographs, we average out particular scenes and noise to get a smoothly varying image. *Thus, a simple prior for bad pixel detection is that the average image should be smooth*; bad pixels should be identifiable as causing discontinuities in the average image.

We grouped photographs by the Flickr users who uploaded them, so that each group has pictures from the same camera instance. We then computed the average of each group. Figure 6 (a) shows the contrast enhanced luminance of the average of 1,186 photographs from a particular Canon S1IS camera. In this image, bad pixels clearly stand out, as can be seen in the zoomed-in portions shown in Figure 6 (b). We identify a pixel as bad if the difference between its average value and the median of the average values in a neighborhood around it is greater than a threshold (7 gray-levels). This technique can also be used to rank camera models by the number of bad pixels in each instance. The table in Figure 6(c) presents results from such a study, for which we picked camera instances which had at least 500 photographs in our collection.

## 5 Conclusion

In this paper, we have presented priors on two aggregate statistics of large photo collections, and exploited these statistics to recover the radiometric properties of camera models entirely from publicly available photographs, without physical access to the cameras themselves. In future work, we would like to develop statistics that reveal other camera properties such as radial distortion, chromatic aberration, spatially varying lens softness, etc.. There are, of course, a number of powerful and accurate approaches to camera calibration, and these existing techniques have both advantages and disadvantages relative to ours. In that

light, our primary contribution is a new approach to exploiting the statistics of large photo collections to reveal information about scenes, cameras, and photographers. Recovering camera properties is only one possible application, and we hope that our work inspires others to exploit this approach in new and interesting ways. For example, differences in scene-specific statistics and scene-independent priors could yield information about a scene's geometry, lighting, weather, and motion. A photographer's photo collection could yield information on propensity for camera shake, typical field of view, and preferred camera orientation.

Statistical priors for single images have been useful for a number of computer vision tasks [16,17]. We argue that priors on the statistics of photo collections have the potential to be similarly powerful, since the deviation from these priors of a slice of the photo collection that holds one factor constant should reveal information about that factor. Computer vision problems that operate on a single image are often ill-posed because they must tease apart the influence of several confounding factors of the scene, the camera, and the photographer. For example, vignetting calibration is challenging because it is hard to know if darkening is caused by vignetting or changes in the scene. In effect, a photo collection allows us to *marginalize* over the factors that confound the task at hand. We believe that our work is only the first step in this exciting direction.

# References

1. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3D. ACM Transactions on Graphics (SIGGRAPH), 835–846 (2006)
2. Goesele, M., Snavely, N., Curless, B., Hoppe, H., Seitz, S.M.: Multi-View Stereo for Community Photo Collections. In: ICCV (2007)
3. Hays, J., Efros, A.A.: Scene Completion Using Millions of Photographs. ACM Transactions on Graphics (SIGGRAPH) (2007)
4. Lalonde, J.-F., Hoiem, D., Efros, A.A., Rother, C., Winn, J., Criminisi, A.: Photo Clip Art. ACM Transactions on Graphics (SIGGRAPH) (2007)
5. Torralba, A., Fergus, R., Freeman, W.: Tiny Images. MIT Tech Report (2007)
6. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for nature images. In: Nature, pp. 607–609 (1996)
7. Simoncelli, E.: Statistical Models for Images: Compression, Restoration and Synthesis. In: Asilomar Conference on Signals, Systems and Computers, pp. 673–678 (1997)
8. Switkes, E., Mayer, M.J., Sloan, J.A.: Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis. Vision Research, 1393–1399 (1978)
9. Baddeley, R.: The Correlational Structure of Natural Images and the Calibration of Spatial Representations. Cognitive Science, 351–372 (1997)
10. Burton, G.J., Moorhead, I.R.: Color and spatial structure in natural scenes. Applied Optics, 157–170 (1987)

11. Field, D.: Relations between the statistics of natural images and the response properties of cortical cells. J. of the Optical Society of America, 2379–2394 (1987)
12. Wackrow, R., Chandler, J.H., Bryan, P.: Geometric consistency and stability of consumer-grade digital cameras for accurate spatial measurement. The Photogrammetric Record, 121–134 (2007)
13. DxO Labs: www.dxo.com
14. PTLens: www.epaperpress.com/ptlens
15. Weiss, Y.: Deriving intrinsic images from image sequences. In: ICCV, pp. 68–75 (2001)
16. Tappen, M.F., Russell, B.C., Freeman, W.T.: Exploiting the sparse derivative prior for super-resolution and image demosaicing. In: Workshop on Statistical and Computational Theories of Vision (2003)
17. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing Camera Shake From A Single Photograph. SIGGRAPH, 787–794 (2006)
18. Torralba, A., Oliva, A.: Statistics of Natural Images Categories. Network: Computation in Neural Systems 14, 391–412 (2003)
19. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. SIGGRAPH, 369–378 (1997)
20. Mitsunaga, T., Nayar, S.K.: Radiometric self calibration. CVPR, 1374–1380 (1999)
21. Grossberg, M.D., Nayar, S.K.: Determining the Camera Response from Images: What is Knowable?. PAMI, 1455–1467 (2003)
22. Manders, C., Aimone, C., Mann, S.: Camera response function recovery from different illuminations of identical subject matter. ICIP, 2965–2968 (2004)
23. Farid, H.: Blind Inverse Gamma Correction. IEEE Transactions on Image Processing, 1428–1433 (2001)
24. Lin, S., Gu, J., Yamazaki, S., Shum, H.-Y.: Radiometric Calibration Using a Single Image. CVPR, 938–945 (2004)
25. Sawchuk, A.: Real-time correction of intensity nonlinearities in imaging systems. IEEE Transactions on Computers, 34–39 (1977)
26. Kang, S.B., Weiss, R.: Can we calibrate a camera using an image of a flat textureless lambertian surface? In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 640–653. Springer, Heidelberg (2000)
27. Stumpfel, J., Jones, A., Wenger, A., Debevec, P.: Direct HDR capture of the sun and sky. Afrigraph, 145–149 (2004)
28. Goldman, D.B., Chen, J.H.: Vignette and exposure calibration and compensation. In: ICCV, pp. 899–906 (2005)
29. Litvinov, A., Schechner, Y.Y.: Addressing radiometric nonidealities: A unified framework. CVPR, 52–59 (2005)
30. Jia, J., Tang, C.K.: Tensor voting for image correction by global and local intensity alignment. IEEE Transactions PAMI 27(1), 36–50 (2005)
31. Zheng, Y., Lin, S., Kang, S.B.: Single-Image Vignetting Correction. CVPR (2006)
32. Dudas, J., Jung, C., Wu, L., Chapman, G.H., Koren, I., Koren, Z.: On-Line Mapping of In-Field Defects in Image Sensor Arrays. In: International Symposium on Defect and Fault-Tolerance in VLSI Systems, pp. 439–447 (2006)
33. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical Recipes in C: The Art of Scientific Computing (1992)
34. Grossberg, M.D., Nayar, S.K.: What is the Space of Camera Response Functions?. CVPR, 602–609 (2003)
35. http://www.cs.columbia.edu/CAVE/projects/photo_priors/