Camera Shading Calibration Using a Spatially Modulated Field

Johann A. Schoonees  G. Terry Palmer
Industrial Research Ltd
PO Box 2225
Auckland 1140
New Zealand
Email: {j.schoonees,t.palmer}@irl.cri.nz

Abstract—Camera shading calibration is investigated using an illumination field with sinusoidal spatial intensity modulation, instead of using an idealized flat field. The modulated field can be set up simply with a relatively imperfect display device such as an LCD computer monitor, or as a printed card under constant illumination. Several images are taken with different horizontal and vertical translations of the camera position with respect to the field. The calibration algorithm simultaneously estimates the field and the camera shading by estimating the camera translations and performing equation iteration. Simulation and experiment show that shading correction images can be obtained to an accuracy dominated only by temporal pixel noise.

I. INTRODUCTION

The optical elements responsible in cameras for focusing images on the image plane generally introduce intensity non-uniformities across the image field. Such deviations from ideal behaviour include vignetting from the lens surrounds, distortion characteristics of the lens optics, and smudges or dirt on the photosensitive surface itself.

Digital image sensors (CCD or CMOS) give rise to further unwanted structure, namely uneven pixel black levels (fixed pattern noise caused by integration of dark currents in the sensor wells) and uneven pixel gain or responsivity (photoreceptor non-uniformity).

Figure 1 shows an image of a flat field (uniform illumination) taken with a camera-lens combination exhibiting most of the shading degradations mentioned above.

The cumulative effect of these degradations manifests itself as an apparent pixel-wise variation in responsivity. As such it can be measured and compensated through camera shading calibration. The aim of calibration is effectively to turn each pixel into an independent linear photometer.

Digital sensors also produce temporal noise from sources like photon shot noise and the read-out electronics. Temporal noise results in per-image variation and cannot be predicted by shading calibration. Calibration accuracy can however be improved by averaging as many test images as is practical at each test position.

There are various existing approaches to shading correction. Methods such as Leong [1] rely on intrinsic properties of the image, with little or no prior calibration of the camera. Others assume or estimate shading or vignetting models [2], [3]. A common method attempts pixel-by-pixel calibration of a particular fixed camera configuration under controlled illumination [4]. This paper elaborates such a conventional method, addressing the case of having calibration access to an individual camera characterized by as-yet unknown shading, and having a fast correction calculation.

II. CONVENTIONAL SHADING CALIBRATION

Camera attributes such as the gain and offset of the digital image sensor, lens and filters, zoom, aperture and focus settings are assumed fixed for the calibration; or else the calibration is replicated for each configuration. Controlled illumination typically means a zero-light condition (a dark environment with the lens cap on the camera), and a homogeneous illumination called a flat field [4].

Shading calibration aims to obtain two correction images: (1) a bias frame \( B \) of pixel black levels (also called fixed pattern noise or pixel offsets) which provides a zero-light reference; and (2) a shading correction frame \( F \), nominally an image of a bright flat field, which is used to equalize the pixel sensitivities.
The bias frame $B$ is obtained by simply obtaining an image taken in zero illumination. Temporal pixel noise can be suppressed by averaging several images (10 to 100 would be a good number); the temporal noise variance is then reduced by a factor equal to the number of images averaged. There may also be a small increase in black level with increasing exposure time. This effect can be included in the calibration by setting an exposure time of the same order as that expected in operational use, even when acquiring zero-light images. If the camera has a black level (or "brightness") setting, it should theoretically be set so that all pixel black levels are just above zero in zero light. Doing so may however in practice result in some black levels being so high as to degrade the affected pixels’ dynamic ranges if the distribution of black levels has wide tails. The requirement may therefore be relaxed so that a fraction (such as 10%) of the pixel black levels is allowed to become zero, on the understanding that their shading corrections will be imperfect to some extent.

The bias frame and the correction frame are subtracted from and divided into each input image $I$ pixel-by-pixel to obtain a corrected output image $C$:

$$C = \frac{I - B}{F - B} \quad (1)$$

This calculation assumes that the photosensitive response is a straight line (constant offset and gain slope). The assumption is closely met in the case of CCD sensors [5] and approximately met in CMOS sensors [6].

If the bias frame is determined first, it can be subtracted from all subsequent images to linearize pixel response. The rest of this paper is concerned with estimating a linear shading correction image $G = F - B$ and will for simplicity omit further mention of the bias frame.

III. PROPOSED SHADING CALIBRATION

In what follows $G(x, y)$, $H_k(x, y)$, etc. refer to images, where integers $x \in \{0, 1, \ldots, N - 1\}$ and $y \in \{0, 1, \ldots, M - 1\}$ refer to pixel column and row indices. The image size is $(N, M)$ and index $(0, 0)$ is the origin in the top left corner. The camera shading $G(x, y)$ is defined as the real-numbered image (Figure 1 is an example) that one would obtain from an ideal flat field and then normalize to its maximum pixel level so that $G(x, y) \in [0, 1) \forall x, y$.

The inputs to the algorithm are $K$ test images $H_k(x, y)$, $k = 1, 2, \ldots, K$, which may be considered subimages of a larger test image $T(x, y)$ with size $(N', M')$, $N' > N, M' > M$. Each $H_k(x, y)$ is therefore a rectangular region of size $(N, M)$ cut from $T(x, y)$ without rotation.

A. Spatially modulated field

Flat fields are relatively easy to produce in specialized applications such as optical microscopy [7], [8] and astronomy [9] where the light surface can be placed well outside of the region of focus of the camera, and fluorescent microscopy [10] where homogeneous fluorescent fields are available.

For more general photography, constructing a light source which approaches the uniformity of an ideal flat field is challenging, because cameras are capable of resolving in-focus inhomogeneities. Practical devices like light boxes for viewing transparencies or X-ray films vary in intensity by several percent across their light surfaces. Exacting applications usually require purchase or construction of an integrating sphere for producing reasonably good flat fields.

Instead of attempting to construct a good flat field, we use a test field $T(x, y)$ with intentional spatial intensity variations. The variations are periodic in both X and Y directions to avoid correlation with the expected shading and to aid estimation of translation distances between images taken of the field. The amplitude of the intensity variations is small (about 5%) relative to the average intensity but much larger than any field display noise and temporal camera noise. We found that sinusoidal intensity variations work well in the presence of temporal noise:

$$T(x, y) = b + \frac{a}{2} \left\{ \frac{\cos 2\pi f_x |x-(N'-1)/2|}{N} + \frac{\cos 2\pi f_y |y-(M'-1)/2|}{M'} \right\}$$

\[ x \in \{0, 1, \ldots, N'-1\} \]
\[ y \in \{0, 1, \ldots, M'-1\} \]

where $b$ is the average field intensity, $a$ is the modulation amplitude, and $f_x$ and $f_y$ are the number of cycles of the periodic modulation in the X and Y directions respectively of an image of the field. Example values for 256-level images with 5:4 aspect ratio would be $b = 250$, $a = 5$, $f_x = 5$ and $f_y = 4$. Such a field is shown in Figure 2. (These example values for $f_x$ and $f_y$ are minimal; lower values degrade the accuracy of the estimated translations.)

The spatially modulated field does not have to be produced to any great accuracy. The calibration algorithm does not depend on any particular shape of modulation, nor on any absolute intensity levels. It is forgiving of common display
imperfections such as intensity slopes across the field, uneven coloration, and stuck display pixels. The requirements of the field are that:

1) it should have a well-defined central 2D autocorrelation peak;
2) it should be no more than weakly correlated with the expected shading correction image;
3) it should be relatively bright at the wanted exposure time, near the pixel saturation level, and for colour cameras it should have a colour temperature that produces approximately equal responses in the colour channels; and
4) it should be time-invariant with the camera exposure times of interest, although a certain amount of intensity fluctuation with mains voltage is tolerated as long as the whole field is uniformly affected.

CRTs, plasma screens and the like are not good devices to use because their pixel intensities vary spatially with time. LCD screens are suitable, although models which have strong view-angle dependencies are best avoided with wide-angle lenses.

The field may also be printed and imaged under spatially constant illumination. In this latter case, care should be taken when mixing different illumination sources such as fluorescent and incandescent lights (for example in order to achieve a wanted illumination colour temperature) because their constant illumination. In this latter case, care should be taken when mixing different illumination sources such as fluorescent and incandescent lights (for example in order to achieve a wanted illumination colour temperature) because their different intensity variations within the camera integration time may result in temporal colour fluctuations across the test field.

B. Calibration procedure

The camera is translated across the field with its image plane parallel to the field at a constant distance from it and as little rotation about the optical axis as possible. About twenty test images are acquired, each of which may be the average of several images acquired at each translation to reduce the effects of temporal image noise. The first image should be taken in the centre of the field because the calibration algorithm estimates the shifts of subsequents images relative to the first one.

The fact that the spatial modulation of the test field is periodic means that there is a limit to the distance the camera may be moved without introducing ambiguities in the image registrations. Translation distances should be limited to less than \(1/(2f_X)\) and \(1/(2f_Y)\) of the image size in the X and Y directions respectively.

Images from sensors with colour filter arrays (such as Bayer-tiled sensors) are treated no differently to monochromatic images, because the calibration is done for each sensor cell individually. In multispectral cameras, such as three-chip colour cameras, the calibration should ideally provide a separate shading correction image for each channel. The shading images could be combined into an average if it is found that the channels have similar shadings.

The algorithm described below is then carried out to obtain \(G(x, y)\) and, as a side effect, the estimated field \(T(x, y)\).

### Input:
Test images \(H_k\), shading guess \(G_0\)

### Calculate \(T \) and \(G\) from \(H_k\):
Initial shading \(G = G_0\) (or 1 if no \(G_0\))
While \(G\) is converging

#### Calculate \(T\) from \(H_k\) and \(G\):
For each \(k = 1, 2, \ldots, K\)
Shifted field \(T_k = H_k/G\)
Mean pixel level \(m_k = \text{mean}(T_k)\)
Normalized \(T_k = T_k/m_k\)
Shift \(s_k = \text{est_shift}(T_1, T_k)\)
Unshifted field \(U_k = \text{shift}(T_k, -s_k)\)
Weights \(W_k = \text{shift}(1, -s_k)\)
End for
Average field \(T = \sum_k U_k / \sum_k W_k\)

#### Calculate \(G\) from \(H_k\) and \(T\):
For each \(k = 1, 2, \ldots, K\)
Reshifted field \(V_k = \text{shift}(T, s_k)\)
Shading \(G_k = H_k/(m_k V_k)\)
End for
Accumulate \(G = \sum_k G_k\)
Normalize \(G = G/\text{max}(G)\)
End while

### Output:
Field \(T\), shading \(G\)

---

C. Shading estimation algorithm

Given \(K\) test images \(H_k(x, y), k = 1, 2, \ldots, K\), of an unknown but fixed test field \(T(x, y)\), the calibration algorithm returns the shading \(G(x, y)\) which minimizes the differences between the \(H_k(x, y)\) and \(T(x - u_k, y - v_k)G(x, y)\) for \(K\) unknown shifts \(s_k = (u_k, v_k)\).

Figure 3 shows pseudocode for the algorithm. To keep it simple, the first test image \(H_1(x, y)\) is assumed to be taken in the centre of the field, and the shifts of all subsequent test images are estimated relative to it. An initial guess \(G_0(x, y)\) is provided for \(G(x, y)\) (perhaps from imaging a roughly flat field), or else all its pixels could be set to unity. The calculation loops until some convergence criterion is met, such as the mean or maximum pixel change in \(G(x, y)\) falling below a small threshold.

In Figure 3, “\(\text{mean}\)” means the arithmetic mean of all the pixel levels in the image; “\(\text{shift}()\)” means an image 2D shifted by the given distance which could be non-integer and hence demand pixel interpolation; and “\(\text{max}\)” means the maximum pixel level in an image.

The “\(\text{est}\_\text{shift}\)” in Figure 3 means the estimated 2D shift \(s = (u, v)\) of two images relative to one another. We

---

1Matlab\textsuperscript{TM} code of the calibration algorithm and a simulation environment is available from the first author’s web page http://kauri.auck.cri.nz/~johanss/publications/publications.html.
at first did registration by locating the central peak of the cross correlation of the images (corrected for edge effects by dividing the result by the pyramid-shaped autocorrelation of a unit image of the same size) and then doing parabolic interpolations to refine the estimate to sub-pixel resolution. In a subsequent refinement, we modified est_shift() to use a faster sub-pixel registration method based on Ghanbari’s logarithmic-time search algorithm [11].

D. Practical considerations

The shading estimation algorithm as described is not particularly fast. Some things can be done to speed it up:

1) A momentum $\rho \in [0, 1]$, typically 0.5, can be added to the final calculation of $G$ (see Figure 3): $G = G + \rho(G - G_{\text{prev}})$ where $G_{\text{prev}}$ is the value of $G$ from the previous iteration.

2) The estimates of shifts $s_k$ between the various test images $H_k(x, y)$ converges long before the shading estimate $G(x, y)$ has converged. The number of expensive calculations of est_shift() can be reduced by skipping unnecessary shift re-estimations in every iteration. One heuristic for the number of est_shift() invocations to skip is the negative logarithm of the maximum absolute difference between current and previous shift estimates $[-\log \max |s - s_{\text{prev}}|]$.

Once a shading correction image is obtained, its use in the process of shading correction can be speed up by storing it in the right form. For practical shading correction it is usually faster to multiply with a correction image $1/G(x, y) \geq 1 \forall x, y$ instead of dividing by $G(x, y)$ as in (1).

IV. RESULTS

Obtaining a ground truth for the shading correction image is equivalent to having a perfect flat field available. We therefore simulated a hypothetical camera shading and a spatially modulated field (or rather images of such a field) to evaluate the proposed calibration method.

Figure 4 shows sections (a single central column) from images from the simulation of a camera of 160 by 120 pixels. (Larger images would have made running the simulation too slow and would not have given much more information because the calibration is done per-pixel.) The test field ground truth (Fig. 4(a)) and camera shading ground truth (Fig. 4(b)) both show some irregularity in the form of added spatial noise. $K = 20$ input test images $H_k(x, y)$ were generated from the ground truth (Fig. 4(c)) with shifts of 0, ±7 and ±14 pixels in the X direction, and shifts of 0, ±3 and ±10 pixels in the Y direction. Noise with normalized variance of 0.001 was added to simulate temporal noise (peak SNR of 60dB). Random global intensity variations simulate mains fluctuation of the field display or illumination. A shading guess $G_0(x, y)$ (Fig. 4(d)) with added pixel-wise noise and an intentional intensity slope was provided to simulate a first shading estimate from a poor flat field.

Figure 5 shows the state of the simulation at the first iteration (refer to the pseudocode in Figure 3). The shading guess $G_0(x, y)$ had been divided into each input image $H_k(x, y)$ and $K = 19$ shifts estimated relative to the first image $H_1(x, y)$, to generate 20 unshifted estimates $U_k(x, y)$ of the field. Only the first four are shown in Fig. 5(a) for clarity. These estimates were averaged to provide the first estimate of the test field $T(x, y)$ in Fig. 5(b).

The second half of the algorithm produced 20 estimates $G_k(x, y)$ of the shading (Fig. 5(c)), by dividing reshifted versions of the averaged test field into the test images $H_k(x, y)$. Finally, an averaged and normalized first estimate of the shading $G(x, y)$ was produced (Fig. 5(d)).

Figure 6 shows the same image variables after the 20th iteration, showing how the 20 estimates of field and shading started clustering more closely, and how both the test field and the shading were converging towards ground truth. There was little change in the estimated $G$ after about 200 iterations for this image size.

Figure 7 shows the resulting camera shading. Illustrative figures of central-column sections of the final estimated test field and estimated shading are indistinguishable from Figures 4(a) and 4(b).

After convergence, the estimated camera shading $G(x, y)$ had a mean absolute error of $6.8 \times 10^{-3}$ (approximately 0.07%) compared to ground truth. The simulation was rerun with different amounts of temporal pixel noise in the input test images $H_k(x, y)$ and the results were collecting in Table I. The image peak signal-to-noise ratio (PSNR) was calculated as $20 \log(p/\sigma)$ where $p$ was the pixel saturation level and $\sigma$ was the standard deviation of the temporal noise. Figure 8 shows the same information graphically.

The numerical results show, for common image PSNR, the mean absolute estimation error of $G(x, y)$ tracking at about half to 0.6 of the standard deviation of temporal noise normalized to saturation level, $\sigma/p$. The maximum absolute estimation error (for this image size) is about 50% to 60% more than the normalized standard deviation of noise. The estimated camera shading may therefore be taken to be about as good as could be obtained with an ideal flat field.

The simulation was repeated with other combinations of camera shadings, test fields, different numbers of test images and their shifts, $f_X$ and $f_Y$, with similar results. Real images of areas of even illumination acquired with a scientific-industrial digital camera, corrected with the method described in this paper, were as flat as could be measured.

<table>
<thead>
<tr>
<th>PSNR</th>
<th>$\sigma/p$</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.0316</td>
<td>0.01911</td>
<td>0.0504</td>
</tr>
<tr>
<td>40</td>
<td>0.01</td>
<td>0.00537</td>
<td>0.0152</td>
</tr>
<tr>
<td>50</td>
<td>0.00316</td>
<td>0.00172</td>
<td>0.0050</td>
</tr>
<tr>
<td>60</td>
<td>0.001</td>
<td>0.00068</td>
<td>0.0022</td>
</tr>
<tr>
<td>70</td>
<td>0.000316</td>
<td>0.00052</td>
<td>0.0015</td>
</tr>
<tr>
<td>80</td>
<td>0.0001</td>
<td>0.00051</td>
<td>0.0013</td>
</tr>
</tbody>
</table>
Fig. 4. Simulation of a 160x120 camera in front of a spatially modulated test field. Only a central column from each image is shown: (a) Test field ground truth $T$, (b) camera shading ground truth $G$, (c) input test images $H_k$ (for clarity only first 4 of 20 are shown), and (d) input guess for shading $G_0$.

Fig. 5. Simulation image variables after the first iteration (see pseudocode in Figure 3). Only a central column is shown from each image: (a) Unshifted field estimates $U_k$ (only first 4 shown), (b) averaged field estimate $T$, (c) shading estimates $G_k$ after reshifting (only first 4 shown), and (d) averaged and normalized shading estimate $G$.

Fig. 6. Simulation image variables after the 20th iteration.

V. CONCLUSION

The value of the work is in the avoidance of constructing an accurate flat field for camera shading calibration. The proposed method uses common display devices like LCD monitors or printed cards as calibration fields. It works to sufficient accuracy for many practical applications while being tolerant of typical field imperfections.

As future work, the main algorithm could be improved by modifying or replacing the simple equation iteration proposed here.

The est_shift() procedure dominates calculation time but care has to be taken in attempts at speeding it up because its reliability is central to the correct functioning of the calibration algorithm. Cross correlation in est_shift() is slow but sure, while faster methods may trade certainty of locating the global extremum for speed.

One may also experiment with different shapes of spatial modulation of the test field, especially if that would allow bigger translations of the test images without introducing periodic ambiguity.

ACKNOWLEDGMENT

This work was supported by funding from the New Zealand Foundation for Research, Science and Technology.

REFERENCES


Fig. 7. Final camera shading estimated by simulation has 0.07% mean absolute deviation from ground truth with image peak SNR of 60dB.


Fig. 8. Plot of the simulated mean absolute deviation of estimated shading from ground truth, versus image PSNR (data from Table I). It shows that shading estimation error is dominated by the temporal pixel noise.


