Automatic Compensation for Camera Settings for Images Taken under Different Illuminants

Cheng Lu and Mark S. Drew; School of Computing Science, Simon Fraser University, Vancouver, British Columbia, Canada V5A 1S6
{*clu,mark*} *@cs.sfu.ca*

Abstract

The combination of two images shot for the same scene but under different illumination has been used in wide applications ranging from estimating scene illumination, to enhancing photographs shot in dark environments, to shadow removal. An example is the use of a pair of images shot with and without a flash. However, for consumer-grade digital cameras, due to the different illumination conditions the two images usually have different camera settings when they are taken, such as exposure time and white balance. Thus adjusting (registering) the two images becomes a necessary step prior to combining the two images. Unfortunately, how to register these two images has not been investigated fully. In this paper, we propose a method which can parametrically adjust the two images so as to compensate for the difference in exposure speed, ISO, aperture size, and white balance. This is accomplished by training a 2nd-order masking model on a set of image pairs to predict the model parameters. This trained model can then be used to register two images. In the training phase, we wish to develop a scheme for adjusting the magnitude in each color channel of one image to register with the other image, for each image pair. The problem is difficult because the difference between the two images is a composite of both camera settings and illumination. Here, we use the simple fact that a shadow effect *should be caused purely by the changes of illumination. Suppose we have two images, one of which is taken under illuminant 1 and the other is taken under illuminant 1 plus illuminant 2. If we subtract the first image from the second, a shadow caused by illuminant 1 should disappear in the resulting difference. By adjusting the RGB pixel values of one image so as to completely remove the shadow in the difference image, compensating magnitudes for each color channel can be computed and used to train a masking model. This masking model can then accurately compensate for camera settings for any two new images such that the difference between compensated images reflects only the difference in illumination.*

Introduction

Work on image pairs taken for the same scene but under different illumination was initiated in [1]. This work combined flash and no-flash images to estimate surface reflectance and illumination in the no-flash image. The no-flash image was first subtracted from the with-flash image to create a pure flash image, which appears as if it were taken under only light from the flash. Petschnigg et al. also used flash and no-flash image pairs, to enhance photographs shot in dark environments and remove red-eye effects [2]. Drew et al. [4] suggested a method to remove the ambient shadow from the no-flash image through the aid of the flash image. These works all analyze the difference image: *(with-flash image − no-flash image)*, to infer the contribution of the flash light to the scene. But to make this computation meaningful, the images must be in the same camera settings. These include: exposure time, ISO, aperture, and white balance. Since different lighting conditions usually cause changes of camera settings, an effective registering method to compensate for the difference, for the image pair, is necessary. Unfortunately, this problem has never been investigated carefully. This is partly because the difference between the two images is a composite of the difference of camera settings and of the light arriving at the camera, and they are difficult to separate.

In this paper, we present a method which uses a masking model to compute the compensation between the two images given the camera settings for the two images. This model assumes additivity and proportionality of the different factors involved. Here we use a 2nd-order masking model, with 9 parameters. To train the model for a digital camera, we collect a set of image pairs. For each image pair, two images are taken for the same scene, under different lighting conditions. In the training phase, we restrict our attention to Lambertian surfaces. For such a reflectance map, at the surface point corresponding to each pixel, all lighting is added up into a single effective light [5]. Suppose we have a pair of images A and B. Both are taken for the same scene but under a different lighting situation: only one light source, illuminant 1, illuminates the scene for A; there are two light sources, illuminant 1 and illuminant 2, illuminating the scene for B. An example of this situation would be taking the first image under sunlight, and the second image with flash added to the sunlight, and such pairs form our training set. Fig. 1 shows such a no-flash (*ambient* or A) image and with-flash (*both* or B) image: Figs. 1(a,b) are respectively the ambient-light image, unscaled, and the same image scaled for display; and Figs. 1(c,d) are the "both" (with-flash) image unscaled and scaled.

This additivity property leads to the fact that for the two images, A and B, subtracting the first from the second leads to the shadow caused by illuminant 1 disappearing in the difference image, i.e. as if it were taken under illuminant 2 only. For the above situation, the difference image will be a pure flash image, and the shadow caused by sunlight will disappear. This will be true for image pairs under any ambient lighting plus another image that includes light from a flash.

The use of a shadow helps us to be able to separate the contributions from the camera settings from illumination in the difference of the two images. That is, the shadow should be caused purely by the changes in illumination. We thus can use this fact to compute by what magnitude we have to adjust A so that illuminant 1 causes no shadows to appear in $B-A$. This magnitude will then be used to train the parameters for a masking model. In the training phase, we first collect image pairs for which we set up an imaging environment which has two light sources. There are shadow regions caused by one of illuminants in each pair of images. For each image pair, we first adjust the magnitude in each color channel of image A, so that the difference image, which is obtained by subtracting the adjusted image A' from B, has the shadow removed. Then using the adjusted images, the parameters of the masking model can be computed given the camera settings of the two images. Once we obtain the parameters of the model, we can use this masking model to adjust new image pairs (A,B) as if they are taken under the same camera settings, such that the difference between the two images will be totally controlled only by the light arriving at the camera. The masking model delivers a 3-vector adjustment to generate A' from A.

Camera Settings and Image Acquisition

We have designed an algorithm that works with images acquired even using just consumer-grade digital cameras, with specialized equipment not required. All of the images were acquired in a RAW format and then conversion software was used to convert them into 12-bit linear TIFF images.

Consumer-grade cameras typically have the following settings which can be adjusted automatically or by user control:

1. Focal length.
2. Exposure time (shutter speed).
3. Aperture (f-number).
4. ISO (film speed).
5. White balance.

We fix the focal length so that the camera's focus remains constant. We also use a tripod when taking images to ensure that the image pair capture exactly the same points in the scene (else spatial registration is required — mutual information is the standard approach to this problem). For other settings, we turn on the 'auto' function and let the camera decide how to set the exposure time, aperture, ISO and white balance for different lighting situations.

The size of the aperture and the brightness of the scene control the amount of light that enters the camera during a period of time, and the shutter controls the length of time that the light hits the recording surface. In photography, exposure value (EV) is a value given to all combinations of camera shutter speed and aperture that give the same exposure. In the Additive Photographic Exposure System (APEX) [3], the exposure value is the sum of the Aperture Value (AV) and the Time Value (TV):

$$EV = AV + TV \tag{1}$$

If N is the f-number, the Aperture Value (AV) is

$$AV = log_2 N^2 = 2 log_2 N \tag{2}$$

If t is the shutter time in seconds, the Time Value (TV) is

$$TV = log_2 \frac{1}{t} = -log_2 t \tag{3}$$

Film speed (ISO) is the measure of a photographic film's sensitivity to light. Our test camera has two ISO settings: 100 and 200. All of the values, shutter speed, aperture, and ISO can be read in the conversion software.

Our test camera has four preset white balance settings: Auto, Daylight, Fluorescent, and Tungsten. The white balance algorithm looks for a white patch in the image, the chromaticity of which will be then taken to be the chromaticity of the illuminant. For automatic white balance, the white patch is usually evaluated as the maximum or average found in each of the three image bands separately. Scaling coefficients are then obtained by comparing the chosen white patch with the values of the three channels of a reference white. For the captured images, the scaling coefficients are not known. Here we encapsulate the effect of white balancing by using use the mean value for each RGB channel in the masking model.

A Masking Model for Compensating for Camera Settings

Our goal now is to find a model that can accurately describe the relation between the difference of the two images and the camera settings. As described above, the problem of compensating for camera settings for two images taken under different illuminations reduces to finding a 3-vector of scaling coefficients such that one of the two images, A is transformed to A' such that the shadows caused by illuminant 1 will be removed in the difference image $B-A'$. Suppose we have a shadow region s in image A and B which is caused by illuminant 1, and an out-of-shadow region ns. So the light reaches the regions s and ns as follows:

- In A, neither illuminant 1 nor illuminant 2 reach region s,
- In A, illuminant 1 reaches region ns,
- In B, illuminant 2 reaches region s,
- In B, both illuminant 1 and 2 reach region ns.

Suppose we manually select two areas s and ns that arise from a material with the same reflectance. Then we want to transform A to A' via a 3-coefficient vector M, a coefficient for each color channel, to compensate for the different camera settings so that the difference in the shadow region s between A' and B should be equal to the difference in the out-of-shadow region ns, i.e. the shadow disappears in $B-A'$.:

$$\begin{aligned} & B(ns) - A'(ns) = B(s) - A'(s) \quad \text{no shadow effect} \\ & \Rightarrow B(ns) - B(s) = A'(ns) - A'(s) \quad . \\ & \text{Let } B(ns) - B(s) = M\,(\,A(ns) - A(s)\,) \\ & \Rightarrow M = \frac{B(ns) - B(s)}{A(ns) - A(s)} \text{ ,} M \text{ is a 3-vector.} \end{aligned} \tag{4}$$

In Fig. 2, we show in-shadow and out-of-shadow regions as white areas.

As M is used to compensate the camera settings of the two images, according to the above equation, M should be a function of the ratios of exposure value, ISO, and white balance between the two images.

Here we choose a 2nd-order masking model [6] to describe the difference brought about by camera settings. Such a model, originally proposed for characterizing color printers, uses logarithms and assumes additivity and proportionality of variables. We use this to establish the amount of settings variables required to match the difference between the two images. The form of the

2nd-order model for our application is proposed to be as follows:

$$
\begin{aligned}
& log\left(\frac{B(ns)_i - B(s)_i}{A(ns)_i - A(s)_i}\right) \equiv logM_i \;\; i = 1..3 \\
& = a_1 \, log\left(\frac{EV_B}{EV_A}\right) + a_2 \, log\left(\frac{ISO_B}{ISO_A}\right) + a_3 \, log\left(\frac{mean_{Bi}}{mean_{Ai}}\right) \\
& + b_1 \, log\left(\frac{EV_B}{EV_A}\right)^2 + b_2 \, log\left(\frac{ISO_B}{ISO_A}\right)^2 + b_3 \, log\left(\frac{mean_{Bi}}{mean_{Ai}}\right)^2 \\
& + c_1 \, log\left(\frac{EV_B}{EV_A}\right) \, log\left(\frac{ISO_B}{ISO_A}\right) \\
& + c_2 \, log\left(\frac{ISO_B}{ISO_A}\right) \, log\left(\frac{mean_{Bi}}{mean_{Ai}}\right) \\
& + c_3 \, log\left(\frac{mean_{Bi}}{mean_{Ai}}\right) \, log\left(\frac{EV_B}{EV_A}\right)
\end{aligned}
\tag{5}
$$

where in the training phase M_i is determined as in (4). Subscripts A and B represent images A and B. The estimation of parameters a_1,a_2,a_3,b_1,b_2,b_3,c_1,c_2 and c_3 is accomplished by a least-squares procedure for all color channels at once, yielding a 9-parameter model. I.e., if there are n training pairs of images, then eq. (4) gives $3n$ values M_i. If we line up the $3n \times 9$ float values on the right hand side of eq. (5), including 3 values each for quantities with subscript i, then we can solve for the 9 unknowns using the pseudoinverse.

Once we obtain the 9 parameters, this same masking model can then be used to register two new images, given their camera settings, by deriving a new vector M for the new pair using eq. (5). For the input pair in Figs. 1(a,b), an uncompensated difference is shown in Fig. 3(a); the compensated difference in Fig. 3(b) has the ambient-shadow correctly removed.

Experiments and Results

In our experimental imaging environment, five lighting sources are used: direct sunlight, cloudy daylight, a tungsten light lamp, an incandescent lamp, and a xenon flash light. We used five objects with different colors to create shadows on five different tablecloths. The environment is shown in Fig. 4. We captured images under the following five situations:

- Using direct sunlight as illuminant 1 to create shadows and adding the flash as illuminant 2.
- Using tungsten light as illuminant 1 and adding the flash as illuminant 2.
- Using tungsten light as illuminant 1 and adding cloudy daylight as illuminant 2.
- Using incandescent light as illuminant 1 and adding the flash as illuminant 2.
- Using incandescent light as illuminant 1 and adding cloudy daylight as illuminant 2.

In each situation, we captured 25 image pairs by different combinations of objects and tablecloths. Overall, we collected 125 image pairs. All of the images were acquired in a RAW format using a consumer-grade digital camera and conversion software was used to convert them into 12-bit linear TIFF images. We fixed the focal length, and using the conversion software for each image recovered the exposure time, aperture, and ISO values from the image metadata.

We performed 125 tests on the method by first taking out one of the training pairs from the data set: this would form our test pair. Then we computed the 9-parameter model using the remaining 124 image pairs. The masking model delivers a 3-vector M for our test image pair, and applying this to the color channels of test image A is meant to eliminate the ambient shadow from the difference image $B - A'$. We already know the s and ns regions, as in Fig. 2. So we can compare the color difference, in the shadow region and out-of-shadow region, for the difference image using a compensated image A'.

Examples of the results are shown in Figs. 5, 6. The first image has shadows caused by direct sunlight, and the second image has flash light in addition. After compensation, the difference image has no sunlight (ambient) shadows at all.

Summary

In this paper, we address the problem of compensation for camera settings for image pairs which are taken for the same scene but under different lighting conditions. The difficulty of this problem is that the difference between the two images is a composite of camera settings and scene illumination. We solve this problem by using the fact that shadow effects in images should be entirely caused by illumination. By removing shadows in the difference image, we achieve a separation of camera settings from illumination. Here we propose using a simple masking model to describe camera settings. Via this model, the effect of camera settings can be easily eliminated, making algorithms that use difference-images much more dependable.

References

[1] J.M. DiCarlo, F. Xiao and B.A Wandell, Illuminating illumination. *9th Color Imaging Conference*, pp. 27–34, 2001.

[2] G. Petschnigg, R. Szeliski, M. Agrawala, M.F. Cohen, H. Hoppe and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph*, 23(3): 664-672, 2004.

[3] http://en.wikipedia.org/wiki/Exposure_value.

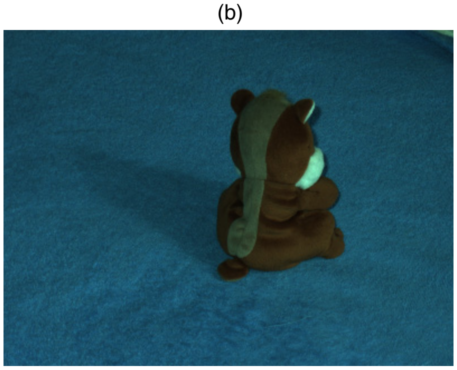
[4] M.S. Drew, C. Lu, and G.D. Finlayson, Removing shadows using flash/noflash image edges, *IEEE Conf. on Multimedia and Expo: ICME*, Toronto, July 2006.

[5] W. Silver, Determining shape and reflectance using multiple images, Ph.D. Dissertation, MIT, 1980.

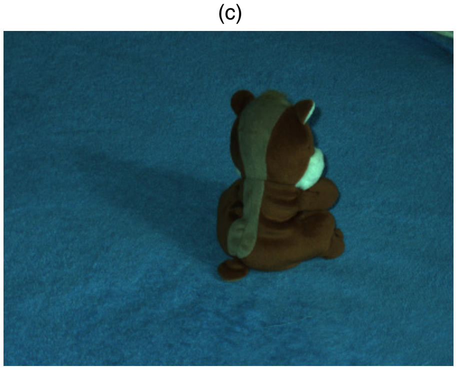
[6] F.R. Clapper, An empirical determination of half-tone colour reproduction requirements, TAGA proc.,p31, 1961.

(a)

(b)



(c)



(d)

Figure 1. *(a,b): Ambient-light image A, unscaled, and scaled for display. (c,d): "Both" image B (ambient $+$ flash), unscaled, and scaled.*

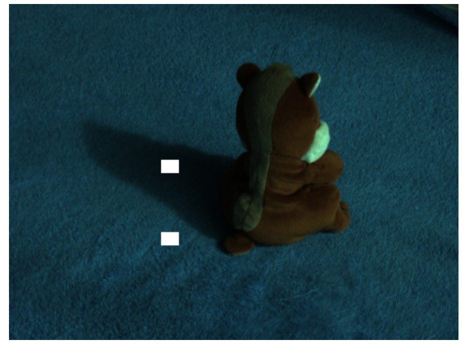
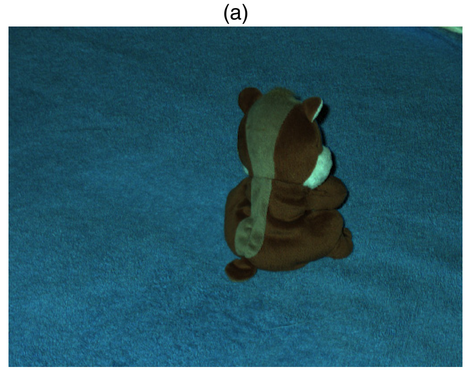


Figure 2. *Ambient-light image A with in-shadow and out-of-shadow regions.*

(a)



(b)

Figure 3. *(a): Pure-flash image $(B - A)$ without compensation; (b): With compensation.*

Figure 4. *Experimental imaging environment.*

Figure 5. *Results: Example 1. Images* A, B *and* $B-A'$.

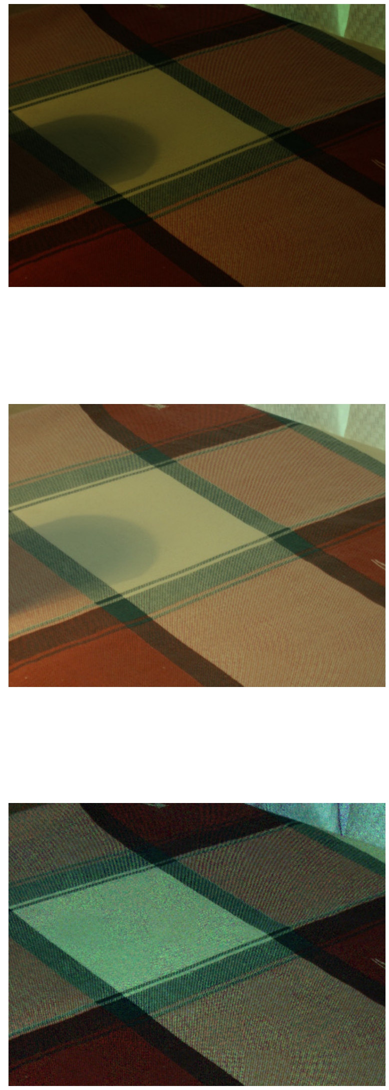


Figure 6. *Results: Example 2. Image* A, B *and* $B-A'$.