FLASH LIGHTENS GRAY PIXELS

Yanlin Qian[⋆], *Song Yan*[⋆], *Joni-Kristian Kämäräinen*[⋆], *Jiri Matas*[†]

[⋆]Computing Sciences, Tampere University
[†]Center for Machine Perception, Czech Technical University in Prague

ABSTRACT

In the real world, a scene is usually cast by multiple illuminants and herein we address the problem of spatial illumination estimation. Our solution is based on detecting gray pixels with the help of flash photography. We show that flash photography significantly improves the performance of gray pixel detection without illuminant prior, training data or calibration of the flash. We also introduce a novel flash photography dataset generated from the MIT intrinsic dataset.

Index Terms— spatial illumination estimation, gray pixel, flash photography, color constancy

1. INTRODUCTION

We address the *illumination estimation* problem which aims to measure the chroma of illumination in order to remove the color-bias from a captured image [1]. Illumination estimation can help in high-level vision tasks, *e.g.* object recognition, tracking [2] and intrinsic image decomposition. There exists a large number of related works, from the traditional non-learning approaches [3] to recent deep learning based approaches [4, 5, 6]. However, the vast majority of these works concentrate on the case of a single global illumination which is often an invalid assumption [7]. In this paper, we explore a more-complex less-optimistic setting – mixed illumination [1].

Spatially-varying illumination refers to that on a captured scene, each pixel captures different number of light phantoms when the camera shutter is on. In other words, all pixels do not share the same configuration of lights [7], which is the default assumption for single-illumination estimation. Compared to the single global illumination setting, the mixed illumination setting better corresponds to the real world [8], but is more challenging clearly, as it extends the ill-posed problem from one point to a spatial map [7], without extra knowledge or input.

To circumvent the hardness that spatially-varying illumination brings, efforts are put as follows: user guidance or human interaction is given as a supervisory signal [9, 10]; small patch is assumed to be cast by only one light [11]; light color

---

[1]We traverse related works and there exists multiple terms referring to the same thing (which may confuse readers), i.e. mixed illumination, spatially-varying illumination, multiple illumination, mixed lighting condition.
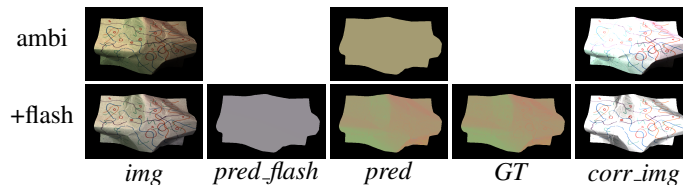

Fig. 1. A multi-illuminant image (*ambi*), Gray Pixel [13] outputs an erroneous prediction (single value, *pred*) compared to the groundtruth (*GT*). By virtue of the flash image (+*flash*) the proposed flash photography gray pixel provides an accurate spatial estimate (*pred*) and faithful corrected image (*corr_img*).

and number need to be known before experiments [12]. Unlike these methods, we make use of flash photography.

Flash photography refers to image processing techniques which use non-flash/flash image pairs. This technique is well adopted to spatial illumination estimation [14, 15], which assume each patch illuminated by one light and obtain decent results. Hui *et al.* [7] proposed a closed-form solution of spatial illumination for the case of a calibrated flash. In essence, flash calibration in [7] equals to knowing the "groundtruth" surface albedo in a flash-only image (thus yield "groundtruth" illumination). What's more, flash may appear in other forms, *i.e.* varying sunlight, cast shadow, which may be hard to calibrate.

In this paper, we propose a novel spatial illumination estimation method, using flash photography, **without** need of flash calibration and any other illuminant prior. Our method relies on gray pixel detection [13]. The original work assumes Lambertian surfaces, and then revisited and improved by [16, 17]. The original and extended gray pixel methods however fail in the case of mixed illumination (the top row in Fig. 1), but we analytically show how flash photography circumvents the problem and "lightens" gray pixel photometrically and in performance (bottom row in Fig. 1). The interplay of flash photography and gray pixel enables us to achieve a largely increased performance on our synthesized laboratory dataset and some real-world images, than running gray pixel methods alone, without knowing the flash color.

Our contributions are three-fold:

- We revisit and revise the gray pixel methods for the case of mixed illumination.
- Leveraging flash photography and the gray pixel method, we propose a novel learning-free and well-performing method for spatial illumination estimation.
- We propose a novel flash-photography dataset for benchmarking multi-illuminant methods. The dataset is based on the MIT intrinsic dataset.

The rest of this paper is organized as follows. Section 2 revisits gray pixel and its variants before we introduce the flash-photography gray pixel in Section 3. In Section 4 we describe the MIT intrinsic based dataset for our task. Section 5 covers the experiments and results. We conclude in Section 6.

2. GRAY PIXEL

Assuming one light source and narrow sensor response, Gray Pixel [13] is derived from the Lambertian model, given as:

$$I^c(p) = R^c(p) \max(n(p)^\mathsf{T} s, 0) \, l^c, \tag{1}$$

which shows the color channel c at the location p in image I is a function of a surface albedo R, surface normal n, light direction s and illumination color l.

Following the procedure in [13], applying *log* and a Mexican hat filter δ on Eq. 1 yields:

$$\delta \log I^c(p) = \delta \log R^c(p) + \delta \log \max(n(p)^\mathsf{T} s, 0) + \delta \log l^c. \tag{2}$$

A single light casting a small local neighborhood (the same color and direction), Eq. 2 simplifies to:

$$\delta \log I^c(p) = \delta \log R^c(p), \tag{3}$$

which is the core of gray pixel. $\delta \log I^c(p) = \delta \log I^{c'}(p)$, $\forall c, c' \in \{R, G, B\}$ defines a "pure gray pixel". To rank pixels *w.r.t.* "grayness", [13] defines the following grayness function, up to a scale:

$$g(p) = \sum_{c \in R,G,B} (\delta \log I^c(p) - \bar{\delta} \log I(p))^2 / \bar{\delta} \log I(p), \tag{4}$$

where $\bar{\delta} \log I(p)$ is the mean value of $\delta \log I^c(p)$. This method works robustly with single-illumination scenes where diffuse reflection (the Lambertian assumption) is dominant.

Then [16] augments the above Gray Pixel by replacing Eq. 4 with the following luminance-independent function:

$$g'(p) = \cos^{-1}\left(\frac{1}{\sqrt{3}} \frac{\|\delta \log I(p)\|_1}{\|\delta \log I(p)\|_2} \right), \tag{5}$$

where $\| \cdot \|_n$ refers to the ℓn norm. $g'(p) = 0$ refers to pure gray pixel. To remove spurious color pixels, [16] applies a mean shift clustering to choose the strongest mode – dominant illumination vector.

The mechanism to detect gray pixel is further improved in [17]. Different to [13, 16], which are based on the Lambertian model, Qian *et al.* [17] uses the *dichromatic reflection model* [18] to derive a set of more strict constraints for gray pixels:

$$g''(p) = \|[\, \delta(\log(I^R) - \log(|I|)), \delta(\log(I^G) - \log(|I|))] \,\|_2, \tag{6}$$

where $|I|$ refers to $(I^R + I^G + I^B)$. We refer readers to the original paper for more details. In Section 5 we report their performance in a mixed-lighting dataset and show how flash photography improves them.

3. FLASH GRAY PIXEL

In the sequel, we first investigate what will happen to the existing gray pixel methods in the case of mixed illumination. Then we propose a novel gray pixel method using flash photography, termed as *Flash Gray Pixel*.

Here we do analysis on the original Gray Pixel [13], but similar conclusion for [16, 17] can be inferred in an analogue manner. A scene is illuminated by N light sources or arbitrary type and color. To describe the image formation process in this case, Eq. 1 is modified to:

$$I^c(p) = R^c(p) \sum_i \lambda_i(p) \, l_i^c \, , \tag{7}$$

where $\lambda_i(p)$ represents the shading term $\max(n(p)^\mathsf{T} s_i, 0)$. Eq. 3 changes to:

$$\delta \log I^c(p) = \delta \log R^c(p) + \delta \log(\sum_i \lambda_i(p) \, l_i^c) \, . \tag{8}$$

Since that is a mixed illumination image, the light configuration varies from pixel to pixel, making the right most term in Eq. 8 non-zero and therefore Eq. 3 fails. This finding explains that mixed illumination hinders the performance of the gray pixel method, which motivates us to leverage flash photography.

When a flash light is present, the flash image I_f is expressed as:

$$I_f^c(p) = R^c(p)(\sum_i \lambda_i(p) \, l_i^c + \lambda_f(p) \, l_f^c), \tag{9}$$

where $\lambda_f(p)$ is the shading term of flash light at the position p and l_f the unit-norm chroma vector. Subtracting Eq. 7 from Eq. 9, we get a *flash-only image* I_{fo}:

$$I_{fo}^c(p) = R^c(p)\lambda_f(p) \, l_f^c, \tag{10}$$

which is a more solid ground for searching gray pixels. In other terms, flash image helps to remove the negative effect of spatially-varying light configurations and gray pixels are now *flash gray pixels*.
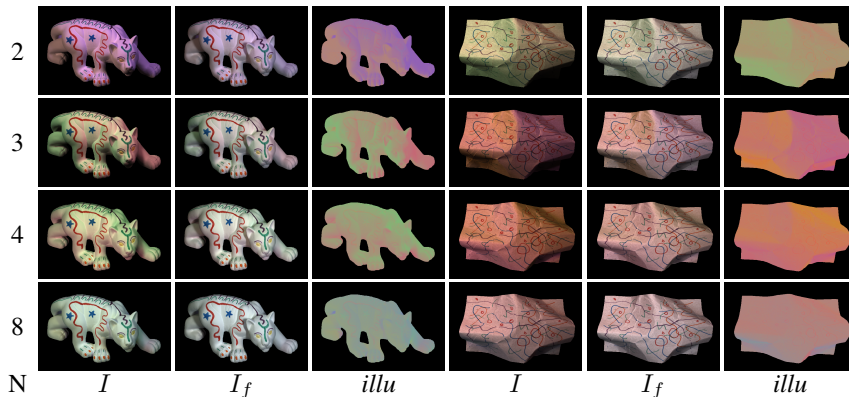
Fig. 2. Example images from the generated flash photography dataset. Each sample is a triplet: {no-flash image I, flash-on image I_f, illumination ground truth map}. Images are generated from the MIT intrinsic dataset by changing the colors and mixing the original directional illumination images.

Flash gray pixels can be found from the grayness map of a residual flash/no-flash image. To select flash gray pixels robustly, we follow [13] to compute flash-only illumination component for each pixel: K-means is used to cluster the top $N\%$ gray pixels into preset M clusters; then the illumination at location p is computed using:

$$L_{fo}^c(p) = \sum_{m=1}^{M} \omega_m L_m^c, \tag{11}$$

where L_m refers to the average illumination for the cluster m and ω_m controls the connection between the pixel $I(x,y)$ to the cluster m, unfolded as:

$$\omega_m = \frac{e^{-\frac{D_m}{2\sigma^2}}}{\sum_{n=1}^{M} e^{-\frac{D_n}{2\sigma^2}}}, \tag{12}$$

where D_m is the Euclidean distance from the pixel to the centroid of the cluster m, encouraging nearby pixels to share a similar flash-only illumination.

Combining the flash-only illumination $L_{fo}^c(p)$ with Eq. 9 allows to color-corrected the image I_{fo} by

$$I_{gray}^c(p) = I_{fo}^c(p)/L_{fo}^c(p), \tag{13}$$

and the mixed illumination is:

$$L^c(p) = I^c(p)/I_{gray}^c(p) . \tag{14}$$

By estimating flash-only illumination, color-corrected image and the spatial illumination in a sequential order, we can capture illumination boundaries (Sec. 5). Flash gray pixel can be filled by more advanced gray pixel methods for further improvement.

4. DATASET

Following [7] (while their dataset is not public), we adapted the MIT Intrinsic benchmark [19] for our task [2]. This dataset

was originally collected for the intrinsic image decomposition task, containing 20 single objects illuminated by uncalibrated whitish light sources from 10 different directions. This property allows us to render each image in the combinations of $1 - 9$ directed lights with arbitrary chroma to compose a new no-flash image I. The 6-th direction is always roughly frontal and was thus used as the flash source and together with I forms the flash image I_f. We generated I and I_f with varying number N of light sources, from 2 to 8. Note that even for the easiest case ($N = 2$), an obviously-gray patch can be simultaneously affected by two light sources, violating the global illumination assumption and failing the original gray pixel methods as demonstrated in Fig. 2 (proven in Sec.5).

Considering the fact that gray pixel methods are designed for realistic consumer images "in the wild" [13] that contain at least a few gray pixels, we left out the 5 chromatic objects[3]. In total, the new dataset contains 105 flash/no-flash image pairs with spatial illumination map ground truth ((15 objects and 7 choices of N).

5. EXPERIMENTS

Our setup is the following: we run the three variants of Gray Pixel methods with flash/no-flash image pairs. On each image the top 10% gray pixels are selected to cover enough area. The cluster number is set to $M = N$, which is the number of illuminants.

Evaluation metric is the standard average angular error [13].

Results are summarized in Table 1 which shows the performance of flash gray pixel on the dataset in Section 4. "GP+f" refers to flash gray pixel on the basis of the original gray pixel method "GP". Results show that flash photography extensions of all gray pixel variants [13, 16, 17] systematically improve

---

[2]There are recent datasets that provide flash/no-flash image pairs [20], but these are unsuitable for our purposes due to unknown illumination number.

[3]"apple, pear, frog2, potato, turtle"

| Method/N	2	3	4	5	6	7	8	all
GP [13]	5.86 6.15	5.33 6.65	4.01 5.74	3.73 5.34	3.04 4.93	3.59 4.80	3.37 4.80	4.07 5.49
GP+f	2.37 3.49	2.32 3.85	2.35 3.79	2.39 3.94	2.39 3.80	2.39 3.85	2.39 3.88	2.37 3.80
MSGP [16]	4.91 5.66	5.25 6.03	4.10 5.23	3.52 4.81	3.17 4.42	3.54 4.35	2.87 4.52	4.10 5.00
MSGP +f	2.34 3.17	2.29 3.32	2.34 3.31	2.34 3.40	2.33 3.38	2.50 3.72	2.30 3.46	2.34 3.40
DGP [17]	6.03 6.26	5.26 6.72	4.52 5.89	4.08 5.47	3.42 5.09	3.97 55.07	3.50 4.99	4.13 5.64
DGP +f	2.37 3.73	2.34 4.00	2.37 3.96	2.39 4.01	2.40 3.93	2.40 4.02	2.39 3.99	2.37 3.93
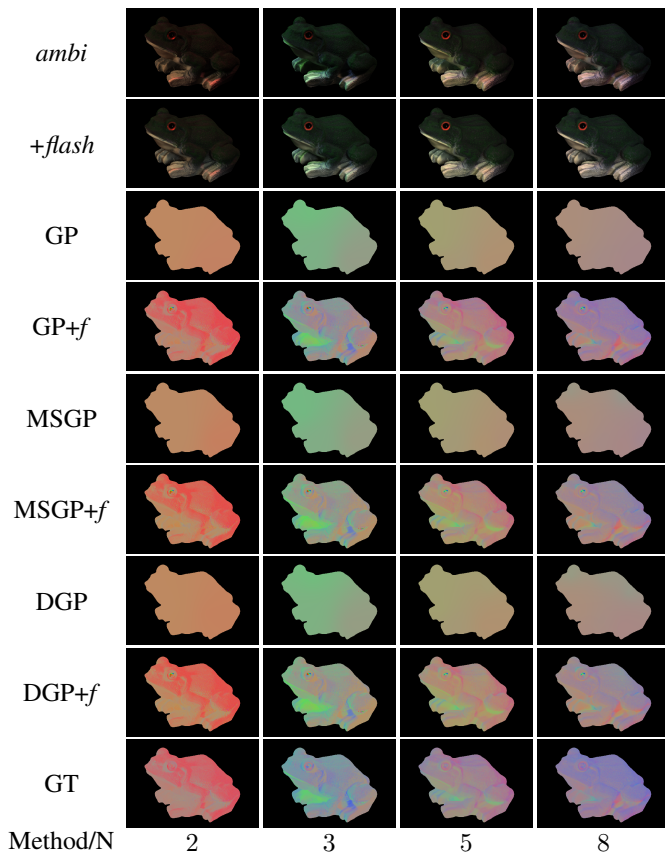
Fig. 3. Qualitative comparison on estimated illumination of flash gray pixels, with N light sources (2,3,5,8).
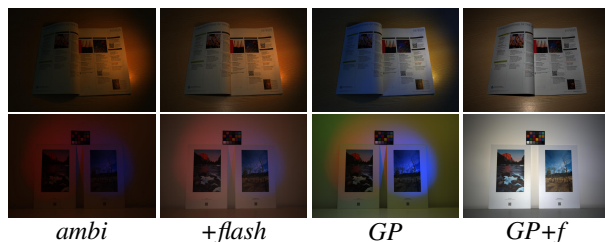

Fig. 4. Qualitative comparison on color-corrected image of flash gray pixels in real-world cases. Macbeth ColorChecker is excluded for illumination estimation.

the results. For all three methods, the improvement over the whole dataset is over 40% in median and 30% in mean. The flash photography variants achieve the sufficient color constancy accuracy ($\leq 3.0°$) in almost all cases. Fig. 3 illustrates predicted illumination between gray pixel methods and their flash variants. It is clear that the original versions cannot find the fine-grained details of mixed illumination. For example, the frog back and stomach are cast with different colors that confuses the original GP methods.

Among all gray pixel methods, the original GP [13] suffers from the mixed illumination the most. MSGP [16] and DGP [17] perform slightly better, but not to a satisfying degree without the flash. With flash gray pixel, all three methods perform similarly thus verifying the efficiency of flash photography. DGP [17] does not perform better than the original GP which is due to the fact that our dataset does not contain specular reflectance components.

As the number of diverse illuminants increases, all GP methods performs better. This can be explained by the fact that a large number of lights with various colors additively mix toward a whitish color. Flash gray pixel variants are not affected by this, obtaining consistent results.

Images from real-world scenes (Fig. 4, images retrieved from [7]), show that the proposed methods also produce high quality results (*e.g.* the white wall is white).

6. CONCLUSION

In this paper we reconsider gray pixel detection through the medium of flash photography. We find that computing a residual map from the flash/no-flash pair, gray pixel methods can perform at full capacity, measuring pixel-wise grayness effectively and providing spatial color constancy. The method is pragmatic – it is lightweight, does not need any illuminant prior, training, flash light calibration or user input. In future, we will move one step further to what camera manufacturers are looking for – exploring indoor-outdoor mixed illumination.

7. REFERENCES

[1] Graham D Finlayson, "Colour and illumination in computer vision," *Interface focus*, vol. 8, no. 4, pp. 20180008, 2018.

[2] David H Foster, "Color constancy," *Vision research*, vol. 51, no. 7, pp. 674–700, 2011.

[3] David H Brainard and Brian A Wandell, "Analysis of the retinex theory of color vision," *JOSA A*, vol. 3, no. 10, pp. 1651–1661, 1986.

[4] Yanlin Qian, Ke Chen, Joni Kämäräinen, Jarno Nikkanen, and Jiri Matas, "Deep structured-output regression learning for computational color constancy," in *ICPR*, 2016.

[5] YuanMing Hu, Baoyuan Wang, and Stephen Lin, "Fully convolutional color constancy with confidence-weighted pooling," in *CVPR*, 2017.

[6] Yanlin Qian, Ke Chen, Joni Kämäräinen, Jarno Nikkanen, and Jiri Matas, "Recurrent color constancy," in *ICCV*, 2017.

[7] Zhuo Hui, Aswin C Sankaranarayanan, Kalyan Sunkavalli, and Sunil Hadap, "White balance under mixed illumination using flash photography," in *2016 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2016, pp. 1–10.

[8] Hamid Reza Vaezi Joze and Mark S Drew, "Exemplar-based color constancy and multiple illumination," *TPAMI*, vol. 36, no. 5, pp. 860–873, 2014.

[9] Dani Lischinski, Zeev Farbman, Matt Uyttendaele, and Richard Szeliski, "Interactive local adjustment of tonal values," in *ACM Transactions on Graphics (TOG)*. ACM, 2006, vol. 25, pp. 646–653.

[10] Ivaylo Boyadzhiev, Kavita Bala, Sylvain Paris, and Frédo Durand, "User-guided white balance for mixed lighting conditions," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, pp. 200, 2012.

[11] Christian Riess, Eva Eibenberger, and Elli Angelopoulou, "Illuminant color estimation for real-world mixed-illuminant scenes," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 782–789.

[12] Eugene Hsu, Tom Mertens, Sylvain Paris, Shai Avidan, and Frédo Durand, "Light mixture estimation for spatially varying white balance," in *ACM Transactions on Graphics (TOG)*. ACM, 2008, vol. 27, p. 70.

[13] Kai-Fu Yang, Shao-Bing Gao, and Yong-Jie Li, "Efficient illuminant estimation for color constancy using grey pixels," in *CVPR*, 2015.

[14] Jeffrey M DiCarlo, Feng Xiao, and Brian A Wandell, "Illuminating illumination," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2001, vol. 2001, pp. 27–34.

[15] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama, "Digital photography with flash and no-flash image pairs," in *ACM transactions on graphics (TOG)*. ACM, 2004, vol. 23, pp. 664–672.

[16] Yanlin Qian, Said Pertuz, Jarno Nikkanen, Joni-Kristian Kämäräinen, and Jiri Matas, "Revisiting gray pixel for statistical illumination estimation," in *International Conference on Computer Vision Theory and Applications*, 2019.

[17] Yanlin Qian, Jarno Nikkanen, Joni-Kristian Kämäräinen, and Jiri Matas, "On finding gray pixels," in *CVPR*, 2019.

[18] Steven A Shafer, "Using color to separate reflection components," *Color Research & Application*, vol. 10, no. 4, pp. 210–218, 1985.

[19] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *CVPR*, 2009.

[20] Y. Aksoy, C. Kim, P. Kellnhofer, S. Paris, M. Elgharib, M. Pollefeys, and W. Matusik, "A dataset of flash and ambient illumination pairs from the crowd," in *ECCV*, 2018.