

What does Plate Glass Reveal about Camera Calibration?

Qian Zheng^{1*} Jinnan Chen¹ Zhan Lu¹ Boxin Shi^{2,3*} Xudong Jiang¹ Kim-Hui Yap¹ Ling-Yu Duan^{2,3} Alex C. Kot¹

¹School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

²National Engineering Laboratory for Video Technology, Department of CS, Peking University, Beijing, China

³Peng Cheng Laboratory, Shenzhen, China

{zhengqian, exdjiang, ekhyap, eackot}@ntu.edu.sg, {jinnan001, zlu015}@e.ntu.edu.sg, {shiboxin, lingyu}@pku.edu.cn

Abstract

This paper aims to calibrate the orientation of glass and the field of view of the camera from a single reflection-contaminated image. We show how a reflective amplitude coefficient map can be used as a calibration cue. Different from existing methods, the proposed solution is free from image contents. To reduce the impact of a noisy calibration cue estimated from a reflection-contaminated image, we propose two strategies: an optimization-based method that imposes part of though reliable entries on the map and a learning-based method that fully exploits all entries. We collect a dataset containing 320 samples as well as their camera parameters for evaluation. We demonstrate that our method not only facilitates a general single image camera calibration method that leverages image contents but also contributes to improving the performance of single image reflection removal. Furthermore, we show our byproduct output helps alleviate the ill-posed problem of estimating the panorama from a single image.

1. Introduction

Camera calibration [65] has been a widely studied problem, due to its indispensability for a broad category of vision tasks such as scene understanding [50, 42], metrology [6], 3D inference [55, 8], and augmented reality [13, 15, 20]. Classic camera calibration requires multiple images containing specific objects or patterns (e.g., [26, 36]). Calibrating a complete camera model using a single image is a rather challenging problem. With comprehensive priors (e.g., horizon line [16]) learned by deep neural networks, recent progress shows a successful estimation of extrinsic and intrinsic parameters (e.g., [32]) from a single noise-free image in the wild.

The classical physically based formation of a reflection-contaminated image can be formulated as [25, 53]¹,

$$\mathbf{I} = (1 - \Omega) \circ \mathbf{T} + \Omega \circ \mathbf{R}, \quad (1)$$

*Corresponding authors.

¹For simplicity, we consider the reflection image \mathbf{R} can be blur or contain ghost effects which is consistent with prior works (e.g., [64, 49]).

where \mathbf{I} , \mathbf{T} , \mathbf{R} are the reflection-contaminated image, transmission image, and reflection image, respectively. ‘ \circ ’ denotes an element-wise multiplication operator. Ω is the reflective amplitude coefficient map. The Fresnel equations describe the relative amplitude of transmission and reflection, based on which Kong *et al.* [25] derive a formation of Ω for glass with double surfaces. One of the properties in their formation is that Ω is a monotonically increasing function of the angle of incidence Θ (AoI, w.r.t. glass plane). This property is widely studied in the topic of multi-image reflection separation using polarization (e.g., [25, 54, 33]), while seldom considered in the context of single image reflection separation (e.g., [7, 53]). On the other hand, thanks to the rapid progress in studying single image reflection removal, recent methods (e.g., [64, 53]) provide a reliable estimation of \mathbf{T} , \mathbf{R}) can be used to calculate Ω pixel-wisely.

In this paper, we observe that Ω or Θ reflects useful cues for camera calibration which motivates us to estimate camera parameters from a single reflection-contaminated image (formed by putting a piece of plate glass between the left and right scenes in Figure 1 row ①). More specifically, we attempt to solve for extrinsic (tilt, roll, w.r.t. the glass plane) and intrinsic (focal length) parameters. In this paper, these camera parameters are represented as the orientation of glass \mathbf{n} and the horizontal field of view (FoV) of camera H (Figure 1 row ②). Figure 1 row ③ intuitively shows our observations about how these two parameters affect the reflection. As can be observed, the orientation of glass \mathbf{n} affects the direction of reflective intensity change (left and middle figures), the angle of H affects the magnitude of reflective intensity change (middle and right figures). Different from previous works of camera calibration that leverage image contents from a noise-free image (e.g., [56, 16, 32]), our method exploits Ω (free from image contents, Figure 1 row ④) as the calibration cue.

We solve for \mathbf{n} and H by incorporating a reflection separation method (taking \mathbf{I} as input) into a calibration algorithm (taking an estimated Ω as the input). The major difficulty is brought by the calibration cue Ω that may be inaccurate. Based on our observation that estimating \mathbf{n} and H from Ω is to solve an over-determined system, we present two strategies to achieve robust calibration, *i.e.*, an optimization-based method that imposes part of though reliable entries of

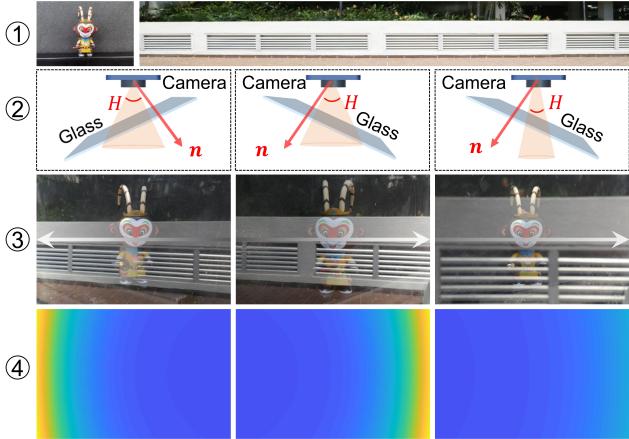


Figure 1. ① An object with simple background as transmission (left) and the scene with (approximately) uniform illumination as reflection (right). ② Geometric camera models with parameters \mathbf{n} and H for reflection-contaminated images capture. ③ Reflection-contaminated images \mathbf{I} captured with geometric camera models above. White arrows with different orientations and gray scale gradients indicate reflective intensity change with different directions (left and middle) and magnitude (middle and right). ④ Visualizations of the corresponding reflective amplitude coefficient maps Ω , simulated according to formation model in [25]. Blue to yellow means small to large values.

an estimated Ω and a learning-based method to indiscriminately exploit all entries of it. To evaluate our methods, we collect a dataset containing 320 reflection-contaminated images as well as their calibration parameters, with various scenes, glass types, \mathbf{n} , and H . To summarize, our major contributions are:

- We propose an optimization-based method to recover extrinsic (glass orientation) and intrinsic (horizontal FoV) parameters of the camera from a reflection-contaminated image. To the best of our knowledge, this is the first work to address the single image camera calibration problem in the context of glass reflection.
- We further propose a learning-based method to jointly estimate the reflective amplitude coefficient map and camera parameters. We collect a validation dataset containing 320 reflection-contaminated images as well as their calibration parameters to evaluate our methods.
- We demonstrate that our calibration method, which is free from image contents, facilitates both single image camera calibration and single image reflection removal. As an additional application, it helps alleviate the ill-posedness of single image panorama estimation.

2. Related Work

The key differences of our method from existing single image camera calibration methods are that we take a

reflection-contaminated image instead of a noise-free one as inputs and we exploit the calibration cue of the reflective amplitude coefficient map instead of image contents. To the best of our knowledge, there is little related work directly addressing the problem of single image camera calibration in the context of plate glass reflection. We thereby briefly review the literature of two relevant topics: single image reflection removal and single image camera calibration.

2.1. Single Image Reflection Removal

Single image reflection separation or removal is a typical ill-posed problem that estimates more than two unknowns from a single input. To make this problem tractable, several priors are made in the literature. Some of them are based on the physical form of the reflection, *e.g.*, ghost effects due to double surfaces of glass [40], or blurry reflection due to out-of-focus and limited depth of field of a camera [30, 31, 62]. Some other priors are observed from transmission and reflection, *e.g.*, the sparsity of gradient due to natural image priors [29, 2], or image contents priors due to known scenes or objects [46]. Data-driven methods are proposed to narrow down the solution space of separation results. Some of them constrain outputs by a reconstruction loss function [53], or introduce additional constraints for information such as edge [7], gradient [48], perceptual metric [64], context [49]. Some others design specific units, *e.g.*, a cascade network [61], or context encoding modules [52].

These methods formulate Ω as a scalar (*e.g.*, [61, 49]), or a spatially-varying variable (*e.g.*, modeled by a Gaussian distribution [7, 64], generated by a GAN [53]). In contrast, we consider Ω as a function of Θ which involves constraints from the geometric camera model.

2.2. Single Image Camera Calibration

Earlier work for the problem of single image camera calibration relies on a manually inserted calibration target [14, 44, 65]. Geometric based methods calibrate a camera through detecting some specific patterns, such as vanishing lines or points [39, 27], concentric circles [21], coplanar circles [5, 59], plumb-lines [34], repeated patterns [36], or lighting cues [26, 57]. Recently, learning-based methods are proposed to deal with a single image in the wild. These methods solve for different components of calibration parameters, such as vanishing points [63] (combined with geometric based methods), FoV [56], horizon line [58] (to estimate the extrinsic rotation matrix), the radial distortion parameter [38], the extrinsic rotation matrix and FoV [16], or the extrinsic rotation matrix together with intrinsic parameters of FoV and radial distortion [32]. Different from these methods that leverage image contents from a noise-free image for calibration, our method exploits the calibration cue of the reflective amplitude coefficient map from a reflection-contaminated image.

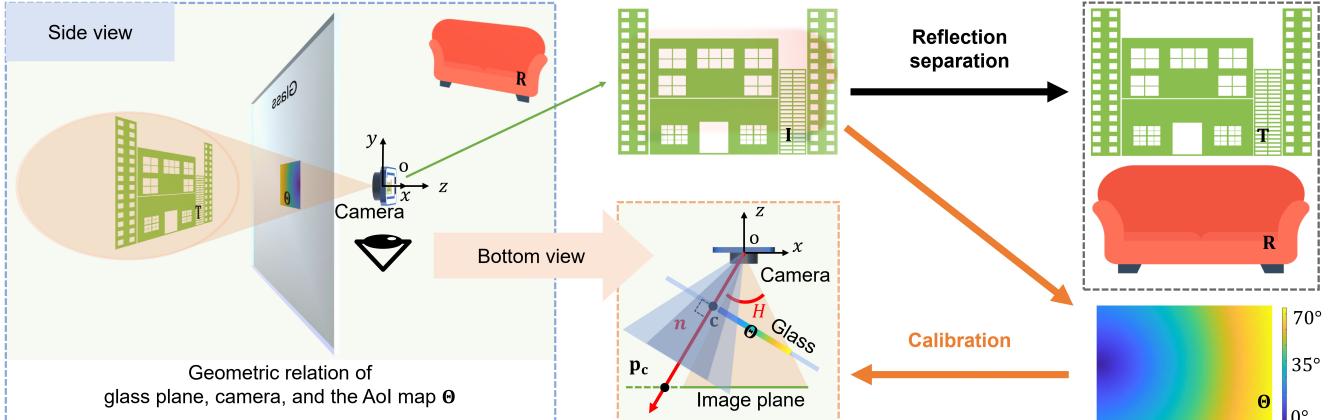


Figure 2. Illustrations of geometric relation between Θ and n, H , and the Cartesian coordinate system used in this paper where x - y plane is parallel to the image plane. For bottom view: each blue cone indicates light direction with the same Angle of Incidence (AoI). The vector of glass orientation n intersects the glass plane and image plane at c and p_c .

Our calibration cue Ω comes from the pattern of imaged concentric circles (to be introduced in Section 3.2). There are several previous works calibrating a camera based on imaged concentric circles, *e.g.*, multiple views [23, 22, 18], multiple pairs [21], or known radii [9, 1]. However, these methods cannot be directly used to our problem because the imaged concentric circles from Ω are always incomplete, noisy, with only one view, one pair, and unknown radii.

3. Proposed Methods

In this paper, we assume that reflection occurs over a piece of plate glass [43], which is homogeneous, isotropic [3], and fills the whole FoV. We adopt the pin-hole camera model with square pixels and centered principal points, similar to a recent work [32] for single image camera calibration.

3.1. Preliminary

We study camera parameters of the extrinsic rotation matrix w.r.t. the glass plane and the intrinsic FoV, from a single image with a size of $h \times w$. We represent these two parameters as the orientation of glass plane n , and H , respectively. The geometric relation of the AoI map Θ and camera parameters is illustrated in Figure 2, which can be expressed as $\Theta = P(n, H)$. To be more specific, considering a point $p = (u, v)^\top$ on the image plane, its coordinate in 3D space is $(u, v, -f)^\top$, where $f = \frac{w}{2\tan(\frac{H}{2})}$. The AoI $\theta \in \Theta$ of p can be computed as $\theta = \arccos \frac{\langle p, n \rangle}{\|p\|_2}$. Inspired by the analytic formation of Ω in [25], we represent $\Omega = F(\Theta, \kappa)$, where κ is the refractive index. Therefore, the calibration problem in this paper is to inversely recover n and H from an input I with a formation model such that,

$$I = (1 - \alpha(n, H, \kappa)) \circ T + \alpha(n, H, \kappa) \circ R, \quad (2)$$

where $\Omega = \alpha(n, H, \kappa) = F(P(n, H), \kappa)$.

Our basic idea to solve this problem is to incorporate a reflection separation method (taking I as input) into a calibration algorithm (taking an estimated Ω as the input). Without losing of generality, we can adopt any state-of-the-art reflection separation method (*e.g.*, [64, 53]) to approximately fit Ω based on Equation (1). We then focus on how to robustly solve the inverse problem with $\Omega = \alpha(n, H, \kappa)$, *i.e.*, recovering n, H from an estimated Ω . We observe that this inverse problem forms an over-determined system because all entries of Ω are expected to be determined by about 5 unknowns (*e.g.*, κ , radial distortion, n , and H) while the number of equations is much larger ($h \times w$). However, we only have a noisy estimate of Ω , which is the biggest challenge of accurate calibration. To this end, we propose two strategies to achieve robust calibration, *i.e.*, an optimization-based method that imposes part of though reliable entries of an estimated Ω and a learning-based method to indiscriminately exploit all.

We achieve the calibration through estimating the coordinate of $p_c = (u_c, v_c, -f)^\top$, which is the intersection of n and image plane, as shown in Figure 2. Once p_c is obtained, the glass orientation n and horizontal FoV H can be calculated by

$$\begin{aligned} n &= \frac{p}{\|p\|_2}, \\ H &= 2 \arctan \frac{w}{2f}. \end{aligned} \quad (3)$$

3.2. An Optimization-based Method

As can be observed from the bottom view of Figure 2, Ω contains specific patterns of conics (intersections of two blue cone surfaces and image plane in this figure), whose corresponding cones share the same apex (origin o) and the same axis (n). Such a pattern is called imaged or projected concentric circles [21, 22]. Our optimization-based method is to impose the imaged concentric circles to estimate p_c .

Assumption. Previous calibration methods using imaged concentric circles cannot be directly applied to our problem due to the under-constrained input, as mentioned in Section 2.2. To make this problem more tractable, we take the analytic formation in [25] for Ω (glass with double surfaces). Given an entry $\alpha \in \Omega$ and its $\theta \in \Theta$, we have²

$$\alpha = F(\theta, \kappa) = \frac{1}{2}(\mathcal{R}_{\perp} + \mathcal{R}_{\parallel}) = \frac{1}{2}\left(\frac{2\mathcal{R}_{\perp}^s}{1 + \mathcal{R}_{\perp}^s} + \frac{2\mathcal{R}_{\parallel}^s}{1 + \mathcal{R}_{\parallel}^s}\right), \quad (4)$$

where refractive index $\kappa = 1.474$ as suggested by [25], subscripts ‘ \perp ’ and ‘ \parallel ’ represent polarized components perpendicular and parallel to the glass plane, \mathcal{R}_{\perp}^s and $\mathcal{R}_{\parallel}^s$ are the polarized components for single surface,

$$\mathcal{R}_{\perp}^s = \left(\frac{\cos \theta - \kappa \cos \theta_t}{\cos \theta + \kappa \cos \theta_t}\right)^2, \quad \mathcal{R}_{\parallel}^s = \left(\frac{\cos \theta_t - \kappa \cos \theta}{\cos \theta_t + \kappa \cos \theta}\right)^2, \quad (5)$$

where $\theta_t = \arcsin(\frac{1}{\kappa} \sin \theta)$ according to Snell’s law. Figure 3 shows that both $F(\theta, \kappa)$ and $F_{\theta}(\theta, \kappa)$ are monotonically increasing functions of θ . Kindly note that polarized images are not required at all and the formation above is just for the analytic formation of Ω in our method.

Detecting reliable points to fit conics. Note that points on the contour of a conic sharing the same value of θ (or α due to their monotonically increasing relation). Therefore, a contour is a set of points sharing the same α , or fall within a given small interval. A larger interval is expected to include points of a contour tolerating larger errors of noisy Ω . However, it may also include outliers from other contours. Considering that $F(\theta, \kappa)$ and $\frac{\partial F(\theta, \kappa)}{\partial \theta}$ are monotonically increasing functions, we obtain two sets of reliable points by a fixed size of an interval decreasing from large numbers. This procedure is outlined in Algorithm 1. Parameter η is set to 80 in our implementation. We then fit two conics from these points according to method [45].

Solving for \mathbf{p}_c . We first estimate an initial guess of coordinate of \mathbf{p}_c w.r.t. image plane, and represent it as $(u_c^0, v_c^0)^\top$. As shown in Figure 2, point \mathbf{p}_c is the projected circular point [12], therefore, $(u_c^0, v_c^0)^\top$ can be estimated by method [22]. We then estimate \mathbf{p}_c by solving

$$\begin{aligned} \min_{\mathbf{p}_c} & \sum_{i=1}^2 \sum_{j=1}^{m_i} (F(\theta_{ij}, \kappa) - \alpha_{ij})^2, \\ \theta_{ij} &= \arccos \frac{\langle \mathbf{p}_c, \mathbf{p}_{i,j} \rangle}{\|\mathbf{p}_c\|_2 \cdot \|\mathbf{p}_{i,j}\|_2}, \quad \text{s.t., } f > 0, \end{aligned} \quad (6)$$

where $\mathbf{p}_{ij}, i = 1, 2, j = 1, 2, \dots, m_i$ are points in the i -th point sets (m_i is points number) we obtained through Algorithm 1, α_{ij} is their corresponding reflective amplitude coefficients. Note that the x - and y -axis coordinates of \mathbf{p}_{ij} and α_{ij} are known, and the z -axis coordinate of \mathbf{p}_{ij} is the same to that of \mathbf{p}_c . We solve Equation (6) based on the Quasi-Newton method [35] initialized by $\mathbf{p}_c^0 = (u_c^0, v_c^0, -w)^\top$.

²We consider function F takes either a scalar θ or matrix Θ as the input and outputs either a scalar α or matrix Ω accordingly.

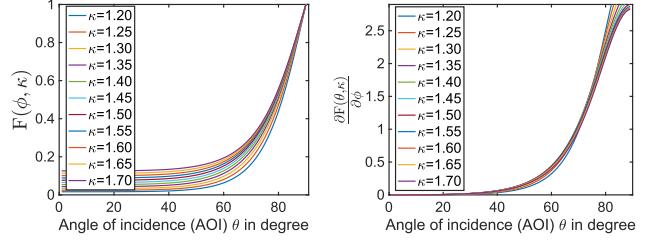


Figure 3. Curves of $F(\theta, \kappa)$ and $\frac{\partial F(\theta, \kappa)}{\partial \theta}$ varying with AoI θ falling in an interval of $[0^\circ, 90^\circ]$ regarding different κ .

Algorithm 1 Detecting two sets of reliable points.

```

set  $a = \frac{\max(\Omega) - \min(\Omega)}{w+h}$ ,  $b = \max(\Omega) - \eta a$ ,
repeat
    repeat
        set an interval  $J = (b-a, b+a]$ ,
         $b = b-a$ ,
    until more than  $(h+w)$  points whose  $\alpha$  fall within  $J$ ,
    obtain a set of points,
     $b = b - \eta a$ ,
until obtain two sets of points or  $b-a < \min(\Omega)$ 

```

3.3. A Learning-based Method

The optimization-based method above achieves our goal for clean data, but it is hardly applicable in a real scenario since it requires a known formation of α with a given refractive index κ . Besides, the fitted conics may cause accumulative errors for the estimation of \mathbf{p}_c . Considering that all entries of Ω are expected to be determined by about 5 unknowns (including the radial distortion parameter), we propose a learning-based method that fully exploits all entries of Ω to overcome above limitations. We jointly optimize a reflection separation network (denoted by \mathcal{S}) and a calibration network (denoted by \mathcal{C}) such that,

$$(\mathbf{T}, \mathbf{R}, \Omega) = \mathcal{S}(\mathbf{I}), \quad \mathbf{p}_c = \mathcal{C}(\Omega). \quad (7)$$

We borrow the method in [53] for our reflection separation network, which can also be replaced by other reflection separation methods such as [64]. An overview of our framework is shown in Figure 4³. We use notations with subscript ‘est’ to represent estimated variables and those with ‘gt’ to represent their ground truth.

Calibration network. Considering \mathbf{p}_c and Ω are spatially consistent, our calibration network has the property of pseudo-invariance regarding to rotation and flipping:

$$G(\mathbf{p}_c) = \mathcal{C}(G(\Omega)), \quad (8)$$

³The down-sampling or up-sampling block in reflection separation network is concatenated by units of convolution or deconvolution, normalization, and ReLU. For implementation details, please refer their paper [53] and code (<https://github.com/csqiangwen/Single-Image-Reflection-Removal-Beyond-Linearity>).

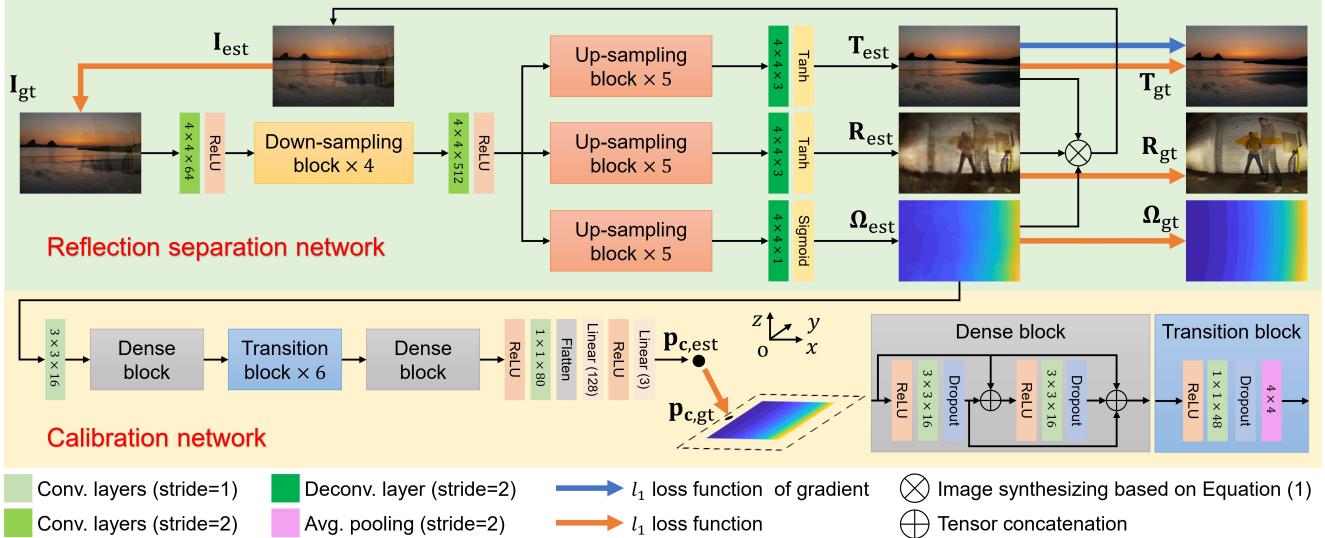


Figure 4. The framework of our learning-based method. The reflection separation network predicts a reflective amplitude coefficient map Ω taking a reflection-contaminated image \mathbf{I} as input. The calibration network estimates coordinate of \mathbf{p}_c taking the predicted Ω as input.

where G represents a flipping or rotation operation. To avoid invalid regions from rotation, this paper only considers flipping operation, *i.e.*, left-right G_1 , up-down G_2 , and their combination G_3 . For simplicity of notation, we use G_0 to represent the operation without flipping, *i.e.*, $\Omega = G_0(\Omega)$, $\mathbf{p}_c = G_0(\mathbf{p}_c)$.

Loss functions. We alternatively optimize S and C . The loss functions of \mathcal{L}^S and \mathcal{L}^C are,

$$\mathcal{L}^S = \mathcal{L}_{WT} + \mathcal{L}_e, \quad \mathcal{L}^C = \mathcal{L}_e + \mathcal{L}_g, \quad (9)$$

where \mathcal{L}_{WT} is the same loss function as in [53], which is composed of five ℓ_1 loss functions for transmission and its gradient, reflection, reconstructed or synthesized image, and reflective amplitude coefficient map. \mathcal{L}_e and \mathcal{L}_g are ℓ_1 loss functions for \mathbf{p}_c ,

$$\begin{aligned} \mathcal{L}_e &= \frac{1}{4} \sum_{i=0}^3 \|\mathcal{C}(G_i(\Omega_{\text{est}})) - G_i(\mathbf{p}_{c,\text{gt}})\|_1, \\ \mathcal{L}_g &= \frac{1}{4} \sum_{i=0}^3 \|\mathcal{C}(G_i(\Omega_{\text{gt}})) - G_i(\mathbf{p}_{c,\text{gt}})\|_1, \end{aligned} \quad (10)$$

where Ω_{est} is estimated from reflection separation network, *i.e.*, $\Omega_{\text{est}} = S(\mathbf{I}_{\text{gt}})$. Both of these neural networks are trained using Adam solver [24] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The batch size is set to 4, and learning rates are 0.0002 and 0.0001 for the reflection separation network and calibration network, respectively.

3.4. Data Preparation

Training dataset. Following strategies adopted in learning-based methods for single reflection separation (*e.g.*, [49, 53]) and single image camera calibration

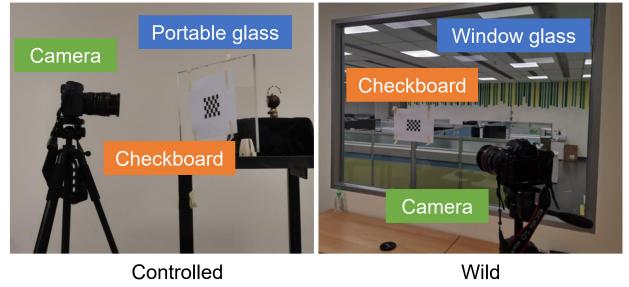


Figure 5. Data collection for controlled and wild scenarios.

(*e.g.*, [32, 56]), we synthesize our training data. We randomly select 22082 images with a focal length between 24mm and 105mm from the FocalLens dataset [60] and divide these images into two sets with paired focal length. We then select \mathbf{T} and \mathbf{R} with same focal length from these two sets respectively, and randomly add noise of blurry, ghost effects, or luminance decay to \mathbf{R} by convoluting it with a Gaussian kernel, one-pulse or two-pulse shaping filter to simulate different types of reflection [47]. We calculate the AoI map Θ according to a randomly generated variables of glass orientation \mathbf{n} and horizontal FoV H . Then Ω is calculated based on Equation (4) with a random κ . Finally, we synthesize 11041 reflection-contaminated images \mathbf{I} according to Equation (1).

Validation dataset. To demonstrate the generality and practicability of the proposed method, we collect 320 real data for validation. The focal length or FoV is automatically provided by the camera and we calibrate the orientation of glass through a checkboard pasting on it (Figure 5). These 320 images are captured with 8 different commonly used focal lengths, each for about 40 images (between 35 and 45).

Regarding the glass, 160 out of 320 images are captured in a controlled scenario and the remaining in the wild scenario. Figure 5 shows the data collection for these scenarios. For the controlled one, we capture images with portable glass. Therefore, we are able to obtain the transmission (by removing the glass) and control the glass orientations (in 5 levels, w.r.t. the angle of the pan). Transmission or reflection images in this scenario are also controlled to be simple. For the wild one, we capture images with fixed natural glass found in daily scenarios, such as glass doors or walls, shop windows, and glass-paneled cabinets or cases.

4. Experimental Results

In this section, we report our evaluation for camera calibration and reflection separation. Moreover, we provide an interesting byproduct to estimate a panorama based on the results of our method.

4.1. Evaluation for Camera Calibration

We conduct experiments for camera calibration on our validation dataset with 320 samples. We compute errors of the estimated glass orientation \mathbf{n} and horizontal FoV H by comparing them with their ground truth, with metrics of the angular error e_n (in degree) and the absolute error e_H (in degree), respectively.

4.1.1 Overall Performance

Considering camera distortion is not addressed in our methods, we report results regarding inputs with or without distortion correction. Table 1 displays the quantitative performance of our optimization-based method (represented as ‘Baseline’) and learning-based method (represented as ‘Ours’). As can be observed, both our optimization-based method and learning-based method achieve better performance with UNDISTORTED inputs, because our optimization-based method requires undistorted images to maintain patterns of conics for a better initialization of Equation (6) and our learning-based method is trained with undistorted Ω . Our learning-based method outperforms our optimization-based method, because it fully exploits all entries of an estimated reflective amplitude coefficient map Ω and implicitly fits unknowns for function $\alpha(\mathbf{n}, H, \kappa)$ in Equation (2). Considering the superior performance of our learning-based method and the applicability to real data results in the following sections are from our learning-based method with distorted inputs.

4.1.2 Free from Image Contents

Different from existing methods for single image camera calibration that leverage image contents for calibration, our method takes the reflective amplitude map as the calibration cue. Therefore, the performance of our method is expected to be free from image contents or scenarios. Be-

Table 1. Quantitative performance of our methods in terms of e_n and e_H with distorted and undistorted inputs.

	DISTORTED		UNDISTORTED	
	Baseline	Ours	Baseline	Ours
e_n	28.33	27.61	28.27	27.31
e_H	18.76	16.30	18.49	15.96

Table 2. Quantitative performance of HS [16], our methods, and their fusion in terms of e_H .

	CONTROLLED			WILD		
	HS [16]	Ours	Fusion	HS [16]	Ours	Fusion
e_H	14.38	16.48	13.91	12.48	16.11	12.43

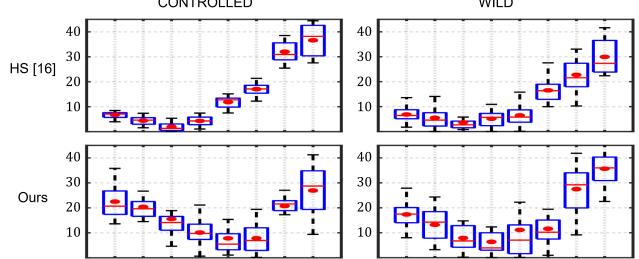


Figure 6. Results of HS [16] and our method. Each subplot shows the results from HS [16] or our method on CONTROLLED or WILD subset; the X -axis is the horizontal FoV of inputs in ascending order; the Y -axis is the absolute angular error e_H in degrees; statistics of e_H for all testing with a same FoV are displayed using the box-and-whisker plot. The red dot indicates the mean value, the red line is the median, the top and bottom bounds of the blue box indicate the first and third quartile values, and the top and bottom ends of the black lines indicate the minimum and maximum e_H .



Figure 7. Images from our CONTROLLED and WILD subsets. Table on the right shows the predicted horizontal FoVs and its ground truth (GT) in degree.

sides, due to the fundamental difference of observations for calibration cues, our method can be combined with general calibration methods (*e.g.*, [16, 32]), to improve the performance of camera calibration. We investigate these arguments by taking the performance of a state-of-the-art calibration method HS [16]⁴ (for a single image, leverage image contents for calibration) as a reference and conduct experiments on data from different scenarios, *i.e.*, CONTROLLED and WILD subsets. The inputs of HS [16] are transmission images recovered by our method.

As can be observed from Table 2 and Figure 6, the overall performance of our method is not as good as that of HS [16] due to challenging calibration cue used. HS [16] shows a larger performance divergence against that of ours

⁴<http://rachmaninoff.gel.ulaval.ca:8003/>

Table 3. Comparisons of quantitative results in terms of SSIM and PSNR on SIR² [47], ZN18-DATA [64], PB19-DATA [37], and our CONTROLLED datasets for reflection removal. Green to red means small to large errors.

Methods		ZN18 [64]	YM19 [62]	WS19 [49]	WY19 [52]	WT19 [53]	Ours
SIR ² [47]	SSIM	0.8453	0.8547	0.9003	0.8885	0.8550	0.8895
	PSNR	19.24	21.42	24.19	23.35	23.71	23.74
ZN18-DATA [64]	SSIM	0.8210	0.7220	0.7654	0.7193	0.8180	0.7510
	PSNR	21.30	18.22	19.07	18.99	21.28	19.08
PB19-DATA [37]	SSIM	0.7970	0.8308	0.8580	0.8713	0.8359	0.8736
	PSNR	15.57	17.06	19.82	20.23	17.02	16.88
CONTROLLED	SSIM	0.8115	0.7877	0.8283	0.7950	0.8004	0.8445
	PSNR	21.93	21.50	20.69	21.27	22.48	23.93

over different scenarios and achieves much better results for WILD scenario against CONTROLLED scenario. This is because images in our WILD subset contain more contents for calibration against those in CONTROLLED subset, as shown in Figure 7. Our superior performance for CONTROLLED data against WILD data indicates the difficulty brought by a variety of glass in WILD data.

As can also be observed from Figure 6, HS [16] and our method achieve their own best performance with different FoVs of inputs, which reflects that they fundamentally use different calibration cues. We conduct the experiment to fuse outputs from ours and those from HS [16], with a simple weighted summation strategy (weights are 0.2 and 0.8, respectively). Table 2 displays that the performance of such a simple fusion. Figure 7 shows the fusion results of two examples. The performance improvement for both subsets demonstrates the fundamental difference between our method and HS [16] that leverages image contents for camera calibration.

4.2. Evaluation for Reflection Removal

Considering our calibration network implicitly constrains the output (*i.e.*, Ω) of the separation network, we investigate this impact on the performance of single image reflection removal. Specifically, we compare the performance with five state-of-the-art single image reflection removal methods, including an optimization-based method (*i.e.*, YM19 [62]) and four learning-based methods (*i.e.*, ZN18 [64], WS19 [49], WY19 [52]). We conduct experiments on three publicly available datasets as well as our CONTROLLED data with commonly used metrics (*i.e.*, SSIM [51] and PSNR [19]) for this topic. The numbers of testing images from these datasets are, SIR² [47] (454), ZN18-DATA [64] (109), PB19-DATA [37] (150), and CONTROLLED (160), respectively. For all learning-based methods, we retrain their models using our training data. To perform a fair comparison with these methods, we report the results with better performance by comparing using the pre-trained models provided by authors and our re-trained models for each dataset.

As can be observed from Table 3, our method achieves state-of-the-art performance. Considering we borrow the reflection separation network from WT19 [53], our supe-



Figure 8. From top to bottom are samples from SIR² [47], ZN18-DATA [64], PB19-DATA [37], and CONTROLLED. From left to right: inputs of reflection-contaminated images, ground truth of transmission, results from ZN18 [64], YM19 [62], WS19 [49], WY19 [52], WT19 [53], and ours. Zoom in for better details.

rior performance over WT19 [53] validates the effectiveness of considering camera parameters for the problem of single image reflection removal. The visual quality comparison is shown in Figure 8. As can be observed, ZN18 [64] and WS19 [49] produce darker results, YM19 [62] and WY19 [52] output over-smooth ones, and results from WT19 [53] contain unexpected artifacts. Our method alleviates these problems and hence produces better results as compared with other methods.

4.3. Byproduct for Panorama Estimation

In this section, we show how our byproduct output contributes to the problem of panorama estimation. Inferencing of a panoramic photograph from a single view is quite challenging due to its limited and unknown FoV. Recent progress leverages deep learning-based methods and shows the successful estimate of the panorama for indoor and outdoor scenes, however, they use parametric representations that generate environment maps lacking sufficient high-frequency information (*i.e.*, [17, 10, 28]), due to the ill-posedness of this problem. Taking a single reflection-contaminated image as an input, our method produces more knowledge of a scene than a single image by taking plate glass as a lighting probe, *i.e.*, two views of a panoramic photograph with geometric relation. We thereby study to alleviate the ill-posed problem of panorama estimation from a single reflection-contaminated image. Similar to previous works [11, 41] for this topic, we consider the indoor scene and regard the problem of panorama estimation as to complete a panoramic photograph.

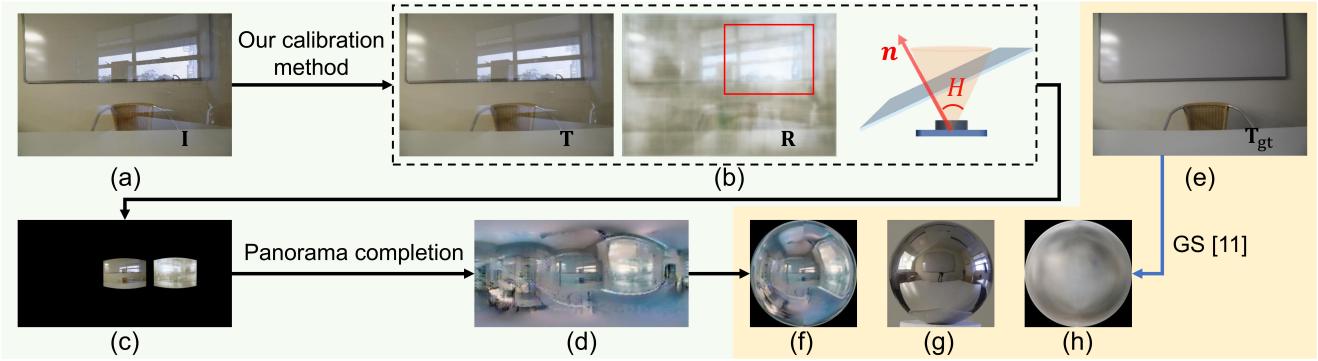


Figure 9. Left: The framework to estimate a panorama from (a) a single reflection-contaminated image I . (b) Our calibration method estimates transmission T and reflection R as well as their camera parameters n, H . (c) An incomplete panorama is then obtained by embedding T and R based on the geometric relation determined by n, H , i.e., fix T at the middle, the spatial position of R in this incomplete panorama is determined by n , the size of embedding views is determined by H . The panorama completion network takes the incomplete panorama as input and outputs (d) a panorama. Right: Visual quality comparison of light probes from (f) our method, (g) ground truth, and (h) GS [11]. GS [11] takes the (e) transmission image T_{gt} as the input.

The framework of estimating panorama from a reflection-contaminated image is outlined in Figure 9 (left). To train the panorama completion network, we borrow an existing image inpainting network [66] and train it with 10,800 indoor panoramas from Matterport3D dataset [4] as outputs. We generated inputs of training data from these panoramas by locating two regions computed from a random n, H and masking out remaining regions.

Figure 9 (right) shows our estimated result as well as the visual quality comparison with GS [11]⁵. Note that GS [11] takes reflection-free T_{gt} as the input. For easier comparison, we provide the panorama captured by a lighting probe for reference and transform panoramas as light probes through HDRshop⁶. As can be observed, our result contains more details as compared to that from GS [11]. Moreover, we also provide a more plausible prediction of the dominant lighting, thanks to the additional information from the separated reflection (red rectangle in Figure 9 (b)).

5. Conclusion

This paper answers the question ‘What plate glass reveals about calibration cues’, by proposing a new calibration method that estimates camera parameters taking the reflective amplitude coefficient map as the calibration cue. Different from previous works that leverage image contents from a noise-free image, our method deals with the calibration problem in the context of plate glass reflection and takes a reflection-contaminated image as the input. It facilitates single image reflection removal by involving the constraint of geometric relation to the reflective amplitude

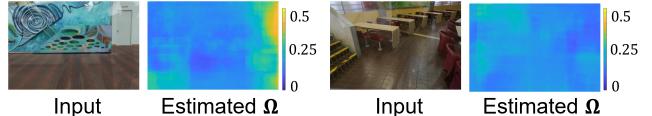


Figure 10. Two failure examples from our validation data with a small and a large FoVs. Left: ground truth $H_{gt} = 39.6^\circ$, estimated $H_{est} = 123.4^\circ$. Right: ground truth $H_{gt} = 73.4^\circ$, estimated $H_{est} = 23.0^\circ$.

coefficient map. Based on outputs from our method, our byproduct shows a solution to the problem of single image panorama estimation by considering reflection as an additional view. Our promising performance on problems of camera calibration and reflection removal benefits to solving high-level vision tasks such as food recognition behind showcases or face recognition behind vehicle windows.

Limitations. Although our method is free from image contents, it is sensitive to incorrectly estimated calibration cue Ω . Figure 10 shows two failure examples. We leave the prediction of a robust representation of calibration cue Ω from a single contaminated image as our future work.

Acknowledgment

This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore, under the NRF-NSFC grant NRF2016NRF-NSFC001-098 and NTU-PKU Joint Research Institute with the donation from Ng Teng Fong Charitable Foundation. It was done at the Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore. It is supported in part by the National Natural Science Foundation of China under Grants U1611461 and 61872012, National Key R&D Program of China (2019YFF0302902), Shenzhen Municipal Science and Technology Program under Grant JCYJ20170818141146428, and Beijing Academy of Artificial Intelligence (BAAI).

⁵<http://rachmaninoff.gel.ulaval.ca:8001/>

⁶Note that the images of light probes are left-right flipped for easier comparison. HDRshop: <http://ict.debevec.org/~debevec/HDRShop/main-pages/tutorials.html>

References

- [1] Francisco Abad, Emilio Camahort, and Roberto Vivó. Camera calibration using two concentric circles. In *International Conference Image Analysis and Recognition*, pages 688–696, 2004. 3
- [2] Nikolaos Arvanitopoulos, Radhakrishna Achanta, and Sabine Süsstrunk. Single image reflection suppression. In *Proc. of Computer Vision and Pattern Recognition*, pages 1752–1760, 2017. 2
- [3] Max Born and Emil Wolf. Principles of optics: electromagnetic theory of propagation, interference and diffraction of light. *Elsevier*, 2013. 3
- [4] Angel Chang, Angela Dai, Thomas Allen Funkhouser, Maciej Halber, Matthias Niebner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3D: Learning from RGB-D data in indoor environments. In *Proc. of International Conference on 3D Vision*, pages 667–676. Institute of Electrical and Electronics Engineers Inc., 2018. 8
- [5] Qian Chen, Haiyuan Wu, and Toshikazu Wada. Camera calibration with two arbitrary coplanar circles. In *Proc. of European Conference on Computer Vision*, pages 521–532, 2004. 2
- [6] Antonio Criminisi, Ian Reid, and Andrew Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000. 1
- [7] Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David P Wipf. A generic deep architecture for single image reflection removal and image smoothing. In *Proc. of Internatoinal Conference on Computer Vision*, pages 3258–3267, 2017. 1, 2
- [8] David F Fouhey, Abhinav Gupta, and Martial Hebert. Data-driven 3D primitives for single image understanding. In *Proc. of Internatoinal Conference on Computer Vision*, pages 3392–3399, 2013. 1
- [9] Vincent Fremont and Ryad Chellali. Direct camera calibration using two concentric circles from a single view. In *International Conference on Artificial Reality and Telexistence*, pages 93–98, 2002. 3
- [10] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-François Lalonde. Deep parametric indoor lighting estimation. In *Proc. of Computer Vision and Pattern Recognition*, pages 7175–7183, 2019. 7
- [11] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. Learning to predict indoor illumination from a single image. *ACM Transactions on Graphics*, 36(6):176, 2017. 7, 8
- [12] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. *Cambridge university press*, 2003. 4
- [13] Varsha Hedau, Derek Hoiem, and David Forsyth. Recovering the spatial layout of cluttered rooms. In *Proc. of Internatoinal Conference on Computer Vision*, pages 1849–1856, 2009. 1
- [14] Janne Heikkila, Olli Silven, et al. A four-step camera calibration procedure with implicit image correction. In *Proc. of Computer Vision and Pattern Recognition*, volume 97, page 1106, 1997. 2
- [15] Derek Hoiem, Alexei A Efros, and Martial Hebert. Putting objects in perspective. *International Journal of Computer Vision*, 80(1):3–15, 2008. 1
- [16] Yannick Hold-Geoffroy, Kalyan Sunkavalli, Jonathan Eisenmann, Matthew Fisher, Emiliano Gambaretto, Sunil Hadap, and Jean-François Lalonde. A perceptual measure for deep single image camera calibration. In *Proc. of Computer Vision and Pattern Recognition*, pages 2354–2363, 2018. 1, 2, 6, 7
- [17] Yannick Hold-Geoffroy, Kalyan Sunkavalli, Sunil Hadap, Emiliano Gambaretto, and Jean-François Lalonde. Deep outdoor illumination estimation. In *Proc. of Computer Vision and Pattern Recognition*, pages 7312–7321, 2017. 7
- [18] Haifei Huang, Hui Zhang, and Yiu-ming Cheung. The common self-polar triangle of concentric circles and its application to camera calibration. In *Proc. of Computer Vision and Pattern Recognition*, pages 4065–4072, 2015. 3
- [19] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 7
- [20] Hamid Izadinia, Qi Shan, and Steven M Seitz. Im2cad. In *Proc. of Computer Vision and Pattern Recognition*, pages 5134–5143, 2017. 1
- [21] Guang Jiang and Long Quan. Detection of concentric circles for camera calibration. In *Proc. of Internatioinal Conference on Computer Vision*, volume 1, pages 333–340, 2005. 2, 3
- [22] Jun-Sik Kim, Pierre Gurdjos, and In-So Kweon. Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):637–642, 2005. 3, 4
- [23] Jun-Sik Kim, Ho-Won Kim, and In So Kweon. A camera calibration method using concentric circles for vision applications. In *Proc. of Asian Conference on Computer Vision*, 2002. 3
- [24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [25] Naejin Kong, Yu-Wing Tai, and Joseph S Shin. A physically-based approach to reflection separation: from physical modeling to constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2):209–221, 2013. 1, 2, 3, 4
- [26] Jean-François Lalonde, Srinivasa G Narasimhan, and Alexei A Efros. What do the sun and the sky tell us about the camera? *International Journal of Computer Vision*, 88(1):24–51, 2010. 1, 2
- [27] Hyunjoon Lee, Eli Shechtman, Jue Wang, and Seungyong Lee. Automatic upright adjustment of photographs with robust camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5):833–844, 2013. 2
- [28] Chloe LeGendre, Wan-Chun Ma, Graham Fyffe, John Flynn, Laurent Charbonnel, Jay Busch, and Paul Debevec. Deep-light: Learning illumination for unconstrained mobile mixed reality. In *Proc. of Computer Vision and Pattern Recognition*, pages 5918–5928, 2019. 7

- [29] Anat Levin and Yair Weiss. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1647–1654, 2007. 2
- [30] Yu Li and Michael S Brown. Exploiting reflection change for automatic reflection removal. In *Proc. of Internatoinal Conference on Computer Vision*, pages 2432–2439, 2013. 2
- [31] Yu Li and Michael S Brown. Single image layer separation using relative smoothness. In *Proc. of Computer Vision and Pattern Recognition*, pages 2752–2759, 2014. 2
- [32] Manuel Lopez, Roger Mari, Pau Gargallo, Yubin Kuang, Javier Gonzalez-Jimenez, and Gloria Haro. Deep single image camera calibration with radial distortion. In *Proc. of Computer Vision and Pattern Recognition*, pages 11817–11825, 2019. 1, 2, 3, 5, 6
- [33] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Reflection separation using a pair of unpolarized and polarized images. In *Advances in Neural Information Processing Systems*, pages 14532–14542, 2019. 1
- [34] Rui Melo, Michel Antunes, João Pedro Barreto, Gabriel Falcão, and Nuno Gonçalves. Unsupervised intrinsic calibration from a single frame using a. In *Proc. of Internatoinal Conference on Computer Vision*, pages 537–544, 2013. 2
- [35] Jorge Nocedal and Stephen Wright. Numerical optimization. *Springer Science & Business Media*, 2006. 4
- [36] James Pritts, Ondrej Chum, and Jiri Matas. Detection, rectification and segmentation of coplanar repeated patterns. In *Proc. of Computer Vision and Pattern Recognition*, pages 2973–2980, 2014. 1, 2
- [37] Abhijith Punnappurath and Michael S Brown. Reflection removal using a dual-pixel sensor. In *Proc. of Computer Vision and Pattern Recognition*, pages 1556–1565, 2019. 7
- [38] Jiangpeng Rong, Shiyao Huang, Zeyu Shang, and Xianghua Ying. Radial lens distortion correction using convolutional neural networks trained with synthesized images. In *Proc. of Asian Conference on Computer Vision*, pages 35–49, 2016. 2
- [39] Carsten Rother. A new approach to vanishing point detection in architectural environments. *Image and Vision Computing*, 20(9-10):647–655, 2002. 2
- [40] YiChang Shih, Dilip Krishnan, Fredo Durand, and William T Freeman. Reflection removal using ghosting cues. In *Proc. of Computer Vision and Pattern Recognition*, pages 3193–3201, 2015. 2
- [41] Shuran Song and Thomas Funkhouser. Neural illumination: Lighting prediction for indoor environments. In *Proc. of Computer Vision and Pattern Recognition*, pages 6918–6926, 2019. 7
- [42] Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proc. of Computer Vision and Pattern Recognition*, pages 567–576, 2015. 1
- [43] William P Spence and Eva Kultermann. Construction materials, methods and techniques. *Cengage Learning*, 2016. 3
- [44] Peter F Sturm and Stephen J Maybank. On plane-based camera calibration: A general algorithm, singularities, applica-
- tions. In *Proc. of Computer Vision and Pattern Recognition*, volume 1, pages 432–437, 1999. 2
- [45] Gabriel Taubin. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (11):1115–1138, 1991. 4
- [46] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, Wen Gao, and Alex C Kot. Region-aware reflection removal with unified content and gradient priors. *IEEE Transactions on Image Processing*, 27(6):2927–2941, 2018. 2
- [47] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Benchmarking single-image reflection removal algorithms. In *Proc. of Internatoinal Conference on Computer Vision*, 2017. 5, 7
- [48] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Crrn: Multi-scale guided concurrent reflection removal network. In *Proc. of Computer Vision and Pattern Recognition*, pages 4777–4785, 2018. 2
- [49] Renjie Wan, Boxin Shi, Haoliang Li, Ling-Yu Duan, Ah-Hwee Tan, and Alex Kot Chichung. Corrn: Cooperative reflection removal network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 1, 2, 5, 7
- [50] Shenlong Wang, Sanja Fidler, and Raquel Urtasun. Holistic 3d scene understanding from a single geo-tagged image. In *Proc. of Computer Vision and Pattern Recognition*, pages 3964–3972, 2015. 1
- [51] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7
- [52] Kaixuan Wei, Jiaolong Yang, Ying Fu, David Wipf, and Hua Huang. Single image reflection removal exploiting misaligned training data and network enhancements. In *Proc. of Computer Vision and Pattern Recognition*, pages 8178–8187, 2019. 2, 7
- [53] Qiang Wen, Yinjie Tan, Jing Qin, Wenxi Liu, Guoqiang Han, and Shengfeng He. Single image reflection removal beyond linearity. In *Proc. of Computer Vision and Pattern Recognition*, pages 3771–3779, 2019. 1, 2, 3, 4, 5, 7
- [54] Patrick Wieschollek, Orazio Gallo, Jinwei Gu, and Jan Kautz. Separating reflection and transmission images in the wild. In *Proc. of European Conference on Computer Vision*, pages 89–104, 2018. 1
- [55] Marta Wilczkowiak, Edmond Boyer, and Peter Sturm. Camera calibration and 3D reconstruction from single images using parallelepipeds. In *Proc. of Internatoinal Conference on Computer Vision*, volume 1, pages 142–148, 2001. 1
- [56] Scott Workman, Connor Greenwell, Menghua Zhai, Ryan Baltenberger, and Nathan Jacobs. Deepfocal: A method for direct focal length estimation. In *Proc. of Internatoinal Conference on Image Processing*, pages 1369–1373, 2015. 1, 2, 5
- [57] Scott Workman, Radu Paul Mihail, and Nathan Jacobs. A pot of gold: Rainbows as a calibration cue. In *Proc. of European Conference on Computer Vision*, pages 820–835, 2014. 2
- [58] Scott Workman, Menghua Zhai, and Nathan Jacobs. Horizon lines in the wild. *arXiv preprint arXiv:1604.02129*, 2016. 2

- [59] Yihong Wu, Haijiang Zhu, Zhanyi Hu, and Fuchao Wu. Camera calibration from the quasi-affine invariance of two parallel circles. In *Proc. of European Conference on Computer Vision*, pages 190–202, 2004. [2](#)
- [60] Han Yan. Focalens dataset. figshare. dataset. <https://figshare.com/articles/FocalLens/3399169/2>, 2016. [5](#)
- [61] Jie Yang, Dong Gong, Lingqiao Liu, and Qinfeng Shi. Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In *Proc. of European Conference on Computer Vision*, pages 654–669, 2018. [2](#)
- [62] Yang Yang, Wenye Ma, Yin Zheng, Jian-Feng Cai, and Weiyu Xu. Fast single image reflection suppression via convex optimization. In *Proc. of Computer Vision and Pattern Recognition*, pages 8141–8149, 2019. [2, 7](#)
- [63] Menghua Zhai, Scott Workman, and Nathan Jacobs. Detecting vanishing points using global image context in a non-manhattan world. In *Proc. of Computer Vision and Pattern Recognition*, pages 5657–5665, 2016. [2](#)
- [64] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *Proc. of Computer Vision and Pattern Recognition*, pages 4786–4794, 2018. [1, 2, 3, 4, 7](#)
- [65] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 2000. [1, 2](#)
- [66] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. Pluralistic image completion. In *Proc. of Computer Vision and Pattern Recognition*, pages 1438–1447, 2019. [8](#)