Single image reflection removal with camera calibration

Chen, Jinnan

2021

Chen, J. (2021). Single image reflection removal with camera calibration. Master's thesis, Nanyang Technological University, Singapore. <https://hdl.handle.net/10356/150528>

<https://hdl.handle.net/10356/150528>

<https://doi.org/10.32657/10356/150528>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 18:19:29 SGT

SINGLE IMAGE REFLECTION REMOVAL WITH CAMERA CALIBRATION

CHEN JINNAN

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Master of Engineering

2021 January

STATEMENT OF ORIGINALITY

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

6/1/2021

.....

Date

CHEN JINNAN

.....

CHEN JINNAN

SUPERVISOR DECLARATION STATEMENT

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

11 Jan 2021

.....

Date

.....

JIANG XUDONG

AUTHORSHIP ATTRIBUTION STATEMENT

This thesis contains material from one paper published in the following peer-reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.

Part of the content is published as Q. Zheng, J. Chen, Z. Lu, B. Shi, X. Jiang, K. Yap, L. Duan, A. Kot, “What does Plate Glass Reveal about Camera Calibration?,” IEEE Conference on Computer Vision and Pattern Recognition, (CVPR’20), Jun 2020

The contributions of the co-authors are as follows:

Dr Qian Zheng initiated guided the whole project, was responsible for the math equation demonstration and the main part of the manuscripts, participating in the experiments design,

I proposed the initial idea to embed the physical information in both of the data synthesis and model training process. I was responsible for designing and building the network for reflection removal and camera calibration, participating in all of the experiments and data collection and part of the manuscripts.

Zhan Lu participated in programming panorama estimation experiments and data collection.

Prof Boxin Shi, Xudong Jiang, Kim-Hui Yap, Ling-Yu Duan and Alex C. Kot participated in its design and coordination and helped to draft the manuscripts.

All the authors read and approved the final manuscripts.

6/1/2021

.....

Date

CHEN JINNAN

.....

CHEN JINNAN

ACKNOWLEDGEMENT

I would like to acknowledge everyone who provided me supports during my MEng research. First of all, I appreciate my supervisor, Associate Professor Xudong Jiang, who has provided a lot of guidance during the past three semesters in both my courses and research. He offered me an opportunity to work with his post-doctoral Dr. Qian Zheng to complete two research work on the topic of single image reflection removal and good publications. Also, he gave me great suggestions about my thesis. I also appreciate Dr. Qian Zheng for his guidance in my research. What's more, I would like to appreciate Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore, to provide me a chance to do research here. Finally, I'm so grateful that my family provided me financial support and encouraged me during these years.

SUMMARY

With the significant development of deep learning in computer vision, it is possible to solve some ill-posed problems with only limited known inputs with specific prior knowledge from training data which could alleviate the ill-posedness.

However, in some cases like reflection removal, it is impossible to collect an enormous amount of paired real data for training for the difficulty to keep the whole scene the same when removing the glass plane as well as the time-consuming process. Therefore, how to simulate the reflection formation process and embed it to the data synthesis process deserves great focus and investigation. In such a way, the trained model could approach the real data distribution, which means better generalization capacity. Inspired by this idea and our observation in daily life, a method that jointly solves the camera calibration and reflection removal problem has been proposed for the first time in the literature, which jointly optimizes two neural networks for camera calibration and reflection removal.

To be more specific, we build the link between these two tasks by estimating the reflection amplitude coefficient map, which has direct physical symbolic to these two tasks. Before training, we simulate the reflection formation in the data generation process by setting camera parameters and glass parameters within certain distributions and employ physical rules and constraints to make it close to the real reflection-contaminated images captured in the camera. During the training time, with the ground truth of such physical parameters, we could design some loss to make our model capable of estimating such physical parameters from the unseen input data. These physical constraints could not only teach the model to find the location of the strong reflection, which deserves more focus, but also make it possible to do the camera calibration from these cues. Besides, with the estimation of the above parameters of the camera as well as the separated background and reflected image, a byproduct that could do a panorama estimation has been proposed and shows promising visual results.

In order to validate our idea, we also collect the first reflection removal dataset, which contains camera extrinsic in the controlled and wild scene. Substantial experimental results and analysis show the effectiveness and superiority of the proposed methods in all three subtasks and in the different real scenarios compared with other state-of-

the-art methods available. This research could have a broad impact on further research in the computer vision community.

In the end, we analyze the experimental results in different scenarios and find the limitation of our method, and then further suggest some possible future works to be investigated.

TABLE OF CONTENTS

STATEMENT OF ORIGINALITY	2
SUPERVISOR DECLARATION STATEMENT	3
AUTHORSHIP ATTRIBUTION STATEMENT	4
ACKNOWLEDGEMENT	5
SUMMARY	6
TABLE OF CONTENTS.....	8
ACRONYMS.....	10
LIST OF FIGURES.....	11
LIST OF TABLES	11
CHAPTER 1.....	12
INTRODUCTION	12
1.1. Background	12
1.2. Motivation.....	14
1.3. Objectives and Major contribution of the Thesis.....	15
1.3.1. Objectives	15
1.3.2. Major contribution	15
1.4. Organization of the Thesis.....	16
CHAPTER 2.....	18
LITERATURE REVIEW	18
2.1. Reflection Removal.....	18
2.2. Single Image Reflection Removal	19
2.2. Single Image Camera Calibration	20
2.3. Panorama estimation.....	21
2.4. Conclusion	22
CHAPTER 3.....	23
CALIBRATED REFLECTION REMOVAL	23
3.1. Preliminary	23
3.2. Optimization-based Method	24
3.3. Learning-based Method	27
3.4. Data Preparation.....	28
3.5. Panorama Estimation.....	38
CHAPTER 4.....	40
EXPERIMENTAL RESULTS.....	40
4.1. Evaluation for Reflection Removal	40
4.2. Evaluation for Camera Calibration	44
4.4. Byproduct of panorama estimation	48
4.5. Ablation Study of Pseudo-invariance.....	49
CHAPTER 5.....	51
CONCLUSION AND FUTURE WORK	51
5.1. Conclusion	51
5.2. Broad impact	52
5.3. Limitation	52

5.4. *Future work* 53

AUTHOR'S PUBLICATION..... 55

REFERENCES..... 56

ACRONYMS

FoV	Field of view
T	Transmission image
R	Reflection image
I	Input image
AoI	Angle of incidence
Ω	Reflective amplitude coefficient map
Θ	Angle of incidence map
U	Uniform distribution
N	Gaussian distribution
S	Separation network
C	Calibration network
κ	Glass reflective index
$\mathbf{n}$	Glass normal vector
H	Horizontal field of view

LIST OF FIGURES

Figure 1. Illustration of our observation	15
Figure 2. Illustrations of geometric relation	24
Figure 3. Curves of $F(\theta, \kappa)$ and $\partial F(\theta, \kappa) / \partial \theta$,	26
Figure 4. Exmaple of Detecting reliable points to fit conic.	26
Figure 5. The framework of our learning-based method.	28
Figure 6. Map of angle of incidence, Reflection Map, Refraction Map	30
Figure 7. Simulation of AoI in different FoV and glass orientations	30
Figure 8. Two paired Synthesized training data	32
Figure 9. Data collection settings	35
Figure 10. Data collection for controlled and wild scene	36
Figure 11. An illustration of images used for distortion estimation	36
Figure 12. Illustrations of distortion correction	37
Figure 13. An overview of our validation data containing 320 images	37
Figure 14. An architecture in [66] we used to train our panorama estimation network	39
Figure 15. Visual quality comparison	43
Figure 16. Error distribution of the validation Controlled dataset	46
Figure 17. Ten examples from our validation Wild dataset	47
Figure 18. Results of HS [16] and our method	47
Figure 19. Results from the CONTROLLED and WILD subset	48
Figure 20. Panorama completion pipeline and visual results	49
Figure 21. Two failure examples from our validation data	53

LIST OF TABLES

Table 1 Numbers of images regarding different FoVs	37
Table 2. Quantitative results	41
Table 3. Quantitative performance of our methods	45
Table 4. Overall results of HS [16], our methods, and fusion	46
Table 5. Ablation study of the property of pseudo-invariance	50

Chapter 1

Introduction

1.1. Background

Reflection removal from single image is an ill-posed low-level vision problem but it is of great importance in computer vision community. On the one hand, reflection always happen when light meets screens and glass everywhere in our daily life, so recovering the reflection-free image could improve the vision quality in the artificial intelligence systems. On the other hand, it is always hard to get multiple images with same content in the real scenario for application. Before the deep learning era, researchers tried to solve this problem as an image denoising problem because the reflection part has some common specific features which could be distinguished from the background part. However, most of such methods end up with destructing the desired background part as well, which is unwanted. Also, for the reflection-contaminated photos captured under different lighting conditions, color distortion could happen after processed by such algorithms, so the applicable scenario is quite limited.

Deep neural network gives a potential solution by encoding some prior knowledge into the model which could adapt to more complicated scenarios with various lighting conditions and reflection amplitude. Down-sampling and up-sampling blocks are proposed for such ‘image to image’ tasks which shows astonishing performance like image segmentation and noise deduction. Lots of previous works [7,48,49,64] on single image reflection removal also used these basic blocks as the foundation for their own idea implementation so the baseline performance could be guaranteed.

Note that large amount of paired training data of noise-free transmission image and reflection-contaminated image are typically expensive to get, so we focus on how to better synthesize reflection-contaminated image to approach the distribution of those in the real scenario as much as possible. The classical physically based model of a reflection-contaminated image can be formulated as [25, 53],

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

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. $\mathbf{\Omega}$ is the reflective amplitude coefficient map. According to the Fresnel rule equations, the amplitude coefficient of each pixel in transmission and reflection images could be accurately defined, based on which Kong et al. [25] derive a formation of $\mathbf{\Omega}$ for glass with double surfaces. One of the properties in their formation is that $\mathbf{\Omega}$ 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]). 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 $\mathbf{\Omega}$ pixel-wisely. However, they failed to consider the geometric constraints during the data synthesis process.

On the other hand, camera calibration is also a classic computer vision problem which draw lots of focus since it is the foundation of abroad category of vision tasks like 3D reconstruction, scene understanding, augmented reality. Classic camera calibration requires multiple input images with same specific patterns which is accurate but time-consuming, so it is not applicable in some cases. Calibrating a complete camera model using single image as input is quite challenging but non-trivial. There are several related works about single image camera calibration from noise-free images with the advancement of deep learning.

1.2. Motivation

We observe that Ω contains useful cues for camera calibration which motivates us to estimate camera parameters from a single reflection-contaminated image. We did an experiment to explain the details by putting a piece of plate glass between the left and right scenes in Figure 1. Row 1: An object with simple background as transmission (left) and the scene with approximately uniform illumination as reflection (right). Row 2: Geometric camera models with parameters $\mathbf{n}$ and H for reflection-contaminated images capture. Here, we use $\mathbf{n}$ and H represent the orientation of the glass plane and horizontal field of view (FoV). Row 3: Reflection-contaminated images 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). Row 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.

Figure 1 Row 3 intuitively shows our observations about how these two parameters affect the reflection part. As can be observed, the orientation of glass $\mathbf{n}$ affects the direction of reflective intensity change (left and middle figures) and 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 to estimate camera parameters.

Here we use a simplified version formula to synthesize training data which only considers the first reflection when lights transmit the glass to utilize our two observations above. Note this is the first work in the literature to compute the reflection amplitude coefficient map pixel-wisely for the training data simulation. This information actually guides our image separation network to be aware of which part of the image has stronger reflection. Thus, it could guarantee the good reflection removal performance of our network. What is more, we found this reflection amplitude coefficient map could be a cue for camera calibration to accurately estimate the solve for extrinsic (tilt, roll of the glass plane) and intrinsic (focal length) parameters.

We then present an optimization-based method that imposes part of though reliable entries of an estimated Ω and learning-based method to indiscriminately exploit all

entries of an estimated Ω to achieve robust calibration. 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 .

Figure 1. Illustration of our observation

1.3. Objectives and Major contribution of the Thesis

1.3.1. Objectives

The thesis aims to introduce the physical-rule-based physical models to simulate the reflection formation process and corresponding optimization-based and deep learning-based models. What is more, through experimental results in the different datasets in the real scenarios, we validate that the proposed method can address three ill-posed subtasks: reflection removal, camera calibration, panorama estimation in the reflection contaminated image input setting and give some in-depth analysis and conclusion.

1.3.2. Major contribution

The proposed method successfully relates reflection removal problem with camera calibration problem and focuses on how to use the pixel-wise reflection coefficient

map to guide reflection removal process as well as calibrate the camera parameters at the same time.

A data-driven method to recover extrinsic (glass orientation) and intrinsic (horizontal Field of view) parameters of the camera from a reflection-contaminated image is proposed. To the best of our knowledge, this is the first work to jointly address the single image camera calibration as well as single image reflection removal problem in single reflection-contaminated input setting.

Then a data-driven method as well as an optimization-based method to compute camera parameters from the reflective amplitude coefficient map are proposed. The two methods have been tested in our collected dataset and a detailed comparison has been made. Our proposed camera calibration method could also be demonstrated to be free from image contents and able to facilitate both single image reflection removal problem.

Also, as an additional application, it helps alleviate the ill-posedness of single image panorama estimation.

At last, a validation dataset containing 320 reflection-contaminated images as well as their calibration parameters is collected for further research use.

1.4. Organization of the Thesis

This report consists of five parts: chapter 1: introduction, chapter 2: literature review, chapter 3: the detailed explanation and formulation of our proposed technique, chapter 4: experimental results and corresponding analysis, chapter 5: conclusion and future work.

The first chapter gives the background of our tasks including the reflection removal, camera calibration as well as our intuition for our novel solution. It also gives a brief introduction to our whole project.

The second chapter is about the previous related work on single image reflection removal and camera calibration. We discuss the different prior knowledge utilized in different models and thus different effect and current research gap to demonstrate the significance of our work.

The following chapter includes mathematical derivation our physical data synthesis model as well as the pipeline of the proposed optimization-based and deep learning-based method.

The experimental chapter presents both comprehensive numerical and visual results of the three subtasks of our proposed model. Besides, we use our training data to finetune other methods and make some meaningful analysis for a fair comparison. Our collected data in both controlled and in the wild scenarios are also shown in that chapter.

The last chapter is to discuss the broad impact of our work as well as the limitation in some real scenarios which could guide our future work.

Chapter 2

Literature Review

2.1. Reflection Removal

Reflection Removal is a classic low-level computer vision task which has been studied for a long time. Researchers proposed to solve this hard problem mainly in two types of methods, single input and multiple input images. Multiple-image-based reflection removal methods commonly leverage constraints from content-free variables to alleviate the ill-posedness of the solution, e.g., polarization angles [54, 25, 33] or reflection disparity [37]. These methods usually could get satisfying results, but they have huge application limitations. First, for most cases, it is impossible to get multiple inputs with the same glass and camera in the same scene of the shooting environment. Second, compared with the single reflection removal, multiple image reflection removal methods would take much more time during the testing period. Therefore, single reflection removal has more application potential and deserves more research focus. However, priors from content-free variables are seldom considered in the context of single image. Most single image reflection removal methods impose

priors from image contents and they can be roughly divided into optimization-based and deep learning-based methods.

2.2. Single Image Reflection Removal

2.1.1. Deep learning methods

Single image reflection removal is a typical ill-posed computer vision problem that estimates more than two unknowns from a single input. To make this problem more tractable, several priors have been proposed 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 effect due to out-of-focus and limited depth of field of the 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]. Most recently, Li et al. [74] designs a multi-scale residual reconstruction loss to encode the influence of the synthesis method across the whole network. Lei [75] utilize the polarization information and proposes a polarized reflection removal model with a two-stage architecture to solve the ill-posed problem. Li [76] further proposes to a polarization-guided raytracing model and loss function design to separate reflection images.

All of these methods show impressive improvements. These methods formulate Ω as a scalar (e.g., [61, 49]), or a spatial-varying variable (e.g., modeled by a Gaussian distribution [7, 64], or generated by a GAN [53]). However, never have scholars considered the reflection amplitude distribution which is directly correlated to the orientation of the glass plane. Our methodology is based on this observation and we consider Ω as a function of Θ which involves constraints from the geometric camera model.

2.1.2. Optimization-based methods

Assumptions have been made in the literature to make the problem of single image reflection removal tractable due to its massive ill-posedness. Optimization-based methods exploit the statistics of natural images to make explicit assumptions. For

example, Levin and Weiss [29] separate transmission and reflection images based on a gradient sparsity prior with manual annotations. Li and Brown [67] utilize a smooth gradient prior since reflection images are likely to be out-of-focus and blurry. Shih et al. [40] remove reflection effect based on the observation of ghosting cues. Wan et al. [68] perform edge classification for transmission and reflection images through multi-scale depth of fields. Arvanitopoulos et al. [2] suppress the reflection by an ℓ_0 gradient penalty on recovered transmission images. Wan et al. [46] employ both content and gradient priors to jointly restore missing contents and recover transmission images [46]. Yang et al. [62] suppress the reflection by solving a partial differential equation. Optimization-based methods work well on the specific case. However, they are not expected to generalize to various types of reflections in real-world when their assumptions are not held or that are far from the prior knowledge utilized which indicates the poor generalization ability of such methods.

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.

Our calibration cue Ω comes from the pattern of imaged concentric circles in the reflective amplitude coefficient map. 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.

2.3. Panorama estimation

Illumination or panorama estimation has been a long-standing problem in both computer vision and graphics. Reconstructing a scene and all its properties including geometry, surface reflectance, and illumination from a single image is one of the long-standing goals of computer vision and graphics, and has been extensively studied in the literature. Because these properties are intrinsically tied to each other, estimating each of them often relies on reconstructing the others too [11].

In this section, we briefly review work most relevant to this thesis.

More recently, deep neural network-based techniques have been proposed for estimating reflectance maps the convolution of a surface bidirectional reflectance distribution function (BRDF) with the incident illumination from a single image [71]. These reflectance maps can then be separated into reflectance and illumination estimates [71]. However, these techniques focus on objects of a known class with approximately known shape and spatially constant reflectance. In contrast, the indoor scenes we focus on have more diverse and complex shapes and reflectance variations. CNN-based methods have also been proposed for estimating scene depth [72], the surface normal [73; 72], and intrinsic decompositions for indoor scenes. While challenging, these problems involve recovering properties that are directly observed in the image, unlike lighting which can lie out of the field-of-view and only indirectly effects the scene appearance [11].

For illumination estimation, Gardner et al. [11] propose to formulate the illumination estimation function as an end-to-end neural network which automatically predicts the high-dynamic-range (HDR) lighting conditions from a single image. It relies on a database of low-dynamic-range (LDR) panoramas. To compensate for the low dynamic range, light sources are detected and the panoramas are warped to generate target light masks, which can be used to train the CNN to predict light directions with the help of crops extracted from the panoramas. Then, the network is fine-tuned on a dataset of HDR panoramas, which allows it to learn to predict light intensities. However, since the input and output representation of their network architecture does not share any immediate notion of pixel-to-pixel spatial correspondence, their model tends to generate illumination maps that reflect general color in the training dataset, which is essentially a severe overfitting problem as opposed to important high-

frequency lighting details in reality. In contrast, [70] Song predicts the 3D geometric structure of the scene and then uses it to warp the observed image into an intermediate representation to align the output illumination map to fully utilize input information and preserve high-frequency lighting details. However, awkward to directly use these methods in our reflection contaminated setting, we instead simply take the separated images as input and formulate it as an image inpainting problem.

2.4. Conclusion

In the single input image with reflection setting, we explore to estimate the implicit information of camera intrinsic and extrinsic through the cues of the reflection distribution in the input images to jointly solve the two highly ill-posed problems, i.e., single image reflection removal, and single image calibration in a novel way. Besides, we also fully utilize the information we get from solving the above two problems to do a panorama estimation as our byproduct.

Compared with current deep learning-based reflection removal methods, the superiority of our method lies in encoding the physical rule to constrain our network during the training process.

Different from current image calibration methods exploiting image contents in a noise-free setting, our approach uses the calibration cue of the reflective amplitude coefficient map from a reflection-contaminated image.

Compared with image inpainting-based single panorama estimation methods, our proposed method could inherently benefit from the extra separated reflection image for the panorama estimation process.

Chapter 3

Calibrated reflection removal

3.1. Preliminary

We study camera parameters of the extrinsic parameter rotation matrix w.r.t. the glass plane and the camera intrinsic parameter horizontal field of view, from a single image with the size of $h \times w$ (height $\times$ width). We represent these two parameters as the orientation of glass plane: $\mathbf{n}$, and horizontal FoV: H , respectively. Figure 2 illustrates of geometric relation between Θ and $\mathbf{n}$, H , which can be expressed as $\Theta = P(\mathbf{n}, H)$ 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 $\mathbf{n}$ intersects the glass plane and image plane at c and pc . To be more specific, considering a point $p = (u, v)^T$ on the image plane, its coordinate in 3D space is $(u, v, -f)^T$, where $f = \frac{\omega}{2 \tan \frac{H}{2}}$. The AoI $\theta \in \Theta$ of p can be computed as $\theta = \arccos \frac{\langle p, \mathbf{n} \rangle}{\|p\|_2}$. Inspired by the analytic formation of $\mathbf{\Omega}$ in [25], we represent $\mathbf{\Omega} = F(\Theta, \kappa)$, where κ is the refractive index. Therefore, the calibration problem in this

paper is to inversely recover $\mathbf{n}$ and H from an input $\mathbf{I}$ with a formation model such that,

$$\mathbf{I} = (\mathbf{1} - \alpha(\mathbf{n}, H, \kappa)) \circ \mathbf{T} + \alpha(\mathbf{n}, H, \kappa) \circ \mathbf{R},$$

$$\text{where } \mathbf{\Omega} = \alpha(\mathbf{n}, H, \kappa) = F(P(\mathbf{n}, H), \kappa). \quad (2)$$

Our basic idea to solve this problem is to incorporate a reflection separation method (taking $\mathbf{I}$ as input) into a calibration algorithm (taking an estimated $\mathbf{\Omega}$ 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 $\mathbf{\Omega}$ based on Equation (1). We then focus on how to robustly solve the inverse problem with $\mathbf{\Omega} = \alpha(\mathbf{n}, H, \kappa)$, i.e., recovering $\mathbf{n}$, H from an estimated $\mathbf{\Omega}$. We observe that this inverse problem forms an over-determined system because all entries of $\mathbf{\Omega}$ are expected to be determined by about 5 unknowns (e.g., κ , radial distortion, $\mathbf{n}$ and H) while the number of equations is much larger ($h \times w$). However, we only have a noisy estimate of $\mathbf{\Omega}$, which is the biggest challenge of accurate calibration. To this end, we propose a learning-based method to indiscriminately exploit all entries of the map.

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

$$\mathbf{n} = \frac{p_c}{\|p_c\|_2},$$

$$H = 2 \arctan \frac{\omega}{2f}. \quad (3)$$

Figure 2. Illustrations of geometric relation

3.2. Optimization-based Method

As can be observed from the bottom view of Figure 2, $\mathbf{\Omega}$ contains 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 .

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 in Figure 3, $F(\theta, \kappa)$ and $\frac{\partial F(\theta, \kappa)}{\partial \theta}$ are monotonically increasing functions varying with AoI θ falling in an interval of $[0^\circ, 90^\circ]$ regarding different κ , 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 p_c . We first estimate an initial guess of co-ordinate of p_c . w.r.t. image plane, and represent it as (u_c^0, v_c^0) . As shown in Figure 2, point p_c is the projected circular point [12], therefore, $(u_c^0, v_c^0)^T$ can be estimated by method [22]. We then estimate p_c by solving

$$\sum_{i=1}^2 \sum_{j=1}^{m_i} (F(\theta_{ij}, \kappa) - \alpha_{ij})^2, \quad \theta_{ij} = \arccos \frac{\langle p_c, p_{ij} \rangle}{\|p_c\|_2 \cdot \|p_{ij}\|_2}, \text{ s.t., } f > 0, \quad (4)$$

where p_{ij} , $i = 1, 2, j = 1, 2, \dots, m_i$ is point j in the i -th point set (m_i is the number of points in i -th point set) we obtained through Algorithm 1, α_{ij} are the corresponding reflective amplitude coefficients. Note that the x-axis and y-axis coordinates of p_{ij} and α_{ij} are known, and the z-axis coordinate of p_{ij} is the same to that of p_c .

We consider function F takes either a scalar θ or matrix Θ as the input and outputs either a scalar α or a matrix Ω . We solve Equation (4) based on the Quasi-Newton method [35] initialized by $p_c^0 = (u_c^0, v_c^0, -\omega)$. Figure 4 gives an example to detect the reliable points to fit conics.

Figure 3. Curves of $F(\theta, \kappa)$ and $\frac{\partial F(\theta, \kappa)}{\partial \theta}$

Algorithm 1. Detecting two sets of reliable points

```

set  $a = \frac{\max(\Omega) - \min(\Omega)}{\omega + h}$ ,  $b = \max(\Omega) - \eta a$ ,

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  $b - a < \min(\Omega)$ 

```

Figure 4. Exmaple of Detecting reliable points to fit conic.

3.3. Learning-based Method

3.3.1. Pipeline

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 p_c . Considering that all entries of Ω are expected to be determined by about 5 un-knowns (including the radial distortion parameter), we propose a learning-based method that fully exploits all entries of Ω to overcome above limitations.

With all entries contained in the reflection amplitude coefficient map, we jointly optimize a reflection separation network (denoted by S) and a calibration network (denoted by C) which should follow such constrains:

$$(\mathbf{T}, \mathbf{R}, \Omega) = S(\mathbf{I}), p_c = C(\Omega). \quad (5)$$

Our image separation network is based on the method of [53]. An overview of our framework is shown in Figure 5. We use notations with subscript ‘est’ to represent estimated variables and those with ‘gt’ to represent their ground truth.

The 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 our calibration network has the property of pseudo-invariance regarding to rotation and flipping:

$$G(p_c) = C(G(\Omega)), \quad (6)$$

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., $\Omega = G_0(\Omega)$, $p_c = G_0(p_c)$. The detailed network architecture is shown in the Figure 5.

Figure 5. The framework of our learning-based method.

3.3.2. Loss functions

We alternatively optimize S and C. The loss functions of L^S and L^C are,

$$L^S = LWT + L_e, L^C = L_e + L_g, \quad (7)$$

where LWT 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. l_e and l_g are l_1 loss functions for p_c ,

$$\begin{aligned} \mathcal{L}_e &= \frac{1}{4} \sum_{i=0}^3 \|\mathcal{C}(Gi(\Omega_{est})) - Gi(p_c, gt)\|_1, \\ \mathcal{L}_g &= \frac{1}{4} \sum_{i=0}^3 \|\mathcal{C}(Gi(\Omega_{gt})) - Gi(p_c, gt)\|_1, \end{aligned} \quad (8)$$

where Ω_{est} is estimated from reflection separation network, *i.e.*, $\Omega_{est} = S(I_{gt})$.

3.3.3. Implementation details

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. The two networks are jointly trained at single Nvidia 1080i GPU in Linux Ubuntu 16.04 environment for 200 epochs with our synthesized training data. The best epoch was chosen for the validation and comparison.

3.4. Data Preparation

3.4.1. Training dataset

3.4.1.1. Assumption

Previous calibration methods using imaged concentric circles cannot be directly applied to our problem due to the under-constrained input, as mentioned before. 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 have

$$\alpha = F(\theta, \kappa) = \frac{1}{2}(\mathcal{R}_{\perp} + \mathcal{R}_{\parallel}) = \frac{1}{2}\left(\frac{2\mathcal{R}_{\perp}^s}{1+\mathcal{R}_{\perp}^s} + \frac{2\mathcal{R}_{\parallel}^s}{1+\mathcal{R}_{\parallel}^s}\right), \quad (9)$$

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

$$\mathcal{R}_{\perp}^s = \left(\frac{\cos \theta - \kappa \cos \theta_t}{\cos \theta + \kappa \cos \theta_t}\right)^2, \quad \mathcal{R}_{\parallel}^s = \left(\frac{\cos \theta_t - \kappa \cos \theta}{\cos \theta_t + \kappa \cos \theta}\right)^2, \quad (10)$$

where $\theta_t = \arcsin\left(\frac{1}{\kappa} \sin \theta\right)$ according to Snell’s law. 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.

Following strategies adopted in learning-based methods for single reflection separation (*e.g.*, [49, 53]) and single image camera calibration (*e.g.*, [32, 56]), we synthesize our training data, but in a physical rule-based way. 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 transmission and reflection image 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 n and horizontal FoV H . Then Ω is calculated based on Equation (9) with a random κ . Figure 6 shows an example of such reflection and refraction map synthesized. Finally, we synthesize 11041 reflection-contaminated images $\mathbf{I}$ in total for training according to Equation (1) which are all visually similar to the reflection-contaminated images in our daily life.

FocaLens database [60] consists of training and testing subsets with numbers of 173, 000 and 61, 000 respectively. All images are captured with horizontal FoVs between 6.87° and 117.2° , then resized to 320×480 .

Figure 6. Map of angle of incidence, Reflection Map, Refraction Map

Figure 7. Simulation of AoI in different FoV and glass orientations

3.4.1.2. Obtaining $\mathbf{T}$ and $\mathbf{R}$

To obtain subsets of $\mathbf{T}$ and $\mathbf{R}$, we choose images with horizontal FoVs between 19.3° (105mm) and 73.7° (24mm) in the training subset of FocaLens database [60], because these FoVs are expected to be more commonly found in our daily life. Note that numbers of images are not uniform regarding different FoVs in FocaLens database [60]. To balance image numbers with different FoVs in our training data, we utilize part of images in the FocaLens database [60]. More specifically, we randomly select 400 images if the image number is larger than 400 and neglect them if the number is smaller than 25, for each FoV. We also neglect all gray images. Finally, we obtain 22, 082 images with 75 different horizontal FoVs. These images are evenly divided into two subsets of transmission ($\mathbf{T}$) and reflection (11, 041 for each), according to their FoVs. Therefore, each transmission image and reflection image are paired with the same FoV. To generalize more realistic reflection images, we process each image in the reflection subset and obtain the final subset of reflection images ($\mathbf{R}$).

3.4.1.3. Generating reflection coefficient map

We use functions F and P in Equation (2) to generate $\mathbf{\Omega}$. More specifically, given a Generating $\mathbf{\Omega}$. pair of $\mathbf{T}$ and $\mathbf{R}$ as well as their horizontal FoV H , a random orientation of glass $\mathbf{n}$ is generated by

$$\mathbf{n} = (\sin\theta_1 \cos\theta_2, \sin\theta_1 \sin\theta_2, \cos\theta_1)^T, \quad (11)$$

$$\theta_1 \sim U(0, m), m = \text{arccot}\left(\frac{\sqrt{h^2 + w^2} \tan\frac{H}{2}}{\omega}\right), \quad (12)$$

$$\theta_2 = a + \max\left\{\min\left\{b, \frac{\pi}{2}\right\}, -\frac{\pi}{2}\right\}, a \sim U\{0, \pi\}, b \sim N(0, \delta), \quad (13)$$

where $\mathcal{U}$ and $\mathcal{N}$ represent Uniform distribution and Gaussian distribution, respectively. δ is set to $\frac{\pi}{9}$ in our implementation. $h \times w$ (320×480) are height and width of images. m is the maximum angle of θ_1 which is to guarantee the glass plane filling the whole image plane. We use a Gaussian distribution to model θ_2 based on the observation that the dominant variation of glass orientation is more commonly observed in the direction of x-axis than that of y-axis (the Cartesian coordinate system is defined in Figure 2 in the real world. With H and $\mathbf{n}$, the angle of incident map Θ can be computed according to function P introduced in Section 3.1. Finally, we generate $\mathbf{\Omega}$ according Equation (9) with $\kappa \sim U(1.4, 1.6)$. With $\mathbf{T}$, $\mathbf{R}$, and $\mathbf{\Omega}$, we

synthesize our training data based on Equation (1). Figure 6 shows some examples of paired reflection contaminated image I by our synthesized method and the corresponding $(1 - \Omega) \circ T, \Omega \circ R, \Omega$.

Figure 8. Two paired Synthesized training data

3.4.2. Validation dataset

3.4.2.1. Basic settings

To demonstrate the generality and reliability of our proposed method, we collect 320 real data in the wild 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 10) using classic calibration algorithm. These 320 images are captured with 8 different commonly used focal lengths, for each focal length our dataset contains between 35 and 45 images.

All validation data are captured by a Canon EOS 5D Mark III camera with apertures between $\frac{f}{2}$ and $\frac{f}{16}$. The original or raw images are with a size of 3840×5760 . We resize them to the size of 320×480 for our validation. An overview of our validation dataset is illustrated in Figure 13. Subfigures (a)-(d) illustrate our data from 4 controlled scenes, i.e., garage, classroom, building. Subfigure (e) illustrates our data from wild scenario. Images are arranged according to glass orientation n and horizontal FoV H , i.e., vertical direction: the angle of pan of camera regarding to vector $\mathbf{n}$, horizontal direction: horizontal FoVs. Numbers above images represent their horizontal FoVs. Regarding the glass, 160 out of 320 images were captured in a controlled scenario and the remaining ones were collected in the wild scenario. For the controlled one, we capture images with portable glass, so we are able to obtain the ground truth transmission image by removing the glass and control the glass orientations (in roughly 5 levels from left to the right, 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, so we could not get the corresponding ground truth transmission image for numerical validation for this part. But this could provide us with more general cases for visual validation. Figure 9 gives an example of our shooting facilities and settings. The first three pictures are the corresponding images captured in the camera of the following three pictures.

Figure 9. Data collection settings

3.4.2.2. Calibrating glass orientation

The ground truth of glass orientation can be calculated by calibrating the extrinsic rotation matrix of a camera regarding a checkboard pasted on the glass (as shown in Figure 4). Note that the camera is strictly fixed for the capture of a reflection-contaminated image and the image with the checkboard. In order to achieve accurate glass orientations, distortion has to be corrected at first. We estimate distortion parameters by calibrating the camera with each focal length. Figure 11 shows images used for camera calibration with focal length 24mm. Our implementation is based on the OpenCV. Figure 12 illustrates images before and after distortion correction. Note that such distortion correction is also applied to obtain our UNDISTORTED data.

3.4.2.3. Aligning images for reflection removal

We capture transmission images for our CONTROLLED data by removing the portable glass. The exposure settings before and after removing the glass are kept unchanged. To alleviate the spatial shift between a reflection-contaminated image and its transmission image caused by the glass, we align them based on the method in [47].

Specifically, we crop a center region of transmission images and align their reflection-contaminated images regarding this region.

To maintain uniform FoVs, the size of cropped images is fixed (5040×3360 in our data). All cropped images are scaled to the size of 320×480 . Horizontal FoVs of cropped images are recalculated accordingly,

$$H_{crop} = 2 \arctan \left(\frac{\omega_c}{\omega_0} \cdot \tan \frac{H_{original}}{2} \right), \quad (14)$$

where $\omega_c = 5040$ and $\omega_0 = 5760$ are the width of the cropped images and the raw images respectively.

Table 1 displays the distribution of image numbers regarding different horizontal FoVs for CONTROLLED data (after cropping) and WILD data. Note that these numbers are consistent with coordinates of X-axis in Figure 18 which is in ascending order of horizontal FoVs.

Figure 10. Data collection for controlled and wild scene

Figure 11. An illustration of images used for distortion estimation

Figure 12. Illustrations of distortion correction

Figure 13. An overview of our validation data containing 320 images

CONTROLLED	17.06°	20.29°	26.46°	29.42°	34.97°	42.99°	58.72°	66.55°	Total
Image number	20	20	20	20	20	20	20	20	160
WILD	19.45°	23.12°	30.08°	33.40°	39.60°	48.46°	65.47°	73.74°	Total
Image number	15	16	18	21	25	25	20	20	160

Table 1 Numbers of images regarding different FoVs

3.5. 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 the indoor and the 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.

The framework of estimating panorama from a reflection-contaminated image is outlined in Figure 19. 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 $\mathbf{n}$, H and masking out remaining regions.

We use the same architecture and default parameters as those in [66] to train our panorama estimation network like an image completion task like Figure 14. Training options of ‘-no flip’ and ‘-no rotation’ are modified by the value of ‘True’ and ‘-mask type’ is set to leverage provided masks. Remaining settings are kept exactly unchanged as those in [66]. We generate masks with inputs panorama images with H and $\mathbf{n}$ from our training data. Note that this generation is based on the assumption that appearances of reflection image observed by the camera and the virtual one (regarding the glass, along direction $\mathbf{n}$) are similar, based on which the reflection image can be embedded into a panorama based on simple geometric relation. The image size of inputs and outputs is 128×256 .

Figure 14. An architecture in [66] we used to train our panorama estimation network

Chapter 4

Experimental Results

In this section, we report our evaluation for camera calibration and reflection separation in both numerical comparison as well as visual quality comparison with other state-of-the-art methods. Moreover, we provide an interesting byproduct to estimate a panorama based on the results of our method.

4.1. Evaluation for Reflection Removal

Since our separation network and calibration network are jointly optimized, the calibration network implicitly constrains the output (i.e., Ω) of the separation network to improve the reflection removal results.

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.

Note that for all learning-based methods, we retrain their models using our training data from the scratch. To perform a fair comparison with these methods, we report the results with better performance between the pre-trained models provided by authors and our retrained models for each dataset.

As can be observed from Table 2, our method achieves satisfactory performance in all of the four datasets, where we can see 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.

Zoom in for better details. As can be observed, ZN18 [64] and WS19 [49] produce darker results which could be seen as color distortion, YM19 [62] and WY19 [52] output over-smooth ones which could destruct the original background image for further use, and results from WT19 [53] contain unexpected artifacts. Our method alleviates these problems and hence produces better results as compared with other methods.

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

Table 2. Quantitative results

Considering we borrow the reflection separation network from WT19 [53], our superior 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 14 which contains three comparison images.

For the first comparison, we show the whole structure and color comparison. 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.

For the second and the third comparison, we show the details comparison. We select two from each dataset in the second image and four images from CONTROLLED dataset in the last image of Figure 14.

In the second comparison: from top to bottom are samples from SIR2 [47], ZN18-DATA [64], PB19-DATA [37]. 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.

In the third comparison: the four rows are all from CONTROLLED dataset.

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.

Rough comparison for all of the four datasets

Comparison for detailed SIR2 [47], ZN18- DATA [64], PB19-DATA [37]

Comparison for detailed CONTROLLED dataset

Figure 15. Visual quality comparison

4.2. 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 $e_{\mathbf{n}}$ (in degree) and the absolute error e_H (in degree), respectively.

4.2.1. Overall Performance

Considering camera distortion is not addressed in our methods, we report results regarding inputs with or without distortion correction. Table 3 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 (4) 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.

4.2.2. 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. Besides, 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] <http://rachmaninoff.gel.ulaval.ca:8003/> (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 4 and Figure 18 (where 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 .), 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 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 11 (Table on the right shows the predicted horizontal FoVs and its ground truth (GT) in degree). 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 18, 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 4 displays that the performance of such a simple fusion. Figure 19 shows the fusion results of two examples.

The performance improvement for both sub-sets demonstrate the fundamental difference between our method and HS [16] that leverages image contents for camera calibration.

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

Table 3. Quantitative performance of our methods

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

Table 4. Overall results of HS [16], our methods, and fusion

4.2.3. Results Analysis

For Controlled data, Figure 15 shows the error distributions for each scene of our CONTROLLED data. Each subfigure contains error distributions of our method (left) and HS [16] (right), and corresponds to 4 scenes in subfigure (a)-(d) in Figure 13. Each color block represents an estimation error e_H (in degree) for an image in Figure 13 (a)-(d) with the same arrangement. Blue to yellow means small to large errors.

Both of our method and HS [16] achieve good performance when the FoV of inputs is small for all scenes, which is consistent with our observation in Figure 16. For scenes of Garage and Classroom, HS [16] achieves poor estimation when the FoV is large because these inputs contain less discriminative features to estimate horizon line (as illustrated in Figure 13 (a)-(b)). However, our method produces much more accurate predictions for these two scenes, thanks to the simple transmission or reflection that facilitates accurate estimation of our calibration cue Ω . For two building scenes, HS [16] outperforms our method. Because the horizon line is more likely to be estimated from building images (as illustrated in Figure 13 (c)-(d)), while complicated image contents increase the ambiguity of reflection separation resulting in the unreliable estimation of our calibration cue Ω . These results are consistent with our analysis for Figure 16. Observations above intuitively demonstrate that our method is free from image contents which is fundamentally different from existing methods such as HS [16].

Figure 16. Error distribution of the validation Controlled dataset

Figure 17 shows some results as well as their comparisons with HS [16]. Ω in the third row are simulated according to Equation (8) with their ground truth of $\mathbf{n}$, H , and $\kappa = 1.474$ as inputs. All Ω in this figure are min-max normalized for easy comparison. Table at the bottom displays the ground truth of horizontal FoVs, predicted results from our method and HS [16], and their errors e_H . Images are arranged in ascending order of errors e_H by our method. Note that different from the visualization of Ω in Figure 1, all Ω in Figure 10 are min-max normalized for easy comparison. Errors of estimated FoVs by our method become large from column 1 to column 10. This is consistent with the similarity varying between a predicted Ω and the one simulated based on the ground truth of $\mathbf{n}$, H . The magnitude of intensity changing is more likely to be observed for an accurately predicted Ω , e.g., in a dominant direction (column1-column3) or from boundary to center (column4). In contrast, the intensity of inaccurately predicted Ω tends to be uniform or irregular (column6-10). Such inaccurate predictions of FoVs are brought by the inaccurately predicted Ω , due to the challenge of single image reflection removal.

Figure 17. Ten examples from our validation Wild dataset

Figure 18. Results of HS [16] and our method

Figure 19. Results from the CONTROLLED and WILD subset

4.4. Byproduct of panorama estimation

Figure 20 (right) shows our estimated result as well as the visual quality comparison with GS [11] (<http://rachmaninoff.gel.ulaval.ca:8001/>). In Figure 20, Left: The framework to estimate a panorama from (a) a single reflection-contaminated image $\mathbf{I}$. (b) Our calibration method estimates transmission $\mathbf{T}$ and reflection $\mathbf{R}$ as well as their camera parameters $\mathbf{n}$, H . (c) An incomplete panorama is then obtained by embedding $\mathbf{T}$ and $\mathbf{R}$ based on the geometric relation determined by $\mathbf{n}$, H , i.e., fix $\mathbf{T}$ at the middle, the spatial position of $\mathbf{R}$ in this incomplete panorama is determined by $\mathbf{n}$, 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 $\mathbf{T}_{\text{gt}}$ as the input.

Note that GS [11] takes reflection-free $\mathbf{T}_{\text{gt}}$ as the input. For easier comparison, we provide the panorama captured by a lighting probe for reference and transform panoramas as light probes through HDRshop. (Note that the images of light probes are left-right flipped for easier comparison. HDRshop: <http://ict.debevec.org/~debevec/HDRShop/main-pages/tutorials.html>) 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 the middle part of Figure 20 (b))

Figure 20. Panorama completion pipeline and visual results

4.5. Ablation Study of Pseudo-invariance

In order to validate the effectiveness of considering the property of pseudo-invariance, we compare our method with that without considering such a property. This comparison is achieved by replacing loss function in Equation (8) by

$$\begin{aligned}\mathcal{L}_e &= \|\mathcal{C}(Gi(\Omega_{est})) - Gi(p_c, gt)\|_1, \\ \mathcal{L}_g &= \|\mathcal{C}(Gi(\Omega_{gt})) - Gi(p_c, gt)\|_1,\end{aligned}\quad (15)$$

The remaining settings (including training data and parameters) are kept unchanged. The superior performance of our method in Table 5 demonstrates the effectiveness of considering the property of pseudo-invariance. It shows quantitative performance of our method and that without considering the property of pseudo-invariance (ours w/o property). e_n , e_H (in degree) are from experiments of camera calibration on our dataset, and SSIM, PSNR are from experiments of reflection removal on SIR2 [47], ZN18-DATA [64], PB19-DATA [37], and our CONTROLLED datasets.

	e_n e_H	SIR ²	ZN18- DATA	PB19- DATA	CONTROLLED
		SSIM PSNR	SSIM PSNR	SSIM PSNR	SSIM PSNR
Ours	27.61 16.30	0.8895 23.74	0.7510 19.08	0.8736 16.88	0.8445 23.93
Ours w/o property	27.94 18.23	0.8688 23.04	0.7458 18.78	0.8509 16.81	0.8133 23.29

Table 5. Ablation study of the property of pseudo-invariance

Chapter 5

Conclusion and Future work

5.1. Conclusion

This work successfully gives a reliable answer to the question of ‘what plate glass relates about calibration cues’ which utilizes the estimated reflection amplitude coefficient map as a to guide the reflection removal as well as the camera calibration process. For the reflection removal, the proposed model facilitates the reflection removal process since the estimated coefficient map explicitly tells the model where reflection is strong and where it is weak like attention mechanism, but in an implicit manner. The numerical and visual quality comparison with other five state of the art methods in the same training setting demonstrate the superiority of proposed method. It also proves the camera calibration network could be a positive constraint to regularize the image separation model and beneficial to the whole model performance.

For the camera calibration, we first collect two datasets in controlled and wild settings with 320 images in total. The controlled dataset was collected with portable glass, so we get the ground truth background images for reflection removal validation. For the wild dataset, we took pictures before the natural glass like the windows and glass

doors that is impossible to get the ground truth background images, but those reflection could be more various and closer to the daily reflection-contaminated images in the real applications.

After the separation model, the proposed model successfully predicts the reflection coefficient map for most cases. Based on the estimated reflection coefficient map, both of the optimization-based and the learning-based method get satisfying calibration results. Because the learning-based method could fully utilize all entries of the input map, the estimation error is lower than the optimization-based method in terms of both glass orientation and FoV estimation. Although our proposed method could not outperform state of the art single camera calibration method in terms of FoV estimation, through experimental analysis, we find our method essentially uses different prior knowledge which is content free, so combined with our method results, the compared method could be improved which demonstrate the effectiveness of our method.

What’s more, by considering reflection as an additional view, our method could better solve the single image panorama estimation problem.

5.2. Broad impact

Our work shows that by adding more geometric constraints in the data-driven methods, we could approach the real data distribution more, so the ill-posedness of such problems could be alleviated and the model performance could be guaranteed. What’s more, some geometric cues always have some physical symbolic which may facilitate other related computer vision problems. Our work is the first one to take glass orientation into consideration to jointly solve the camera calibration and reflection removal problems which could bring some inspiration for further research.

5.3. Limitation

Although our method is image contents agnostic, it is sensitive to the calibration cue Ω , so it would be not so robust in some cases. For example, Figure 21 show two failure scenarios 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 $H_{est} = 23.0^\circ$. In this scenario, illumination is not uniformly distributed, which could have a reverse effect to our assumption and model. Also, our model is not so robust in some similar scenarios with the same image content which deserve further investigation. We did

not take absorption effect into consideration in our data synthesis process and this effect could have some impact on the reflection removal process.

Figure 21. Two failure examples from our validation data

5.4. Future work

Based on the above limitations and our analysis, we would separate our future work into three part to make our research more significant.

5.4.1. Comprehensive reflection formation

In the current thesis, we keep the reflection formation process simple without considering too much physical phenomenon. In the next step, we would make our physical rule-based data synthesis process more comprehensive and completed. Currently, we have successfully validated the superiority of our idea with geometric constraints, but it is not comprehensive because we neglect absorption effect of the energy loss when light transmits the glass, and also some parameters such as the glass material have been ignored, some special effects like ghost effect and blurry effect could also be simulated with some camera or glass parameters. To be more specific, we would like to study the absorption effect, ghost effect and blurry effect and see how they could have effect on our reflection removal model based on the observation that the more physical information we encode in our model, the more our model could approach the real data distribution to facilitate the down-stream tasks.

Based on the current investigation, the absorption effect controls the attenuation of content-free variables and is variable according to different colors, thicknesses, or orientations of glass in real-world. We aim to find some parameters that can be numerically approximate the absorption effect.

5.4.2. Robustness investigation

Deep learning models are vulnerable to little input perturbation which has been focused for a long time especially in computer vision community. For this reflection

removal task, we would like to make some experiments to investigate the generalization ability and robustness of our reflection removal model with little perturbation in terms of some physical parameters in the contaminated-image synthesis process and try to find a reliable solution by adding some constraints to our network to make it capable of tolerating such perturbations. This work is essential because as we know, current deep learning model trained on the specific training datasets may not be applied to real scenarios and often fail to give robust results when it comes to deal with similar inputs with human-unobservable perturbations, thus we aim to improve the generalization and robustness of our model in the future.

5.4.3. Model improvement in terms of quality and speed

Considering the rapid development of the deep learning models, we aim to further improve the performance of our reflection removal model which is the core task. We will try to reformulate the network structure or part of image formation model which could adapt the additional physical information above to improve the reflection removal quality and reach the state-of-the-art performance. On the other hand, the current network is not available for the real-time inference with several seconds for each input image. Based on the current research in model compression and model distillation technique, it is of great importance to make the proposed model lighter to deploy on the real-time systems. To be more specific, we could train our teacher model with full physical information embedded and complex model structure as well as designed loss such as different CNN layer perceptual loss and GAN loss and gradient loss to achieve desired performance. Then we aim to train a student network which just use the necessary parameters and the basic network structure but add one more loss to mimic the teacher model behaviors that is regarding the teacher model output as a soft supervision label for training but keep the model light during the inference period thus our model could be more compact to be deployed in the real time for more practical use.

AUTHOR’S PUBLICATION

Q. Zheng, **J. Chen**, Z. Lu, B. Shi, X. Jiang, K. Yap, L. Duan, A. Kot, “What does Plate Glass Reveal about Camera Calibration?,” IEEE Conference on Computer Vision and Pattern Recognition, (CVPR’20), Jun 2020

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 International 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 International 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 International 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 LongQuan. Detection of concentric circles for camera calibration. In *Proc. of International 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. Deeplight: 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 International 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 Falcao, and Nuno Gonçalves. Unsupervised intrinsic calibration from a single frame using a. In *Proc. of International 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] Yi Chang 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, applications. 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 International Conference on Computer Vision*, 2017. 5, 7
 - [48] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Crnr: 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, AlanCBovik, HamidRSheikh, and EeroPSimoncelli. 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 mis-aligned 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 International 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 International 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/FocaLens/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
- [67] Y. Li and M. S. Brown, “Exploiting reflection change for automatic reflection removal,” in *Proc. of International Conference on Computer Vision, 2013*, pp. 2432–2439.
- [68] R. Wan, B. Shi, T. A. Hwee, and A. C. Kot, “Depth of field guided reflection removal,” in *Proc. of International Conference on Image Processing, 2016*, pp. 21–25.
- [69] Zhe Zhu, Ralph R. Martin and Shi-Min Hu, “Panorama completion for street views” in *Computational Visual Media Vol. 1, No. 1, March 2015*, 49–57
- [70] Shuran Song Thomas Funkhouser, “Neural Illumination: Lighting Prediction for Indoor Environments” In *Proc. of Computer Vision and Pattern Recognition*, pages 6918–6926, 2019
- [71] Stamatios Georgoulis, Konstantinos Rematas, Tobias Ritschel, Mario Fritz, Luc J. Van Gool, and Tinne Tuytelaars, “DeLight-Net: Decomposing Reflectance Maps into Specular Materials and Natural Illumination” CoRR abs/1603.08240, 2016
- [72] David Eigen and Rob Fergus, “Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture” In *Proc. of International Conference on Computer Vision 2015*.
- [73] Aayush Bansal, Bryan Russell, and Abhinav Gupta. Marr, “Revisited: 2D-3D Model Alignment via Surface Normal” In *Proc. of Computer Vision and Pattern Recognition, 2016*

- [74] Chao Li, Yixiao Yang, Kun He, Stephen Lin, and John E. Hopcroft, “Single Image Reflection Removal through Cascaded Refinement” *In Proc. of Computer Vision and Pattern Recognition, 2020*
- [75] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen, “Polarized Reflection Removal with Perfect Alignment in the Wild” *In Proc. of Computer Vision and Pattern Recognition, 2020*
- [76] Rui Li, Simeng Qiu, Guangming Zang, and Wolfgang , “Reflection Separation via Multi-bounce Polarization State Tracing” *In Proc. of European Conference on Computer Vision, 2020*