An Investigation into Automatic Photometric Calibration

by

Feng Chun-Cheng

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2024

© Feng Chun Cheng 2024

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Photometric calibration is a critical process that ensures uniformity in brightness across images captured by a camera. It entails the identification of a function that converts the scene radiance into the pixel values in an image. The goal of the process is to estimate the three photometric parameters - camera response function, vignette, and exposure.

A significant challenge in this field is the heavy reliance on ground truth information in current photometric calibration methods, which is often unavailable in general scenarios. To address this, we investigate our proposed simple method, New Photometric Calibration (NPC), which eliminates the need for ground truth data. Firstly, we integrated our photometric calibration algorithm with long-term pixelwise trackers, MFT, enhancing the system's robustness and reliability. Since the MFT effectively handles occlusion and reduces drifting, it results in a more stable trajectory. By incorporating MFT to track feature points across frames and using the trajectory as corresponding points, we can utilize the pixel intensity of corresponding points to forgo the need for exposure ground truth during initialization. Subsequently, we independently optimize the photometric parameters to sidestep the exponential ambiguity problem.

Our experiments demonstrate that our method achieves results comparable to those utilizing ground truth information, as evidenced by comparable root mean square errors (RMSE) of the three photometric parameters. In scenarios without ground truth data, NPC outperforms existing methods. This indicates that our approach maintains the accuracy of photometric calibration and can be applied to arbitrary videos where ground truth information is not provided.

In conclusion, our research represents a significant advancement in the field of photometric calibration. We investigate a novel and effective method that requires no ground truth information during the photometric calibration process. Our approach incorporates the use of a robust tracker, enhancing the trajectories of feature points, thereby improving the overall performance of our method. Furthermore, our model not only bypasses the exponential ambiguity problem inherent in the optimization process but also addresses the challenges associated with the traditional reliance on ground truth information, outperforming previous photometric calibration methods when the input lacks ground truth data.

Acknowledgements

I would like to express my sincere appreciation to everyone who supported me in completing this thesis and my degree. First of all, thanks to my parents and my brother who greatly supported me mentally. Despite being thousands of kilometers away from home, with only one video call per week and visiting them only once, I am still grateful for their support. Secondly, Ting has always helped, encouraged, and supported me throughout this journey. I am grateful that I made the choice and took the opportunity to study in Canada.

I also want to thank Professor John Zelek, who made this amazing journey begin. From the first day of being a member of the VIP Lab, I have always been grateful and appreciative for providing me with the opportunity to work with an extraordinary group of people here. This has definitely been a life-changing opportunity for me and my family. I will never forget the day I received the offer and my first visit to EC4. It was a rainy and very cold day, with temperatures around 5 degrees Celsius, and I had traveled from Mississauga. It was a two-hour drive on the Go bus, and I couldn't wait to see the University of Waterloo. The emotion was so touching that I will always remember it. Despite the weather, it was definitely one of the best days of my life. I couldn't be more grateful and luckier. Thank you, Professor Zelek, thank for your guidance, insights, and time devoted to my research. Moreover, I would like to thank Professor Wang and Professor Azzi for their time to review my thesis.

Last but not least, I want to thank all the members of the VIP Lab who helped me during my program. Your suggestions and discussions have seriously impressed me. Especially, Ryan, Yuxiang, Matthew and Kevin. All of your advice and support have been valuable assets to me and absolutely helpful for my research. Besides, I want to thank all my friends that I met during this unforgettable journey. Thank you to all of you for making this terrific journey even more colorful.

Dedication

This is dedicated to my family and friends.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgements iv

Dedication v

List of Figures ix

List of Tables xii

List of Abbreviations xiii

1 Introduction 1

1.1 Overview 1

1.2 Problem Scope 3

1.3 Contribution 4

1.4 Thesis Outline 5

2 Background 7

2.1 Photometric Calibration 7

2.1.1 Joint Optimization 11

2.1.2 Sequential Optimization . . . 12
2.1.3 Related Work . . . 17
2.2 Feature Tracker . . . 19
2.3 Image and Video Quality Assessment . . . 21
2.4 Datasets . . . 23
2.5 Summary . . . 24

3 Methodology **25**
3.1 Introduction . . . 25
3.2 Feature Tracker Evaluation . . . 25
3.2.1 Experiment Setup . . . 26
3.2.2 Experiment Results . . . 28
3.2.3 Summary . . . 32
3.3 Photometric Calibration . . . 32
3.4 Evaluation . . . 36

4 Experimental Results **37**
4.1 Analysis with Ground Truth Information . . . 37
4.1.1 Photometric Parameters . . . 37
4.1.2 Image and Video Quality Assessment . . . 44
4.1.3 DSO results . . . 51
4.2 Analysis without Ground Truth Information . . . 57
4.2.1 Photometric Parameters . . . 57
4.2.2 Image and Video Quality Assessment . . . 61
4.2.3 DSO results . . . 65
4.3 Discussion . . . 68

5 Conclusion **69**
5.1 Summary . . . 69
5.2 Future Work . . . 70

References **71**

APPENDICES **79**

A Supplement results **80**

A.1 Stats for Photometric Calibration on TUM MONO Dataset 80

A.2 Stats for Photometric Calibration Without Ground Truth 113

List of Figures

1.1 3D reconstruction with and without photometric calibration 6

2.1 CRF and ICRF Diagram 9
2.2 Illustration of the vignetting effect 10
2.3 Schematic diagram of photometric formation process. 10
2.4 The main steps of OSPC approach 13

3.1 Feature extraction 26
3.2 Flowchart of tracking 27
3.3 Photometric parameters results from various trackers 29
3.4 Reconstructed frames using different trackers 30
3.5 Flowchart for our proposed method 34
3.6 SIFT and MFT comparison 35

4.1 Diagram of RMSE for ICRF across various methods 38
4.2 Diagram of RMSE for vignette across various methods 40
4.3 Diagram of RMSE for exposure across various methods 41
4.4 FR-IQA score for three methods 45
4.5 NR-IQA score for three methods 46
4.6 VQA score for three methods 47
4.7 Reconstructed frames in Sq 16 for various methods 48
4.8 Reconstructed frames in Sq 23 for various methods 49

4.9 Reconstructed frames in Sq 49 for various methods 50
4.10 DSO scene for Sq 07 across various methods 52
4.11 DSO trajectory of Sq 07 across various methods 53
4.12 DSO results for Sq 26 across different methods 54
4.13 DSO results for Sq 29 across different methods 55
4.14 DSO results for Sq 30 across different methods 56
4.15 Photometric calibration without ground truth on ICL-NUIM Dataset . . . 58
4.16 Photometric calibration without ground truth on EuRoc Dataset 59
4.17 Photometric calibration without ground truth on KITTI Dataset 60
4.18 IQA and VQA diagrams for OPC and NPC without ground truth 62
4.19 Reconstructed frames in MH-01-easy 63
4.20 Reconstructed frames in KITTI 05 63
4.21 Reconstructed frames in V2-03-difficult 64
4.22 DSO trajectory on EuRoc MH-01-easy 66
4.23 DSO scene across different sequences 67

A.1 Sequence 07. 87
A.2 Sequence 08. 88
A.3 Sequence 16. 89
A.4 Sequence 23. 90
A.5 Sequence 25. 91
A.6 Sequence 26. 92
A.7 Sequence 28. 93
A.8 Sequence 29. 94
A.9 Sequence 30. 95
A.10 Sequence 31. 96
A.11 Sequence 32. 97
A.12 Sequence 35. 98

A.13 Sequence 36. 99
A.14 Sequence 37. 100
A.15 Sequence 38. 101
A.16 Sequence 39. 102
A.17 Sequence 40. 103
A.18 Sequence 41. 104
A.19 Sequence 42. 105
A.20 Sequence 43. 106
A.21 Sequence 44. 107
A.22 Sequence 45. 108
A.23 Sequence 46. 109
A.24 Sequence 47. 110
A.25 Sequence 48. 111
A.26 Sequence 49. 112

List of Tables

3.1 IQA results for different trackers 31
3.2 VQA results for different trackers 31

4.1 Average and STD of RMSE for ICRF across various methods 39
4.2 Average and STD of RMSE for vignette across various methods 39
4.3 Average and STD of RMSE for exposure across various methods 41
4.4 RMSE of exposure across different patch sizes 42
4.5 RMSE results for SIFT and MFT 42
4.6 Computational time 43
4.7 Average and STD of FR-IQA for three methods 44
4.8 Average and STD of NR-IQA for three methods 46
4.9 Average and STD of VQA for three methods 48
4.10 Average and STD of IQA and VQA across all sequences for OPC and NPC 61
4.11 Lookup table for the sequences and corresponding numbers for the charts. 62

A.1 RMSE of the photometric parameters on TUM MONO Dataset 80
A.2 VQA of photometric calibrated video on TUM MONO Dataset 83
A.3 VQA of photometric calibrated video without ground truth. 113

List of Abbreviations

CLIP ontrastive Language-Image Pretraining 21

CLIPIQA CLIP-based Image Quality Assessment 21, 31

CNN Convolutional Neural Network 21, 22

CNNIQA Convolutional Neural Network-based Image Quality Assessment 21, 31

CRF Camera Response Function 2–4, 7–9, 12, 17–19, 23, 33, 70

DLM Detail Loss Metric 22

DSO Direct Sparse Odometry 2, 6, 11, 24, 35–37, 51–57, 65–68, 70

e Exposure 2, 7

EMoR Empiric Model of Response 7, 18

FR-IQA Full-reference Image Quality Assessment 21, 37, 44, 45, 61, 62

HSO Hybrid Sparse Visual Odometry 2

ICRF Inverse Camera Response Function 8, 9, 12, 14–18, 28, 29, 32, 34–40, 42, 43, 50, 57–60, 68, 80

IMU Inertial Measurement Unit 23

IQA Image Quality Assessment 5, 21, 22, 24, 30, 31, 36, 37, 51, 61

KLT Kanade-Lucas-Tomasi 3, 11, 19, 20, 25, 28–32, 69

LPIPS Learned Perceptual Image Patch Similarity 21, 31, 44

MANIQA Multi-task Attention-based No-reference Image Quality Assessment 21, 31

MAVs Micro Aerial Vehicles 23

MFT Multi-Flow dense Tracker 20, 21, 25, 27–35, 42, 43, 64, 68, 69

NIMA eural Image Assessment 22, 31

NIQE Natural Image Quality Evaluator 21, 31, 45

NPC New Photometric Calibration 5, 32, 37–70, 80–114

NR-IQA No-reference Image Quality Assessment 21, 37, 45–47, 61, 62, 69

OPC Online Photometric Calibration 11, 24, 25, 28, 30, 37–41, 43, 44, 46–69, 80–114

OSPC Online Sequential Photometric Calibration 12–14, 16, 24, 32, 33, 35, 37–57, 68, 80–112

PCA Principal Component Analysis 7

PIPs Persistent Independent Particles 20, 25, 27–32

PSNR Peak Signal-to-Noise Ratio 21, 31, 44, 45

RMSE Root Mean Square Error 28, 29, 36–43, 51, 53–56, 65, 66, 80–82, 87

RR-IQA Reduced-reference Image Quality Assessment 21, 22

SIFT Scale-Invariant Feature Transform 12, 15, 32, 33, 35, 42

SLAM Simultaneous Localization and Mapping 2, 11, 23, 24, 35, 68, 70

SSIM Structural Similarity Index Measure 21, 31

STD Standard Deviation 37, 39–41, 44–48, 61

SVM Support Vector Machine 22

TAPIR Tracking Any Point with per-frame Initialization and temporal Refinement 20, 21, 25, 27–32

V Vignette 2, 7, 50

VIF Visual Information Fidelity 22, 31, 83

VMAF Video Multi-Method Assessment Fusion 22, 31, 47, 83

VO Visual Odometry 2, 23

VQA Video Quality Assessment 5, 21, 22, 24, 31, 36, 37, 47, 48, 50, 51, 61–63, 83–86, 113, 114

Chapter 1

Introduction

1.1 Overview

Photometric calibration is a pivotal process that determining the relationship between the brightness of a scene and the pixel intensity values captured by the camera [23]. Here is a concise overview of this thesis.

The process of photometric calibration encompasses the following tasks:

- the conversion of image intensities into sensor irradiance values,
- the management of intensity variations, and
- the estimation of photometric parameters.

The key steps involved typically comprise:

- the extraction of feature points from video frames,
- the initial estimation of parameters,
- the modeling of the camera response function and vignetting,
- the optimization process,
- the validation and refinement stages, and

- the reconstruction of a photometrically calibrated video.

The photometric calibration process involves identifying a function that translates pixel values into standardized light measurements, known as SI light units. The objective of this procedure is to ascertain the values of the three photometric parameters - Camera Response Function (CRF), Vignette (V), and Exposure (e) [3]. This process allows for the transformation of pixel values, generated by a specific digital camera, into universally recognized photometric quantities [74]. It serves as an indispensable preprocessing step for optical flow computation, Visual Odometry (VO) [57], Direct Sparse Odometry (DSO) [11], and Simultaneous Localization and Mapping (SLAM) systems [10]. All these systems operate based on the fundamental assumption that the brightness of a given point in the scene remains consistent across various images [3]. Additionally, photometric calibration is critical for computational techniques that depend on understanding the light intensity of a scene. This includes methods like photometric stereo, high dynamic range image formation, shape from shading, and color constancy [18].

VO [57] and DSO [11], for instance, use photometric calibration to track camera poses and estimate 3D points. They are fundamental building block for many emerging technologies such as autonomous vehicles, drones, and robot navigation. SLAM [10], on the other hand, utilizes photometric calibration for autonomous operation in unknown environments. SLAM has a wide range of applications, including consumer robotics (such as drones or vacuum cleaners), self-driving cars, augmented reality applications on smartphones, mobile 3D mapping systems, and deploying robots in unexplored environments (such as disaster sites, underwater, or extraterrestrial locations).

The aim of conducting photometric calibration is to improve the efficacy of SLAM [10], VO [57], and DSO [11], particularly in scenarios with drastic changes in brightness. With the help of photometric calibration, the overall performance can be significantly improved, not only in direct method [3] but Hybrid Sparse Visual Odometry (HSO) [51]. Fig. 1.1 demonstrates an instance of with and without photometric calibration in DSO [11]. Yang *et al.* [70] conducted thorough experimental research on photometric calibration, focusing on state-of-the-art direct, feature-based, and semi-direct methods. Their findings revealed that the application of photometric calibration significantly boosts the performance of direct methods. However, for semi-direct and feature-based methods, the effect is not uniform and is contingent upon the specific feature used, the response function of the camera, and the overall light intensity of the scene. Recent direct methods strive to adjust for changes in automatic exposure, for instance, by refining an affine mapping to correlate brightness values across distinct images [13]. However, the majority of direct methods proceed under the presumption of a steady exposure time [12, 35]. Besides, photometric

calibration facilitates the easing of the assumption of constant brightness that is the foundation of widely used feature trackers like Kanade-Lucas-Tomasi (KLT) and other direct algorithms [31, 64]. Accordingly, the purpose of this thesis is to explore the intricacies of photometric calibration. We aim to contribute to the existing body of knowledge by exploring novel methodologies and addressing potential challenges in the field. This research is expected to have significant implications for the advancement of computer vision technologies.

1.2 Problem Scope

Essentially, there are two primary approaches to photometric calibration - the joint optimization method [3] and the sequential method [23]. However, both of these approaches have their own set of challenges.

The joint optimization method, as the name suggests, optimizes three photometric parameters and the tracker simultaneously, leading to an inherent issue known as exponential ambiguity. This ambiguity poses a significant hurdle in photometric calibration, especially when attempting to recover the camera response function and exposure ratios from the intensity mapping, which can have multiple potential solutions [19]. Additionally, the joint optimization method employs the ground truth value of the CRF to tackle the issue of exponential ambiguity. Moreover, it can only compute exposure on a relative scale, meaning it estimates the exposure ratios. To determine precise exposure for each frame, it relies on the ground truth exposure values.

On the other hand, the sequential method estimates three parameters independently [23]. This independent optimization, coupled with the assumption that the inverse response function is a polynomial function, effectively circumvents the exponential ambiguity inherent in the optimization process [19]. However, previous implementations of the sequential method still relied on the ground truth of exposure time [23]. For many videos, auto exposure settings dictate the exposure times, which are typically selected automatically and are beyond the user's control. When viewing a new video, it is generally impossible to access the ground truth information regarding the camera's photometric parameters. In response to this, our model builds upon the sequential method, incorporating further enhancements to eliminate the need of ground truth information and boost its performance.

Furthermore, feature trackers are essential tools in computer vision [66], playing a crucial role in tasks such as image registration, motion analysis, and robot navigation [72]. Most of the currently available feature trackers are based on optical flow, which operates under the assumption of brightness consistency [47]. This assumption posits that the

brightness remains consistent across different frames in a video sequence, which may not always hold true, particularly in datasets with photometric changes. Therefore, the performance can be significantly impacted by photometric changes in datasets, such as variations in lighting conditions. These changes can lead to inconsistencies in feature detection and tracking, thereby affecting the overall accuracy of computer vision systems.

Given the importance of feature trackers and the potential impact of photometric changes, it is of paramount significance to investigate and evaluate their performance under these conditions. In our research, we conducted an in-depth study of the details of this survey, discuss our evaluation methodology, and present our findings. One if our aims is to provide insights into the performance of different feature trackers under varying photometric conditions and identify potential improvements to enhance their robustness and accuracy.

To sum up, a notable challenge with the methods of photometric calibration off the shelf is their heavy reliance on ground truth information of CRF and exposure. This dependency can be problematic, as obtaining accurate ground truth information can be complex and resource-intensive. Moreover, in most cases, the videos we aim to work on do not come with ground truth. To address this, our research first investigates long-term trackers, which are designed to handle various challenges in tracking, including occlusions, appearance changes, and illumination changes [56]. We then investigate a novel approach to photometric calibration that discards the need for ground truth information by initializing the exposure ratio through pixel intensity. This innovative approach aims to overcome the limitations of existing methods and enhance the robustness and accuracy of photometric calibration.

1.3 Contribution

The primary contributions of this thesis are twofold:

- We assess the performance of long-term feature trackers under datasets with photometric changes.
- We investigate a simple method for photometric calibration that utilizes pixel intensity to eliminate the need for ground truth information.

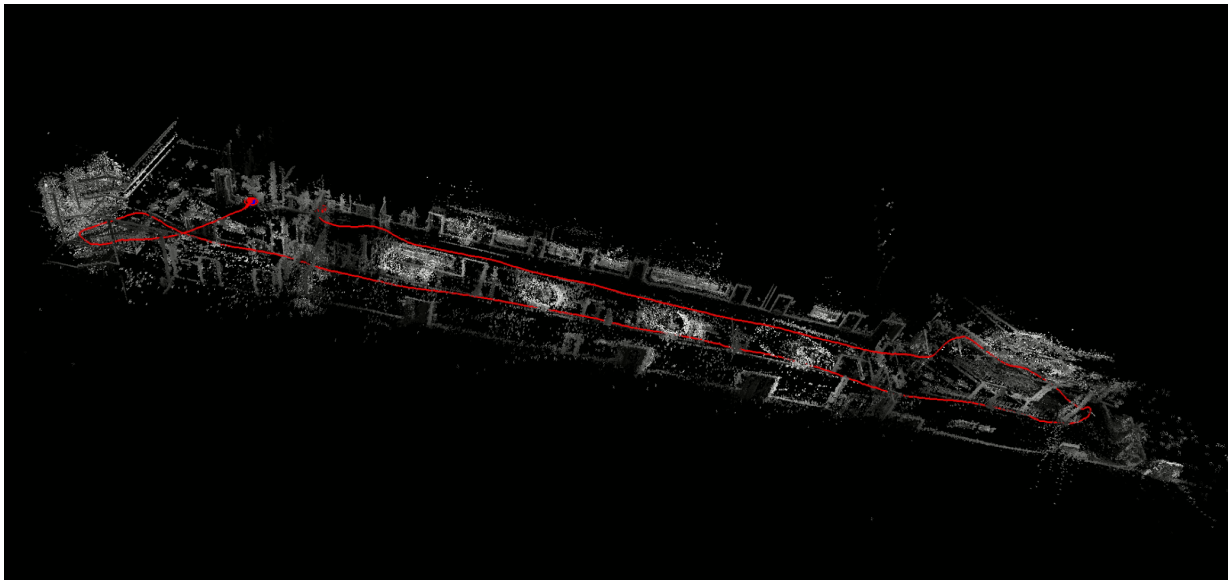
In our research, we first aim to gain a deeper understanding of how different feature trackers perform under various photometric conditions. One of our goals is to identify potential

enhancements that could improve their robustness and accuracy. To achieve this, we explore the details of our investigation, discuss the methodology we employed for evaluation, and share the findings of our research. As a second key point, we investigate a New Photometric Calibration (NPC) method, which does not require any ground truth data by using pixel intensity from the MFT [56]-tracked feature points to initialize the exposure ratio. We will also conduct a comprehensive evaluation of its performance across various datasets. This evaluation will allow us to understand the effectiveness and applicability of our method in different contexts and conditions. It's an essential step in demonstrating the robustness and versatility of our approach.

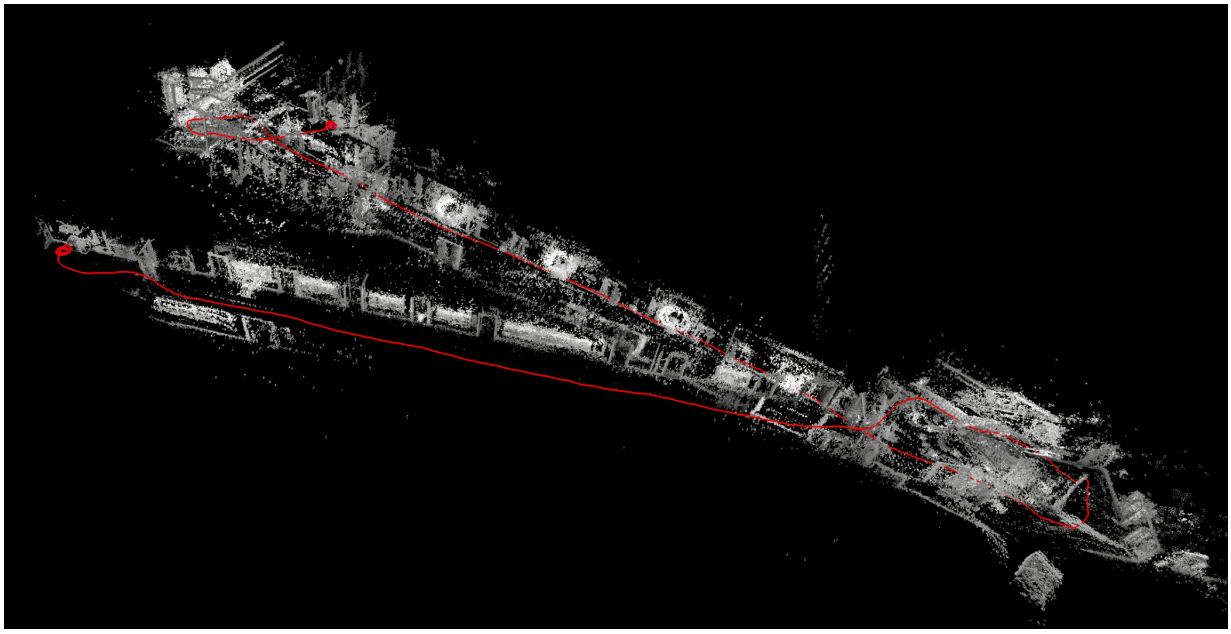
1.4 Thesis Outline

This thesis is organized as follows.

- Chapter 2 provides a comprehensive literature review and background information on photometric calibration and feature trackers. The concepts of Image Quality Assessment (IQA) and Video Quality Assessment (VQA), both of which are pivotal in understanding the performance and effectiveness of feature trackers, are discussed. Additionally, this chapter also introduces the datasets that we utilize in our research.
- In Chapter 3, we present our proposed methodology. This includes a detailed explanation of our research design and the evaluation techniques we employ to assess the performance of our approach. This chapter serves as the foundation of our research, outlining the steps we take to ensure a rigorous and thorough investigation.
- Chapter 4 is dedicated to the presentation of our experimental results. Here, we provide a detailed analysis of our findings, interpreting the data and discussing its implications. This chapter presents understanding the impact and effectiveness of our proposed methodology.
- Lastly, Chapter 5 states and summarizes our results and conclusions as well as identifying future research opportunities.



(a) With photometric calibration.



(b) Without photometric calibration.

Figure 1.1: An example of 3D reconstruction with and without photometric calibration in DSO [11] on the TUM Mono Dataset [14] sequence 40. Without photometric calibration, there is a significant drift, particularly noticeable between the start and end segments. In contrast, when photometric calibration is applied, the drift is considerably reduced.

Chapter 2

Background

2.1 Photometric Calibration

Photometric calibration enables the retrieval of accurate sensor irradiance values from measured intensities, which is a crucial pre-processing step for various vision algorithms. The primary inputs for a photometric calibration process are the corresponding points in frames [23], and the ultimate goal of this process is to recover three main parameters: exposure (e), the CRF, and the attenuation factors of the sensor irradiance due to vignetting (V) [3].

Exposure time, denoted as (t), is the interval during which a camera's sensor is exposed to light, critically influencing the resultant image's brightness. The exposure, represented by (e), is a function of both the aperture size (d) and the exposure time t, following the formula:

$$e = (\pi \cdot d^2) \cdot t \tag{2.1}$$

Typically, exposure times are selected through an automatic process known as auto exposure, which is not subject to user modification. The significance of exposure in image luminance has been well-documented in the literature, such as in Grossberg and Nayar's work [19].

The CRF describes how a camera maps irradiance to pixel values, and it is the main reason that the mapping is nonlinear [30]. It is unique to each camera and can significantly affect the appearance of images. Currently, most methods are built on the Empiric Model of Response (EMoR) introduced by Grossberg and Nayar [20]. Principal Component Analysis (PCA) is utilized to determine the mean response $f_0(x)$ and the basis functions $h_K(x)$.

These elements can be linearly combined to construct the camera response function $f_G(x)$ by selecting appropriate parameters $c_k \in \mathbb{R}$. The CRF at the location x can be written as follow:

$$f_G(x) = f_0(x) + \sum_{k=1}^{n} c_k h_k(x) \tag{2.2}$$

When examining a video frame, the only accessible value is the image intensity O. Consequently, in photometric calibration, actually the goal is to estimate the Inverse Camera Response Function (ICRF), represented as f^{-1}. Grossberg *et al.* [20] have demonstrated that the ICRF can be modeled as a second-degree polynomial with the property of $f_G(0) = 0$, $f_G(1) = 255$. To identify the three parameters, this serves as an effective approximation in terms of both accuracy and the time needed. With pixel intensity O, the ICRF can hence be written as:

$$f^{-1}(O) = c_0 + c_1 O + c_2 O^2 \tag{2.3}$$

Fig. 2.1 presents an example of both the CRF and ICRF derived from sequence 30 in the TUM MONO Dataset [14]. The ground truth provided for all 50 sequences in this dataset is the ICRF. The figure demonstrates the non-linear mapping characteristic of this function. The horizontal axis represents intensity, while the vertical axis signifies irradiance. This arrangement illustrates that the CRF maps irradiance to intensity and vice versa.

Vignetting, which is caused by the optical system of the camera, refers to radiative attenuation of pixel intensity towards image boundaries [18]. Vignetting is commonly perceived as a radial and symmetric phenomenon centered around the image, which can be mathematically represented by the following equation, where $R(x)$ is the radial distance at the position x:

$$V(x) = 1 + v_1 R(x)^2 + v_2 R(x)^4 + v_3 R(x)^6 \tag{2.4}$$

Fig. 2.2a illustrates an example of the vignette effect, which is fundamentally represented as radial. Correspondingly, Fig. 2.2b demonstrates that the vignetting factor diminishes from the center of the image towards the corner.

Given the knowledge of the three photometric parameters, the photometric formation process can be described, as shown in Fig. 2.3. A scene point on a Lambertian surface is illuminated by a light source, reflects the light, and is then captured by a camera sensor. Radiance L, is the measure of light reflected from a point in a scene. It represents the energy that a surface emits in a specific direction per unit area, encapsulated within a

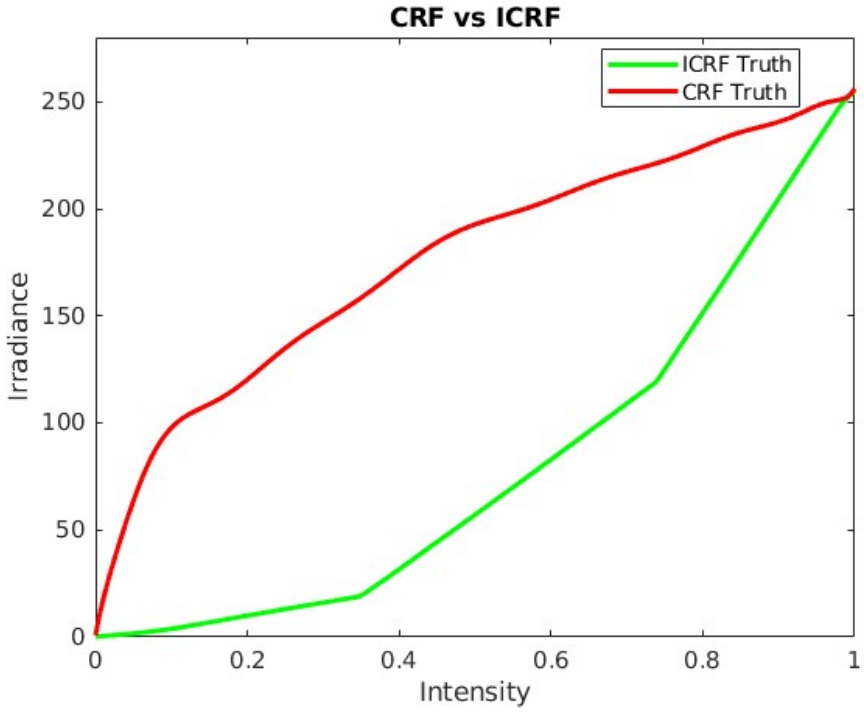


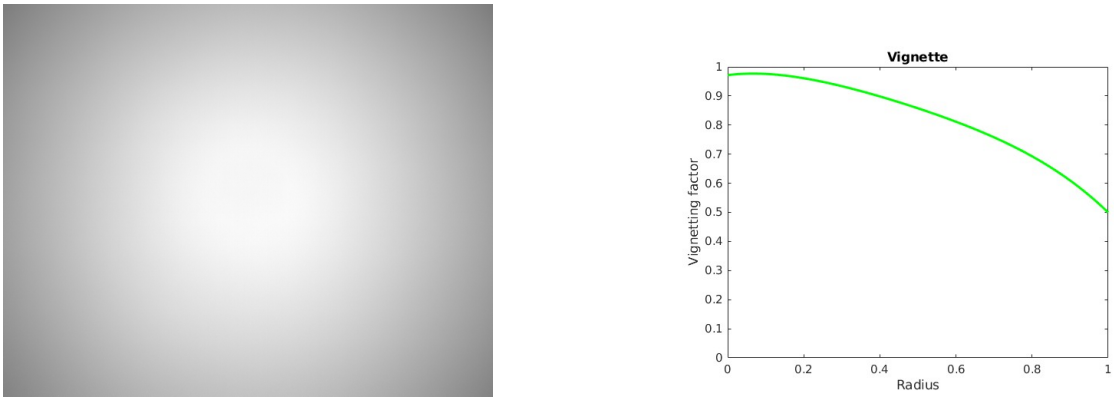
Figure 2.1: The function of both the CRF and ICRF for sequence 30 in the TUM MONO Dataset [14].

unit solid angle [27]. The total amount of energy received at the sensor location x is termed as the Irradiance $I(x)$. The irradiance $I(x)$ is influenced by vignetting based on its location, expressed as $I(x) = V(x)L$. The sensor irradiance is integrated over a time window determined by the camera's exposure time. Consequently, the total irradiance value is given as:

$$I_t(x) = eI(x) \tag{2.5}$$

Finally, the camera response function maps I_t into pixel intensity O. This transformation is expressed as Eq. 2.6, for the CRF f maps total irradiance to pixel intensity O:

$$O(x) = f(eV(x)L) \tag{2.6}$$



(a) Vignette (b) Vignette function

Figure 2.2: Illustration of the vignetting effect for sequence 30 in the TUM MONO Dataset [14]. Fig. 2.2a demonstrates the gradual decrease in brightness towards the edges of the image, showcasing the vignetting effect. Fig. 2.2b depicts the vignette function, where the radius represents the radial distance from the image center, and the curve illustrates how the vignetting intensity diminishes from the center outward.

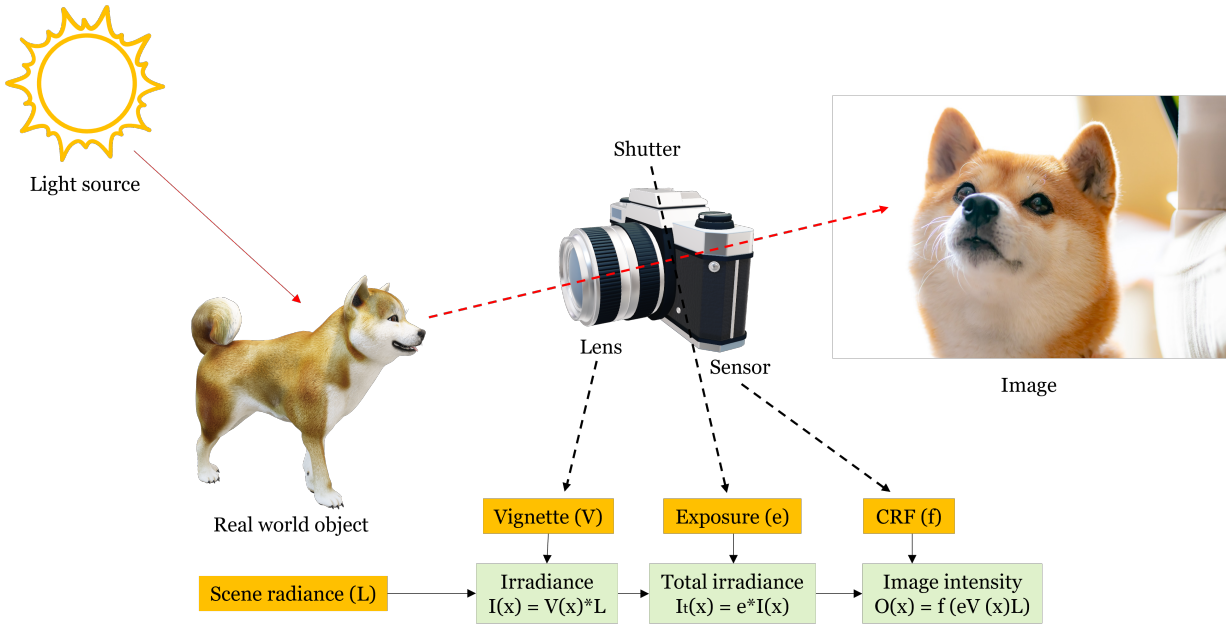


Figure 2.3: Schematic diagram of photometric formation process.

2.1.1 Joint Optimization

One of the most well-known joint optimization method is Online Photometric Calibration (OPC), proposed by Bergmann *et al.* [3] and is adopted and verified by many, such as [28, 46, 47]. Hu *et al.* [28] exploited OPC to address the light-sensitivity issue inherent in Visual-Inertial SLAM based on the sparse optical flow method. They enhanced the stability of front-end visual tracking and improve the accuracy of sliding window optimization at the back-end. Liu *et al.* [46] incorporated a joint method into the DSO with loop-closure algorithm to address its issue of poor positioning and mapping robustness. This design dynamically adjusts the exposure of the input sequence, enhancing the algorithm's robustness under varying illumination conditions. Liu *et al.* [47] proposed a gain-adaptive KLT optical flow algorithm with OPC. They designed a monocular visual-inertial odometry that is insensitive to brightness changes. This method can calibrate the photometric parameters online in real time, meet the assumption of constant brightness in practical applications, and make the algorithm more robust and accurate in the case of dynamic changes in brightness.

In the joint optimization process, the three photometric parameters are optimized simultaneously using the following objective function Eq. 2.7. Consider a collection of scene points P, each tracked over a series of images. For each point $p \in P$, we apply Eq. 2.6 to establish an energy function expressed as Eq. 2.7. In this context, O_i^p, the only known parameter, represents the measured intensity of point p in image i among all the frames F, and it is the only known variable. e_i denotes the exposure for image i, L_p indicates the radiance of point p, and x_i^p specifies the spatial coordinates where point p is projected onto image i. The term w_i^p is a weighting factor, and the Huber norm, which provides robustness to outliers, is employed, symbolized by $|\cdot|_h$.

$$E = \min_{f,e,V,L} \sum_{p \in P} \sum_{i \in F} w_i^p \left| O_i^p - f(e_i V(x_i^p) L^p) \right|_h \tag{2.7}$$

Subsequently, Eq. 2.7 is optimized through the application of the Gauss-Newton algorithm, enhanced with Levenberg-Marquardt damping, and the utilization of analytic Jacobians. The update of radiance L_p for each point in the scene can be independently calculated using Eq. 2.8.

$$\Delta L = \frac{-J_p^T W_p r_p}{(1+\lambda) J_p^T W_p J_p} \tag{2.8}$$

where J_p is the jacobian matrix that is created for each point p, and it comprises the partial derivatives of its residuals with respect to the scene points radiance. The matrix

W_p encompasses the weights for optimization, r_p represents the stacked vector of residuals, and $\lambda \in \Re$ is the damping factor within the LM iteration process.

It has been observed that there is an exponential ambiguity in the recovery of parameters during the optimization process when the response function is unknown. This ambiguity can lead to the optimization process settling into local optima rather than finding the global optimum and there exists infinite possible solutions [19]. The majority of current photometric calibration techniques employ a joint optimization approach to determine all parameters. While this approach is thorough, it is also computationally demanding and introduces a level of uncertainty. This ambiguity is typically mitigated by incorporating ground truth data into the process. For instance, Bergmann *et al.* [3] addressed this issue by employing CRF ground truth into their optimization process. They constrained the response to pass through a specific point and recovered a 2^8-valued lookup table, similar to the method used by Debevec *et al.* [7]. The ground truth information was provided by TUM MONO Dataset [14]. By integrating real-world information into the process, they were able to significantly reduce the ambiguity inherent in parameter recovery. However, the ground truth information is not available in most cases.

2.1.2 Sequential Optimization

Haidar *et al.* [23] proposed Online Sequential Photometric Calibration (OSPC), which adopted a sequential estimation approach to solve for photometric parameters. Fig. 2.4 demonstrates the main steps of OSPC method. This approach independently computes the ICRF, vignette, and the exposure values. The reason it is ICRF here rather than CRF is that the only known variable is the image pixel intensity O. Therefore, we need ICRF to inversely calculate irradiance. By assuming that the ICRF is a polynomial function, we can effectively avoid the exponential ambiguity that usually occurs during the joint optimization process. It has been mathematically demonstrated by Grossberg *et al.* [19] that this sequential photometric calibration method is devoid of exponential ambiguity. Haidar *et al.* rely on the ground truth of exposure time as the initial value. Our objective is to enhance this method and eliminate the reliance on ground truth data.

The sequential optimization process has three steps. The first step is to solve the ICRF, denotes as f^{-1}. Initially, identify the corresponding points on two frames using the Scale-Invariant Feature Transform (SIFT) algorithm [49] and ensure their proximity to neglect the vignetting effect. The radial map is calculated beforehand to accelerate the process and the radial distance is set to be smaller than the threshold $t = 0.1$. When these points are near each other, they share almost the same location, resulting in a nearly

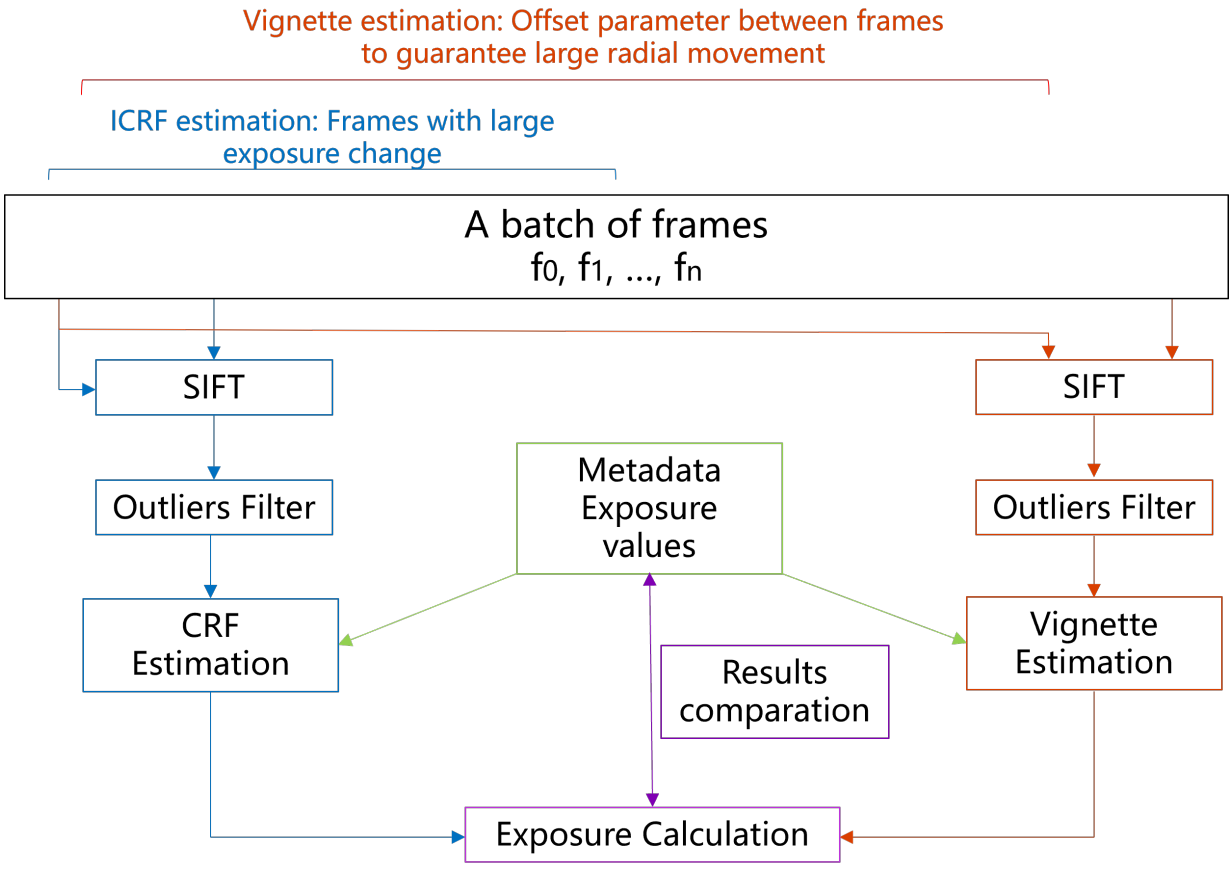


Figure 2.4: The main steps of OSPC [23] approach.

identical vignetting factor. The two frames need not be successive but require sufficient large exposure changes so that the exposure ratio is large or small enough to possess independent vectors, namely, full rank, in the optimization process. Furthermore, assume that the corresponding points are the exact same point in the scene on a Lambertian surface, they should have same radiance and hence cancelled. Building upon photometric formation function as Eq. 2.6, the division of two corresponding points across two frames with vignette at the location $V(x)$ and exposure of each frame e can be express in Eq. 2.9.

$$\frac{f^{-1}(O_1)}{f^{-1}(O_2)} = \frac{e_1 V(x_1) L_1}{e_2 V(x_2) L_2} \tag{2.9}$$

where O_1 and O_2 are the pixel intensities of the corresponding points on two frames.

In OSPC, e_1 and e_2 are the exposure values of the two frames provided by ground truth data. The ICRF is represented by a second-degree polynomial as Eq. 2.3, offering a balance between computational efficiency and precision in determining its three coefficients [20]. Moreover, it's assumed that the corresponding points are situated on a Lambertian surface. Given that they represent the exact same point in the real-world scene, they should exhibit the same radiance, irrespective of the camera's viewing angle. This assumption allows for the cancellation of radiance. Consequently, this leads to Eq. 2.10, where k represents the exposure ratio.

$$\frac{f^{-1}(O_1)}{f^{-1}(O_2)} = \frac{e_1}{e_2} = k \tag{2.10}$$

By integrating the polynomial of the ICRF, Eq. 2.3, into Eq. 2.10, we deduce Eq. 2.11.

$$\frac{c_0 + c_1 O_1 + c_2 O_1^2}{c_0 + c_1 O_2 + c_2 O_2^2} = k \tag{2.11}$$

Eq. 2.11 can be further written in the vector form as Eq. 2.12.

$$\begin{bmatrix} 1-k & O_1 - kO_2 & O_1^2 - kO_2^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = 0 \tag{2.12}$$

The result of each pair of corresponding points are stored in matrix A, and the ICRF coefficients are denoted as vector c. Therefore, Eq. 2.12 can be simplified to Eq. 2.13.

$$Ac = 0 \tag{2.13}$$

This process ultimately results in a set of homogeneous equations that have an infinite solutions. However, all inverse camera response functions fall within the function space that adheres to the constraints $f^{-1}(0) = 0$ and $f^{-1}(1) = 1$. The addition of these equality constraints paves the way to a unique solution. The objective function can then be reformulated into the following function as shown in Eq. 2.14, with the constraint in Eq. 2.15.

$$\min_c \|Ac\|^2 \tag{2.14}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} c_i = 1 \qquad c_0 = 0 \tag{2.15}$$

Subsequently, reformulate Eq. 2.14 and Eq. 2.15 to a more compact form as Eq. 2.16 and Eq. 2.17 respectively.

$$\min_{c} \quad c^T A^T A c \tag{2.16}$$

$$\text{s.t.} \quad Pc = q \tag{2.17}$$

In this context, P is a matrix composed of two rows. The first row is entirely filled with ones. In the second row, only the first element is one, while the rest of the elements are zeros. On the other hand, q is a vector that contains two elements, with the first element being one and the second element being zero. In this way, the objective function can be rearranged to Eq. 2.18 with LM damping λ.

$$\min_{c,\lambda} c^T A^T A c + \lambda^T (Pc - q) \tag{2.18}$$

In order to address this optimization problem, the Lagrangian function Eq. 2.18 is minimized. By computing the second derivative with respect to c, a positive semi-definite matrix $A^T A$ is obtained. This affirms the convex nature of this formulation, implying the existence of a global minimum for the problem. This minimum can be accessed by setting the gradients, with respect to c and λ, to zero. The final expression can be represented as Eq. 2.19, which can be conveniently solved using a pseudo inverse.

$$\begin{bmatrix} 2(A^T A) & P^T \\ P & 0 \end{bmatrix} \begin{bmatrix} c \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ q \end{bmatrix} \tag{2.19}$$

The vignetting estimation is performed in the second step of the analysis. Again, corresponding points are identified in two frames using SIFT [49]. However, it is crucial that the radial distance is sufficiently large to ensure an accurate estimation of the vignette. To satisfy this requirement, any pair of points with a displacement under the radial distance threshold of $t = 0.3$ is not processed. Radius map, which consists of the distance for every pixel position to the image center, is calculated beforehand to speed up the process.

The methodology for determining the vignette parameters mirrors the process used in the previous step for developing the ICRF. The process commences by dividing the irradiance measured at corresponding points on a pair of images. As derived from the photometric formation function Eq. 2.6, the division of two corresponding points is represented as Eq. 2.9. Given that they are the same point on the Lambertian surface, they should have the same amount of radiance, i.e., $L_1 = L_2$. Consequently, this can be expressed in the following equation Eq. 2.20.

$$\frac{f^{-1}(O_1)}{f^{-1}(O_2)} = \frac{e_1 V(x_1)}{e_2 V(x_2)} = k \frac{V(x_1)}{V(x_2)} \tag{2.20}$$

By substituting the vignette model Eq. 2.4 into Eq. 2.20 and rearranging the terms, we can derive the following equation Eq. 2.21. $R(x)$ denotes the radius distance for each feature location x.

$$\frac{1 + R(x_1)^2 v_1 + R(x_1)^4 v_2 + R(x_1)^6 v_3}{1 + R(x_2)^2 v_1 + R(x_2)^4 v_2 + R(x_2)^6 v_3} = \frac{1}{k} \frac{f^{-1}(O_1)}{f^{-1}(O_2)} = \kappa \tag{2.21}$$

Furthermore, Eq. 2.21 can be further simplified into the following vector form.

$$\begin{bmatrix} R(x_1)^2 - \kappa R(x_2)^2 & R(x_1)^4 - \kappa R(x_2)^4 & R(x_1)^6 - \kappa R(x_2)^6 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \kappa - 1 \tag{2.22}$$

Revisiting the process, it's important to note that in OSPC method, the ground truth exposure are used for computing the exposure ratio. Moreover, by incorporating the ICRF coefficients from the previous step, the only unknown variables in Eq. 2.22 become the vignetting coefficients. This simplifies the optimization problem to a linear form, which can be conveniently solved using a least squares approach. Eq. 2.22 can be further written in the matrix from:

$$Ax = b \tag{2.23}$$

The least squares solution to this system is given by solving the normal equations:

$$A^T A x = A^T b \tag{2.24}$$

This can be solved for x to get the least squares solution:

$$x = (A^T A)^{-1} A^T b \tag{2.25}$$

Upon obtaining the estimated results for both the ICRF and vignette, the last step involves calculating the exposure ratio. The goal is to retrieve the exposure ratio for the entire video sequence, encompassing every frame. To achieve this, consecutive frames are used as a pair for computation. The exposure ratio is computed by performing straightforward algebraic manipulations based on Eq. 2.9, which can be expressed in the following formulation.

$$\frac{f^{-1}(O_1) V(x_2) L_2}{f^{-1}(O_2) V(x_1) L_1} = \frac{e_1}{e_2} = k \tag{2.26}$$

Again, each corresponding pair of pixels represents the same object point in the real-world scene. Therefore, given that they lie on Lambertian surface, their radiance should

be the same. Hence, the radiance $L_1 = L_2$ can be cancelled in Eq. 2.26. Consequently, the exposure ratio can be effortlessly computed using the subsequent formula. Here, N represents the total number of corresponding pixels on the patch for each image pair. O denotes the pixel intensity for each pixel, and x is the position of the pixel.

$$k = \frac{1}{N} \sum_{i=1}^{N} \frac{f^{-1}(O_1)V(x_2)}{f^{-1}(O_2)V(x_1)} \tag{2.27}$$

2.1.3 Related Work

Early research concentrated on the techniques of image stitching and mosaicking, which necessitate the derivation of essential calibration parameters from a limited collection of overlapping images [18, 37]. Some research may require manual control exposure of the camera [7] or need a chart [5] to help calibration. For instance, Pilet *et al.* [59] presents an unique approach for recovering geometric and photometric calibration parameters of multiple cameras whose fields of view may not all overlap. The technique requires waving an arbitrary textured planar pattern in front of the cameras and neither vignette nor exposure is considered. Besides, to estimate vignette, the background of the image require a uniform colored surface, such as a white wall [34]. In the study conducted by Alexandrov *et al.* [2], a white paper or sheet is utilized as a calibration tool to rectify the vignetting effect. These are not the common instances when taking video in a real world scene, for example most datasets lack photometric calibration and do not grant access to the camera. On the other hand, the early research can only estimate for one of these photometric parameters. In [42], Lin *et al.* developed a technique to deduce the camera response function. This method is grounded on the correlation between the inverse response and the non-linear distribution of observed color intensities near the edges within a single RGB image. Lin *et al.* [43] proposed a method to calculate the response function, without consider vignette and exposure, by analyzing the variations using skewed histogram distribution in light intensity near edges in one grayscale image. In [7], Debevec *et al.* utilized a non-parametric model to build a radiance map of the scene for estimating the ICRF up to a scale. Krawczyk *et al.* [38] introduced a calibration method for the CRF specifically designed for high dynamic range cameras. This paper elaborates on the utilization of intricate camera response curves and the choice of suitable calibration targets. However, the setup is notably complex and does not take vignetting into account. By calculating CRF in preprocessing step, Zheng *et al.* [75] developed an approach using single image for vignetting function estimation by pinpointing areas within the scene that exhibit the same radiance, and then they assess how this brightness diminishes as it approaches the edges of the image. Since only a single

image is utilized, estimating the exposure time is not possible [3]. Besides, for both [42] and [43], EMoR is employed for recovering better ICRF results.

Litvinov *et al.* [45, 44] propose a linear optimization method that simultaneously addresses the response and vignetting functions using an inverse logarithmic approach. Nevertheless, this method requires the manual introduction of a model prior to exclude trivial solutions and implement model-based assumptions. Mitsunaga *et al.* [53] solve ICRF and exposure ratio using a high-order polynomial model. This process began with initial approximations of the ratios, supplemented by measurements of pixel brightness. However, this method is applied on the static scene, not suitable for dynamic video. Besides, the known exposure value is required for initialization. Sun *et al.* [65] applied photometric calibration in their approach to target detection. They construct the CRF using an N-degree polynomial model and consider a compensation function for the vignetting effect. The exposure ratio is calculated in each iteration of their algorithm. However, the exposure time of the initial frame is assumed to be known and is used as the ground truth. Kim *et al.* [36] jointly optimize the short tracker with radiometric calibration, yet the vignette is ignored and not recovered due to the small displacement across consecutive frames. Grundmann *et al.* [21] adopted a long-term tracker, but again assuming that the tracks are not affected by vignetting. Mo *et al.* [55] managed to estimate ambiguity-free CRF utilizing the linear edge color blending constraint, but vignette is again neglected. Goldman *et al.* [18] use the Gauss-Newton algorithm to jointly optimize the response function, vignetting, exposure ratios, and scene point radiances. Nonetheless, their focus on panorama stitching, which relies on many precise matches from few images, makes their method unsuitable for auto-exposure videos.

Bergmann *et al.* [3] proposed a significant advancement that resolving the least squares problem associated with feature tracking and all photometric parameters simultaneously, and their proposed approach is able to run online. As mentioned in the previous section, the joint optimization cannot avoid the exponential ambiguity problem, and a point from the CRF ground truth information is utilized. Furthermore, to tackle the issue of ambiguity, Grossberg *et al.* [20] determined a specific value for the response function at a designated image intensity value, which was obtained from ground truth data. Kim *et al.* [37] proposed a comprehensive radiometric calibration algorithm that robustly estimates the radiometric response function, exposures, and vignetting. This approach can be seemed as a semi-sequential method, which decouples the vignette from CRF and exposure optimization. Nevertheless, it requires only a few images for panorama stitching, and the runtime of the algorithm is not considered. Furthermore, they cannot get away from ambiguity issue during the solving jointly for both the CRF and the exposure. Quenzel *et al.* [60] introduced a real-time online technique for calibrating vignetting and response,

along with an additional exposure estimation for global-shutter color cameras. However, it's important to note that their experimental setup utilized a synchronized stereo camera system, as opposed to a single monocular camera. Haidar *et al.* [23] adopted the sequential optimization method to address the ambiguity problem that encounter in joint method, but they still use metadata of exposure values for initialization.

Worth noticing, photometric calibration is also used in the field of astronomy. Nonetheless, the focus of photometric calibration differs from the field of computer vision. In astronomy, photometric calibration is primarily concerned with converting the arbitrary units in which CCD images are recorded into standard, reproducible units. This process is crucial for accurately measuring the brightness of celestial objects and comparing these measurements across different observations or telescopes. These factors can include the redshift of distant objects, which can shift the observed colors, and the characteristics of the telescope and instruments themselves, which can introduce systematic biases [33]. By carefully calibrating the photometry, researchers can ensure that their measurements are accurate and comparable across different observations and instruments. This is essential for many areas of astronomical research, from studying the properties of individual stars to measuring the large-scale structure of the universe [58]. Huang *et al.* [29] discuss various modern photometric calibration methods. They elucidate the fundamental principles of photometric calibration, which encompass the general principles of photometric calibration, which involve defining transformations and applying them to convert instrumental magnitudes into calibrated magnitudes.

On the other hand, photometric calibration and uneven illumination correction, despite their apparent similarities in handling illumination and color, are distinct processes. Uneven illumination correction employs shading or lighting models, such as ATF [1], to compute the light intensity reflected from a specific surface point. This correction method includes a penalty term that manages intensity distribution and transitions between predefined uniformly and non-uniformly illuminated sections of the input scale image [8]. However, it does not take into account the CRF, vignette, and exposure, which distinguishes it from photometric calibration.

2.2 Feature Tracker

KLT tracker was frist proposed by Lucas and Kanade [50]. They developed a local search method using gradients, which is weighted by an approximation of the second derivative of the image. This method was primarily designed to address the typically high cost of

traditional image registration techniques. The KLT tracker uses spatial intensity information to guide the search for the best match. It is faster than traditional techniques because it checks for fewer potential matches. Bergmann *et al.* [3] introduced a method known as the gain robust KLT tracker, designed for online photometric calibration. This method is essentially a pyramidal implementation of a standard KLT tracker. Given the significant exposure variations between frames, both the tracking updates and the gain ratio across frames are simultaneously optimized to derive the photometric parameters.

Persistent Independent Particles (PIPs) [26] method provides a reliable way to track feature points in videos, even when they are temporarily obscured. It is a long-term tracking technique that revisits the concept of "particle video" [62], which is an optical flow based method. This means that it is built on brightness consistency assumption. In this method, each pixel is assigned a trajectory that predicts its location in several future frames. This makes it possible to track the movement of particles over time, even when they are not visible in every frame. To train the PIPs method, the researchers used long-range point trajectories from existing optical flow data. They enhanced these data with synthetic multiframe occlusions, creating a more challenging training environment that better reflects real-world conditions. Compared to the gain robust KLT tracker, which tracks frame by frame, PIPs tracks eight frames in each batch and iterates through the entire video sequence.

Tracking Any Point with per-frame Initialization and temporal Refinement (TAPIR) [9] is a dual-phase algorithm specifically engineered for tracking points in videos. The initial phase independently locates an appropriate match for the query point in each frame. The subsequent phase refines this by updating both the trajectory and query features, taking local correlations into account. TAPIR is an extension of the PIPs method, and its training process also utilizes long-range amodal point trajectories from existing optical flow data, artificially augmented with multi-frame occlusions. This equips TAPIR with a strong capability to handle occlusions and track points over long duration. In addition, TAPIR exhibits improved resilience when dealing with occlusions. TAPIR is an excellent long-term tracker that is not limited to eight frames in a batch, but it requires significant memory during computation. We have set the batch size to ten frames to achieve a balance between time and memory usage.

Multi-Flow dense Tracker (MFT) [56] is a novel tracking method that maintains dense pixel-level tracking over extended periods. The method is unique in that it exploits optical flows estimated not only between consecutive frames but also between pairs of frames at logarithmically spaced intervals. This allows for more robust tracking even when the frame-to-frame motion is large or complex. The method addresses occlusions by selecting the most reliable sequence of flows based on estimates of geometric accuracy and the probability of

occlusion. These estimates are provided by a pre-trained Convolutional Neural Network (CNN), which assesses the likelihood of occlusion for each pixel. This integration with a CNN allows MFT to make informed decisions about which optical flows are most reliable for tracking through occlusions and other challenging conditions. By estimating flow with respect to a reference frame, MFT reduces drift, which is a common issue in long-term tracking where small errors accumulate over time. Although MFT can track points frame by frame and is not as computationally expensive as TAPIR, we still process it in batches of ten frames for convenience.

2.3 Image and Video Quality Assessment

We employ IQA and VQA to evaluate the reconstructed video. These assessments aim to evaluate the perceptual quality of visual content from a human perspective. IQA works by using computational models to measure the perceptual quality of images. Depending on the reference image's availability, these models fall into three categories: Full-reference Image Quality Assessment (FR-IQA), No-reference Image Quality Assessment (NR-IQA), and Reduced-reference Image Quality Assessment (RR-IQA) [52].

FR-IQA requires an unaltered reference image for comparison with the distorted image. The full-reference matrices employed include Learned Perceptual Image Patch Similarity (LPIPS), Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR). LPIPS [73] leverages deep neural networks to evaluate perceptual differences between two images through their high-level features, with lower scores indicating superior quality. SSIM [69] assesses changes in structural information, luminance, and contrast, where higher scores reflect enhanced quality. PSNR [15] calculates the ratio of the maximum possible signal power to the corrupting noise power, with higher values representing improved image quality.

NR-IQA evaluates image quality based on features that correlate with perceived quality, without requiring a reference image. Natural Image Quality Evaluator (NIQE) [54] utilizes natural scene statistics and a set of quality-aware features to measure image quality; lower scores indicate higher quality. Multi-task Attention-based No-reference Image Quality Assessment (MANIQA) [71] employs a multi-task learning framework with attention mechanisms to estimate image quality scores; lower scores denote better quality. CLIP-based Image Quality Assessment (CLIPIQA) [68] uses the ontrastive Language-Image Pretraining (CLIP) model to evaluate image quality by understanding the content and context, with lower scores suggesting higher quality. Convolutional Neural Network-based Image Quality Assessment (CNNIQA) [32] relies on a pre-trained CNN to predict image quality

from learned image data features; lower scores imply higher quality. eural Image Assessment (NIMA) [67] rates images in technical and aesthetic quality using a deep CNN trained on human-annotated datasets; higher scores indicate better quality. RR-IQA requires only partial information about the reference image. We utilize the PyTorch Toolbox for Image Quality Assessment [6] to evaluate the results with the aforementioned matrices by averaging the scores of each frame.

VQA extends the principles of IQA to video, offering a holistic evaluation of video quality [48]. It considers smoothness, encompassing frame rate and motion continuity for fluid playback, and spatial quality, which examines clarity and detail in each frame. VQA tackles temporal artifacts like ghosting and blurring using advanced models and temporal pooling techniques. It also addresses compression issues like blocking and banding by employing visual saliency models to detect and reduce these artifacts. In addition, it evaluates the color fidelity for accurate representation and dynamic range to ensure visibility under various lighting conditions. These factors are crucial for a seamless multimedia experience. The goal of VQA is to develop metrics that resonate with human subjective perception of video quality [16]. For this purpose, we employ Video Multi-Method Assessment Fusion (VMAF) [61] to assess the quality of reconstructed videos.

VMAF [61], a full-reference video quality metric developed by Netflix, assesses subjective video quality by comparing a reference video with a distorted sequence [41]. It integrates various basic video quality metrics using machine learning methods. One of the key features of VMAF is the Visual Information Fidelity (VIF) [63], which evaluates information fidelity loss across spatial scales. The VIF score ranges from 0 to 1, with 1 indicating perfect information fidelity. Another important feature is the Detail Loss Metric (DLM) [39], which quantifies detail loss and distracting impairments. It is named as ADM2 score in VMAF, which measures the structural similarity between the reference and distorted videos. This score ranges from 0 to 1, with 1 indicating perfect structural similarity. The Motion2 score is a measure of the temporal difference between adjacent frames in a video. It is calculated by finding the average absolute pixel difference for the luminance component, and its scale varies from 0 to 20. These features are combined through a Support Vector Machine (SVM)-based regression to produce a per-frame score, with 100 representing perfect quality. The final output score of the VMAF model is computed by incorporating a 'Motion' feature along with the per-frame score. In general, the higher the VMAF score, the higher the perceived quality of the video [40].

2.4 Datasets

The following datasets are utilized in our experiments. First of all, the TUM MONO Dataset [14], provided by the Computer Vision Group at the Technical University of Munich, consists of 50 sequences of real-world video footage with ground truth timestamp. We use 26 sequences that were captured using a standard camera, not a fisheye lens. These sequences capture a variety of scenes, ranging from narrow indoor corridors to expansive outdoor environments. Each sequence features substantial exploratory camera motion and begins and ends at the same location. This setup facilitates the evaluation of tracking accuracy by assessing cumulative drift from start to end, eliminating the need for ground-truth verification across the entire sequence. Notably, all sequences within this dataset have been subjected to photometric calibration, thereby affording exposure times for each frame as recorded by the sensor, CRF, and vignetting.

The EuRoC MAV Dataset [4], created by the Autonomous Systems Lab at ETH Zürich, is an essential benchmark for evaluating visual-inertial localization algorithms with real aerial flight data. It includes 11 stereo-inertial sequences, captured from a quadrocopter in flight across three distinct indoor settings. These sequences are complemented by synchronized Inertial Measurement Unit (IMU) readings and meticulously accurate ground truth data pertaining to both motion and structure, all of which have been gathered from Micro Aerial Vehicles (MAVs). The dataset's primary aim is to support the development and assessment of visual-inertial localization algorithms, offering high-quality motion capture and laser tracker data to ensure accurate 6D pose and 3D position information. Although comprehensive for visual-inertial system benchmarking, it does not specify the inclusion of photometric parameters such as ground truth exposure values.

The ICL-NUIM Dataset [25], contributed by Imperial College London, is a benchmarking resource for RGB-D, VO, and SLAM algorithms. It features two distinct scenes: a living room and an office room. The living room scene is equipped with 3D surface ground truth, depth maps, and camera poses, facilitating the benchmarking of camera trajectories and reconstruction capabilities. Conversely, the office room scene solely provides trajectory data and lacks an explicit 3D model. Designed specifically for indoor environments, the ICL-NUIM dataset offers detailed ground truth for 3D surface reconstruction, a feature that is not consistently available in the TUM MONO Dataset [14]. Notably, the dataset does not include ground truth information for photometric parameters, which presents a significant challenge in our context.

The KITTI Odometry Dataset [17] is a comprehensive resource comprising 22 stereo sequences captured from a moving vehicle. The dataset, which primarily focuses on forward-motion patterns and street scenes, is particularly useful for the development and evaluation

of odometry and SLAM algorithms. It includes pre-rectified images and GPS-INS ground truth poses for all frames, but does not provide raw sensor measurements or calibration datasets. The dataset is divided into two sets: 11 sequences with ground truth trajectories for training, and 11 sequences without ground truth for evaluation. The benchmark allows for results from various methods, including monocular or stereo visual odometry, laser-based SLAM, or algorithms that combine visual and LIDAR information, provided they are fully automatic and use the same parameter set for all sequences. This dataset does not contain any photometric information.

2.5 Summary

Photometric calibration is a fundamental task in computer vision that has evolved significantly, branching into two distinct categories: joint methods and sequential methods. The most well-known joint method is OPC [3], which has been widely adopted in many works. This approach jointly optimizes all photometric parameters but suffers from exponential ambiguity. To address this, ground truth data is used to solve the optimization problem. Conversely, the sequential method, such as OSPC [23], independently optimizes photometric parameters, thereby avoiding the exponential ambiguity issue. However, the initialization of the exposure ratio in this method still relies on ground truth data. Additionally, existing sequential methods only estimate exposure values up to a scale, meaning they can only retrieve the exposure ratio. While this is sufficient for applications like DSO [11] and SLAM [10], it is inadequate for reconstructing photometrically calibrated frames without exposure values for each frame. Despite the progress made, the field faces challenges such as those mentioned. Future research aims to integrate a more robust tracker and eliminate the need for photometric ground truth. The research will be comprehensively conducted on several datasets and will utilize IQA and VQA to evaluate the calibrated frame.

Chapter 3

Methodology

3.1 Introduction

This research is structured around two pivotal experiments with the aim of enhancing the photometric calibration process. The first objective is to assess the robustness and performance of feature trackers under diverse photometric conditions. We argue that the gain robust KLT tracker, which is used in OPC [3], lacks sufficient reliability even though the pyramid method is employed [24]. Subsequently, our primary goal is to investigate a simple method that eliminates dependence on photometric ground truth information. The following section will break down the details of the methodology algorithm. Our evaluation will commence with the TUM MONO Dataset [14], in which the photometric ground truth information is provided. Following this, we will assess a dataset devoid of ground truth information, as well as a video that we have independently created.

3.2 Feature Tracker Evaluation

The input for the photometric calibration process are the corresponding points in a video sequence. Therefore, finding robust corresponding points is an important task. We observe that the robust KLT tracker used in OPC [3] is not sufficient robust to occlusion [24]. Accordingly, we aim to investigate whether long-term trackers, such as PIPs [26], TAPIR [9], and MFT [56], could offer better performance. We integrate them with the OPC [3] and evaluated the results on the TUM MONO Dataset [14] sequence 30. The goal is to gain

insights into the performance of different feature trackers and identify potential improvements.

3.2.1 Experiment Setup



Figure 3.1: This figure illustrates feature points extracted using the Shi-Tomasi corner detector [64]. The distribution map ensures that feature points are evenly distributed throughout the frame, and points that surpass the frame boundary are removed.

The process begins by resizing all frames to a resolution of 256×256 pixels to accelerate the subsequent processes. The first step is to find initial feature points on the first frame of the video sequence. The distribution map, which divides the image into a 10×10 grid, ensures a uniform spread of feature points across the frame for feature extraction, as shown in Fig. 3.1. This prevents the concentration of feature points in a small area. The Shi-Tomasi corner detector [64] is used to identify the initial feature points within these cells. The number of active points is set at $N = 100$, implying that one point is found within

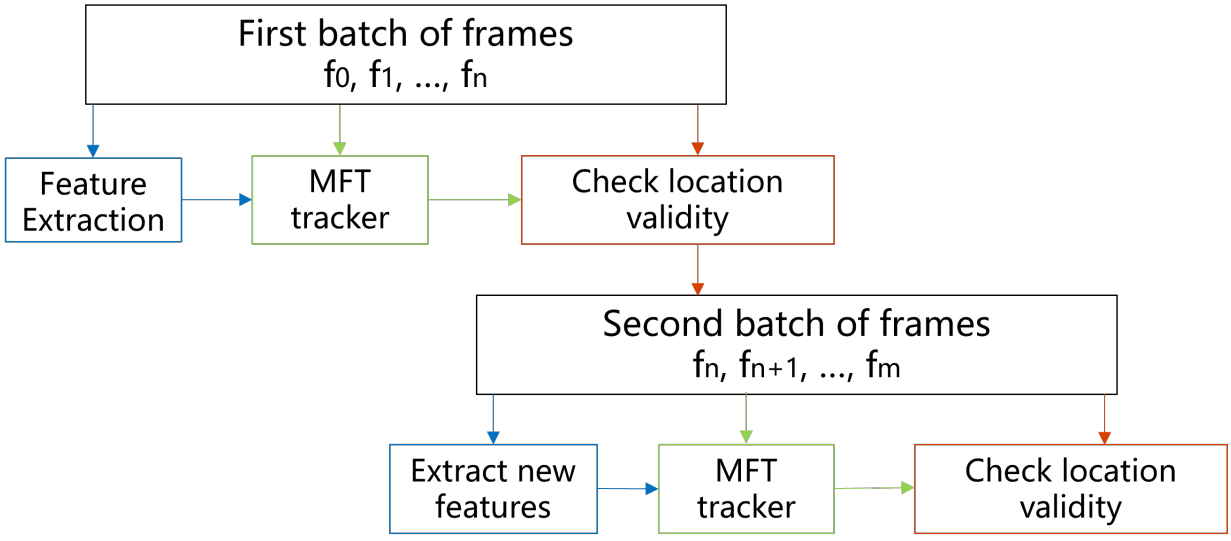


Figure 3.2: This flowchart illustrates the process of tracking active points across video frames. It begins by extracting 100 active points in the first frame, then retains trajectories within frame boundaries, and uses the points from the final frame as starting points for the next series. If the number of points in the series is insufficient, new ones are added. This process continues until the end of the video.

each cell. These points are then introduced into the long-term trackers and monitored throughout the sequence.

Each batch consists of 8 frames for PIPs and 10 frames for both TAPIR and MFT. At the end of each batch, the trajectories are examined by a validation check. A padding, equivalent to 5 percent of the image size, is added to the image boundary. If any points extend beyond this padding, the corresponding trajectory is removed. When a new batch arrives, new features are extracted from the first frame, which is also the last frame from the previous batch, using the same map. The number of new features to extract is determined by subtracting the count of existing features from the maximum number of active features, which makes sure the number of current features around 100. If there exists a sufficient number of active points, then the extraction of new features is not necessary. The new features are extracted from a region of interest within the frame that corresponds to the current cell. If there are any empty cells, the extraction process iterates over each of them. The list of new features is then compiled and combined with the existing features. Both sets of features are then fed to the long-term tracker. This procedure is repeated

iteratively to track features throughout the entire video sequence, ensuring a consistent and efficient tracking process. To integrate the trajectories with the OPC [3], it's necessary to ensure that consecutive frames have the same number of tracked feature points. The implementation is demonstrated in Fig. 3.2.

3.2.2 Experiment Results

We conducted a comparative analysis of the photometric parameter results generated by four different trackers. These include the gain robust KLT tracker, which is inherent in the OPC [3] algorithm, and three other long-term trackers: PIPs, TAPIR, and MFT.

Fig. 3.3 presents the estimated results of three photometric parameters using four trackers. Root Mean Square Error (RMSE) between the ground truth and estimated results is calculated. From left to right in each column, they represent the results obtained using the gain robust KLT tracker [3], PIPs [26], TAPIR [9], and MFT [56]. In these figures, the green line represents the ground truth, while the red line represents the estimated results obtained employing each tracker. The first row in Fig. 3.3 illustrates the ICRF, where the horizontal axis represents the image brightness value and the vertical axis denotes the sensor irradiance. We can observe that the comprehensive analysis indicates that the outcomes are generally favorable. However, MFT [56] provides the most accurate results in terms of aligning with the ground truth. This could be attributed to MFT's [56] advanced tracking algorithms which may be more effective at handling complex scenarios compared to the other two trackers.

The second row in Fig. 3.3 depicts the outcome of the the vignetting correction. The horizontal axis denotes the radius, while the vertical axis represents the vignetting factor. Here again, MFT [56] outperforms all trackers and is significantly better than the robust KLT tracker [3]. Lastly, the third row in Fig. 3.3 presents the outcomes of exposure time estimation, with the horizontal axis indicating the frame index and the vertical axis representing the exposure time. Note that the figure only contains the exposure of key frames. Given the inherent design of long-term trackers to monitor objects over extended durations and across diverse conditions, they accumulate a greater number of tracking points. Consequently, the use of long-term trackers results in a higher density of optimization blocks, leading to more apparent oscillations or fluctuations in the results. Despite having comparable results, MFT exhibits a slight advantage over all other trackers, as evidenced by the RMSE results shown in Fig. 3.3.

Fig. 3.4 displays the qualitatively results of photometrically calibrated images from the video sequence. From left to right, the first column shows the input sequence, the second

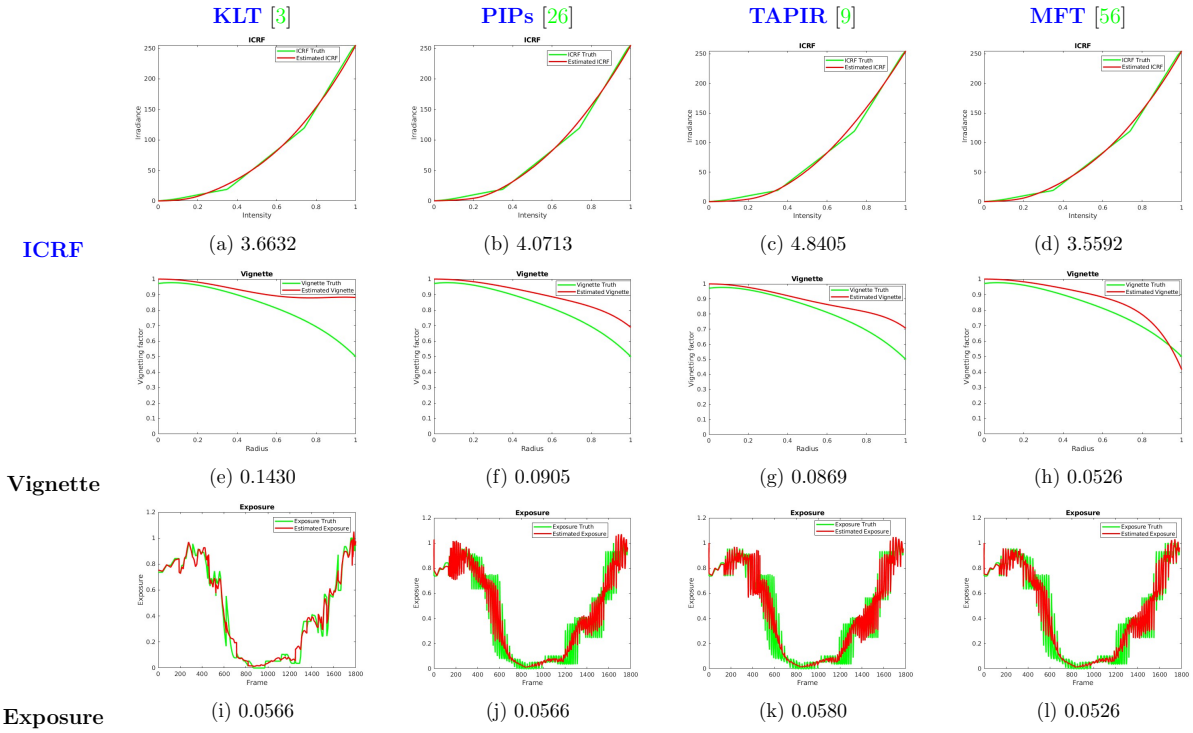


Figure 3.3: The figure presents the outcomes of three photometric parameters derived from various trackers. In this context, KLT represents the robust KLT tracker [3]. We computed the RMSE, i.e., numbers next to index, to measure the discrepancy between the estimated results and the ground truth. The estimated ICRF and exposure from all trackers exhibit close proximity. In terms of vignette, the long-term trackers demonstrate superior performance over the KLT tracker [3]. MFT [56] performs the best in all three parameters.

column presents the gain robust KLT tracker [3], the third column follows with PIPs [26], the fourth with TAPIR [9], and lastly, MFT [56]. The first row in Fig. 3.4 represents the initial frame of the video sequence. While the second row illustrates a brighter frame from the middle of the sequence. The last row depicts the final frame, returning to the original position as the first frame. This design is intended for loop closure, but it also allows us to compare the photometrically calibrated results. While limited information can be derived from the first frame, using MFT [56] provides a clearer image in the brighter region in the second row, particularly noticeable in the top left corner. In the last frame, the gain robust KLT tracker [3] underperforms, whereas using the long-term trackers yields slightly improved results compared to it, as they generate a brighter photometrically calibrated

image.

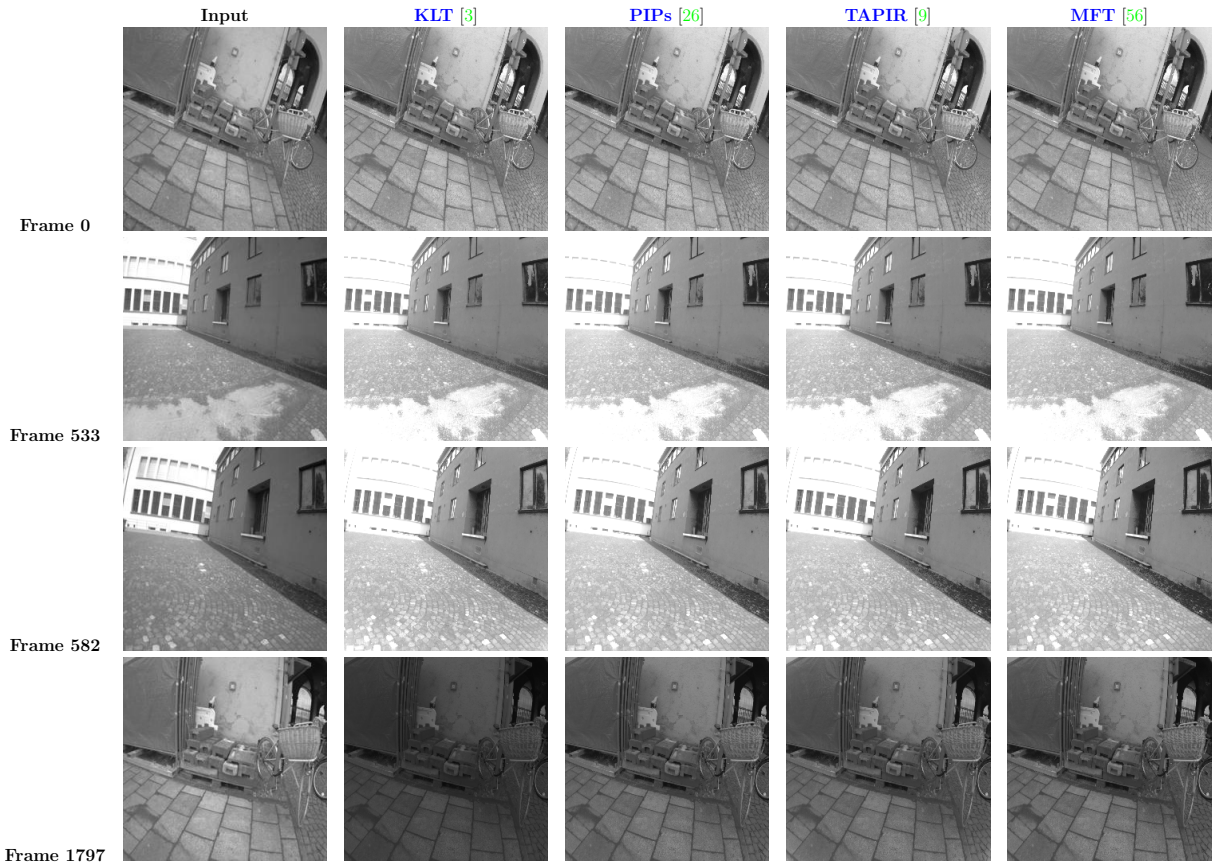


Figure 3.4: The frames, reconstructed by estimated photometric parameters, are obtained using various trackers integrated with the OPC [3] method. Qualitative analysis shows that MFT [56] and PIPs [26] yield superior results, especially evident in Frame 582, where the bright area in the top left corner stands out. This observation is supported by IQA results.

The estimated results of the three photometric parameters are used to reconstruct the photometrically calibrated image. IQA matrices are then employed to calculate a score for each photometrically calibrated frame of the video. These scores are subsequently summed and averaged over the total number of frames. The results of the IQA matrices are presented in Table 3.1. An arrow next to the matrix name indicates whether higher or lower scores are preferable. It is evident that the IQA scores of the long-term trackers significantly surpass those of the robust KLT tracker [3]. Moreover, MFT [56] yields

superior results in the majority of the matrices.

The original video has the highest score in NIQE. NIQE, designed to assess image quality based on natural scene statistics, favors images closer to its definition of 'natural.' This could explain why the original video outperforms in NIQE scoring. Conversely, metrics like LPIPS, SSIM, PSNR, MANIQA, CLIPIQA, and NIMA, which might be more sensitive to brightness changes introduced by video trackers, evaluate different aspects of image quality. These aspects, such as structural similarity, contrast, sharpness, or noise level, may not align with NIQE's focus on 'naturalness,' leading to varied results.

Subsequently, videos are constructed from the photometrically calibrated frames of each tracker, and their performance is evaluated using the VMAF metric. As demonstrated in Table 3.2, the MFT [56] tracker generally exhibits superior performance across most metrics when compared to the other trackers.

Table 3.1: The table presents the results of IQA metrics for different trackers. The calibrated sequence using MFT [56] performs well in most matrices.

Score	Original	KLT [3]	PIPs [26]	TAPIR [9]	MFT [56]
LPIPS ↓	-	0.1268	**0.1019**	0.1062	0.1035
SSIM ↑	-	0.7784	**0.8388**	0.8355	0.8335
PSNR ↑	-	17.4138	20.5126	20.0663	**20.9026**
NIQE ↓	**3.7394**	6.0860	6.1745	6.2092	6.241
MANIQA ↑	0.3252	0.4805	0.5105	0.5107	**0.515**
CLIPIQA ↑	0.2388	0.3777	0.4075	0.4042	**0.4152**
CNNIQA ↑	0.5585	0.6008	0.6408	0.6433	**0.6468**
NIMA ↑	3.7475	4.2757	4.3952	4.4097	**4.4288**

Table 3.2: The table presents the results of VQA metrics for different trackers. The calibrated video using MFT [56] has the highest VMAF score.

Score	KLT [3]	PIPs [26]	TAPIR [9]	MFT [56]
VIF ↑	0.3802	0.3924	0.3868	**0.3953**
Motion2 ↑	7.0572	7.0572	7.0572	7.0572
ADM2 ↑	0.8050	0.8980	0.8980	**0.9244**
VMAF ↑	45.0238	65.9138	63.3139	**75.0219**

3.2.3 Summary

Our study reveals that the utilization of long-term trackers generally surpasses the gain robust KLT tracker [3] in terms of both quantitative and qualitative results. Despite the minimal presence of occlusion issues in our case, the long-term trackers demonstrate superior performance in handling such scenarios. TAPIR [9], which is built upon the PIPs [26] model, produce comparable outcomes. MFT [56] distinguishes itself in the estimation of vignette and exposure. Interestingly, all four trackers yield similar results in ICRF. It is noteworthy that among all, TAPIR's [9] tracking process is the most time-consuming. On the other hand, MFT [56] excels in the recovery of frames in bright areas and maintains relatively good quality in the final frame. Consequently, we have opted to employ MFT [56] in subsequent experiments.

3.3 Photometric Calibration

We aim to investigate our proposed method, termed as New Photometric Calibration (NPC), which is an extension of the OSPC [23] method. NPC also optimizes three parameters independently in a sequence of steps. The advantage of this approach is that it productively bypasses the exponential ambiguity problem inherent in the joint method. Fig. 3.5 illustrates an algorithmic overview of our proposed approach, NPC, and Algorithm 1 presents its pseudo code. The main contributions and changes we have made include:

- Replacing SIFT [49] with MFT [56] to enhance robustness.
- Using pixel intensity to initialize the exposure ratio, thereby eliminating the need for ground truth exposure.
- Retrieving the exposure value for each frame, rather than computing it up to a scale.
- Reconstructing calibrated frames, which is not performed in OSPC [23].

Algorithm 1 New Photometric Calibration

features ← CornerDetector(frames)
trajectories ← MFT(features)
exposureRatio ← ComputeExposureRatio(frames, trajectories)
icrfCoefficients ← OptimizeICRF(frames, trajectories, exposureRatio)
vignetteCoefficients ← OptimizeVignette(frames, trajectories, exposureRatio, icrfCoefficients)
exposureEstimation ← CalculateExposure(frames, trajectories, icrfCoefficients, vignetteCoefficients)
reconstructedFrame ← Reconstruction(frames, icrfCoefficients, vignetteCoefficients, exposureEstimation)

First of all, instead of using SIFT [49] to identify corresponding points between two frames, as done in OSPC [23], we employ Shi-Tomasi corner detector [64] for feature extraction and utilize MFT [56] to acquire trajectories across frames. MFT [56], identified from previous experiments, is employed in our method to ensure robust trajectories, thereby facilitating an efficient calibration process. As shown in Fig. 3.6, SIFT [49] often produces poor corresponding pairs of points, whereas MFT [56] generates more and robust corresponding points. In addition, the distribution map, as outlined in the preceding section, is utilized along with the same feature tracker to obtain new feature points. Furthermore, a 3×3 patch is applied to each pair of corresponding point to ensure robustness.

To eliminate the need of exposure ground truth data for selecting image pairs, we adopt a similar method as [53, 65] to utilize the image intensity to initialize the exposure ratio k in Eq. 3.1. Under the assumption that a CRF is determined by the camera sensor and remains consistent throughout the entire video. As well as with the small movement, the vignette can be ignored and the feature points are located on a Lambertian surface so that the reflected energy is independent to the viewing angle. Furthermore, the effect of depth is negligible. Given these conditions, exposure emerges as the only variable that can cause variations in pixel intensity values. This implies that the pair of frames does not need to be successive. This method provides a more direct and practical way to determine the exposure ratio without the need for ground truth exposure.

$$k = \frac{1}{N} \sum_{i=1}^{N} \frac{I_1(x_i)}{I_2(x_i')} \tag{3.1}$$

Where N is the number of corresponding points. I_1 and I_2 are the frame pair. x and

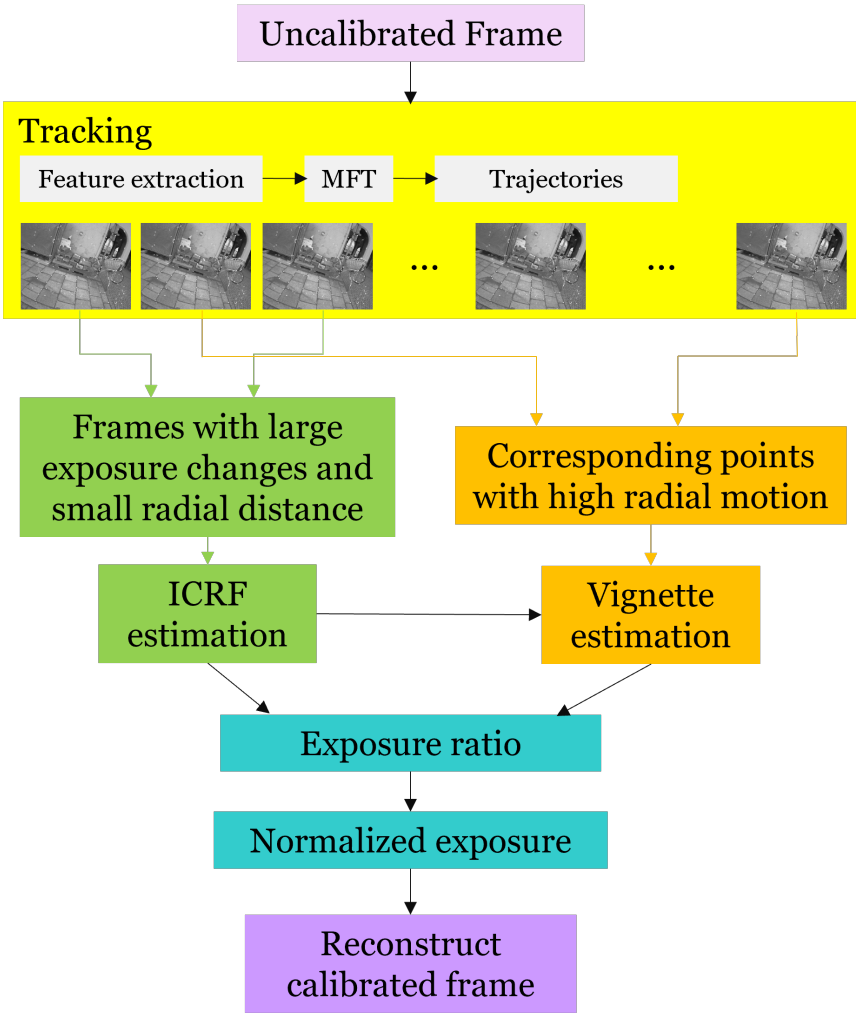
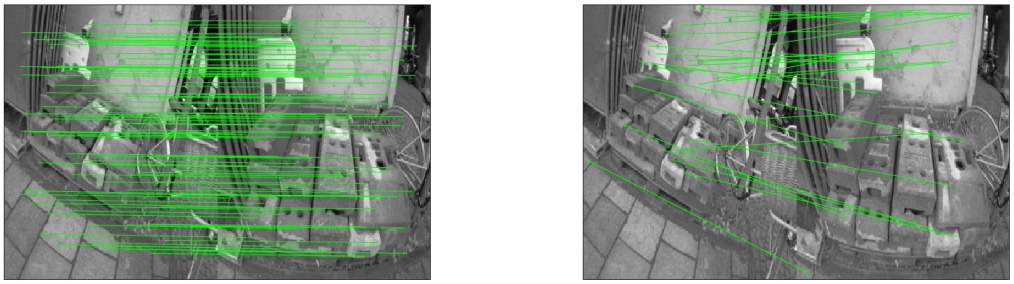


Figure 3.5: Flowchart illustrating the methodology proposed in this study and its key steps. The process begins with feature extraction and tracking by MFT [56] to acquire trajectories. Frames with large exposure changes are then used for ICRF estimation and vignette estimation. The exposure ratios are computed by estimates ICRF and vignette. To calculate normalized exposure for each frame, the ground truth data of the first frame is utilized. This leads to the final step of reconstructing calibrated frames.

x' denotes the location of corresponding points.

To retrieve the exposure value of each frame, we assume the exposure value on the



(a) Finding corresponding points using MFT [56].

(b) Finding corresponding points using SIFT [49].

Figure 3.6: A comparison of finding corresponding points across two frames using MFT [56] and SIFT [49] reveals that SIFT [49] produces fewer and less accurate corresponding points.

first frame is 1, then divide this by the exposure ratio k to determine the exposure for the second frame, and continue this process for the entire video. Subsequently, the exposure is normalized to fall into the range of $[0, 1]$. In OSPC [23], the exposure is only compute up to a scale. In the context of DSO [11] or SLAM [10], the exposure ratio alone is sufficient. However, to reconstruct images that are photometrically calibrated, it's necessary to have the exposure value for each individual frame. In other words, OSPC [23] method does not provide the calibrated frame.

Once obtaining the ICRF, vignette, and exposure, the reconstruction of the image can be accomplished. This is achieved by calibrating each pixel on the frame using the following formula (Eq. 3.2) to adjusts pixel values, based on the estimated results of the three photometric parameters. Here, p represents the pixel coordinates.

$$L(p) = \frac{f^{-1}(O(p))}{eV(p)} \tag{3.2}$$

The reconstructed frames are then further assembled to form a complete photometrically calibrated video.

Note that the pixel intensity is normalized to the range $[0, 1]$. Accordingly, three parameters are all normalized. The ICRF is configured to convert brightness from the range $[0, 1]$ to radiance within the range $[0, 255]$. The vignetting factor also falls within the range $[0, 1]$. And the exposure is normalized to ensure it falls within the range $[0, 1]$. When reconstruct the image, the minimum vignetting value is set to 0.3, as in practice, the optical

system nowadays is unlikely to reach such a lower value. As a result of these conversions, the computed radiance can become quite large. To facilitate visualization of the results, the gamma correction is applied. This process is crucial for maintaining the integrity of the data and ensuring accurate representation in the visual output.

3.4 Evaluation

The evaluation is based on the RMSE of three photometric parameters: the ICRF, vignette, and exposure. The RMSE is computed by comparing the estimated results with the ground truth. The ground truth information for the ICRF, vignette, and exposure is sourced from the TUM MONO Dataset [14]. The choice of RMSE as an evaluation metric is due to its effectiveness in quantifying the average magnitude of errors between estimated and actual observations, providing a comprehensive measure of accuracy.

Additionally, to assess the quality of the photometrically calibrated video, both VQA and IQA are employed, which include no-reference and full-reference metrics. IQA, in particular, is utilized to evaluate the quality of individual frames. The IQA score for each frame is computed and then averaged over the total number of frames to provide an overall quality assessment. The use of VQA and IQA allows for a thorough and detailed evaluation of the video quality.

In certain instances, some datasets do not provide photometric ground truth information, presenting a unique challenge in the evaluation process. Without photometric ground truth data, it is impossible to compute the RMSE. Therefore, the assessment of the results relies solely on IQA and VQA. Despite this limitation, these assessments still offer valuable insights into the quality of the photometric calibration for photometrically calibrated frames.

The photometrically reconstructed videos are fed into DSO [11], allowing for a visual representation of the 3D reconstruction scene. Additionally, the evaluation of the trajectory generated by DSO [11] is performed using the EVO framework [22]. EVO [22] offers the mean, deviation, and plotting for various trajectories, providing valuable insights. The output results can be compared by using the TUM MONO Dataset [14], which provides the necessary ground truth information for accurate assessment.

Chapter 4

Experimental Results

The results that follow are in two parts: analyzing datasets both with and without ground truth. For each experiment, we compare the evaluation of photometric parameters, image and video quality assessment, and DSO [11] results. Our study contrasts the proposed NPC method with the joint method (OPC [3]) and the sequential method (OSPC [23]). To evaluate the results, we compute the RMSE for three photometric parameters: ICRF, vignette, and exposure. Subsequently, IQA and VQA are used to evaluate the quality of the photometric calibrated frame. Additionally, we reconstruct the scene using DSO [11] and employ EVO [22] to evaluate and compare the trajectory output.

4.1 Analysis with Ground Truth Information

In this section, we evaluate 26 video sequences from the TUM MONO Dataset [14] taken with a non-fisheye camera. The ground truth information provided by the TUM MONO Dataset [14] are used to compute the RMSE for the estimated results. To assess the quality of the calibrated frames, we compare them with the original video sequences using NR-IQA, FR-IQA, and VQA. The original sequences, combined with the ground truth data, are used as references in the DSO [11] evaluation.

4.1.1 Photometric Parameters

Fig. 4.1 presents the RMSE plots of ICRF for all sequences and Table 4.1 summarizes the average and Standard Deviation (STD) of RMSE. Generally, OPC [3] outperforms

OSPC [23] and NPC, as shown in the figure, with OPC [3] consistently achieving lower RMSE values. However, in a few cases, such as sequence 16 and sequence 35, NPC achieves even better results than both OPC [3] and OSPC [23]. Fig. A.3 illustrates a specific instance where OPC [3] performs poorly, while OSPC [23] and NPC can still achieve good results. This is due to the model's limitation in retrieving only certain polynomial patterns, resulting in a lack of flexibility. It is also evident that the sequential approaches are more unstable. In some cases, our model achieves lower RMSE than OSPC [23], while in others it does not. High deviation in NPC results may be resulted from the selection of image pairs, which is based on pixel intensity. This rough method for exposure ratio initialization can lead to severe errors. In a later experiment (Table 4.5), when we used ground truth exposure values, we observed that the RMSE dropped significantly. As photometric calibration is quite sensitive, small changes can lead to significant differences in RMSE. Therefore, a good selection of image pairs and corresponding points is important.

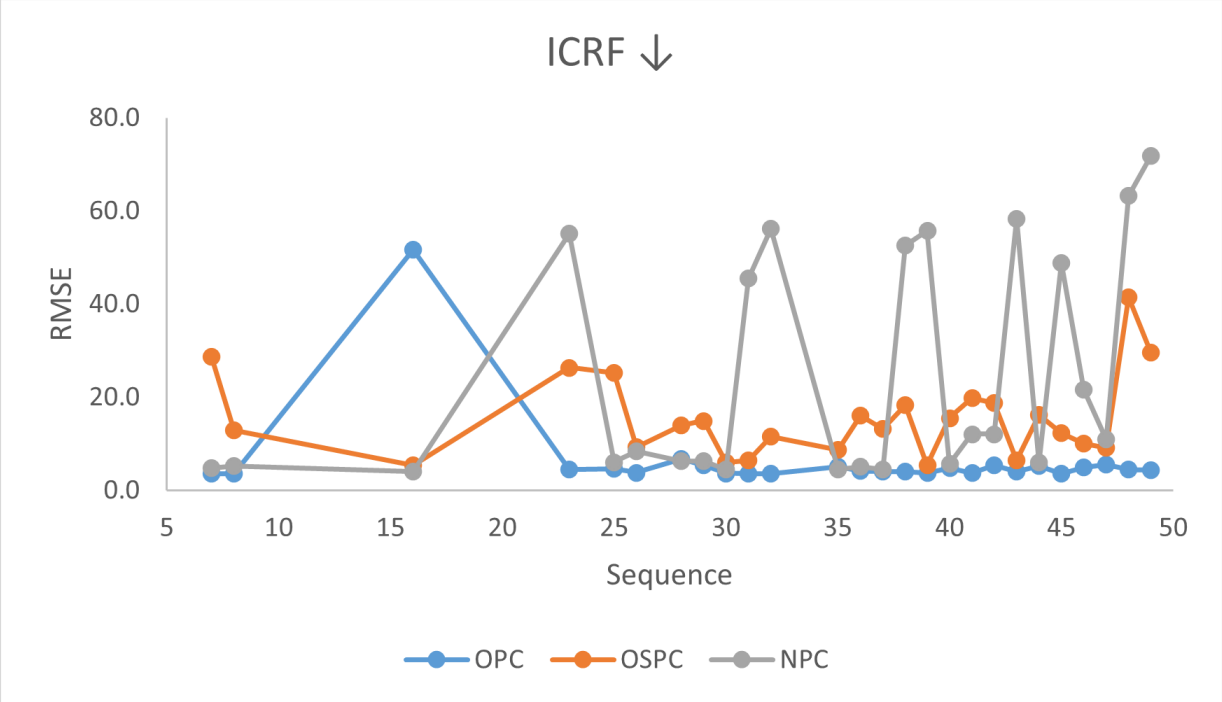


Figure 4.1: Comparative analysis of RMSE for ICRF across different sequences for OPC [3], OSPC [23], and NPC methods. In most sequences, OPC [3] achieves lower RMSE values than OSPC [23] and NPC. Although NPC appears to be unstable, it still outperforms in a few sequences.

Table 4.1: Average and STD of RMSE for ICRF using OPC [3], OSPC [23], and NPC methods. The high average and STD for NPC may be due to the poor selection of image pairs. This suggests that careful selection of image pairs is crucial for achieving more accurate results.

ICRF	OPC [3]	OSPC [23]	NPC
Average ↓	**6.2183**	12.4419	24.4300
STD	9.3032	**8.8380**	24.3724

Fig. 4.2 shows the RMSE plots of vignette for each sequence, and Table 4.2 displays the average and STD of RMSE. OPC [3] steadily outperforms OSPC [23] and NPC in most sequences, whereas sequential approaches exhibit more oscillation. During the experiment, we observed that slow or static movements, such as sequence 07, result in poor and challenging vignette estimations. This is because our method for estimating vignette requires significant radial movement, where the threshold is set to be $t = 0.3$. When the scene is static, it is difficult to find suitable image pairs and corresponding points. Additionally, when utilizing ground truth exposure data (Table 4.5), the RMSE values for the vignette further decrease. This provides clear evidence that using a rough method for initializing the exposure ratio negatively affects the optimization. For sequences 07, 37, and 47, the RMSE is particularly high because the curve of the estimated vignette goes beyond the range of $[0, 1]$. However, for exposure estimation and frame reconstruction, we constrain the vignetting factor to a maximum of 1 and a minimum of 0.3.

Table 4.2: Average and STD of RMSE for vignette using OPC [3], OSPC [23], and NPC methods. OPC [3] achieves the lowest average and STD of RMSE , demonstrating its robustness.

Vignette	OPC [3]	OSPC [23]	NPC
Average ↓	**0.0966**	0.1363	0.1576
STD	**0.0263**	0.0986	0.1423

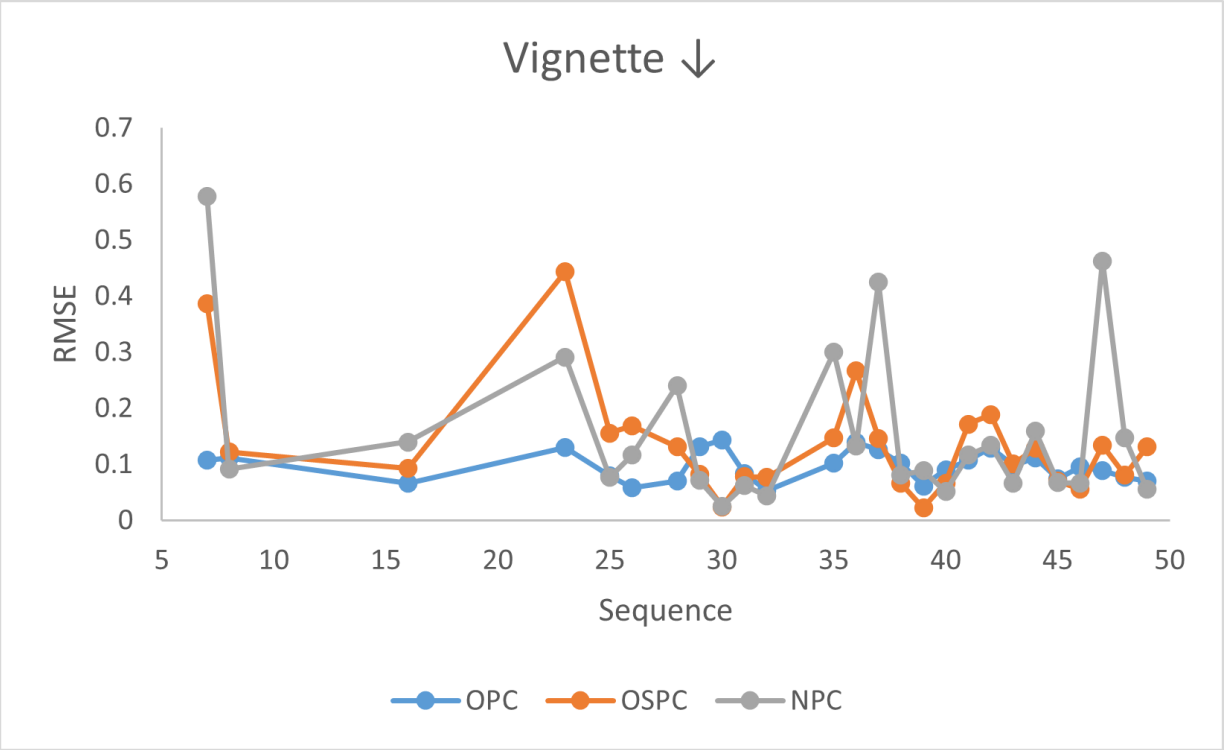


Figure 4.2: Comparative analysis of RMSE for vignette across different sequences for OPC [3], OSPC [23], and NPC methods. OPC [3] consistently delivers better performance compared to OSPC [23] and NPC in the majority of sequences.

Fig. 4.3 exhibits the RMSE plots for exposure, while Table 4.3 displays the average and STD of RMSE. For exposure estimation, OPC [3] reliably achieves the lowest RMSE across all sequences, likely due to its intricate optimization method and the use of ground truth data. Conversely, even without ground truth information, our method can still produce results comparable to, and sometimes better than, OSPC [23]. We observed that when exposure changes are minimal, such as in sequences 26 (Fig. A.6), 35 (Fig. A.12), and 36 (Fig. A.13), where the exposure ratio is close to 1, both OSPC [23] and our proposed method struggle to achieve good results. This difficulty arises from the assumption in the first step of estimating ICRF, which requires significant exposure changes.

As previously mentioned, our method calculates each parameter based on the estimated result of the preceding parameter. It is observed that even if the RMSE of ICRF or vignette is large, the final RMSE of exposure can still perform better than that of OSPC [23]. This is

because the exposure ratio is calculated first, which offsets the error, and then the exposure value of the first frame is used to estimate the exposure for each subsequent frame.

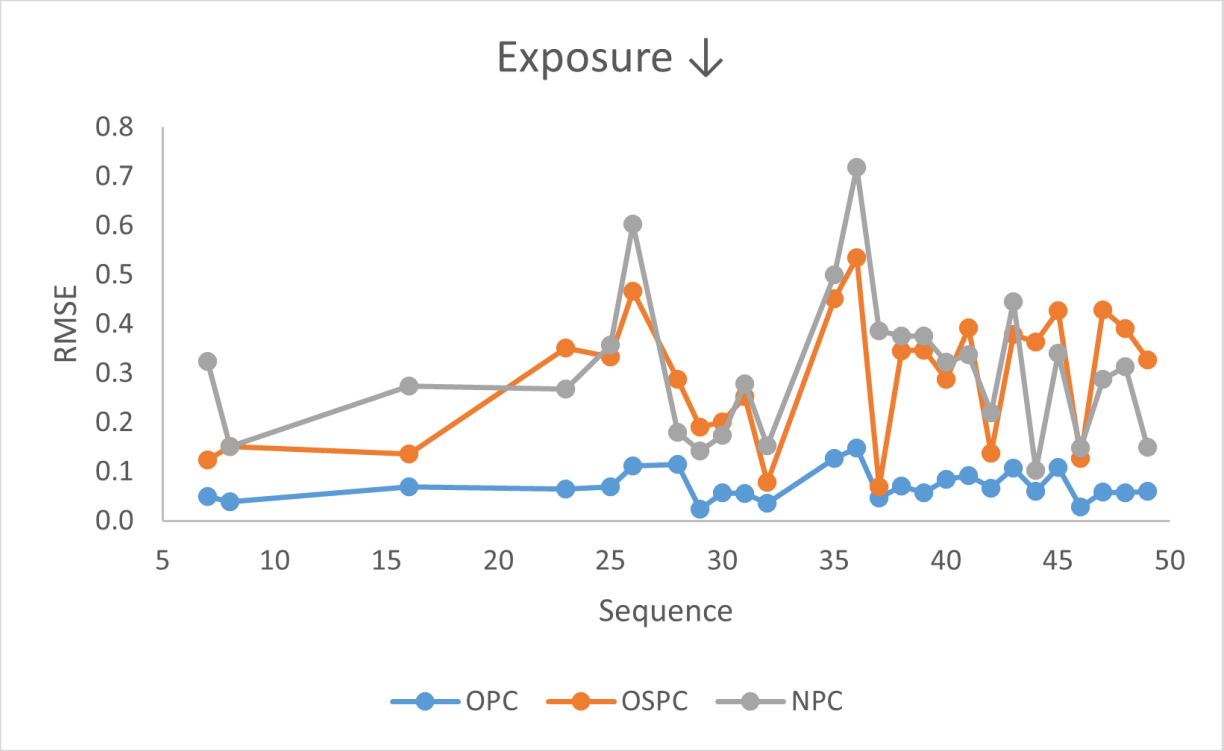


Figure 4.3: Comparative analysis of RMSE for exposure across different sequences for OPC [3], OSPC [23], and NPC methods. OPC [3] consistently achieves the lowest RMSE across all sequences, due to its intricate optimization and use of ground truth data. Additionally, even without ground truth, NPC often matches or surpasses OSPC [23].

Table 4.3: Average and STD of RMSE for exposure using OPC [3], OSPC [23], and NPC methods. The average shows that OPC produces the best results, whereas NPC can achieve results very close to OSPC [23].

Exposure	OPC [3]	OSPC [23]	NPC
Average ↓	**0.0715**	0.2916	0.3049
STD	**0.0314**	0.1320	0.1479

An experiment was conducted to evaluate the effect of patch size on the retrieval of exposure values. Several sequences were randomly selected for this purpose. The patch sizes tested were 1 pixel, 3×3, 5×5, 7×7, and 9×9. The results, as shown in Table 4.4, indicate that increasing the patch size to 9×9 yields the smallest RMSE across all sequences. We notice that RMSE go up first when increase the size from a single pixel to 3×3 then back down. This may due to a single pixel relies on subpixel accuracy, which can be very precise. A 3×3 patch might lose this precision, while larger patches might regain it by averaging over more pixels. Besides, the Shi-Tomasi corner [64] detector identifies strong corners, but a 3×3 patch might not capture enough context around the feature point, making it more susceptible to edge effects and leading to higher errors. Larger patches might mitigate this by including more of the feature's context.

Table 4.4: RMSE results of exposure for different sequences with varying patch sizes. Increasing the patch size to 9 yields the lowest RMSE for all sequences.

Patch size	Sq 16	Sq 30	Sq 48	Sq 49
1	0.2681	0.3218	0.3068	0.1480
3×3	0.2736	0.3237	0.3136	0.1496
5×5	0.2550	0.2971	0.3114	0.1501
7×7	0.2439	0.2859	0.3094	0.1466
9×9	**0.2327**	**0.2809**	**0.3070**	**0.1444**

An additional experiment was conducted to test the performance of our method using ground truth exposure. In other words, we replaced SIFT [49] in OSPC [23] with MFT [56]. As shown in Table. 4.5, the RMSE of the photometric parameters indicates that, in most cases, using MFT [56] yields better results. This improvement is likely due to the superior ability of MFT [56] to identify corresponding points compared to SIFT [49].

Table 4.5: Performance comparison of SIFT [49] vs. MFT [56]. The RMSE of the photometric parameters is displayed in the table.MFT [56] demonstrates superior results in identifying corresponding points and achieves lower RMSE in most cases.

RMSE	OSPC [23]	NPC	OSPC [23]	NPC	OSPC [23]	NPC
Sequence	ICRF		Vignette		Exposure	
08	13.3228	**9.0761**	0.1244	**0.0662**	0.1513	**0.1266**
16	20.6559	**12.1422**	0.0908	**0.0824**	0.1361	**0.1022**
38	18.3018	**18.2028**	0.0789	**0.0531**	**0.3453**	0.3813
45	**10.2637**	13.5300	0.0755	**0.0631**	0.4264	**0.2845**

The computational time for each key step was monitored. Table. 4.6 presents the timing results per frame of our method, with all steps independently timed. Although the extraction process is quite fast, the MFT tracker is relatively slow for tracking each frame. For photometric calibration, using the Lagrange method to solve the ICRF and the M estimator to solve the vignette, combined with exposure estimation and reconstruction, requires approximately 0.0373 seconds. This indicates that we are close to achieving real-time performance at 30 fps.

Table 4.6: Execution time per frame for the key steps. While the MFT tracker exhibits a slower performance, the aggregated execution time for photometric calibration is approximately 0.0373 seconds. This suggests that achieving real-time performance at 30 fps is within reach.

Step	Action	Execution time (s)
Tracking	Extract on first frame	0.0066
	Extract on later frame	0.0034
	Tracking	0.2360
	Check validity	0.0001
ICRF	Remove huge distance	0.000314
	CVX solver	0.1112
	Lagrange method	0.000007
Vignette	remove huge distance	0.000311
	CVX solver	0.1131
	M estimator	0.000395
Exposure	Retrieving exposure	0.0223
Reconstruction	Time per frame	0.0140

In summary, OPC [3] generally outperforms OSPC [23] and NPC in estimating the three parameters. However, for some sequences, our method surpasses OSPC [23] across all parameters, demonstrating the feasibility of photometric calibration based on a sequential method without ground truth information. NPC exhibits high variability due to the initialization of exposure, which depends on pixel intensity. Consequently, poor image pairs can lead to worse estimations. Although NPC achieves good results in some sequences, its higher deviation indicates that the method is less consistent and more affected by variations in image pairs, resulting in less reliable outcomes.

For more details, Table A.1 presents the RMSE results for different methods and parameters, including ICRF, vignette, and exposure, from the TUM MONO Dataset [14].

Additional charts can be found in Appendix A.

4.1.2 Image and Video Quality Assessment

We first use the original sequence as the reference frame for calculating FR-IQA. Figure 4.4 shows the plot of the average score for each sequence across different methods. The average score is calculated by taking the negative of LPIPS (since lower values indicate better quality) and averaging the sum of all scores from the full reference matrices. Therefore, a higher average score indicates closer quality to the original frame. Table 4.7 presents the average and STD of the scores across all sequences. The chart and average scores show that all methods achieve similar outcomes. Some high peaks, such as in sequences 26 and 35, result from high PSNR scores. Since we average the scores from all FR-IQA matrices without weighting or normalization, this may lead to some significant results.

Table 4.7: Average and STD of FR-IQA across all sequences for three methods. The average FR-IQA score indicates that OSPC [23] produces the highest image quality among the three methods, with NPC also achieving better quality than OPC [3]. Higher quality refers to being closer to the original video sequences.

FR-IQA	OPC [3]	OSPC [23]	NPC
Average ↑	6.4015	**7.1688**	6.8292
STD	**0.8339**	1.0615	1.0468

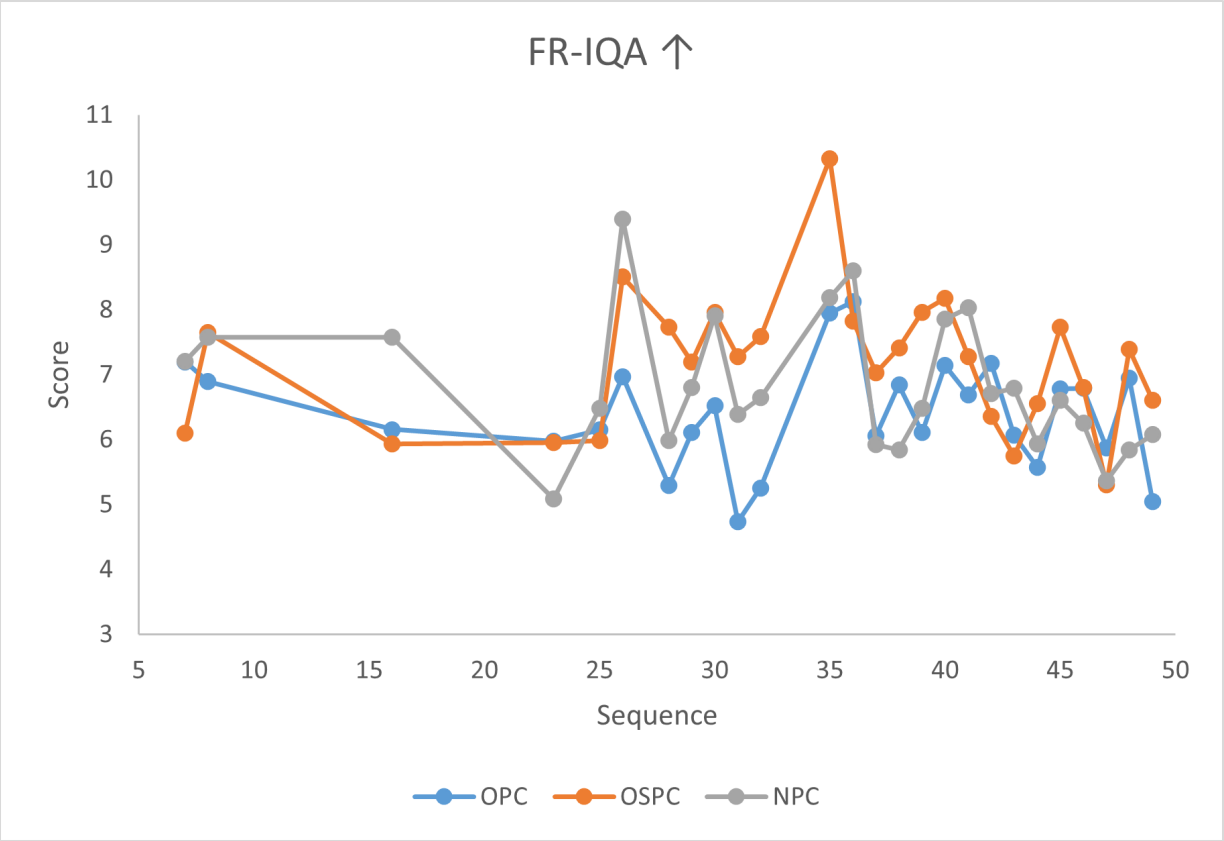


Figure 4.4: FR-IQA scores for three methods on TUM MONO Dataset [14]. The chart shows that the three methods yield similar results. The high peaks in sequences 26 and 35 are due to high PSNR scores. Averaging scores from all FR-IQA matrices without weighting or normalization may lead to significant results.

To gain insight into image quality, we utilize five NR-IQA matrices, same as the previous mentioned matrices. Fig. 4.5 shows a chart of NR-IQA score across all sequences for three approaches and the original video sequences. Table 4.8 provides the average and STD of NR-IQA score. Since NIQE is a metric where lower values indicate better quality, we take the negative of NIQE scores, sum all the scores across the matrices, and calculate the average score. Thus, a higher average score represents better image quality, while the negative NIQE score results in a negative average score. We observe that for the majority of sequences, OSPC [23] and NPC have higher scores. This may be due to our reconstruction method, where we set a threshold for the vignette factor, constraining it between 0.3 and 1,

as modern optical designs do not produce extremely poor imaging. Additionally, we apply gamma correction, which is not used in OPC [3], resulting in OPC [3] often generating very dark reconstructed frames.

Table 4.8: Average and STD of NR-IQA across all sequences for three methods. The average score reveals that OSPC [23] and NPC may produce better quality than OPC [3] reconstructed frames.

NR-IQA	Original	OPC [3]	OSPC [23]	NPC
Average ↑	-0.2016	-0.3686	**-0.1373**	-0.1558
STD	0.4544	**0.2778**	0.3964	0.4080

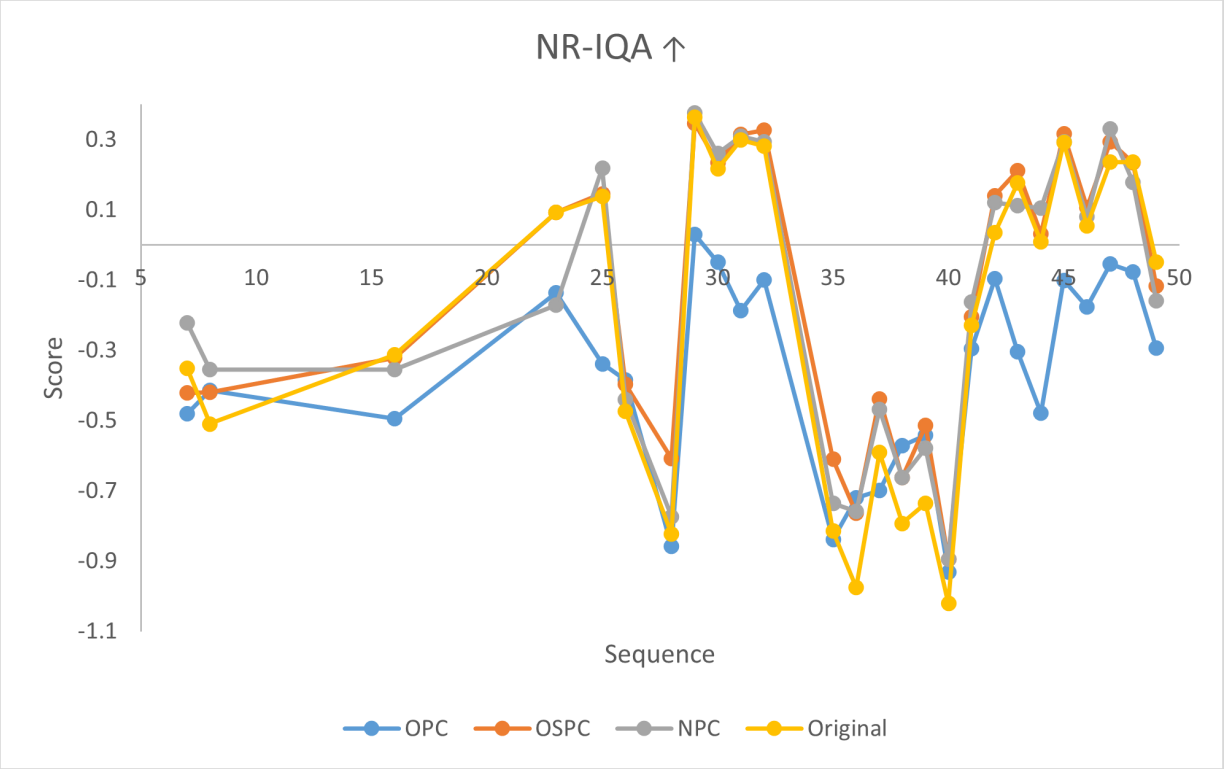


Figure 4.5: NR-IQA scores for three methods on TUM MONO Dataset [14]. OSPC [23] and NPC achieve higher scores for the majority of sequences, likely due to our reconstruction method. We use gamma correction, which OPC [3] does not, leading to OPC [3] frequently producing very dark reconstructed frames and causing low scores.

We further convert the calibrated frames into videos and use VMAF to evaluate their quality compared to the original videos. Fig. 4.6 presents the VQA scores and Table 4.9 displays the average and STD, where the metric is higher-the-better. OPC [3] generally outperforms the other methods in most cases. However, in some sequences, our proposed NPC achieves the highest score. Interestingly, in sequences 16 and 23, NPC receives the lowest scores. As shown in Fig. 4.7, the image quality for NPC in sequence 16 was actually acceptable but appeared dim. The low score may be due to the high contrast in the reference frame, which OPC [3] preserves, while OSPC [23] and NPC use gamma correction that reduces the contrast. This is supported by NR-IQA, where both OSPC [23] and NPC achieve higher scores. For sequence 23, shown in Fig. 4.8, the recovered frame exhibits a strong vignetting effect and relatively low overall brightness, which lowers the quality.

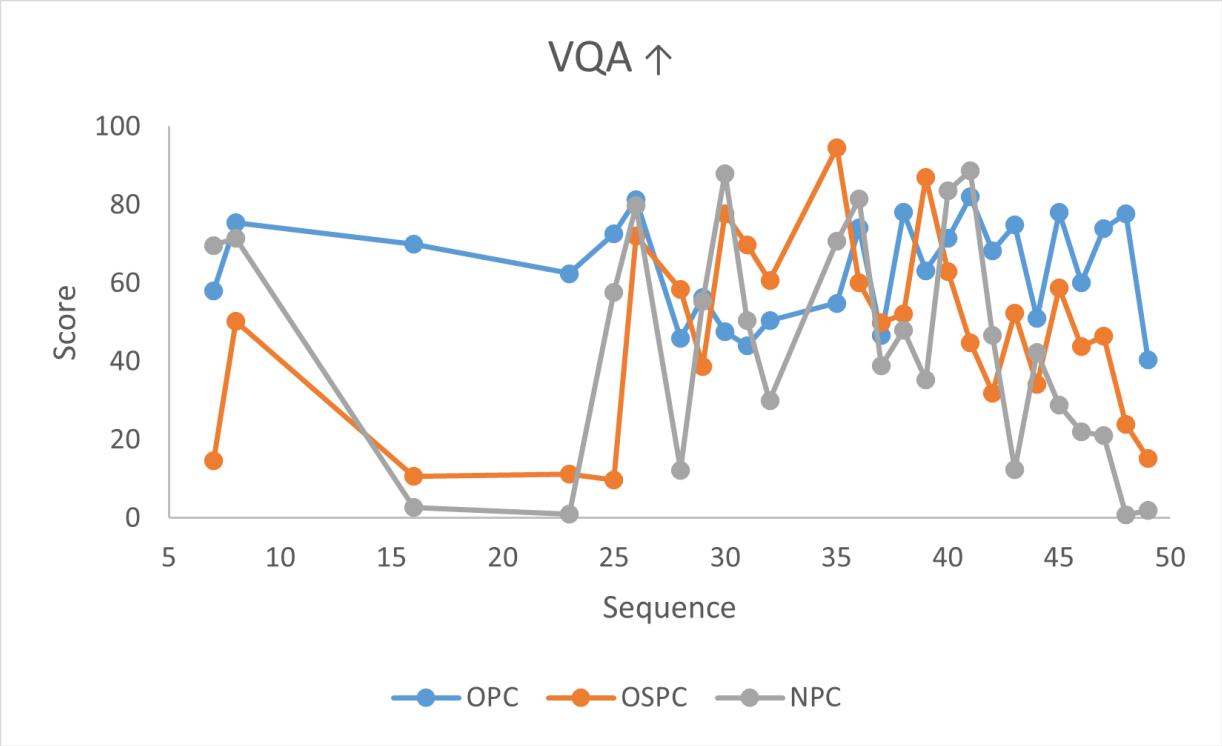


Figure 4.6: VQA scores for three methods on TUM MONO Dataset [14]. In this chart, OPC [3] appears to have a higher VQA score, indicating that its quality is closer to the original video.

Table 4.9: Average and STD of VQA across all sequences for three methods. OPC [3] outperforms OSPC [23] and NPC, with the significant gap between them likely due to the low scores for OSPC [23] and NPC in some sequences. However, the high VQA score indicates that the quality is similar to original video.

VQA	OPC [3]	OSPC [23]	NPC
Average ↑	**63.8077**	47.3873	43.8933
STD	**13.0400**	23.6207	29.2996

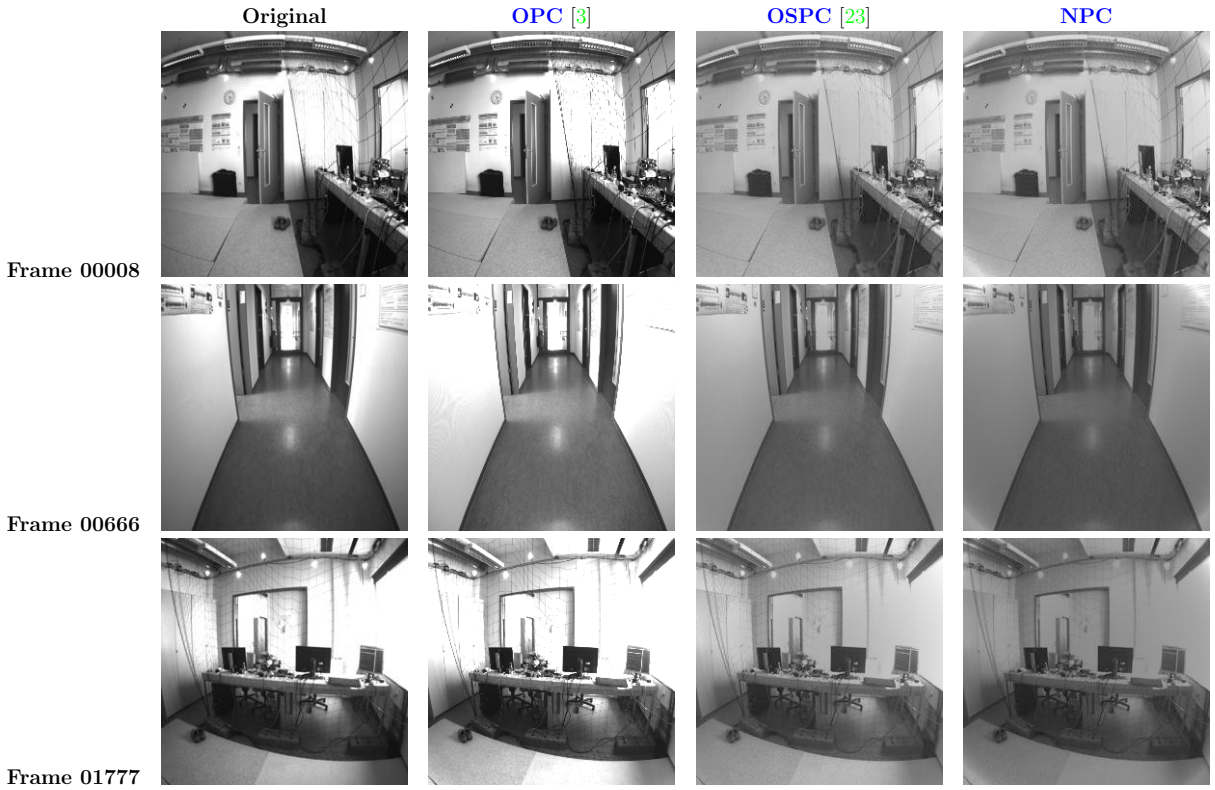


Figure 4.7: Example of reconstructed frames in Sq 16. OSPC [23] and NPC results appear more uniform but dimmer, whereas the original and OPC [3] frames have higher contrast. This difference leads to the low VQA scores in OSPC [23] and NPC. However, the reconstructed frames are still acceptable.

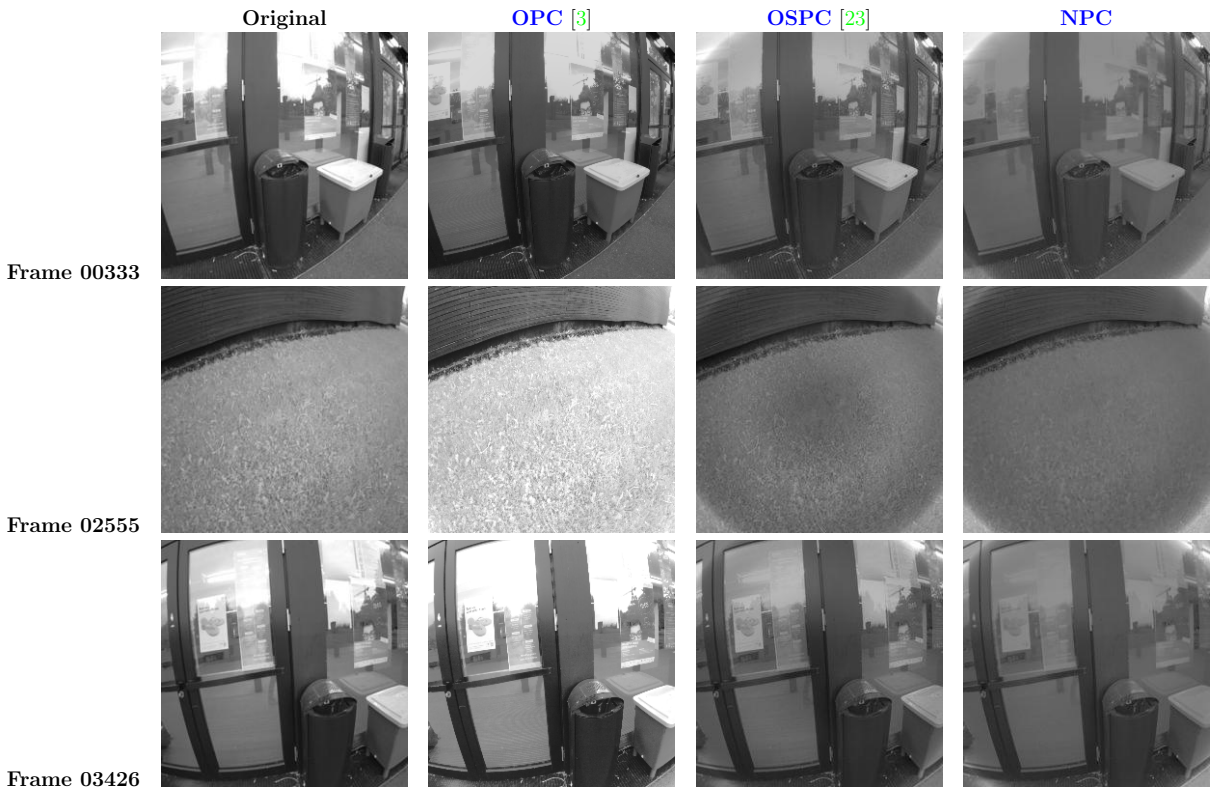


Figure 4.8: Example of reconstructed frames in Sq 23. The recovered frame shows a pronounced vignetting effect and relatively low overall brightness, which reduces the quality in OSPC [23] and NPC. In contrast, the reconstructed frames of OPC [3] appear brighter.

Additionally, we found that the parameter V significantly impacts overall visual performance. High ICRF errors typically result in low-contrast images with high average intensity, as observed in Fig. 4.9. We also notice that when a large area is plain (e.g. Frame 02555 in sequence 23, shown in Fig. 4.8), the vignetting factor is more pronounced, significantly affecting the quality.

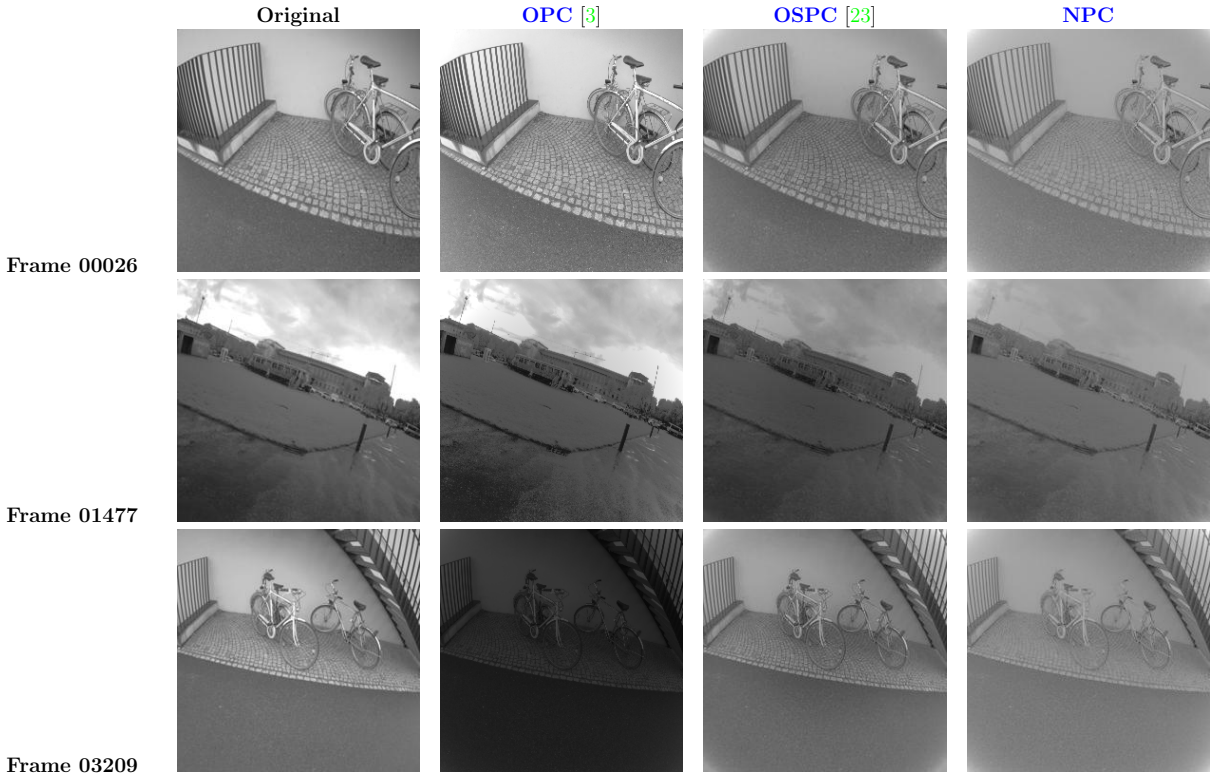


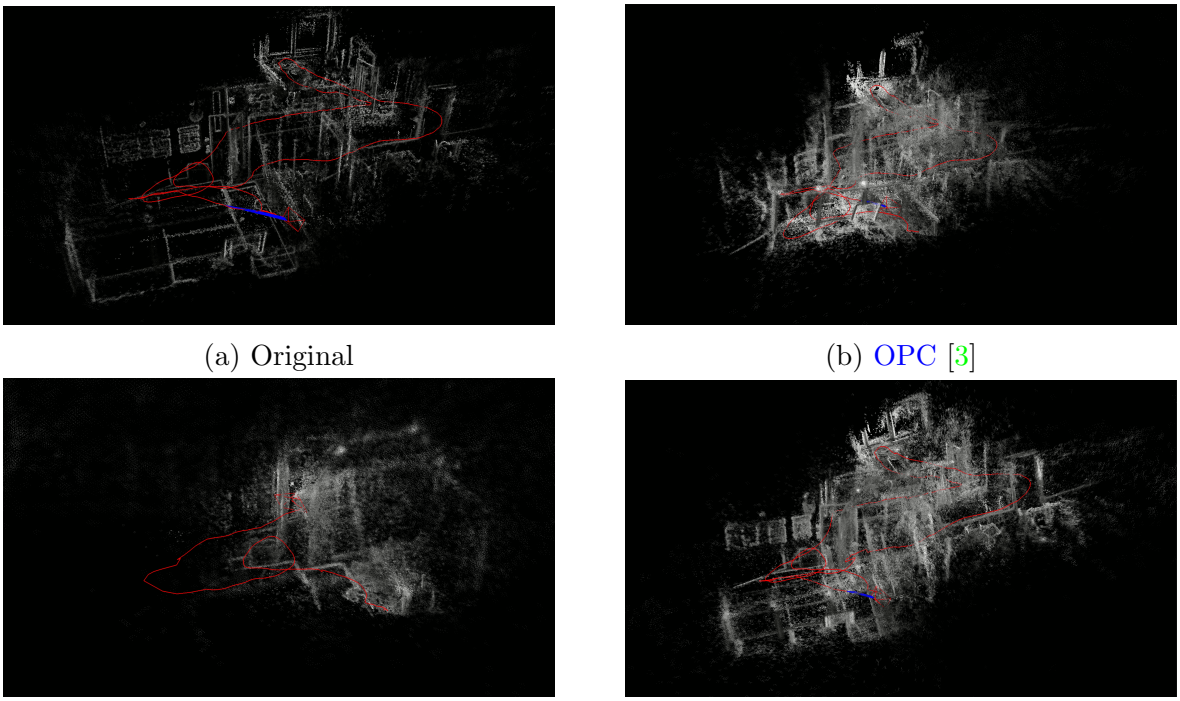
Figure 4.9: Example of reconstructed frames in Sq 49. High error in ICRF for NPC leads to low-contrast frames, resulting in a low VQA score.

4.1.3 DSO results

The photometrically calibrated frames are fed into DSO [11] to evaluate performance. To assess the trajectory derived from DSO [11], EVO [22] is employed. Unfortunately, for most sequences, due to the downsizing of the images and the unsatisfactory quality of the reconstructed frames, DSO [11] breaks down during execution. Consequently, we cannot compare all three methods simultaneously for most sequences. However, we can still compare with OPC [3] and OSPC [23] respectively when they work.

Fig. 4.10 exhibits the reconstructed scene of sequence 07 with the trajectory across different methods. It is evident that OSPC [23] produces the worst scene, whereas NPC and OPC [3] generate similar results. On the other hand, Fig. 4.11 illustrates the DSO [11] trajectory across different approaches. The RMSE values for OPC, OSPC, and NPC are 0.9308, 2.7452, and 0.2978, respectively. Our results are the closest to the original trajectory, which uses ground truth information. This can be evidenced by combining the previous evaluations, including parameters, IQA, and VQA, demonstrating that our method has the potential to achieve satisfactory results without any ground truth information. Fig. 4.12, 4.13, 4.14 show the DSO [11] results for sequence 26, 29 and 30, respectively. The RMSE values and the reconstructed scenes indicate that our proposed method, NPC, performs worse than OPC [3] but outperforms OSPC [23].

These findings suggest that the current photometric calibration techniques lack robustness, particularly for images that have been significantly downsized. During our experiments, we observed that DSO [11] fails when there are substantial changes in exposure between consecutive frames. This indicates that achieving accurate photometric calibration is challenging when the exposure ratio is close to 1.



(a) Original (b) OPC [3]

(c) OSPC [23] (d) NPC

Figure 4.10: DSO [11] scene using calibrated frames from different approaches for sequence 07. The reconstructed scene by NPC is the most similar to the original, while OPC [3] and OSPC [23] exhibit more noise.

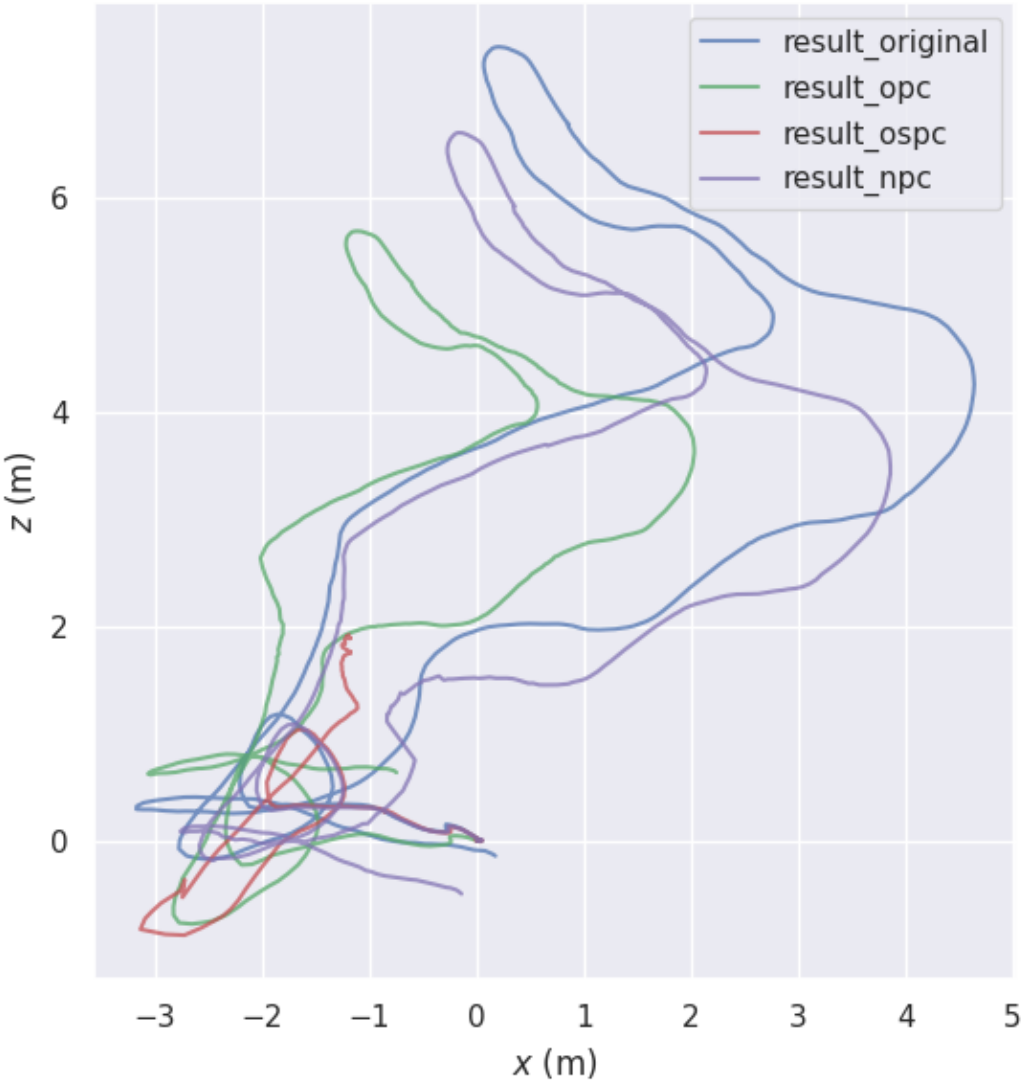
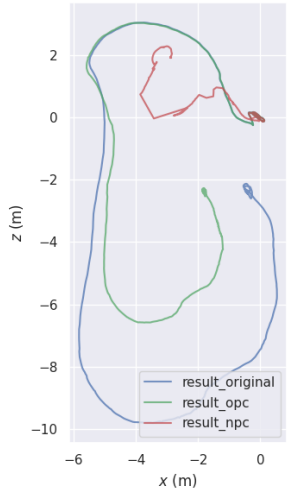
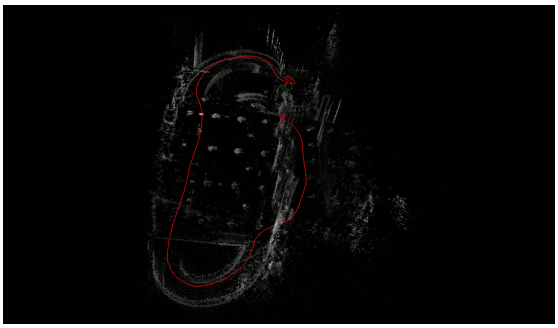


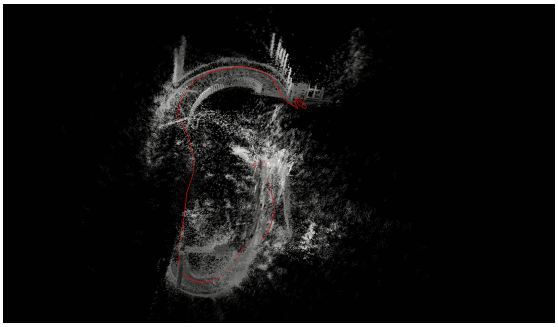
Figure 4.11: DSO [11] trajectory using reconstructed frame from different methods on TUM MONO Dataset [14] sequence 07. The RMSE values for OPC [3], OSPC [23], and NPC are 0.9308, 2.7452, and 0.2978, respectively. The trajectory of NPC is the closest to the original, which uses photometric ground truth.



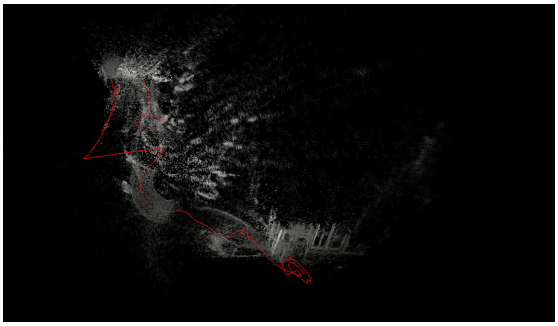
(a) Trajectory



(b) Original

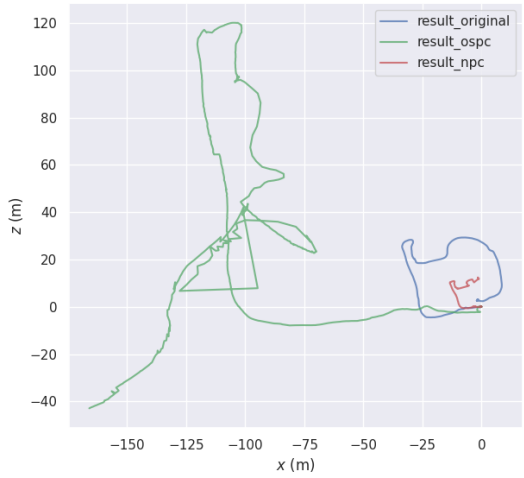


(c) OPC [3]

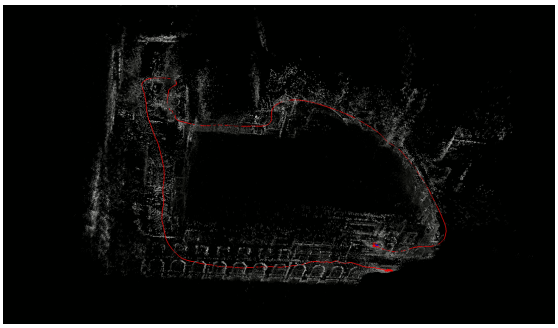


(d) NPC

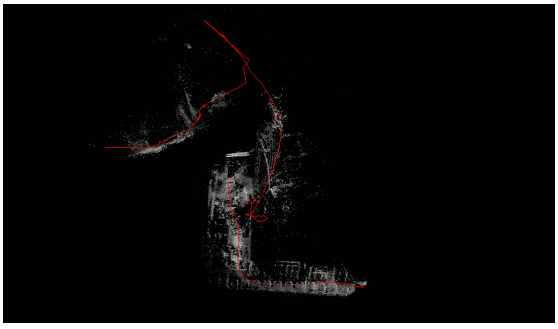
Figure 4.12: DSO [11] scene using calibrated frames from different approaches for sequence 26. DSO [11] cannot operate on OSPC [23] calibrated frames, and both OPC [3] and NPC exhibit significant noise. The RMSE values for OPC and NPC are 1.6935 and 4.1066, respectively. The OPC [3] trajectory is closer to the original trajectory.



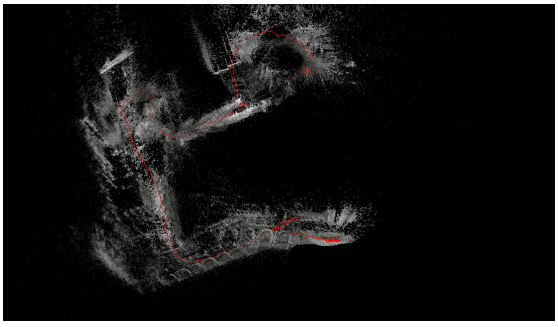
(a) Trajectory



(b) Original

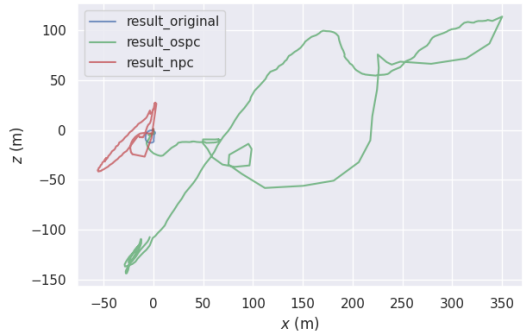


(c) OSPC [23]

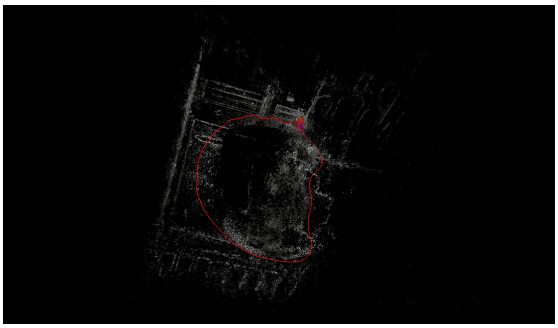


(d) NPC

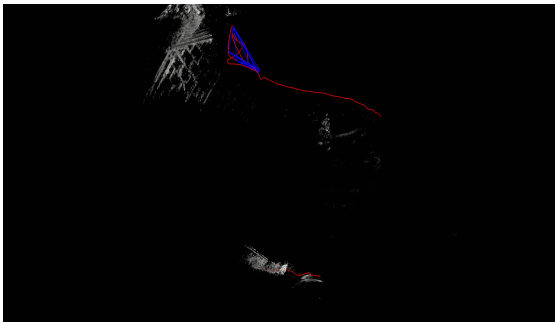
Figure 4.13: DSO [11] scene using calibrated frames from different approaches for sequence 29. The calibrated frames by OPC [3] fail to execute DSO [11]. The DSO [11] reconstructed scene using OSPC [23] almost completely fails. NPC's results are the closest to the original, but the scale appears much smaller than the original sequence. The RMSE values for OSPC [23] and NPC are 66.3572 and 11.0565, respectively.



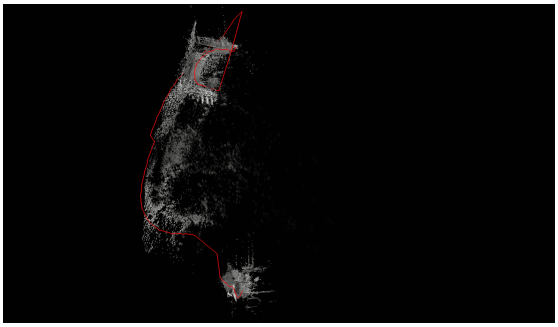
(a) Trajectory



(b) Original



(c) OSPC [23]



(d) NPC

Figure 4.14: DSO [11] scene using calibrated frames from different approaches for sequence 30. DSO [11] failed when executing the calibrated sequence by OPC [3]. The DSO [11] reconstructed scene using OSPC [23] is very noisy and almost completely fails. The DSO [11] results using NPC also seem to fail, reconstructing only half of the scene. The RMSE values for [23] and NPC are 170.2470 and 83.7905, respectively.

4.2 Analysis without Ground Truth Information

In this section, we randomly selected four sequences from the EuRoC Dataset [4], two videos from the ICL-NUIM Dataset [25], and four sequences from the KITTI Dataset [17]. These datasets do not provide photometric ground truth. All images were maintained in their original form without scaling. As usual, we evaluated three aspects: photometric parameters, image and video quality assessment, and DSO [11] trajectory. Due to the absence of ground truth information, OSPC [23] could not be utilized; therefore, we only compared our results with OPC [3].

4.2.1 Photometric Parameters

Figures 4.15, 4.16, and 4.17 present the estimated photometric parameters for OPC [3] and NPC across the ICL-NUIM [25], EuRoC [4], and KITTI Datasets [17], respectively. The blue line indicates OPC results and the red line is NPC. It is evident that the exposure results using OPC [3] are highly unstable, oscillating between 0 and 1. In contrast, the results obtained using NPC are more reasonable. On the other hand, the estimated ICRF for both methods are similar and appear stable. For vignette correction, NPC seems to produce worse results than OPC [3], with some outcomes being irrationally low. Particularly for the KITTI Dataset [17], both approaches yield poor results due to the rapid movement, leading to inaccuracies.

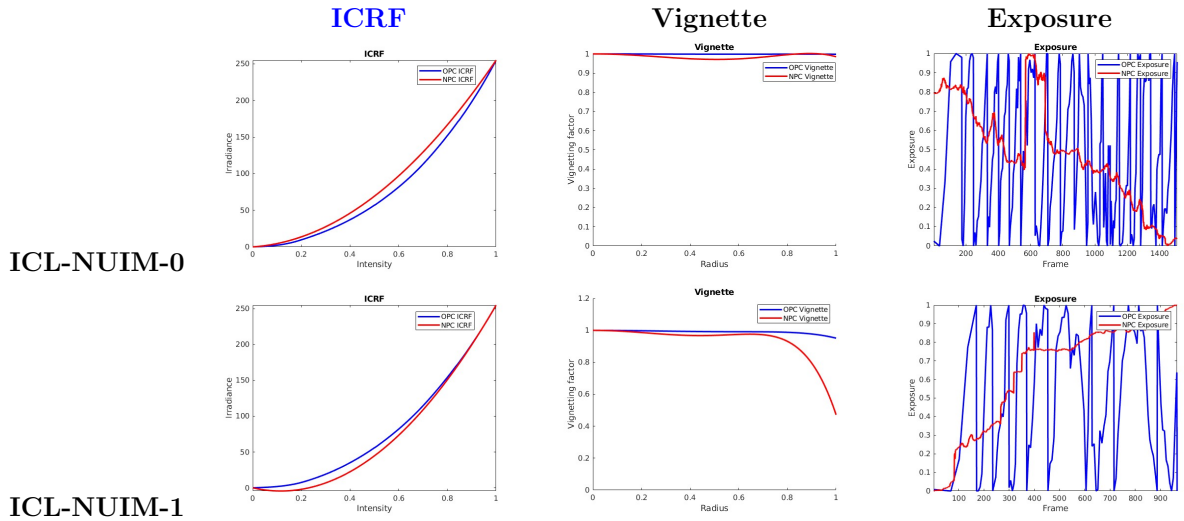


Figure 4.15: Photometric calibration without ground truth on ICL-NUIM Dataset [25]. OPC [3] and NPC show similar ICRF results for both scenes. For vignetting, both methods yield very similar results in ICL-NUIM-0, indicating no vignetting effect. However, in the other scene, their estimations differ. In the exposure chart, it is evident that OPC [3] results are poor, with strong oscillations.

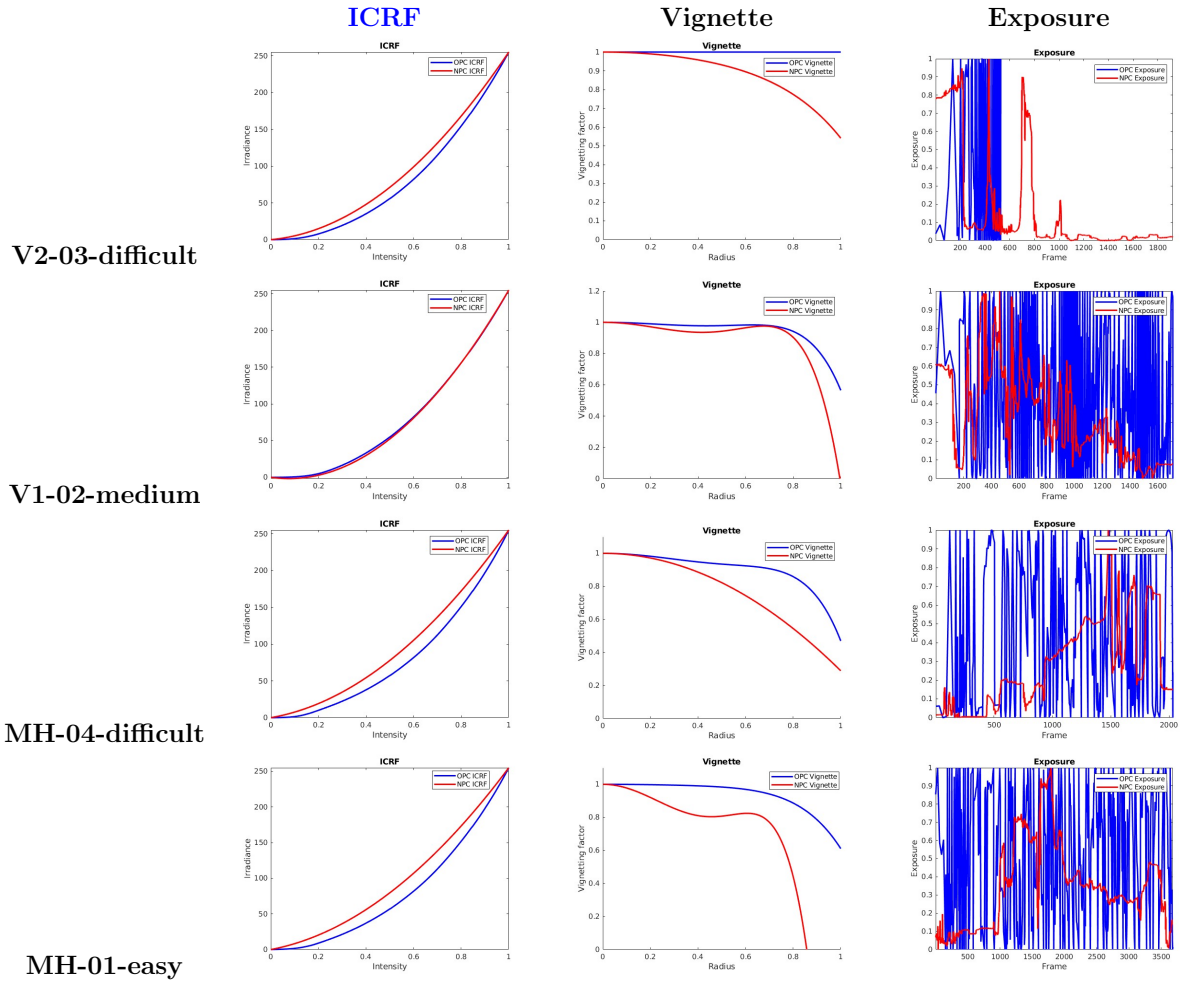


Figure 4.16: Photometric calibration without ground truth on EuRoc Dataset [4]. Without any ground truth information, OPC [3] struggles to accurately retrieve exposure, whereas NPC still retrieve relatively normal results.

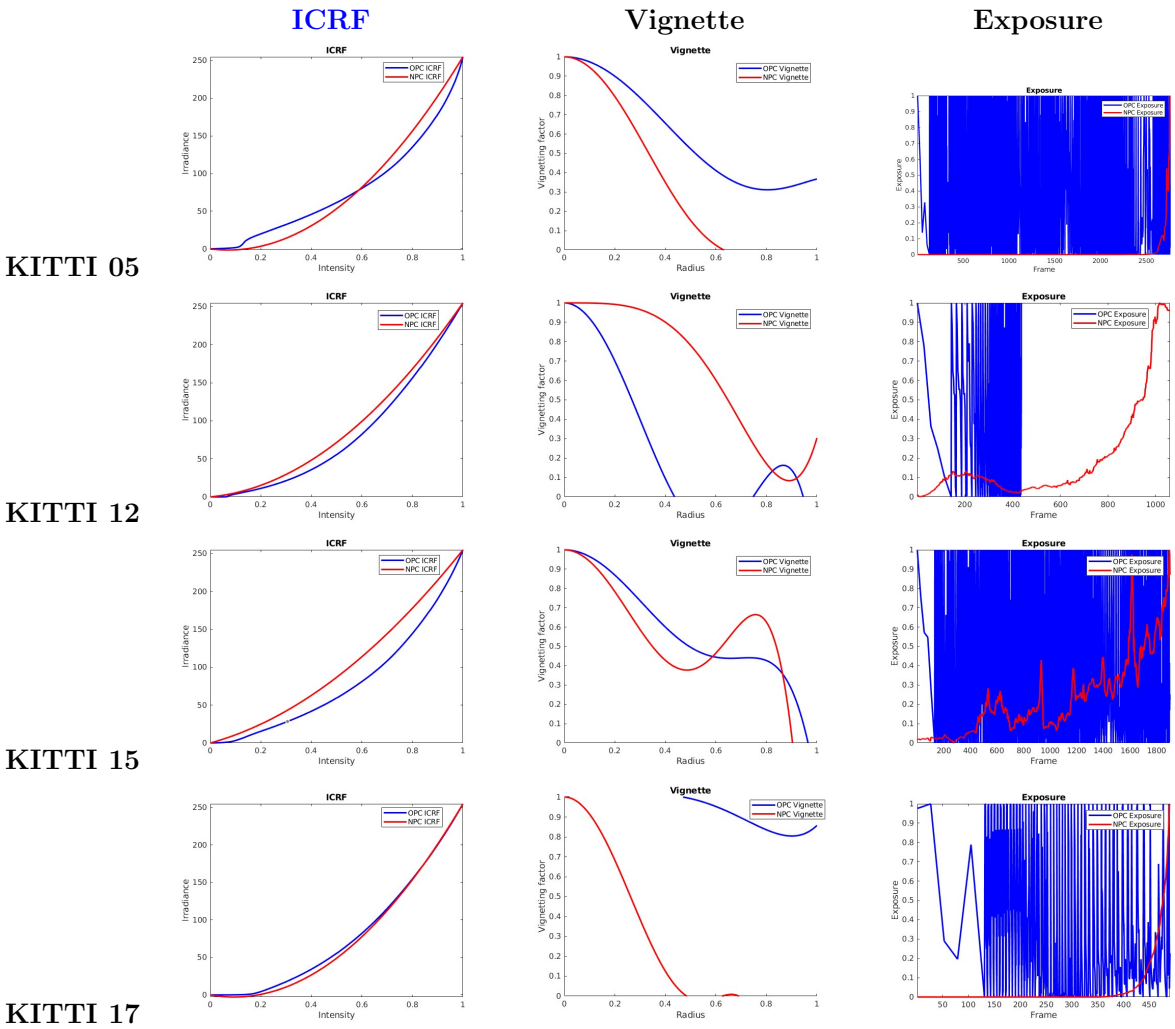


Figure 4.17: Photometric calibration without ground truth on KITTI Dataset [17]. Although the estimation in the ICRF is accurate, the vignette and exposure results are notably poor, particularly for OPC [3]. In contrast, the NPC exposure results are more reasonable.

4.2.2 Image and Video Quality Assessment

Table 4.10 lists the average scores and STD of all sequences for OPC [3] and NPC calibrated frames. For VQA scores, NPC is slightly lower than OPC [3]; however, for FR-IQA and NR-IQA, NPC outperforms OPC [3] and exhibits lower STD. This may be due to the NPC calibrated frames of MH-01-easy (# 4) being too dim, as shown in Fig. 4.19, and lowered the overall average. Nonetheless, the overall performance of NPC on the ICL-NUIM [25] and KITTI Datasets [17] is superior than OPC [3]. Additionally, the image size is not scaled, resulting in NR-IQA scores greater than zero due to the higher resolution.

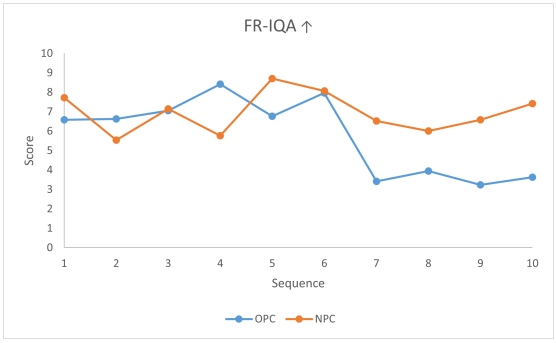
Table 4.10: Average and STD of IQA and VQA across all sequences for OPC [3] and NPC. The FR-IQA and NR-IQA scores indicate that NPC reconstructs higher quality frames. Although the VQA score for OPC [3] is higher, it only considers the successfully reconstructed frames. During the execution of OPC [3], the photometric calibration process failed in V2-03-difficult and KITTI 12. Additionally, the frames reconstructed by OPC [3] often exhibit a strong vignetting effect. Therefore, the overall quality of OPC [3] should be considered inferior to NPC.

Score	Original	OPC [3]	NPC
FR-IQA Average ↑	-	5.7627	**6.9449**
FR-IQA STD	-	1.9909	**1.0443**
NR-IQA Average ↑	**0.3584**	0.2556	0.3433
NR-IQA STD	0.2208	1.7678	**0.2191**
VQA Average ↑	-	**64.4839**	60.2539
VQA STD	-	**13.5745**	32.1182

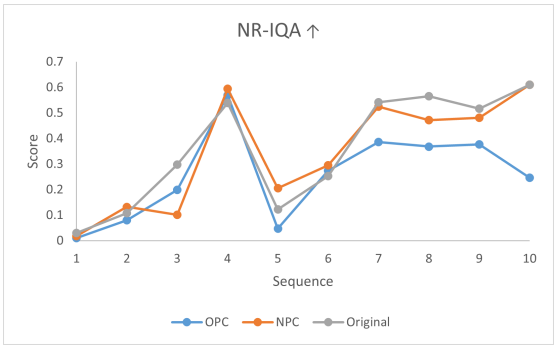
Fig. 4.18 illustrates the average FR-IQA, NR-IQA, and VQA scores for the two methods across the ten calibrated sequences that lack photometric ground truth. Please refer to Table 4.11 for the corresponding sequences. The NR-IQA results indicate that, for most sequences, the photometrically calibrated images from NPC achieve higher scores and, consequently, better quality. It is important to note that OPC [3] crashes on some sequences, so the scores are computed only for the successfully calibrated frames. In summary, NPC achieves significantly better quality for calibrated frame without ground truth data.

Table 4.11: Lookup table for the sequences and corresponding numbers for the charts.

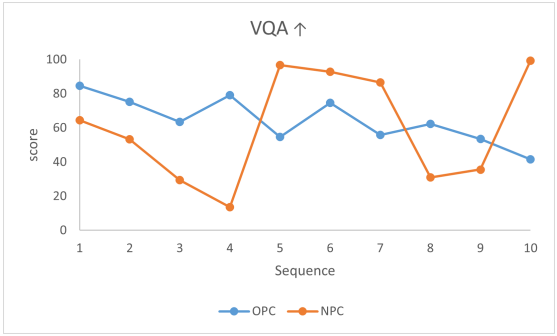
Number	1	2	3	4	5
Sequence	V2-03-difficult	V1-02-medium	MH-04-difficult	MH-01-easy	ICL-NUIM-0
Number	6	7	8	9	10
Sequence	ICL-NUIM-1	KITTI 05	KITTI 12	KITTI 15	KITTI 17



(a) FR-IQA



(b) NR-IQA



(c) VQA

Figure 4.18: FR-IQA, NR-IQA, and VQA scores for the calibrated frame using OPC [3] and NPC without ground truth information. FR-IQA and NR-IQA scores indicate that the reconstructed frames of NPC are of better quality. The low VQA score for NPC in index 4 is due to low intensity in the reconstructed sequence. The VQA score does not accurately reflect that OPC's [3] overall quality is inferior to NPC, due to strong vignetting and the failed execution of the entire video sequence.

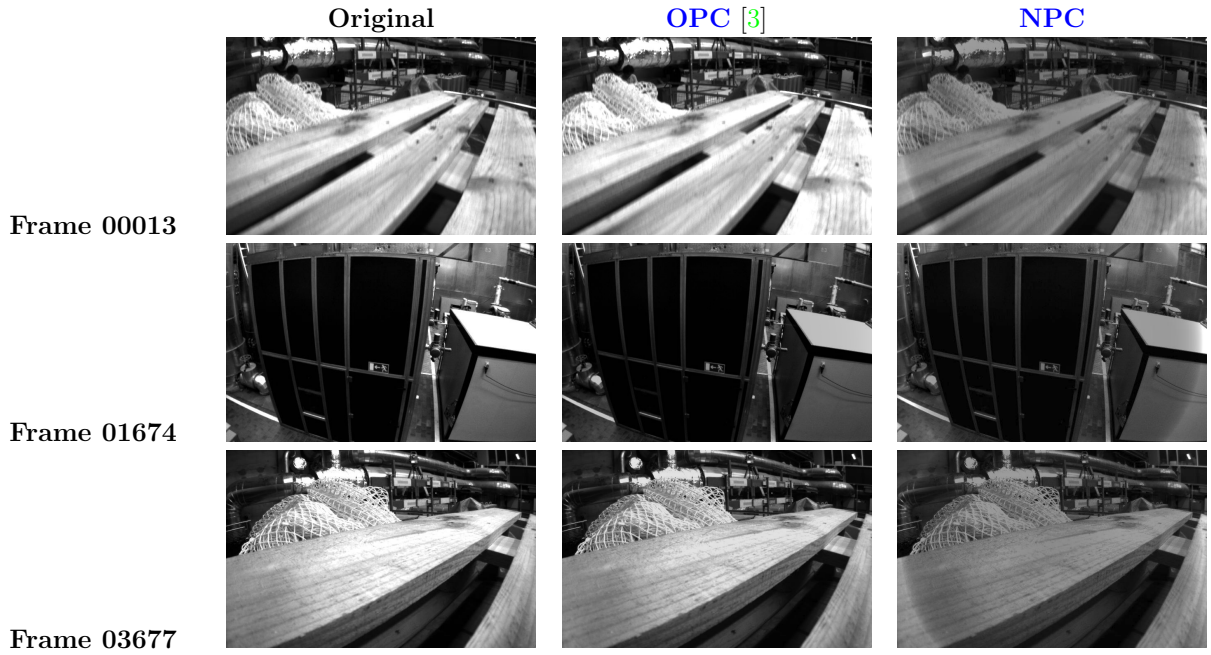


Figure 4.19: Example of reconstructed frames in MH-01-easy. The photometric calibrated frames by NPC are darker thorough our whole sequence, which leads to lower VQA score.

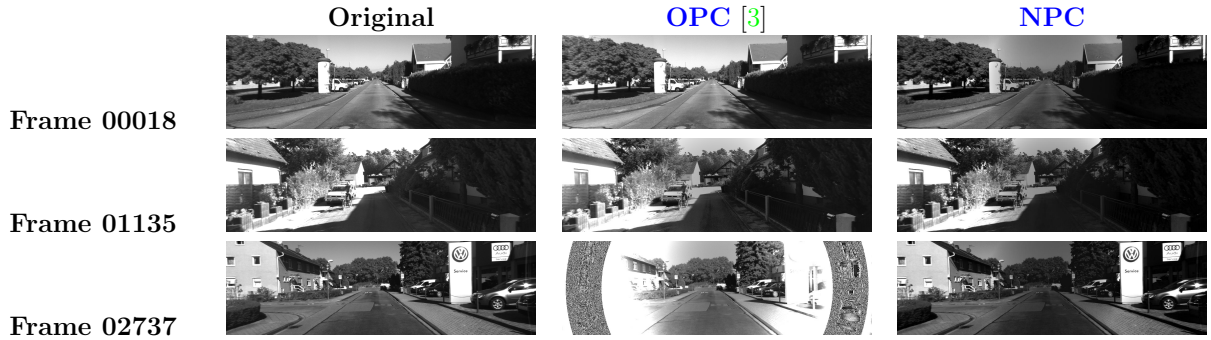


Figure 4.20: Example of reconstructed frames in KITTI 05. The faster movement in the KITTI Dataset [17] poses challenges for both OPC [3] and NPC. Frame 2737 exhibits strong artifacts due to vignetting for OPC [3]. However, NPC retrieves relatively better results compared to OPC [3].

Fig. 4.20 and Fig. 4.21 illustrate some of the poorly calibrated results from OPC [3]. The reconstructed frames exhibit strong artifacts, primarily due to the vignetting factor.

In contrast, NPC not only successfully calibrates the entire video but also achieves quite satisfactory results. The rapid movement in the KITTI Dataset [17] often leads to damage and failure in OPC [3]. However, by employing MFT [56] in our method, normal operation can still be maintained.

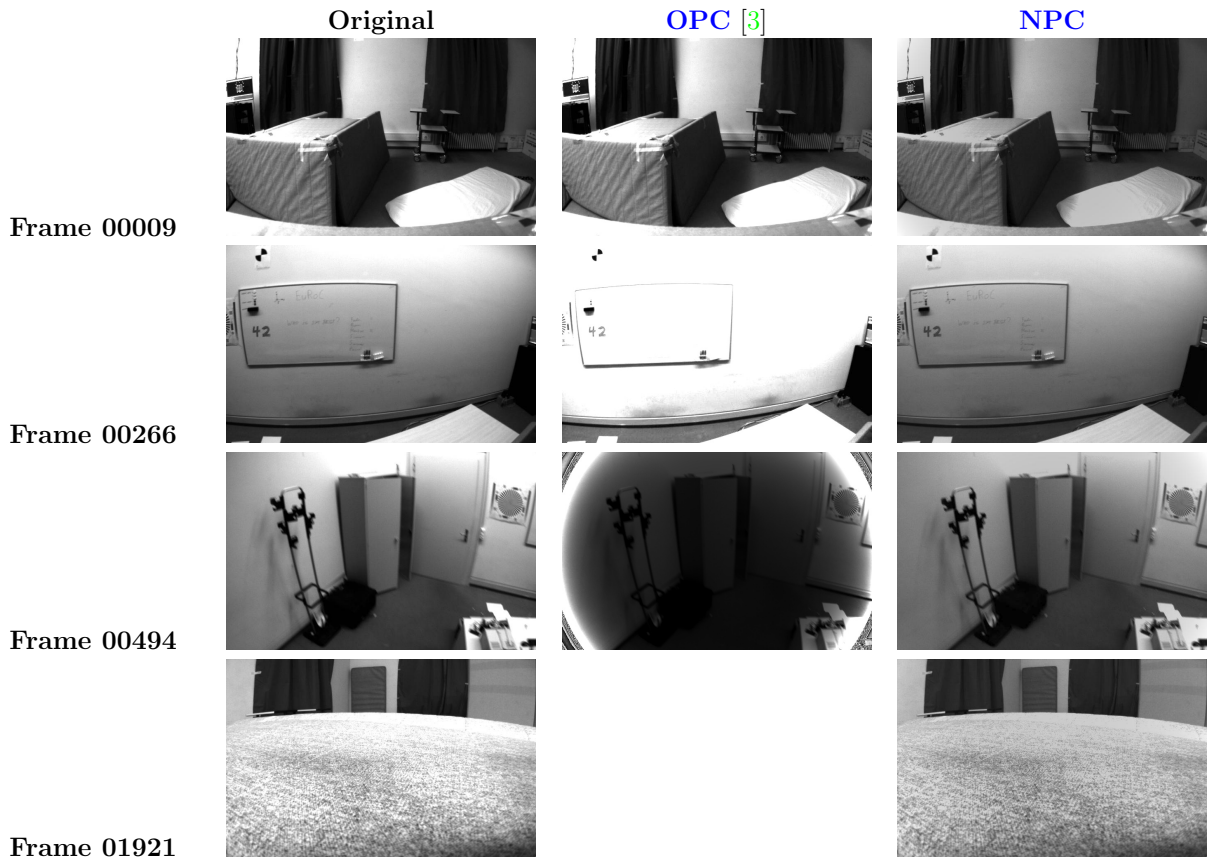


Figure 4.21: Example of reconstructed frames in V2-03-difficult. Since OPC [3] breaks down in the middle, the frame 01921 remains blank. Frame 266 in OPC [3] appears overly bright, while Frame 494 exhibits a strong vignetting effect with a very dark scene. In contrast, NPC produces much better quality throughout the entire video sequence.

4.2.3 DSO results

DSO [11] can successfully execute in most sequences, owing to the large, unscaled image size. However, it fails in the KITTI Dataset [17] due to high motion. Additionally, the trajectory and reconstructed scenes are suboptimal, primarily due to the instability of exposure, which causes flashing. In Figure 4.22, the trajectory plot shows that OPC [3] is closer to the original trajectory. The RMSE values for OPC and NPC are 0.0828 and 0.2069, respectively. This may due to our vignette result is much worse and cause the trajectory off the track. Figure 4.23 presents the reconstructed scenes across different datasets. OPC [3] yields relatively poorer results due to drastic changes in exposure, leading to DSO [11] losing track. Additionally, for the EuRoC Dataset [4], NPC performs better than OPC [3], with MF-04-difficult NPC presenting a much better outcome. The RMSE values for OPC [3] and NPC are 37.6857 and 13.3399, respectively.

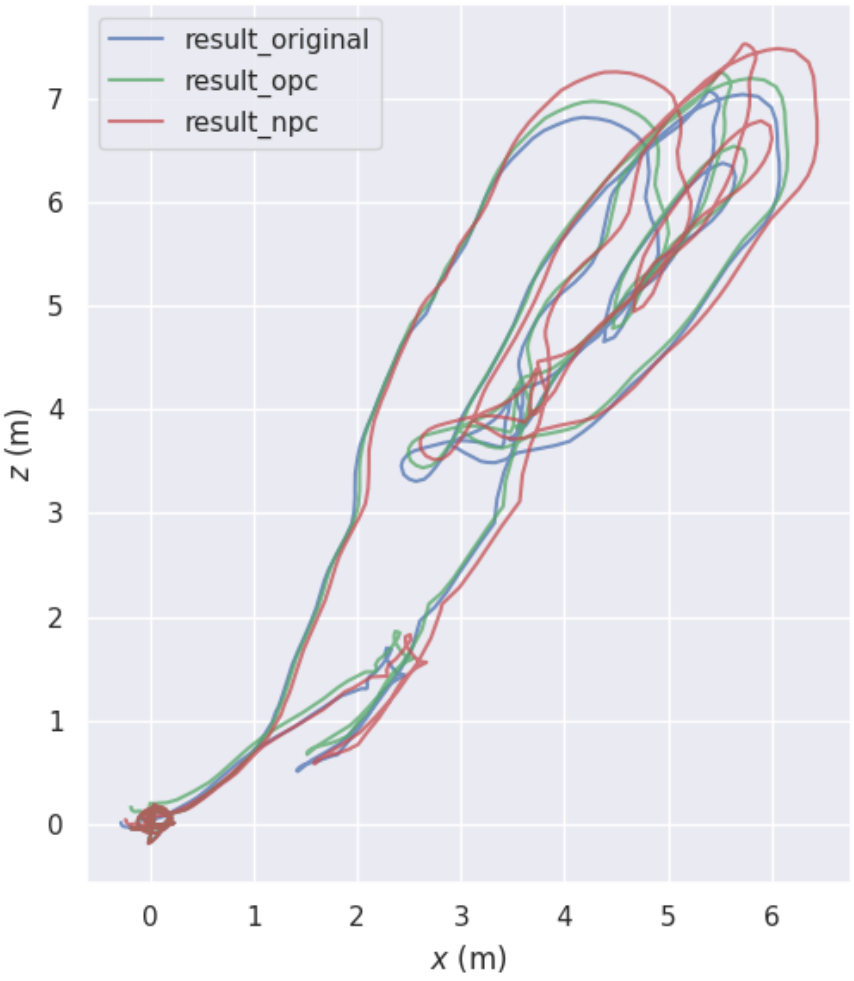


Figure 4.22: DSO [11] trajectory using reconstructed frame from different methods on EuRoc [4] MH-01-easy. The RMSE values for OPC [3] and NPC are 0.0828 and 0.2069, respectively. The OPC [3] trajectory is closer to original than NPC.

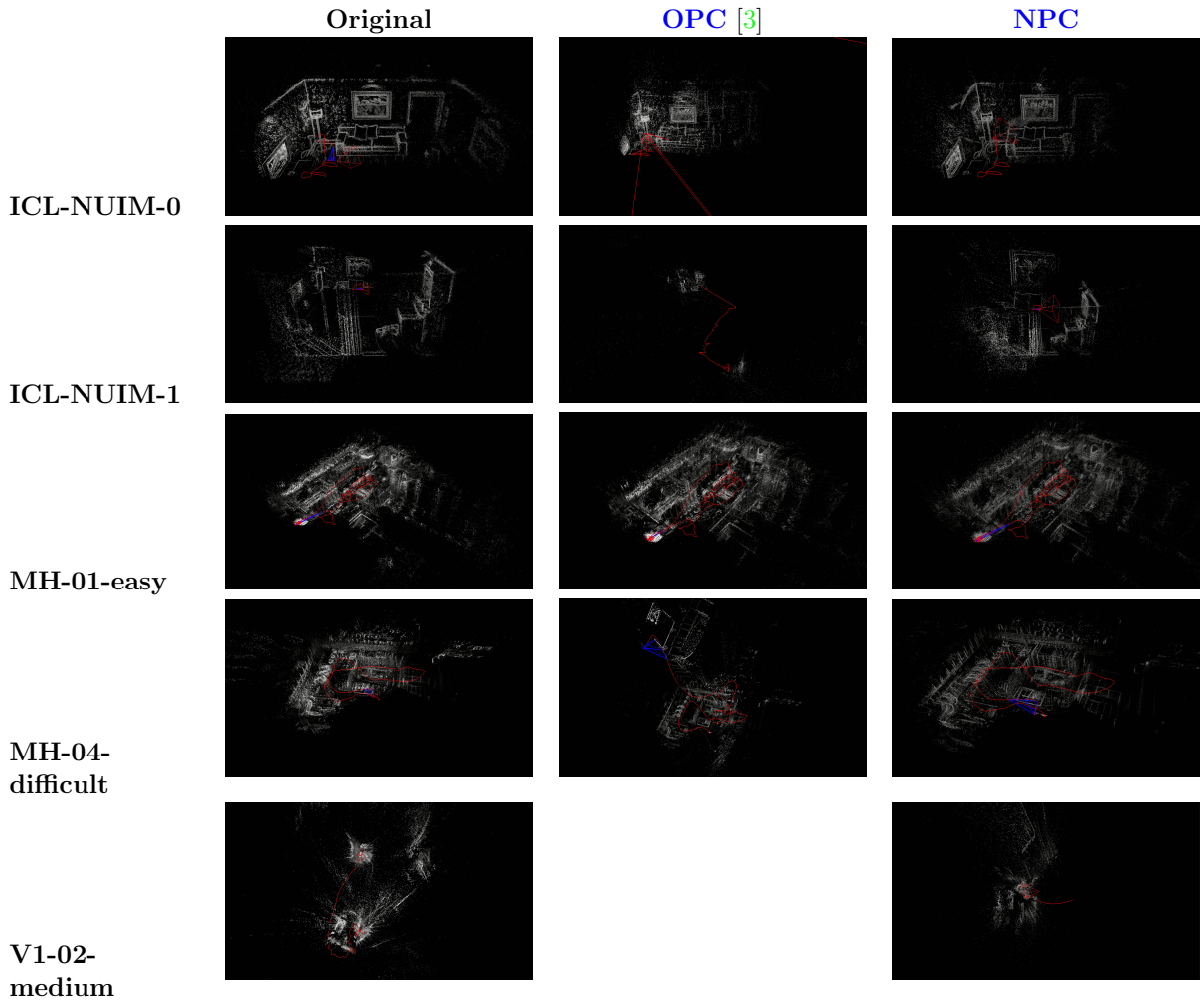


Figure 4.23: Reconstructed scenes across different sequences. DSO [11] fails to operate for OPC [3] in V1-01-medium, resulting in a blank block. Overall, NPC's reconstructed scenes are superior to those of OPC [3].

4.3 Discussion

Previous research has demonstrated that photometric calibration can enhance the performance of DSO [11] and visual SLAM [10] systems. However, our findings indicate that if the calibrated results do not closely match the ground truth data, the frame quality is often compromised. High ICRF errors result in low contrast and relatively bright images, while vignetting factors significantly degrade image quality by creating a circular rim. Additionally, drastic changes in exposure between consecutive frames lead to varying brightness levels, causing DSO [11] to lose track. Furthermore, the rough estimation of the exposure ratio for initialization may result in poor image pair selection. Additionally, since the trajectory from the MFT [56] tracker is accurate to subpixel levels, rounding the location to integer pixels can lead to inaccurate pixel intensity measurements. A better approach may involve interpolating from surrounding pixels.

Despite these challenges, OPC [3] outperforms OSPC [23] and NPC when the dataset includes ground truth data, significantly improving optimization results. However, in the absence of ground truth, our proposed method, NPC, demonstrates superior performance. OSPC [23] is unable to function without exposure ground truth, and OPC [3] produces severely flawed results. Our experiments revealed that sequential methods, OSPC [23] and NPC, struggle when connecting frames have constant exposure. Nevertheless, in high-motion scenarios such as the KITTI Dataset [17] , OPC [3] fails to operate effectively, whereas the long-term tracker, MFT [56], integrated into our model ensures that NPC can still track feature points and execute normally.

Chapter 5

Conclusion

5.1 Summary

In this thesis, we conduct an investigation into a new photometric calibration (NPC) approach. This simple method eliminates the dependency on ground truth data by exploiting pixel intensity to initialize the exposure ratio in the photometric calibration process. Additionally, building upon the sequential method, NPC effectively circumvents the exponential ambiguity problem typically associated with joint optimization methods. The study begins by assessing the efficacy of long-term feature trackers in photometrically variable conditions by integrating them into the joint optimization approach. It is discovered that long-term trackers, particularly the MFT tracker, outperform the gain robust KLT tracker in a joint optimization method. Subsequently, the MFT tracker is employed within a sequential optimization framework to address the ambiguity issue and to initialize the exposure ratio, thereby obviating the need for ground truth information regarding the camera response function and exposure values. Through extensive evaluation across diverse datasets, this research shows that while OPC [3] excels with ground truth, our proposed method, NPC, surpasses previous photometric calibration techniques without any ground truth data. The experimental results highlight the potential and feasibility of NPC in photometric calibration without ground truth. Notably, the exposure estimation of OPC [3] simply just oscillates between 0 and 1. Additionally, the average NR-IQA scores for OPC [3] and NPC are 0.2556 and 0.34433, respectively, indicating that our proposed method can retrieve higher quality frames.

5.2 Future Work

Looking ahead, this research opens several avenues for advancement. A primary objective is to integrate the tracker with our photometric calibration process to facilitate real-time application, or directly integrating NPC into visual SLAM [10] or DSO [11], thereby broadening the scope of its application. Currently, the tracker is used to gather feature point trajectories, followed by independent CRF and vignette optimization, and exposure value computation. Enhancing this workflow to operate in real-time could significantly improve practical usability. Additionally, refining the optimization process remains a critical goal. A robust method for selecting image pairs and a more effective optimization for retrieving the vignetting factor can enhance the reconstructed frames. The optimization process is activated only when the exposure ratio of the image pairs is sufficiently large or small. Our method uses the pixel intensity of corresponding points to calculate the exposure ratio. However, this approach may have some uncertainties. For example, depth is not considered, which can lead to inaccurate estimations and poor selection of image pairs.

Filtering outliers of corresponding points may also improve the results. Developing a finer approach for estimating exposure values under constant exposure in the input data is another important task. Moreover, combining the proposed method with a lighting model, such as ATF [1] model, may further refine the reconstructed frames. Alternatively, the image can be improved by rectifying small regions that are too bright or too dark, such as when capturing video through a door or window. This may be done by using adaptive histogram equalization [76] with in a small region.

References

[1] Zohair Al-Ameen. Adapted type-ii fuzzy algorithm to process images with non-uniform illumination. *Signal, Image and Video Processing*, pages 1–14, 2024.

[2] Sergey V Alexandrov, Johann Prankl, Michael Zillich, and Markus Vincze. Calibration and correction of vignetting effects with an application to 3d mapping. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4217–4223. IEEE, 2016.

[3] Paul Bergmann, Rui Wang, and Daniel Cremers. Online photometric calibration of auto exposure video for realtime visual odometry and slam. *IEEE Robotics and Automation Letters*, 3(2):627–634, 2017.

[4] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10):1157–1163, 2016.

[5] Young-Chang Chang and John F Reid. Rgb calibration for color image analysis in machine vision. *IEEE Transactions on image processing*, 5(10):1414–1422, 1996.

[6] Chaofeng Chen and Jiadi Mo. IQA-PyTorch: Pytorch toolbox for image quality assessment. [Online]. Available: https://github.com/chaofengc/IQA-PyTorch, 2022.

[7] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 643–652. 2023.

[8] Nilanjan Dey. Uneven illumination correction of digital images: A survey of the state-of-the-art. *Optik*, 183:483–495, 2019.

[9] Carl Doersch, Yi Yang, Mel Vecerik, Dilara Gokay, Ankush Gupta, Yusuf Aytar, Joao Carreira, and Andrew Zisserman. Tapir: Tracking any point with per-frame initialization and temporal refinement. *arXiv preprint arXiv:2306.08637*, 2023.

[10] Hugh Durrant-Whyte and Tim Bailey. Simultaneous localization and mapping: part i. *IEEE robotics & automation magazine*, 13(2):99–110, 2006.

[11] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):611–625, 2017.

[12] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *European conference on computer vision*, pages 834–849. Springer, 2014.

[13] Jakob Engel, Jörg Stückler, and Daniel Cremers. Large-scale direct slam with stereo cameras. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 1935–1942. IEEE, 2015.

[14] Jakob Engel, Vladyslav Usenko, and Daniel Cremers. A photometrically calibrated benchmark for monocular visual odometry. *arXiv preprint arXiv:1607.02555*, 2016.

[15] Fernando A Fardo, Victor H Conforto, Francisco C de Oliveira, and Paulo S Rodrigues. A formal evaluation of psnr as quality measurement parameter for image segmentation algorithms. *arXiv preprint arXiv:1605.07116*, 2016.

[16] Tulsi Pawan Fowdur and Lavesh Babooram. *Video Quality Assessment*, pages 97–140. Apress, Berkeley, CA, 2024.

[17] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012.

[18] Daniel B Goldman. Vignette and exposure calibration and compensation. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2276–2288, 2010.

[19] Michael D Grossberg and Shree K Nayar. Determining the camera response from images: What is knowable? *IEEE Transactions on pattern analysis and machine intelligence*, 25(11):1455–1467, 2003.

[20] Michael D Grossberg and Shree K Nayar. What is the space of camera response functions? In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, pages II–602. IEEE, 2003.

[21] Matthias Grundmann, Chris McClanahan, Sing Bing Kang, and Irfan Essa. Post-processing approach for radiometric self-calibration of video. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2013.

[22] Michael Grupp. evo: Python package for the evaluation of odometry and slam. https://github.com/MichaelGrupp/evo, 2017.

[23] Jawad Haidar, Douaa Khalil, and Daniel Asmar. Ospc: Online sequential photometric calibration. *arXiv preprint arXiv:2305.17673*, 2023.

[24] Youngsoo Han, Changhyeon Kim, Youngseok Jang, and H Jin Kim. Parametric analysis of klt algorithm in autonomous driving. In *2020 20th International Conference on Control, Automation and Systems (ICCAS)*, pages 184–189. IEEE, 2020.

[25] Ankur Handa, Thomas Whelan, John McDonald, and Andrew J Davison. A benchmark for rgb-d visual odometry, 3d reconstruction and slam. In *2014 IEEE international conference on Robotics and automation (ICRA)*, pages 1524–1531. IEEE, 2014.

[26] Adam W Harley, Zhaoyuan Fang, and Katerina Fragkiadaki. Particle video revisited: Tracking through occlusions using point trajectories. In *European Conference on Computer Vision*, pages 59–75. Springer, 2022.

[27] Berthold Horn. *Robot vision*. MIT press, 1986.

[28] Liying Hu, Lingling Sun, Yucong Wang, Keqiang Yue, Zhenghui Li, and Zehao Yan. Online photometric calibration of optical flow visual-inertial slam system. In *2020 12th International Conference on Communication Software and Networks (ICCSN)*, pages 13–17. IEEE, 2020.

[29] Bowen Huang, Kai Xiao, and Haibo Yuan. Photometric calibration methods for wide-field photometric surveys. *arXiv preprint arXiv:2206.01007*, 2022.

[30] Yongqing Huo and Xudong Zhang. Single image-based hdr image generation with camera response function estimation. *IET Image Processing*, 11(12):1317–1324, 2017.

[31] Takeo Kanade. An iterative image registration technique with an application to stereo vision (ijcai).

[32] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1733–1740, 2014.

[33] Patrick L Kelly, Anja von der Linden, Douglas E Applegate, Mark T Allen, Steven W Allen, Patricia R Burchat, David L Burke, Harald Ebeling, Peter Capak, Oliver Czoske, et al. Weighing the giants–ii. improved calibration of photometry from stellar colours and accurate photometric redshifts. *Monthly Notices of the Royal Astronomical Society*, 439(1):28–47, 2014.

[34] Christian Kerl, Mohamed Souiai, Jürgen Sturm, and Daniel Cremers. Towards illumination-invariant 3d reconstruction using tof rgb-d cameras. In *2014 2nd International Conference on 3D Vision*, volume 1, pages 39–46. IEEE, 2014.

[35] Christian Kerl, Jürgen Sturm, and Daniel Cremers. Dense visual slam for rgb-d cameras. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2100–2106. IEEE, 2013.

[36] Seon Joo Kim, David Gallup, Jan-Michael Frahm, and Marc Pollefeys. Joint radiometric calibration and feature tracking system with an application to stereo. *Computer Vision and Image Understanding*, 114(5):574–582, 2010.

[37] Seon Joo Kim and Marc Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE transactions on pattern analysis and machine intelligence*, 30(4):562–576, 2008.

[38] Grzegorz Krawczyk, Michael Goesele, and Hans-Peter Seidel. Photometric calibration of high dynamic range cameras. 2005.

[39] Songnan Li, Fan Zhang, Lin Ma, and King Ngi Ngan. Image quality assessment by separately evaluating detail losses and additive impairments. *IEEE Transactions on Multimedia*, 13(5):935–949, 2011.

[40] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara. Toward a practical perceptual video quality metric. [Online]. Available: https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652, 2016.

[41] Zhi Li, Christos Bampis, Julie Novak, Anne Aaron, Kyle Swanson, Anush Moorthy, and JD Cock. Vmaf: The journey continues. *Netflix Technology Blog*, 25(1), 2018.

[42] Stephen Lin, Jinwei Gu, Shuntaro Yamazaki, and Heung-Yeung Shum. Radiometric calibration from a single image. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–II. IEEE, 2004.

[43] Stephen Lin and Lei Zhang. Determining the radiometric response function from a single grayscale image. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 66–73. IEEE, 2005.

[44] Anatoly Litvinov and Yoav Y Schechner. Addressing radiometric nonidealities: A unified framework. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 52–59. IEEE, 2005.

[45] Anatoly Litvinov and Yoav Y Schechner. Radiometric framework for image mosaicking. *JOSA A*, 22(5):839–848, 2005.

[46] Peixin Liu, Xianfeng Yuan, Chengjin Zhang, Yong Song, Chuanzheng Liu, and Ziyan Li. Real-time photometric calibrated monocular direct visual slam. *Sensors*, 19(16):3604, 2019.

[47] Quanpan Liu, Zhengjie Wang, and Huan Wang. Optical flow monocular visual-inertial odometry with online photometric calibration. In *Journal of Physics: Conference Series*, volume 1828, page 012163. IOP Publishing, 2021.

[48] Tsung-Jung Liu, Yu-Chieh Lin, Weisi Lin, and C-C Jay Kuo. Visual quality assessment: recent developments, coding applications and future trends. *APSIPA Transactions on Signal and Information Processing*, 2:e4, 2013.

[49] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999.

[50] Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81: 7th international joint conference on Artificial intelligence*, volume 2, pages 674–679, 1981.

[51] Dongting Luo, Yan Zhuang, and Sen Wang. Hybrid sparse monocular visual odometry with online photometric calibration. *The International Journal of Robotics Research*, 41(11-12):993–1021, 2022.

[52] Kede Ma and Yuming Fang. Image quality assessment in the modern age. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 5664–5666, 2021.

[53] Tomoo Mitsunaga and Shree K Nayar. Radiometric self calibration. In *Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149)*, volume 1, pages 374–380. IEEE, 1999.

[54] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[55] Zhipeng Mo, Boxin Shi, Sai-Kit Yeung, and Yasuyuki Matsushita. Ambiguity-free radiometric calibration for internet photo collections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7):1670–1684, 2019.

[56] Michal Neoral, Jonáš Šerỳch, and Jiří Matas. Mft: Long-term tracking of every pixel. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6837–6847, 2024.

[57] David Nistér, Oleg Naroditsky, and James Bergen. Visual odometry. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. Ieee, 2004.

[58] Nikhil Padmanabhan, David J Schlegel, Douglas P Finkbeiner, JC Barentine, Michael R Blanton, Howard J Brewington, James E Gunn, Michael Harvanek, David W Hogg, Željko Ivezić, et al. An improved photometric calibration of the sloan digital sky survey imaging data. *The Astrophysical Journal*, 674(2):1217, 2008.

[59] Julien Pilet, Andreas Geiger, Pascal Lagger, Vincent Lepetit, and Pascal Fua. An all-in-one solution to geometric and photometric calibration. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 69–78. IEEE, 2006.

[60] Jan Quenzel, Jannis Horn, Sebastian Houben, and Sven Behnke. Keyframe-based photometric online calibration and color correction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–8. IEEE, 2018.

[61] Reza Rassool. Vmaf reproducibility: Validating a perceptual practical video quality metric. In *2017 IEEE international symposium on broadband multimedia systems and broadcasting (BMSB)*, pages 1–2. IEEE, 2017.

[62] P Sand. *Long-range video motion estimation using point trajectories*. PhD thesis, Ph. D. dissertation, Cambridge, MA, USA, 2006, adviser-Teller, Seth.

[63] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on image processing*, 15(2):430–444, 2006.

[64] Jianbo Shi et al. Good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, pages 593–600. IEEE, 1994.

[65] Guoqing Sun, Haibo Zhou, Chenming Li, and Yijiao Zhao. Research on vision detection technology based on photometric calibration. In *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, pages 138–144. IEEE, 2021.

[66] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Nature, 2022.

[67] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011, 2018.

[68] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2555–2563, 2023.

[69] 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.

[70] Nan Yang, Rui Wang, Xiang Gao, and Daniel Cremers. Challenges in monocular visual odometry: Photometric calibration, motion bias, and rolling shutter effect. *IEEE Robotics and Automation Letters*, 3(4):2878–2885, 2018.

[71] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1191–1200, 2022.

[72] Khalid Yousif, Alireza Bab-Hadiashar, and Reza Hoseinnezhad. An overview to visual odometry and visual slam: Applications to mobile robotics. *Intelligent Industrial Systems*, 1(4):289–311, 2015.

[73] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[74] Zhengyou Zhang. *Camera Calibration*, pages 76–77. Springer US, Boston, MA, 2014.

[75] Yuanjie Zheng, Stephen Lin, Chandra Kambhamettu, Jingyi Yu, and Sing Bing Kang. Single-image vignetting correction. *IEEE transactions on pattern analysis and machine intelligence*, 31(12):2243–2256, 2008.

[76] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. 1994.

APPENDICES

Appendix A

Supplement results

A.1 Stats for Photometric Calibration on TUM MONO Dataset

Table A.1: RMSE of the photometric parameters on TUM MONO Dataset [14].

Result (RMSE)		OPC	OSPC	NPC
sq07	ICRF	**3.5971**	28.6887	4.7418
	Vignette	**0.1075**	0.3866	0.5777
	Exposure	**0.0500**	0.1234	0.3243
sq08	ICRF	**3.5954**	12.9026	5.2302
	Vignette	0.1112	0.1225	**0.0913**
	Exposure	**0.0382**	0.1513	0.1509
sq16	ICRF	51.6623	5.4498	**4.0562**
	Vignette	**0.0669**	0.0925	0.1402
	Exposure	**0.0686**	0.1361	0.2736
sq23	ICRF	**4.4659**	26.3571	55.2017
	Vignette	**0.1307**	0.4443	0.2916
	Exposure	**0.0646**	0.3511	0.2674
sq25	ICRF	**4.6302**	25.1617	6.0630
	Vignette	0.0793	0.1561	**0.0772**
	Exposure	**0.0688**	0.3327	0.3578

Continued on next page

Table A.1 – continued from previous page

Result (RMSE)		OPC	OSPC	NPC
sq26	ICRF	**3.8127**	9.2874	8.3557
	Vignette	**0.0584**	0.1692	0.1167
	Exposure	**0.1116**	0.4671	0.6032
sq28	ICRF	**6.7164**	13.9268	50.1849
	Vignette	**0.0705**	0.1313	0.2402
	Exposure	**0.1146**	0.2871	0.1803
sq29	ICRF	**5.4574**	14.9362	6.2502
	Vignette	0.1313	0.0825	**0.0715**
	Exposure	**0.0235**	0.1910	0.1424
sq30	ICRF	**3.6632**	5.9996	4.4504
	Vignette	0.1430	**0.0246**	0.0252
	Exposure	**0.0566**	0.2017	0.3237
sq31	ICRF	**3.5359**	6.4124	45.5713
	Vignette	0.0831	0.0758	**0.0629**
	Exposure	**0.0561**	0.2524	0.2783
sq32	ICRF	**3.5906**	11.4993	56.2036
	Vignette	0.0536	0.0768	**0.0438**
	Exposure	**0.0353**	0.0785	0.1521
sq35	ICRF	5.0316	8.7054	**4.4935**
	Vignette	**0.1024**	0.1469	0.3008
	Exposure	**0.1266**	0.4508	0.4995
sq36	ICRF	**4.1253**	16.0218	5.0433
	Vignette	0.1396	0.2672	**0.1332**
	Exposure	**0.1477**	0.5355	0.7185
sq37	ICRF	**4.0162**	13.1646	4.4718
	Vignette	**0.1258**	0.1468	0.4250
	Exposure	**0.0467**	0.0694	0.3855
sq38	ICRF	**3.9994**	18.3018	52.5412
	Vignette	0.1025	**0.0658**	0..0811
	Exposure	**0.0703**	0.3453	0.3757
sq39	ICRF	**3.7599**	5.3584	55.6898
	Vignette	0.0617	**0.0223**	0.0894
	Exposure	**0.0574**	0.3440	0.3750
sq40	ICRF	**4.7290**	15.5420	5.7549

Continued on next page

Table A.1 – continued from previous page

Result (RMSE)		OPC	OSPC	NPC
	Vignette	0.0899	0.0671	**0.0522**
	Exposure	**0.0845**	0.2874	0.3226
	ICRF	**3.7096**	19.7618	5.8026
sq41	Vignette	**0.1078**	0.1720	0.1169
	Exposure	**0.0921**	0.3928	0.3373
	ICRF	**5.3675**	18.7268	12.0863
sq42	Vignette	**0.1295**	0.1890	0.1343
	Exposure	**0.0669**	0.1368	0.2191
	ICRF	**4.0377**	6.4380	58.3024
sq43	Vignette	0.0976	0.1008	**0.0659**
	Exposure	**0.1072**	0.3791	0.4460
	ICRF	**5.2442**	16.1951	5.9678
sq44	Vignette	**0.1118**	0.1291	0.1600
	Exposure	**0.0602**	0.3631	0.1019
	ICRF	**3.5947**	12.3519	48.7689
sq45	Vignette	0.0741	0.0707	**0.0674**
	Exposure	**0.1087**	0.4264	0.3410
	ICRF	**4.9487**	10.0388	21.6188
sq46	Vignette	0.0961	**0.0554**	0.0666
	Exposure	**0.0281**	0.1276	0.1486
	ICRF	**5.5471**	9.1259	10.8857
sq47	Vignette	**0.0888**	0.1340	0.4630
	Exposure	**0.0583**	0.4285	0.2878
	ICRF	**4.4952**	41.4901	63.2512
sq48	Vignette	**0.0776**	0.0805	0.1470
	Exposure	**0.0566**	0.3910	0.3136
	ICRF	**4.3427**	29.6449	71.8436
sq49	Vignette	0.0699	0.1314	**0.0558**
	Exposure	**0.0597**	0.3277	0.1496

Table A.2: VQA of photometric calibrated video on TUM MONO Dataset [14].

VQA		OPC	OSPC	NPC
sq07	VIF	0.4847	0.4933	0.8043
	Motion2	3.3255	3.3224	3.3224
	ADM2	0.8698	0.7646	0.9496
	VMAF	58.1056	14.6973	**69.6719**
sq08	VIF	0.5467	0.7703	0.8466
	Motion2	2.9168	2.9172	2.9172
	ADM2	0.9369	0.8768	0.9496
	VMAF	**75.4974**	50.2614	71.5093
sq16	VIF	0.4771	0.5756	0.4840
	Motion2	6.9247	6.9286	6.9048
	ADM2	0.9392	0.6849	0.5911
	VMAF	**69.9682**	10.7355	2.6639
sq23	VIF	0.3521	0.5813	0.4033
	Motion2	3.8211	3.8208	3.8208
	ADM2	0.8904	0.7027	0.5544
	VMAF	**62.4821**	11.1798	0.9683
sq25	VIF	0.3718	0.5779	0.7884
	Motion2	4.9489	4.9457	4.9457
	ADM2	0.9315	0.6879	0.8602
	VMAF	**72.5436**	9.7799	57.6170
sq26	VIF	0.5957	0.8551	0.8858
	Motion2	2.7539	2.7557	2.7557
	ADM2	0.9687	0.9338	0.9519
	VMAF	**81.3235**	72.0618	79.8978
sq28	VIF	0.5066	0.7984	0.6044
	Motion2	3.1383	3.1381	3.1381
	ADM2	0.8235	0.8975	0.6950
	VMAF	45.8361	**58.3720**	12.2863
sq29	VIF	0.3824	0.7111	0.7761
	Motion2	5.4805	5.4799	5.4799
	ADM2	0.8533	0.8009	0.8569
	VMAF	**56.2960**	38.7959	55.4498
sq30	VIF	0.3834	0.7101	0.8936

Continued on next page

Table A.2 – continued from previous page

Sequence	VQA	OPC	OSPC	NPC
	Motion2	7.0572	7.0576	7.0566
	ADM2	0.8204	0.9336	0.9695
	VMAF	47.6165	77.6465	**88.0520**
sq31	VIF	0.3393	0.8132	0.7325
	Motion2	8.2599	8.2637	8.2637
	ADM2	0.7278	0.9096	0.8363
	VMAF	43.9645	**69.7477**	50.5103
sq32	VIF	0.3470	0.7784	0.6223
	Motion2	8.3104	8.3114	8.3114
	ADM2	0.7770	0.8823	0.7650
	VMAF	50.3895	**60.7035**	29.9994
sq35	VIF	0.5754	1.0069	0.8020
	Motion2	1.9580	1.9571	1.9571
	ADM2	0.8750	1.0877	0.9602
	VMAF	54.9077	**94.5136**	70.7609
sq36	VIF	0.6102	0.8333	0.8760
	Motion2	2.6153	2.6166	2.6166
	ADM2	0.9404	0.8991	0.9776
	VMAF	74.1229	60.1337	**81.4603**
sq37	VIF	0.5166	0.7577	0.6976
	Motion2	3.1479	3.1469	3.1469
	ADM2	0.8102	0.8522	0.8499
	VMAF	46.6653	**49.8562**	38.9640
sq38	VIF	0.5910	0.7759	0.7574
	Motion2	3.2813	3.2821	3.2821
	ADM2	0.9453	0.8776	0.8467
	VMAF	**78.0844**	52.1959	47.9321
sq39	VIF	0.5711	0.9042	0.6969
	Motion2	3.1270	3.1270	3.1270
	ADM2	0.9041	0.9873	0.8057
	VMAF	63.2309	**87.0259**	35.2224
sq40	VIF	0.5918	0.8161	0.8928
	Motion2	3.0102	3.0107	3.0107
	ADM2	0.9253	0.9075	0.9951

Continued on next page

Table A.2 – continued from previous page

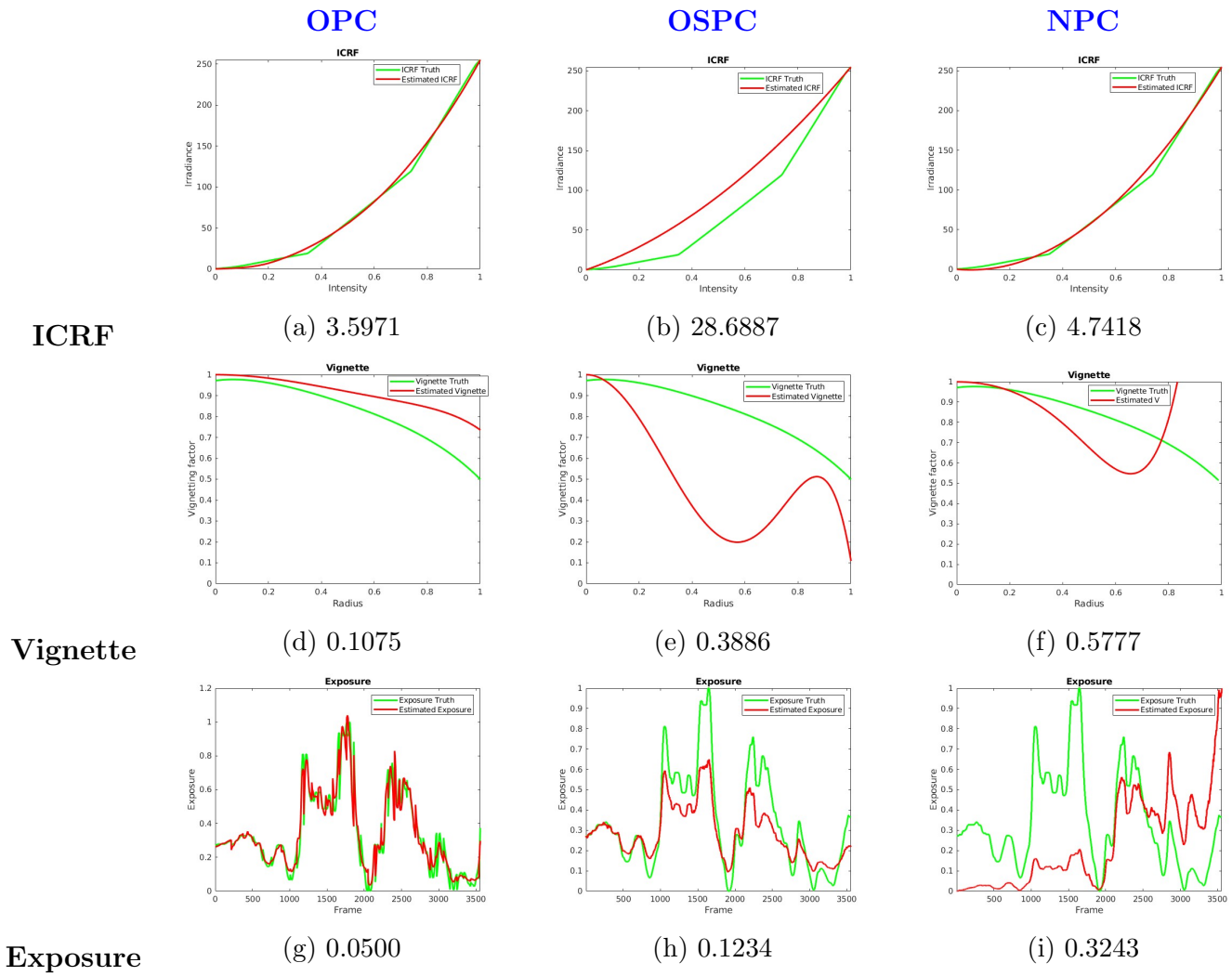
Sequence	VQA	OPC	OSPC	NPC
	VMAF	71.5114	62.9259	**83.5297**
sq41	VIF	0.5428	0.7554	0.8979
	Motion2	4.5874	4.5856	4.5856
	ADM2	0.9637	0.8448	0.9970
	VMAF	82.0078	44.8674	**88.8116**
sq42	VIF	0.4528	0.6854	0.7444
	Motion2	4.7501	4.7482	4.7482
	ADM2	0.8995	0.7834	0.8409
	VMAF	**68.3426**	31.9800	46.7014
sq42	VIF	0.4528	0.6854	0.7444
	Motion2	4.7501	4.7482	4.7482
	ADM2	0.8995	0.7834	0.8409
	VMAF	**68.3426**	31.9800	46.7014
sq43	VIF	0.3653	0.7601	0.5664
	Motion2	5.8276	5.8294	5.8294
	ADM2	0.9127	0.8465	0.6814
	VMAF	**74.8701**	52.3753	12.3817
sq44	VIF	0.3722	0.6996	0.72234
	Motion2	5.2107	5.2132	5.2132
	ADM2	0.7377	0.7831	0.8089
	VMAF	**50.9350**	34.0774	42.2243
sq45	VIF	0.3517	0.7935	0.6459
	Motion2	5.7505	5.7534	5.7534
	ADM2	0.9466	0.8856	0.7689
	VMAF	**78.1480**	58.8021	28.7772
sq46	VIF	0.4071	0.7416	0.6423
	Motion2	4.5016	4.5027	4.5068
	ADM2	0.8754	0.8243	0.7318
	VMAF	**60.1162**	43.8387	22.0238
sq47	VIF	0.3675	0.7099	0.6152
	Motion2	7.3456	7.3445	7.3445
	ADM2	0.8213	0.8199	0.7361
	VMAF	**73.8696**	46.664	21.1244
	VIF	0.3975	0.6344	0.4388
sq48				

Continued on next page

Table A.2 – continued from previous page

Sequence	VQA	OPC	OSPC	NPC
	Motion2	8.2104	8.2109	8.2109
	ADM2	0.9243	0.7359	0.5555
	VMAF	**77.7335**	23.8918	0.7896
sq49	VIF	0.3842	0.6075	0.4883
	Motion2	5.4475	5.4456	5.4456
	ADM2	0.7494	0.7103	0.6086
	VMAF	**40.4326**	15.1393	1.8967

The following plots show the output of three parameters across three methods on TUM MONO Dataset [14]. The green line is the ground truth and the red line is the estimated result. The number next to index is RMSE value.



ICRF

(a) 3.5971 (b) 28.6887 (c) 4.7418

Vignette

(d) 0.1075 (e) 0.3886 (f) 0.5777

Exposure

(g) 0.0500 (h) 0.1234 (i) 0.3243

Figure A.1: Sequence 07.

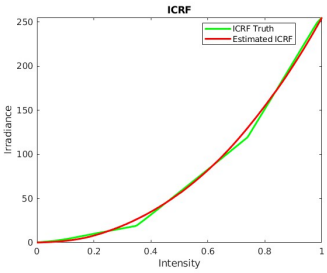
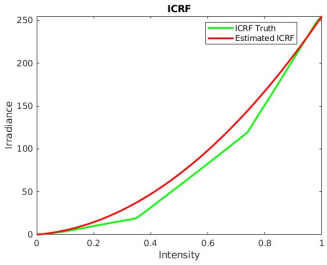
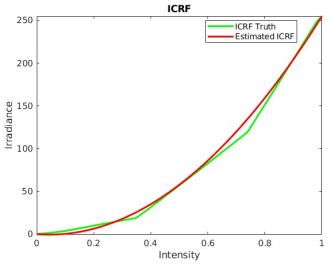
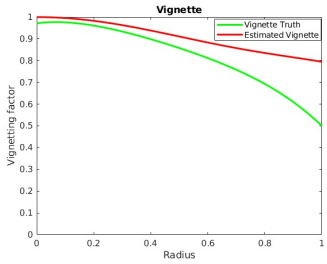
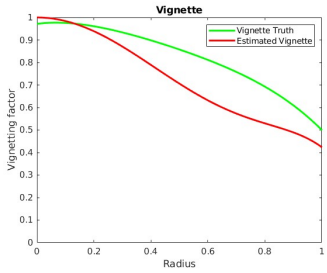
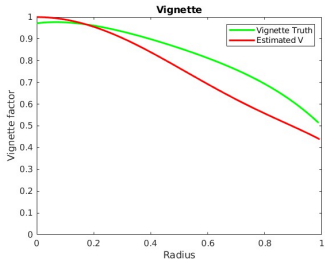
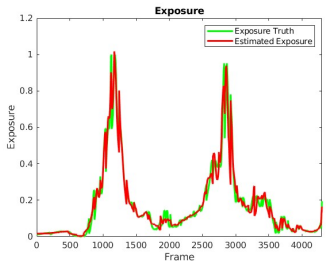
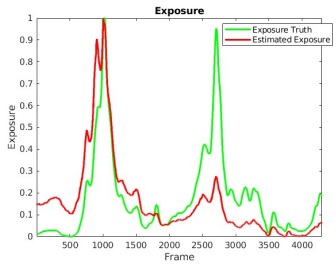
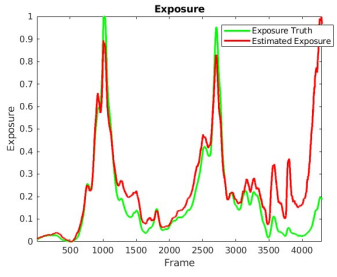
	OPC	OSPC	NPC
ICRF	 (a) 3.5954	 (b) 12.9026	 (c) 5.2302
Vignette	 (d) 0.1112	 (e) 0.1225	 (f) 0.0913
Exposure	 (g) 0.0382	 (h) 0.1513	 (i) 0.1509

Figure A.2: Sequence 08.

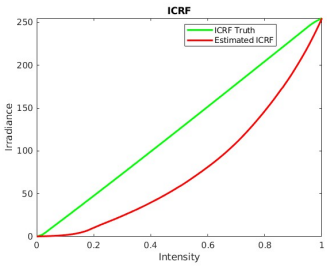
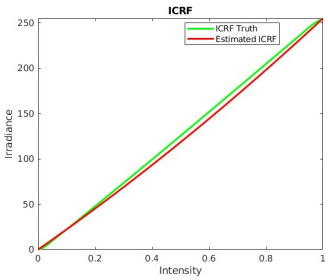
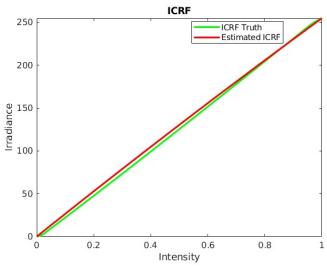
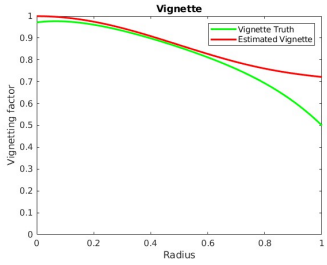
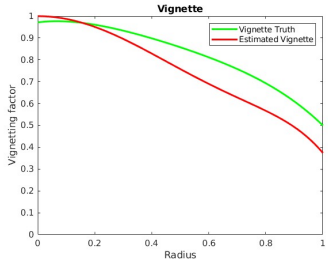
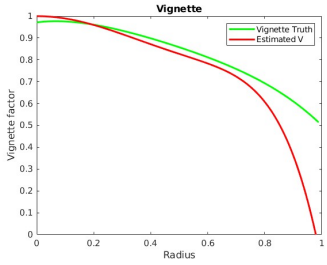
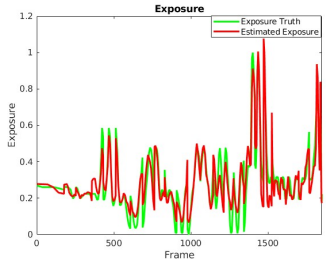
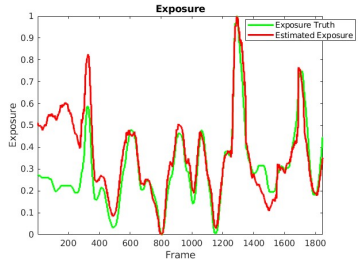
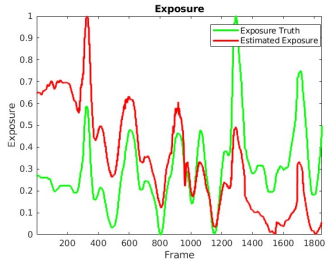
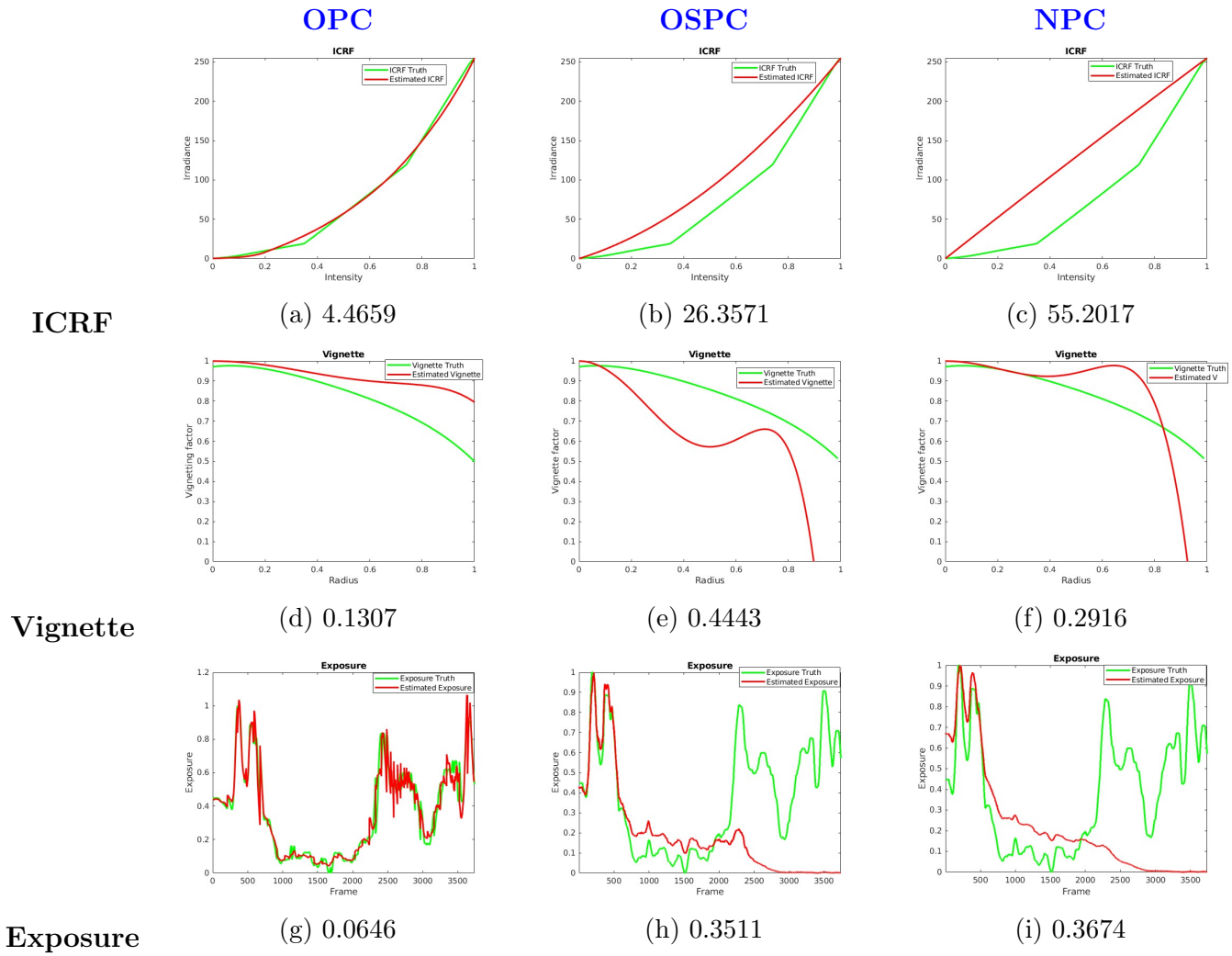
	OPC	OSPC	NPC
ICRF	 (a) 51.6623	 (b) 5.4498	 (c) 4.0562
Vignette	 (d) 0.0669	 (e) 0.0925	 (f) 0.1402
Exposure	 (g) 0.0686	 (h) 0.1361	 (i) 0.2736

Figure A.3: Sequence 16.



ICRF

(a) 4.4659 (b) 26.3571 (c) 55.2017

Vignette

(d) 0.1307 (e) 0.4443 (f) 0.2916

Exposure

(g) 0.0646 (h) 0.3511 (i) 0.3674

Figure A.4: Sequence 23.

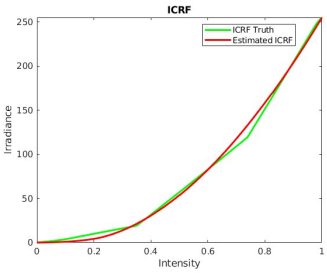
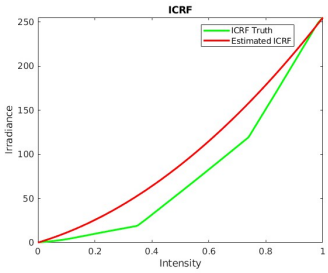
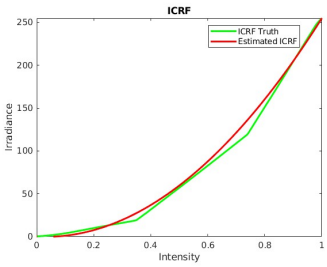
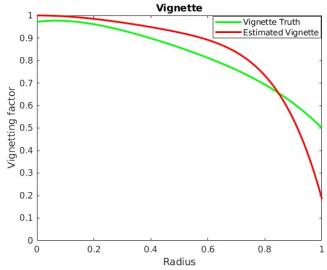
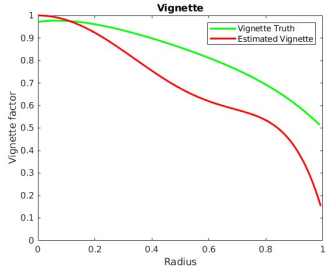
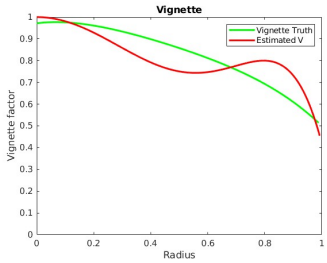
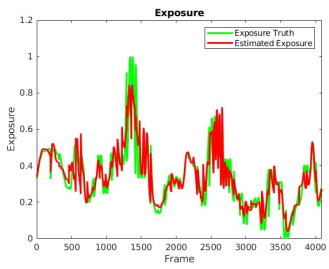
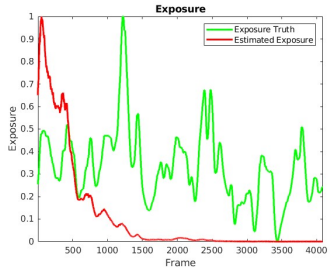
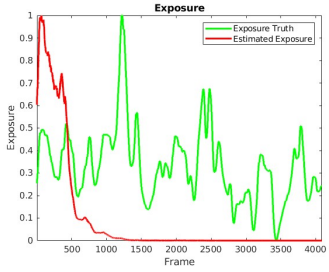
	OPC	OSPC	NPC
ICRF	 (a) 4.6302	 (b) 25.1617	 (c) 6.0630
Vignette	 (d) 0.0793	 (e) 0.1561	 (f) 0.0772
Exposure	 (g) 0.0688	 (h) 0.3327	 (i) 0.3578

Figure A.5: Sequence 25.

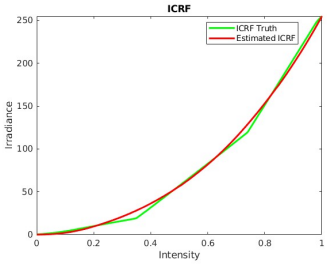
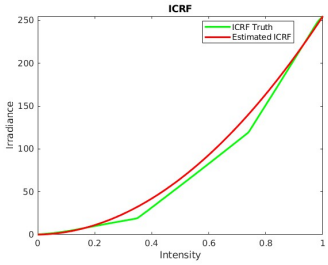
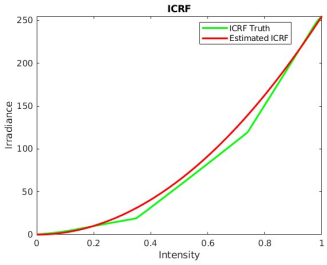
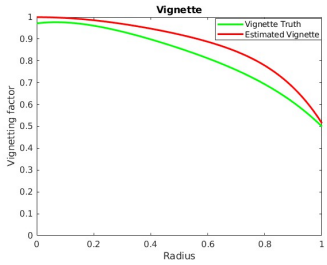
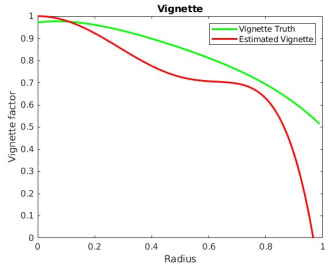
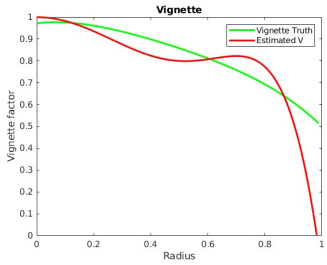
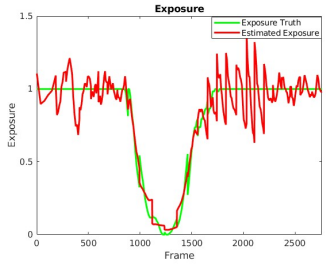
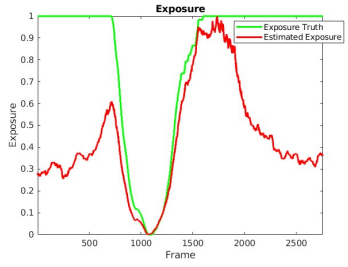
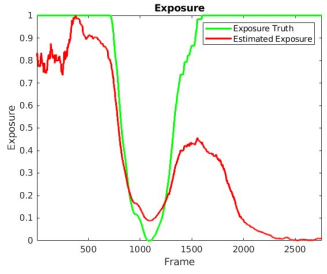
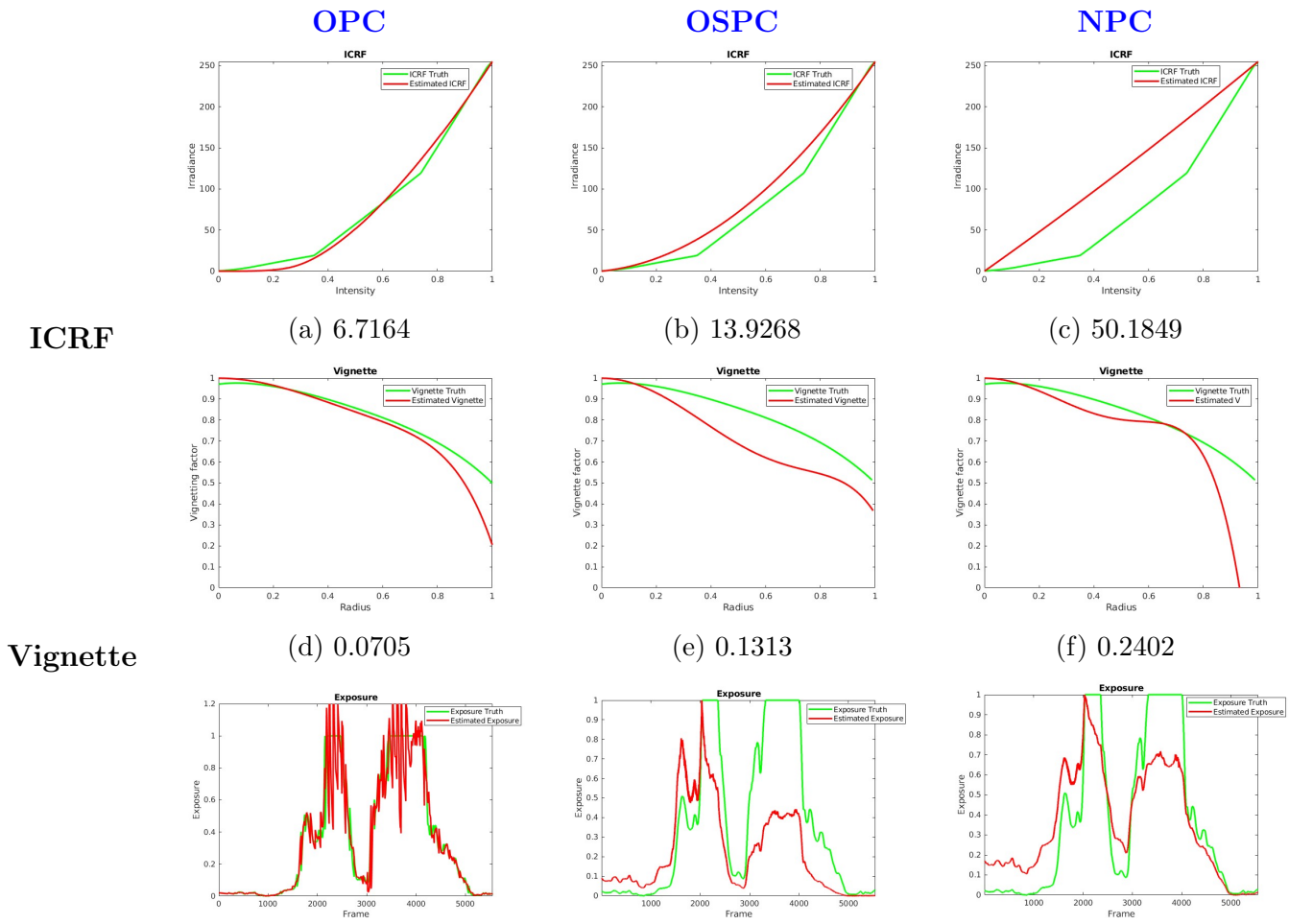
	OPC	OSPC	NPC
ICRF	 (a) 3.8127	 (b) 9.2874	 (c) 8.3557
Vignette	 (d) 0.0584	 (e) 0.1692	 (f) 0.1167
Exposure	 (g) 0.1116	 (h) 0.4671	 (i) 0.6032

Figure A.6: Sequence 26.



(a) 6.7164 (b) 13.9268 (c) 50.1849

(d) 0.0705 (e) 0.1313 (f) 0.2402

(g) 0.1146 (h) 0.2871 (i) 0.1803

Figure A.7: Sequence 28.

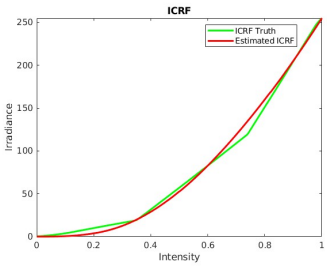
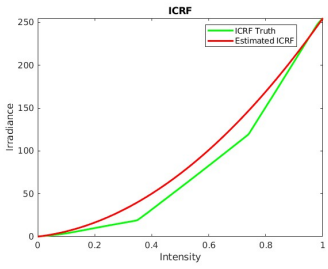
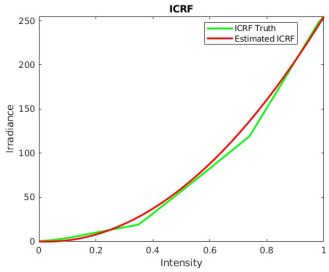
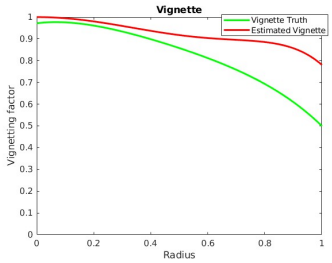
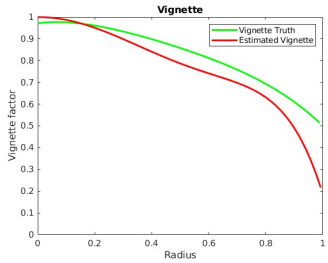
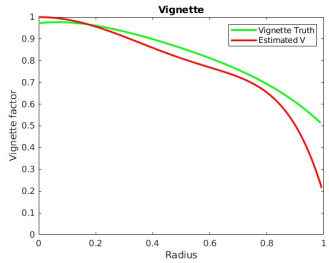
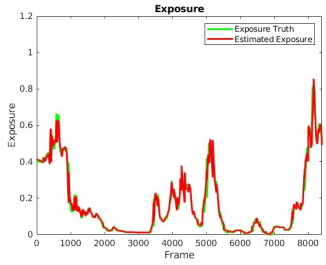
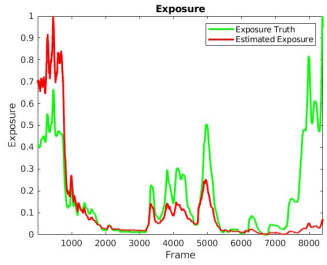
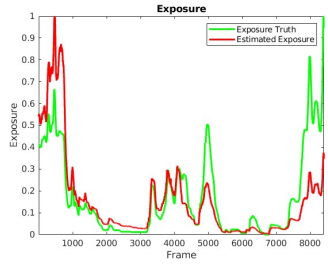
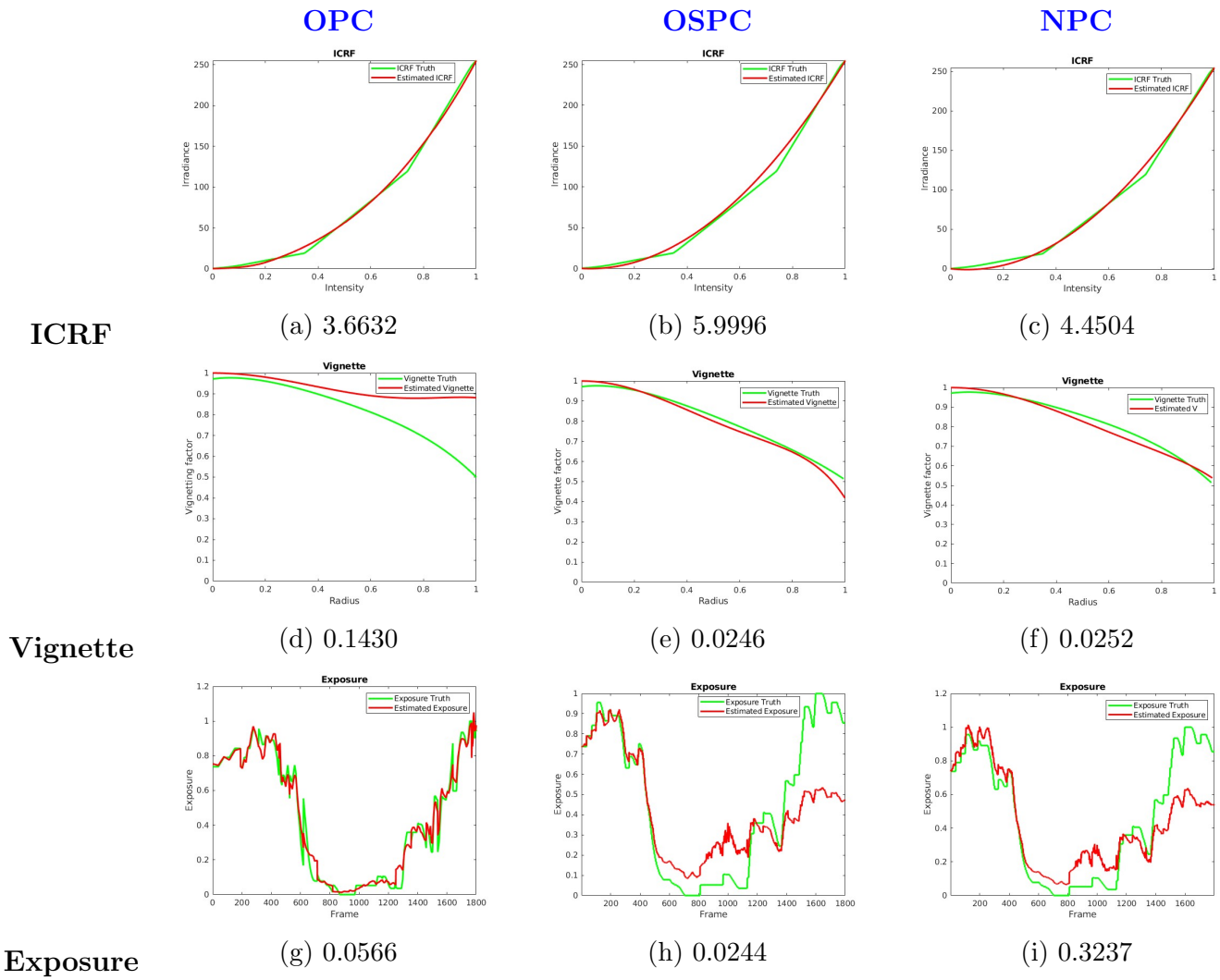
	OPC	OSPC	NPC
ICRF			
	(a) 5.4574	(b) 14.9362	(c) 6.2502
Vignette			
	(d) 0.1313	(e) 0.0825	(f) 0.0715
Exposure			
	(g) 0.0235	(h) 0.1910	(i) 0.1424

Figure A.8: Sequence 29.



(a) 3.6632 (b) 5.9996 (c) 4.4504

(d) 0.1430 (e) 0.0246 (f) 0.0252

(g) 0.0566 (h) 0.0244 (i) 0.3237

Figure A.9: Sequence 30.

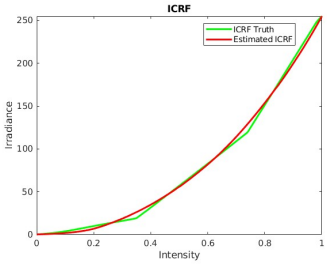
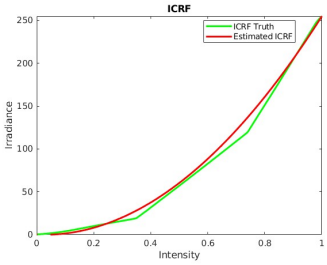
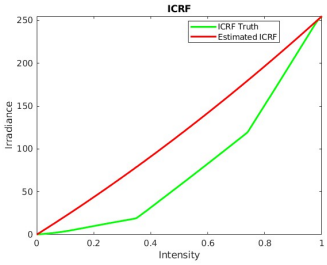
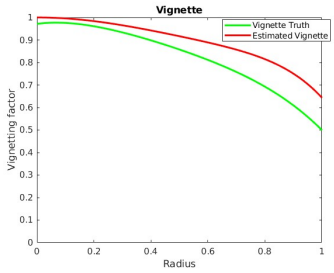
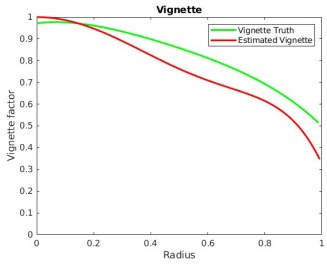
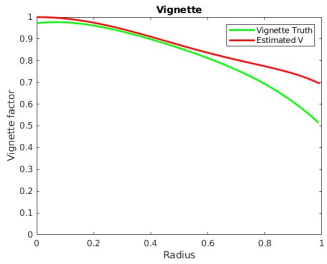
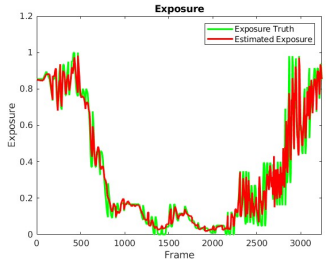
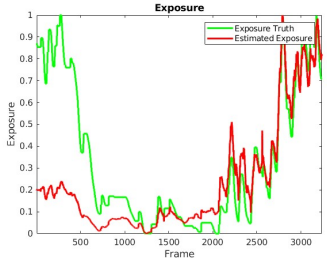
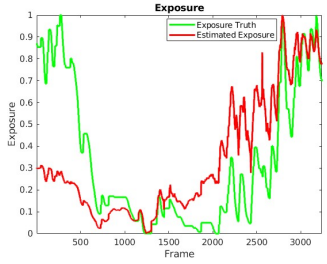
	OPC	OSPC	NPC
ICRF	 (a) 3.5359	 (b) 6.4124	 (c) 45.5713
Vignette	 (d) 0.0831	 (e) 0.0785	 (f) 0.0629
Exposure	 (g) 0.0561	 (h) 0.2524	 (i) 0.2783

Figure A.10: Sequence 31.

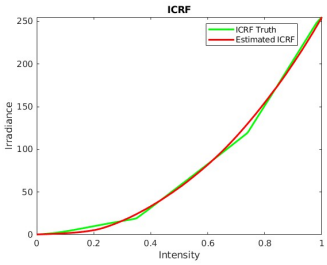
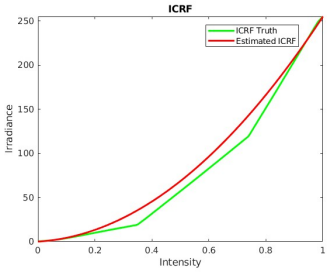
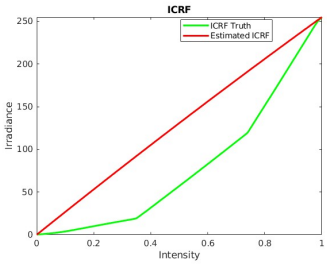
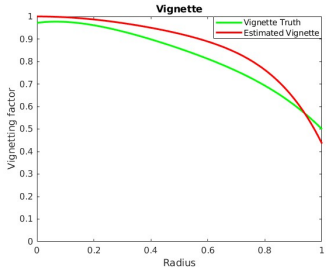
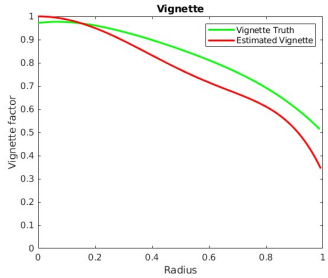
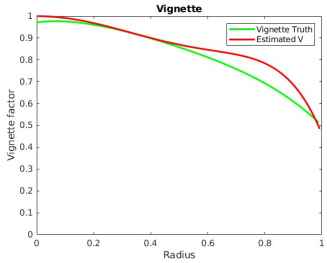
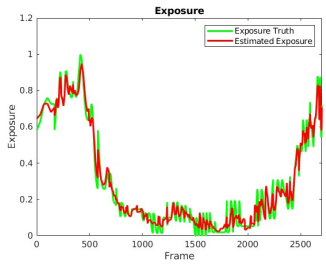
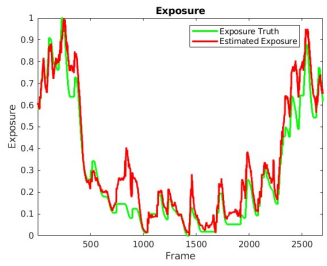
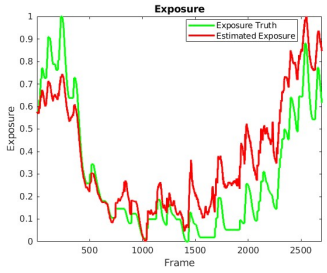
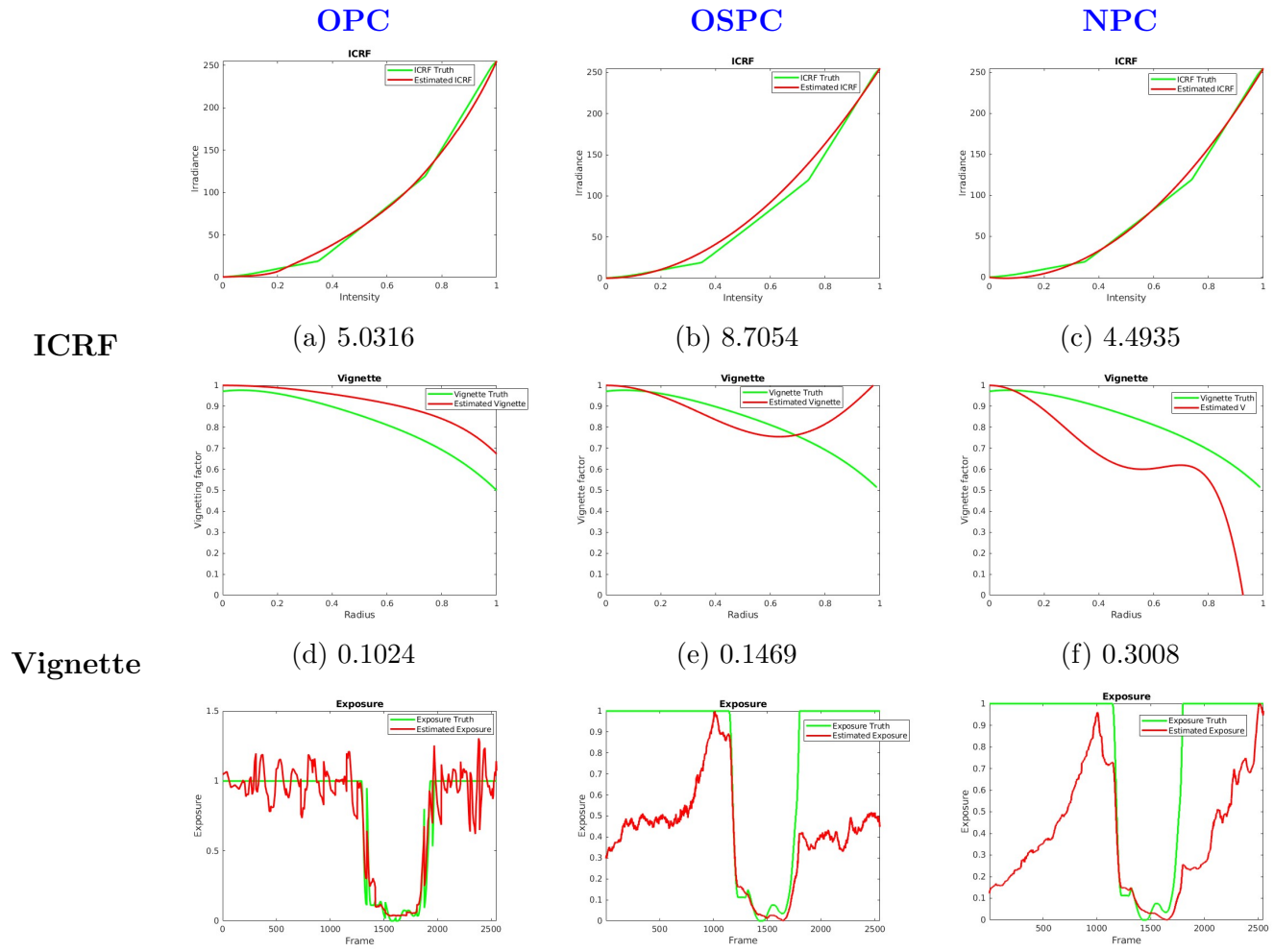
	OPC	OSPC	NPC
ICRF	 (a) 3.5906	 (b) 11.4993	 (c) 56.2036
Vignette	 (d) 0.0536	 (e) 0.0768	 (f) 0.0438
Exposure	 (g) 0.0353	 (h) 0.0785	 (i) 0.1521

Figure A.11: Sequence 32.

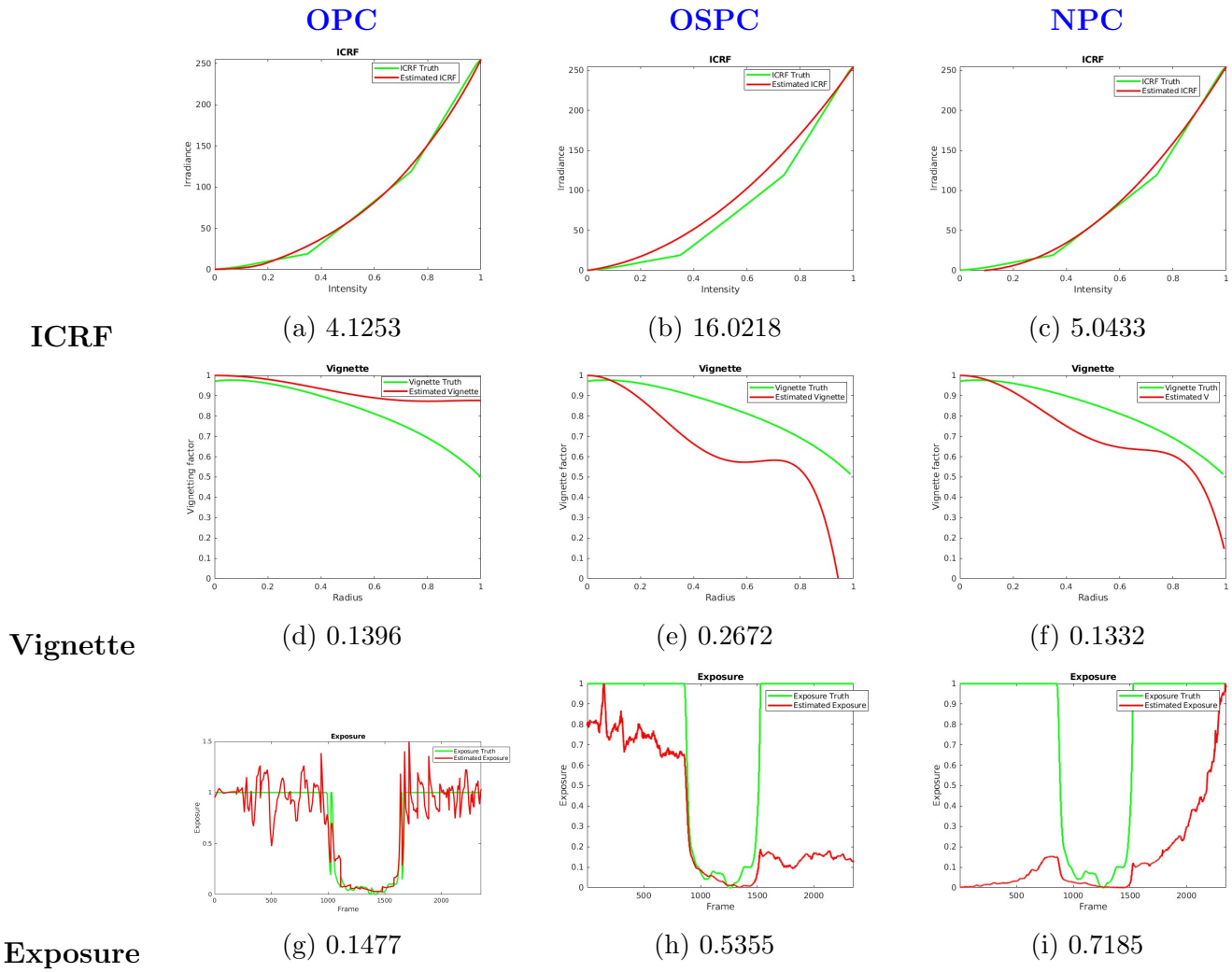


(a) 5.0316 (b) 8.7054 (c) 4.4935

(d) 0.1024 (e) 0.1469 (f) 0.3008

(g) 0.1266 (h) 0.4508 (i) 0.4995

Figure A.12: Sequence 35.



(a) 4.1253 (b) 16.0218 (c) 5.0433

(d) 0.1396 (e) 0.2672 (f) 0.1332

(g) 0.1477 (h) 0.5355 (i) 0.7185

Figure A.13: Sequence 36.

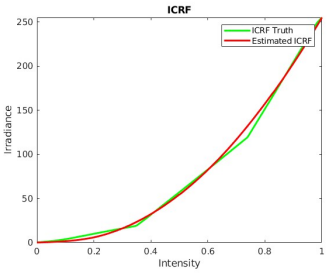
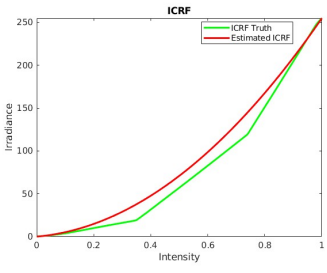
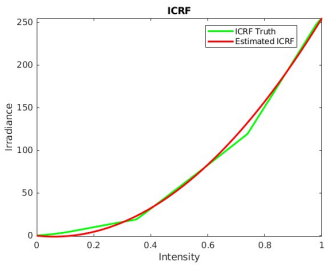
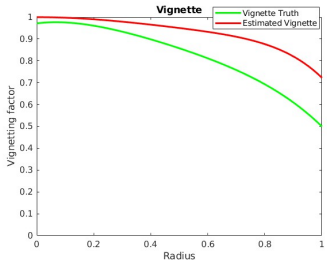
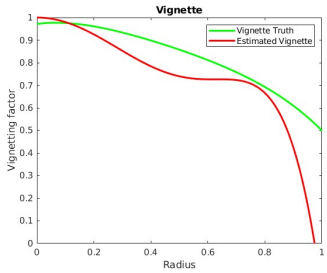
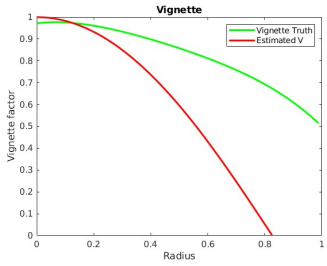
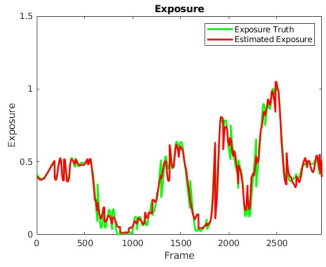
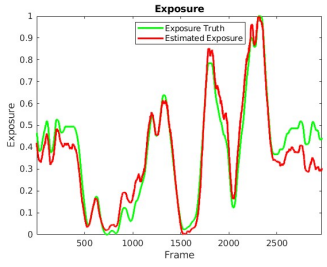
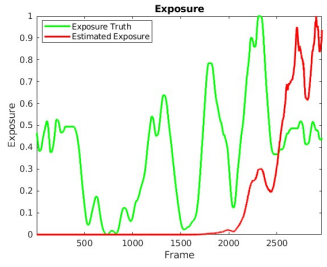
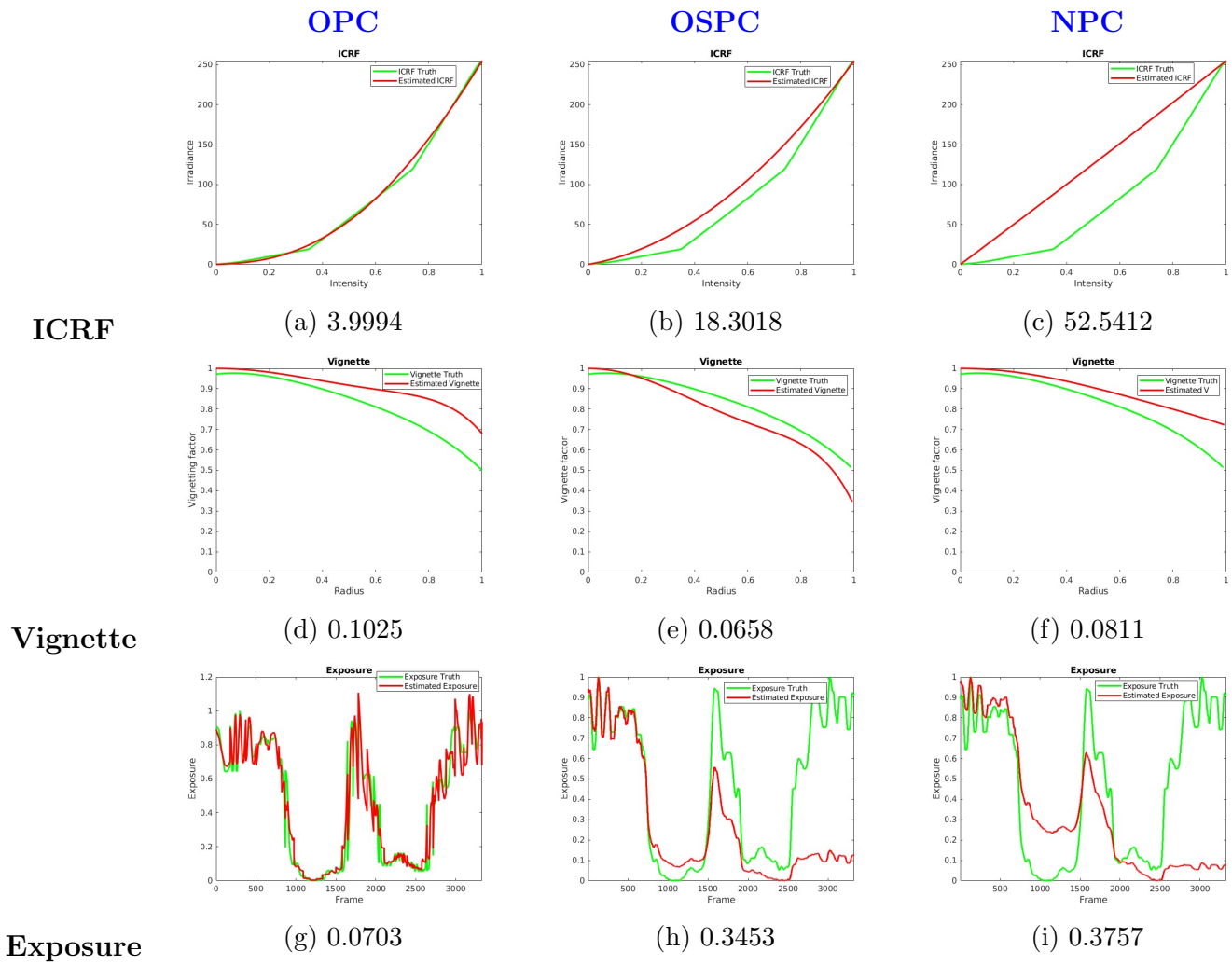
	OPC	OSPC	NPC
ICRF	 (a) 4.0162	 (b) 13.1646	 (c) 4.4718
Vignette	 (d) 0.1258	 (e) 0.1468	 (f) 0.4250
Exposure	 (g) 0.0469	 (h) 0.0694	 (i) 0.3855

Figure A.14: Sequence 37.



ICRF

(a) 3.9994 (b) 18.3018 (c) 52.5412

Vignette

(d) 0.1025 (e) 0.0658 (f) 0.0811

Exposure

(g) 0.0703 (h) 0.3453 (i) 0.3757

Figure A.15: Sequence 38.

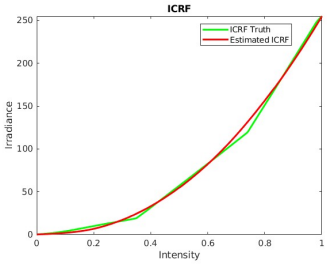
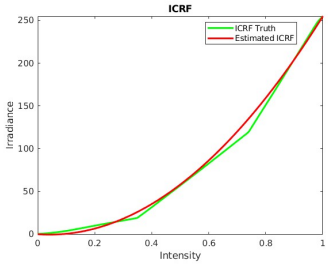
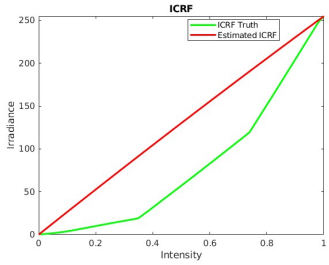
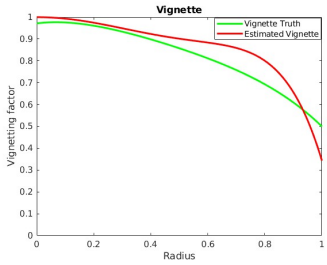
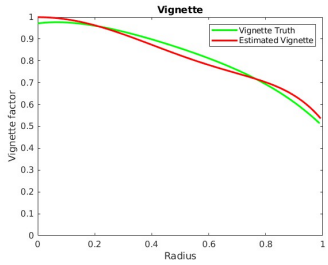
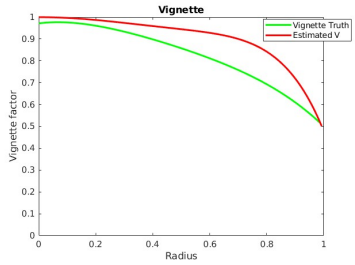
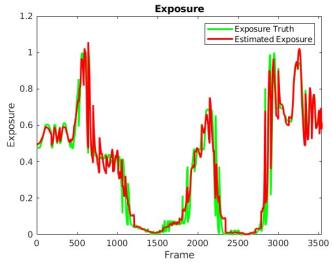
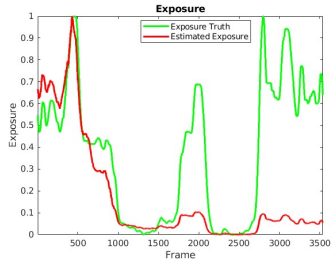
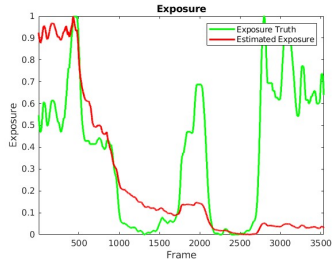
	OPC	OSPC	NPC
ICRF			
	(a) 3.7599	(b) 5.3584	(c) 55.6898
Vignette			
	(d) 0.0617	(e) 0.0223	(f) 0.0894
Exposure			
	(g) 0.0574	(h) 0.3474	(i) 0.3750

Figure A.16: Sequence 39.

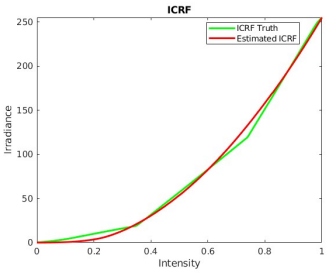
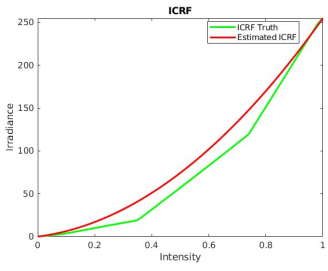
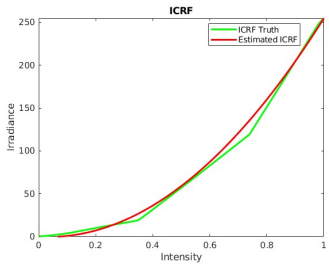
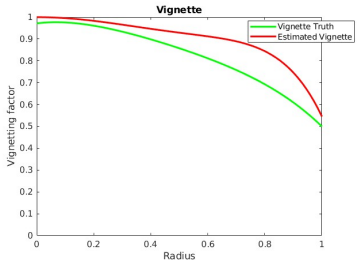
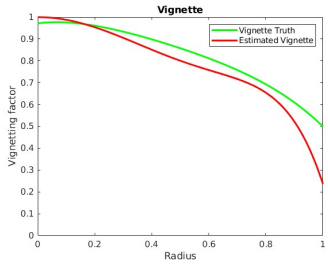
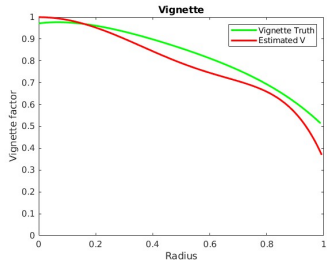
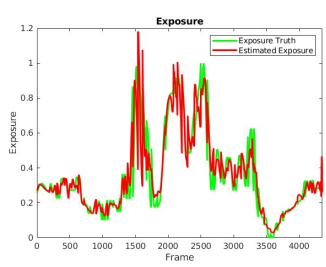
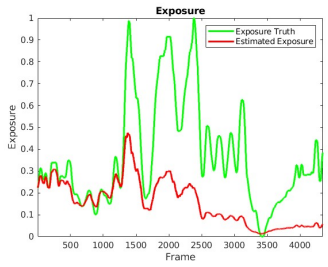
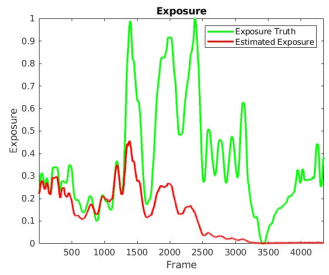
	OPC	OSPC	NPC
ICRF			
	(a) 4.7290	(b) 15.5420	(c) 5.7549
Vignette			
	(d) 0.0899	(e) 0.0671	(f) 0.0522
Exposure			
	(g) 0.0845	(h) 0.2874	(i) 0.3226

Figure A.17: Sequence 40.

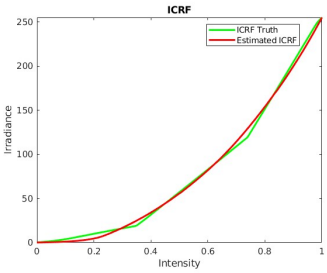
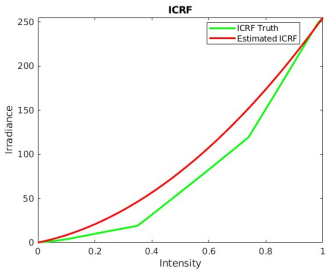
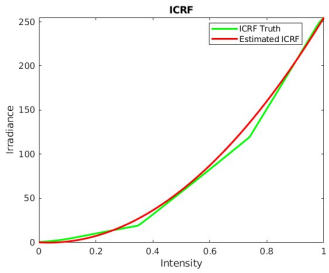
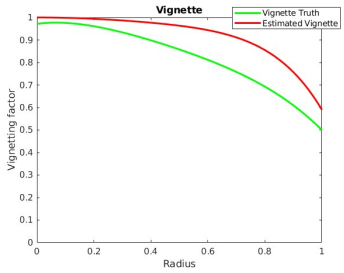
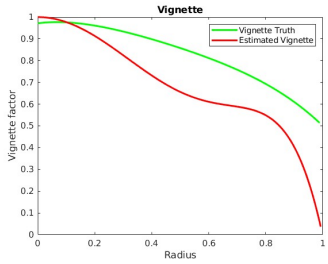
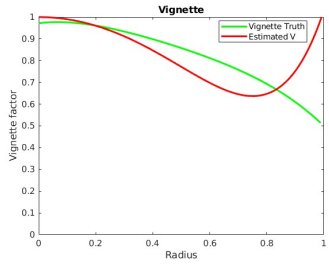
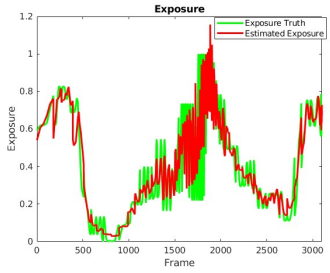
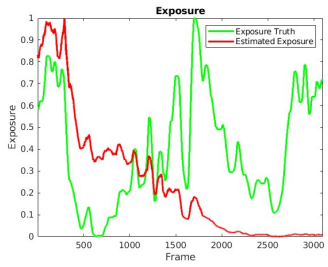
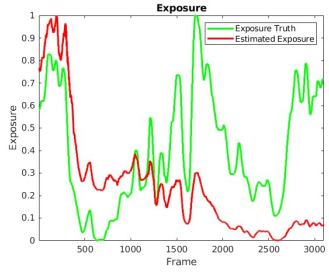
	OPC	OSPC	NPC
ICRF	 (a) 3.7096	 (b) 19.7618	 (c) 5.8026
Vignette	 (d) 0.1078	 (e) 0.1720	 (f) 0.1169
Exposure	 (g) 0.0921	 (h) 0.3928	 (i) 0.3373

Figure A.18: Sequence 41.

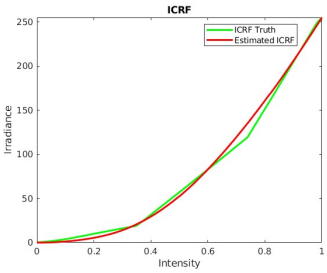
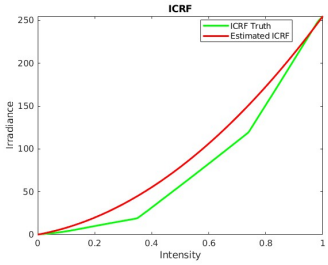
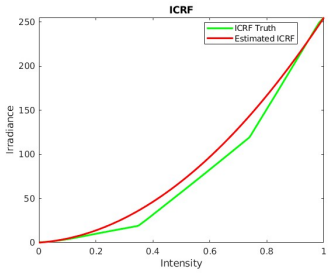
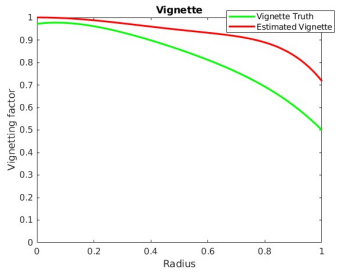
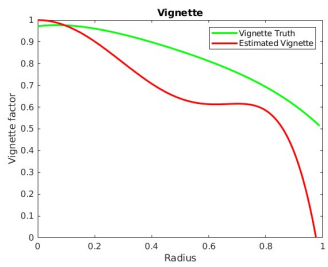
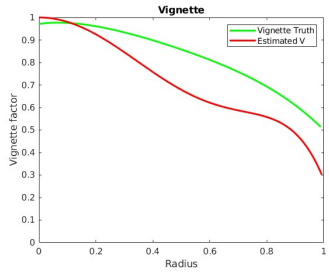
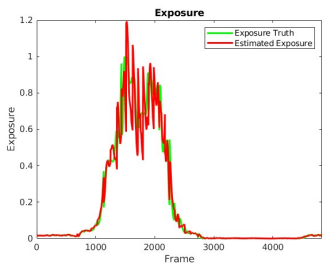
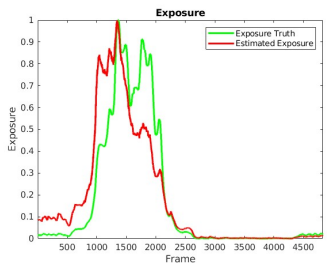
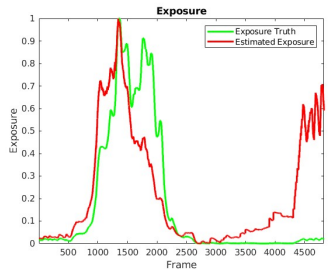
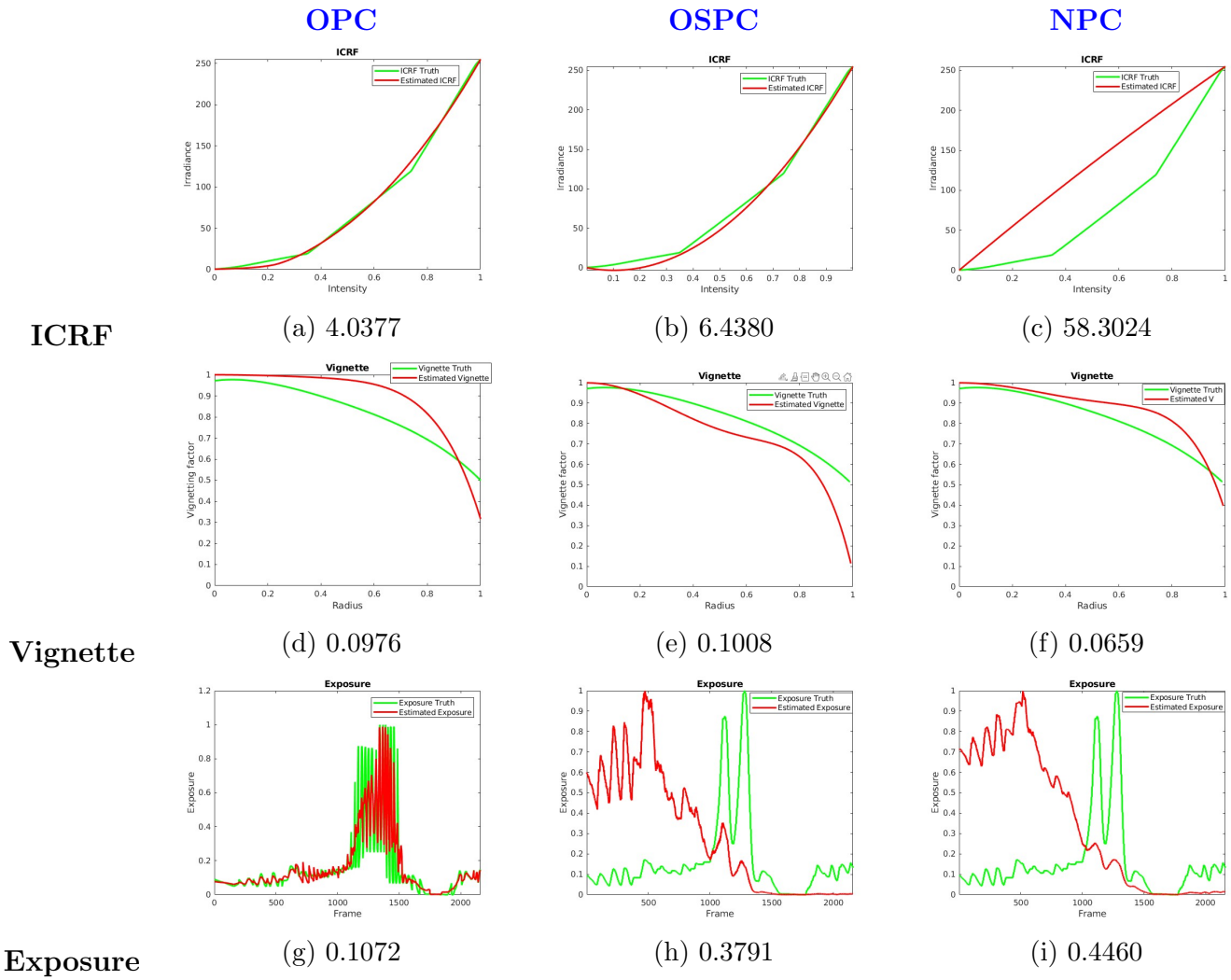
	OPC	OSPC	NPC
ICRF	 (a) 5.3675	 (b) 18.7268	 (c) 12.0863
Vignette	 (d) 0.1295	 (e) 0.1890	 (f) 0.1343
Exposure	 (g) 0.0669	 (h) 0.1368	 (i) 0.2191

Figure A.19: Sequence 42.



(a) 4.0377 (b) 6.4380 (c) 58.3024

(d) 0.0976 (e) 0.1008 (f) 0.0659

(g) 0.1072 (h) 0.3791 (i) 0.4460

Figure A.20: Sequence 43.

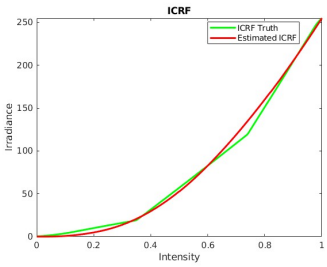
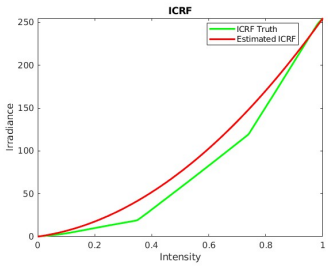
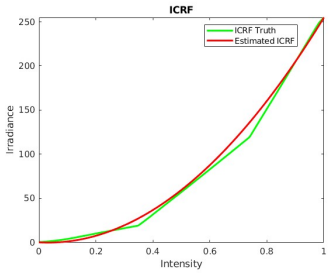
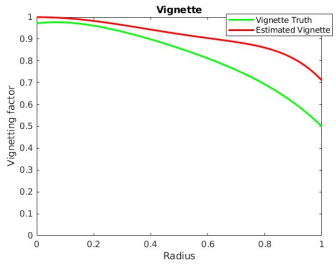
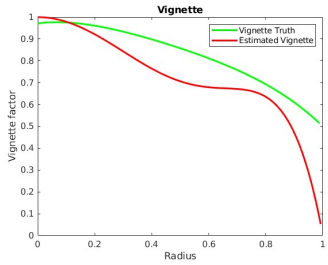
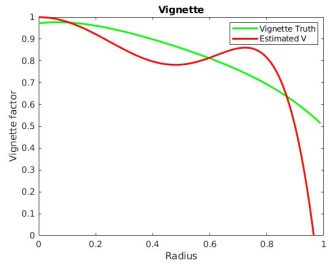
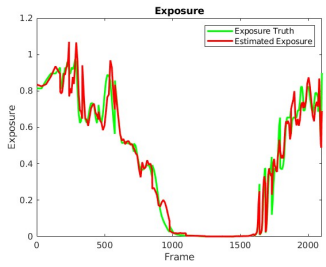
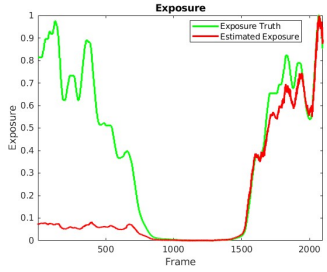
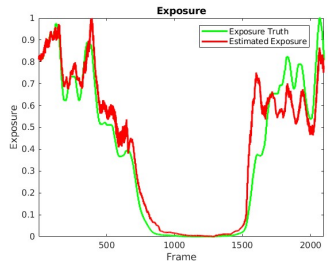
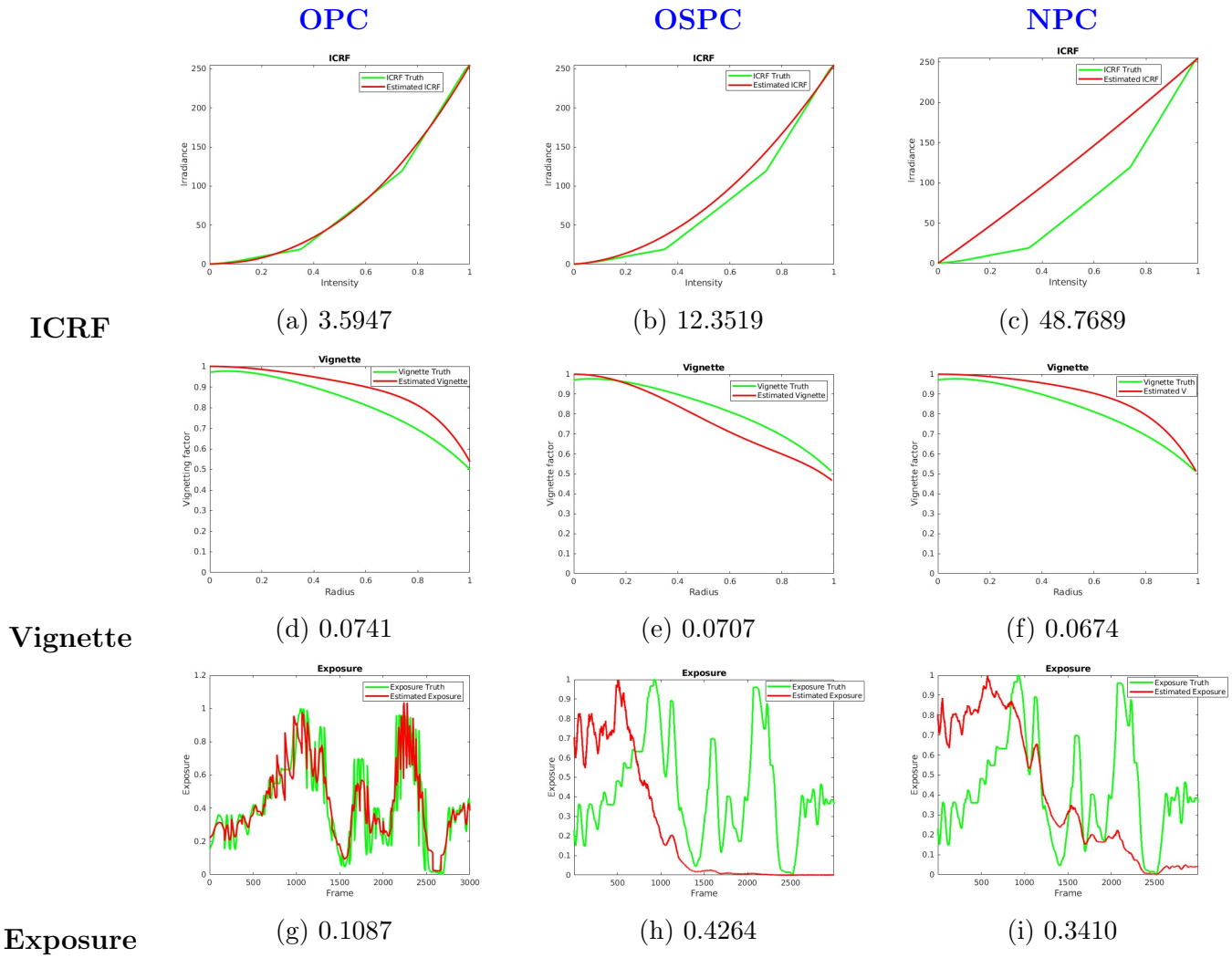
	OPC	OSPC	NPC
ICRF	 (a) 5.2442	 (b) 16.1951	 (c) 5.9678
Vignette	 (d) 0.1118	 (e) 0.1291	 (f) 0.1600
Exposure	 (g) 0.0602	 (h) 0.3631	 (i) 0.1019

Figure A.21: Sequence 44.

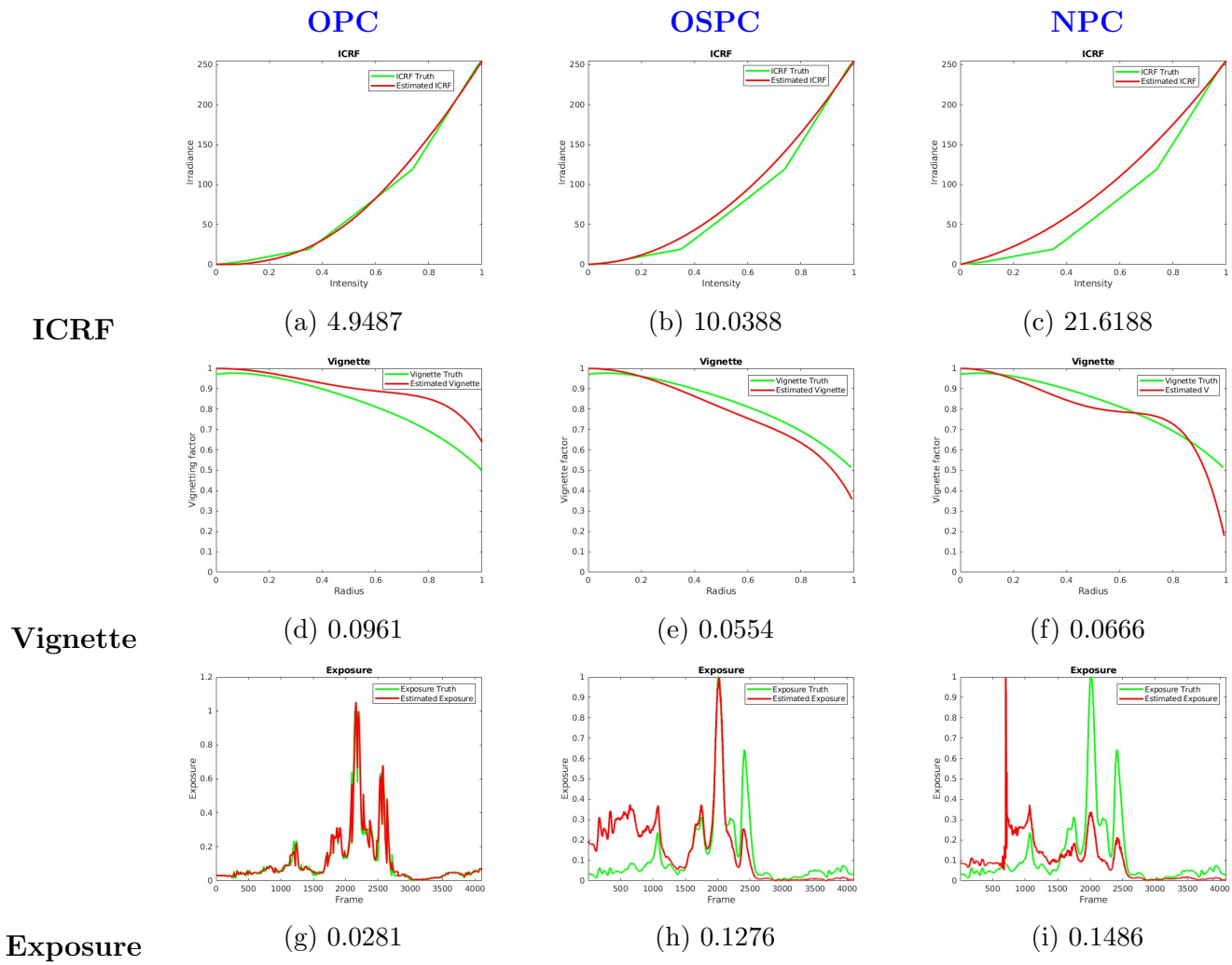


(a) 3.5947 (b) 12.3519 (c) 48.7689

(d) 0.0741 (e) 0.0707 (f) 0.0674

(g) 0.1087 (h) 0.4264 (i) 0.3410

Figure A.22: Sequence 45.



(a) 4.9487 (b) 10.0388 (c) 21.6188

(d) 0.0961 (e) 0.0554 (f) 0.0666

(g) 0.0281 (h) 0.1276 (i) 0.1486

Figure A.23: Sequence 46.

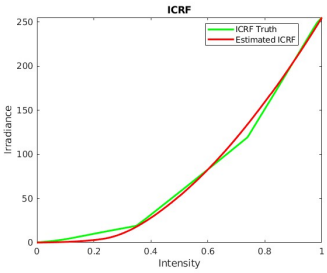
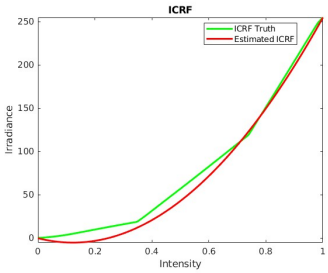
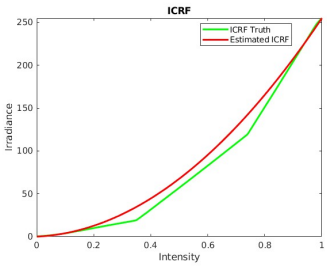
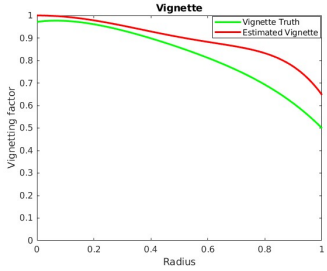
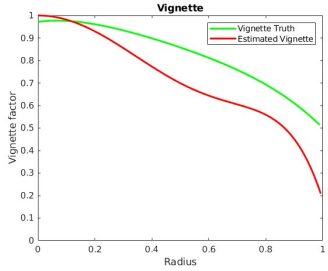
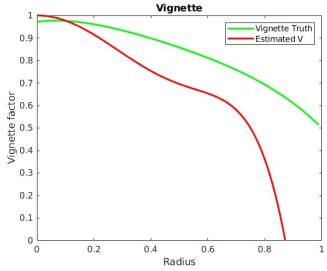
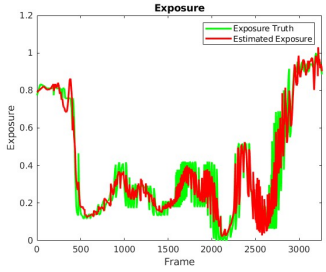
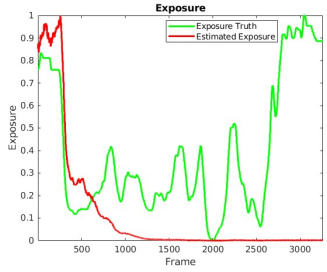
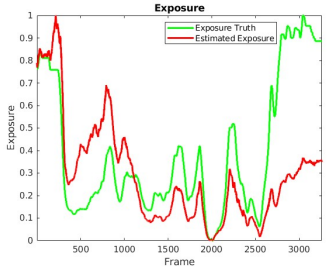
	OPC	OSPC	NPC
ICRF	 (a) 5.5471	 (b) 9.1259	 (c) 10.8857
Vignette	 (d) 0.0888	 (e) 0.1340	 (f) 0.4630
Exposure	 (g) 0.0583	 (h) 0.4285	 (i) 0.2878

Figure A.24: Sequence 47.

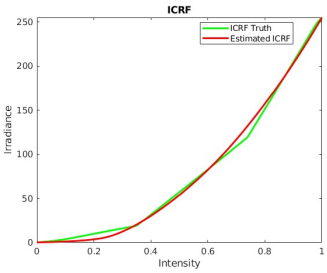
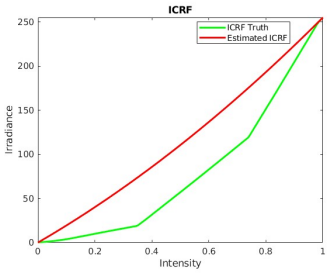
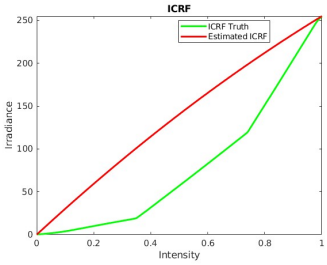
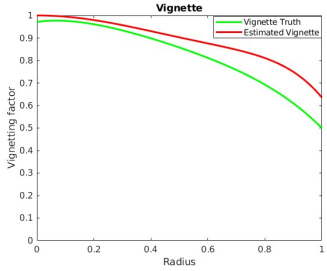
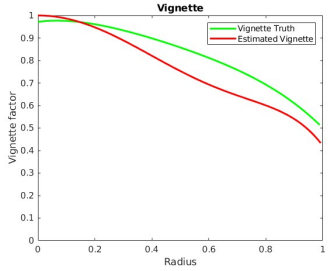
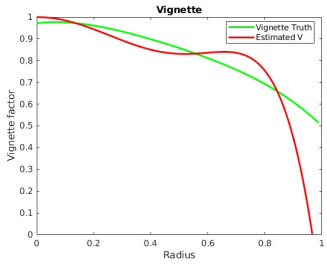
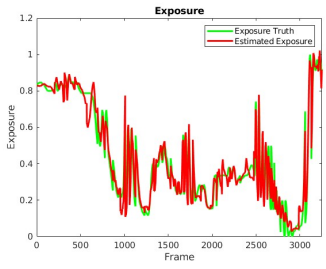
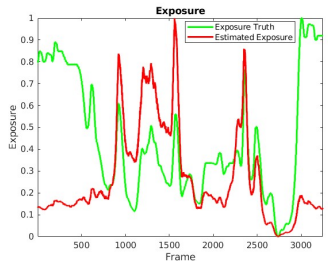
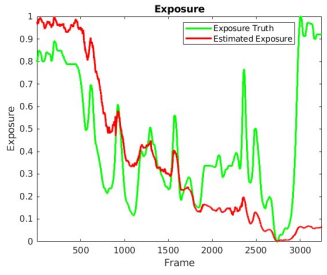
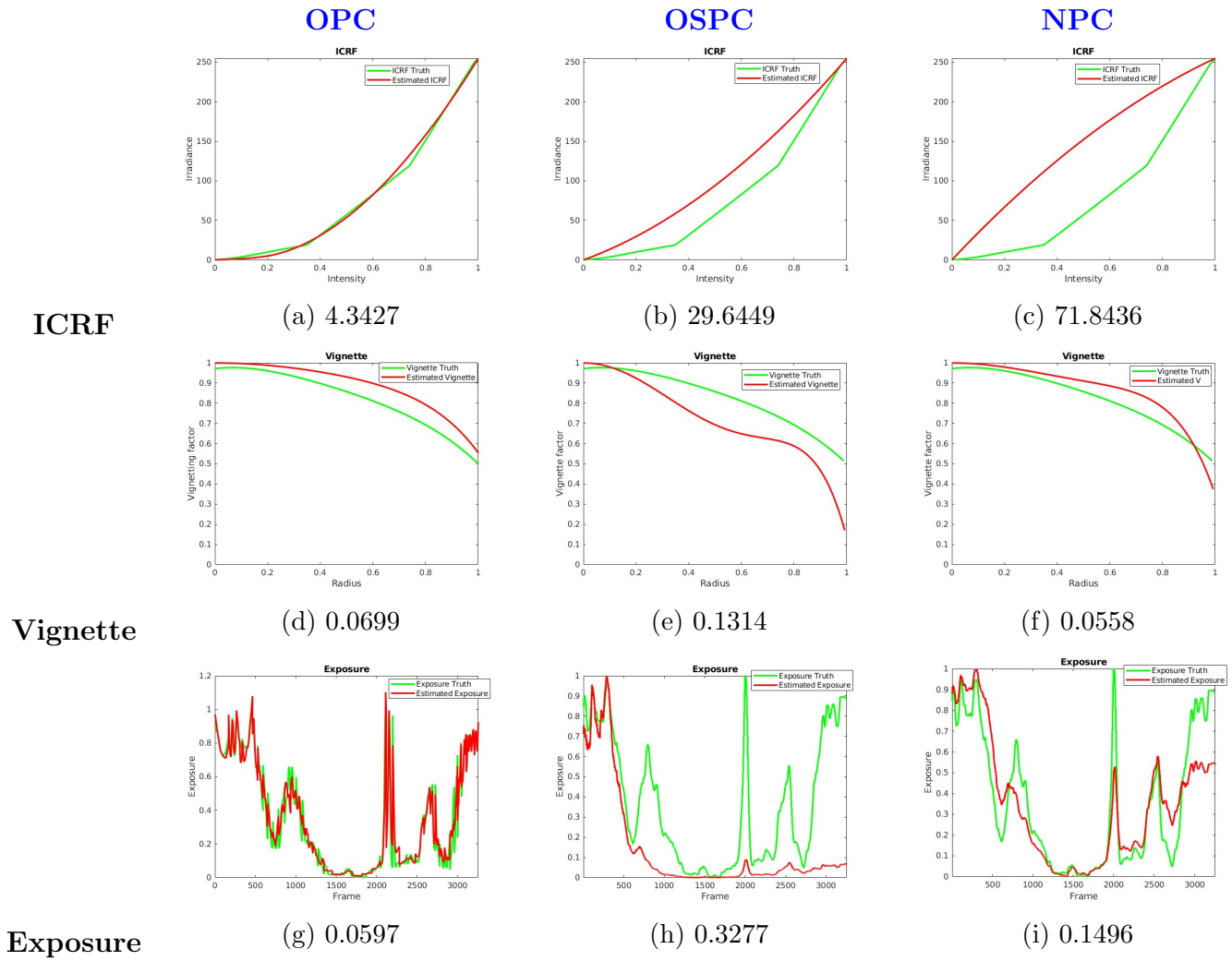
	OPC	OSPC	NPC
ICRF	 (a) 4.4952	 (b) 41.4901	 (c) 63.2512
Vignette	 (d) 0.0776	 (e) 0.0805	 (f) 0.1470
Exposure	 (g) 0.0566	 (h) 0.3910	 (i) 0.3136

Figure A.25: Sequence 48.



ICRF

(a) 4.3427 (b) 29.6449 (c) 71.8436

Vignette

(d) 0.0699 (e) 0.1314 (f) 0.0558

Exposure

(g) 0.0597 (h) 0.3277 (i) 0.1496

Figure A.26: Sequence 49.

A.2 Stats for Photometric Calibration Without Ground Truth

Table A.3: VQA of photometric calibrated video without ground truth.

VQA		OPC	NPC
V2-03-difficult (1)	VIF	0.7945	0.8241
	Motion2	9.7838	14.0445
	ADM2	0.9713	0.8373
	VMAF	**84.5634**	64.4400
V1-02-medium (2)	VIF	0.8143	0.7606
	Motion2	13.6464	13.6464
	ADM2	0.9039	0.8063
	VMAF	**75.2637**	53.3342
MH-04-difficult (3)	VIF	0.7989	0.7251
	Motion2	11.4122	11.4122
	ADM2	0.8419	0.6959
	VMAF	**63.4007**	29.3343
MH-01-easy (4)	VIF	0.8587	0.6363
	Motion2	12.9605	12.8605
	ADM2	0.9201	0.6268
	VMAF	**79.1161**	13.4839
ICL-NUIM-0 (5)	VIF	0.8402	0.9747
	Motion2	3.3232	3.3232
	ADM2	0.8393	1.0378
	VMAF	54.6609	**96.6945**
ICL-NUIM-1 (6)	VIF	0.9019	0.9359
	Motion2	3.4427	3.4427
	ADM2	0.9156	1.0597
	VMAF	74.6281	**92.9006**
KITTI 05 (7)	VIF	0.4493	0.5575
	Motion2	19.0970	19.0970
	ADM2	0.8304	0.9164
	VMAF	55.8037	**86.6157**
KITTI 12 (8)	VIF	0.5474	0.4707
	Motion2	9.9582	8.9093

Continued on next page

Table A.3 – continued from previous page

Sequence	VQA	OPC	NPC
	ADM2	0.8862	0.7266
	VMAF	**62.3080**	30.9490
KITTI 15 (9)	VIF	0.4795	0.5299
	Motion2	18.2838	18.2838
	ADM2	0.8145	0.7166
	VMAF	**53.4860**	35.5663
KITTI 17 (10)	VIF	0.4156	0.5456
	Motion2	14.2005	14.2005
	ADM2	0.6670	0.9978
	VMAF	41.6080	**99.2206**