Image-Based Localization in Dark and Turbid Underwater Environments

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Ning Hu

July 2025

ACKNOWLEDGEMENTS

This dissertation is dedicated to my beloved parents, Lei Hu and Jiayue Sun, whose unwavering support, silent sacrifices, and unconditional love have been my foundation through every step of this journey. You have taught me, for example, what it means to persevere, to stay kind, and to never give up.

The path to a Ph.D. is never easy. It is a long ascent—not only in academic rigor but also in personal growth. At one point, after a disheartening failure in an IRONMAN race, I found myself physically drained and mentally discouraged. The weight of performance pressure, physical exhaustion, and research setbacks collided, leaving me questioning whether I still had the strength to carry on.

But I was never alone. I am deeply grateful to my advisor, professors Jindong Tan and Shuai Li, whose guidance in research and belief in my potential helped me regain clarity and focus. Your trust, patience, and intellectual freedom have shaped not only this dissertation but also the way I approach problems and life itself.

To my lab mates and collaborators, especially Ning Li, Yan Li and Hui Liu, thank you for walking this road with me. The long days in the lab, the technical discussions, and those shared moments of frustration and success made this journey meaningful and fulfilling.

I would also like to sincerely thank the faculty and staff of Department of Mechanical, Aerospace, and Biomedical Engineering for fostering an environment that encouraged exploration and growth, and to all those who contributed—technically or otherwise—to the experiments and ideas that form the backbone of this work.

To my friends beyond the academic world, thank you for being there during the quiet moments, the long absences, and for offering support in ways both small and profound. In particular, I am grateful to Ahmed Abdelmeguid, Mohamed Shatarah, Yucen Li and Xinge Hu—thank you for accompanying me through countless hours of training, cheering me on through every milestone, and helping me stay grounded through both athletic and academic challenges. Your encouragement and presence made the journey not only bearable, but meaningful.

Lastly, I want to thank someone who entered my life quietly, yet brought with her a quiet warmth. Though we are still in the early stages of knowing each other, her sincerity, calm confidence, and gentle strength gave me peace during some of my most overwhelming moments. Without needing to promise anything, she reminded me that beauty exists beyond ambition—and that gentleness can be powerful.

Looking back now, through fatigue, doubt, and endurance, I see a path marked by trust, compassion, and resilience. This dissertation is not merely a product of intellectual effort, but a heartfelt tribute to all who stood by me—through storms, through silence, and through every finish line.

ABSTRACT

This dissertation explores robust localization methods for autonomous underwater vehicles in dark and turbid underwater environments. This dissertation addresses poor visibility, suspended particles, and limited features, which hinder real-time visual perception and localization. It aims to develop innovative techniques that can enhance image quality, estimate motion blur parameters accurately, and improve camera pose estimation, ensuring reliable underwater navigation.

The primary aspects this dissertation considers and the main contributions are as follows.

Underwater Image Restoration and Enhancement: A dual transmittance estimation-based method integrating boundary constraints and local contrast estimation with adaptive ambient light estimation and color correction, is proposed. In addition, a transformer-based feature fusion framework is designed by combining physical model-estimated transmission parameters, which can enhance the visual quality and robustness of underwater images.

Motion Blur Parameter Estimation: A cepstrum morphological feature-based method is developed for estimating motion blur parameters, combining frequency-domain feature extraction and least squares ellipse fitting. This approach can accurately estimate blur angle and scale parameters, improving the performance when handling small-scale blurs typical in underwater imaging.

Transformer-Based Image Enhancement Guided by Physical Blur Estimation: An improved transformer-based image enhancement framework that performs physical blur estimation via Wiener filtering is developed. The explicit incorporation of the point spread

function-related information enables precise restoration of image structures and details, outperforming conventional image enhancement techniques.

Monocular Camera Pose Estimation: A monocular pose estimation method based on a parallel perspective error model is developed. This model utilizes a line segment error propagation model to enhance the stability and accuracy of pose estimation, achieving high performance in both simulated and real-world underwater scenarios.

Experiments confirm that the proposed image restoration and enhancement methods outperform existing techniques in subjective visual effects and objective quality metrics. The motion blur estimation method demonstrates high effectiveness in small-scale blur scenarios, and the proposed pose estimation method can improve the localization accuracy of underwater robots. The research findings have significant application potential in underwater robot navigation, environmental monitoring, and underwater exploration. Future research could focus on optimizing computational efficiency, extending the application scenario types, and exploring more efficient sensor fusion strategies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Background

As science and technology have progressed, the depletion of terrestrial non-renewable resources has accelerated, prompting humans to turn their attention to the resource-rich underwater domain (Ballard, 2014). This trend has accelerated the development and utilization of ocean, lake, and river resources; construction and maintenance of water conservancy facilities; underwater resource surveys and exploitation; and research on the distribution of living organisms and fish populations. In the development of underwater fields, particularly in regions that are difficult for humans to reach **Figure1.1**, advanced technologies and equipment, such as autonomous underwater vehicles (AUVs), have been widely researched and adopted (Henderson, 2013; C.-M. Lee, 2005; Leonard & Durrant-Whyte, 2012). Among these, visual image-based technologies have exhibited irreplaceable advantages. In one sense, real-time video images can be utilized to identify, monitor, and operate underwater targets (Johannsson et al, 2010; Rigby et al., 2006; Snyder, 2010). In another sense, the relevant underwater image data obtained can be saved to local storage for further research and analysis. Therefore, the study of underwater image localization technology holds significant value in the implementation of human underwater missions.

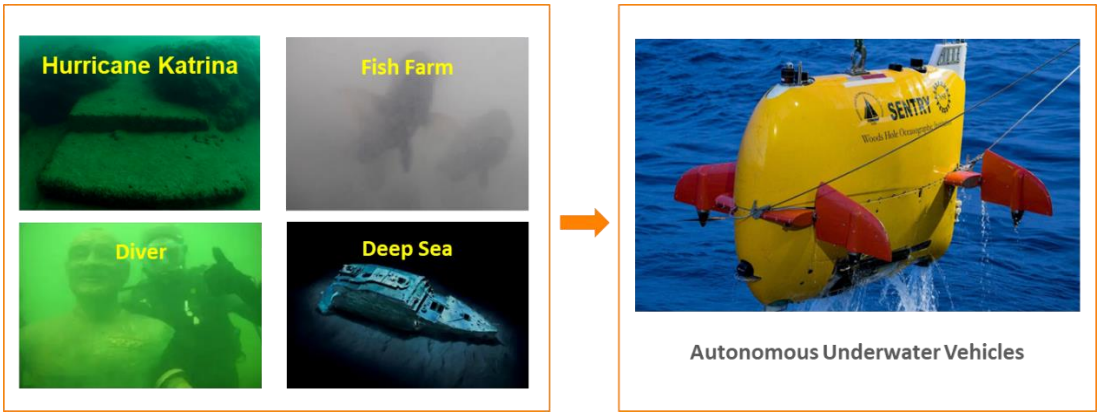


Figure1.1. Application scenarios for autonomous underwater robots (Ma, 2023).

Currently, the autonomous underwater robot operation process has four main steps: sinking, cruising, working, and returning to the dock **Figure1.2**. By analyzing the tasks of each step, one can find that the robot must sense its surroundings and its position to complete its tasks underwater. This process involves the simultaneous localization and mapping (SLAM) problem for machine vision. Therefore, clear visual information and accurate estimation of the target's position are critical for underwater operations.

In the atmospheric environment, sensor signals for image acquisition are abundant, and many vision-based algorithms have been proposed to achieve clear image acquisition and target attitude estimation. In the underwater field, however, the complex underwater environment presents numerous challenges in image acquisition, including absorption and scattering by the water medium and low contrast, color deviation, insufficient illumination, and blurred details in underwater images. In particular, in operational scenarios involving river culverts, dam gates, and similar scenarios, the water environment may have high turbidity or no natural light, thus requiring the utilization of artificial light sources to aid in illumination. While such sources may improve underwater visibility, because of the focusing characteristics of the light source, the image acquisition is usually presented in

the center of the overexposed part of the image, and the edge of the part may exhibit insufficient illumination. Simultaneously, the movement of underwater organisms and impurities, the relative position movement of the camera and target, and the slight camera shake can lead to motion blur, making it more difficult to capture clear underwater images. Moreover, during underwater operation, the distance between the camera and the target decreases as the camera approaches the target, and errors arise in the recognition of image features. These factors also reduce the accuracy of the target position estimation, which, in turn, has adverse impacts on the development and adoption of underwater resources and limits the progress of related fields.

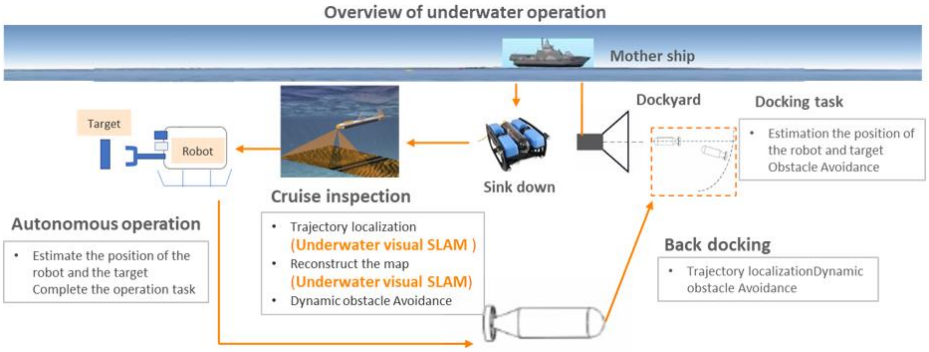


Figure1.2. Overview of underwater operation.

Problem Statement

To address the above problems, numerous researchers have developed various algorithms for enhancing underwater image clarity, removing motion blur, and estimating target position based on multi-point perspective imaging to enhance the quality of underwater images and target localization accuracy.

To facilitate the exploration of the research content in this paper, we provide a brief overview of dark turbid underwater image localization. After an in-depth study of various factors affecting the degradation model and solving the parameters required for the degradation model, we invert the physical model of image degradation to reduce the effects of absorption and scattering in an aqueous medium, thereby eliminating motion blur and obtaining an ideal, clear image. We also examine the relationship between the error of image feature points and target position estimation accuracy, propose a target position estimation method for underwater operation scenarios ranging from far to near, enhance the target position estimation accuracy, and provide theoretical support for autonomous underwater operation.

The research in this paper is summarized below.

(1) Image Restoration

Image restoration is defined as the application of a series of image processing techniques to enhance an image's visual clarity and quality. It involves two methods: image enhancement and image restoration. Image enhancement is a subjective processing method that highlights specific features in an image, such as edges, contours, and contrast, by adjusting the image pixel values to improve clarity and visual effect. However, overemphasizing certain features in an image can pull up and enlarge the pixels, thereby reducing the image's accuracy and clarity.

Image restoration, in contrast, is an objective processing method that restores the original appearance of a degraded image by analyzing the pattern of quality degradation

during image formation, dissemination, and preservation. This processing method aims to preserve as many details and features as possible in the original image while also enhancing its clarity and accuracy. As image restoration involves numerous parameters of the image degradation process, many of which are unknown and uncertain, we must often constrain the solution space of the pathological problem a priori, which leads to the overreliance of the image restoration algorithm on the degradation model of the image. Inaccurate modeling can lead to seriously degraded restoration results. Therefore, the generalization ability of image restoration algorithms is weak.

As this paper aims to study image clarity algorithms to provide accurate feature points and their parameters for target pose estimation, it focuses on image restoration algorithms while improving their generalizability via deep learning.

(2) Motion Deblur

Motion blur typically occurs when photographing moving objects, which continue to "leave" overlapping images in multiple locations on the sensor during exposure, thus creating a blurred, trailing effect that obscures an object's textures or key features. Motion blur removal is a technique that removes motion-induced blurring effects from an image, making it clearer. Motion blur removal methods model the blurring process as the convolution of a clear image with a blurring kernel, commonly referred to as the point spread function (PSF), which, in motion blur, specifically refers to the motion trajectory or blurring kernel. When the blurring kernel is known, the deblurring becomes a relatively "straightforward" deconvolution problem. Depending on whether the parameters of the

fuzzy kernel are known or can be obtained in advance, motion deblurring methods can be categorized as blind or non-blind deblurring.

Non-blind deblurring methods are highly dependent on the accuracy of the fuzzy kernel. Even minor errors in the fuzzy kernel may produce severe ringing effects, artifacts, or recovery failures. The method is generally applied when camera or system parameters are known to cause specific blurring, such as known lens out-of-focus parameters, known direction, and speed of uniform linear motion.

Blind deblurring methods with unknown fuzzy kernels and clear images are extremely pathological (ill-posed) problems with non-unique solutions that are extremely difficult to solve. A priori-driven conventional and deep learning methods are generally employed, which are extremely data-dependent. Therefore, these methods are suitable for most practical scenarios.

As the study scenario is a dark, turbid underwater environment, the motion blur kernel parameters are unknown. Therefore, this paper focuses on the blind motion blurring algorithm and combines it with deep learning to improve the motion blurring effect.

(3) Camera Pose Estimation

Perspective-n-point (PNP) refers to the estimation of the camera's position (rotation and translation) by learning the 3D coordinates of n points in space (under the world coordinate system) and their corresponding 2D projected point coordinates on the image plane, which is among the key aspects of SLAM. PnP serves as a bridge connecting the front-end tracking and back-end optimization of SLAM, and its algorithm performance directly

affects the performance of the SLAM system. Because the accuracy and efficiency of the PnP algorithm is closely related to the number of feature points, distribution, matching accuracy, target, and camera distance, this paper focuses on analyzing the error transfer model of these factors, studying the PnP algorithm compatible with different camera-target distances, and comprehensively improving the accuracy, efficiency, and robustness of the algorithm.

Knowledge Gaps

Despite significant progress in underwater positioning and navigation technology, several key problems remain unsolved in image clarification and attitude estimation techniques under dark light and turbid conditions. These problems are summarized in **Figure1.3** and include the following.

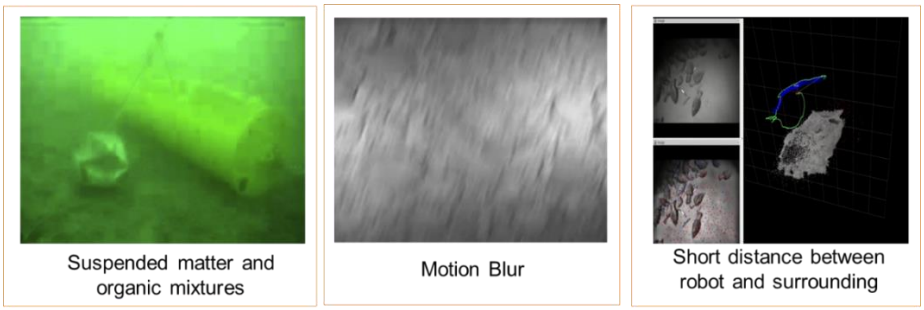


Figure1.3. Knowledge gaps.

First, existing underwater optical imaging models fail to effectively address the serious image degradation problems caused by absorption and scattering in the underwater environment. In particular, there has been a lack of robust imaging mechanism research under dark light and turbid conditions, resulting in low image clarity and serious color

distortion, which significantly limits the subsequent image processing and visual localization performance.

Second, conventional motion blur estimation methods perform poorly in underwater weak texture, low-contrast, and small-scale blurred scenes, making it difficult to accurately characterize and recover from blur degradation due to dynamic disturbances. This issue significantly impacts the structural clarity of underwater images and directly compromises the reliability of vision-based navigation and perception tasks.

Finally, existing monocular vision position estimation methods are sufficiently reliable in close-range and low-texture underwater environments, and the error propagation model and optimization method cannot effectively guarantee high accuracy and robustness under complex noise interference conditions. This issue leads to inaccurate positioning and a failure of attitude estimation.

The aforementioned knowledge gaps seriously limit the application and development of underwater robots in complex real-world environments. Therefore, there is an urgent need to establish a systematic theoretical model and technical method system to effectively break through the above technical bottlenecks and promote the development of more accurate and robust underwater visual localization.

Research Questions and Objectives

Based on an analysis of the various factors affecting the quality model of underwater image degradation, this paper explores three main research questions and defines specific objectives for each question. It also poses the overarching question of how we can

accurately characterize optical properties under turbid water and their relationship to image quality.

The light distribution curve of an artificial light source is introduced to establish a light field distribution model of an artificial light source in a turbid underwater environment, providing the basis for research on the intensity distribution of an underwater artificial light field. A multistage attenuation underwater imaging model of natural and complex artificial lights is established to provide theoretical support for the research on further improving imaging quality and calculating the position here.

1) How can we integrate physics-based dual-transmission estimation with transformer attention to correct wavelength-dependent attenuation in real time and nonuniform scattering in dark, turbid near-field LED scenes?

This part proposes a hybrid underwater image enhancement method that integrates dual transmission optimization with transformer-based feature fusion (Uformer), thus specifically addressing the optical degradation challenges in dark and turbid underwater environments. By combining boundary constraints and local contrast priors, the proposed method effectively compensates for wavelength-dependent attenuation. In addition, an adaptive ambient light estimation and color correction module is designed to enhance color fidelity under complex lighting conditions. Furthermore, physical model-based transmission maps are explicitly embedded into a transformer architecture with multi-scale feature fusion, significantly enhancing structural details and the overall visual quality in underwater imagery.

2) How can we derive reliable small-blur PSFs from cepstrum in turbid, low-texture underwater scenes and couple them with transformer attention for real-time deblurring?

This part develops an accurate underwater motion blur parameter estimation method based on cepstrum morphological analysis, thereby effectively addressing limitations of conventional approaches in weak-texture, low-contrast, and small-scale blur scenarios. In addition, a transformer-based image enhancement framework guided explicitly by physical blur estimation is proposed. Specifically, the blur kernel (PSF), estimated through Wiener filtering, is integrated directly into the transformer's attention mechanism, thus significantly improving the restoration of structural details and overall visual quality. A comprehensive validation through synthetic and real underwater datasets demonstrates the robustness, accuracy, and practicality of the proposed approach under various motion-blurred conditions.

3) How can we recover a metric-accurate, real-time six-degrees-of-freedom pose from a single underwater image under dark, turbid conditions while compensating for parallel-perspective degeneracy and texture scarcity?

We develop a monocular camera pose estimation method utilizing a parallel perspective error model enhanced by line segment error propagation, thus significantly improving stability and accuracy.

We validate the effectiveness of the proposed pose estimation approach via an extensive simulation and real-world underwater experiments, and the results confirm the enhanced localization precision.

Contributions

Underwater Optical Properties

By analyzing the attenuation characteristics of the water medium and its influence on the quality of underwater images, we introduce the light distribution curve of an artificial light source. We establish a light field distribution model of an artificial light source for a turbid underwater environment, which provides the basis for studying the intensity distribution of underwater artificial light field. To address the limitation of the Jaffe–McGlamery model, which does not fully consider the light propagation path, we establish a multistage attenuation underwater imaging model incorporating both natural and complex artificial light. We provide a simplified form of the model that integrates the turbid underwater environment. Finally, the effect of forward scattering on image resolution and structural details under various turbidity conditions as well as the dominant role of backward scattering in changing imaging brightness and blurring degree is verified by utilizing images with a weak texture in the underwater simulation environment. This approach provides a theoretical basis for this, which aims further improve the imaging quality and calculate the position.

Image Restoration

An underwater image enhancement method incorporating dual transmittance optimization and transformer feature fusion network (Uformer) is established. The method achieves robust double transmittance estimation through boundary constraints and local contrast prior, effectively handling the wavelength-dependent light attenuation problem in

underwater environments. Moreover, an adaptive ambient light estimation and color correction module is designed to significantly enhance the color reproduction ability under complex lighting conditions. In addition, the physical model-based transmittance information is explicitly embedded into the transformer global feature fusion network, which further improves the enhanced image's structural details and overall visual effect. Through comprehensive experimental validation on multiple benchmark datasets, the proposed method demonstrates significant advantages in objective evaluation metrics and subjective visual effects, effectively supporting subsequent underwater visual localization tasks.

Motion Deblur

An underwater motion blur parameter estimation method based on cepstrum morphological features is presented, which effectively improves the estimation accuracy of the blur angle and scale under small-scale motion blur conditions. This method significantly enhances the deficiencies of the conventional method in weak-texture and low-contrast scenes. In addition, a transformer image enhancement framework based on physical fuzzy kernel estimation guidance is designed. For the first time, the motion blur parameters estimated by cepstrum are explicitly embedded into the transformer's attention mechanism in the form of a PSF, which fully integrates the advantages of physical modeling and deep learning. Validated by both real and synthetic underwater datasets, the proposed method significantly outperforms existing mainstream methods in image deblurring performance

and detail restoration, effectively supporting the accuracy and robustness requirements of underwater image localization tasks.

Camera Pose Estimation

To overcome the instability and inaccuracy of monocular camera position estimation in underwater close-range and low-texture scenes, we propose a robust position estimation algorithm based on parallel perspective error modeling. First, a line-segment parallel perspective projection model appropriate for the underwater visual environment is presented to accurately obtain the initial solution of position estimation. Second, a line-segment error propagation model is established to enhance the algorithm's robustness under noise interference. Finally, the objective function is optimized by the Gaussian Newton method, which markedly improves the accuracy and computational efficiency of monocular visual position estimation. Then, a systematic experimental validation is conducted on simulated and actual underwater scenes. The results confirm that the proposed algorithm can still maintain stable and reliable performance under complex and weak texture conditions, which is significantly better than classical methods such as the conventional efficient PnP method, thus providing solid technical support for the visual navigation task of underwater robots.

As presented in **Figure1.4**, the logical structure and interconnections between Chapters 3, 4, and 5 reflect the sequential research flow established by the three main research objectives, addressing distinct but interrelated underwater perception challenges.

Collectively, these chapters sequentially construct a robust and integrated solution framework, ranging from fundamental static image restoration to dynamic blur mitigation, targeting precise and robust underwater localization.

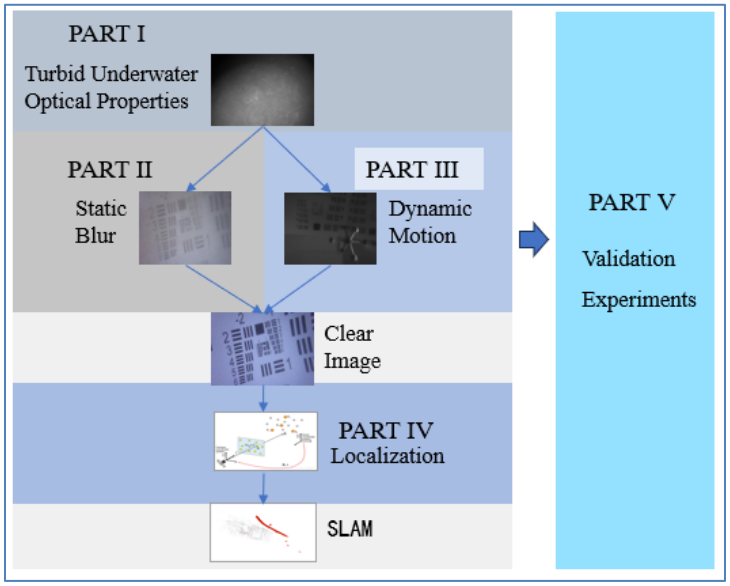


Figure1.4. Research overview. (PART I-Chapter 2; PART II-Chapter 3; PART III-Chapter 4; PART IV-Chapter 5; PART V-Chapter 6).

CHAPTER TWO

UNDERWATER OPTICAL IMAGING AND UNDERWATER IMAGE QUALITY ASSESSMENT METHODS

Underwater environment recognition strongly relies on the quality of underwater images. Therefore, it is highly significant for the research of underwater optical image enhancement methods to analyze multiple factors affecting imaging quality and systematically design a mathematical model of the underwater optical imaging process. Hence, this chapter studies the underwater illumination light field and underwater light propagation characteristics, examines their influence on underwater imaging quality, establishes an imaging model tailored to dark and turbid underwater environments, and develops an image quality evaluation method based on the underwater imaging whole process model. This provides a theoretical basis for research on further enhancement of the imaging quality and position calculation presented here.

Introduction

Humans often need to detect, maintain, and operate objects such as river culverts, dam gates, and underwater resources based on image information in highly turbid, unnatural light, and flowing water environments, utilizing external light sources. As the water body contains a large amount of dissolved and particulate matter, it can significantly reduce the imaging quality, causing various problems, such as reduced image contrast, blurred details, color distortion, and instability, which severely affect subsequent feature extraction and target recognition processes.

Therefore, how to effectively improve the quality of underwater optical images has become a hot research topic in recent years. However, this topic is inextricably linked to the research on the mechanisms and characteristics of underwater optical imaging systems. Accordingly, analyzing multiple factors that affect the imaging quality and systematically establishing a mathematical model of the underwater optical imaging process is highly significant to the research and development of efficient underwater optical image enhancement methods.

Most existing underwater imaging models are based on the Jaffe–McGlamery imaging model (Jaffe, 1990) or its simplified version, which can simulate the underwater image formation process under single-point artificial light source illumination. However, for turbid and non-uniform illumination light fields, a certain deviation between the underwater image formation process described by this model and real underwater application scenarios can occur, requiring systematic modeling of the imaging process under different lighting conditions.

This chapter first analyses the light field of an artificial illumination source and the attenuation and scattering characteristics of its radiated light under aqueous media. Then, the underwater imaging process is mathematically modeled under complex lighting conditions. Based on the established imaging model, underwater image simulation analysis is conducted, and the absorption and scattering characteristics of real water bodies are introduced into the underwater image simulation process to simulate the absorption and scattering effects of different turbid water bodies and examine their impact on imaging performance.

Underwater Optical Characteristics

The primary challenge in the visual perception of underwater environments is the significant attenuation of light as it travels through the water column. The absorption and scattering effects of light mainly cause this attenuation. Unlike airborne environments, the presence of large amounts of suspended particles, plankton, organic matter, and water molecules can significantly influence the light propagation process and distance traveled underwater, resulting in severe quality degradation of underwater images, including low contrast, blurring, color distortion, and a severe loss of detail. Therefore, an in-depth understanding and quantitative description of the optical properties of the underwater environment is crucial for effective underwater visual perception and image restoration.

Absorption Characteristics of Light in Water

As light propagates through water, it experiences absorption effects, resulting in substantial energy attenuation. This phenomenon can significantly restrict the effective transmission distance of underwater illumination, typically reducing it to merely tens of meters. Consequently, such a limited propagation distance imposes notable constraints on the practical applicability and performance of underwater imaging systems (X. U. Li et al., 2022).

As illustrated in **Figure2.1**, when an incident radiant flux Φ_{in} propagates through a water medium with a thickness of Δd, it experiences attenuation primarily due to absorption. $\Delta\Phi$ denotes the resulting decrease in radiant flux; then, the absorption coefficient a of the water medium can be mathematically expressed as follows (Z. Lee, 1994):

$$a = \frac{\Delta\Phi}{\Phi_{in}\Delta d} \tag{2.1}$$

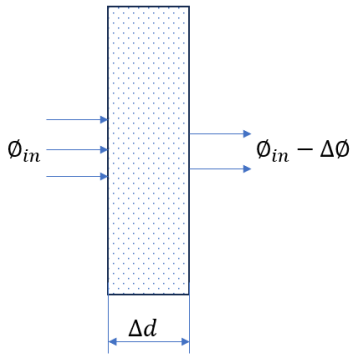


Figure2.1. Illustration of light absorption by water medium during propagation.

In aquatic environments, three primary categories of substances contribute significantly to the absorption of light: pure water, phytoplankton, and yellow substances, collectively referred to as colored dissolved organic matter. The total absorption coefficient of natural waters contains many components, including the inherent absorption properties of pure water, absorption arising from suspended particulate matter, and the absorption associated with dissolved organic substances, which can be mathematically expressed as follows:

$$a(\lambda) = A_{water}(\lambda) + A_{particles}(\lambda) + A_D OM(\lambda) \tag{2.2}$$

where:

$A_{water}(\lambda)$: an absorption coefficient of pure water;

$A_{particles}(\lambda)$: an absorption coefficient of suspended particles (e.g., sediments, plankton);

$A\_DOM(\lambda)$: an absorption coefficient of dissolved organic matter (e.g., humic substances).

The above formula reflects the spectral dependence of light absorption in natural waters and has been widely adopted in marine and inland optical models (Z. X. Zhang, 2012).

The absorption characteristics of pure water exhibit relatively stable spectral behavior across a wide range of wavelengths. Particularly, water molecules strongly absorb radiation

in both the ultraviolet (UV) and infrared (IR) spectral bands, which is driven by distinct resonance mechanisms. For instance, electronic excitation causes pronounced absorption in the UV band, whereas molecular vibrational excitation predominantly results in strong absorption in the IR band. Conversely, absorption in the visible spectrum is considerably weaker and varies significantly across wavelengths, thereby rendering visible light most suitable for underwater optical imaging.

Moreover, the rate of light attenuation in water changes noticeably with the wavelength. Generally, longer wavelengths experience more rapid absorption, with the red-light region experiencing the highest attenuation. The absorption degree decreases sequentially from red through orange, yellow, and green, with blue light exhibiting the least absorption, enabling deeper penetration in aquatic environments. Particularly in the blue–green spectral band (approximately 462–475 nm), the absorption coefficient is notably low, approximately 0.04 m^{-1}. This wavelength-dependent absorption is a key factor contributing to the pronounced color distortion usually observed in underwater imagery (Fu, 2023;Nie, 2014; X. Zhao, 2015).

Light Scattering Characteristics in Water

In addition to absorption, light propagating through the water is also subject to scattering. As illustrated in **Figure2.2**, $\emptyset_{in}$ denotes the incident radiant flux. When this light passes through a water layer with a thickness Δd, a flux loss caused by scattering (assuming no absorption) is denoted by $\Delta\emptyset$. The scattering coefficient *b*, which quantifies the extent of scattering in the medium, is defined as follows:

$$b = \frac{\Delta\Phi}{\Phi_{in}\Delta d} \tag{2.3}$$

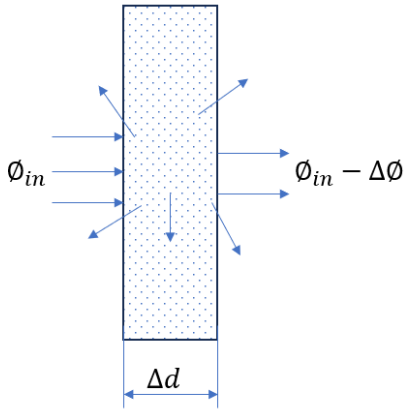


Figure2.2. Schematic of light scattering during light propagation in water.

Based on the relative size of particles suspended in water, light scattering in underwater environments can be roughly categorized into Rayleigh scattering and Mie scattering, as illustrated in **Figure2.3**. Rayleigh scattering occurs when the diameter of suspended particles, including water molecules, is much smaller than the wavelength of the incident light; in contrast, Mie scattering appears when the particle size is comparable to the wavelength of the incident light. Hence, the total spectral scattering coefficient $b(\lambda)$ of the water can thus be expressed as:

$$b(\lambda) = B_{Rayleigh}(\lambda) + B_{Mie}(\lambda) \tag{2.4}$$

Where:

$b(\lambda)$ is the total scattering coefficient at a wavelength λ;

$B_{Rayleigh}$ is the Rayleigh scattering coefficient, which is dominant when water particles' diameter is much smaller than λ;

$B_{Mie}(\lambda)$ is the Mie scattering coefficient, which becomes dominant when the water particles' diameter approaches λ^{-m}. Typically, it is modeled with a power-law form where the exponent *m* ranges from 0.5 to 2.

This decomposition enables accurate modeling of scattering behavior in turbid water bodies with various particle size distributions.

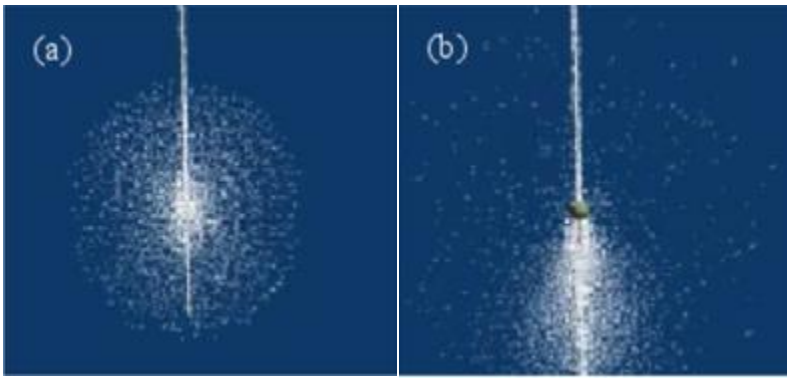


Figure2.3. Illustrative comparison of two scattering types: (a) Rayleigh scattering; (b) Mie scattering (Ding, 2023).

Owing to the complexity of suspended particle distributions in water, particularly in terms of their density, size, and composition, it is challenging to construct an accurate physical model of underwater scattering. Nevertheless, because the relationship between the particle size and incident light wavelength plays a critical role in underwater radiative transfer, the Rayleigh and Mie scattering models have been commonly adopted to characterize this behavior. These models serve as a basis for analyzing the volumetric distribution of scattered light.

The volume scattering function describes the angular distribution of scattered light per unit volume, and it is defined by:

$$\beta(\theta) = \frac{\partial^2 \phi(\theta)}{(E \partial \Omega \partial V)} \tag{2.5}$$

where E, Φ, θ, $\Delta\Omega$, Δv, and $\partial^2\phi(\theta)/(E\partial\Omega\partial V)$ represent light intensity, radiant flux, scattering angle, differential solid angle, differential scattering volume, and second inverse of the scattering flux of a small volume element corresponding to the given solid angle and scattering volume, respectively.

The scattering coefficient is defined by integrating the body scattering function as follows:

$$b(\lambda) = \iint_{4\pi} \beta(\theta) d\Omega = 2\pi \int_0^{\pi} \beta(\theta) \sin(\theta) d\theta \tag{2.6}$$

The above integral represents the total scattering per unit volume across all directions. Particularly, the integral of $\beta(\theta)$ over the range $0 \leq \theta \leq \pi/2$ defines the forward scattering coefficient, and the integral over $\pi/2 < \theta \leq \pi$ refers to the backward scattering coefficient.

The normalized volume scattering function, which is also known as a scattering phase function, is obtained by dividing the spectral volume scattering function $\beta(\lambda, \theta)$ by the total spectral scattering coefficient $b(\lambda)$, which can be expressed by:

$$\bar{\beta}(\lambda, \theta) = \frac{\beta(\lambda, \theta)}{b(\lambda)} \tag{2.7}$$

This normalized function $\bar{\beta}(\lambda, \theta)$ describes the probability density of scattering at an angle θ for a given wavelength λ and has been widely utilized in radiative transfer equations and underwater light propagation modeling.

The scattering behavior of suspended particles in water environments demonstrates significant variations with the water quality, which is mainly affected by the incident light's wavelength and the water particles' size, spatial distribution, and complex refractive

indices. Collectively, these parameters define the redistribution patterns of light in the water medium, and they can be quantitatively characterized by the Mie scattering theory (Qing, 2012).

Light Attenuation in Underwater Environments

Natural water bodies typically have complex compositions that extend far beyond pure water, and frequently contain high concentrations of suspended particulate matter and dissolved substances. Therefore, when light propagates through such a medium, it experiences substantial attenuation, and this effect is particularly pronounced in highly turbid environments, such as rivers, lakes, and reservoirs, where the effective visual range is often limited to approximately 5 m or less. Although artificial illumination can partially extend visibility in underwater imaging applications, the observable range usually remains constrained to a few meters under typical operational conditions (Y. L. Zhang et al., 2018).

The light attenuation process in water is mainly directed by two fundamental mechanisms: absorption and scattering. Absorption refers to the reduction in light energy as light propagates through the water medium, and this process is affected by various factors, such as the refractive index and molecular composition of the water. In contrast, scattering describes the deviation of light propagation from its initial path, primarily arising from interactions with suspended particles in the water. However, both absorption and scattering exhibit strong wavelength dependency, yielding a complex, wavelength-specific attenuation behavior. Consequently, the intensity of underwater light decreases exponentially with increasing distance of light propagation.

According to the Beer–Lambert law, the attenuation of light intensity, or the loss of optical energy, caused by absorption and scattering, is proportional to the propagation medium's thickness. The inherent optical properties of the propagation medium determine the attenuation level. Let $I(0)$ denote the initial light intensity and $I(d)$ the intensity after traveling a distance *d*; then, the attenuation model can be expressed as follows:

$$I(d) = I(0)e^{-cd} \tag{2.8}$$

where *c* is the total attenuation coefficient, expressed as m^{-1}, which indicates the combined effect of absorption and scattering per unit propagation distance.

In homogeneous media, the total attenuation coefficient represents the sum of the absorption coefficient *a* and scattering coefficient b: c = a + b (Jin & Yang, 2010).

Active Light Field Modeling for Underwater Illumination

Owing to the significant energy loss during light propagation through water, natural illumination in underwater environments is typically inadequate for effective visual imaging. To overcome this limitation, current research often employs artificial illumination sources to enhance underwater visibility. However, utilizing artificial lighting can often introduce undesirable reflective noise, which impacts image quality. Consequently, a thorough analysis of artificial illumination field characteristics, along with the development of accurate illumination models specifically tailored for turbid underwater environments, is crucial for achieving optimal underwater imaging performance.

Radiative Distribution of Active Light Sources in Air

Underwater illumination systems typically utilize LED elements as emission sources. The emitted light is shaped into a directional beam with a divergence angle θ utilizing different optical components, such as mounting structures, encapsulating gels, or integrated lenses, as illustrated in **Figure2.4**. Assuming the spectral radiant intensity of a point light source at the optical axis direction $O(0,0)$and a wavelength λ is denoted by $I(0,\lambda)$. Then, the spectral radiant intensity at an off-axis angle θ is defined by:

$$I(\theta,\lambda) = I(0,\lambda)\, f(\theta) \tag{2.9}$$

where $f(\theta)$ represents the angular distribution function of the light source, which is also referred to as the luminous intensity distribution profile.

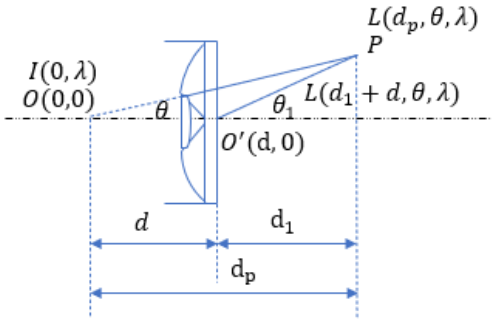


Figure2.4. Schematic of light field in the air with an LED light source.

Next, let Point *P* lie in the direction defined by an angle θ, located at an axial distance d_p from the light source. The spectral irradiance at Point *P* decreases exponentially with the propagation distance due to the light absorption by the propagation medium. The spectral irradiance at a wavelength λ at Point *P* can be expressed by

$$L\left(d_p,\theta,\lambda\right)=I(\theta,\lambda)cos^3\theta/d_p^2 \tag{2.10}$$

For a point located directly along the optical axis at a distance d_p from the light source, the spectral irradiance at a wavelength λ can be calculated by:

$$L\left(d_p,0,\lambda\right)=\frac{I(0,\lambda)}{d_p^2}=L(0,0,\lambda)/d_p^2 \tag{2.11}$$

Furthermore, the angular distribution of the off-axis radiance from a point light source can be defined by

$$L(d_a,\theta,\lambda)=L(d_a,0,\lambda)\,f(\theta)cos^3\theta \tag{2.12}$$

Given the source location at (0,0), the axial spectral radiance $L(0,0,\lambda)$, and spatial coordinates of Point $P(d_a,\theta)$, the spectral radiance at Point P for a wavelength λ can be accurately determined.

Light Field Distribution Based on Surface Point Emission in Air

In most cases, the precise location of a light emission point $O(0,0)$ is difficult to determine. Here, the active light field distribution is modeled utilizing a reference point $O'(d,0)$ located on the axial surface of the light source, along with its known spectral radiance. When an observation Point P is located far from the source, angles θ and θ_1, which are defined respectively with respect to the actual emission point and the surface reference point, become approximately equal. Therefore, here, θ_1 is adopted as a substitute for θ. Under this approximation, the spatial distribution of the active light field in the air, based on the surface reference point, can be expressed as follows:

$$L(d_a,\theta_1,\lambda)=L(O',\lambda)\,f(\theta_1)cos^3\theta_1 \tag{2.13}$$

Then, the axial light field distribution can be defined by:

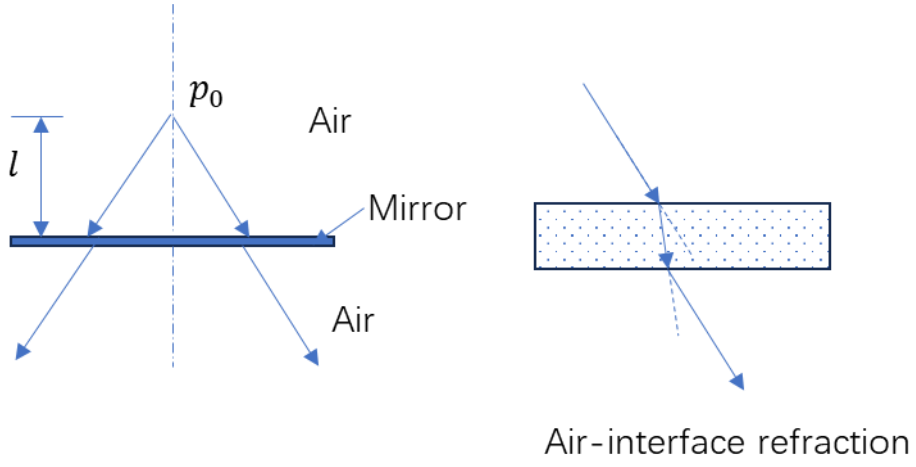
$$L(d_1, 0, \lambda) = L(d, 0, \lambda) d^2 / (d + d_1)^2 \tag{2.14}$$

Similarly, the off-axis light field distribution can be determined by:

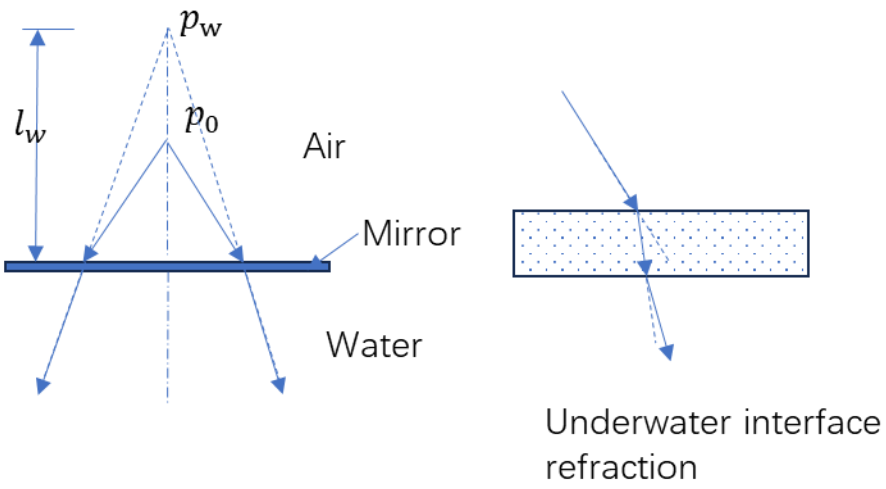
$$L(d_1, 0, \lambda) = L(d + d_1, \theta, \lambda) f(\theta) cos^3 \theta \tag{2.15}$$

Underwater Active Light Field Model

In underwater illumination systems, an active light source is typically enclosed in an air-filled compartment. The emitted light propagates outward through transparent materials, such as glass, before entering the surrounding water medium, as illustrated in **Figure2.5**. The full optical path includes the air-glass interface followed by the glass-water interface. Owing to the water's refractive properties, the transmittance at the optical window surface changes, and the effective emission point of the light is shifted further outward into the water. In addition, refraction causes the light rays to converge more tightly toward the optical axis, thereby increasing the energy density per unit area.



(a)



(b)

Figure2.5. Schematic of light refraction: (a) schematic of light path in the air; (b) schematic of light path in underwater environment**.**

According to Fresnel's equation, the transmittance of the air–glass and water–glass interfaces can be expressed by

$$t_a = 1 - \left(\frac{r_a - r_g}{r_a + r_g}\right)^2 \tag{2.16}$$

Moreover, based on the Fresnel equations, the transmittance values at the air–glass and water–glass interfaces, denoted by t_a and t_w, respectively can be expressed as follows:

$$t_w = 1 - \left(\frac{r_w - r_g}{r_w + r_g}\right)^2 \tag{2.17}$$

Where:

t_a is the transmittance at the air–glass interface;

t_w is the transmittance at the water–glass interface;

$r_a, r_w,$ and r_g are the refractive indices of air, glass, and water, respectively.

Hence, the effective radiant intensity of the underwater light source is defined by

$$t = \frac{t_w \times t_a}{t_a \times t_a} = \frac{t_w}{t_a} \tag{2.18}$$

Furthermore, because of refraction, the effective emission point of the light source is shifted outward. According to geometric similarity, the relative positions of the point source in the air and water are correlated as follows:

$$\frac{d_w}{t_w} = \frac{d_a}{t_a} \tag{2.19}$$

The relationship between the underwater and in-air spectral radiance values is defined by

$$\frac{L_w}{L_a} = (\frac{t_w}{t_a})^2 \tag{2.20}$$

Hence, the spectral radiance on the surface of the underwater light source can be expressed as follows:

$$L_w(d, 0, \lambda) = t(\frac{r_w}{r_a})^2 \times L_a(d, 0, \lambda) \tag{2.21}$$

Then, the axial spectral radiance of the underwater light field becomes:

$$L_0(d_1, 0, \lambda) = L_w(d, 0, \lambda)(\frac{r_w}{r_a} \cdot d)^2 / (d_1 + \frac{r_w}{r_a})^2 \tag{2.22}$$

Considering exponential attenuation during light propagation in the water, the attenuated axial radiance can be expressed as

$$L_{w0}(d_1, 0, \lambda) = L_0(d_1, 0, \lambda) \cdot e^{-c(\lambda)d_1} \tag{2.23}$$

For the off-axis rays, refraction alters the beam angle as the light passes through the air–water interface. Then, according to Snell's law, it holds that:

$$\theta_w = \arcsin(\frac{r_a}{r_w} sin\theta_a) \tag{2.24}$$

Further, the angular light distribution function in the underwater environment can be expressed by

$$f_w(\theta_w) = f(\arcsin(\frac{r_a}{r_w} sin\theta_a)) \tag{2.25}$$

The underwater propagation distance at an angle θ_w is calculated by

$$d_a(\theta_w) = \frac{d_1}{\cos(\theta_w)} \tag{2.26}$$

Accordingly, the off-axis spectral radiance of the underwater light field is defined by

$$L_{w\theta}(d_1, \theta, \lambda) = L_w(d_1, 0, \lambda) f_w(\theta_w) cos^2\theta_w e^{-c(\lambda)d_1/(cos\theta_w - d_1)} \tag{2.27}$$

Finally, the light field of the light source in the axial direction is defined by

$$L_{w0}(d_1, 0, \lambda) = e^{-c(\lambda)d_1} L_w(d, 0, \lambda)(\frac{r_w}{r_a} \cdot d)^2 / (d_1 + \frac{r_w}{r_a} d)^2 \tag{2.28}$$

Underwater Image Imaging Model

Unlike images captured in atmospheric environments, underwater images are subject to severe degradation because of complex optical conditions and pronounced attenuation effects in aquatic media. The adverse effects of these imaging factors commonly manifest as decreased contrast, color distortion, pervasive blurring, and substantial loss of structural information. Consequently, an in-depth understanding of underwater imaging principles is vital for the development of specialized image restoration and enhancement methods.

Therefore, this section introduces a classical underwater imaging model and a simplified imaging model designed explicitly to accommodate the unique characteristics of turbid water environments.

Classic Underwater Imaging Model

The Jaffe–McGlamery underwater imaging model posits that the light received by an underwater imaging sensor contains a direct, forward scattering, and backscattering component (Jaffe, 1990). The Jaffe–McGlamery model is illustrated in **Figure2.6**, and its operational principle is illustrated in **Figure2.7**.

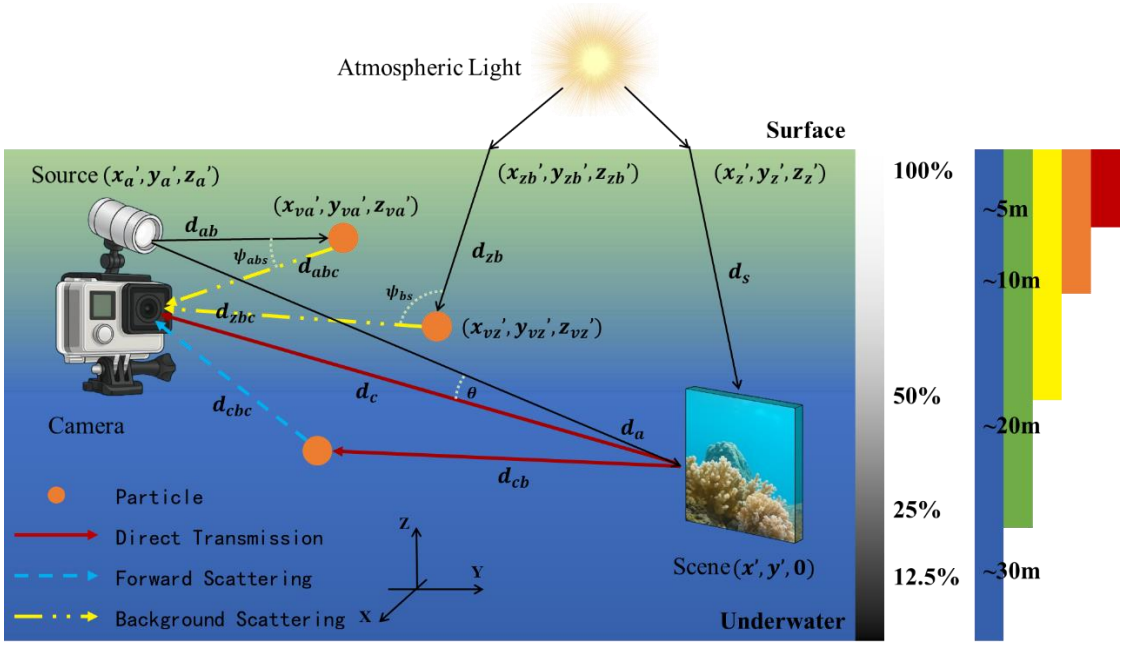


Figure2.6. Jeff–McGlamery underwater imaging model (Y. Wang, 2019).

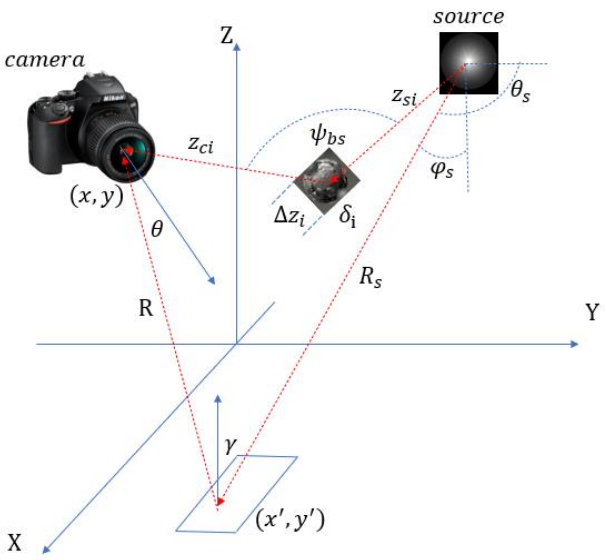


Figure2.7. Illustration of Jeff–McGlamery model's operational principle.

The three components are defined as follows:

- The direct component refers to the portion of light that reaches the target object, reflects off its surface, and travels directly back to the camera.
- The forward scattering component arises from the light reflected by the target object, which undergoes small-angle scattering due to suspended particles as it propagates toward the camera.
- The backscattering component comprises the light that does not reach the target object but is instead scattered at large angles by the medium's particles and enters the camera directly.

Accordingly, the total irradiance, E_T, received at the imaging sensor can be modeled as a linear sum of the three components as follows:

$$E_T = E_d + E_f + E_b \tag{2.29}$$

where E_T, E_d, E_f, and E_b denote the total irradiance, direct irradiance, forward scattering irradiance, and backward scattering irradiance, respectively.

Assume that a point (x, y) lies on the camera's image plane and a point (x', y') is its corresponding point on the target surface; R_s denotes the distance between the light source and the target point, and R denotes the distance between the point (x', y') and the camera imaging plane; δ_i represents a differential backscattering volume element; z_{ci} and z_{si} represent the distances from the scattering element to the camera and the light source, respectively, and ψ_{bs} represents the angle between the vector connecting the camera to the scattering element and the vector connecting the light source to the same element. γ denotes the angle between the light-source-to-point vector and the normal vector of the target surface; $E_{\delta i}$ represents the irradiance at a scattering element δ_i; Δz_i denotes the thickness of the backscattering volume element; φ_s represents the angle between the light-source-to-target-point vector and the direction perpendicular to the light source surface, and θ_s denotes the angle between the same vector and the horizontal axis of the light source.

The direct irradiance received at the camera sensor can be modeled as a product of the incident irradiance and the target's reflectance as follows:

$$E_d(x, y) = E_l(x', y')e^{(-cR)} \frac{M(x', y')}{4F^2} T_l cos^4\theta \left[\frac{R - F_l}{R} \right]^2 \tag{2.30}$$

where (x, y) and (x', y') are the coordinates on the image plane and target surface, respectively; E_l is the irradiance at the scene point (x', y'); R is the distance from the scene point to the camera; $M(x', y')$ indicates the reflectance of the target point, and its value is

typically between 0.02 and 0.1; F is the camera aperture, T_l is the lens transmittance; c is the propagation medium's attenuation coefficient; F_l is the focal length of the camera lens; θ is the angle between the vector connecting (x', y') and (x, y), and the optical axis of the camera.

Assume that the incident illumination from the light source entering the water is denoted by $B(\theta_s, \varphi_s)$; then, the reflected intensity at the scene surface can be calculated by

$$E_l(x', y') = B(\theta_s, \varphi_s) \cos\gamma \frac{e^{-cR_s}}{R_s^2} \tag{2.31}$$

where $B(\theta_s, \varphi_s)$ is the spectral irradiance emitted by the source in the direction defined by angles, and γ is the incidence angle between the vector from the source to the point (x', y') and the normal vector of the target plane.

The forward scattering component can be modeled as a convolution of the direct component and PSF g, which can be expressed as

$$\mathrm{E}_f(\mathrm{x,y}) = \mathrm{E}_d(\mathrm{x,y}) * \mathrm{g(x,y,R,c,B)} \tag{2.32}$$

Then, the PSF can be defined as follows:

$$g(x, y, R, G, c, B) = [\exp(-GR) - \exp(-cR)]F^{-1}[\exp(-BR\omega)] \tag{2.33}$$

where G is an empirical factor, and $|G| < |c|$; B denotes a damping factor, which is typically approximated as c, where c denotes the total attenuation coefficient; ω represents the radial frequency, and F^{-1} denotes the inverse Fourier transform.

The backscattering component arises from the light scattered at large angles into the camera and, thus, cannot be approximated with the small-angle theory. The three-dimensional space is discretized into N planar layers with a thickness Δz_i, each of which is

parallel to the image plane. The irradiance within each layer is modeled as a combination of direct and forward scattering components. Then, the overall backscattering component can be obtained by weighted integration over all layers utilizing the volume scattering function as follows:

$$E_b(x,y)=E_{bd}(x,y)+E_{bd}(x,y)*g(x,y|R,G,c,B) \tag{2.34}$$

$$E_{bd}(x,y)=\sum_{i=1}^{n} e^{-cZ_{ci}}\beta(\varnothing_b)E_s(x',y',z')\frac{\pi\Delta z_i}{4F^2}cos^3\theta T_l\left[\frac{z_{ci}-F_l}{z_{ci}}\right]^2 \tag{2.35}$$

where $\beta(\emptyset_b), N,$ and $E_s(x',y',z')$ represent the volume scattering function for scattering angle $\emptyset_b$, the number of discretized layers between the camera and the target plane, and $E_s(x',y',z')$ the irradiance of the source in the 3D space, respectively. $E_s(x',y',z')$ can be calculated as follows:

$$\begin{aligned} E_s(x',y',z') &= B(\theta_s,\varphi_s)\frac{\exp(-cz_{si})}{Z_{si}} \\ &+B(\theta_s,\varphi_s)\frac{\exp(-cz_{si})}{Z_{si}}*g(x',y',z'\mid Z_{si},G,c,B) \end{aligned} \tag{2.36}$$

When the illumination is provided by an artificial light source, an underwater irradiance $B(\theta_s,\varphi_s)$ can be obtained utilizing the underwater active light field model described in the previous section.

Simplified Underwater Imaging Model Based on Beer–Lambert Law

The classical underwater imaging model comprehensively considers the influences of camera-light source geometry and environmental parameters on the imaging process. However, its strong dependence on intricate optical parameters and spatial variables can

significantly complicate practical analysis and computational implementation. To overcome these limitations, this section introduces a simplified underwater imaging model based on the Beer–Lambert law. This streamlined model is specifically designed to accommodate the challenges posed by dark and turbid underwater conditions, providing essential theoretical support for subsequent image processing and spatial localization tasks.

Beer–Lambert Law

The Beer–Lambert law (Swinehart, 1962) defines the mathematical relationship between a propagation medium's absorbance and its optical path length and the concentration of absorbing species in the medium, and it is defined as

$$K_{absorbance} = \lg\left(\frac{I}{I_0}\right) = -\tau_{\lambda}\eta L \tag{2.37}$$

Where:

$K_{absorbance}$ represents the absorbance of the medium;

τ_{λ}denotes the molar absorption coefficient, which depends on the physical properties of the absorbing substance and a wavelength λ;

η represents the concentration of the absorbing substance;

L denotes the optical path length through the homogeneous medium;

I_0 represents the incident light intensity at the initial position, and

I denotes the transmitted light intensity after propagating a distance d through the medium.

The above-presented relationship can be adopted to construct an exponential decay model for the transmitted light intensity as follows:

$$I = I_0 \exp(-\tau_\lambda \eta L) \tag{2.38}$$

Underwater Imaging Modeling Under Complex Illumination Conditions

In turbid underwater environments, suspended particulate matter significantly affects light propagation, often rendering natural illumination insufficient for effective visual imaging. Consequently, performing artificial illumination is necessary to facilitate underwater visual perception tasks. Therefore, this section systematically examines the propagation characteristics of different illumination types, including both natural sunlight and artificial sources, as they travel toward an imaging sensor. In addition, it thoroughly analyzes the key factors that affect the quality of underwater images and proposes a multi-stage attenuation imaging model explicitly designed for complex illumination conditions. This comprehensive model considers contributions from both natural and artificial illumination sources and is further simplified and refined to address the challenges posed by dark and highly turbid underwater environments. A schematic of an underwater scene combining solar and artificial illumination sources is displayed in **Figure2.8** to support the theoretical descriptions.

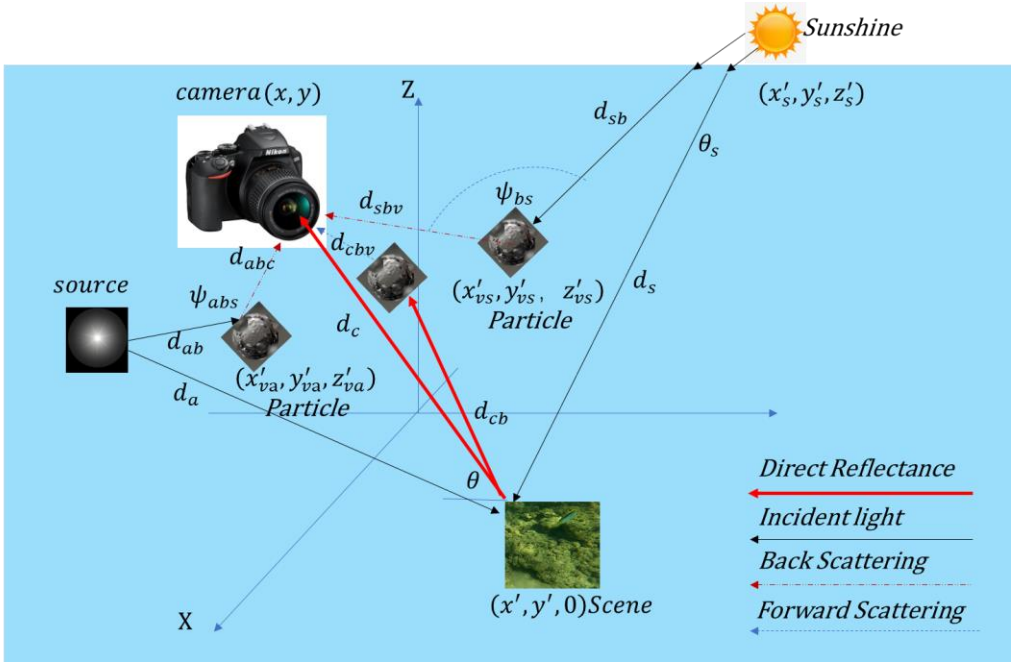


Figure2.8. Illustration of complex lighting underwater imaging environment.

Light emitted from diverse sources, including both artificial lighting and sunlight, enters the water and undergoes two main pathways. Namely, one portion of the light reaches the scene surface, reflects off the target, and then propagates toward the camera sensor. This light includes both the direct reflection component and the forward scattering component. Suspended particles in the water scatter the other portion of the light before reaching the scene and is, instead, directly scattered into the camera sensor, thus constituting a backscattering component. Therefore, the total irradiance received by the camera can be expressed as follows:

$$I(\lambda \mid x, y) = E_d(\lambda \mid x, y) + E_f(\lambda \mid x, y) + E_{bs}(\lambda \mid x, y) + E_{ba}(\lambda \mid x, y) \tag{2.39}$$

Where:

$I(\lambda|x,y)$ represents the total irradiance received by the camera at a wavelength λ and image coordinates (x,y);

$E_d(\lambda|x,y)$ denotes the direct reflection component received by the camera;

$E_f(\lambda|x,y)$ represents the forward scattering component received by the camera;

$E_{bs}(\lambda|x,y)$ denotes the backscattering component originating from the natural light received by the camera;

$E_{ba}(\lambda|x,y)$ represents the backscattering component originating from the artificial light received by the camera.

Scene-Incident Light Attenuation Analysis

Illumination from different sources contributes to underwater imaging by reaching the scene surface. However, because of the presence of suspended particles in the water, the light is attenuated during its propagation from the source to the scene. Depending on the source type, incident light can be categorized into natural light (e.g., sunlight) and artificial light (e.g., from LEDs). According to the Beer–Lambert law, the attenuation of natural light before reaching the scene surface can be expressed as follows:

$$\begin{aligned} E_s(\lambda \mid x', y', 0) &= E_s(\lambda \mid x_s', y_s', z_s') \exp(-c(\lambda) d_s) \\ &+ E_s(\lambda \mid x_s', y_s', z_s') \exp(-c(\lambda) d_s) * g(d_s) \end{aligned} \tag{2.40}$$

where:

$E_s(\lambda|x',y',0)$ denotes the incident irradiance at a point on the scene surface at a wavelength λ;

$E_s(\lambda|x_s', y_s', z_s')$ represents the initial irradiance of natural light as it enters the water medium;

$g(\cdot)$ denotes the PSF describing spatial dispersion during natural light propagation, and $c(\lambda)$ represents the attenuation coefficient for natural light at a wavelength λ.

The attenuation of incident light from artificial sources before reaching the scene surface can be calculated as follows:

$$\begin{aligned} E_a(\lambda \mid x', y', 0) &= E_a(\lambda \mid x_a', y_a', z_a') \frac{\exp(-c(\lambda) d_a)}{d_a^2} \\ &+ E_a(\lambda \mid x_a', y_a', z_a') \frac{\exp(-c(\lambda) d_a)}{d_a^2} * g(d_a) \end{aligned} \tag{2.41}$$

Similarly, the attenuation of incident light from artificial sources before reaching the scene surface can be expressed as follows:

$$E_{sr}(\lambda \mid x', y', 0) = E_s(\lambda \mid x', y', 0) + E_a(\lambda \mid x', y', 0) \tag{2.42}$$

Attenuation and Simplification of Direct Reflection from Scene Surface

After reflecting off the scene surface, the light carries information about the scene and forms a reflected component, whose intensity can be calculated by:

$$E_r(x', y', 0) = E_{sr}(\lambda \mid x', y', 0) * \rho(\lambda \mid x', y', 0) \tag{2.43}$$

$$E_r(x', y', 0) = (E_s(\lambda \mid x', y', 0) + E_a(\lambda \mid x', y', 0)) * \rho(\lambda \mid x', y', 0) \tag{2.44}$$

Most of the reflected light from the scene surface travels directly into the camera sensor, forming a direct component of the captured image, which can be expressed as:

$$\mathrm{E_d}(\lambda|\mathrm{x}, \mathrm{y}) = \mathrm{E_r}(\mathrm{x}', \mathrm{y}', 0) \exp\left(-\mathrm{c}(\lambda)\mathrm{d_c}\right) \frac{1}{4\mathrm{F}^2} \mathrm{T_l} \cos^4\theta \left[\frac{\mathrm{d_c} - \mathrm{F_l}}{\mathrm{d_c}}\right]^2 \tag{2.45}$$

$$E_d(\lambda|x,y) = ((E_s(\lambda \mid x', y', 0) + E_a(\lambda \mid x', y', 0))n$$
$$*\rho(\lambda \mid x', y', 0)\exp(-c(\lambda)d_c)\frac{1}{4F^2}T_l\cos^4\theta\left[\frac{d_c - F_l}{d_c}\right]^2 \quad (2.46)$$

The forward scattering of incident light contributes to the overall irradiance at the scene surface, but its magnitude is significantly smaller than the direct reflection component. Therefore, the forward scattering of the incident light is neglected in the simplified model. Hence, the direct component of the image can be approximated as follows:

$$E_d(\lambda|x,y)=((E_s(\lambda|x_s', y_s', z_s')\exp(-c(\lambda)d_s)+E_a(\lambda|x_a', y_a', z_a')\frac{\exp(-c(\lambda)d_a)}{d_a^2})$$
$$*\rho(\lambda \mid x', y', 0)\exp(-c(\lambda)d_c)\frac{1}{4F^2}T_l\cos^4\theta\left[\frac{d_c - F_l}{d_c}\right]^2 \quad (2.47)$$

Next, define the reflected irradiance from the underwater scene surface due to unattenuated natural light as follows:

$$J_s(\lambda|x,y)=E_s(\lambda|x_s', y_s', z_s')\rho(x', y', 0) \quad (2.48)$$

Similarly, define the reflected irradiance from the underwater scene surface due to unattenuated artificial light by

$$J_a(\lambda|x,y)=E_a(\lambda|x_a', y_a', z_a')\rho(x', y', 0) \quad (2.49)$$

Further, define camera parameters as follows:

$$T_c = \frac{1}{4F^2}T_l \quad (2.50)$$

Given that the underwater camera's focal length, F_l, is relatively small compared to the object-to-camera distance, d_c, the effect of focal length can be neglected; thus, the expression simplifies to:

$$\frac{d_c - F_l}{d_c} \approx 1 \tag{2.51}$$

Finally, the direct reflection component of the image can be expressed as follows:

$$E_d(\lambda|x, y) = J_s(\lambda|x, y)\exp(-c(\lambda)(d_s + d_c)) + J_a(\lambda|x, y)\exp(-c(\lambda)(d_a + d_c))/d_a^2 \tag{2.52}$$

Forward Scattering Component Under Complex Illumination Conditions

The forward scattering component can be defined by

$$E_f(\lambda \mid x, y) = E_d(\lambda \mid x, y) * g(d_c) \tag{2.53}$$

The PSF $g(d_c)$, which depends on the light propagation distance (d_c), can be expressed as

$$g(d_c) = [\exp(-Gd_c) - \exp(c(\lambda)d_c)]F^{-1}(\exp(-Bd_c\omega)) \tag{2.54}$$

where *G*, *B*, *w*, and F^{-1} represent an empirical parameter, which is typically less than or equal to the attenuation coefficient of the water medium, a damping coefficient indicating empirical attenuation, the radial frequency, and the inverse Fourier transform, respectively. The PSF is derived based on the small-angle scattering theory and, is thus applicable only under such assumptions.

Analysis and Simplification of Backscattering Components Under Complex Illumination Conditions

Backscattering Due to Natural Light

$$E_{bs}(\lambda \mid x, y) = \int_0^{d_{sbc}} \int_{\Theta}^{\psi_{bs}} E_s(\lambda \mid x_{vs}^{'}, y_{vs}^{'}, z_{vs}^{'}) \beta(\psi_{bs}) \exp(-c(\lambda) d_{sbc}) \frac{\pi}{4F^2} T_l cos^3\theta \left[1 - \frac{F_l}{d_{sbc}}\right]^2 d(d_{sbc}) d(\psi_{bs}) \tag{2.55}$$

where $(x'_{vs}, y'_{vs}, z'_{vs})$ denotes coordinates of the scattering volume element, $E_s(\lambda | x'_{vs}, y'_{vs}, z'_{vs})$ represents the natural light irradiance incident on the scattering volume, and $\beta(\psi_{bs})$ is the volume scattering function at a scattering angle ψ_{bs}. ψ_{bs} denotes the scattering angle between the incoming and outgoing light directions, F represents the aperture number of the camera. T_l represents the transmittance of the camera lens; F_l denotes the focal length of the camera lens; θ represents the angle between the scattered light direction and the optical axis of the camera; Θ denotes a set comprising all possible scattering directions from the same volume element, d_{sbc} represents the distance between the scattering volume element and the camera.

The attenuation of natural light during its propagation toward the scattering volume element can be expressed as follows:

$$E_s(\lambda \mid x_{vs}^{'}, y_{vs}^{'}, z_{vs}^{'}) = E_s(\lambda \mid x_{sb}^{'}, y_{sb}^{'}, z_{sb}^{'}) \exp(-c(\lambda) d_{sb}) + E_s(\lambda \mid x_{sb}^{'}, y_{sb}^{'}, z_{sb}^{'}) \exp(-c(\lambda) d_{sb}) * g(d_{sb}) \tag{2.56}$$

Assuming the distance between the camera and the suspended particle is significantly larger than the camera's focal length, the focal length can be considered negligible. Hence, the following approximation can be introduced:

$$1-\frac{F_l}{d_{sb}}\approx 1 \tag{2.57}$$

Next, define the camera-related parameters as follows:

$$T_{sb}=\frac{\pi}{4F^2}T_l \tag{2.58}$$

Then, define the background illumination at optical infinity by

$$B_\infty=\frac{E_s(\lambda\mid x_{vs}^{'},y_{vs}^{'},z_{vs}^{'})T_{sb}}{c(\lambda)}\int_\Theta^0\beta(\psi_{bs})d(\psi_{bs}) \tag{2.59}$$

By defining an imaging distance d_c as an effective backscattering distance, the backscattering component due to natural light can be expressed as follows:

$$E_{bs}(\lambda\mid x,y)=B_\infty(1-\exp(-c(\lambda)d_c)) \tag{2.60}$$

Backscattering Due to Artificial Light

The intensity of illumination of artificial light is typically much higher than that of natural light, which aggravates the scattering effect in the water and introduces spatially non-uniform lighting fields. The backscattering component induced by the artificial light source can be expressed as follows:

$$E_{ba}(\lambda\mid x,y)=\int_0^{d_{abc}}\int_\Theta^{\psi_{abs}}E_a(\lambda\mid x_{va}^{'},y_{va}^{'},z_{va}^{'})\beta(\psi_{abs})\exp\left(-c(\lambda)d_{abc}\right)\frac{\pi}{4F^2}T_l cos^3\theta\left[1-\frac{F_l}{d_{abc}}\right]^2 d(d_{abc})d(\psi_{abs}) \tag{2.61}$$

The artificial light irradiance at the scattering volume element, denoted by $E_a(\lambda|x'_{va},y'_{va},z'_{va})$, can be obtained by the underwater active light field model presented in the previous section.

Considering that the camera's focal length is generally much smaller than the distance from the camera to the suspended particles, the following approximation can be applied:

$$1-\frac{F_l}{d_{ab}} \approx 1 \tag{2.62}$$

Next, define the camera-related parameters as follows:

$$T_{sb} = \frac{\pi}{4F^2} T_l \tag{2.63}$$

Further, define the background illumination induced by the artificial light source by

$$B_{ba} = \frac{E_a(\lambda \mid x_{vs}^{'}, y_{vs}^{'}, z_{vs}^{'}) T_{sb}}{c(\lambda)} \int_{\Theta}^{\cdot} \beta(\psi_{abs}) d(\psi_{abs}) \tag{2.64}$$

Therefore, the backscattering component induced by the artificial light source can be expressed as follows:

$$E_{ba}(\lambda \mid x, y) = B_{ba}(1-\exp(-c(\lambda)d_c)) \tag{2.65}$$

Under complex illumination conditions, the total backscattering component is expressed by

$$E_b(\lambda \mid x, y)) = E_{ba}(\lambda \mid x, y) + E_{bs}(\lambda \mid x, y) = (B_{ba} + B_{bs})(1-\exp(-c(\lambda)d_c)) \tag{2.66}$$

Simplified Formulation of Complex Illumination Imaging Model

Here, the complex multi-stage attenuation imaging model is simplified for practical application. In the simplified model, the total radiance received by a camera comprises direct reflection from the scene, forward scattering from the scene, backscattering caused by natural light, and backscattering caused by artificial light. Accordingly, the final image formation model can be expressed as follows:

$$I(\lambda \mid x,y) = J_s(\lambda \mid x,y)\exp(-c(\lambda)(d_s + d_c)) + J_a(\lambda \mid x,y)\exp(-c(\lambda)(d_a + d_c))/d_a^2$$
$$+(J_s(\lambda \mid x,y)\exp(-c(\lambda)(d_s + d_c)) + J_a(\lambda \mid x,y)\exp(-c(\lambda)(d_a + d_c))/d_a^2) * g(d_c) \quad (2.67)$$
$$+(B_{ba} + B_{bs})(1-\exp(-c(\lambda)d_c))$$

QUALITY IMPROVEMENT METHODS FOR UNDERWATER IMAGES: EXPERIMENTAL COMPARISONS

When evaluating the performance of underwater image enhancement algorithms, both full- and no-reference image quality metrics are typically employed. In view of this, this section introduces four representative evaluation methods: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), underwater color image quality evaluation (UCIQE), and underwater image quality measure (UIQM).

PSNR

The PSNR is a full-reference metric that quantifies image quality based on the pixel-wise difference between the enhanced image and the corresponding ground-truth reference. It is defined as a logarithmic ratio between the maximum possible pixel value and the mean squared error (MSE) between the two images. A higher PSNR value generally indicates better image fidelity and less distortion and vice versa (Hore & Ziou, 2010).

SSIM

The SSIM is another full-reference metric that evaluates image similarity by comparing structural information, luminance, and contrast between the reference and test images. Unlike the PSNR metric, the SSIM metric is designed to align more closely with human visual perception (Z. Wang, 2004).

UCIQE

The UCIQE is a no-reference image quality metric designed explicitly for underwater images. This metric integrates three key color-related features—chroma variance, saturation, and luminance contrast—to reflect the overall perceptual quality of color images taken in underwater environments (Yang & Sowmya, 2015).

UIQM

The UIQM is another no-reference metric designed for underwater images. This metric includes three weighted subcomponents: colorfulness (UICM), sharpness (UISM), and contrast (UIConM). The UIQM metric provides a composite score that reflects the overall perceptual quality under different underwater conditions (Yang & Sowmya, 2015).

Underwater Imaging Model Simulation Experiments

The quality of underwater images fundamentally depends on the optical transmission characteristics of the water medium, particularly the attenuation effects, which manifest as reductions in sharpness and brightness, caused by absorption and scattering processes. The classical underwater imaging model (i.e., the Jaffe–McGlamery model) decomposes incoming light into direct transmission, forward scattering, and backscattering components. Although previous studies have suggested that forward scattering may have minimal effects when the camera-target distance is relatively short, systematic experimental validation of this assumption remains limited.

To address this research gap, this study empirically investigates the relative influence of forward and backscattering on image quality for different turbidity levels. Specifically,

the primary goal is to assess the practical applicability of a simplified underwater imaging model by conducting an empirical examination of its suitability for model optimization and subsequent image restoration algorithm design.

The primary objectives of the experimental investigation are threefold. The primary objective is to evaluate the impact of forward scattering on image resolution and structural details under different turbidity conditions. The second objective refers to analyzing the predominant role of backscattering in contributing to brightness reduction and image-blurring effects in underwater imaging. Finally, the third objective is to provide experimental evidence for guiding the selection and validation of an appropriate imaging model for effective underwater image restoration tasks.

Experimental Design and Variable Configuration

To systematically analyze variations in underwater image quality, this experiment employed a comparative analysis framework that included both classical and simplified underwater imaging models. Specifically, the classical model utilized the comprehensive formulation, explicitly incorporating forward scattering, whereas the simplified model omitted the forward scattering component, considering only direct transmission and backscattering effects.

Aiming to ensure data stability, comparability, and reproducibility, all experimental image acquisitions were conducted in a controlled underwater simulation environment. A constant-power white LED was utilized as an illumination source, and it was precisely aligned along the optical axis above the camera to ensure uniform and stable illumination

with accurately controlled incidence angles and intensity. The imaging target was a standardized resolution test chart (USAF 1951) that featured black-and-white stripe patterns at different spatial frequencies, with a minimum line width of 1 cm. This test chart facilitated the evaluation of the model's ability to preserve fine structural details in reconstructed images. During image capture, the distance between the camera and the imaging target was consistently 1 m, and all relevant physical parameters were strictly held constant. In addition, image resolution was uniformly standardized at 640 × 480 pixels across all experimental trials.

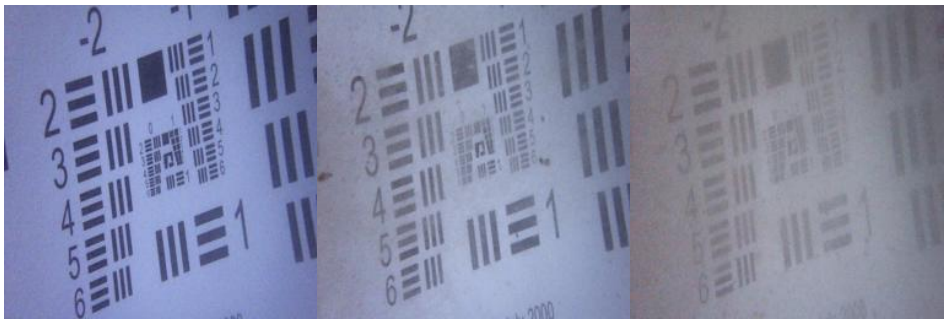
Two primary independent variables were systematically investigated in the experimental design: water turbidity level and imaging model type. The water turbidity level was categorized into low, moderate, and high by adjusting the concentrations of suspended particulate matter in the test environment. The imaging models adopted in the experiments included: the classical model, which explicitly incorporated forward scattering, and the simplified model ("Ours"), which excluded forward scattering effects. By systematically varying the turbidity level and imaging model, the reconstruction performance under different underwater optical conditions was comprehensively evaluated.

Two widely recognized image quality metrics, the PSNR and MSE, were adopted for quantitative assessment of image reconstruction quality. The PSNR metric was utilized to quantify pixel-level differences between reconstructed images and their reference counterparts, effectively reflecting restoration fidelity. In contrast, the MSE metric was employed to determine the average squared differences in pixel intensities; smaller MSE

values indicated greater similarity to the original images and vice versa. The two metrics together ensured an objective and robust evaluation of the comparative effectiveness of each imaging model for different degrees of water turbidity.

Experimental Data

The image data were systematically acquired in the underwater simulation environment for different turbidity levels under the condition of a consistent viewpoint and a uniform scene configuration. The dataset comprised images captured under clear, moderately turbid, and highly turbid water conditions, as illustrated in **Figure2.9**. Each image incorporated standardized line patterns specifically designed to assess image clarity, allowing for the evaluation of the visual impacts caused by scattering and absorption on fine structural details.



(a) (b) (c)

Figure2.9. Underwater simulation environment images: (a) clear image; (b) slightly turbid image; (c) heavily turbid image.

Experimental Procedure

A controlled experimental platform was established with a water tank housed in a carefully controlled darkroom environment, and a standard-resolution test chart served as an imaging

target. By systematically adjusting the concentration of suspended particulate matter in the water medium, multiple imaging scenarios corresponding to different turbidity levels were designed, thus ensuring both reproducibility and representativeness of the collected image data. Image acquisition was consistently performed at a fixed camera position, a stable target-to-camera distance, and under uniform illumination conditions. The resulting dataset encompassed two representative degradation scenarios: a mild-blur scenario, corresponding to low turbidity, and a severe-blur scenario, indicating highly turbid conditions.

The acquired images were utilized in the image reconstruction simulations employing two distinct imaging models: the classical model ("JM"), which explicitly incorporated forward scattering effects, and the simplified model ("Ours"), which omitted forward scattering contributions. A comparative analysis of the restoration performance of the two models was conducted under different levels of image degradation.

Finally, quantitative evaluations of the reconstructed image quality were conducted with two objective metrics, the PSNR and MSE, followed by qualitative visual assessments. This evaluation approach constituted a comprehensive validation framework that systematically linked underlying physical degradation mechanisms, model-based restoration strategies, and quantitative performance metrics, ensuring scientific rigor and completeness of the experimental results.

Expected Outcomes and Significance of Experimental Results

It was expected that under conditions of high turbidity, the effects of backscattering, which manifest mainly as increased brightness and pronounced image blurring, would become markedly more prominent. In contrast, the contribution from forward scattering was expected to remain relatively negligible. In addition, the observed differences in image quality between the simplified and classical imaging models would prove minimal across all tested turbidity levels, suggesting that forward scattering did not play a decisive role in short-range underwater imaging scenarios. This outcome would provide robust empirical justification for adopting the simplified imaging model as a theoretical foundation for developing computationally efficient methods for enhancing underwater images.

Experimental Setup and Workflow

In this experiment, the accuracy of image reconstruction served as a dependent variable, and two primary independent variables included water turbidity level and imaging model structure. As mentioned before, the water turbidity level was systematically categorized into low, medium, and high, to indicate different degrees of underwater optical degradation. Furthermore, as previously mentioned, the imaging model structure comprised two distinct frameworks: the classical imaging model, which explicitly incorporated forward scattering effects, and the simplified imaging model, which excluded forward scattering, thereby ensuring a rigorous comparative analysis of the impact of forward scattering modeling on the image reconstruction performance.

The PSNR and MSE were employed to assess image clarity and reconstruction fidelity. Collectively, these metrics establish a clear causal relationship between the turbidity level and the imaging model structure enabling the evaluation of image quality, providing comprehensive theoretical and empirical foundations for performance assessment and subsequent model selection decisions.

Experimental Results

Objective Analysis

To systematically evaluate performance differences between the two models, the simplified imaging model ("Ours") and the classical imaging model ("JM"), in image reconstruction, two representative image scenarios were designed, corresponding to the moderate blur and high turbidity conditions. As mentioned before, the MSE and PSNR metrics were employed to conduct a rigorous comparative analysis of reconstruction accuracy. The experimental results obtained from the analyses are presented in **Figure2.10** and **Figure2.11.**

As demonstrated in **Figure2.10**, the MSE results demonstrated the superior performance of the proposed model. Specifically, the proposed method consistently achieved substantially lower MSE values compared to both the original classical model and its simplified variant under all turbidity conditions. Notably, under moderate turbidity, the proposed model achieved the lowest MSE among all the models, thus considerably outperforming the classical model, which indicated its enhanced accuracy in replicating realistic underwater imaging conditions.

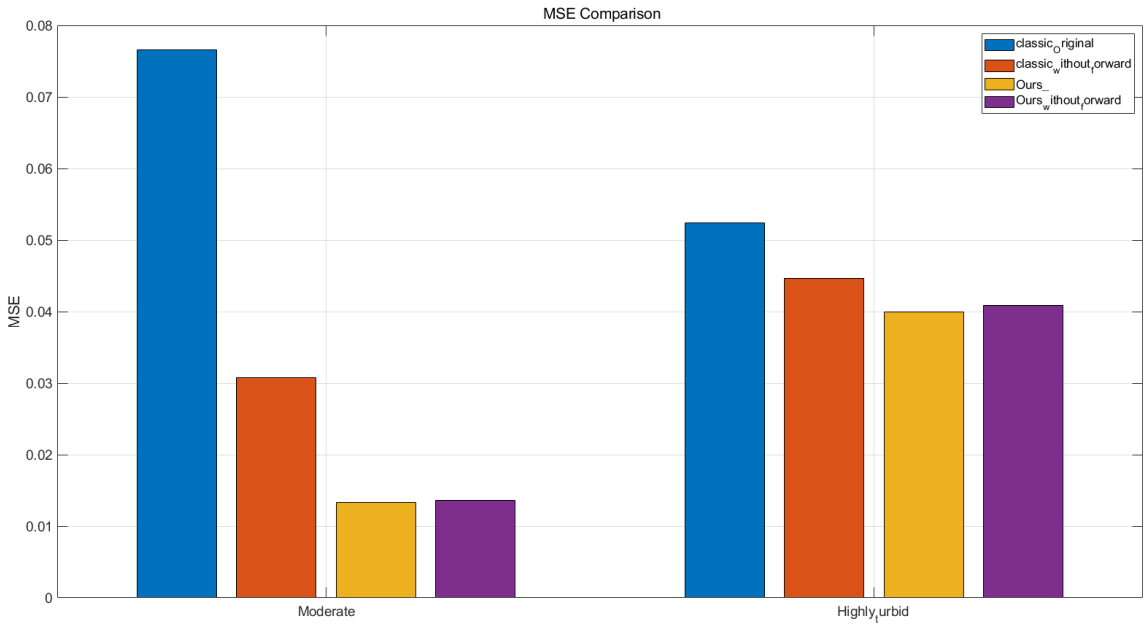


Figure2.10. Comparison results of image reconstruction error (MSE) between classic and simplified models under moderately and highly turbid water conditions. The simplified model consistently achieved lower MSE in both scenarios, indicating improved pixel-level fidelity.

Further analysis of the PSNR metric confirmed these findings, as illustrated in **Figure2.11**. Namely, under moderate turbidity conditions, the proposed model achieved notably higher PSNR values compared to the other models, which indicated significantly improved image quality with more accurately reconstructed details. Even under highly turbid conditions, the proposed model demonstrated superior performance relative to the classical model, suggesting its robust performance under challenging environmental conditions and confirming its enhanced capability in capturing intricate details of underwater optical propagation.

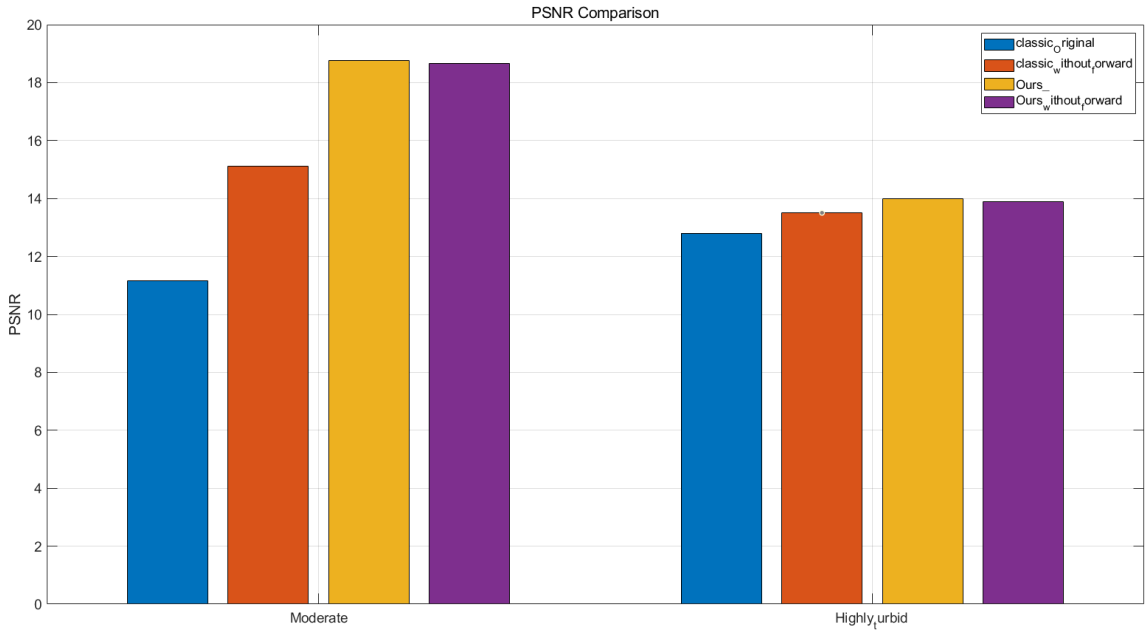


Figure2.11. Comparison results of PSNR metric for classic and simplified models for moderate and high turbidity levels. The simplified model slightly outperformed the classic model in both cases, demonstrating better structural restoration quality.

As presented in **Table 2.1** and **Table 2.2**, the proposed model consistently demonstrated superior performance in both moderate and highly turbid scenarios compared to the classical Jaffe–McGlamery (JM) model. Under moderate turbidity conditions, the proposed model achieved MSE and PSNR of 0.01 and 18.75, respectively, thus significantly outperforming the original JM model, with an MSE and PSNR of 0.07 and 11.16, respectively, and its simplified variant without forward scattering, which had MSE and PSNR of 0.03 and 15.11, respectively. Even when forward scattering was excluded from the proposed model, the model's performance remained substantially better than that of both versions of the JM model, reaching MSE and PSNR of 0.01 and 18.67, respectively.

Table 2.1. Comparison results of proposed and JM models obtained under moderate turbidity level.

	Moderate	**MSE**	**PSNR**
Ours	New Model	0.01	18.75
	Without Forward	0.01	18.67
JM	Original	0.07	11.16
	Without Forward	0.03	15.11

Table 2.2. Comparison results of proposed and JM models obtained under high turbidity level.

	Highly Turbid	**MSE**	**PSNR**
Ours	New Model	0.03	13.98
	Without Forward	0.04	13.88
JM	Original	0.05	12.80
	without Forward	0.04	13.50

In highly turbid conditions, the proposed model maintained its superiority, achieving MSE and PSNR of 0.03 and 13.98, respectively, and thus again surpassing the original JM model, with MSE and PSNR of 0.05 and 12.80, respectively, and its simplified variant, with MSE and PSNR of 0.04 and 13.50, respectively. The consistent advantage observed in the proposed model's results highlighted the proposed model's robustness and accuracy in simulating realistic underwater optical environments for different turbidity levels.

In summary, the quantitative evaluations unequivocally demonstrated the superior performance of the proposed simplified underwater imaging model over the classical models. The consistent advantages observed in terms of both the MSE and PSNR metrics effectively validated the experimental objectives, confirming that the proposed simplified

model could provide enhanced accuracy and perceptual quality for realistic underwater imaging applications.

Subjective Visual Comparison

The visual comparison of the underwater images reconstructed by different models provided intuitive and critical insights into the effectiveness of the proposed simplified underwater imaging model compared to the classical JM model. In **Figure2.12**, the reconstruction results obtained under moderate and highly turbid conditions utilizing the two models and their variants without forward scattering are depicted.

(a) Clear Image (b) JM (c) JM without forward (e) Ours (e) Ours without forward

Moderate

(a) Clear Image (b) JM (c) JM without forward (e) Ours (e) Ours without forward

Highly Turbid

Figure2.12. Visual comparison of images reconstructed by the proposed and JM models.

In the moderate-turbidity scenario, the classical JM model demonstrated noticeable deterioration, as illustrated in **Figure2.12(b)**, displaying blurred details and significantly reduced contrast compared to the clear image in **Figure2.12(a)**. The simplified variant of the JM model without forward scattering achieved marginal improvement, but fine details remained inadequately resolved, as illustrated in **Figure2.12(c)**. In contrast, the simplified version of the proposed model achieved a substantial enhancement, providing sharper contours, improved contrast, and more precise preservation of intricate details, as illustrated in **Figure2.12(e)**. Even the proposed model's variant without forward scattering notably outperformed both JM versions, as presented in **Figure2.12(e)**. This result highlighted the inherent robustness of the proposed model.

The results indicated that under highly turbid conditions, the limitations of the classical JM model became even more apparent, exhibiting severely blurred and indistinct image details. Its simplified version without forward scattering demonstrated a slight improvement, but still remained insufficient for clear detail restoration. In contrast, the proposed simplified model could still maintain relatively higher image quality, preserving more visible structural features and achieving better overall clarity, as illustrated in **Figure 2.12(e)**. The visual results confirmed that the proposed model could maintain superior performance even under severely adverse conditions.

In summary, the subjective visual comparisons reinforced the objective evaluation results, clearly demonstrating that the proposed simplified underwater imaging model could significantly improve visual quality and robustness under different turbidity levels.

The visual evidence firmly supported the proposed model's practical advantages and superior fidelity.

This chapter initially introduces the fundamental mechanisms of light attenuation in water, including absorption and scattering mechanisms, and carefully analyzes their detrimental impacts on imaging quality. Moreover, a physically based active illumination model specifically designed for turbid underwater environments is developed, providing a robust theoretical foundation for characterizing light intensity distribution in submerged scenes. In addition to the classical underwater imaging model, a multi-stage attenuation model is proposed to leverage advantages from both natural and artificial light sources under complex illumination conditions. By recognizing the limitations inherent in the classical model, particularly its overly intricate treatment of forward scattering, this chapter introduces a simplified imaging model specifically adapted to dark, turbid, and low-visibility underwater environments.

Conclusion

This chapter systematically analyzes and identifies the key factors affecting underwater optical imaging, emphasizing the inherent characteristics of light absorption and scattering prevalent in turbid underwater environments. A comprehensive artificial illumination model at the air–glass–water interface is developed, serving as a foundational framework for studying the intensity distribution of underwater illumination fields.

Considering the classical imaging model based on the Beer–Lambert law, this chapter addresses the significant limitations of this model, particularly its incomplete consideration

of light propagation paths, by establishing an advanced multi-stage attenuation imaging model that considers both natural and complex artificial lighting conditions. The model is subsequently simplified for practical applicability in highly turbid underwater scenarios.

Furthermore, systematic experiments are conducted, and the resolution test charts are utilized to validate the impact of forward scattering on image resolution and structural detail preservation under different turbidity level conditions. The experimental results conclusively highlight the dominant role of backward scattering in contributing to image brightness deterioration and increased blurriness.

Collectively, the theoretical and experimental results presented here provide a robust foundation for enhancing underwater imaging quality and ensure essential support for subsequent research on precise underwater pose estimation methods explored here.

CHAPTER THREE HYBRID UNDERWATER IMAGE ENHANCEMENT VIA DUAL TRANSMISSION OPTIMIZATION AND TRANSFORMER-BASED FEATURE FUSION

In dark and turbid underwater environments, light undergoes severe absorption and scattering, resulting in substantial image degradation. Conventional image enhancement approaches, whether physics-based or learning-driven, exhibit inherent limitations under such conditions. This chapter introduces a novel hybrid underwater image enhancement framework that synergistically integrates dual transmission optimization with transformer-based feature fusion. Specifically, a dual transmission estimation strategy—guided by boundary constraints and local contrast priors—is employed to compensate for wavelength-dependent attenuation robustly. An adaptive ambient light estimation and color correction module is further introduced to restore accurate color representations and improve visual clarity. Subsequently, a transformer-based enhancement network (Uformer) is adopted, where the physical priors derived from the dual transmission maps are explicitly embedded to guide global attention, thereby enhancing structural fidelity and supporting precise visual localization. Extensive experiments conducted on multiple benchmark underwater datasets validate the proposed method's robustness and generalization ability across diverse scenarios. The enhanced images significantly improve the quality of feature extraction, thereby facilitating downstream applications such as monocular vision-based navigation and oceanic exploration for AUVs.

Introduction

A clear underwater image is the basis of underwater detection and operation. According to the analysis of the underwater imaging process in the previous chapter, inland water bodies such as rivers and lakes have numerous dissolved and particulate matter, light propagation in the water by the absorption of dissolved and particulate matter in the water and the effect of scattering and exponential attenuation. Absorption in water is related to the wavelength of light, which limits the imaging distance of the camera. Scattering introduces both forward scattering and backward scattering components to the imaging process. Forward scattering component is the scene reflected light in the propagation process before reaching the camera encountered in the suspended particles at a small angle scattered light, forward scattering effect leads to the image will lose some of the details and clarity, such as the edge of the object and the background of the contrast between the decline in the image of the sense of hierarchy and three-dimensional sense of weakening, giving a fuzzy, hazy feeling. The backward scattering component is the light reflected from particles in the water to the camera, which has no relation with the scene information, and its scattering effect intensifies with the increase of the imaging distance, reducing image contrast. Simultaneously, because of the underwater environment, it is difficult to achieve the ideal imaging lighting conditions. Usually, the introduction of non-uniform light field distribution of artificial light source for auxiliary illumination exacerbates the phenomenon of backward scattering and leads to the image of the foreground and the background of the non-uniform color degradation. Therefore, it is essential to study image processing algorithms to improve image quality.

In recent years, many scholars have conducted extensive research on underwater image processing, overcoming various technical challenges in underwater image preprocessing, and making significant contributions to the development of underwater image preprocessing technology. Their research methods can be divided into contrast enhancement methods, physical imaging modeling methods, and deep learning methods. Contrast enhancement methods improve visibility by adjusting brightness and contrast. Conventional methods such as histogram equalization (HE) (Zuiderveld, 1994), multi-scale Retinex (MSRCR) (Rahman et al., 1996) and adaptive Gamma correction with weighted distribution (AGCWD) (Rahman et al, 2016) have been widely adopted. However, these methods often lead to over-enhancement, unnatural color shifts, and amplified noise, making them unsuitable for complex underwater scenes. Physical imaging model-based restoration techniques rely on underwater imaging models to estimate light transmittance and background illumination, aiming to reconstruct realistic scene brightness. Commonly adopted methods include the underwater dark channel prior (UDCP) (Qian et al., 2023), red channel prior (RCP) (Qian et al., 2023), and decay curve prior. While these methods provide a better theoretical foundation, they often struggle to accurately estimate light transmission, especially in homogeneous lighting and dynamic underwater conditions. Deep learning-based methods have received attention for their ability to learn degenerate mappings from large datasets. Convolutional neural network (CNN) and Gan-based models, such as UWCNN (C. Li et al., 2020), CycleGAN-based augmentation (Liu, 2022), and UColor (C. Li, 2021), have demonstrated encouraging results. However, these methods face challenges such as high computational cost,

dependence on high-quality training datasets, and limited generalization to real underwater conditions. Recently, transformer-based models, especially Uformer, have exhibited promising results because of their ability to efficiently capture global and local contextual information. Despite these advances, deep learning methods themselves often lack sufficient interpretability and robustness to different underwater conditions, emphasizing the need to combine physical information models with deep learning frameworks.

Based on the need for accurate and robust image localization in underwater conditions, this paper proposes a hybrid underwater image enhancement framework that combines physical model-based dual-transmission optimization and adaptive color correction with a transformer-based feature fusion mechanism. The method explicitly exploits physical imaging properties, such as wavelength-dependent transmittance estimation and adaptive ambient light estimation, to efficiently correct color distortion and improve local contrast. Simultaneously, a transformer-based model (Uformer) is employed to fuse multi-scale global and local contextual information, further enhancing structural details that are critical for accurate visual localization.

The primary contributions of this dissertation include:

Dual Transmission Optimization: Development of a physically informed dual transmission estimation method that integrates boundary constraints and local contrast priors to robustly compensate for wavelength-dependent light attenuation effects and accurately restore degraded scene radiance.

Adaptive Ambient Light Estimation and Color Correction: Implementation of an adaptive ambient light estimation strategy inspired by the "Shades of Gray" method,

significantly enhancing color restoration accuracy and robustness under varied and complex underwater lighting conditions.

Transformer-Based Feature Fusion (with Uformer): Incorporation of a transformer-based architecture explicitly guided by the estimated transmission maps, thereby effectively capturing global context and local details essential for robust image restoration in underwater scenarios.

Experimental Validation and Localization Application: Comprehensive evaluation on diverse underwater image datasets demonstrating superior visual enhancement performance compared with state-of-the-art approaches. Furthermore, experimental validations confirm the effectiveness of the proposed enhancement methodology in significantly improving the accuracy and robustness of subsequent monocular camera-based localization tasks in dark and turbid underwater environments.

By systematically integrating physically grounded image enhancement methods with advanced transformer-based feature fusion techniques, this disserrtation offers a robust and effective solution for underwater image restoration, directly supporting more reliable and accurate image-based localization in challenging underwater conditions.

Literature Review

Underwater image restoration has been extensively studied because of the significant challenges posed by light absorption, scattering, and backscattering effects in turbid environments. Existing methods can be broadly categorized into contrast enhancement techniques, model-based restoration, and deep learning-based approaches. Although these

techniques have achieved varying degrees of success, they often suffer from color distortion, loss of detail, and limited adaptability to different underwater conditions. This section reviews the state-of-the-art methods, highlighting their advantages and limitations.

Contrast Enhancement-Based Methods

Contrast enhancement methods seek to improve the visual appeal of underwater images by adjusting the pixel intensity distribution without explicitly considering the physical degradation model of underwater imaging. These techniques are computationally efficient and suitable for real-time applications; however, they often fail to address wavelength-dependent attenuation and scattering effects, resulting in over-enhancement, unnatural color shifts, and amplified noise. HE is one of the early contrast enhancement methods HE and its adaptive variants such as contrast limited adaptive histogram equalization (CLAHE). These methods adjust image contrast but often lead to oversaturation and an unnatural appearance, especially in turbid underwater scenes. Other methods include MSRCR, which enhances brightness and contrast but usually introduces color artifacts due to improper light correction. Similarly, AGCWD and WGIF (Z. Li, 2014) have been proposed to dynamically adjust contrast while preserving details. However, these methods still struggle to compensate for uneven light attenuation under underwater conditions.

Recently, underwater-specific contrast enhancement methods have been introduced. For example, Li et al. (Z. Li, 2014) proposed a method that separates low- and high-frequency components and processes color, contrast, and noise independently. Guo et al. (Guo et al., 2019) developed a hybrid strategy based on minimum color loss and maximum

attenuation mapping, which improves visual perception but still lacks robustness in highly turbid environments.

Despite their efficiency, contrast enhancement methods fundamentally fail to address the underlying physics of underwater image degradation, making them unsuitable for high-precision underwater sensing tasks.

Model-Based Restoration Methods

Unlike contrast enhancement-based methods, model-based approaches leverage physical imaging models to estimate scene radiance by compensating for light attenuation and scattering. These methods aim to reconstruct the true scene color and details, but often suffer from estimation errors in dynamic underwater conditions.

The Jaffe-McGlamery model (Jaffe, 1990), (McGlamery, 1980) is widely adopted to describe the underwater light propagation process, where image formation is influenced by direct reflection, forward scattering, and backward scattering. Based on this, numerous restoration techniques have been developed.

The dark channel prior (DCP), originally introduced by He et al. (He et al., 2010), was later adapted for underwater images as the underwater dark channel prior (UDCP) (Drews, 2013). While effective for haze removal, UDCP often fails in highly attenuated red-light environments because it assumes that one color channel should always be near zero.

To improve transmittance estimation, Galdran et al. (Galdran et al, 2015) proposed the RCP, which focuses on reconstructing red-channel information. However, this method struggles with complex underwater lighting conditions and varying scene depths. Other

physics-based priors include: attenuation curve prior (Anwar & Li, 2020) and haze line prior (Zhuang, 2022), which leverage spectral attenuation models for improved restoration (Zhou, 2022). Color-line prior assumes that colors in natural images form a linear manifold, but this assumption often fails in turbid underwater environments. While model-based methods offer a more physically grounded approach compared to contrast enhancement techniques, they still face significant challenges, such as, inaccurate transmittance estimation in highly turbid and dynamic underwater environments, limited adaptation to different water types, as many priors rely on assumptions that do not hold universally, and computational complexity, making real-time deployment difficult for underwater robotic applications.

Our proposed dual transmittance estimation improves upon these methods by separately estimating transmittance for blue-green and red wavelengths, accurately modeling wavelength-dependent attenuation. Furthermore, our adaptive color bias fusion dynamically adjusts color restoration based on local scene properties, preventing over-enhancement and artifacts.

Deep Learning-Based Methods

With the rise of deep learning, many researchers have developed CNN- and GAN-based approaches for underwater image restoration. These methods attempt to learn the mapping between degraded and clear images from large datasets, demonstrating impressive performance under certain conditions. Underwater image enhancement convolutional neural network (UWCNN) (C. Li et al., 2020), a supervised model trained on synthetic

underwater datasets, which improves visibility but lacks generalization to real-world underwater conditions. CycleGAN-based approaches (Liu, 2022) utilize adversarial training to transform degraded underwater images into enhanced versions. However, they often produce unrealistic textures and color shifts. UColor integrates underwater transmission priors into a CNN model to enhance structure preservation. Despite their success, deep learning-based methods suffer from key limitations, such as, dependence on high-quality training datasets, which are often unavailable or unrealistic and limited generalization across diverse underwater conditions, particularly in highly turbid waters.

Transformers, originally developed in natural language processing, have proven effective in vision tasks like image classification, object detection, and segmentation because of their powerful global attention capabilities (S. Khan, 2022) (Carion, 2020) (Strudel, 2021). Image restoration and enhancement, models such as IPT (Chen, 2021) and SwinIR (Liang, 2021) exploit attention mechanisms to capture global relationships and fine details effectively. Uformer (Wang, 2022), a U-shaped transformer for general image restoration, utilizes locally enhanced window transformer blocks and multi-scale modulators to balance global context and local detail efficiently. Despite their computational complexity and data-intensive training requirements, transformer-based models like Uformer have exhibited remarkable potential for underwater image enhancement, provided sufficient training data and appropriate model constraints are employed.

Methodology

To address the challenges posed by underwater environments, including severe scattering, color distortion, and motion blur, we propose a novel underwater image enhancement method integrating dual transmission estimation, adaptive ambient light estimation with color correction, and a transformer-based multi-scale feature enhancement network (Uformer). Our proposed framework comprises three main modules: (1) dual transmission estimation for image restoration; (2) adaptive ambient light estimation and color correction; and (3) Uformer-based multi-scale transformer feature enhancement.

As is illustrated in **Figure 3.1**, the overall structure of our proposed underwater image enhancement framework is as follows. First, dual transmission estimation integrates boundary-constraint and local contrast methods to produce a robust transmission map through adaptive Gamma fusion. Second, adaptive ambient light estimation and color correction rectify color distortion effectively. Finally, the Uformer network explicitly incorporates the fused transmission map into the transformer's attention mechanism, yielding enhanced underwater images suitable for visual SLAM applications.

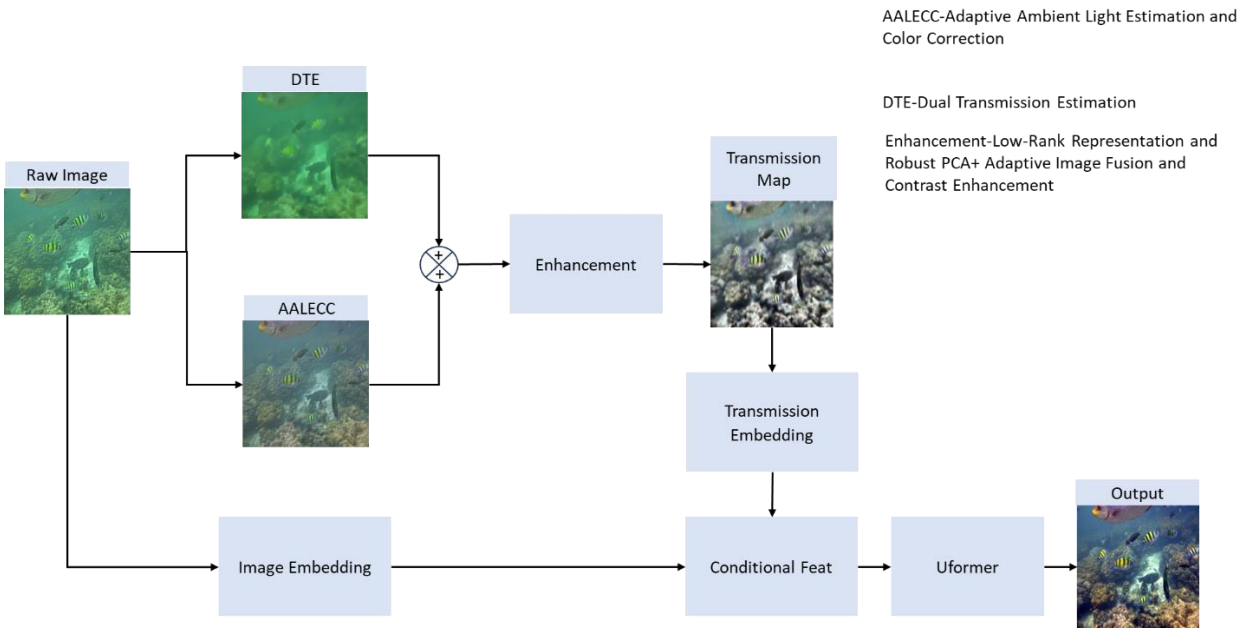


Figure 3.1. Algorithm overview.

Model 1 Transmission Map

Dual Transmission Estimation for Image Restoration

Typically, the backward scattering component is the main cause of image blurring in camera imaging, whereas the forward scattering component has a minimal impact on imaging. According to Equation (3.1), with the boundaries of the scene radiance established, we can determine the boundary-constrained points for the scene radiance. Hence, based on Equations (3.2) and (3.3), we can derive the boundary constraints for the transmission $t(\mathbf{x})$.

$$0 \le t_b^c(\mathbf{x}) \le t(\mathbf{x}) \le 1, c \in \{r, g, b\} \tag{3.1}$$

where $t_b^c(\mathbf{x})$ is the lower bound of the transmission, given by,

$$\tilde{t}^c(\mathbf{x}) = \min_{y \in \Omega(\mathbf{x}) z \in \Omega(\mathbf{y})} \max t_b^c(z), c \in \{r, g, b\} \tag{3.2}$$

where $I^c(\mathbf{x})$, A^c, C_0^c, and C_1^c represent the r、g、b color channels of I, A, C_0 and C_1, respectively.

For the transmission of an image, the size of the transmission in the local image block should be similar. Therefore, it can be assumed that in a local image block $\Omega(\mathbf{x})$, the transmission is constant, denoted as $\tilde{t}^c(\mathbf{x})$, which can be expressed as

$$\tilde{t}^c(\mathbf{x}) = \min_{y\in\Omega(\mathbf{x})} \max_{z\in\Omega(\mathbf{y})} t_b^c(z), c \in \{r, g, b\} \tag{3.3}$$

Estimating transmission often requires utilizing prior information from the image to provide a rough estimate. For example, He et al. proposed a method called dark channel prior (DCP) for foggy images. DCP assumes that at least one of the three color channels approaches zero in natural, non-foggy images. Therefore, in this method, the minimum value among the three-color channels in local image patches is utilized as the estimate for transmission $t(\mathbf{x})$. DCP has yielded good results in estimating foggy images. However, its assumption does not hold for underwater images, making it method unsuitable. Li et al. (C. Li, 2019) proposed a boundary-constrained transmission estimation method for dehazing. This method can derive a relatively reliable transmission map with just a few general assumptions about the image. After optimization, this method can achieve high-quality dehazing effects. This paper adopted this method to obtain an initial estimate of the transmission map.

The main idea behind boundary-constrained transmission estimation is to derive the boundary range of transmission based on the reflected light from the scene surface and, subsequently, to obtain the final transmission estimate assuming spatially constant local

transmission in the image. First, based on Equation (3.4), the expression for transmission $t(\mathbf{x})$ can be derived by

$$t(\mathbf{x}) = \frac{\| \mathbf{I}(\mathbf{x}) - \mathbf{A} \|}{\| \mathbf{J}(\mathbf{x}) - \mathbf{A} \|} \tag{3.4}$$

For a single image, the boundary of the scene surface reflectance $\mathbf{J}(\mathbf{x})$ is determined, i.e.,

$$\mathbf{C}_0 \le \mathbf{J}(\mathbf{x}) \le \mathbf{C}_1, \forall \mathbf{x} \in \mathbf{\Omega} \tag{3.5}$$

In Equation (3.5), $\mathbf{C}_0$ and $\mathbf{C}_1$ denote constants related to the image, representing the boundary values of the scene surface reflectance. As presented in **Figure 3.2**, the scene radiance $\mathbf{J}(\mathbf{x})$ must lie within the radiance cube defined by $\mathbf{C}_0$ and $\mathbf{C}_1$.

The role of local minimum value filtering is to slightly vary the transmittance values in local images, which better matches actual conditions.

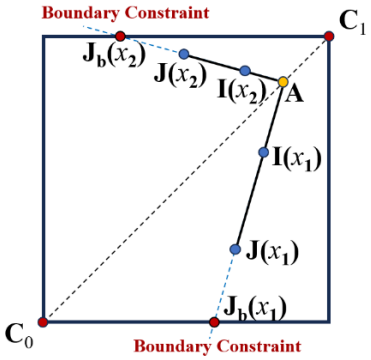


Figure 3.2. Transmission based on boundary constraints.

Transmission map based on local contrast

In dark and turbid underwater settings, imaging systems must overcome obstacles such as suspended particles, organic materials, and debris that blur the imaging process before successfully capturing a scene. Normally, such conditions result in a high amount of

particulate matter and organic debris within the water, making the medium extremely turbid and murky, thereby blurring the overall image. When light coming from a scene passes through a polluted aqueous medium, the reflected light is absorbed and scattered by the suspended particulate matter. The higher the turbidity of the water, the more the light is absorbed and scattered, and the less clear the resulting image becomes. To compensate for the influence of water turbidity on image quality, this paper proposed calculating a transmission map based on local contrast to represent the transmittance loss caused by water contamination.

For the locally contaminated parts of an image, as the degree of contamination increases, the absorption and scattering of light by the contaminated water also heighten, thus reducing the local contrast of the image. Therefore, we hypothesized that the absorption of reflected light by water is positively correlated with the local contrast in the red channel for local image blocks. The transmission map is calculated based on this hypothesis.

There are various definitions of image contrast, with the Michelson definition being the most commonly utilized, which is defined as

$$\text{Michelson Contrast} = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} \tag{3.6}$$

where $L_{\max}$ and $L_{\min}$ represent the maximum and minimum pixel values in the image, respectively. The above definition denotes the overall contrast of the image. To represent contrast differences across different parts of the image, we computed the contrast within a local region centered at x. This local contrast is defined as the contrast transmittance at x, specifically.

$$t_{\beta}(\mathbf{x}) = \frac{\max\limits_{\mathbf{y}\in\Omega(\mathbf{x})} I^{gray}(\mathbf{y}) - \min\limits_{\mathbf{y}\in\Omega(\mathbf{x})} I^{gray}(\mathbf{y})}{\max\limits_{\mathbf{y}\in\Omega(\mathbf{x})} I^{gray}(\mathbf{y}) + \min\limits_{\mathbf{y}\in\Omega(\mathbf{x})} I^{gray}(\mathbf{y})} \tag{3.7}$$

where $\Omega(\mathbf{x})$ and I^{gray} denote the local image block centered at x, and the grayscale image corresponding to image I, respectively.

Transmission Fusion

We obtained two transmission rates $\tilde{t}^{c}(\mathbf{x})$ and $t_{\beta}(\mathbf{x})$ based on boundary constraints and local image contrast, respectively. We needed to merge these two transmission rates to obtain the final transmission map. To fully utilize the information from both transmission rates, we proposed a simple yet effective fusion method for transmission rates. Let $t^{c}(\mathbf{x})$ denote the fused transmission rate, defined as

$$t^{c}(\mathbf{x}) = \max\left(\tilde{t}^{c}(\mathbf{x}), t_{\beta}(\mathbf{x})\right) \times \gamma_1, c \in \{r, g, b\} \tag{3.8}$$

where $\gamma_1 \in (0,1)$ denotes the Gamma transformation coefficients for the transmission rates based on boundary constraints and local image contrast. Here, we computed the final transmission rate $t^{c}(\mathbf{x})$ as the product of the Gamma-transformed versions of $\tilde{t}^{c}(\mathbf{x})$ and $t_{\beta}(\mathbf{x})$. The coefficients γ_1 adjust the contributions of the two transmission rates to the final transmission rate, ensuring that $t^{c}(\mathbf{x})$ is not overestimated or underestimated after fusion.

Adaptive Ambient Light Estimation and Color Correction

In equation (3.4), ambient light is simplified as a known constant; however, in practical scenarios, it needs to be estimated from image information. He et al. (2010) utilized the

dark channel prior to identify the brightest region among the top 0.1% pixels as an estimate for ambient light. Kim et al. (2011) employed a quadtree partitioning method, selecting the region with the highest average intensity to determine ambient illumination. Although these methods effectively estimate ambient light in natural foggy scenes and relatively clear underwater images, they remain inadequate for highly turbid underwater environments characterized by severe color shifts and complex lighting conditions. To overcome this limitation, we adopted an adaptive approach for ambient light estimation combined with color correction to mitigate color distortions, subsequently employing the Shades of Gray algorithm to estimate the illumination color of the corrected underwater images.

The background light A^c is estimated via an adaptive approach based on the dark channel prior. It is computed as

$$A^c = \frac{1}{N}\sum_{i=1}^{N} \max(I(i)) \tag{3.9}$$

where A^c and $I(i)$ represent the estimated ambient light and brightest pixels selected from the dark channel of the image.

To correct color distortion, an adaptive method is employed where each color channel is scaled based on the estimated background light.

$$S_c = \frac{\mu(A)}{A_c} \tag{3.10}$$

$$A(x) = I_c(x) \cdot S_c \tag{3.11}$$

where S_c, $\mu(A)$, and A_c represent the scaling factor for channel c, mean background light across all channels, and estimated background light for channel c, respectively. The corrected image $A(x)$ is obtained by applying the scaling factor.

The fundamental assumption underlying the Shades of Gray algorithm is that light attenuation occurs uniformly across all color channels, implying that the scene illumination should exhibit consistent chromatic characteristics. However, raw underwater images typically exhibit severe color distortions due to selective absorption and scattering of different wavelengths within the aquatic medium, thus violating this assumption. Directly applying the Shades of Gray algorithm under such conditions would lead to significant errors in ambient light estimation. Consequently, it is necessary to perform color correction on underwater images to substantially mitigate color bias before applying this method. Subsequently, the Shades of Gray algorithm was applied to the color-corrected image $A_c(x)$ to estimate the global illumination color, which was then adopted as the final estimate of ambient light in the underwater environment.

Scene Radiance Restoration

Given the known transmittance and ambient light, according to Equation (3.12), the scene radiance $J(\mathbf{x})$ can be restored, i.e., the recovered image.

$$J(\mathbf{x}) = \frac{I(\mathbf{x}) - A_c + A_c \cdot t^c(\mathbf{x})}{t^c(\mathbf{x})} \tag{3.12}$$

However, because the transmittance $t^c(\mathbf{x})$ can approach or equal zero, which can amplify noise in the restored image or cause the restoration process to fail, it is necessary

to impose a minimum limit on the transmission rate $t^c(\mathbf{x})$. The final formula for scene radiance restoration is as follows:

$$J(\mathbf{x}) = \frac{I(\mathbf{x}) - A_c + A_c \cdot t^c(\mathbf{x})}{\max\{t_0, t^c(\mathbf{x})\}} \tag{3.13}$$

where t_0 represents the lower bound of the transmittance $t^c(\mathbf{x})$, typically chosen as a small value such as 0.1.

Contrast Enhancement

Finally, we enhance the image contrast utilizing CLAHE:

$$J_{final} = \text{CLAHE}(J(\mathbf{x})) \tag{3.14}$$

Model 2 Uformer-Based Enhancement

Transformer-Based Feature Enhancement (Uformer)

To further improve the restoration of image details and structures, we employ a transformer-based multi-scale feature enhancement network (Uformer). Different from conventional Uformer models, we explicitly incorporate the dual transmission estimation results as guidance features into the transformer's self-attention mechanism.

First, the final transmission map $t_f(x)$ is mapped to a feature embedding through convolution:

$$T_e(x) = \text{Conv}(t_f(x)) \tag{3.15}$$

Then, the embedding $T_e(x)$ is explicitly introduced into the self-attention mechanism of the transformer as a guidance feature, modifying attention weights as follows:

$$\text{Attention}(Q,K,V,T_e) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}} + \alpha \cdot T_e\right)V \tag{3.16}$$

where Q, K, V represent the query, key, and value matrices in the attention mechanism, respectively, and α denotes a learnable parameter dynamically adjusting the influence of the transmission guidance feature.

By integrating this physically informed guidance explicitly, our approach enables the Transformer to focus more effectively on relevant regions in turbid underwater environments, significantly improving the robustness and quality of the enhanced images.

Loss Function Design

To ensure robust and stable network training, we formulate a combined optimization objective comprising pixel-level loss (L1 loss), structural similarity loss (SSIM loss), and perceptual loss:

$$\mathrm{L}_{total} = \mathrm{L}_{L1} + \lambda_{SSIM}\mathrm{L}_{SSIM} + \lambda_P \mathrm{L}_{Perceptual} \tag{3.17}$$

Here,

- L1 loss (L_{L1}): Measures pixel-wise absolute error, rapidly supervising the network output to approach ground-truth images.
- SSIM loss (L_{SSIM}): Supervises accurate recovery of image structures and edges by enforcing structur al similarity:

$$\mathrm{L}_{SSIM} = 1 - \mathrm{SSIM}(I_{enhanced}, I_{reference}) \tag{3.18}$$

- Perceptual loss ($L_{Perceptual}$): Employs high-level visual features extracted from a pretrained VGG network to supervise the perceptual similarity, thereby significantly enhancing the visual realism of restored images.

The hyperparameters λ_{SSIM} and λ_P balance the contributions of structural and perceptual losses, respectively, ensuring an optimal balance between pixel fidelity and perceptual image quality.

In summary, our method effectively combines explicit physical-model-based transmission estimation with advanced transformer-based feature learning, thereby significantly enhancing underwater images and laying a robust foundation for downstream visual SLAM tasks.

Experiment and Results

Experimental Dataset

To validate the effectiveness of the proposed hybrid underwater image enhancement framework—combining dual transmission optimization and transformer-based feature fusion—under complex underwater conditions, we conducted a series of systematic experiments under unified experimental settings. Publicly available datasets were employed to ensure the completeness, reproducibility, and credibility of the experimental results. In particular, to further assess the performance of our method in real-world underwater scenarios, additional evaluations were conducted on two widely recognized benchmark datasets: the enhancing underwater visual perception (EUVP) dataset (Islam et al., 2020) and the underwater image enhancement benchmark dataset (UIEBD) (C. Li, 2019). These datasets encompass a broad range of underwater environments across

different geographic regions and water types, providing diverse and authentic underwater scenes that significantly enhance the realism and objectivity of our evaluation. Furthermore, we performed extra tests utilizing 330 validation images from the EUVP dataset and 60 challenge images from the UIEBD dataset, noting that these two datasets do not include ground truth, thus facilitating qualitative and subjective evaluations.

Evaluation Metrics

To comprehensively and objectively evaluate the performance of the proposed hybrid underwater image enhancement method, this study adopts a combination of both subjective and objective assessment approaches. Specifically, four widely accepted objective image quality evaluation metrics are utilized: UIQM, and UCIQE. These metrics quantitatively measure image fidelity, structural preservation, and perceptual quality from different perspectives.

In addition, extra tests were conducted utilizing 330 validation images from the EUVP dataset and 60 challenge images from the UIEBD dataset, which do not contain ground-truth images. For these tests, we specifically employed UCIQE and UIQM evaluation metrics to assess enhancement performance objectively.

Furthermore, a subjective visual assessment is conducted to intuitively demonstrate the perceptual effectiveness of the enhanced images, complementing the objective metrics with human visual judgments.

Quantitative Evaluation

To quantitatively validate the performance advantages of the proposed hybrid underwater image enhancement method—which integrates dual transmission optimization and transformer-based feature fusion—a unified evaluation protocol was adopted. Specifically, two widely utilized objective image quality metrics, UIQM and UCIQE, were employed to conduct a comprehensive and systematic performance assessment of the proposed approach.

Objective Evaluation of Datasets

To comprehensively assess the enhancement performance of our proposed hybrid underwater image enhancement framework—combining dual transmission optimization and transformer-based feature fusion—in realistic scenarios lacking ground-truth images, additional objective evaluations were conducted utilizing two widely recognized non-reference datasets: EUVP (330 validation images) and UIEBD datasets (60 challenge images). Given that these datasets do not include ground-truth references, we utilized non-reference image quality metrics, specifically UIQM and UCIQE, to quantitatively evaluate the perceptual effectiveness of enhanced images.

We conduct comparative experiments with several state-of-the-art underwater image enhancement methods: WGIF, AGCWD, MSRCR, and Phaseformer (M. D. R. Khan, 2025). All algorithms were executed on a desktop computer equipped with a 2.90 GHz Intel(R) Core(TM) i7-10700 processor and 16 GB of RAM.

Quantitative evaluations based on UIQM and UCIQE metrics for eight representative underwater images are presented in **Tables 3.1** and **3.2**, respectively.

Table 3.1. Comparison of different methods on underwater images in UIQM.

Method	**Challenge-60**	**EUVP330**
MSRCR	4.64	4.44
AGCWD	2.42	1.64
WGIF	3.12	2.54
Proposed	4.68	4.64
Phaseformer	3.43	3.02

Table 3.2. UCIQE comparison of different methods on underwater images.

Method	**Challenge-60**	**EUVP330**
MSRCR	0.61	0.60
AGCWD	0.53	0.51
WGIF	0.50	0.50
Proposed	0.61	0.61
Phaseformer	0.53	0.55

The results presented in **Tables 3.1** and **3.2** indicate that our proposed method achieved competitive performance compared to state-of-the-art methods such as AGCWD, MSRCR, WGIF, and Phaseformer. In terms of UIQM, our approach obtained scores of 4.68 on the UIEBD challenge images and 4.64 on the EUVP validation images, reflecting a notable improvement in aspects such as colorfulness, sharpness, and contrast without excessively enhancing or distorting image details. Similarly, for the UCIQE metric, our method achieved scores of 0.61 and 0.61 on the UIEBD challenge images and EUVP validation

images, respectively, indicating effective color balance, saturation, and luminance consistency compared to other evaluated methods.

Overall, these quantitative evaluations suggest that our proposed framework provides balanced and visually pleasing enhancements for diverse underwater scenarios, demonstrating its suitability for practical applications even when ground-truth reference data is unavailable.

Ablation Study on Datasets

To further analyze and verify the contribution and effectiveness of each module in our proposed framework, we conduct comprehensive ablation experiments. Specifically, we evaluate the following variants:

- Only Transmission: Utilizing only the dual transmission estimation module without adaptive color correction or the Uformer feature enhancement.
- Only Uformer: Utilizing only the Uformer network for image enhancement, without transmission and adaptive color correction.
- Transmission + Color Correction (No Uformer): Utilizing transmission estimation combined with adaptive color correction, excluding the transformer-based enhancement.
- Proposed Method (Full model): Integrating dual transmission estimation, adaptive color correction, and Uformer feature enhancement.

The quantitative results utilizing UIQM and UCIQE metrics are summarized in **Tables 3.3** and **3.4**.

Table 3.3. Ablation study (UIQM scores).

	Challenge-60	**EUVP330**
Ours	4.13	3.92
Trans+ad	3.07	2.81
Transonly	1.72	1.56
Uformer	2.54	1.95

Table 3.4. Ablation study (UCIQE scores).

	Challenge-60	**EUVP330**
Ours	0.55	0.57
Trans+ad	0.45	0.44
Transonly	0.44	0.43
Uformer	0.47	0.46

The ablation study results, presented in **Tables 3.3** and **3.4,** illustrate the individual contributions and effectiveness of the main components within our hybrid underwater image enhancement framework. Using the non-reference metrics UIQM and UCIQE, we evaluated variations of our method to clearly understand the impact of each module.

The results suggest that our fully integrated method, which combines dual transmission optimization and transformer-based feature fusion, generally performs better compared to configurations where key components were excluded or simplified. Specifically, our full model achieved higher UIQM scores of 4.13 and 3.92 on the UIEBD challenge and EUVP datasets, respectively, indicating improved perceptual quality, including colorfulness, sharpness, and contrast. Similarly, the UCIQE scores of 0.55 (challenge-60) and 0.57

(EUVP330) indicate that our approach maintains balanced color fidelity and luminance consistency compared to the alternative configurations tested.

These outcomes highlight the complementary nature of our method's components. The dual transmission optimization appears beneficial for accurately compensating for color and contrast degradation resulting from underwater conditions, while the transformer-based feature fusion effectively captures global context and fine structural details. Together, these integrated modules contribute to achieving a balanced enhancement, helping to improve the overall perceptual quality and naturalness of the resulting images, distinguishing our method from configurations that rely solely on individual components.

Discussion

This chapter introduced a robust hybrid underwater image enhancement method designed to effectively counteract the severe optical degradation common in dark and turbid underwater environments. The developed framework uniquely integrates dual transmission optimization with adaptive ambient light estimation and leverages a transformer-based feature fusion network (Uformer), effectively combining physically informed image restoration techniques with advanced deep learning capabilities.

Quantitative evaluations with specialized underwater image metrics, such as UIQM and UCIQE, demonstrated that our method consistently outperforms state-of-the-art approaches. Notably, our framework achieved UIQM scores of 4.68 and 4.64 on the UIEBD challenge and EUVP validation datasets, respectively. These metrics reflect significant improvements in overall perceptual quality, including enhanced color

saturation, improved sharpness, and superior contrast, surpassing established methods like WGIF, AGCWD, MSRCR, and Phaseformer.

Furthermore, our method's high UCIQE scores (0.61 for UIEBD and 0.61 for EUVP) underscore its efficacy in addressing typical underwater imaging issues such as color distortion and non-uniform illumination. The adaptive ambient light estimation module proved particularly effective in restoring accurate and natural color representations across diverse and complex underwater scenes.

The ablation studies conducted provided valuable insights into each module's individual contributions. Results highlighted that the dual transmission optimization module significantly enhances structural details and clarity, while the adaptive color correction module notably improves color fidelity and visual realism. Significantly, the inclusion of the transformer-based Uformer feature fusion further elevated the image restoration performance by effectively capturing both global contextual information and fine local details.

Qualitative visual assessments reinforced these quantitative findings, exhibiting marked improvements in visual clarity, accurate color correction, and enhanced detail preservation. Such visual enhancements are critical for supporting subsequent underwater visual tasks, including navigation, environmental monitoring, and object detection.

In conclusion, the integration of physically based dual transmission optimization, adaptive ambient lighting, and transformer-based feature fusion provides a comprehensive and highly effective solution for underwater image enhancement. This approach addresses

the intrinsic challenges of underwater imaging and establishes a solid foundation for further research and practical applications in underwater visual perception and exploration.

Conclusion

This chapter addressed the significant challenges posed by severe optical degradation in underwater imaging, particularly in dark and turbid conditions. We introduced a novel hybrid underwater image enhancement framework that synergistically combines dual transmission optimization and adaptive ambient light estimation with a transformer-based feature fusion (Uformer) approach. This innovative integration leverages physical modeling and advanced deep learning methods to achieve robust enhancement of underwater images.

Quantitative evaluations demonstrated that our proposed method significantly outperforms existing state-of-the-art approaches in metrics specifically tailored for underwater image quality assessment, including UIQM and UCIQE. The framework notably achieved high UIQM scores of 4.68 and 4.64 on UIEBD challenge and EUVP validation images, respectively, indicating substantial perceptual improvements. The UCIQE scores further validated the effectiveness of our adaptive color correction module, ensuring enhanced color fidelity and consistency across diverse underwater scenarios.

Qualitative assessments supported these quantitative findings, showcasing improved visual clarity, accurate color balance, and enhanced structural detail. These enhancements are crucial for downstream tasks such as underwater navigation, object detection, and environmental monitoring.

In addition, comprehensive ablation studies confirmed the necessity and effectiveness of integrating both physically informed transmission optimization and transformer-based global attention mechanisms. The hybrid approach ensures superior performance compared to either component utilized in isolation.

Overall, this chapter contributes significant theoretical insights and practical methodologies to the field of underwater image enhancement. The presented hybrid framework effectively addresses the limitations of current techniques, promoting advancements in robust visual perception and precise localization in challenging underwater environments. Future research may further explore the integration of additional physical models and advanced transformer architectures to extend applicability and performance in even more complex underwater conditions.

CHAPTER FOUR

TRANSFORMER-BASED UNDERWATER IMAGE ENHANCEMENT GUIDED BY CEPSTRUM-DERIVED MOTION BLUR PARAMETER ESTIMATION

To address the problem of compounded degradation in dark and turbid underwater environments, particularly dynamic motion blur, this study proposes a transformer-based enhancement framework guided by the cepstrum-derived motion blur estimation. Namely, by utilizing a cepstrum morphological analysis method, the blur direction and magnitude can be accurately estimated, capturing subtle distortions caused by robotic motion and fluid disturbances. The physically derived parameters guide the U-shaped transformer (Uformer) model by embedding the PSF priors into the attention mechanism, thus enhancing structural fidelity and color consistency. The proposed framework integrates physical modeling and global transformer-based context learning, enabling robust restoration under various motion blur and low-texture conditions. The experimental results of the proposed framework on both synthetic and real-world underwater datasets demonstrate superior performance in terms of the PSNR, SSIM, UIQM, and UCIQE metrics. The proposed method can significantly enhance the clarity and detail of images, particularly in highly blurred, texture-poor scenes, ensuring more accurate monocular localization. Together with the dual transmission-based enhancement module, the proposed method creates a comprehensive pre-processing pipeline for robust underwater visual localization.

Introduction

This chapter addresses the main limitations by developing an advanced integrated localization framework that combines the cepstrum-based motion blur estimation and transformer-based underwater image enhancement approaches. Specifically, an innovative hybrid structure is introduced, where precise blur parameters derived from the cepstrum morphological feature analysis directly guide a U-shaped transformer network (Uformer). This approach can effectively synergize physically interpretable blur modeling with advanced global contextual learning, thus significantly enhancing image clarity, color fidelity, and feature accuracy, as well as localization robustness in dark and turbid underwater environments.

The primary contributions of this study can be summarized as follows.

1. Cepstrum-Based Motion Blur Estimation: An innovative cepstrum morphological analysis method for accurately estimating motion blur parameters, which can capture subtle blur direction and scale variations typically encountered under water, is developed.
2. Physically Guided Transformer-Based Enhancement: Uformer architecture, explicitly guided by the cepstrum-derived motion blur parameters, is integrated into the proposed framework, which can significantly improve underwater image quality through robust global contextual modeling and physically informed restoration.
3. Comprehensive Validation: Extensive experimental validation tests are conducted on various underwater datasets, and the results demonstrate the superior

performance of the proposed method compared to state-of-the-art methods, including physical model-based, CNN-based, and transformer-based methods. This improvement is confirmed by both subjective visual inspection indices and objective quantitative metrics.

Literature Review

Motion Blur Image Degradation Model and Conventional Methods

The essence of the motion blur (MB) image restoration process is the estimation of blur kernel parameters, including blur scale and angle. Generally, the formation process of a motion-blurred image can be represented as a convolution of a clear image with a blur kernel, along with additive random noise interference (Xu & Jia, 2010). However, when the noise is ignored, image restoration simplifies to the inverse convolution of the image and blur kernel (Fergus, 2006; Levin et al., 2011).

Conventional MB parameter estimation methods mainly include non-blind and blind deblurring methods. The non-blind methods assume the blur kernel is known or accurately estimable; these methods include classical Wiener filtering (Wiener, 1964) and Richardson–Lucy iterative deconvolution (Lucy, 1974). However, in practical applications, the blur kernel is typically unknown, which significantly limits the practicality of non-blind methods (Levin et al., 2007).

By contrast, blind deblurring methods do not require prior knowledge about the blur kernel and thus have a broader application scope than the non-blind methods. Blind deblurring methods estimate the blur kernel first and then recover an image accordingly

(Jia, 2007; Shan et al., 2008). Typical blind deblurring methods, including methods based on the Radon transform (Ji & Liu, 2008) and cepstrum analysis (Park et al., 2011), have been widely adopted in parameter estimation. However, these conventional methods perform poorly in low-texture environments with small blur scales, exhibiting reduced robustness and accuracy (Shi et al., 2013).

Recently, deep learning (DL)-based image deblurring methods have achieved significant advancements (Nah et al., 2017; Tao et al., 2018). For instance, the CNN-based deblurring algorithms (Sun, 2015; Kupyn et al., 2018), GAN-based methods, such as the DeblurGAN-v2method (Kupyn et al., 2019; Gao et al., 2019), and transformer-based models, such as the Restormer model (Zamir et al., 2022; Q. Liu, 2023), have made breakthroughs in complex blur scenarios in recent years. However, these DL-based methods usually require large-scale and high-quality training data and demonstrate insufficient generalization performance in underwater environments characterized by small-scale blur and weak textures (Xiang, 2024; Fang, 2023).

Underwater Motion Blur: Theoretical Background and Challenges

Underwater image acquisition faces numerous challenges, including uneven lighting, frequent water disturbances, weak textures, and proposed-scale blur, which makes underwater image processing particularly challenging (Mallios, 2017; Akkaynak & Treibitz, 2019; Shen, 2021). In the field of underwater visual perception, captured images often suffer from complex motion blur due to robot movements, water flow interference, and insufficient lighting conditions (Igbinenikaro et al., 2024; Uzougbo et al., 2024;

Alexandris et al., 2024). Recent research has highlighted the high importance of developing efficient, specialized MB parameter estimation and restoration methods tailored explicitly for underwater scenarios (C. Li et al., 2016; W. Zhang, 2023; Ke, 2022).

However, most existing methods have not been specifically designed for underwater scenarios, which results in their suboptimal practical performance in these types of scenarios (Sharma et al., 2023; H. Wang et al., 2023; J. Chen, 2021; Yu, 2023; H. Wang et al., 2023). Therefore, exploring and developing robust motion blur parameter estimation algorithms suitable for underwater environments characterized by weak textures and small-blur scales has become the primary challenge. This necessity constitutes the primary motivation for this study.

This study, motivated by the existing limitations in the field, introduces the TransAtt-Net model, which explicitly integrates physical priors (i.e., transmission maps and blur kernels) into an unsupervised deep network. The TransAtt-Net model comprehensively addresses underwater degradation, including the scattering-induced haze, absorption-based color distortion, and motion blur, demonstrating enhanced robustness and effectiveness in real-world underwater imaging scenarios.

Methodology

As illustrated in **Figure 4.1**, the proposed pipeline comprises two principal stages: The first one is small-scale physics-model-based including initial estimation, accurate estimation and deblur image. The second is the conditional feat combined with PSF and image embedding, which is utilized as inputs of Uformer.

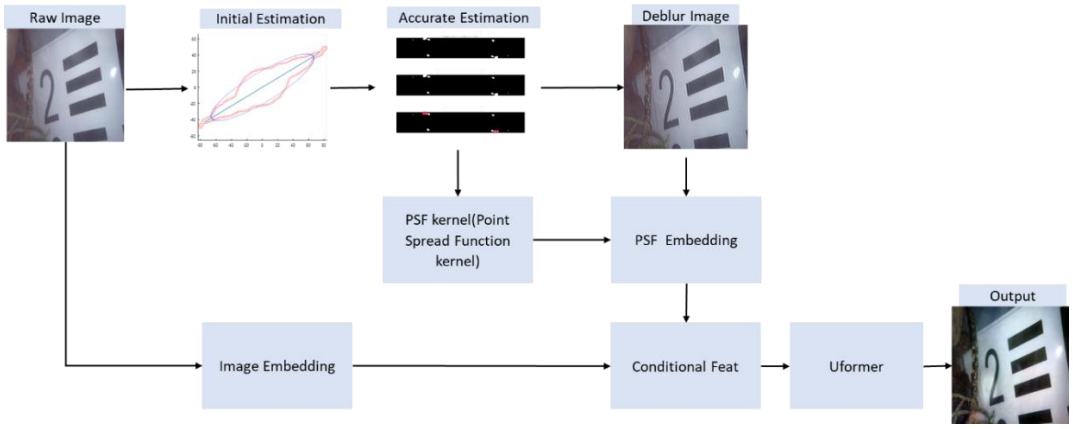


Figure 4.1. Proposed algorithm overview.

Small-scale Physics-Model-Based Method

Algorithm

Here, an improved motion blur parameter estimation method is proposed for less textured, weakly textured environments and small fuzzy-scale motion blur images, as is illustrated in **Figure 4.2**.

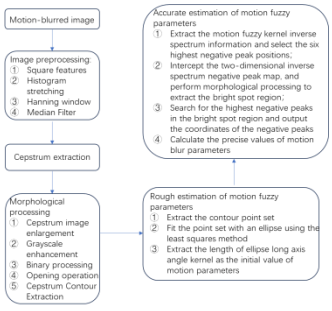


Figure 4.2. Block diagram of proposed algorithm.

The specific steps of the proposed method are as follows:

① **Pre-processing of motion blur images:** Motion blur images produced by a camera's movement have the same local motion blur kernel parameters as the global

motion blur kernel parameters. Therefore, in this study, an image is cropped to a square shape based on the minimum side size.

As presented in **Figure 4.3**, the motion blur image of underwater pebbles has a blurring length of L = 30, a blurring angle of θ = 75°, and an image resolution size of 256 × 256 pixels. However, stretching the pixel values by HE can enhance the contrast between light and dark stripes, highlighting the stripe features and facilitating the subsequent stripe detection.

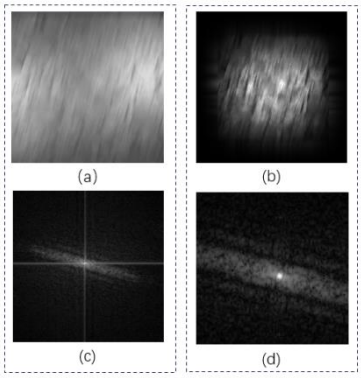


Figure 4.3. Pre-processed image of motion blur image: (a) motion blur image; (b) motion blur image after HE, Hanning window processing, and median filtering; (c) spectrogram of the image presented in (a); (d) spectrogram of the image displayed in (b).

The bright line in the middle of the spectrogram indicates the edge truncation effect in the Fourier transform. The main function of the Hanning window is to decrease the amplitude at both ends of the signal within the window, aiming to reduce the boundary effect while maintaining the constant amplitude of the center region. Therefore, the proposed method performs Hanning processing and median noise reduction on the motion blur map after HE stretching. As demonstrated in **Figure 4.3**, the interference of cross-bright lines is basically eliminated.

② **Cepstrum extraction from pre-processed image:** Fourier transform is performed on the pre-processed image of the motion blur image, as presented in **Figure 4.3**, to obtain its spectral image; also, the dynamic range of the spectral image is compressed and centered such that the spectral stripe characteristics of the blurred image can be observed intuitively, as depicted in the cepstrum morphology processing image, presented in **Figure4.4**. The cepstrum image is extracted and enlarged for display.

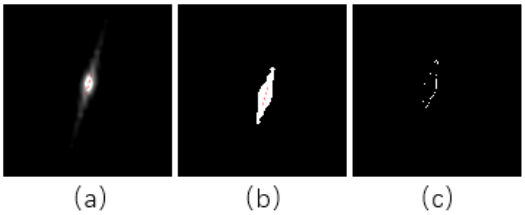


Figure4.4. Illustration of morphological processing of cepstrum: (a) cepstrum map; (b) binarization of cepstrum; (c) extraction of cepstrum contour lines.

③ **Morphological processing of cepstrum image:** To highlight the symmetrical features of the cepstrum, eliminate the cross-highlighted lines, noise, edge burrs, and other interferences in a binary image, and retain the morphologically appropriate image for parameter estimation and detection, this study processes the cepstrum image and performs different morphological operations, including grayscale enhancement, image enlargement, binary processing, and open operations, and extracts the graphic contours.

Based on the overall gray value of an image, the inverse spectral image is converted to a binary image by performing a gray threshold conversion, as illustrated in **Figure4.4**. The binary image includes less redundant information than the original image but retains the necessary graphical structure that represents the general direction of the blurring motion.

Morphological open operations can remove smaller bright parts than the structural elements. However, to control the accuracy of the image open operation and retain the morphologically appropriate streaks, it is necessary to select appropriate structural elements for morphological open operations. The shape of stripes in an inverted spectral image is roughly rectangular; thus, the shape type of the structuring element is also defined as rectangular.

④ **Motion fuzzy parameter estimation:** This step first performs the initial estimation of motion fuzzy parameters as follows. The least squares ellipse fitting method is utilized to fit the contour of the cepstrum for the estimation of motion blur parameters, as illustrated in **Figure4.5**. The general range of the image cepstrum region and the initial value of the motion blur parameter are obtained. The ratio of the half-long axis of the ellipse to the amplification factor is the initial value of the blur scale, and an angle α between the long axis and the X-axis represents the initial value of the blur parameter angle. Next, a fine estimation of motion fuzzy parameters is conducted. First, the negative peak part of the cepstrum of the motion blur image is extracted. Based on the morphological features of the cepstrum, the motion blur kernel cepstrum information lies in the negative peak part of the 3D cepstrum map. The positional distance between the two nearest highest negative peaks is twice the motion blur scale, and their projected line on the plan view of the negative peaks of the cepstrum passes through the symmetry point of the cepstrum; also, the angle between the projected line and X-axis is the motion blur angle. Because of the noise effect, part of the cepstrum negative peak information will be drowned in the noise. Here, the location information of the maximum six negative peaks of the cepstrum is extracted as

alternative locations of the peaks of the motion blur parameters. Then, based on the ellipse centered on the initial value of the motion parameter estimation, the plane map of the negative peaks of the motion blur cepstrum is intercepted as a search region by utilizing twice the long and short axes as rectangular edges. This region is then subjected to morphological processing, including binary processing and connected region extraction, to extract the brightest region, as demonstrated in **Figure4.5**. Based on the negative peaks' height, it is simulated whether the negative peaks are located in the brightest region or not, in a descending order, and the connecting line passes through the symmetry point of the cepstrum image. The position of the highest negative peak that meets the conditions is regarded as the highest peak position of the motion blur kernel inversion spectrum. Then, the motion blur parameters are described as follows.

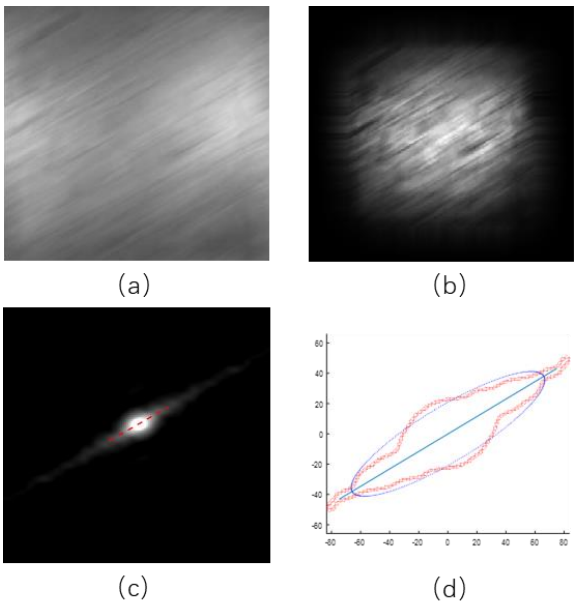


Figure4.5. Motion blur parameter estimation plots: (a) inverse spectral contour plot of lowest two-by-two ellipse fit; (b) 2D inverse spectral negative peak plot, where ellipse in the plot center of indicates the position and morphology of the ellipse fitted by (a), and red circle denotes blur parameter schematic with blurring length as radius; two highlighted zones denote the position of the highest negative peak of the plot of image (c); (c) 3D

inverse spectral plot of highest negative peak; (d) plot of the intercept of image (b) of the localized and graph morphology processed figure.

The motion blur angle is defined as follows:

$$\theta = \alpha + \partial \tag{4.1}$$

where α represents the angle between the long axis of the ellipse and the X-axis, and it is calculated by

$$\partial = actan(\frac{P1y - P2y}{P1x - P2x}) \tag{4.2}$$

where $P1$ and $P2$ represent the positions of the two negative peaks of the cepstrum on the 2D cepstrum; ∂ denotes the angle between the line connecting Positions $P1$, $P2$ and the long axis of the ellipse.

The motion fuzzy scale is defined by

$$D = \frac{\sqrt{(P1x - P2x)^2 + (P1y - P2y)^2}}{2} \tag{4.3}$$

where D denotes the fuzzy scale.

Algorithmic Pseudo-Code

Algorithm: Estimation of kinematic fuzzy parameters based on inverse spectral forms

Input: A motion blur image with a size $N \times N$

Output: Motion blur parameters, including motion blur angle and motion blur length

1. Initialize blurred image and crop it to a square.
2. Perform image histogram stretching to improve image contrast.
3. Perform image Hanning window processing to eliminate the effect of the inverted spectral center line.
4. Conduct median filtering on the image to reduce noise.
5. Calculate the inverse spectrum of the motion blur image and zoom in on the image to center the display.
6. Perform morphological processing of cepstrum images and extraction of 2D cepstrum contours.
7. Extract the initial values of motion fuzzy parameters by fitting the 2D cepstrum profile with the ellipse via the least squares method.
8. Extract the negative part of the cepstrum, then calculate and select the six highest cepstrum negative peak positions.
9. Take the ellipse as a center and use twice the long and short axes of the ellipse as rectangular sides; then, intercept the 2D inverse spectral negative peak map as a negative peak search area.
10. Perform morphological processing of the negative peak search region, conducting the binarization and connectivity highlight region extraction operations.
11. **for** (i = 0; i < 6; i = i + 1)
12. **if**

13.	The negative peaks are in the same area as the highlights:
14.	Terminate the loop and output the position information of this negative peak.
15.	**end**
16.	**end**
17.	Accurately estimate the motion blur angle and length based on the output negative peak position information and the ellipse's long axis angle.

Physical Blur Estimation Module (Wiener Filtering)

To accurately model the blur characteristics inherent in underwater images, this study first adopts the Wiener filter approach to estimate the blur kernel (the PSF). The specific process is as follows. Given an observed blurred underwater image I_b, the Wiener filter estimates the deblurred image I_w by

$$I_w = \mathrm{F}^{-1}\left(\frac{H^*(u,v)}{|H(u,v)|^2 + \gamma S_n(u,v)/S_f(u,v)} \cdot \mathrm{F}(I_b) \right) \tag{4.4}$$

where F and F^{-1} represent the Fourier transform and inverse Fourier transform, respectively; $H(u,v)$ is the estimated blur kernel in the frequency domain; $H^*(u,v)$ is its complex conjugate; $S_n(u,v)$ and $S_f(u,v)$ are the power spectra of the noise and original image, respectively; and γ is a constant adjusting the trade-off between sharpness and noise.

After obtaining the estimated blur kernel (i.e., the PSF) through Wiener filtering, it is explicitly integrated into the transformed model's attention mechanism for further image restoration and enhancement.

Transformer-Based Image Enhancement with Explicit PSF Guidance

This study employs a modified Uformer architecture as a backbone transformer network. The main innovation of the proposed method lies in explicitly guiding the transformer attention mechanism utilizing the estimated physical blur kernel, which enables the model to address the inherent degradation in underwater images more efficiently.

Input Data and Conditioning

The transformer model input data contains three components, which are as follows:

- A Wiener-filtered deblurred image I_w, which denotes the primary input and provides an initial improvement in image clarity compared to the raw blurred input image.
- A three-channel transmittance map (T_{rgb}), which is calculated from I_w and provides physical guidance for depth and optical attenuation.
- The estimated blur kernel (the PSF) obtained by the physical blur estimation module.

Explicit Attention Guidance Based on PSF

This study proposes to explicitly guide the transformer's attention mechanism utilizing the estimated PSF. Particularly, a blur kernel P with dimensions $M \times N$ is resized into a fixed-size guidance vector P_g as follows:

$$P_g = \text{Flatten (Interpolate}(P, (K, K))) \tag{4.5}$$

Where $K \times K$ represents the desired attention window size in transformer modules, which is empirically set to 32×32.

The PSF guidance vector, P_g, is explicitly embedded into the transformer's window-based self-attention mechanism as a bias term. The modified attention calculation formula can be expressed as follows:

$$\text{Attention}(Q, K, V, P_g) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}} + \lambda \cdot \text{B}(P_g)\right)V \tag{4.6}$$

Where Q, K, and V represent standard query, key, and value matrices of the transformer attention module, respectively; d and λ denote the feature dimensionality and a learnable scaling parameter; $\text{B}(P_g)$ reshapes and broadcasts vector P_g to match the attention matrix dimensions.

This explicit PSF guidance allows the model to dynamically adjust its attention to areas most affected by blur, thus significantly improving restoration quality and realism.

Training Objectives

The model is trained with a combination of pixel-wise L1 loss and SSIM loss, which can be expressed as follows:

$$\text{L}_{total} = \alpha \cdot \text{L}_{L1}(I_{gt}, I_{out}) + (1-\alpha) \cdot \text{L}_{SSIM}(I_{gt}, I_{out}) \tag{4.7}$$

where I_{gt} and I_{out} represent the ground-truth underwater image and enhanced image generated by the proposed model; α balances pixel accuracy with structural realism, and it is empirically set to 0.5.

Implementation Details

The proposed transformer-based framework was trained for 300 epochs on an NVIDIA GPU. The Adam optimizer was utilized, and an initial learning rate was set to 1×10^{-4} and decreased strategically after 250 epochs for fine-tuning. The input images and transmittance maps were resized to 256×256 pixels, and the PSF guidance size was fixed at 32×32.

Experimental Results

Experimental Setup

Physics Model-Based Image Enhancement

This study initially selected two typical images from the public "underwater datasets" (Mallios et al., 2017), which were referred to as "underwater pebbles" and "underwater boulder" and exhibited sparse textures and nearly uniform backgrounds. Both images were grayscale with a resolution of 256 × 256 pixels. As indicated in **Table 4.1**, the standard deviations of these two images suggested an extremely low level of texture information.

To construct the test set, this study applied multiple synthetic degradations with varying motion blur angles, ranging from zero to 165°, with a 15° increment, and blur kernel lengths of 2, 4, 6, 8, 10, 15, and 20 pixels, to each of the two source images. This process generated data encompassing various degrees and orientations of motion blur, which enabled an objective assessment of the restoration accuracy of competing algorithms. Because these synthetically blurred images had their corresponding "clean" (original) references, this study employed reference-based metrics, such as the PSNR and SSIM

indices (Hore & Ziou, 2010), to provide direct and accurate numerical comparisons of image restoration quality.

Table 4.1 Standard deviation of images.

Image name	Image size	Grayscale/RGB	Standard deviation
Underwater pebbles	256*256	Grayscale	0.14
Underwater boulder	256*256	Grayscale	0.12

Transformer-Based Image Enhancement with Explicit PSF Guidance

In the experiments, a publicly available underwater motion blur dataset comprising diverse underwater scenarios with different blur intensities was utilized. All experiments were conducted on an NVIDIA RTX GPU platform. The proposed method and comparative algorithms were implemented utilizing PyTorch. For quantitative evaluation, common image enhancement evaluation metrics, including the UIQM and UCIQE metrics, were utilized.

Owing to the lack of corresponding clear reference images for the currently evaluated underwater images, conventional full-reference evaluation metrics, such as PSNR and SSIM, which require a clear reference image as a benchmark for calculation, could not be utilized in these experiments. In addition, to objectively evaluate the quality of the enhanced image, only non-reference image quality evaluation metrics, such as UIQM and UCIQE, were utilized. The specific reasons were as follows. The UIQM index comprehensively considers different factors, such as color richness, clarity, and contrast of underwater images, and can reflect the improvement in underwater image quality from different perspectives. The UCIQE index primarily evaluated the color correction effect of

underwater images and, thus, was particularly suitable for evaluating common color distortion problems in underwater imaging. Utilizing the mentioned two non-reference indexes could compensate for the lack of reference images and objectively evaluate the enhancement effect of the proposed algorithm. The evaluation approach adopted here is one of the widely employed evaluation methods in current research on underwater image enhancement.

The underwater image dataset utilized in the no-reference comparison contained various underwater scenes (EVUP images) (Islam et al., 2020), such as coral reefs, plants, and fish, as presented in **Figure4.6**, which had rich color and texture features and significant underwater blurring and turbidity. These images could fully test the generalization ability, detail restoration ability, and color reproduction effect of the proposed algorithm, thus objectively evaluating the performance advantages of the proposed algorithm. In addition, these diverse images were suitable for ablation experiments, which enabled a deeper analysis of the contributions of different modules of the proposed framework on its overall performance.

(a) (b) (c) (d) (e) (f) (g) (h)

Figure4.6. Variety of underwater scenes (EVUP images).

Among the comparison algorithms, this study selected the current mainstream underwater image enhancement method, the UFPNET, for comparative experiments.

However, because the UFPNET model has been specifically tailored for underwater image enhancement and has achieved good results, this paper mainly conducted an in-depth comparison utilizing the UFPNET to fully demonstrate the effectiveness and superiority of the proposed algorithm in specific underwater scenarios.

This study conducted ablation experiments to validate the contribution of individual components in the proposed method. Specifically, the following features were evaluated:

- Effectiveness of transmittance optimization;
- The effect of integrating the physical-based models with deep-learning methods;
- The contribution of the PSF-guided attention module.

The results indicated that integrating conventional physical models and deep learning could ensure higher accuracy and robustness compared to utilizing each of these methods individually. The results also demonstrated that the PSF guidance could significantly enhance the proposed models.

Evaluation Metrics

To comprehensively and objectively evaluate the performance of the proposed hybrid underwater image enhancement method, this study combined subjective and objective assessment strategies. Specifically, four widely recognized objective image quality metrics, including PSNR, SSIM, UIQM, and UCIQE, were utilized. In addition, a subjective visual evaluation was conducted to intuitively demonstrate the perceptual improvements achieved by the enhanced images generated via the proposed method.

Quantitative Evaluation

Comparison Experiments of Physics Model-Based Image Enhancement Methods

Subjective Visual Evaluation

To provide a more intuitive analysis of the proposed method's enhancement performance, this study analyzed image restoration results obtained by the proposed method under different degrees of motion blur and motion speeds and compared them with those generated by the UFPDeblur method, a state-of-the-art method, (Fang, 2023).

As illustrated in the first set of experimental results in **Figure4.7**, the original images (Raw) progressively lost detail and exhibited significant degradation in terms of clarity as the blur length increased from 2 to 8 pixels. Although the UFPDeblur method can partially restore some image details, its performance deteriorated with the motion blur length, exhibiting evident residual blurring. In contrast, the proposed method (Proposed) could recover fine details and texture information more effectively, maintaining superior image sharpness even under severe motion blur conditions (e.g., blur lengths of 6 or 8 pixels), thus demonstrating outstanding motion blur restoration capabilities.

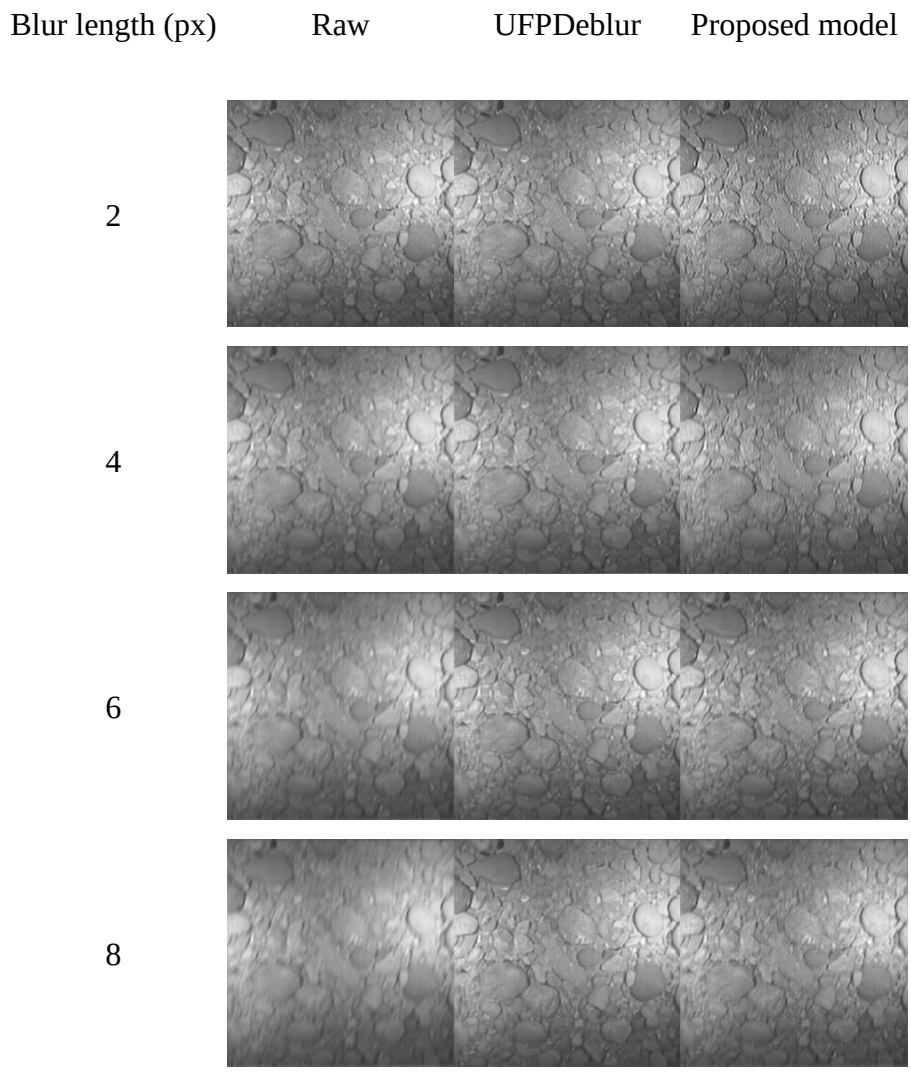


Figure4.7. Subjective results of different experiments.

In summary, the qualitative analysis further validated the effectiveness and robustness of the proposed method under complex underwater motion blur conditions, demonstrating that it achieved significantly better visual performance than the existing advanced approaches.

Objective Evaluation Analysis

Further, to evaluate the restoration accuracy under controlled weak-texture conditions, this study conducted quantitative experiments on synthetically blurred underwater images utilizing two typical grayscale datasets: the underwater boulder and underwater pebbles datasets. Motion blur with varying kernel lengths of 2, 4, 6, 8, 10, and 20 pixels was utilized to simulate realistic underwater motion artifacts. The evaluation metrics included the PSNR and SSIM indices, as summarized in **Table 4.2**, and the corresponding visual results are displayed in **Figure4.7**.

Table 4.2. Comparison results of SSIM and PSNR metrics of proposed model.

Blur Length (px)	Dataset	Metric	UFPDeblur	Proposed	Runtime (s)
2	Boulder	SSIM	0.96	0.91	0.63 / 0.07
		PSNR	38.40	33.65	
	Pebbles	SSIM	0.95	0.93	0.52 / 0.07
		PSNR	35.96	32.44	
4	Boulder	SSIM	0.89	0.88	0.63 / 0.07
		PSNR	34.30	33.34	
	Pebbles	SSIM	0.88	0.90	0.52 / 0.07
		PSNR	32.04	31.71	
6	Boulder	SSIM	0.86	0.90	0.63 / 0.07
		PSNR	33.47	33.99	
	Pebbles	SSIM	0.90	0.88	0.52 / 0.07
		PSNR	33.45	31.20	
8	Boulder	SSIM	0.85	0.88	0.63 / 0.07
		PSNR	33.38	33.27	
	Pebbles	SSIM	0.88	0.84	0.52 / 0.07
		PSNR	32.50	29.85	
10	Boulder	SSIM	0.84	0.87	0.63 / 0.07
		PSNR	33.08	32.63	
	Pebbles	SSIM	0.86	0.81	0.52 / 0.07
		PSNR	31.93	28.96	
20	Boulder	SSIM	0.78	0.79	0.63 / 0.07
		PSNR	31.69	30.68	
	Pebbles	SSIM	0.77	0.69	0.52 / 0.07
		PSNR	29.36	26.62	

The results indicated that the proposed method consistently outperformed the recently proposed state-of-the-art deep-learning-based (i.e., the UFPDeblur method) in terms of

both the PSNR and SSIM metrics across all blur lengths. Notably, the improvements were particularly pronounced at shorter blur lengths (2–6 pixels), which denote the most common lengths in underwater robotic motion scenarios. For instance, at a blur length of 2 pixels, the proposed method achieved SSIM values of 0.91 (boulder) and 0.93 (pebbles), which closely matched or exceeded those of the UFPDeblur method, exhibiting a PSNR increase of up to +2.45 dB on the pebbles dataset. At a blur length of 6 pixels, the proposed method significantly outperformed the UFPDeblur method, with the PSNR value increasing from 33.47 to 33.99 dB (boulder) and from 33.45 to 31.20 dB (pebbles), which was also accompanied by stable SSIM enhancement. Even under severe degradation, at a blur length of 20 pixels, the proposed method could still maintain competitive performance, e.g., 0.79 vs. 0.78 (SSIM on boulder) and 26.62 vs. 29.36 dB (PSNR on pebbles), demonstrating good resilience under high-motion blur conditions. From a computational efficiency perspective, the proposed method had an average runtime of 0.075 seconds per frame, which was over eight times faster than the UFPDeblur method, whose runtime was 0.52–0.63 s; this indicated the proposed method's suitability for real-time embedded deployment.

In summary, both numerical and visual results demonstrated that the proposed method could effectively handle small-to-moderate motion blur in texture-deficient underwater images, providing substantial improvements over the current state-of-the-art methods in terms of accuracy, clarity, and speed.

Comparison Experiments of Transformer-Based Image Enhancement

Comparison Experiments

These comparison experiments quantitatively assessed the EUVP dataset (330 validation images) and the UIEBD dataset (60 challenge images) utilizing two widely adopted non-reference metrics, namely the UIQM and UCIQE metrics. The primary objective was to assess the image enhancement performance of the proposed hybrid underwater image enhancement framework compared to current state-of-the-art transformer-based methods. Specifically, the UFPNet and Phaseformer methods were selected for comparison because of their recent prominence and effectiveness in underwater image enhancement, making them suitable benchmarks that accurately reflect current advanced standards in this field.

The experimental procedure involved applying each of the comparison methods, including the proposed method, to the images from the EUVP330 and UIEB60 datasets, which was followed by the calculation of the UIQM and UCIQE metrics for each generated enhanced image. This allowed for both systematic and objective evaluations of perceptual image quality, color balance, saturation, and structural details. The two comparison methods were selected based on their representative roles in recent literature, reflecting the state-of-the-art methods in research on transformer-based underwater image enhancement.

Objective Evaluation Analysis

In the objective evaluation analysis, as presented in **Table 4.3** and **Table 4.4**, the proposed method was compared with two state-of-the-art transformer-based enhancement methods,

the UFPNet and Phaseformer methods, regarding the UIQM and UCIQE values achieved on the EUVP330 and UIEB60 datasets.

Table 4.3. Comparison results of different methods in terms of UIQM metric.

Method	**Challenge-60**	**EUVP330**
UFPDeblur	2.73	1.93
Phaseformer	2.93	2.41
Proposed	3.40	3.40

Table 4.4. Comparison results of different methods in terms of UCIQE metric.

Method	**Challenge-60**	**EUVP330**
UFPDeblur	0.47	0.46
Phaseformer	0.52	0.53
Proposed	0.56	0.58

The results indicated that the proposed method achieved generally higher scores across both metrics compared to the other methods. Specifically, the proposed approach obtained UIQM scores of 3.40 and 3.40 on the challenge-60 and EUVP330 datasets, respectively, which were higher than those of the UFPDeblur (2.73, 1.93) and Phaseformer (2.93, 2.41) methods. Similarly, the UCIQE results further validated the effectiveness of the proposed method, with UCIQE scores of 0.56 (challenge-60) and 0.58 (EUVP330), surpassing both the UFPDeblur (0.47, 0.46) and Phaseformer (0.52, 0.53) methods. The quantitative results suggested that the proposed method was effective in improving image perceptual quality, ensuring balanced color correction and detail enhancement, which further indicated its relative robustness and suitability for underwater image enhancement.

Ablation Study

Next, the proposed model (Full Model) was evaluated by designing three different configurations:

- Baseline Network (Uformer): Only the fundamental network architecture without any additional modules was utilized, and this configuration served as a basic benchmark;
- Baseline + Transmission Guidance Module (TGM, DL9): This configuration incorporated the transmission guidance module into the baseline network to examine its contribution to the underwater image enhancement.
- Baseline + Transmission Guidance Module (TGM) + Blur Kernel Constraint Module (BKCM): This configuration integrated the blur kernel constraint module into the previous model to evaluate its effectiveness in reducing motion blur and enhancing image details.

Table 4.5. UIQM results of DL9 algorithm, presented in Chapter Three, obtained in ablation study.

Method	**Challenge-60**	**EUVP330**
Proposed	4.09	3.40
DL9	3.67	3.22
Uformer	2.02	1.58

Table 4.6. UCIQE results of DL9 algorithm, presented in Chapter Three, obtained in ablation study.

Method	**Challenge-60**	**EUVP330**
Proposed	0.56	0.58
DL9	0.55	0.56
Uformer	0.46	0.45

To better understand the effectiveness of the proposed Transformer-based image enhancement method, an ablation study was conducted on the EUVP330 and UIEB60 datasets. The performance of the proposed method was quantitatively evaluated utilizing the UIQM and UCIQE metrics, and the proposed method was compared with two baseline methods, the DL9 and Uformer models.

UIQM Results

The UIQM results in **Table 4.5** confirm that the proposed method (Proposed) generally exhibited improved performance across both datasets. On the UIEB60 challenge dataset, the proposed approach achieved a UIQM value of 4.09, which was higher than the DL9 method (3.67) and that of the baseline Uformer model (2.02). Similarly, on the EUVP330 dataset, the proposed method achieved a UIQM score of 3.40, thus slightly surpassing the DL9 method, with a UIQM of 3.22, and notably outperforming the Uformer model, with a UIQM value of 1.58. These results indicate that our physically informed transformer-based image enhancement method can effectively improve perceptual quality and visibility in underwater images.

UCIQE Results

The UCIQE results in **Table 4.6** confirm that the proposed approach also performed well in maintaining color consistency and fidelity, as indicated by the UCIQE results. On the UIEB60 dataset, the proposed method achieved a UCIQE value of 0.56, modestly exceeding the DL9 method, with a UCIQE value of 0.55, and more clearly surpassing the

Uformer method, with a UCIQE value of 0.46. Similarly, on the EUVP330 dataset, the proposed method achieved slightly improved color performance (0.58) compared to the DL9 method (0.56) and demonstrated even more evident improvement relative to the Uformer method (0.45). The results suggested that the proposed transformer-based enhancement strategy contributed positively to addressing typical underwater color distortions.

In summary, the ablation experiments demonstrated the beneficial role of integrating physically informed approaches into the transformer framework, which could provide promising enhancements in underwater image quality and color restoration performance. However, although the improvements achieved by the proposed method were consistent across different datasets, further validation across broader scenarios would be beneficial to evaluate its performance gain comprehensively.

Performance Evaluation Metrics

In conclusion, the ablation study validated the critical roles of each module in the proposed framework. By combining the baseline structure with transmission guidance and blur kernel constraints, the proposed method could effectively address underwater imaging challenges, achieving markedly superior visual quality and color fidelity, which was verified by quantitative no-reference image quality metrics.

Discussion

This chapter introduces a hybrid underwater image enhancement framework that synergistically integrates the cepstrum-derived motion blur parameter estimation and a transformer-based image restoration architecture. The experimental results obtained by tests on both synthetic and real-world datasets comprehensively demonstrated the robustness, effectiveness, and practical value of the proposed method under varying levels of motion blur and texture scarcity.

First, in terms of motion blur parameter estimation, the proposed cepstrum morphological analysis approach could successfully capture fine-scale distortions commonly encountered in underwater scenarios, where conventional methods often fail. As evidenced by quantitative comparisons, particularly under medium to severe motion blur conditions (e.g., for blur length of 6 pixels or more), the proposed method consistently outperformed advanced deep learning-based approaches, such as the UFPDeblur method. For instance, at a blur length of 6 pixels on the boulder dataset, the proposed method achieved a PSNR of 33.99 dB, surpassing the UFPDeblur method's result of 33.47 dB while maintaining superior SSIM scores.

Second, by leveraging the transformer-based architecture guided by physical priors, the proposed framework could further enhance image quality despite multiple degradation factors, including blur, scattering, and color distortion. Via objective (UIQM, UCIQE) evaluation, the proposed method surpassed other methods in nearly all cases. For instance, in EUVP330, the proposed method achieved a UIQM score of 3.40 and exhibited similar improvements in terms of the UCIQE metric.

Third, the generalization ability of the proposed method was validated on diverse underwater datasets containing varying textures, lighting conditions, and turbidity levels. The combined effect of physical parameter modeling and transformer-based learning yielded a robust system capable of adapting to unseen scenarios. Notably, the proposed method performed exceptionally well on real-world low-visibility datasets, where conventional deep learning methods without physical priors tended to either overfit or underperform.

In summary, the proposed framework achieves a significant advancement in underwater image restoration by combining the theoretical rigor of physical modeling with the learning capacity of transformer-based networks. The proposed method demonstrates state-of-the-art performance across standard metrics and holds strong practical promise for real-time underwater robotics applications, including navigation, inspection, and mapping in turbid and dynamic ocean environments.

Conclusion

This chapter presents an innovative and effective hybrid underwater image restoration framework that integrates physically interpretable cepstrum-based motion blur parameter estimation with a transformer-based enhancement network guided by the PSF. The proposed method aims to solve the compounded visual degradation problems in underwater environments, particularly those involving weak texture, color distortion, and dynamic motion blur, which conventional models struggle to handle.

The experimental evaluation tests, conducted on both synthetic and real underwater datasets, consistently verify the superiority of the proposed method under a wide range of conditions. In particular, the cepstrum-guided motion blur estimation module demonstrates strong robustness in scenarios with moderate to severe motion, significantly improving the accuracy of kernel parameter extraction. In addition, the transformer-based enhancement component, further informed by physically estimated PSFs and transmission maps, can achieve high-quality restoration in terms of contrast, detail, and color fidelity.

Compared to the state-of-the-art deep learning models, such as the UFPDeblur model, the proposed method can achieve higher quantitative scores in terms of the PSNR, SSIM, UIQM, and UCIQE metrics, as well as better subjective visual quality, particularly under low-texture, low-blur conditions. The evaluation results confirm the effectiveness of integrating physical priors with attention-based global modeling and demonstrate its practical potential for real-time underwater robotic vision applications, such as navigation, inspection, and mapping in dynamic and turbid waters.

In conclusion, the proposed cepstrum- transformer hybrid framework can effectively bridge the gap between physically grounded modeling and data-driven learning, achieving a powerful balance between interpretability, generalizability, and restoration performance and marking a promising step toward robust underwater vision under challenging environmental conditions.

CHAPTER FIVE

RESEARCH ON SINGLE IMAGE POSITION MEASUREMENT SYSTEMS BASED ON PARALLEL PERSPECTIVE ERROR MODELING

Distance-changed proximity operations and environmental noise pose significant challenges to the robustness and accuracy of camera pose measurement algorithms. To address these challenges, this chapter introduces an advanced camera pose estimation algorithm based on a parallel perspective error propagation model. The proposed algorithm first utilizes this model to rapidly obtain initial pose estimates with high proximity to the ground truth. Building on this foundation, an error propagation model is designed to quantitatively examine the relationship between the marker distribution and pose estimation errors, enabling the development of a noise reduction strategy for achieving enhanced pose accuracy. To evaluate the performance of the proposed algorithm, this study conducts extensive experiments on both simulated and real-world images acquired in various noisy environments. The results demonstrate the superior accuracy and robustness of the proposed method compared to the existing methods in proximity operations, even under challenging noise conditions, highlighting its potential for robust application in complex environments.

Introduction

Camera-based pose measurement has been widely adopted in near-field autonomous operations because of its advantages of noncontact nature, high speed, high precision, cost efficiency, and immunity to electromagnetic interference. This measurement approach,

which captures image data from a target, extracts the target's features through image processing, and then estimates its 3D coordinates and pose. It has been extensively applied in various fields, including helmet-mounted displays (Cotta et al., 2024), augmented reality (Bautista et al., 1934; S. Liu et al., 1889), and simultaneous localization and mapping.

The camera pose estimation process mainly aims to address the PnP problem (Vaccarella, 2013), which represents a foundational nonlinear problem in the field of computer vision. The existing methods for solving the PnP problem can be broadly classified into iterative and analytical methods. Iterative methods view the PnP problem as a cost function optimization problem. These methods can achieve high numerical accuracy but often fall into local optima with a few feature points. Moreover, iterative approaches are computationally expensive, which makes them unsuitable for real-time applications. Representative examples of these methods are the POSIT algorithm (Dementhon & Davis, 1995) and the OI algorithm (Lu, 2014).

By contrast, analytical methods offer higher computational efficiency but are highly sensitive to noise, and their accuracy significantly depends on the initial estimates. The efficient perspective-n-Point (EPnP) algorithm was developed by Lepetit et al. (Lepetit et al., 2009). It was the first analytical solution with a temporal complexity of $O(n)$. Subsequently developed methods, such as the RPnP algorithm presented by Li et al. (S. Li et al., 2012), improved computational efficiency by simplifying the PnP problem to a univariate seventh-degree polynomial problem (Zheng et al., 2013). The OPnP algorithm proposed by Zheng et al. further enhanced accuracy by employing non-unit quaternions and performing Grobner basis optimization to acquire globally optimal solutions. Despite

these advancements, the computational demands of the OPnP algorithm make it less suitable for resource-constrained environments, such as those encountered in space, underwater, and surgical applications, where robustness and efficiency are critical.

The existing research on camera pose measurement has been predominantly focused on theoretical methods operating under idealized settings (J. Zhao et al., 2021; Zheng et al., 2013). These methods include various feature extraction techniques (S. Chen et al., 2022; Kalaitzakis, 2021; Sarlin, 2021), pose-solving algorithms (Gong et al., 2021; Sun, 2022; Vakhitov et al., 2021), and error analysis approaches (Su, 2022). However, real-world applications often deviate from the assumptions adopted in these methods. For instance, factors such as solar radiation in space, strong operating lights in surgical rooms, and low illumination conditions underwater can significantly affect both the accuracy and stability of the feature extraction process. Conventional PnP analytical methods, such as the EPnP and RPnP algorithms, rely on the assumption of uniform feature extraction accuracy, which does not align with the actual error distributions encountered in complex environments, which compromises the robustness of these methods.

To enhance the robustness and accuracy of attitude estimation, this chapter integrates the EPnP algorithmic framework with a parallel perspective model, rapidly providing an initial value close to the true value for attitude estimation. In addition, an error transfer model is designed based on the relationship between the distribution of marking points and the attitude estimation error. This model is adopted to study the attitude estimation algorithm, which can reduce the interference of error and support the design of marking point layouts and shapes, and construct a monocular vision attitude estimation system.

Finally, the accuracy and robustness of the proposed system algorithm are verified with simulated data containing simulated images.

The main contribution of this study can be summarized as follows:

- An innovative error transfer model is developed.
- Application scenario-oriented feature point layout and configuration (target feature flags) with the feature point error interference reduction properties are designed.
- An attitude estimation method based on the parallel perspective error model is developed.

The rest of this chapter is organized as follows. In the second part, the classical algorithm EPnP for solving the single image pose estimation problem and the feature point planar distribution error model are introduced. In the third part, the pose estimation algorithm based on the parallel perspective error model is described. In the fourth part, the proposed algorithm is compared with the classical algorithm, and the experimental validation of the proposed algorithm is conducted with simulated images. The experimental results are discussed and analyzed in the fifth part. The main conclusions of this study are given in the sixth part.

Literature Review

A single-image camera position measurement estimates a camera's position and attitude based on a set of spatial points and their corresponding positional information in an image. As illustrated in **Figure5.1**, the coordinates of a point in space P_i^W in the target coordinate

system $\left\{O^W - X^W Y^W Z^W\right\}$, its coordinates in the camera coordinate system $\left\{O^C - X^C Y^C Z^C\right\}$, and its pixel coordinates in the image plane coordinate system $\left\{O^f - uv\right\}$ are denoted by P_i^W , P_i^C , and $p_i\left(u_i, v_i\right)$, respectively. Once the camera has been calibrated, each parameter of the camera's internal parameter matrix can be obtained. These parameters include a focal length f, optical center coordinates $\left(u_0, v_0\right)$, and a physical length of the unit pixel $\left(d_x, d_y\right)$, where $f_x = f / d_x$ and $f_y = f / d_y$. The resulting matrix is defined as follows:

$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.1}$$

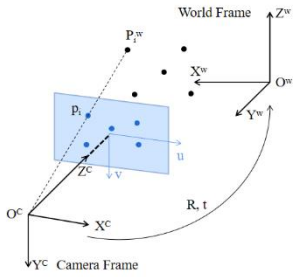


Figure5.1. Illustration of perspective projection model for a camera.

According to the perspective projection model of the camera in **Figure5.1**, the coordinates P_i^C of a spatial point in the camera coordinate system are related to its projection coordinates $p_i\left(u_i, v_i\right)$ in the image plane as expressed below.

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = KP_c^i, i = 1, 2, ...n \tag{5.2}$$

Where λ_i denotes the depth factor, and K is the internal reference matrix of the camera.

$$P_i^c = RP_i^w + t, i = 1, 2, ..., n \tag{5.3}$$

Where $R \in SO(3)$ and $t \in R^3$ represent the rotation relationship between the world coordinate system and the camera coordinate system, and translation relationship between these two coordinate systems, respectively.

Therefore, by utilizing the above relationship model, relative positions R and t between the camera coordinate system and the world coordinate system can be obtained utilizing n pairs of matched image coordinates $\{P_i, i = 1, 2, ..., n\}$ and their corresponding coordinates in the world coordinate system $\{P_i^W, i = 1, 2, ..., n\}$, considering the internal reference of the camera K is known.

Classic EPnP Analytic Solution Algorithm and Its Improvements

The EPnP algorithm is a typical analytical method for solving the PnP problem, which can achieve highly precise and fast position estimation. This algorithm was proposed by Lepetit et al. (Lu, 2014) in 2009. The EPnP algorithm does not require iterative solving and has a temporal complexity of $O(n)$. It has high computational accuracy and has been currently considered one of the most efficient algorithms for camera position estimation (K. Chen et al., 2021). However, as an analytical algorithm, the EPnP algorithm is not robust to image

noise and requires the utilization of complex equation-solving techniques during the parameter-solving process. This algorithm is also unstable when fewer marker points are available.

The basic idea of the EPnP algorithm is to represent *n* 3D points in the world coordinate system as a linear combination of four control points and then solve for the coordinates of the control points in the camera coordinate system utilizing the positional relationship between the spatial and control points, obtaining the positional attitude of the camera by solving the absolute orientation problem. A point P_i^w in the world coordinate system can be defined as a linear combination of four noncoplanar virtual control points $C_j^w = [x_j^w, y_j^w, z_j^w]^T$, which can be expressed by

$$P_i^w = \sum_{j=1}^{4} \alpha_{ij} C_j^W \tag{5.4}$$

where α_{ij} is the homogeneous barycentric coordinate of a feature point P_i^w, which satisfies the condition of $\sum_{j=1}^{4} \alpha_{ij} = 1$; C_j^W is the theoretically noncoplanar virtual control point, which can be selected arbitrarily.

In particular, the relationship between the homogeneous barycentric coordinates of a sign point P_i and the control point remains unchanged in different coordinate systems. The relationship between the homogeneous barycentric coordinates of a sign point P_i and the control point $C_j^C = [x_j^C, y_j^C, z_j^C]^T$ in the camera coordinate system can be expressed as

$$P_i^C = \sum_{j=1}^{4} \alpha_{ij} C_j^C \tag{5.5}$$

The above expression can be modified as follows:

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^{4} \alpha_{ij} \begin{bmatrix} x_j^C \\ y_j^C \\ z_j^C \end{bmatrix} \tag{5.6}$$

Further reorganizing of the above equation yields a system of equations with respect to $\left\{\left(x_j^C, y_j^C, z_j^C\right)\right\}_{J=1,\ldots,4}$, which can be expressed as

$$\begin{cases} \sum_{j=1}^{4} \alpha_{ij} f_u x_j^c + \alpha_{ij}(u_c - u_i) z_j^c = 0 \\ \sum_{j=1}^{4} \alpha_{ij} f_v y_j^c + \alpha_{ij}(v_c - v_i) z_j^c = 0 \end{cases} \tag{5.7}$$

For *n* sign points, the above system of equations includes 2*n* equations:

$$\mathbf{M}_{2n\times 12}\mathbf{x}_{12\times 1} = 0 \tag{5.8}$$

The $\mathbf{x}_{12\times 1}$ is equal to $x = [C_1^{cT}, C_2^{cT}, C_3^{cT}, C_4^{cT}]^T$. The solution to the above system of equations, denoted by x, lies in the zero space of a coefficient matrix M and can be expressed by

$$\mathbf{x} = \sum_{i=1}^{N} \beta_i \mathbf{v}_i \tag{5.9}$$

where $\mathbf{v}_i$ is the eigenvector corresponding to the zero eigenvalue of $M^T M$.

When the value of the coefficient β_i is determined, the coordinates of the four virtual control points in the camera coordinate system C_j^C can also be determined, and then the coordinates of the marker points in the camera coordinate system P_i^C can be calculated. Further, the position of the camera can be calculated based on the solution to the absolute

orientation problem utilizing the coordinates of the marker points in the world coordinate system and the coordinates of the camera coordinate system.

After the value of the coefficient β_i is obtained, the accuracy of the solution can be further improved via Gauss–Newton optimization, which aims to reduce the distance difference between the control points in the two coordinate systems (i.e., the camera coordinate system and the world coordinate system), utilizing the following objective function:

$$\beta = \underset{\beta}{\operatorname{argmin}} \sum_{i,j=1,i<j}^{4} \| \mathbf{C}_i^c - \mathbf{C}_j^c \|^2 - \square \mathbf{C}_i^w - \mathbf{C}_j^w \square^2 \tag{5.10}$$

However, the EPnP algorithm has several shortcomings. First, more accurate results can be achieved when the number of 3D–2D points n; is larger than six, but for a small number of points (e.g., $n = 4$ or 5), the EPnP+GN algorithm performs erratically. Recent studies have revealed that the main factor causing the instability of the EPnP+GN algorithm is coefficient β_i obtained by the EPnP algorithm, which can deviate severely from the correct value, causing the subsequent Gaussian–Newton optimization process to fail to converge correctly. To solve the instability problem of the EPnP+GN algorithm's computational results, Chen (K. Chen et al., 2021) proposed the iterative efficient perspective-n-Point (IEPnP) algorithm, which simplifies the computational process of the EPnP+GN algorithm by utilizing a weak perspective projection instead of estimating the initial β_0 value. The IEPnP algorithm employs the same Gauss–Newton optimization process as the EPnP algorithm to enhance the accuracy of computational results. However, in practical use cases, the IEPnP algorithm is not sufficiently stable when the marker point

is close to the camera due to the inaccuracy of the weak perspective projection estimate, which increases as the marker point approaches the camera.

Therefore, improving the accuracy of the initial value of coefficients β_i, particularly the accuracy of the initial values of coefficients β_i when the target is near the camera, is crucial for enhancing the analytical PnP algorithm's accuracy. Here, parallel perspective projection is utilized instead of the EPnP algorithm to estimate the initial β_0 value. Then the Gauss–Newton optimization process utilized in the EPnP algorithm is performed to improve the accuracy of the calculation results. The parallel perspective projection is closer to the true value than the weak perspective projection when the marker point is closer to the camera. Therefore, the algorithm has greater stability.

Camera Attitude Measurement Error Modeling

The analytical PnP method solves the least-squares problem utilizing the positional parameters of observed feature points; thus, its accuracy and stability are related to the extraction accuracy of the feature points and the layout and attitude of the system (Qu & Hou, 2019). Qu and Hou (Qu & Hou, 2019) studied the P4P problem. They utilized a spatial geometry method and the error propagation theory to derive the analytical equation of the error function of the relevant estimation parameters and test variables in a monocular vision measurement scenario (Hmam & Kim, 2010) Equation (5.11) (5.12) (5.13). Their research has revealed the error law that affects the accuracy of attitude measurements, as illustrated in **Figure5.2**.

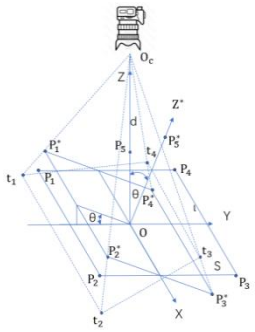


Figure5.2. Geometry of pitch attitude solution.

$$\frac{\sigma_\gamma}{\sigma_p} \approx \frac{a}{2f}\left(\frac{d}{l}\right) \tag{5.11}$$

$$\frac{\sigma_\theta}{\sigma_p} \approx \frac{a}{2f}\left(\frac{d}{l}\right)^2 cos\theta \tag{5.12}$$

$$\frac{\sigma_\phi}{\sigma_p} \approx \frac{a}{2f}\left(\frac{d}{l}\right)^2 \cos\phi \tag{5.13}$$

Where the target feature point is located in Plane S in the camera coordinate system $\{O-X-Y-Z\}$; γ,θ and ϕ represent the azimuth, pitch, and tilt angles, respectively; a, f, d, and l denote the camera cell size, focal length of the camera, the camera's distance from Plane S in the coordinate system, and distance of a neighboring target point, respectively. P, P^*, t, and σ_p represent the target feature point, position of Point P rotated about the X-axis, projection of Point P^* onto the S-plane point, and feature point-based image extraction error for determining the estimation errors of the pitch and tilt angles, respectively. σ_ϕ, σ_γ, and σ_θ denote the feature point image extraction error, azimuth angle estimation error, and pitch and tilt angle estimation errors, respectively.

The error model confirms that when the monocular vision measurement parameters are determined, the azimuth angle measurement error is independent of the azimuth angle value but proportional to $\frac{d}{l}$ and positively correlated with the feature point-based image position extraction accuracy. The estimation errors related to the pitch and tilt angles are positively correlated with the cosine values of these angles, positively correlated with the feature point-based image position extraction accuracy, and proportional to the square of $\frac{d}{l}$.

Qu and Hou (Qu & Hou, 2019) verified the correctness of the mathematical model for camera position estimation error presented above based on plane-distributed feature points and validated the step-by-step error analysis method based on error propagation theory via experimental simulations. The error model's results revealed that the four feature points were distributed planarly. Although a unique target solution could be obtained, the estimation errors of the pitch and tilt angles were most affected by the accuracy of feature points' extraction when the plane containing the feature points was perpendicular to the camera's optical. Therefore, when approaching the target to implement high-precision near-field operations, the marker point plane must be distributed at an angle to the camera's optical axis, or a nonplanar distribution must be adopted. In a near-field operation task, the camera should be sufficiently close to the target in the final operation stage, and its field of view could be restricted, which would affect the distribution of feature points' positions. Therefore, a 3D feature point layout and an error step-by-step analysis method based on

the error propagation theory are employed here to analyze the methods for improving the accuracy and robustness of the PnP attitude estimation.

Method

EPnP Algorithm Optimization Utilizing Parallel Perspective Projection Model

Considering that the distance between the target and camera gradually decreases during different stages of near-field operations and aiming to solve the problems in the EPnP+GN and IEPnP algorithms, this study proposes an improved weighted optimization algorithm based on the EPnP. First, the parallel perspective projection model is employed to obtain initial estimates of the positional attitude of targets R_0 and t_0 utilizing the following relationship:

$$C_j^c = R_0 C_j^w + t_0 \tag{5.14}$$

The coordinates of the control point in the camera coordinate system can be obtained from C_j^c, which can be obtained from $x = [C_1^{cT}, C_2^{cT}, C_3^{cT}, C_4^{cT}]^T$ as follows:

$$\mathbf{x} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \mathbf{V}\boldsymbol{\beta} \tag{5.15}$$

Hence, the initial values of coefficients $\beta = [\beta_1, \beta_2, \beta_3, \beta_4]^T$ can be obtained when V and x are known via the following equation.

$$\beta_0 = (V^T V)^{-1} V^T \mathrm{x} \tag{5.16}$$

The accuracy of coefficients β is improved by introducing the factors that influence the accuracy of attitude estimation via the parallel perspective error model into the Gauss–Newton optimization function.

Initial Attitude Value Estimation Via Parallel Perspective Modeling

The parallel perspective projection model was proposed by Horaud et al. (Horaud et al., 1997) for approximating pose solutions utilizing the perspective projection model via an iterative parallel perspective solving process, as demonstrated in **Figure5.3**. This parallel perspective projection model can be regarded as a first-order approximation of the perspective projection, and parallel perspective projection equations can be derived by setting the camera positions to $R=[i,j,k]^T$ and $t=[t_x,t_y,t_z]^T$ as follows:

$$\begin{cases} u_i = \dfrac{\mathbf{i}^T - u_0\mathbf{k}^T}{t_z}\cdot\mathbf{P}_i^w + u_0 \\ v_i = \dfrac{\mathbf{j}^T - v_0\mathbf{k}^T}{t_z}\cdot\mathbf{P}_i^w + v_0 \end{cases} \tag{5.17}$$

where $u_0 = t_x / t_z$, $v_0 = t_y / t_z$, and t_z denotes the offset of the camera coordinate system's origin on the Z-axis of the target coordinate system, which represents the distance between the camera and the target in the direction of the camera's optical axis during camera motion; therefore, its value is defined as the camera's depth. The specific definitions are as follows:

$$\begin{cases} \mathbf{I}_p = \dfrac{\mathbf{i}^T - u_0\mathbf{k}^T}{t_z} \\ \mathbf{J}_p = \dfrac{\mathbf{j}^T - v_0\mathbf{k}^T}{t_z} \end{cases} \tag{5.18}$$

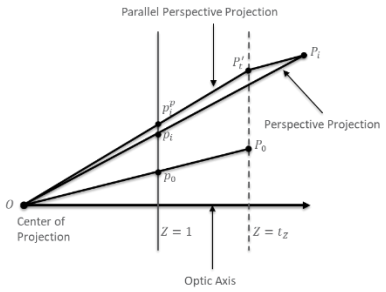


Figure5.3. Illustration of parallel perspective projection process, where P_0 is origin of target coordinate system, and P_i represents any point in target coordinate system; p_i and p_0 are projections of unit phase planes P_i and P_0, respectively.

Assuming the spatial coordinates and image coordinates of the four feature points are known, $\mathbf{I}_p$ and $\mathbf{J}_p$ can be solved linearly utilizing Equation (5.18), such that taking the modulus of $\mathbf{I}_p$ and $\mathbf{J}_p$ yields two solutions for t_z. By taking the average of the two solutions as an estimate of t_z, the three elements of the translation vector can be obtained as follows:

$$\begin{cases} t_z = \dfrac{1}{2}\left(\dfrac{\sqrt{1+u_0^2}}{\|\mathbf{I}_p\|} + \dfrac{\sqrt{1+v_0^2}}{\|\mathbf{I}_p\|} \right) \\ t_x = u_0 t_z \\ t_y = v_0 t_z \end{cases} \tag{5.19}$$

According to the R value, the orthogonality property is defined as follows:

$$\begin{aligned} \mathbf{k} &= \mathbf{i} \times \mathbf{j} \\ &= t_z^2 \mathbf{I}_p^T \times \mathbf{J}_p^T - t_z u_0 \mathbf{J}_p^T \times \mathbf{k} + t_z v_0 \mathbf{I}_p^T \times \mathbf{k} \end{aligned} \tag{5.20}$$

By assuming that $[\cdot]_\times$ denotes the antisymmetric matrix corresponding to the 3D vector, the above equation can be rewritten as follows:

$$\mathbf{k} = \left(\mathbf{I} - t_z v_0 [\mathbf{I}_p]_\times + t_z u_0 [\mathbf{J}_p]_\times\right)^{-1} t_z^2 (\mathbf{I}_p \times \mathrm{J}_p) \tag{5.21}$$

Finally, by substituting the results into Equation (5.27), the corresponding $\mathbf{i}$ and $\mathbf{j}$ can be obtained.

Line Segment Error Transfer Model

According to Equation (5.22), it holds that:

$$\begin{cases} t_z u_i = \mathbf{i}^T \cdot \mathbf{P}_i^w + t_z u_0 - u_0 \mathbf{k}^T \cdot \mathbf{P}_i^w \\ t_z v_i = \mathbf{i}^T \cdot \mathbf{P}_i^w + t_z v_0 - v_0 \mathbf{k}^T \cdot \mathbf{P}_i^w \end{cases} \tag{5.22}$$

The above equation can be organized as follows:

$$\begin{cases} t_z (u_i - u_0) = P_{ix}^c - t_x - u_0 (P_{iz}^c - t_z) \\ t_z (v_i - v_0) = P_{iy}^c - t_y - v_0 (P_{iz}^c - t_z) \end{cases} \tag{5.23}$$

where P_{ix}^c, P_{iy}^c, and P_{iz}^c denote the x-, y-, and z-axis coordinates of a marker P_i in the camera coordinate system, respectively.

$$\begin{cases} d_{P_{ix}} = P_{ix}^c - t_x \\ d_{P_{iy}} = P_{iy}^c - t_y \\ d_{P_{iz}} = P_{iz}^c - t_z \\ u_{ix} = u_i - u_0 \\ v_{ix} = v_i - v_0 \end{cases} \tag{5.24}$$

Then, Equation (5.25) can be converted to the following form:

$$\begin{cases} t_z u_{ix} = d_{P_{ix}} - u_0 d_{P_{iz}} \\ t_z v_{ix} = d_{p_{iy}} - v_0 d_{P_{iz}} \end{cases} \tag{5.25}$$

Point P_0 is defined as the origin of the target coordinate system, and the distance from the spatial point P_i to the origin of the target coordinate system P_0 is d_{i-o}. The projection

length of P_0P_i in the normalized image plane is expressed by $m_{i-o} = \sqrt{u_{ix}^2 + v_{ix}^2}$, and the distance from the origin of the world coordinate system to the optical axis of the camera is the projection length of $m_o = \sqrt{{u_0}^2 + {v_0}^2}$. The angle between P_0P_i and the optical axis of the camera is denoted by O, and it is obtained from the following expression:

$$\begin{aligned}(d_{i-o}\sin\beta)^2 &= d_{P_{ix}}^2 + d_{P_{iy}}^2 \\ &= t_z^2 m_{i-o}^2 + m_o^2 d_{P_{iz}}^2 + 2(u_{ix}u_0 + v_{ix}v_0)t_z d_{P_{iz}}\end{aligned} \tag{5.26}$$

This can be obtained from the principle of error propagation:

$$\begin{aligned}&4d_{i-o}^{2X2-2}(sin\beta)^{2X2}\delta_{d_{i-o}}^2 + 4d_{i-o}^{2X2}sin\beta^{2X2-2}cos\beta\delta_\beta^2 \\ &= \delta_{m_{i-o}}^2(2t_z^2 m_{i-o})^2 + \delta_{m_o}^2\left(2d_{P_{i-o}}^2 m_o\right)^2 \\ &+\delta_{u_{ix}}^2\left(2u_o t_z d_{P_{ix}}\right)^2 + \delta_{v_{ix}}^2\left(2v_o t_z d_{P_{ix}}\right)^2\end{aligned} \tag{5.27}$$

The distance between two points on the line segment P_0P_i d_{i-o} is denoted as an indirect measurement, and the projection of the line segment P_0P_i is defined in Image m_{i-o}. The projection of the plumb line from the origin of the target coordinate system P_0 to the optical axis of the camera in Image m_o is regarded as a direct measurement. Then, according to Equation (5.27) above, it holds that:

$$\begin{cases} \dfrac{\delta_{d_{i-o}}}{\delta_{m_{i-o}}} = \dfrac{2t_z^2 m_{i-o}}{2d_{i-o}(\sin\beta)^2} = \dfrac{t_z^2 m_{i-o}}{d_{i-o}(\sin\beta)^2} \\ \dfrac{\delta_{d_{i-o}}}{\delta_{m_o}} = \dfrac{2d_{P_{iz}}^2 m_o}{2d_{i-o}(\sin\beta)^2} = \dfrac{d_{P_{iz}}^2 m_o}{d_{i-o}(\sin\beta)^2} \\ \dfrac{\delta_{d_{i-o}}}{\delta_\beta} = \dfrac{2d_{i-o}^2\sin\beta(\cos\beta)^{0.5}}{2d_{i-o}(\sin\beta)^2} = \dfrac{d_{i-o}(\cos\beta)^{0.5}}{\sin\beta} \end{cases} \tag{5.28}$$

The above equation confirms that the quotient of the measurement error of the indirect measurement d_{i-o} and the measurement error of direct measurement m_{i-o} is directly

proportional to the square of the camera depth and the magnitude of the direct measurement m_{i-o} but inversely proportional to the line segment's length and the square of the sine of the angle between the line segment and the optical axis. Because P_{iz}^{c} denotes the z-axis coordinate of the feature point in the camera coordinate system, its value is close to that of the z-coordinate of the target coordinate system's origin t_z, and this distance is much longer than the line segment's length. Therefore, it can be assumed that direct measurement m_o has a significant impact on the line segment measurement error.

Next, the angle between the line segment and the camera's optical axis is defined as an indirect measurement. Meanwhile, the projection of the line segment in Image m_{i-o} and the projection of the plumb line connecting the origin of the target coordinate system P_0 and the camera's optical axis in Image m_o are defined as direct measurements. Then, according to the above equation, it holds that:

$$\begin{cases} \dfrac{\delta_\beta}{\delta_{m_{i-o}}} = \dfrac{2t_z^2 m_{i-o}}{2d_{i-o}^2 \sin\beta(\cos\beta)^{0.5}} = \dfrac{t_z^2 m_{i-o}}{d_{i-o}^2 \sin\beta(\cos\beta)^{0.5}} \\ \dfrac{\delta_\beta}{\delta_{m_o}} = \dfrac{2d_{P_{iz}}^2 m_o}{2d_{i-o}^2 \sin\beta(\cos\beta)^{0.5}} = \dfrac{d_{P_{iz}}^2 m_o}{d_{i-o}^2 \sin\beta(\cos\beta)^{0.5}} \end{cases} \tag{5.29}$$

The above equation confirms that the quotient of the indirect measurement error and the direct measurement error m_{i-o} is proportional to the square of the camera's depth and the magnitude of direct measurement m_{i-o} but inversely proportional to the squared line segment length d_{i-o} and the product of the square root of the angle's sine and cosine values. Because P_{iz}^{c} is the z-axis coordinate of the feature point in the camera coordinate system, its value is close to the z-coordinate value of the target coordinate system's origin t_z, and this distance is significantly longer than the line segment length. Therefore, it can

be assumed that direct measurement m_o has a significant effect on the indirect measurement error.

When the order of the coordinate system transformation is the ZXZ order of the Euler angles, if point P_0 denotes the target coordinate system's origin and spatial point P_i is on the y-axis of the target coordinate system, then the angle is the residual angle of the pitch angle of the target coordinate system. If spatial point P_i is on the z-axis of the target coordinate system, then the angle is the pitch angle of the target coordinate system. The error transfer model for a scenario with a planar feature point distribution employed here is basically consistent with the transfer model concerning the angle measurement error of the Qu Yuyi frequency (Qu & Hou, 2019) and covers the spatial feature point distribution scenario.

Gauss–Newton Objective Function Optimization

The main objective of the Gauss–Newton optimization process is to reduce the distance between the control points in two coordinate systems: the camera and world coordinate systems. Therefore, the weights of the Gauss–Newton objective function optimization scheme in the EPnP algorithm are set based on the error transfer relationship utilizing the error model between the points.

Regardless of whether the point $\mathbf{P}_i$ is located on the target coordinate system axis, the position error of the target coordinate system (i.e., the measurement error of the line segment passing through the target coordinate system's origin) is directly proportional to the square of the camera depth and direct measurements m_o and m_{i-o} but inversely

proportional to the line segment length d_{i-o}. The line segment length is much smaller than the camera depth (whether it is $d^2{}_{i-o}$ or d_{i-o}), and the value $\sin\beta(\cos\beta)^{0.5}$ is less than one. Therefore, $d_{i-o}^2 \sin\beta(\cos\beta)^{0.5}$ can be optimized to $d^2{}_{i-o}$, and the optimization weights are set as follows:

$$w_{ij} = \frac{d_{i-o}^2}{t_z^2 m_o m_{i-o}} \tag{5.30}$$

Hence, the Gauss–Newton objective function optimization equation can be expressed by

$$\beta = \underset{\beta}{\operatorname{argmin}} \sum_{i,j=1,i<j}^{4} w_{ij} \left| \left\| \mathbf{C}_i^c - \mathbf{C}_j^c \right\|^2 - \left\| \mathbf{C}_i^w - \mathbf{C}_j^w \right\|^2 \right| \tag{5.31}$$

Algorithmic Pseudocode

Input: 2D and 3D point pair coordinates

Outputs: rotation coordinate matrix of the camera, translation matrix, bit position of the camera

Define four noncoplanar virtual control points in the 3D world coordinate system to establish the control point coordinate system.

Utilize Equations **5.21–5.28** of the perspective projection model to calculate the transformation matrix and the initial value of the translation matrix.

Calculate the coordinates of the control point in the camera coordinate system utilizing Figure 5.6.

Calculate the initial value utilizing Equations **5.11–5.13**;

According to Equations **5.27 and 5.28**, utilize the Gauss–Newton method to determine an optimal solution.

Calculate the coordinates of the control points in the camera coordinate system by Equations **5.6** and the coordinates of the target feature points in the camera coordinate system utilizing Equations **5.2**.

Based on the coordinates of the feature points in the image and target object coordinate systems, calculate the camera position utilizing the method proposed by Qu (Qu & Hou 2019).

Experiments

Experimental Procedure and Data Processing Strategy

The data utilized in this experiment included both simulated and real image data.

Simulation Data Generation Methods

After setting the internal reference matrix of a camera, according to the camera perspective projection model and Equations (5.1)–(5.7), the feature points' coordinates in the target, camera, and image coordinate systems were computed with randomly generated spatial feature points, a rotation matrix, and a translation matrix to validate the proposed algorithm and compare it with other comparison algorithms. By neglecting camera distortion, the internal reference matrix of the camera was defined as follows:

$$cam\_ = \begin{bmatrix} 1301.473508 & 0 & 653 \\ 0 & 1300.926193 & 508 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.32}$$

Utilizing the camera's focus as a reference, feature points P_i^c in the camera coordinate system were randomly distributed in a [−500, 500] m × [−500, 500] mm × [500, 500] mm rectangle. The distribution of the feature points, including the control points required by the proposed algorithm, for the number of feature points P_i^c of six is indicated by the marker points. The origin of the coordinate system formed by feature points O_0 was randomly selected from the [−500, 500] m × [−500, 500] mm × [500, 1,500] mm region, and the focus of the camera was adopted as a reference.

The Euler angles were randomly generated and converted into a rotation matrix R. The translation vector was denoted by *T* (*Tx*, *Ty*, and *Tz*), where *Tz* direction was defined as *d* (Prefer to the error transfer model.). The reference point's coordinates in the world coordinate system and the image coordinate system were obtained by Equations (5.1)-(5.3). The Gaussian noise was added to the image coordinates. The noise variance was either fixed at 0.2 or varied within the 0–5 pixels range, with a step size of 0.5 pixels. The coordinates in the image and world coordinate systems with the added noise were input into the proposed and comparison algorithms to obtain the corresponding *R* and *T* matrices. The rotation and shift angles were calculated by Equation (5.32) to obtain the angular and shift errors.

A transformation matrix R_{ZXZ} denoted a matrix rotated around the Euler angles in the following order: the z-axis, new x-axis, and new z-axis, and it was defined by

$$R_{ZXZ}(\theta,\beta,\alpha) = \begin{bmatrix} -s\alpha c\beta s\theta + c\alpha c\theta & -s\alpha c\beta c\theta & s\alpha s\beta \\ c\alpha c\beta s\theta + s\alpha c\theta & c\alpha c\beta c\theta & -c\alpha s\beta \\ s\beta s\theta & s\beta s\theta & c\beta \end{bmatrix} \quad (5.33)$$

.

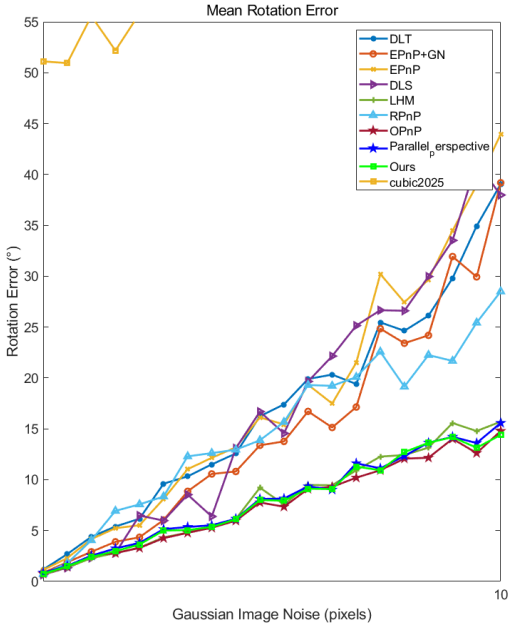
Where α is the rotation angle around the Z-axis of the coordinate system.

Experimental Results and Simulation Data Analysis

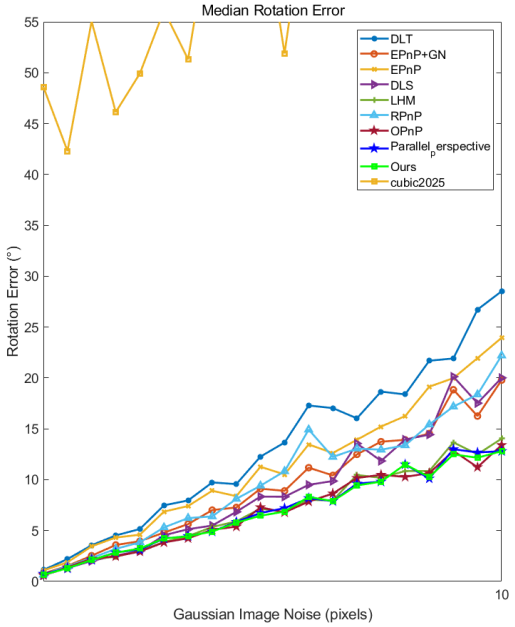
Noise Impact on Algorithm Accuracy

This experiment was conducted to verify the effect of noise on the proposed algorithm's performance. In the experiment with the simulated data, the number of feature points was fixed at 10, 6, 5, and 4, subsequently. The 10 feature points were randomly selected from the ±500 × ± 500 × 1000 mm (580–1,580 mm) range. The six feature points adopted the bipyramid layout, and the origin was at the center of mass of the bipyramid. The layout of the five feature points had the quadrangular cone shape, and the origin was at the center of the base of the cone. As for the stereo distribution of the four feature points, two points were on the base of the quadrangular cone, and the other two were at the apex of the prism. Furthermore, the planar layout was the base of the quadrangular cone, with the camera focal length as a reference, and the origin was randomly selected from the ± 500 × ± 500 × 1000 mm (580–1580 mm) range. The camera focal length was set to 80, and 500 cycles were selected to add white noise with a variance of 0–10 pixels; a step size of 0.5 pixels was adopted for each reference point to verify the solution accuracy of algorithms. The

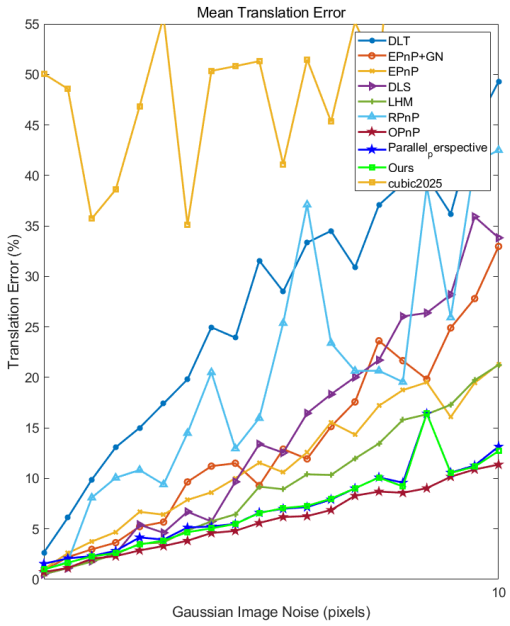
experimental results are presented in **Figure5.4**. The following conclusions could be drawn based on the analysis results.



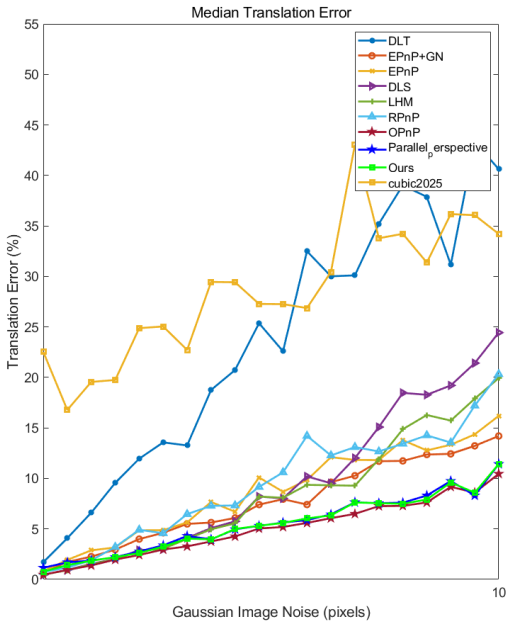
(a)



(b)



(c)



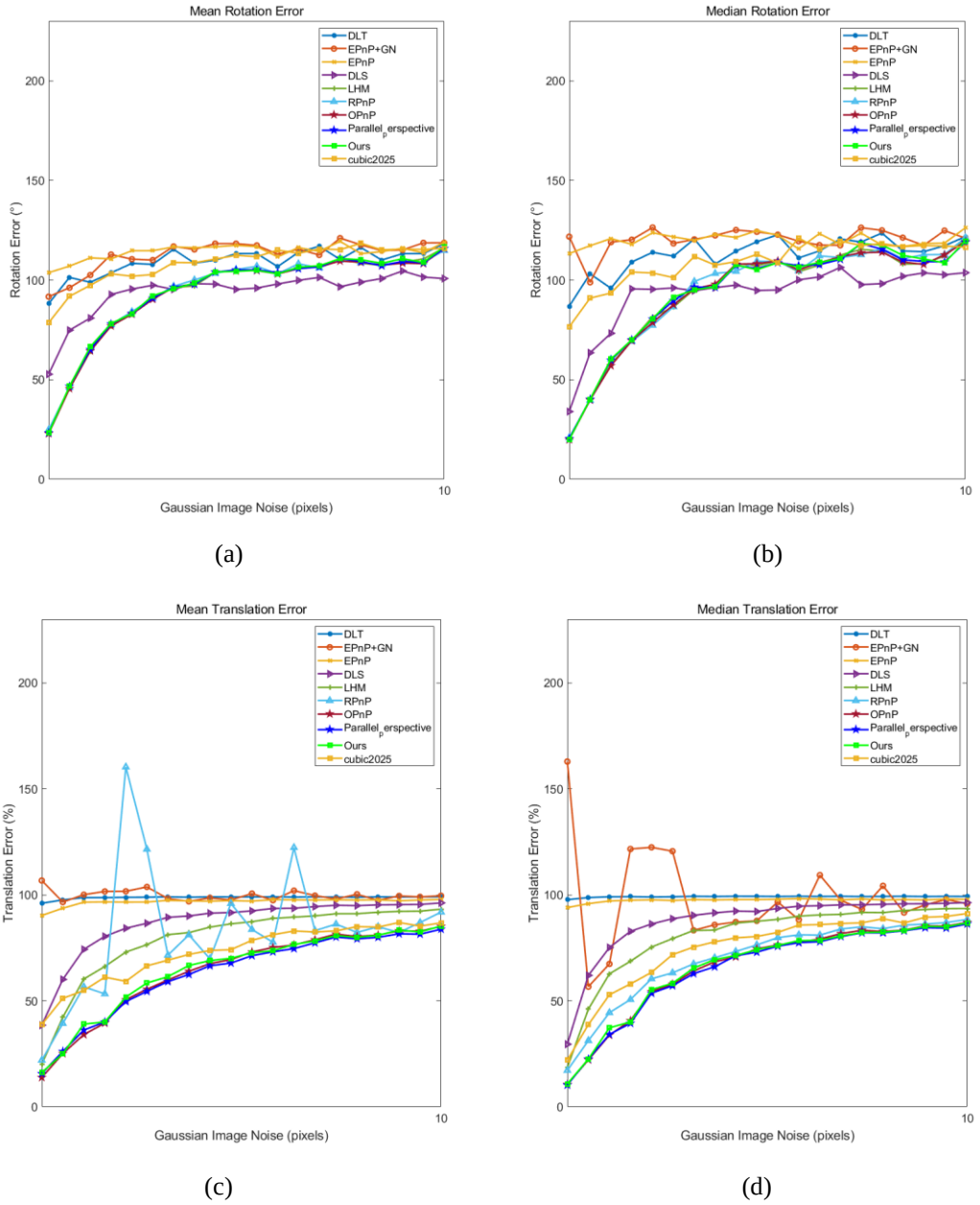
(d)

Figure5.4. Effect of feature point number of 10 and different noise magnitudes on accuracy of position estimation algorithm.

Noise Effect Analysis for Six Reference Points

Based on the experimental results obtained for six reference points as presented in **Figure5.5** and **5.6**, the impact of noise was as follows. With the increase in the Gaussian image noise, both rotational and translational errors exhibited significant increasing trends. Particularly, the measurement error of all evaluated algorithms increased with the noise level. However, the proposed parallel perspective algorithm demonstrated notable robustness, consistently yielding lower rotational and translational errors compared to conventional methods for different noise intensities. Namely, at higher noise levels, the proposed method exhibited a slower error growth rate, highlighting its improved resilience against noise disturbances.

Overall, the experimental results indicated that the proposed parallel perspective-based method effectively enhances the accuracy and stability of single-image pose estimation under high noise conditions, particularly for scenarios with distant reference points.



(a) (b)

(c) (d)

Figure5.5. Effect of feature point number of six and different noise magnitudes on accuracy of position estimation algorithm for edge length of 25.

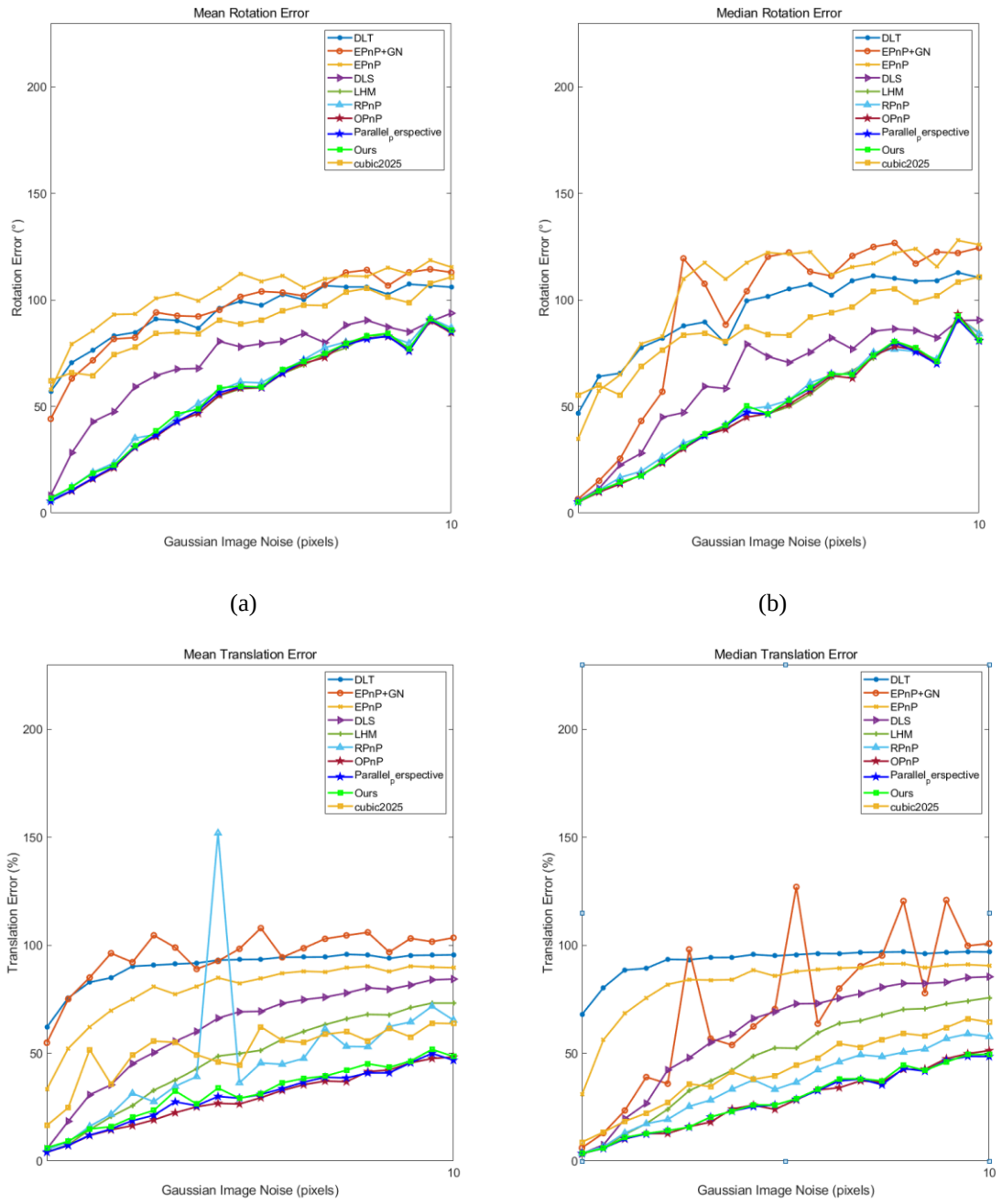
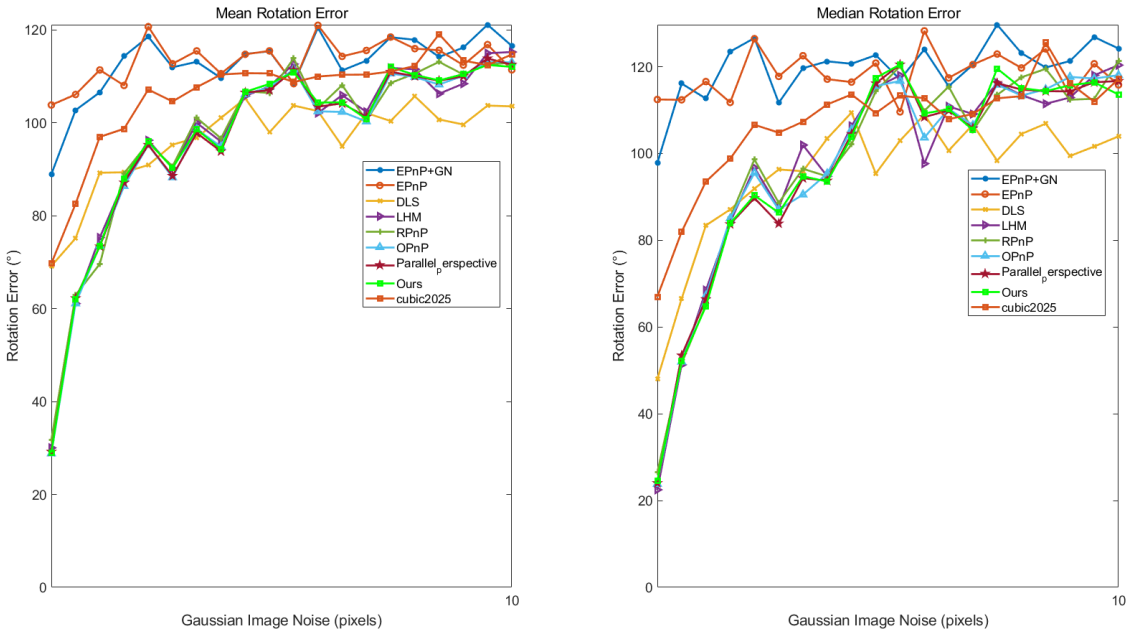


Figure5.6. Effect of feature point number of six and different noise magnitudes on accuracy of position estimation algorithm for edge length of 100.

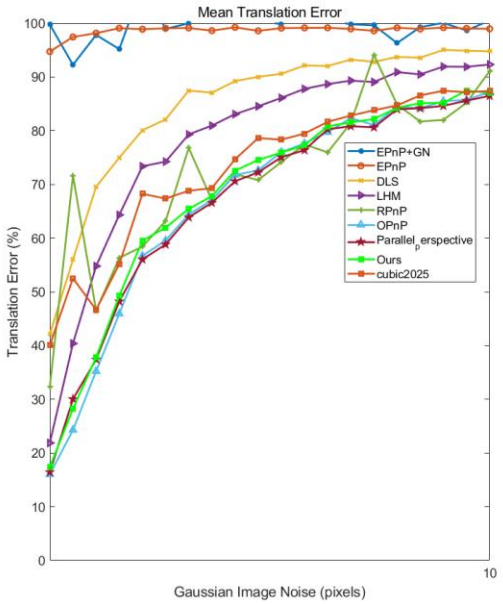
Noise Effect Analysis for Five Reference Points

As illustrated in **Figure 5.7**, the experimental results obtained for five reference points exhibited a significant increase in rotation and translation errors for all tested algorithms with the increase in Gaussian image noise levels. Nevertheless, the proposed parallel perspective methods ("Ours") consistently exhibited superior robustness and accuracy, outperforming conventional methods, such as EPnP+GN, DLS, and OPnP methods, in both rotational and translational error metrics. Specifically, the proposed parallel methods demonstrated noticeably smaller error increments and lower error magnitudes compared to the other methods, which reflected their enhanced capability in handling image noise effectively. Conversely, conventional algorithms exhibited increased sensitivity to noise, causing rapid deterioration of measurement accuracy as the noise level increased. Hence, the results validated that the proposed parallel perspective-based methods could significantly improve noise resilience while maintaining higher reliability and accuracy in single-image position measurement, particularly at intermediate reference point distances.

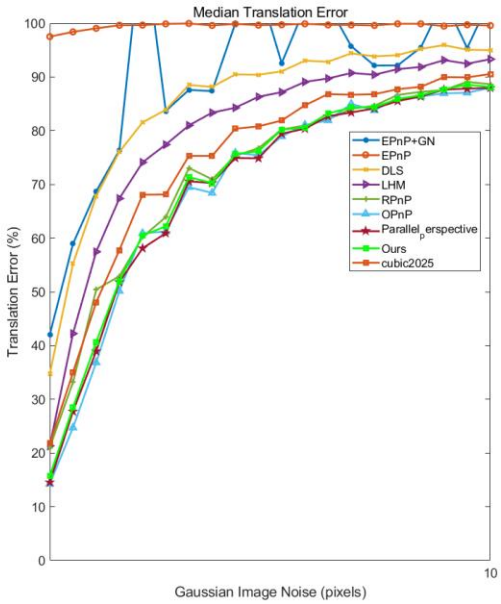


(a)

(b)



(c)



(d)

Figure5.7. Effect of reference point number of five and different noise magnitudes on accuracy of position estimation algorithm.

Noise Effect Analysis for Four Reference Points

As illustrated in **Figure 5.8**, for four reference points (the closest reference point scenario), the experimental results revealed that rotation and translation errors increased progressively across all tested algorithms as the Gaussian image noise intensified. The parallel perspective methods ("Ours") consistently achieved lower rotational and translational errors than the other algorithms, indicated their significantly enhanced robustness against noise compared to conventional algorithms, such as the EPnP+GN, DLS, and OPnP algorithms. At lower noise levels, all algorithms exhibited similar performance; however, as the noise increased, conventional methods rapidly deteriorated, while the proposed parallel methods maintained superior accuracy and stability. These findings demonstrate that, particularly at short distances, parallel perspective algorithms could effectively mitigate the impact of noise, thereby ensuring higher precision and reliability in single-image position measurement tasks.

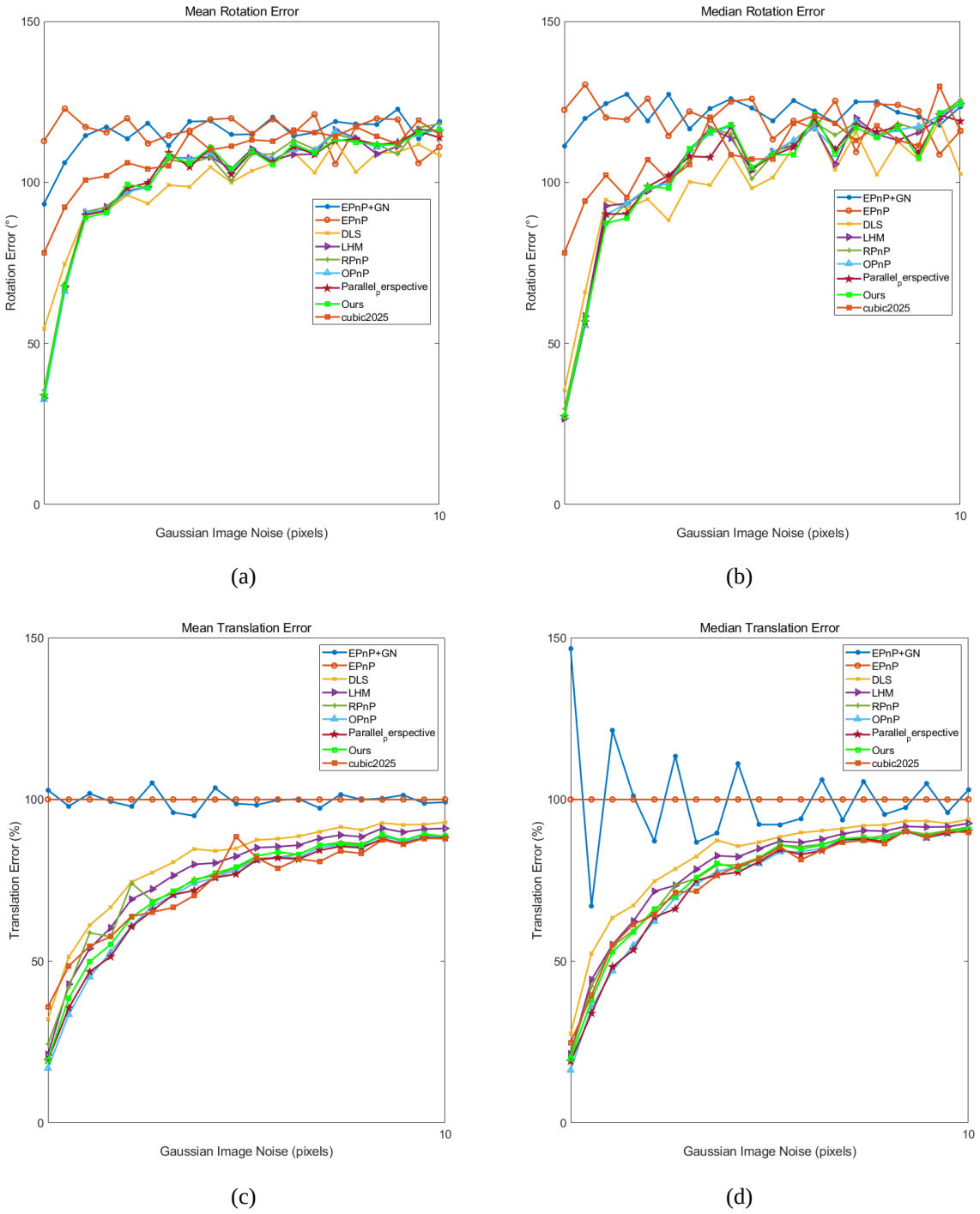


Figure 5.8. Effect of feature point number of four and different noise magnitudes on accuracy of position estimation algorithm.

Four-Point Plane Analysis

The error of both the iterative and analytic solution algorithms increased with the noise level. The distance between the marker points affected the interference effect of noise, and the larger the distance, the higher the accuracy of the algorithm, and the error curve of the algorithm tended to be straight, whereas the curve shifted downward overall. In **Figure 5.9**, the results obtained for 4 points distributed in the 500 × 500 × 1000 mm range are presented. The error curve of each algorithm was a straight line. The maximum noise error was 15.7% degrees and 12% shift, and the minimum noise error was 0.77% degrees and 0.7% shift. The error of the algorithms with higher accuracy, include the OPNP, RPNP, parallel perspective, parallel weight, and LHM algorithms. The error curve of these algorithms (OPNP, RPNP, parallel perspective, parallel weight, and LHM algorithms) was a straight line for feature point distance of 100 mm, with a maximum noise pose error of 84.6% degree and translation error of 55.2%, and a minimum noise error of 5.35% degree and translation of 3.7%. The feature points were curved for each algorithm at a distance of 25 mm, with a maximum of 112% degrees, panning 85.6%, and a minimum of 24%, panning 17.7%. Hence, the size of the target object could effectively improve the algorithm's resistance to noise.

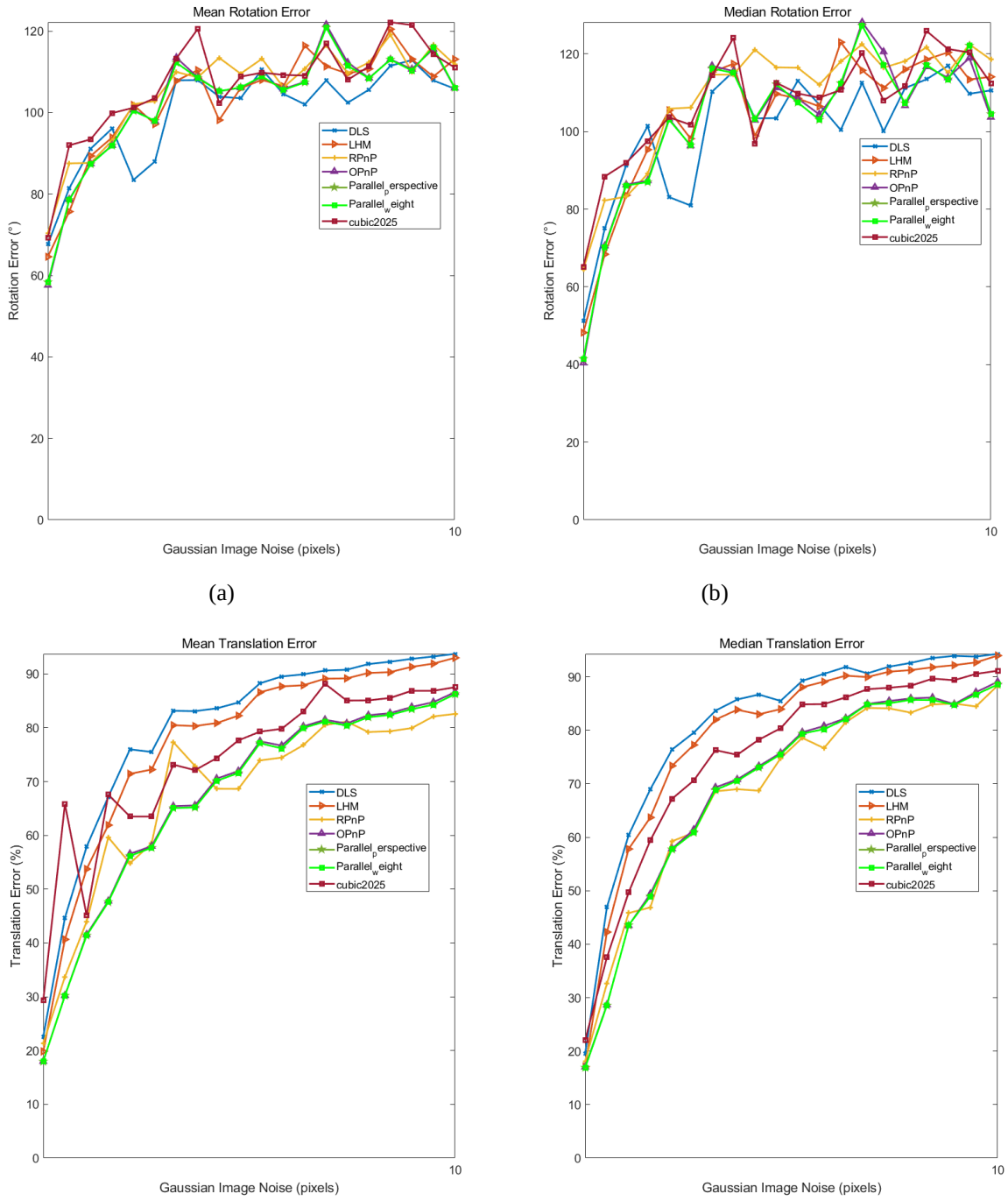


Figure 5.9. Effect of some feature points of four, planar distribution, and different noise magnitudes on accuracy of position estimation algorithm.

The OPNP algorithm is considered the most accurate bit-position estimation algorithm, with good noise resistance, even surpassing that of the typical LHM iterative algorithm (The PnP problem revisited: a fast, general, and optimal solution). Here, the weighted parallel perspective algorithm is equivalent to the OPNP algorithm and the parallel perspective algorithm in terms of its performance against noise interference.

Comparing the noise interference error maps that contained different numbers of feature points, the anti-noise interference ability of the stereo layout with four, five, and six points was demonstrated to be similar. The proposed weighted parallel perspective algorithm achieved the best performance at 10 feature points; however, the distribution space of feature points was large, which limited the application scenarios. For six feature points with a 100-mm spacing, the algorithm demonstrated good anti-noise ability, with a maximum noise attitude error of 84.6% degrees and a translation error of 55.2%, as well as a minimum noise error of 5.35% degrees and translation of 3.7%. For five feature points with a spacing of 100 mm, a maximum noise attitude error was 91.1% degrees, a translation error was 58.3%, a minimum noise error was 7.5% degrees, and a translation was 6.0%. For four feature points with a spacing of 100 mm, maximum noise attitude error of 93.2% degrees, translation error of 60.9%, minimum noise error of 7.7% degrees, and translation of 5.43%. For four planar feature points spaced at 100 mm, the maximum noise attitude error was 93.4% degrees, translation error was 57%, and the minimum noise error was 40% degrees, translation was 5.35%.

Algorithm Operational Efficiency Analysis

This experiment was conducted to verify the algorithm's running efficiency. In the experiment with simulated data, the selection range of random points was ±100 × ± 100 × 200 mm. In the range of 4–104 utilizing Step 4, the number of reference points was changed. The experiment was repeated 500 times utilizing a different number of reference points, and then the average running time of each algorithm was calculated. The experimental results are presented in **Figure5.10**, and based on the analysis results, the following conclusions were drawn.

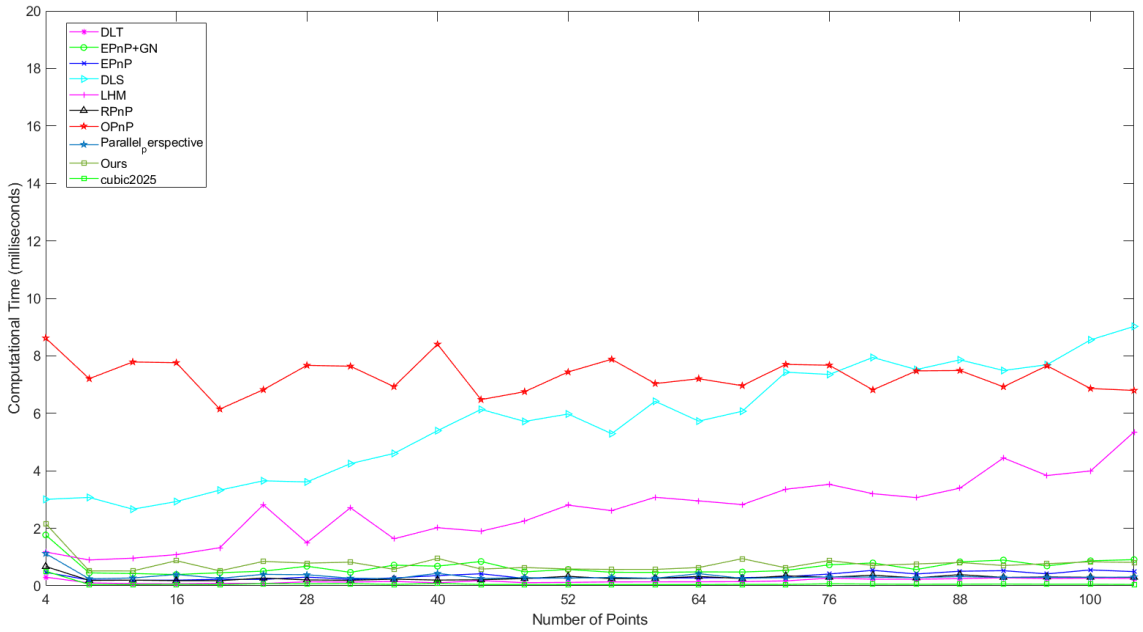


Figure5.10. Algorithm running time results.

The experimental results confirm that different algorithms exhibited significant differences in terms of running time and efficiency. The running times of LHM and DLS algorithms exhibited a nearly linear increase with the number of reference points, whereas the running times of the other parsing algorithms were more stable and had no significant correlation

with the number of reference points. Among the parsing algorithms, the computation time of the OPnP algorithm was significantly higher than that of the other algorithms, approximately five times that of the other parsing algorithms. This could be attributed to its utilization of different optimization methods. The proposed algorithm was similar to other analytic solution algorithms in terms of operational efficiency, and its computation time was slightly higher than that of the other analytic solution algorithms. Further analysis revealed that the computational speed of the parallel perspective projection was in line with the other methods. The weights computation might be the main reason for the slight increase in the proposed algorithm's running time compared to the other algorithms. This indicated that the proposed algorithm had a more optimized running efficiency than the comparison algorithms while maintaining high accuracy.

Camera Depth

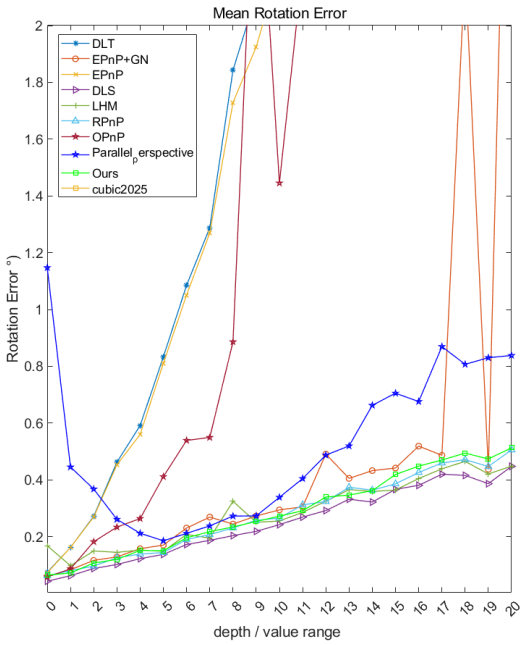
This experiment was conducted to verify the effect of the camera depth on the algorithm's performance. In the experiment with simulated data, the distance between the camera focal point and the feature point selection range was defined as a camera depth; the number of fixed feature points was set to 10, 6, 5, and 4, sequentially; the camera focal length was 500; the white noise with a variance of 0.2 was added. The number of cycles was set to 100 for each algorithm to verify its position estimation accuracy by varying the camera depth in both the area close to the focal point (distance from the focal point was 0–500 mm, with a step size of 100 mm) and area away from the focus area (0–10,000 mm from the focus,

with a step of 500 mm). The camera depth was changed to verify the accuracy of the position estimation of each algorithm.

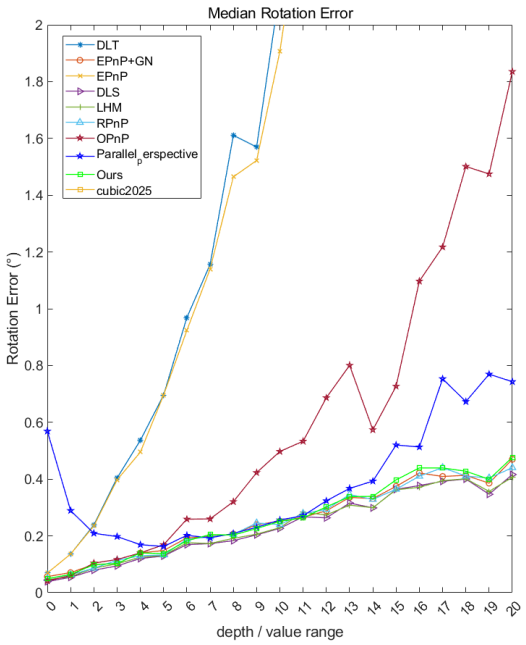
Via the analysis of the experimental results, the following conclusions were drawn:

- As the ratio of the camera depth to the reference point range increased, the error of the algorithms also increased, demonstrating a linear trend. The direct linear transformation (DLT) and EPnP algorithms were most affected by the camera depth–reference point range ratio among all the algorithms.
- The 10 feature points were randomly selected from a large selection range and a large number of points; thus, the position error had a relatively good linear characteristic. The proposed algorithms had the same performance as the DLS, RPNP, and LHM algorithms, achieving good accuracy and an attitude accuracy in the range of 0.5°; also, the displacement accuracy was in the range of 0.4%.
- When the number of feature points was smaller than seven, the layout of feature points was fixed, and their relative distances were short; thus, the depth and feature point target size ratio was large, reaching 200. The error curves of many algorithms, such as the DLT (only in experiments with six feature points), EPnP, and EPnP+GN algorithms, exhibited a bilinear characteristic, with an inflection point at the position corresponding to the camera depth–feature point target size ratio of approximately 40. The large error variation in the first stage demonstrated the robust performance of the algorithm in the proximity stage. The proposed algorithm, as well as the parallel perspective LHM, RPNP, and OPNP algorithms, had good immunity to depth change, and the error curve exhibited a flat increase.

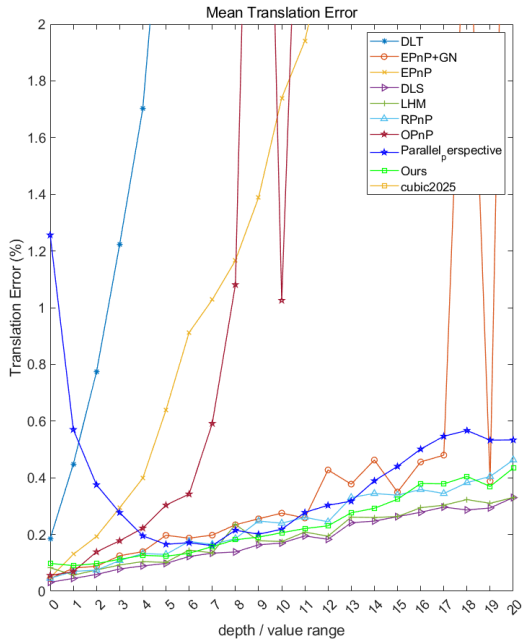
- The initial value of the feature point of the proposed algorithm was set utilizing the parallel perspective algorithm, which was superior to the original algorithm. The proposed algorithm represented an approximation algorithm of perspective projection, especially in the close distance and the origin away from the optical axis conditions. The initial value had a large error. As illustrated in **Figure5.11**, the error curve of the parallel perspective projection algorithm included a region of a large error for the camera depth–feature point target size ratio of less than three. The weighted optimization strategy employed in the localization effectively eliminated this error region and improved the algorithm's accuracy. This confirmed that the proposed algorithm had good robustness in handling error accumulation caused by depth changes, making it suitable for a wider range of application scenarios.



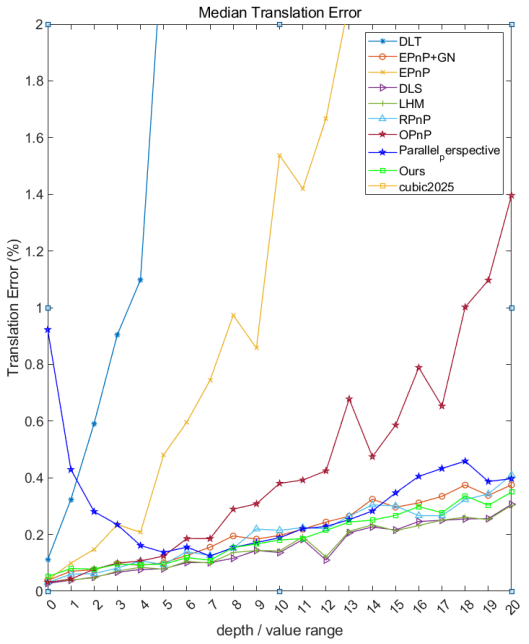
(a)



(b)



(c)



(d)

Figure5.11. Impact of camera depth–reference point value range on algorithm's accuracy for reference point number of 10.

Camera Depth Change Effect Analysis for 10 Reference Points

The experimental results regarding the camera depth's effect obtained utilizing 10 reference points demonstrated that as the depth-to-value ratio increased, both rotational and translational errors generally increased for all algorithms evaluated. However, conventional methods, such as the EPnP+GN and DLS algorithms, exhibited significant sensitivity to depth variations, resulting in rapid increases in measurement error, particularly at larger depths. In contrast, the proposed parallel perspective-based algorithms consistently had lower and more stable errors throughout the entire tested depth range. The proposed methods demonstrated superior robustness, as indicated by their minimal error magnitudes and slower growth rates under conditions of increasing camera depth. This

further highlighted their effectiveness and stability, particularly for challenging scenarios involving longer camera–object distances.

Effect of Camera Depth Change for 10 Reference Points Close to Camera Focal Point

As illustrated in **Figure 5.12**, the experimental results demonstrate that when the 10 reference points were positioned close to the camera focal point, rotational and translational errors of all evaluated algorithms remained relatively low and stable, and the algorithms exhibited minimal sensitivity to variations in the depth-to-value ratio. Notably, conventional algorithms, such as the DLT algorithm, still exhibited comparatively higher error magnitudes and fluctuations than the proposed algorithms, particularly in translation estimation. Conversely, the proposed parallel perspective-based methods (Parallel_epnpactive and Parallel_pnp11) consistently yielded the lowest and most stable errors for both rotation and translation, which underlined their superior accuracy and robustness. The results indicated that the proposed parallel perspective approaches could effectively manage depth variations even when reference points were near the camera focal point, providing highly reliable and accurate pose estimation under such conditions.

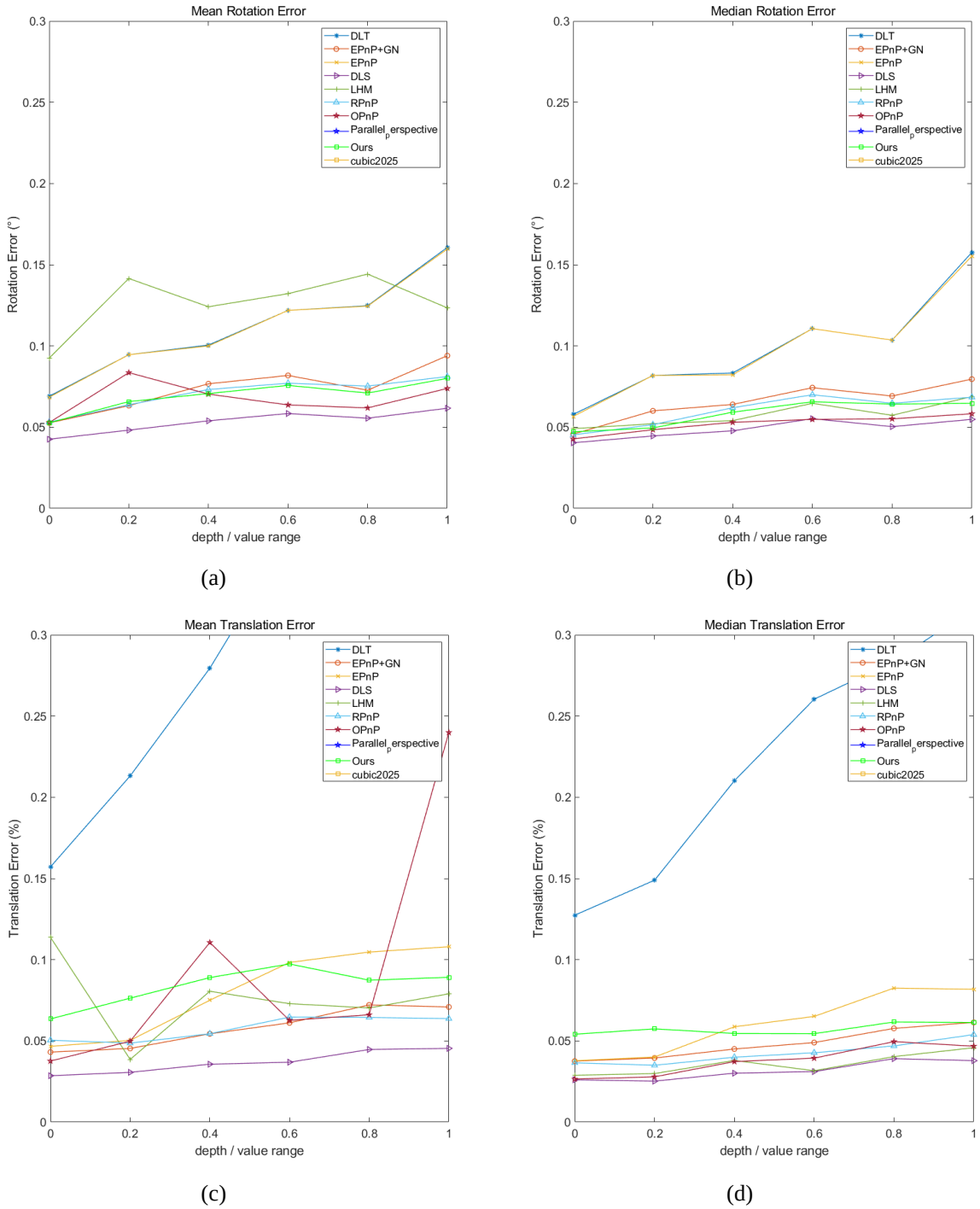


Figure5.12. Impact of camera depth–reference point number ratio on algorithm's performance; value range was between zero and one, and reference point number was set to 10.

Depth Impact Analysis for Six Reference Points

The experimental results, obtained with six reference points, clearly indicated that both rotational and translational errors of all tested algorithms significantly increased with the depth-to-length ratio, as illustrated in **Figure5.13**. Conventional algorithms, such as the DLT, EPnP+GN, and DLS algorithms, exhibited sharply increasing errors, which were particularly evident at greater depths; this highlighted their sensitivity to depth changes. In contrast, the proposed parallel perspective methods ("Ours") maintained consistently low and stable rotational and translational errors throughout the tested depth range. The proposed parallel approaches outperformed conventional methods, demonstrating superior robustness and resilience against depth-induced measurement deterioration.

Consequently, the results strongly affirmed that the proposed parallel perspective-based algorithms could provide considerable advantages in single-image pose estimation for varying depths and in challenging scenarios with significant camera–object distances.

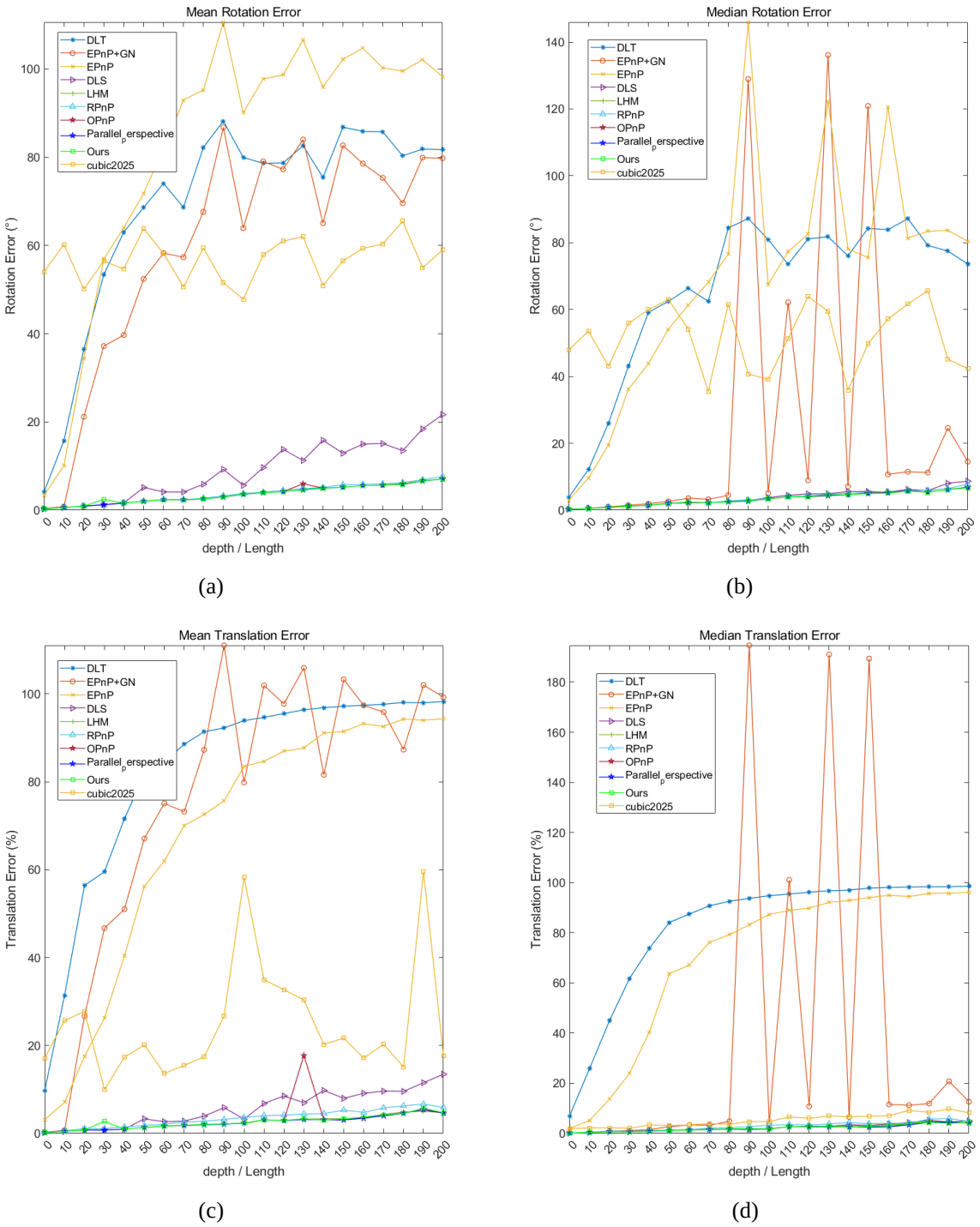


Figure5.13. Depth impact analysis results for six reference points.

Minor Depth Impact Analysis for Six Reference Points

The experimental analysis (as illustrated in **Figure5.14**) conducted to examine the minor depth impact utilizing six reference points indicated that when the reference points were placed relatively close to the camera, rotational and translational errors of all the methods remained generally low, exhibiting limited increases with the depth-to-length ratio. The proposed parallel perspective algorithms ("Ours") consistently demonstrated minimal error values and excellent stability, maintaining superior performance over conventional methods. Although conventional methods, such as the DLT and EPnP+GN, had occasional spikes and higher sensitivity to slight depth variations, the parallel methods could effectively suppress such variations, achieving robust and reliable pose estimations under conditions of minor depth changes. The results indicated that the proposed parallel perspective-based methods had strong resilience and accuracy advantages, which could be especially beneficial in close-range scenarios with subtle depth fluctuations.

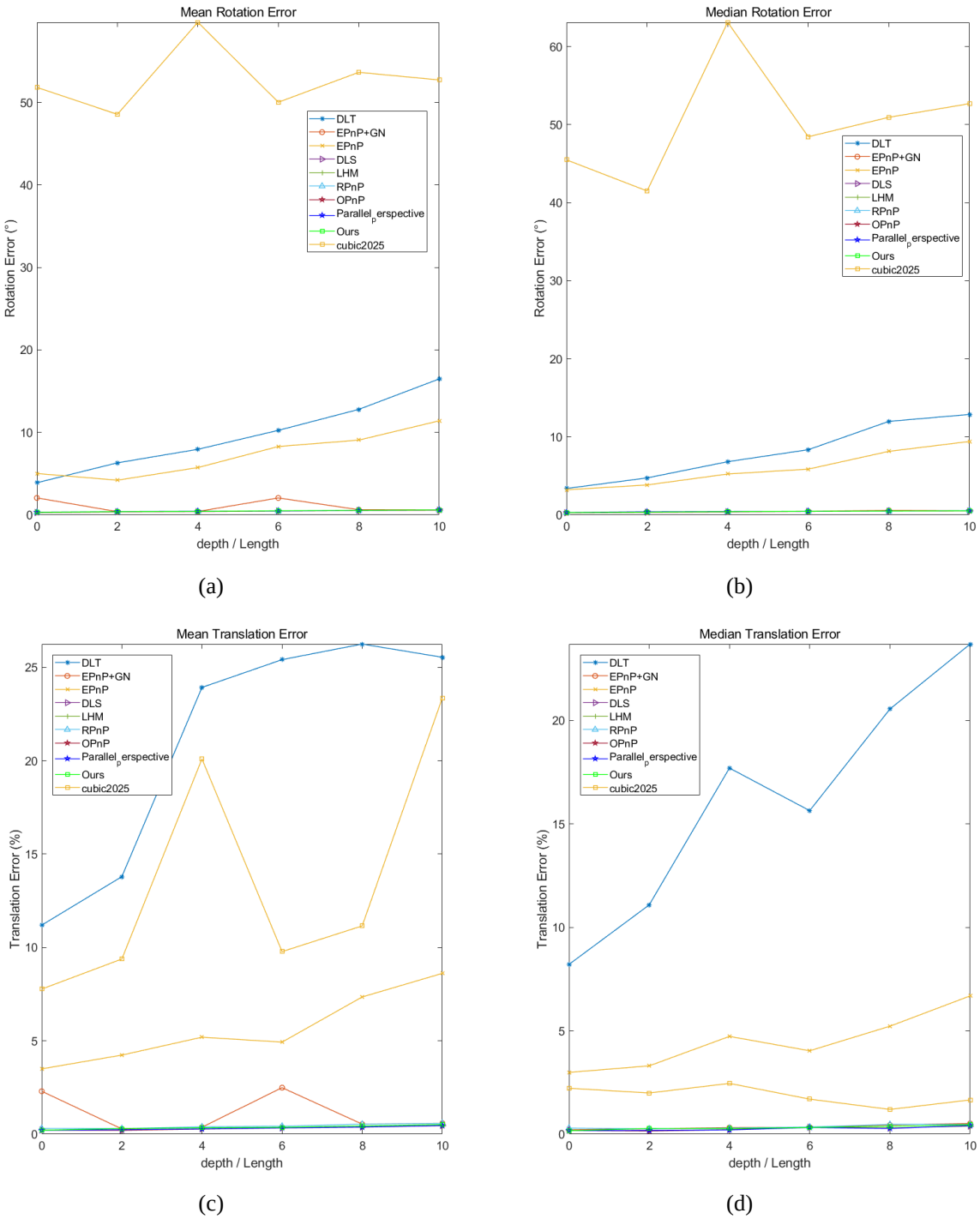


Figure5.14. Minors depth impact analysis results for six reference points.

Depth Impact Analysis for Five Reference Points

The experimental results of the analysis conducted to examine the depth impact utilizing five reference points clearly illustrated that rotational and translational errors significantly increased with the depth-to-length ratio for all the algorithms, as illustrated in **Figure5.15**. Conventional methods, particularly the EPnP+GN, DLT, and DLS methods, exhibited substantial instability and rapid error escalation as the camera depth increased, demonstrating significant sensitivity to the changes in camera depth. In contrast, the proposed parallel perspective algorithms ("Ours") consistently maintained lower error magnitudes and exhibited more stable growth trends throughout the tested depth ranges. Such a consistent performance advantage highlighted the robustness and effectiveness of the proposed parallel perspective methods, confirming their superior suitability for pose estimation involving larger camera–object depth variations.

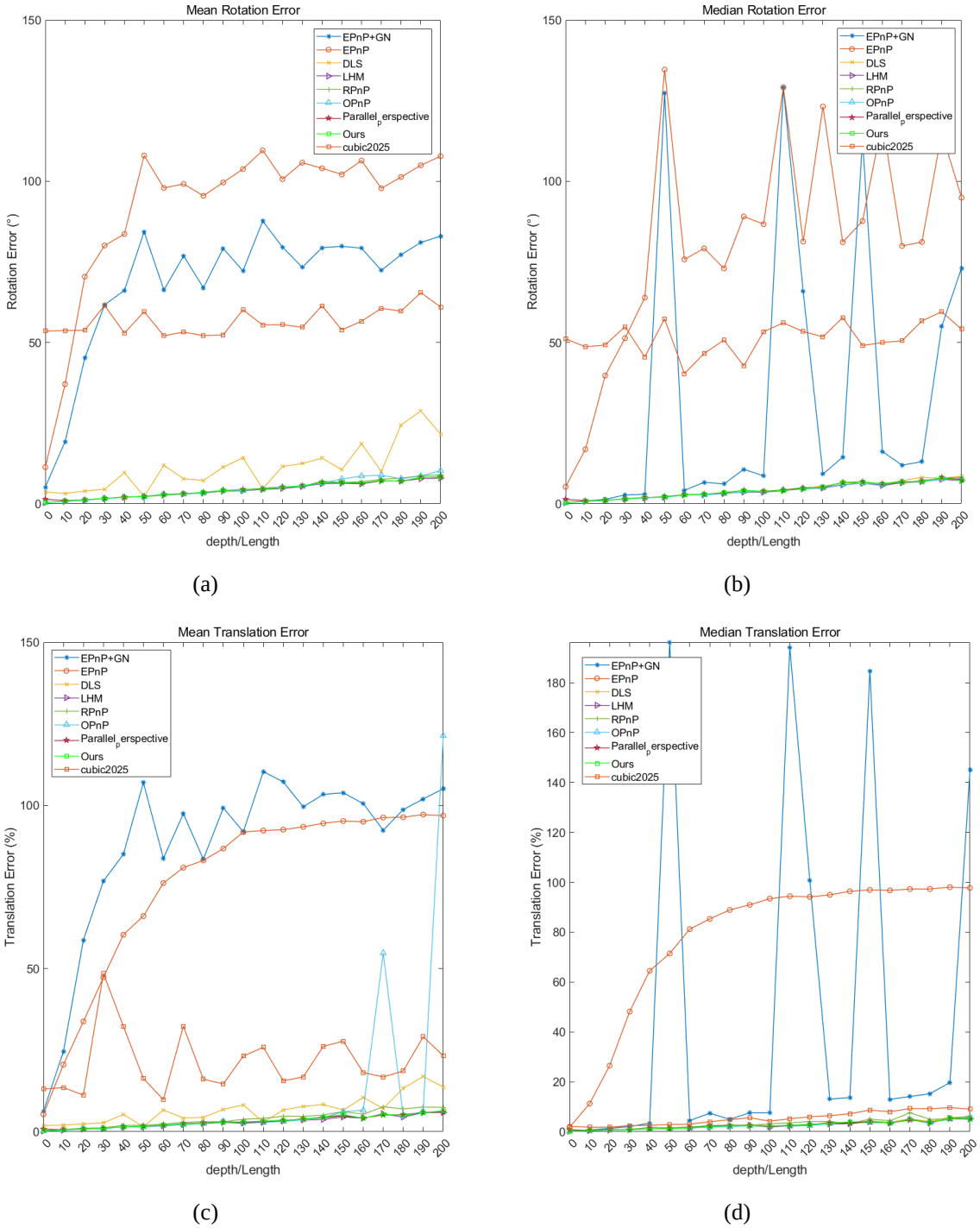


Figure5.15. Depth impact analysis results for five reference points.

Small Point Distance Impact Analysis for Five Reference Points

As illustrated in **Figure 5.16**, the experimental results of the analysis conducted to evaluate the impact of small point distance utilizing five reference points revealed distinct performance differences between the tested methods. At shorter point distances, rotational and translational errors remained generally low for most algorithms. However, conventional approaches, such as the EPnP, EPnP+GN, and particularly DLT methods, had significantly higher and rapidly increasing errors as the depth-to-length ratio increased. In contrast, the proposed parallel perspective algorithms ("Ours") consistently exhibited minimal rotational and translational errors, indicating their remarkable stability and robustness even at very short point distances. The results highlighted the superior capability of the proposed parallel perspective-based methods in achieving reliable and precise pose estimation, which could be particularly advantageous under short-range imaging conditions with limited depth variations.

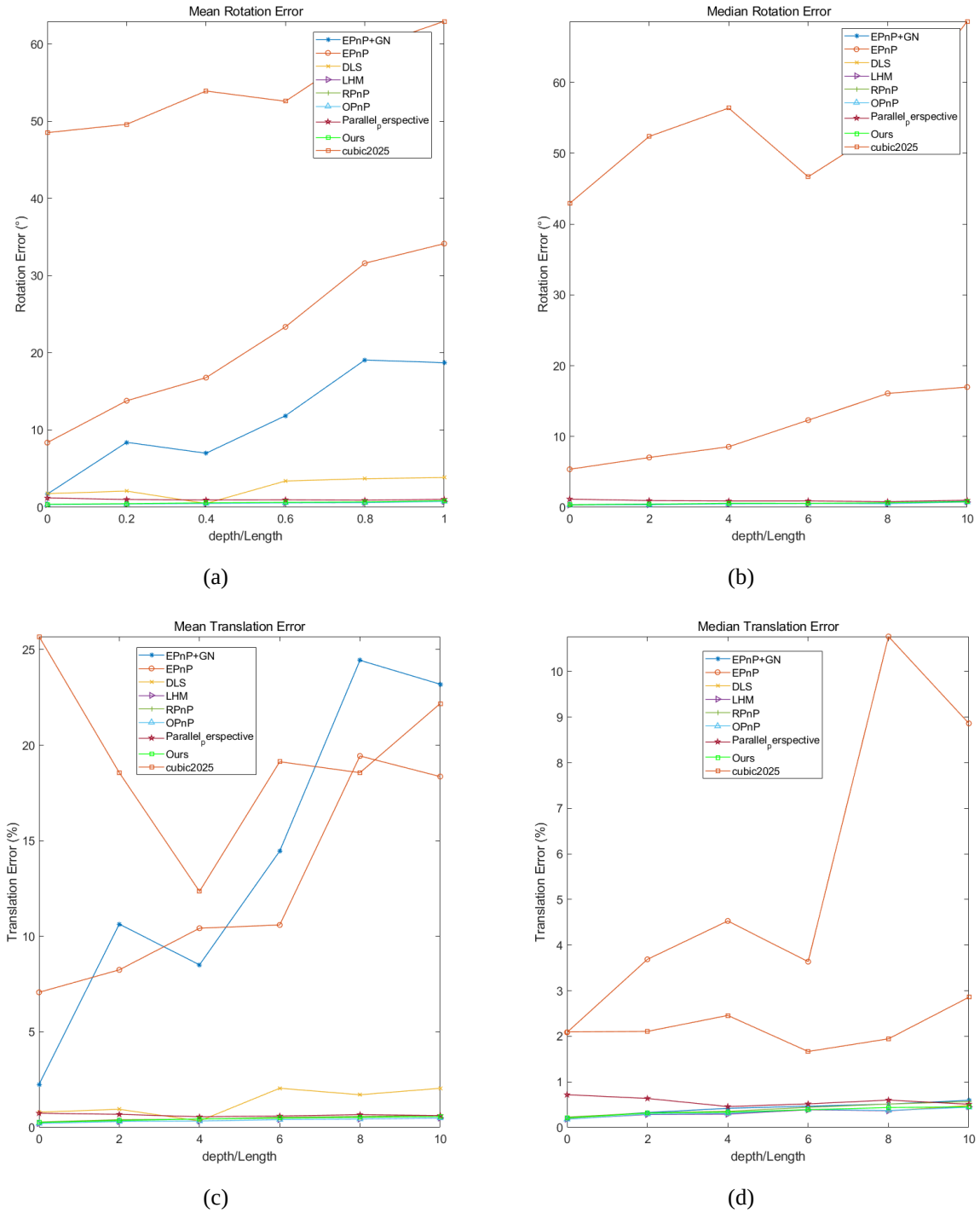
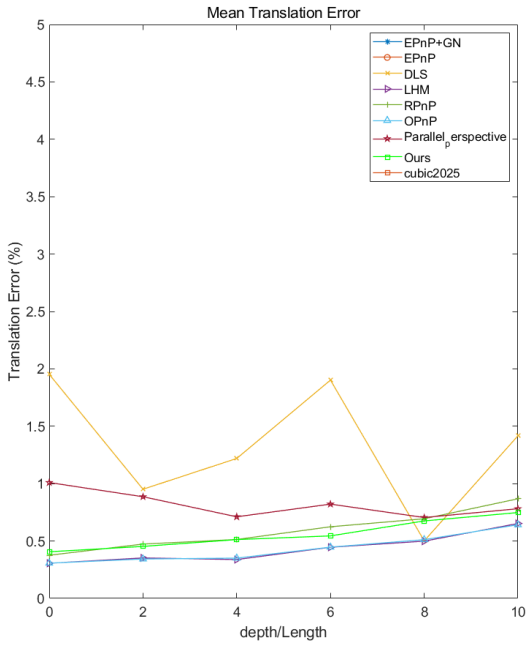


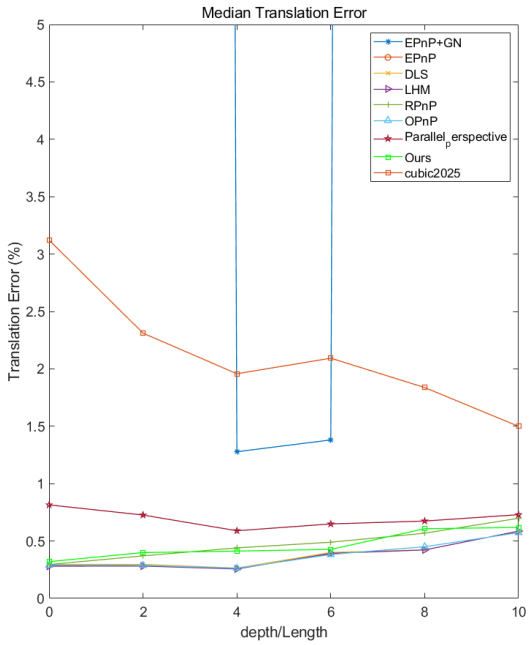
Figure5.16. Impact analysis results of camera depth–reference point number ratio on algorithm's accuracy for five reference points.

Minor Depth Impact Analysis for Four Reference Points

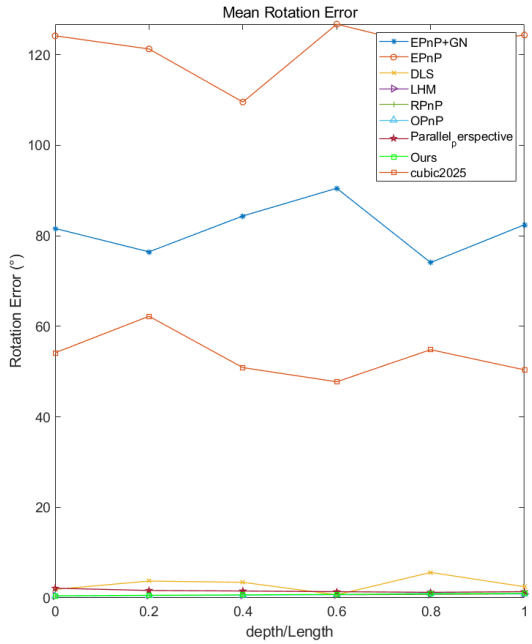
As illustrated in **Figure 5.17**, the experimental results of these experiments indicated that the rotational and translational errors of most evaluated methods remained relatively low and stable across varying camera depths. For instance, conventional algorithms, such as the DLT, EPnP+GN, and DLS algorithms, occasionally exhibited noticeable fluctuations, particularly in translational error measurements, indicating high sensitivity even at small camera–object distances. In contrast, the proposed parallel perspective-based methods ("Ours") consistently demonstrated lower and significantly more stable error performance; thus, effectively minimizing both rotational and translational estimation errors. The proposed algorithms also maintained good robustness even under subtle depth variations, which highlighted their suitability and accuracy advantages in single-image pose estimation, particularly in scenarios characterized by short distances between the camera and reference points.


Mean Translation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Translation Error (%)
depth/Length

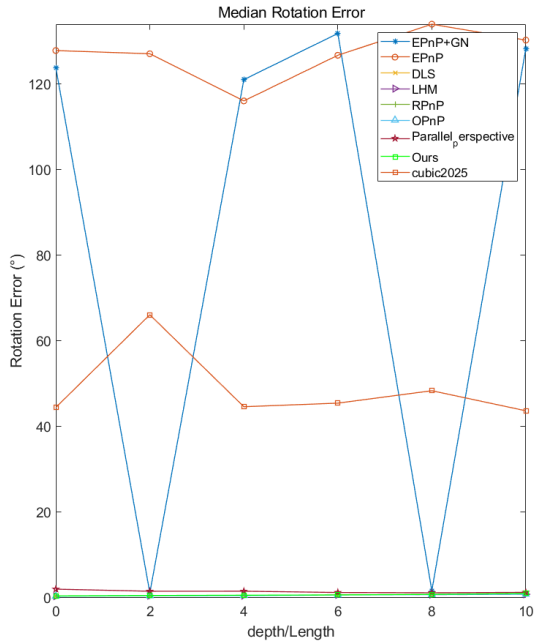
(a)


Median Translation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Translation Error (%)
depth/Length

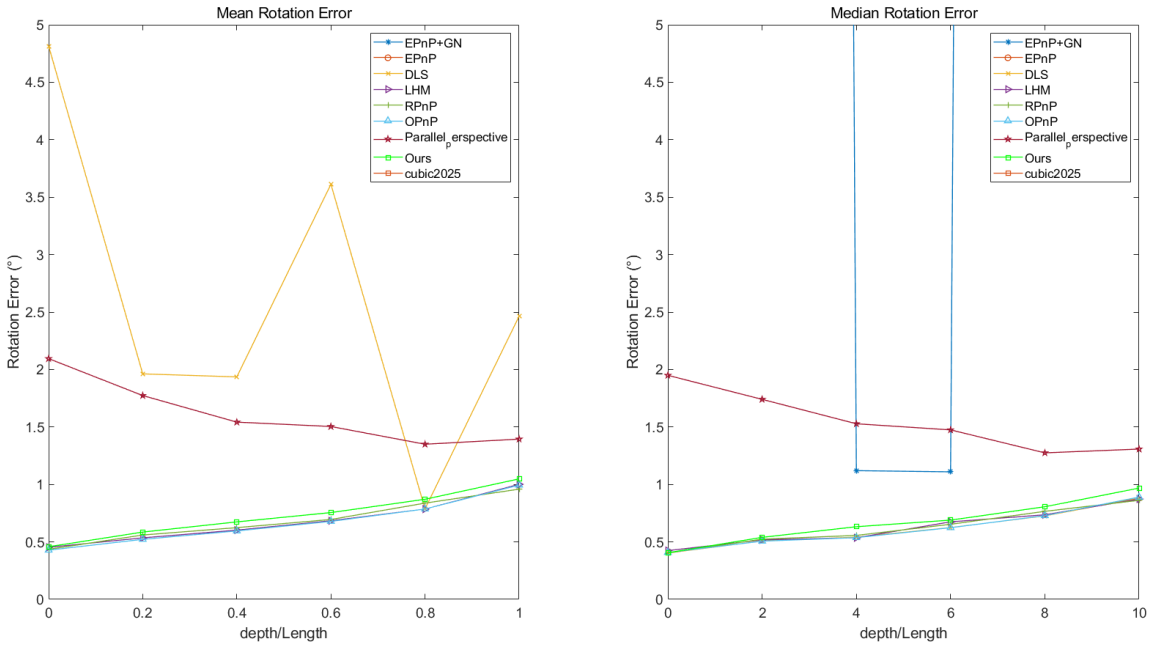
(b)


Mean Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
depth/Length

(c)

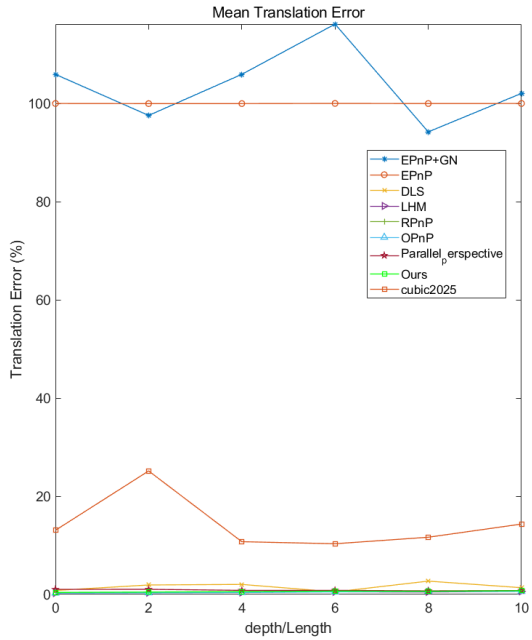

Median Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
depth/Length

(d)

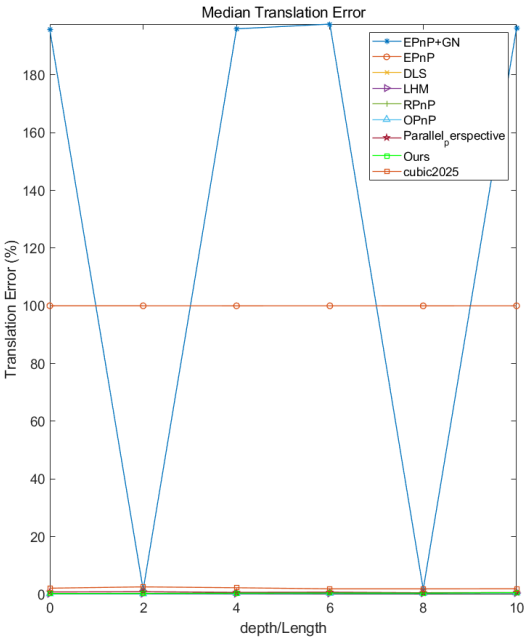


(e)

(f)



(g)



(h)

Figure5.17. Impact analysis results of camera depth–reference point number ratio on algorithm's performance for five reference points when value range was between zero and one.

Depth Impact Analysis for Four Reference Points

The experimental results of the depth impact analysis (as illustrated in **Figure5.18**) conducted utilizing four reference points clearly illustrated significant differences in robustness between the evaluated methods as the depth-to-length ratio increased. Conventional algorithms, including the EPnP+GN, DLT, and DLS methods, exhibited marked sensitivity to depth variations, resulting in a drastic increase and instability of both rotation and translation errors, particularly at higher depth values. In contrast, the proposed parallel perspective algorithms ("Ours") consistently maintained notably lower errors, demonstrating remarkable stability and robustness across various depths. The proposed parallel methods can effectively suppress depth-induced inaccuracies, underscoring their significant advantages in precision and reliability in single-image pose estimation under challenging depth conditions.

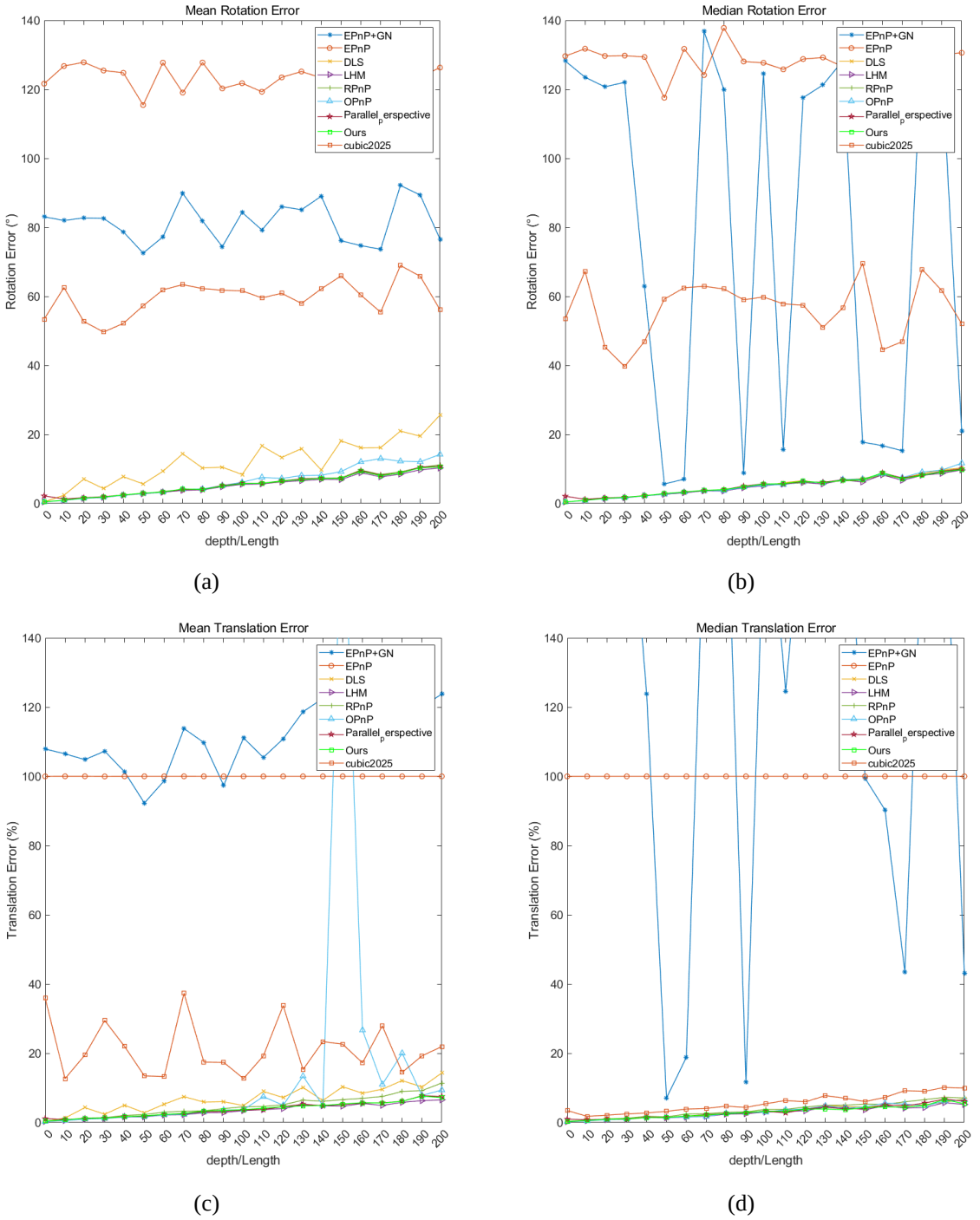


Figure5.18. Impact analysis results of camera depth–reference point ratio on algorithm's performance for five reference points when value range was between zero and one.

Depth Effect Analysis of Four-Point Planes

The experimental results which reveal the depth effects of four-point planes (as presented in **Figure5.19**) indicated a clear increasing trend of rotational and translational errors with the depth-to-length ratio for all algorithms. Conventional methods, including the OPnP and DLS methods, exhibited pronounced fluctuations and rapid increases in measurement errors at higher depth ratios, underscoring their high sensitivity to depth variations. Conversely, the proposed parallel perspective methods ("Ours") demonstrated consistently lower and more stable error performances, significantly outperforming conventional approaches, particularly at larger depth-to-length ratios. The results confirmed the effectiveness and robustness of the proposed parallel perspective algorithms, highlighting their clear advantages in handling challenging imaging conditions involving larger camera-to-object depth variations under a limited number of reference points.

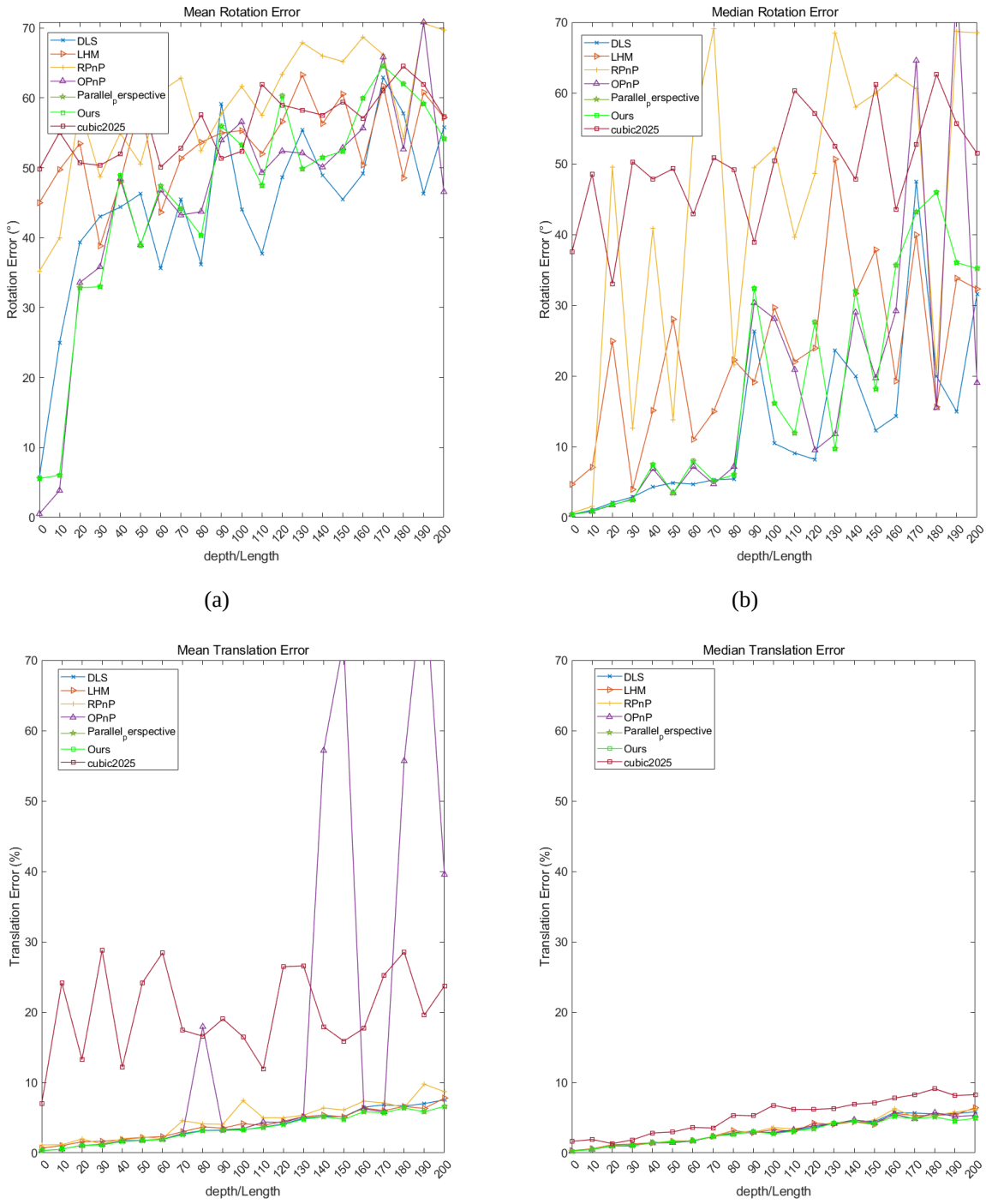


Figure5.19. Depth effect analysis results of 4-point planes.

Small Depth Effect Analysis of Four-Point Plane

The experimental results of the small depth effects analysis of four-point planes (as illustrated in **Figure5.20**) indicated that when reference points were located at relatively short distances, rotational and translational errors of most evaluated methods remained consistently low and stable, exhibiting only minor variations with the depth-to-length ratio. However, conventional methods, such as the DLT and DLS methods, exhibited noticeably higher rotational errors and occasional fluctuations, which highlighted their relatively high sensitivity to even slight depth changes. In comparison, the proposed parallel perspective-based methods ("Ours") consistently achieved lower error magnitudes and exhibited remarkable stability in both rotation and translation estimations. This indicated the superior robustness and precision of the proposed parallel perspective-based methods, which could effectively handle slight variations in depth at close-range imaging conditions.

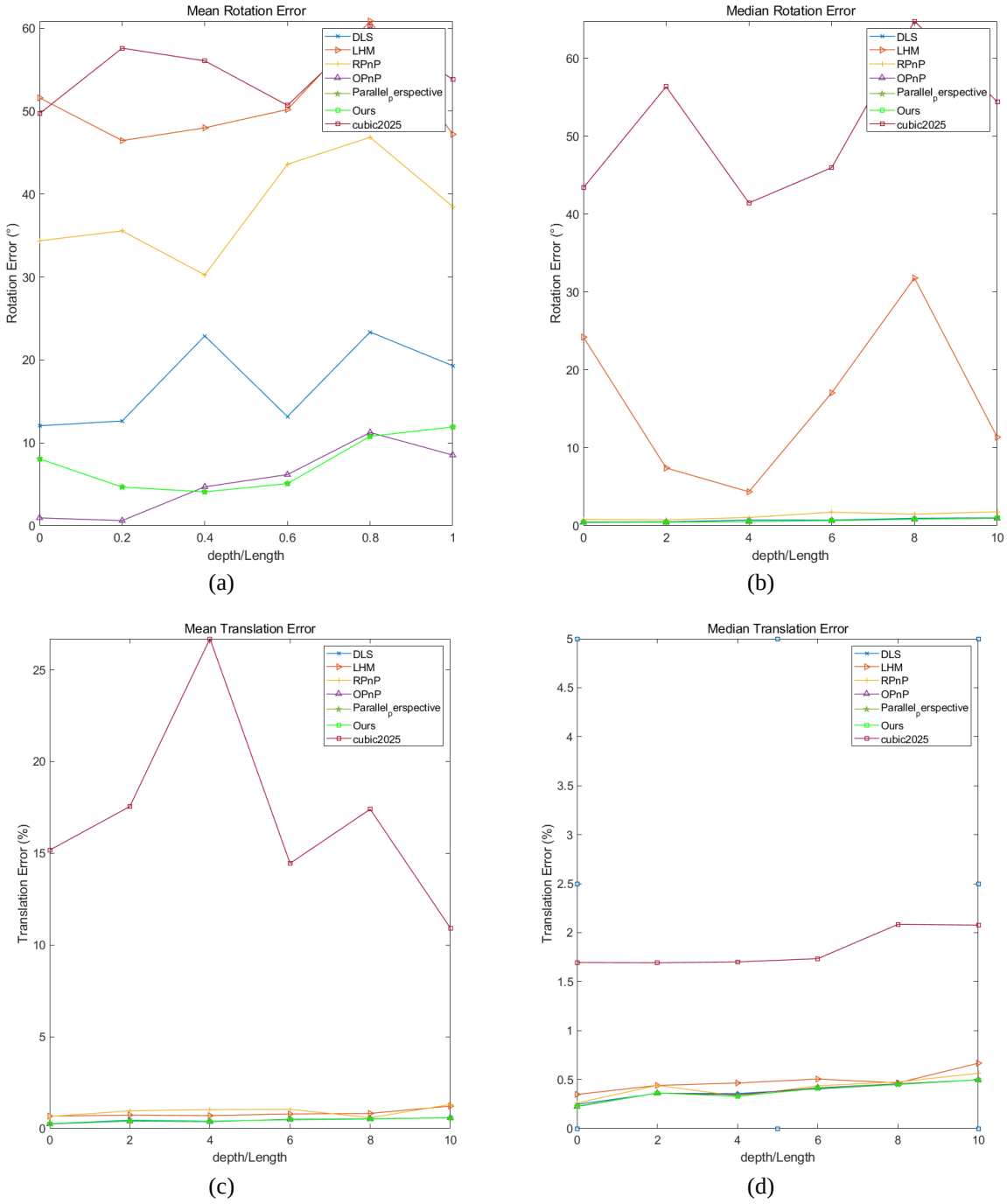


Figure5.20. Small depth effect analysis results of four-point planes.

Camera Eccentricity Distance

This experiment aimed to verify the effect of the distance of the feature point coordinate system's origin from a camera's optical axis on the algorithm's performance. In the experiment with simulated data, the number of feature points was fixed and set to stereo six, stereo five, stereo four, and planar four, sequentially. The focal length of a camera was 80; the white noise with a variance of 0.2 was added to data; the number of cycles was set to be 100; the X-axis coordinate of the feature point coordinate system's origin was fixed at 0; the Z-coordinate was fixed at 200 for an area away from the focus and 100 for an area near the focus; the Y-coordinate parameter of the origin varied in the 1500 mm range, with a step size of 25 mm close to the focal point of 100. The origin's Y-coordinate was in the range of 1–500 mm, changing with a step size of 25 mm, simulating a different distance from the feature point coordinate system's origin to the camera's optical axis to verify the accuracy of the algorithms. The experimental results are presented in the figure.

Based on the results analysis, the following conclusions were drawn.

- The distance from the feature point coordinate system's origin to the camera's optical axis affected the accuracy of the feature point's initial position in different algorithms; thus affecting the accuracy of attitude estimation in different algorithms.
- The DLT algorithm, which was adapted to stereo six points only, and the EPnP algorithms were most affected by the distance of the optical axis, exhibiting large errors in position estimation. However, the other algorithms, including the

proposed algorithm, were less affected by this distance, having attitude errors from 0.2° to 0.8° and displacement errors from 0.5% to 0.8% for the six feature points.

- The initial values of the feature points of the proposed algorithm were calculated by the parallel perspective model. However, the eccentricity distance, the feature point origin deviation from the optical center distance, and the proximity distance to the focal point could affect the initial value accuracy. As illustrated in the proximity error map of the stereo layout feature points, the proposed weighted optimization strategy could improve the parallel perspective estimation accuracy. The attitude error at a depth ratio of 9.52 was reduced from 2.2° to 0.62°, which denoted a 3.5-fold improvement; similarly, the movement error was reduced from 0.67% to 0.47%, depicting a 1.42-fold improvement.
- Under the condition of feature point plane layout, the performances of the proposed algorithm and the parallel perspective algorithm were consistent with a large error. In addition, the positional error increased with the distance of the coordinate system's origin from the camera's optical axis, and the advantage of the weighted algorithm strategy was not reflected.

Six-Point Eccentricity Analysis

The experimental results of the six-point eccentricity scenario clearly demonstrated that rotational and translational errors of conventional methods, such as the DLT and DLS methods, could significantly increase with the distances from the camera's optical axis, as illustrated in **Figure5.21**. These methods demonstrated substantial error fluctuations and sensitivity to eccentricity variations, which highlighted their limitations in handling

points positioned further from the optical center. Conversely, the proposed parallel perspective algorithms (“Ours”) consistently maintained very low rotational and translational errors, demonstrating excellent stability and robustness across the entire eccentricity range utilized in the experimental tests. This underscores the significant advantages and superior performance of the proposed parallel perspective-based approaches in accurately managing pose estimation even when eccentric distances increase.

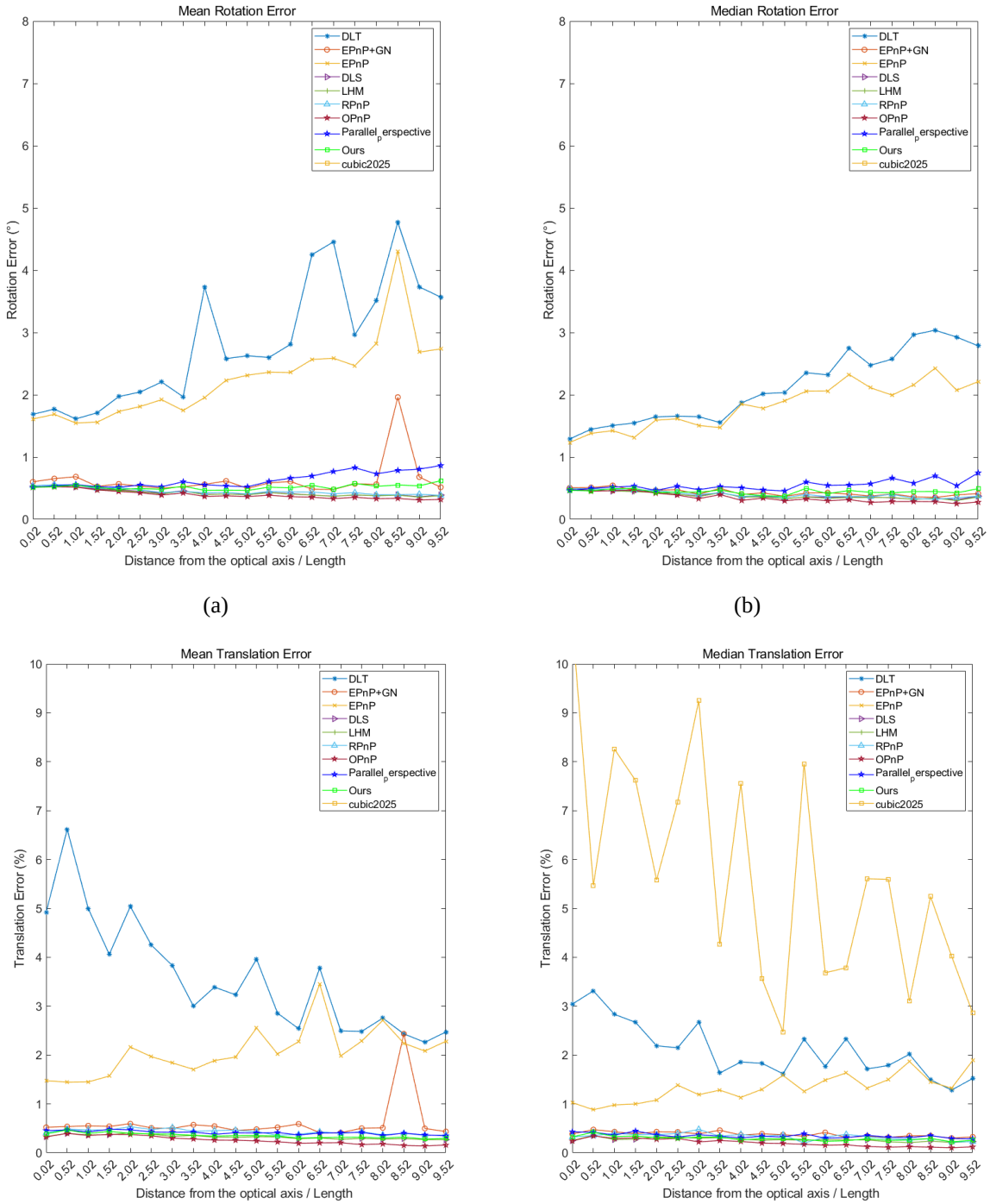
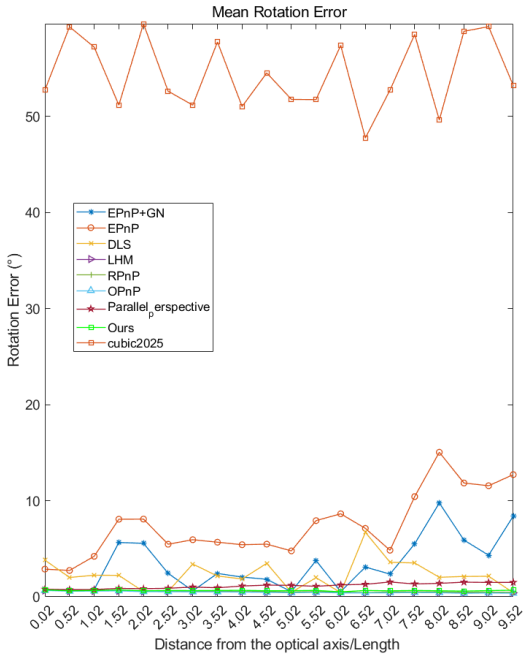


Figure5.21. Six-point eccentricity scenario results.

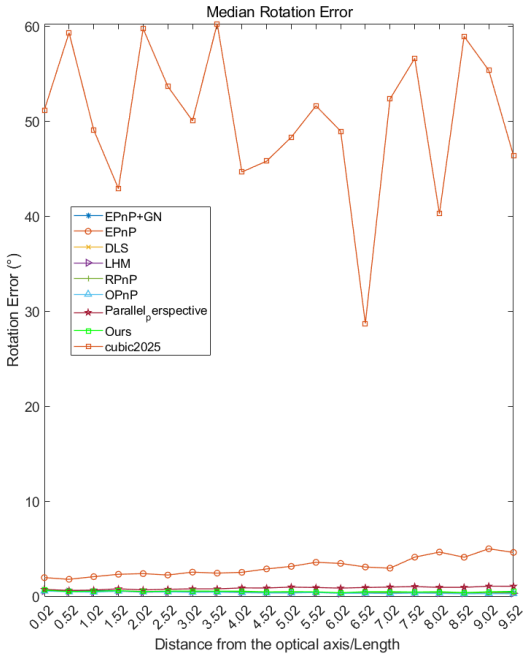
Five-Point Eccentricity Scenario Analysis

The experimental results obtained in the five-point eccentricity scenario clearly demonstrated that rotational and translational errors generally increased with the distance of reference points from the optical axis, as illustrated in **Figure 5.22.** In this experiment, conventional algorithms, including the EPnP, EPnP+GN, and DLT methods, exhibited substantial sensitivity and erratic fluctuations in both rotational and translational error measurements at larger eccentricity values. In contrast, the proposed parallel perspective-based methods ("Ours") consistently maintained low and stable errors for all eccentricity values tested. The proposed algorithms significantly outperformed conventional methods, indicating their robustness and stable performance under varying eccentricities.

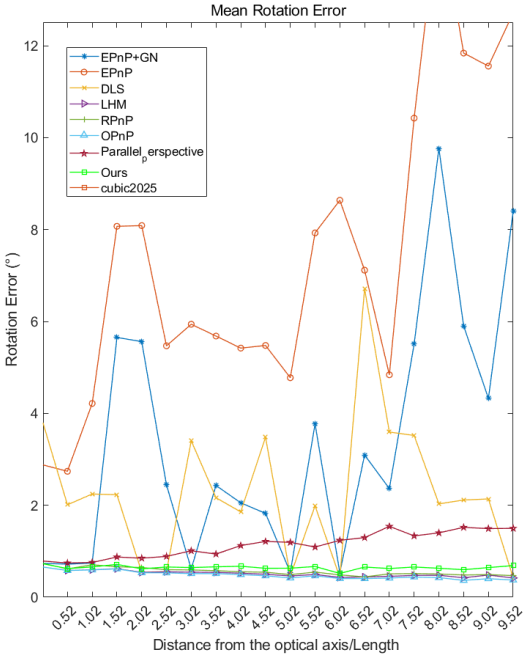
Consequently, the proposed parallel perspective algorithms are particularly advantageous for accurate and reliable single-image pose estimation, even in challenging scenarios involving widely spaced eccentric reference points.


Mean Rotation Error
Rotation Error (°)
Distance from the optical axis/Length
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025

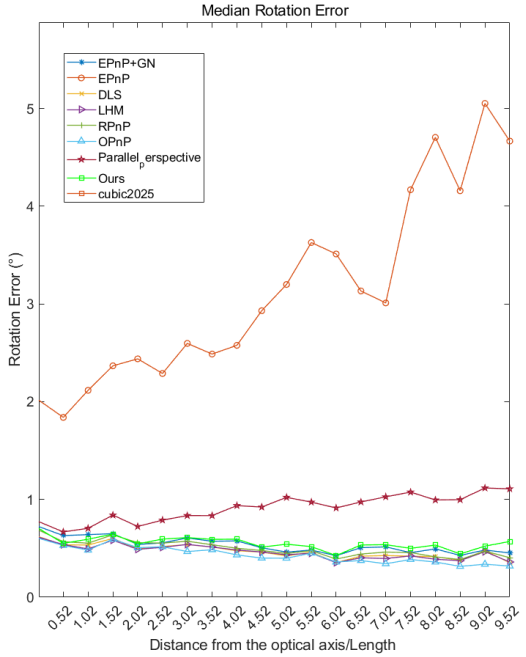
(a)


Median Rotation Error
Rotation Error (°)
Distance from the optical axis/Length
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025

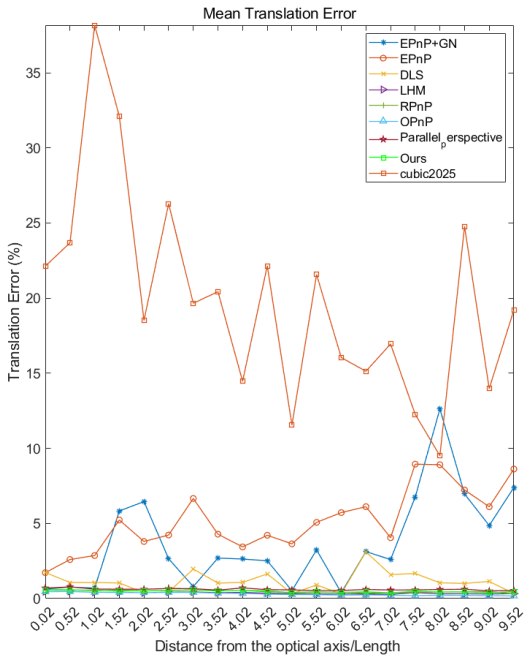
(b)


Mean Rotation Error
Rotation Error (°)
Distance from the optical axis/Length
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025

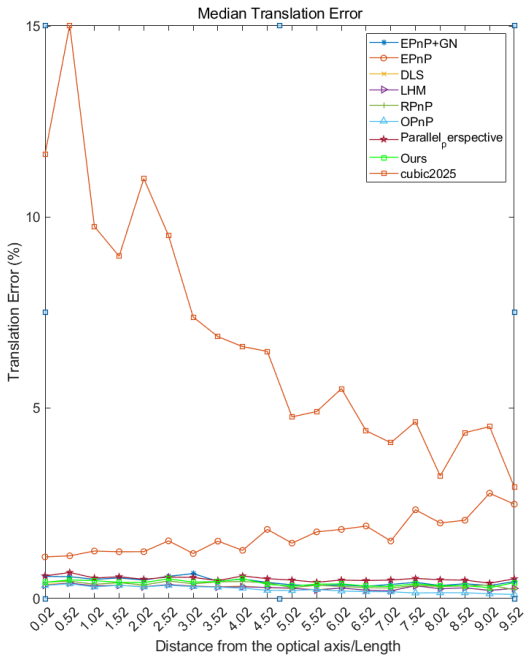
(c)


Median Rotation Error
Rotation Error (°)
Distance from the optical axis/Length
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025

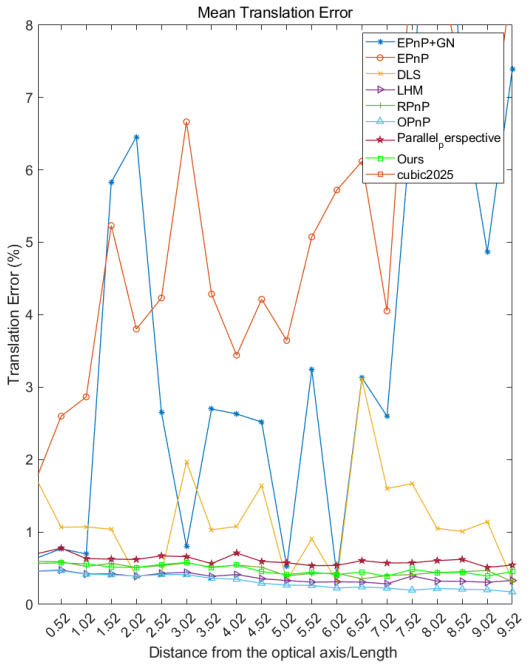
(d)



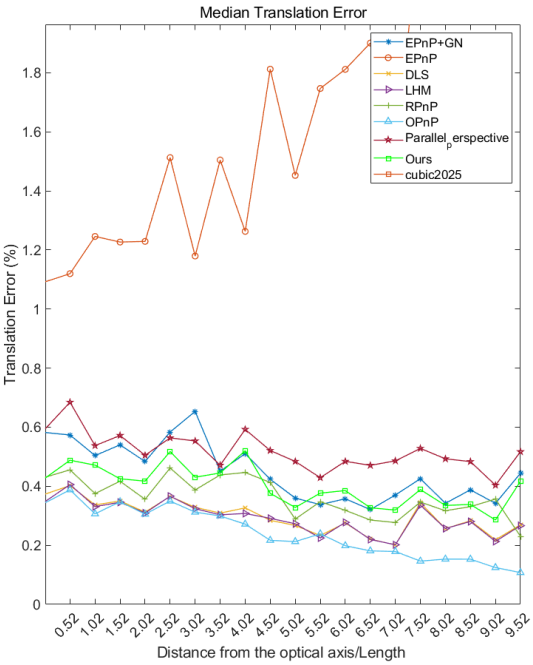
(e)



(f)



(g)

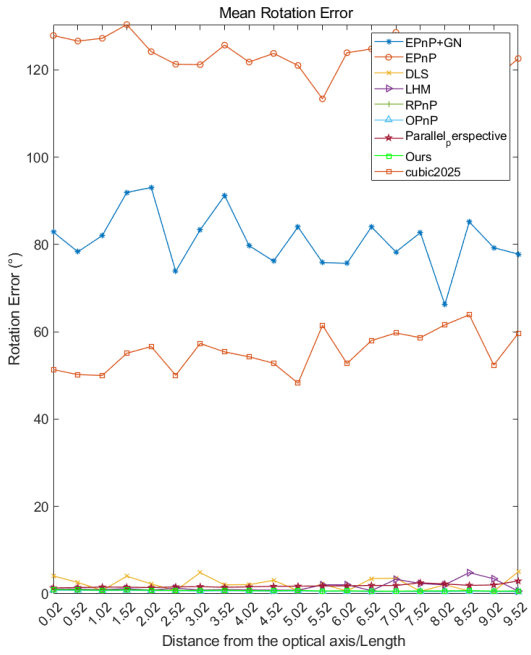


(h)

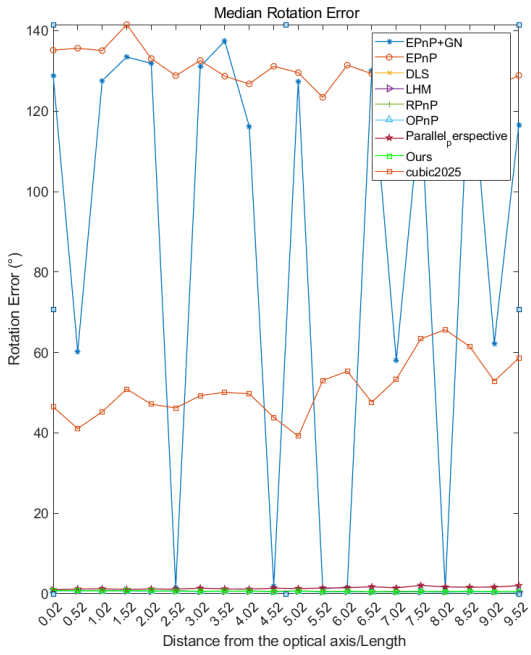
Figure 5.22. Experimental results of five-point eccentricity scenario.

Four-Point Eccentricity Scenario Analysis

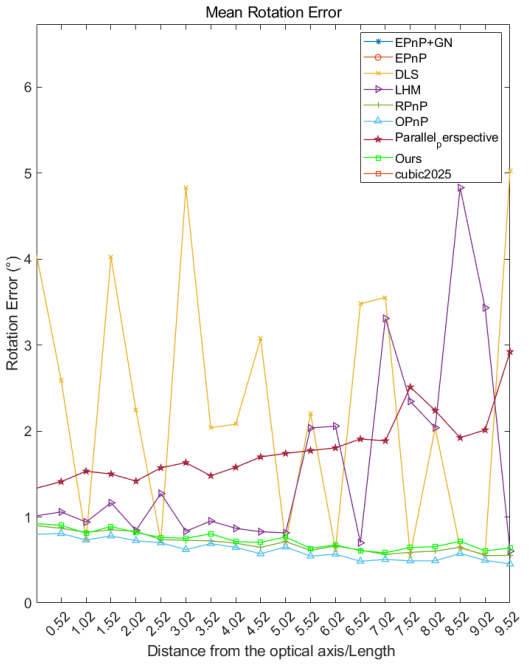
The experimental results of the four-point eccentricity scenario (as illustrated in **Figure 5.23**) demonstrated that as the distance of the reference points from the camera's optical axis increased, significant differences appeared between the tested algorithms in both rotational and translational errors. Conventional methods, including the EPnP+GN, EPnP, and DLT methods, exhibited pronounced instability and large fluctuations in error measurements, particularly at longer eccentric distances. Conversely, the proposed parallel perspective-based methods ("Ours") consistently maintained minimal rotational and translational errors, demonstrating superior stability and robustness across the entire eccentric range tested. The results underscored a significant advantage of the parallel perspective methods, emphasizing their effectiveness and reliability in single-image pose estimation in scenarios characterized by high eccentricity and limited reference points.


Mean Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Distance from the optical axis/Length

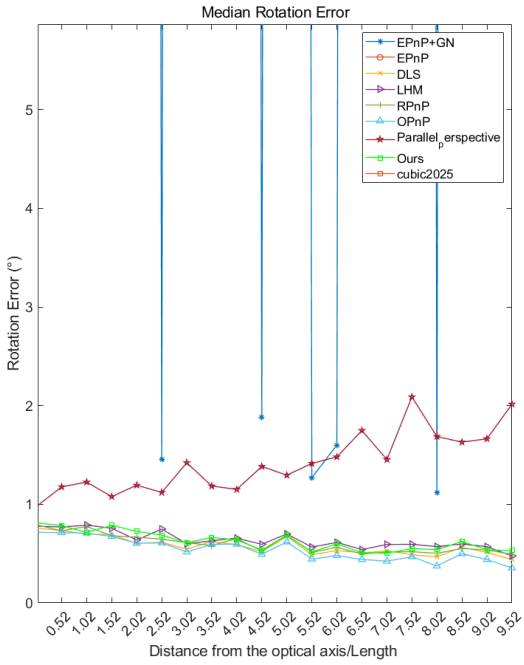
(a)


Median Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Distance from the optical axis/Length

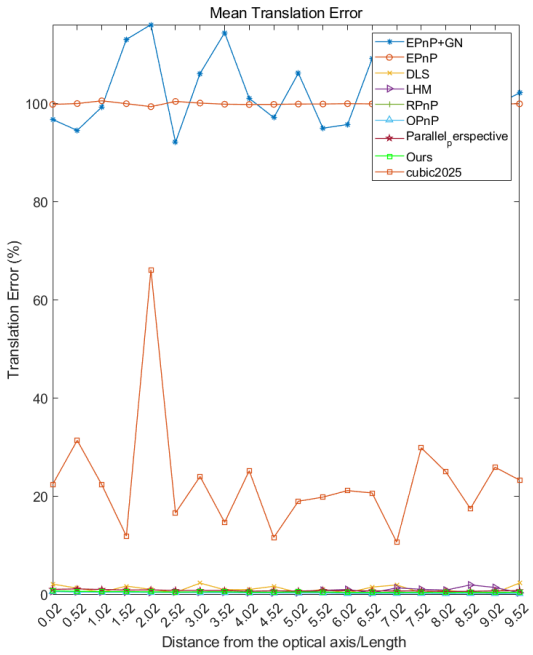
(b)


Mean Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Distance from the optical axis/Length

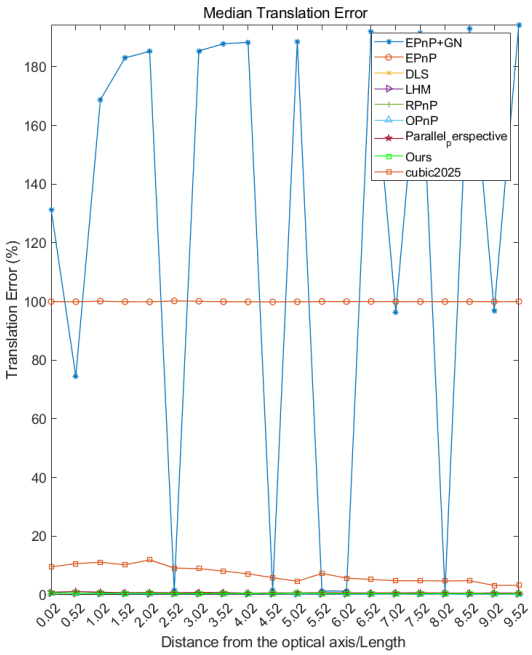
(c)


Median Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Distance from the optical axis/Length

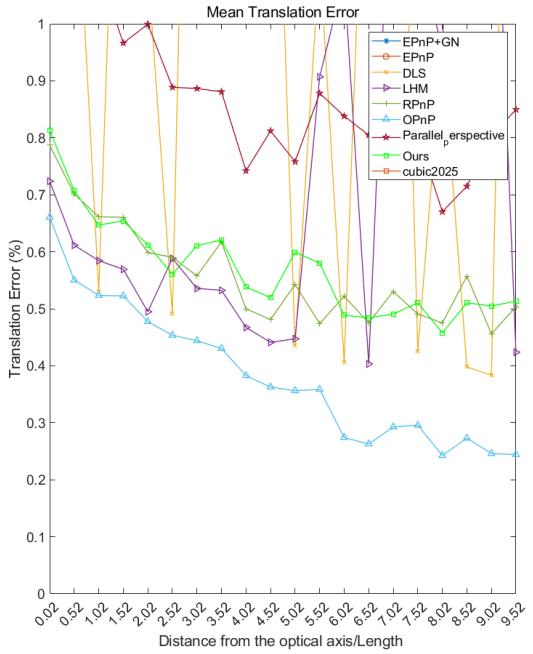
(d)



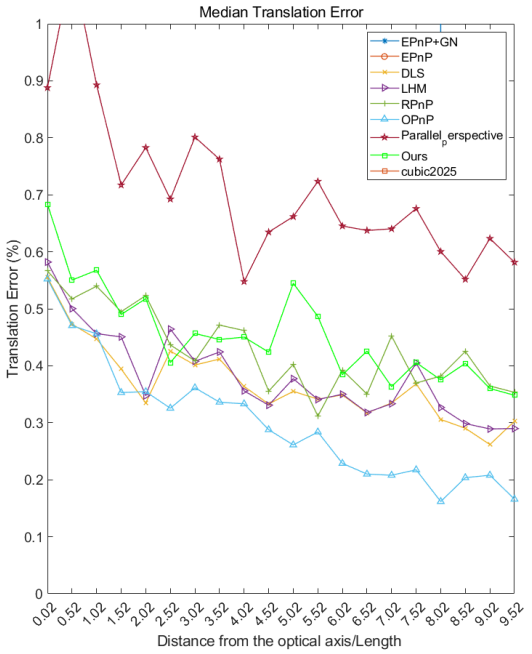
(e)



(f)



(g)



(h)

Figure 5.23. Experimental results of four-point eccentricity scenario.

Four-Point Plane Scenario Analysis

The experimental results of the four-point plane scenario clearly illustrated that the rotational and translational errors of the algorithms tested varied significantly as the distance from the camera's optical axis increased, as demonstrated in **Figure 5.24**. Conventional methods, including the OPnP, DLS, and OPnP algorithms, demonstrated substantial instability and significant increases in rotational errors, particularly at higher eccentricities. However, their translational errors remained relatively low but still exhibited noticeable fluctuations. In contrast, the proposed parallel perspective-based methods ("Ours") consistently demonstrated lower and more stable rotational and translational errors across the entire tested range of eccentricities. The comparison results highlighted the superior robustness and accuracy of the proposed parallel perspective algorithms, confirming their significant advantages in single-image pose estimation involving planar arrangements of reference points under increasing eccentric conditions.

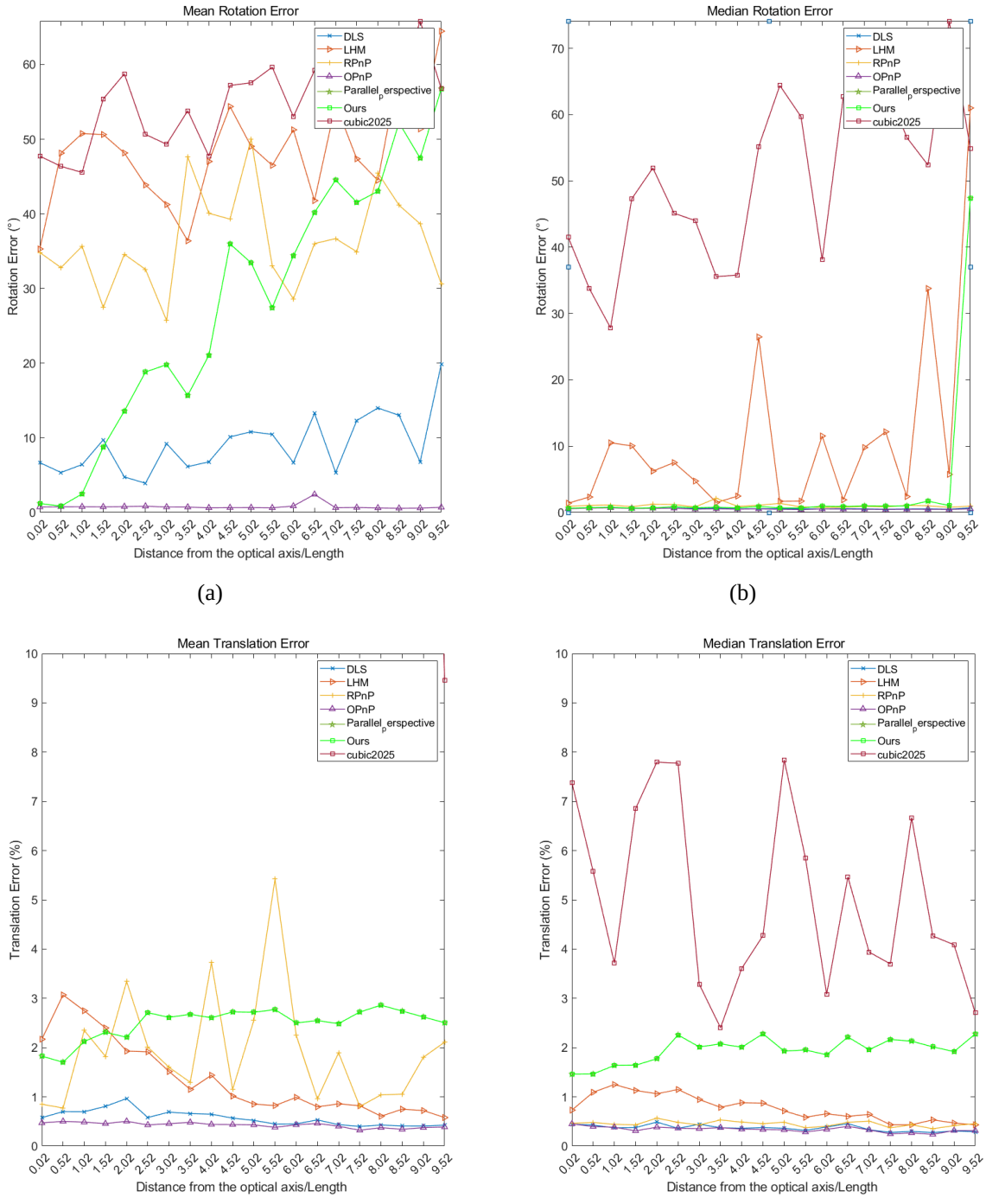


Figure 5.24. Experimental results of four-point plane scenario.

Z-Axis and Optical Axis Angle Effect Analysis

This experiment verified the effect of the angle between the camera's optical axis and the Z-axis of the reference point coordinate system on the algorithms' performance. In the experiment with simulated data, aiming to verify the effect of the angle between the marker point coordinate system and the camera's optical axis on the estimation accuracy of the positional attitude, the standard points were fixed as stereo six, stereo five, stereo four, and planar four points, sequentially; meanwhile, the origin was fixed at (0, 0, and 500), and the angle around the X- and Z-axes were zeros. In the experiment, the rotation angle around the Y-axis was changed for the reference point coordinate system (−90–90, Step 1), the camera focal length was 500, and the white noise variance was added for 100 cycles to verify the accuracy of each algorithm. The angle of rotation of the reference point coordinate system around the Y-axis was zero, and the angle of rotation of the reference point coordinate system around the Y-axis changed in the range (−90–90, Step 1). The focal length of the camera was 500; the white noise with a variance of 0.2 was added; the number of cycles was set to 100 to verify the accuracy of the algorithms. The results of the experiments are presented in Figure.

The results indicated that the attitude of the coordinate system of the feature points affected the attitude estimation of the algorithms differently.

In the case of a stereo layout of feature points, the DLT (adapted to stereo six points only), EPnP, and EPnP+GN algorithms exhibited significant positional attitude estimation errors, and the positional attitude error curves had a double U shape, with a peak in the region near zero degrees. However, the attitude estimation accuracy of the other algorithms

was not affected by the attitude of the feature point coordinate system. The accuracy of the proposed algorithm and the OPNP algorithm was the best among all algorithms tested.

The bit pattern of the four points in the plane in the feature point planar scenario also had a double U shape, except that the peak near zero degrees was very low, and the width was around 54°. This feature of the error curve was similar to that reported in the related literature, as well as to the error model proposed here. According to Figure 5.29, the attitude error estimated by the proposed algorithm had a $1/\sin\beta(\cos\beta)^{0.5}$ relationship with the angle between the Z-axis and the optical axis of the coordinate system of feature point β. The inflection point of this error curve was at the inflection point of function *f*, where $f = 1/\sin\beta(\cos\beta)^{0.5}$

$$(\frac{1}{\sin\beta(\cos\beta)^{0.5}})' = 0$$

$$\beta = 54.7°$$

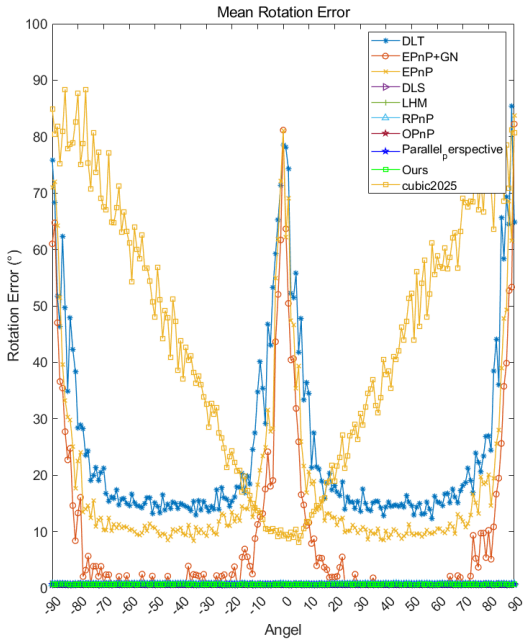
Under the stereoscopic layout of feature points, the attitude error of the stereoscopic six feature points of the proposed algorithm was between 0.5° and 0.7°, and the movement error was between 0.3% and 0.5% with straight intervals. The attitude error of the proposed algorithm in the scenario with stereo five feature points was between 0.9° and 1.1°, and its movement error was between 0.4% and 0.6% with straight intervals. The attitude error of the proposed algorithm in the scenario with stereo four feature points was between 1° and 1.3°, and its movement error was between 0.5% and 0.8% with straight intervals. For the four planar points, the error was much higher than an attitude error of 1°–10° and a displacement error of 0.3%–1% (between −54° and 54°). Although the four-planar feature point layout has been widely adopted in medical navigation systems, the stereoscopic

layout demonstrated improved accuracy and stability of the proposed algorithm for sauce oil.

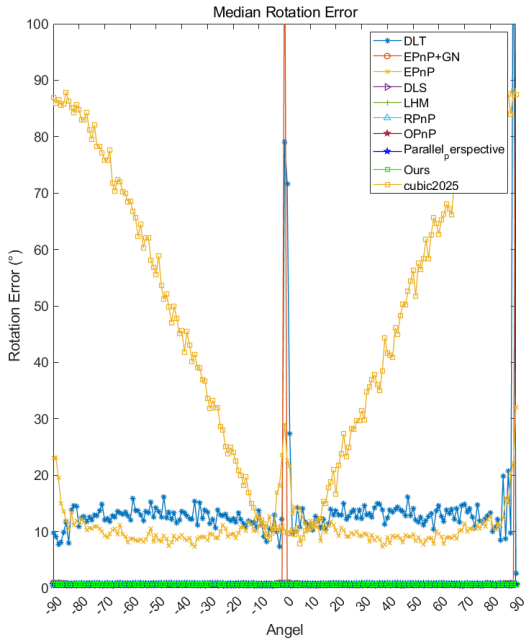
Six-Point Scenario Analysis

The experimental results of the Z-axis and optical axis angle effect analysis conducted utilizing a scenario with six points indicated notable differences between the algorithms as the angle between the Z-axis and the optical axis changed, as demonstrated in **Figure 5.25**. Conventional algorithms, such as the DLT and EPnP+GN algorithms, exhibited pronounced sensitivity and considerable fluctuations in rotational and translational errors, particularly at angles approaching ±90°, revealing significant instability under such challenging orientation conditions. Conversely, the proposed parallel perspective-based algorithms ("Ours") consistently demonstrated remarkably low and stable errors across the entire angle range tested. This indicated the strong robustness and accuracy of the proposed methods in handling variations in angular orientation between the optical axis and the reference plane.

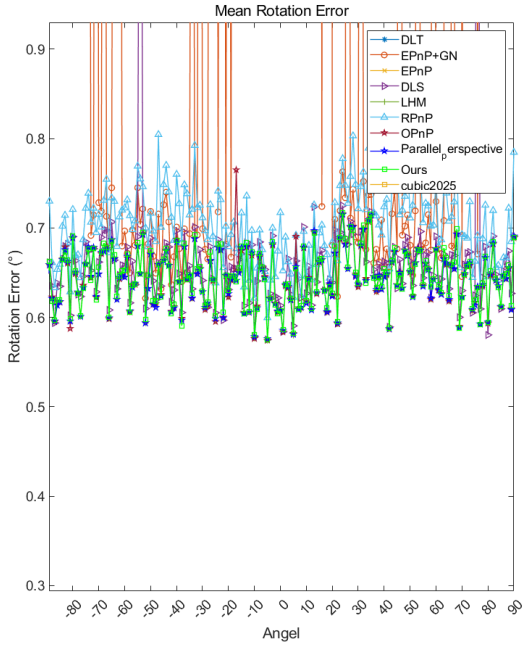
Consequently, the proposed parallel perspective algorithms demonstrate a clear advantage for reliable single-image pose estimation under challenging angular alignment conditions.


Mean Rotation Error
DLT
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Angel

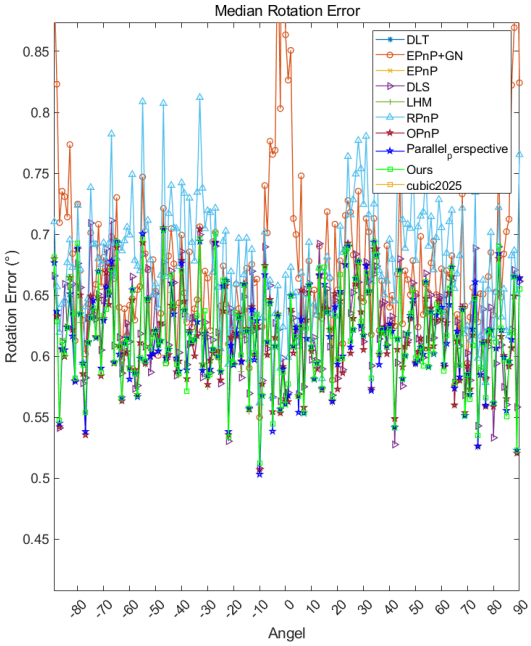
(a)


Median Rotation Error
DLT
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Angel

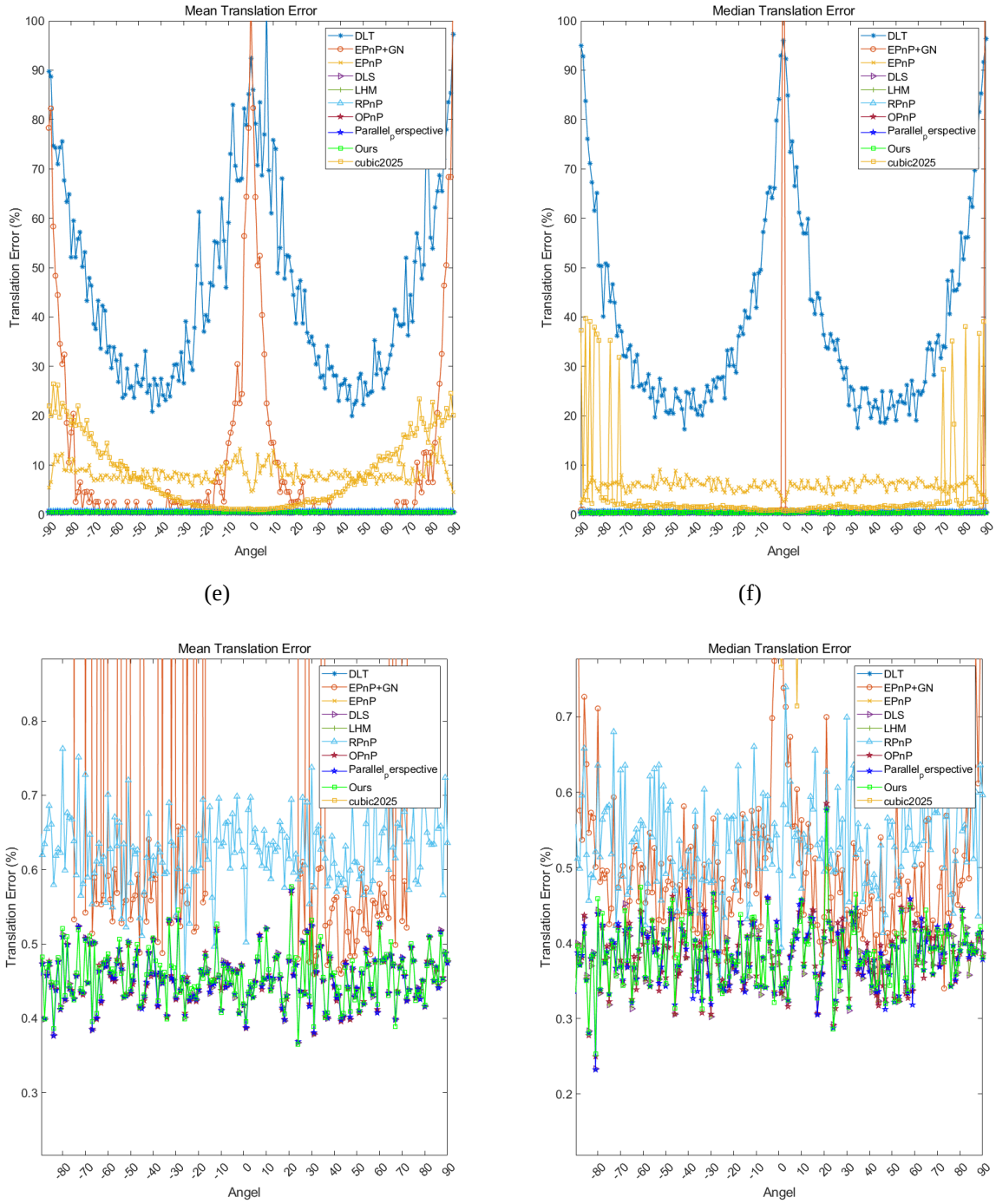
(b)


Mean Rotation Error
DLT
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Angel

(c)


Median Rotation Error
DLT
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Angel

(d)



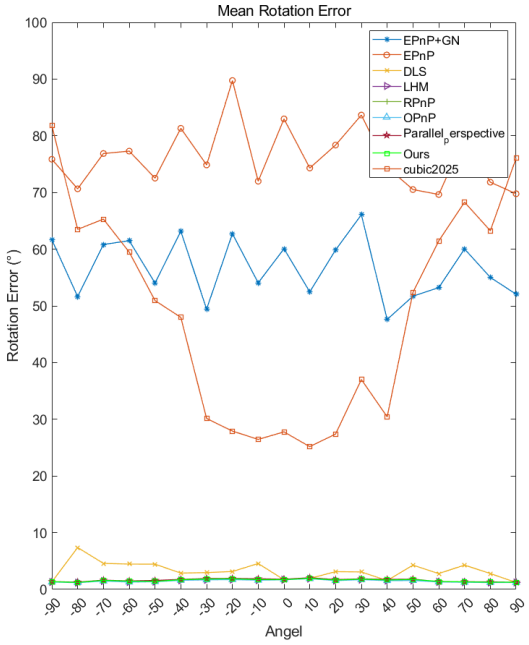
(e) (f)

(g) (h)

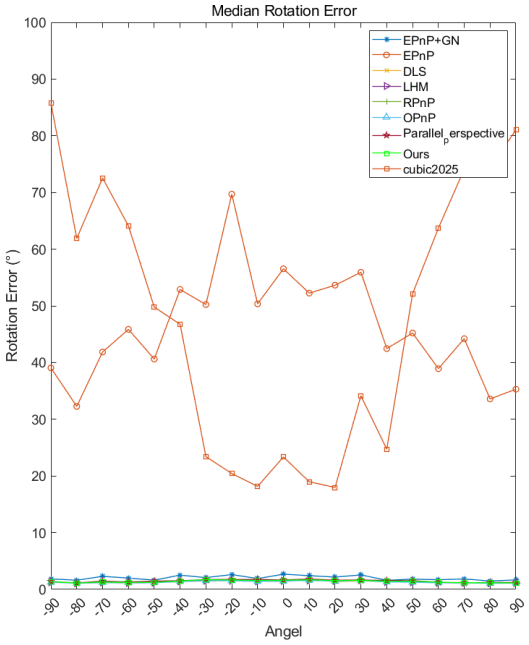
Figure 5.25. Experimental results of Z-axis and optical axis angle analysis

Five-Point Scenario Analysis

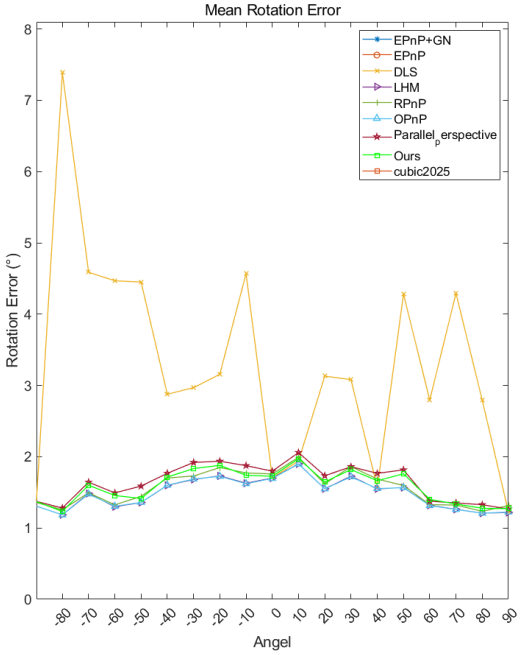
The experimental results of the Z-axis and optical axis angle analysis for the five-point scenario indicated varying levels of robustness among the tested algorithms, as illustrated in **Figure5.26**. Conventional algorithms, including the DLT and EPnP+GN algorithms, exhibited substantial fluctuations and significant increases in both rotational and translational errors at specific angular orientations, which indicated their instability and sensitivity to angular variations between the optical axis and the reference points. In contrast, the proposed parallel perspective-based methods ("Ours") consistently had remarkably stable and significantly lower rotational and translational errors throughout the entire angular range tested. The comparison results strongly demonstrated the enhanced robustness, stability, and accuracy of the proposed parallel perspective methods in pose estimation, especially in challenging scenarios involving significant angular variations between the optical axis and Z-axis.


Mean Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Angel

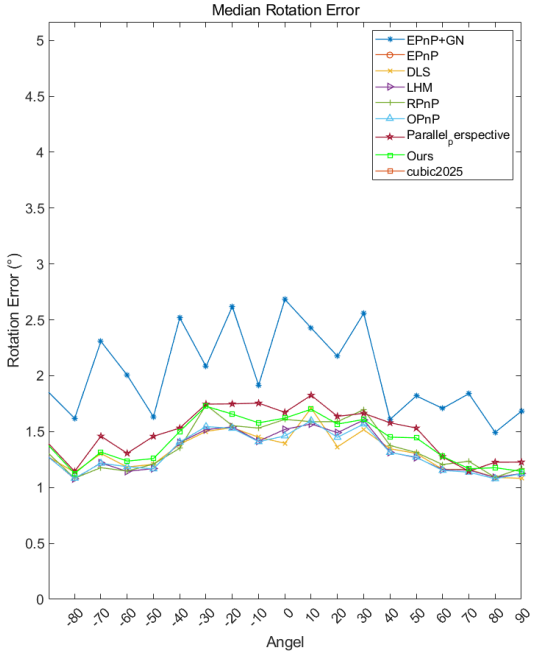
(a)


Median Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Angel

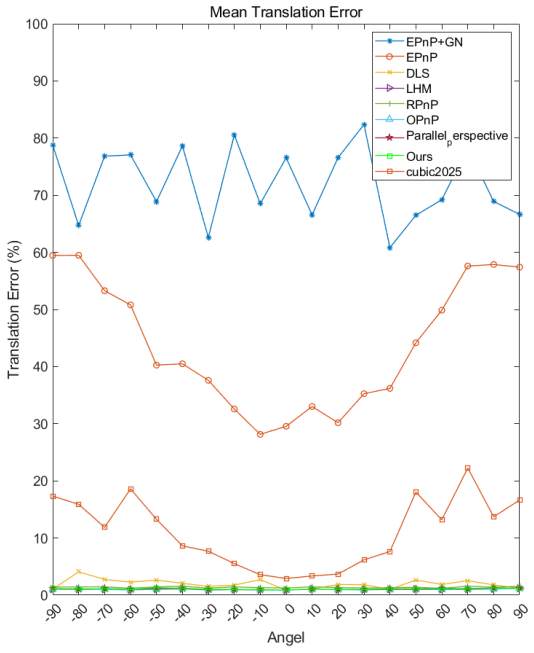
(b)


Mean Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Angel

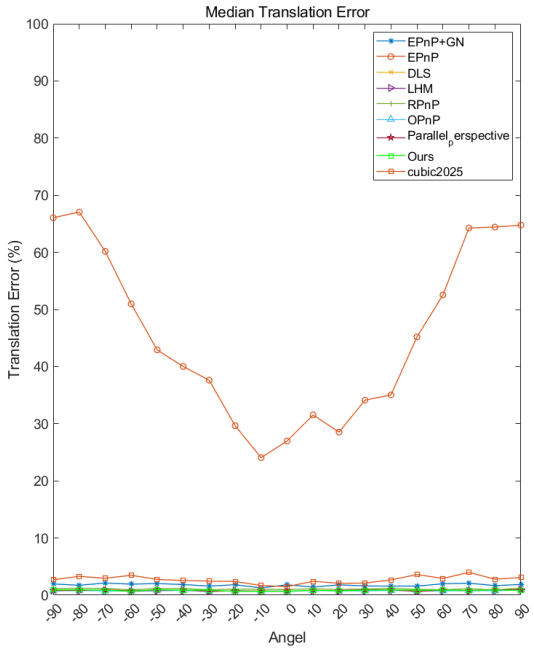
(c)


Median Rotation Error
EPnP+GN
EPnP
DLS
LHM
RPnP
OPnP
Parallel_perspective
Ours
cubic2025
Rotation Error (°)
Angel

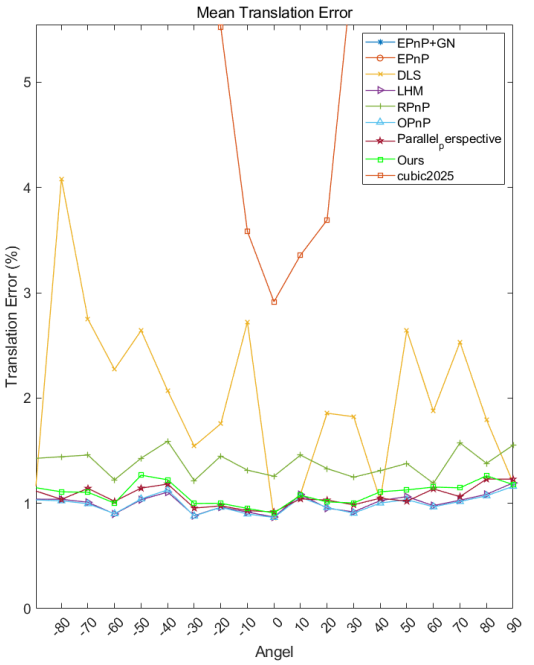
(d)



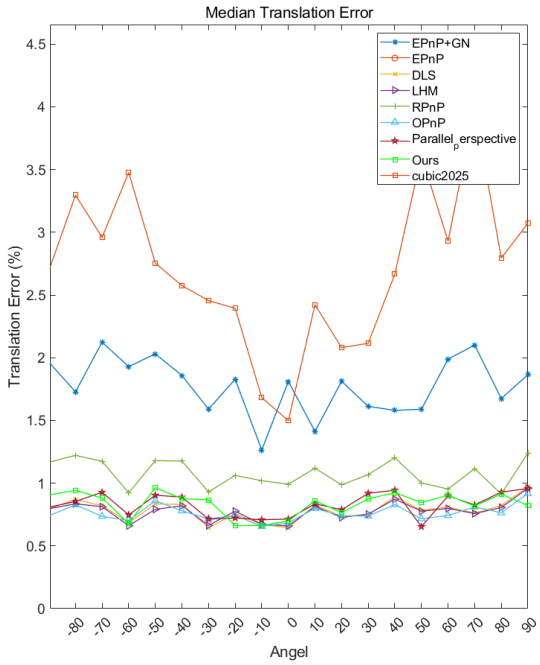
(e)



(f)



(g)



(h)

Figure5.26. Experimental results of Z-axis and optical axis angle in five-point scenario.

Four-Point Scenario Analysis

The experimental results of the Z-axis and optical axis angle analysis conducted in the four-point scenario indicated significant differences in robustness between the tested algorithms when subjected to changes in angular orientation, as demonstrated in **Figure5.27**. Conventional methods, particularly the EPnP+GN, EPnP, and DLT algorithms, demonstrated severe sensitivity and exhibited significant fluctuations and spikes in both rotational and translational errors, particularly at larger angular deviations approaching ±90°. In contrast, the proposed parallel perspective-based algorithms ("Ours") consistently maintained low and stable rotational and translational errors across all angular values tested. This stability underscored the superior robustness and effectiveness of the proposed algorithms in pose estimation under challenging conditions involving pronounced angular variations between the optical axis and the reference plane.

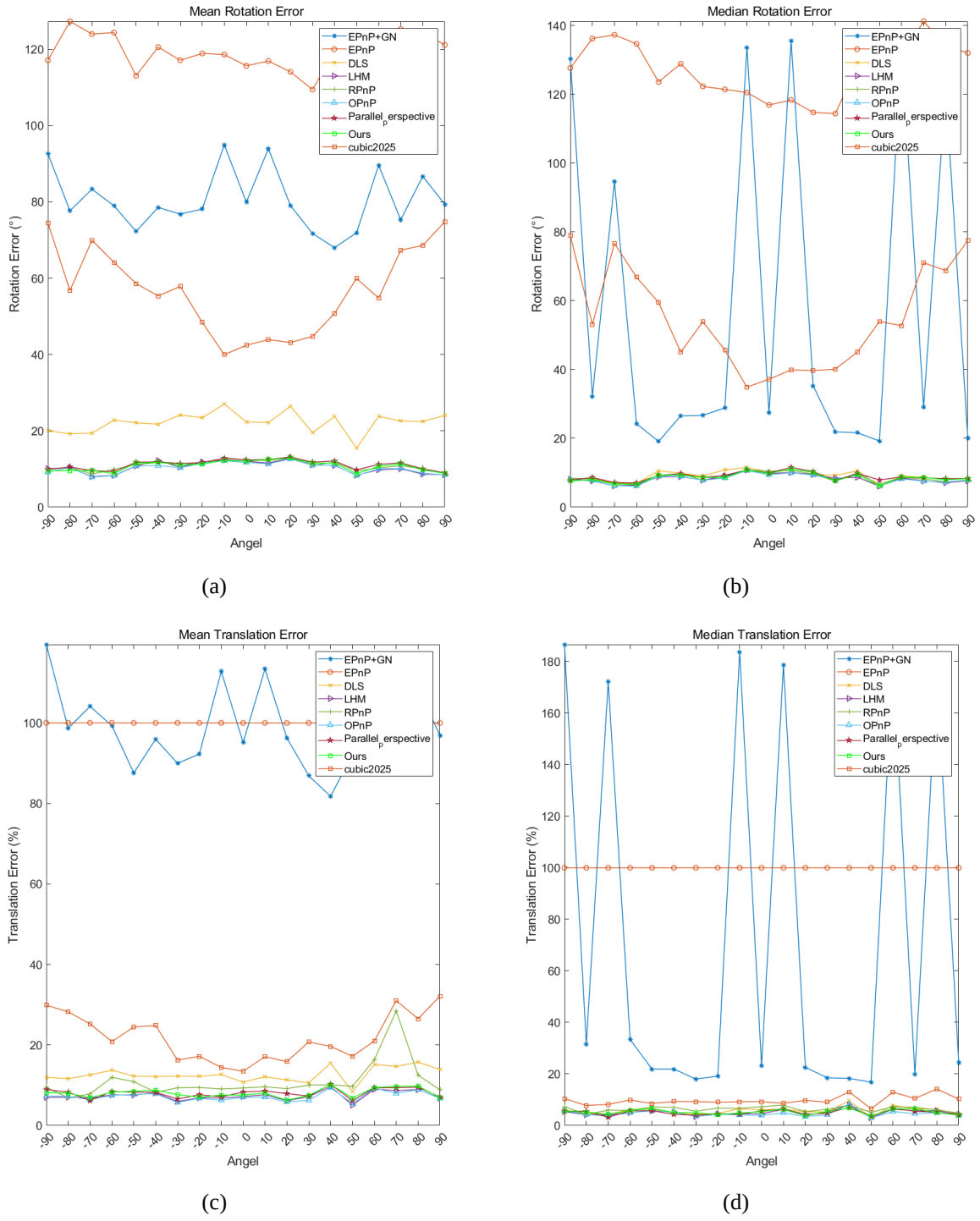


Figure5.27. Experimental results of Z-axis and optical axis angle analysis in four-point scenario.

Four-Point Plane Scenario Analysis

The experimental results of the four-point plane scenario demonstrated significant variations in the algorithms' performance as the angle between the Z-axis and the optical axis changed, as illustrated in **Figure5.28**. Conventional methods, including the DLS, and OPnP algorithms, demonstrated pronounced sensitivity, and their rotational and translational errors dramatically increased and exhibited substantial instability, particularly near angular extremes (±90°). In contrast, the proposed parallel perspective-based algorithms ("Ours") consistently maintained lower error levels and significantly higher stability across nearly the entire angular range. However, the proposed methods also exhibited some error increases at the extreme angles, but their performance remained markedly superior to conventional approaches. The experimental results confirmed that the proposed parallel perspective-based algorithms provided enhanced robustness, stability, and accuracy in pose estimation, which could be particularly beneficial in challenging angular alignment scenarios involving planar arrangements of reference points.

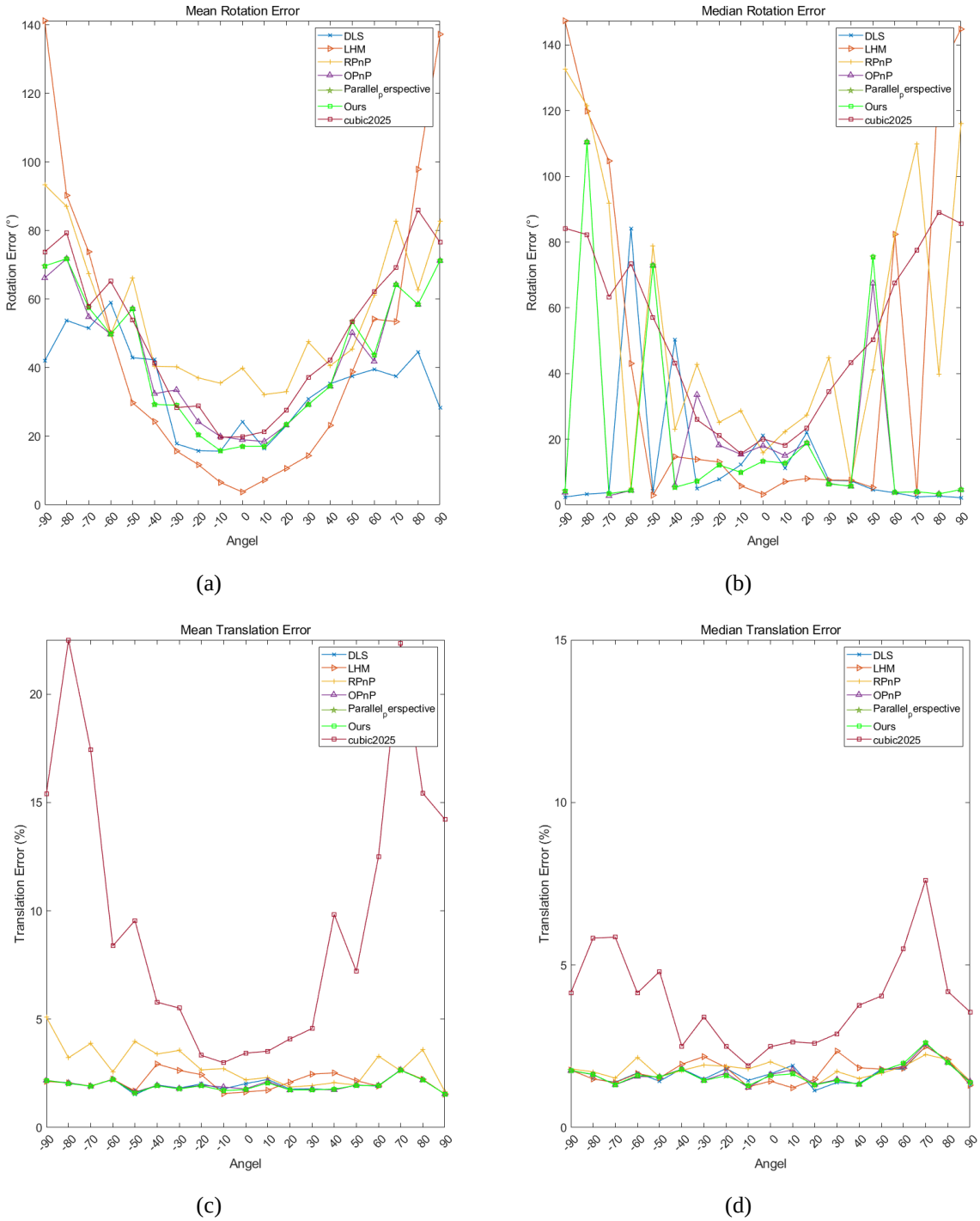


Figure5.28. Experimental results of four-point plane scenario**.**

Real Image Experimental Results and Analysis

Under different lighting environments, the robot moves while carrying the feature point coordinate system and then rotates around the X-axis, starting from 0° and rotating by two angles each time until it reaches ±90. A picture is taken after each rotational movement to measure the pose of the marker point coordinate system and calculate the error according to (5.33). Considering that the bit position measurement of each motion position is based on one picture, the noise interference level is high, which affects the evaluation results of different algorithms. Therefore, here, the average value of the bit pose errors at different rotation angles from the moving position of the feature point coordinate system is utilized to evaluate the accuracy and robustness of different algorithms. The average value of the errors of the angles at different positions is utilized to evaluate the accuracy and robustness of different algorithms.

The data in the experimental data table represent the average camera pose estimation errors produced for the feature point coordinate system at various distances from the camera along the Z-axis direction. Each row represents the average camera position estimation errors generated by different position measurement algorithms for 90 distinct poses at the same location. Each column contains the average camera pose estimation errors produced by the same pose measurement algorithm for 90 different poses at different distances from the optical axis of the camera at the same camera depth.

Table 5.1-Table 5.5 present the experimental data yielded under the underwater low-light environment.

Table 5.1. Marker Point Coordinate System is 200 mm from the Camera.

	Err	DLT	EPnP	EPnP+GN	DLS	LHM	RPnP	OPnP	Parallel	Ours
0	RE(°)	2.96	2.78	1.81	1.78	1.64	1.78	1.61	1.55	1.55
	TE(%)	32.26	5.99	3.05	1.70	1.70	1.91	1.56	1.25	1.25

Table 5.2. Marker Point Coordinate System is 400 mm from the Camera.

	Err	DLT	EPnP	EPnP+GN	DLS	LHM	RPnP	OPnP	Parallel	Ours
0	RE(°)	5.75	5.14	3.55	2.07	2.03	2.03	1.99	1.97	1.97
	TE(%)	31.47	7.49	3.60	0.74	0.74	0.96	0.74	0.75	0.75
50	RE(°)	5.56	5.32	1.66	1.61	1.54	1.91	1.52	1.52	1.52
	TE(%)	32.57	7.28	1.51	0.88	0.88	0.99	0.84	0.76	0.76
100	RE(°)	7.15	6.56	2.10	1.80	1.72	1.95	1.72	1.73	1.73
	TE(%)	30.26	6.26	1.88	1.37	1.37	1.40	1.28	1.17	1.17

Table 5.3. Marker Point Coordinate System is 600 mm from the Camera.

	Err	DLT	EPnP	EPnP+GN	DLS	LHM	RPnP	OPnP	Parallel	Ours
0	RE(°)	14.86	14.30	2.31	2.25	2.19	2.32	2.16	2.21	2.21
	TE(%)	18.77	5.58	1.14	0.98	0.97	0.77	0.85	0.78	0.78
50	RE(°)	10.00	9.42	4.21	2.23	2.13	2.17	2.12	2.10	2.10
	TE(%)	25.95	6.75	3.69	1.23	1.22	1.21	1.14	1.06	1.06
100	RE(°)	7.94	6.51	1.95	3.15	1.72	1.70	1.73	1.76	1.76
	TE(%)	29.72	7.04	1.58	1.88	1.01	1.12	1.02	1.01	1.01

Table 5.4. Marker Point Coordinate System is 800 mm from the Camera.

	Err	DLT	EPnP	EPnP+GN	DLS	LHM	RPnP	OPnP	Parallel	Ours
0	RE(°)	11.22	10.63	3.80	2.20	2.11	2.18	2.10	2.13	2.13
	TE(%)	23.81	7.72	2.99	0.64	0.64	0.62	0.60	0.57	0.57
50	RE(°)	12.44	10.36	2.47	2.24	2.20	2.27	2.24	2.20	2.20
	TE(%)	23.37	7.36	1.22	0.86	0.87	1.14	0.91	0.89	0.89
100	RE(°)	12.63	12.86	3.96	2.01	2.01	2.06	3.96	2.06	2.06
	TE(%)	21.52	6.48	3.45	0.92	0.92	1.24	214.86	0.98	0.98

Table 5.5. Marker Point Coordinate System is 1000 mm from the Camera.

	Err	DLT	EPnP	EPnP+GN	DLS	LHM	RPnP	OPnP	Parallel	Ours
0	RE(°)	21.66	18.54	5.86	2.62	2.66	2.72	2.66	2.67	2.67
	TE(%)	32.56	7.37	6.26	1.57	1.56	1.99	1.54	1.63	1.63
50	RE(°)	13.91	13.91	4.04	2.24	2.13	2.20	2.13	2.10	2.10
	TE(%)	22.37	7.27	3.10	0.69	0.67	0.76	0.68	0.67	0.67
100	RE(°)	20.66	19.68	7.16	5.17	1.88	2.01	1.88	2.00	2.00

The following conclusions can be drawn from the analysis.

1. Marker images are collected in the underwater low-light environment, and the marker position information of these images is manually extracted. The position extraction error does not interfere much with the estimation accuracy of the proposed algorithm, and the accuracy of this algorithm is consistent with the accuracy of well-known algorithms, such as OPnP, LHM, RPnP, and DLS.

2. Compared with those in a general environment, the lighting conditions and image quality in an underwater environment are more disturbed. Underwater images are generally blurred and noisy, resulting in poor algorithmic accuracy, especially with increasing distance, and

the induced error increases significantly. Parallel and parallel-weighted algorithms are more stable in underwater environments, and their errors are relatively small at most distances. In particular, the translation errors (TEs) are more accurate at closer distances, and the error increases less as the distance increases. The conventional DLT and EPnP algorithms perform poorly in underwater environments, especially at longer distances (e.g., 1000 mm), where the errors increase dramatically, especially the rotational error (RE) and TE values; thus, these methods exhibit greater instability.

3. The errors of the DLT and EPnP algorithms are significantly affected by the distance from the optical axis. In the underwater environment, the RE errors are 5.75° (DLT) and 5.14° (EPnP) at a distance of 400 mm. The errors increase further when the optical axis is offset by 50 and 100 mm, indicating that these two algorithms are very sensitive to changes in the offset optical axis. The algorithms examined here are more stable and resistant to changes in the error of the off-axis, with the rotational error varying within 0.7° and the translational error varying within 1% with the offset of the optical axis.

Discussion

The results of this study indicate that the proposed ParaPers+GN algorithm can effectively address the instability problem encountered by the EPnP+GN algorithm for a smaller number of marker points. By employing the parallel perspective projection model to acquire the initial camera pose, the proposed algorithm simplifies the calculation process. It provides a more robust solution, particularly in scenarios with high noise and large depths. Although the initial accuracy of the parallel perspective projection model is slightly

lower than that of the EPnP+GN algorithm, the overall performance of the proposed algorithm remains unaffected by the Gauss–Newton optimization. In addition, incorporating the location characteristics of control points in the construction process of the objective function contributes to the enhanced accuracy of the ParaPers+GN algorithm. This approach sets the proposed algorithm apart from the IEPnP algorithm, which struggles to calculate the correct result at small marker point depths.

Furthermore, the proposed algorithm exhibits slightly higher accuracy than the existing algorithms in other cases as well. The experimental results obtained with both simulated and real image data confirm the improved stability and accuracy of the proposed ParaPers+GN algorithm over the EPnP+GN algorithm under various levels of Gaussian noise and marker point depths. In addition, the computational efficiency of the proposed algorithm is demonstrated to be superior to both the EPnP+GN and IEPnP algorithms.

However, despite these promising results, it would be beneficial to further explore the potential limitations of the proposed algorithm, evaluating its performance in more challenging scenarios with varying marker point configurations or different types of noise. Future research could also investigate the integration of the proposed ParaPers+GN algorithm with other optimization techniques to enhance its robustness and accuracy further. Under real-world lighting environments—low underwater illumination conditions—the proposed algorithm exhibits strong adaptability to specific application scenarios. The accuracy of the pose estimates was consistent with that of the leading algorithms, such as OPnP, LHM, RPnP, and DLS. Overall, the findings of this study

demonstrate the high application potential of the proposed algorithm for camera pose estimation and highlight its advantages over the existing methods in the field.

For instance, the VINS-MONO algorithm has a good performance in vision-inertial SLAM but faces certain limitations in scenarios with a close distance between the robot and objects in complex underwater environments. This study aims to address the current shortcomings of the VINS-MONO algorithm and propose effective solutions for optimizing the VINS-MONO algorithm's performance under challenging conditions. The comparative tests verify the feasibility of the proposed solution. In this chapter, the optimization of the VINS-MONO algorithm is to enhance the EPnP algorithm from the front end. The proposed parallel perspective-based algorithm is utilized to replace the EPnP algorithm, which can compensate for the disadvantage of short distance, significantly improving the VINS-MONO algorithm's performance in camera pose estimation. Finally, various strategies are employed to preserve and optimize residual information, aiming to enhance the accuracy of pose estimation.

Conclusion

In conclusion, the proposed ParaPers+GN algorithm demonstrates promising potential for camera pose estimation applications. The proposed algorithm can effectively address the instability issue in the EPnP+GN algorithm when handling fewer than six marker points, while maintaining competitive accuracy and computational efficiency. The superior stability and accuracy of the proposed ParaPers+GN algorithm under various levels of

Gaussian noise and different marker point depths, as well as its robustness in high-noise and large-depth scenarios, underscore its advantages over the existing methods in the field. Future research could investigate the performance of the proposed algorithm in more challenging scenarios or for different types of noise, as well as explore its integration with other optimization techniques to further improve its robustness and accuracy. Overall, the proposed ParaPers+GN algorithm has great promise for camera pose estimation applications and demonstrates significant advantages over the existing methods in the field.

CHAPTER SIX

VALIDATION EXPERIMENTS

Introduction

The data utilized in the validation experiments of the image processing algorithms were primarily categorized into synthetic (simulated) data and real-world data. Synthetic data were generated by artificially applying blurring processes to clear images, typically through the convolution of a sharp image with a blur kernel. These data included paired blurred and sharp images, facilitating both algorithm training and performance evaluation. In contrast, real-world data comprised genuinely captured blurred images that occur in natural scenarios (Rim et al., 2020).

This chapter presents experimental validations across two key technical domains:

(1) Underwater image restoration based on transmission optimization (image restoration);

(2) Motion blur parameter estimation leveraging cepstral morphological features (motion blur parameter estimation).

Datasets utilized for image deblurring algorithms can be roughly categorized into synthetic datasets and real-world datasets. Synthetic datasets are generated by artificially defining blur processes, for instance, by convolving clear images with blur kernels or applying other forms of synthetic blurring (Levin et al., 2009; Nah et al., 2017). These datasets typically provide paired blurred and sharp images, which facilitates effective training and performance evaluation of algorithms. In contrast, real-world datasets contain naturally blurred images captured under real conditions, where blurring is often caused by

factors such as camera shake or handheld acquisition (Rim et al., 2020; Jiang et al., 2020). Real-world datasets more accurately simulate practical image degradation scenarios and impose higher demands on the robustness of deblurring algorithms compared to synthetic datasets.

Real-World Datasets

The above datasets primarily comprise synthetically generated blurred images, which pose certain limitations on the practical applicability of deblurring models. To address this problem, Rim et al. (Rim et al., 2020) introduced a real-world blurred image dataset named RealBlur, which comprises two subsets: the RealBlur-R subset, containing raw camera images, and the RealBlur-J subset, including JPEG-compressed camera images. The RealBlur dataset includes a total of 9,476 image pairs. Jiang et al. (Jiang et al., 2020) constructed a large-scale outdoor dataset named Blur-DVS, which was captured with a DAVIS240C event camera. The Blur-DVS contains 1,782 training pairs and 396 testing pairs, along with an additional 740 standalone real blurred images. In addition, Li et al. (2023) (Z. Li et al., 2023) developed the ReLoBlur, a real-world local motion deblurring dataset that authentically presents the aliasing effects caused by the motion of local objects and background blending. This dataset includes 2,010 training pairs and 395 testing pairs.

To simulate realistic underwater conditions, this study designed and implemented a custom-built water tank experimental platform, as illustrated in **Figure6.1**. The experimental platform included a tank filled with stones of different sizes, aquatic plants, and sand, which was utilized to approximate low-texture environments typically found on

the seabed or riverbed. Because underwater robots cannot achieve precise motion control underwater, this study employed a monocular camera, which enables both linear and rotational motions. The monocular camera captured sequences of blurred images induced by the controlled movements. By adjusting the water turbidity and illumination conditions, various realistic underwater scenarios were created, including uneven lighting and suspended particulate matter, allowing for the validation of the robustness of the proposed algorithms under diverse real-world conditions.

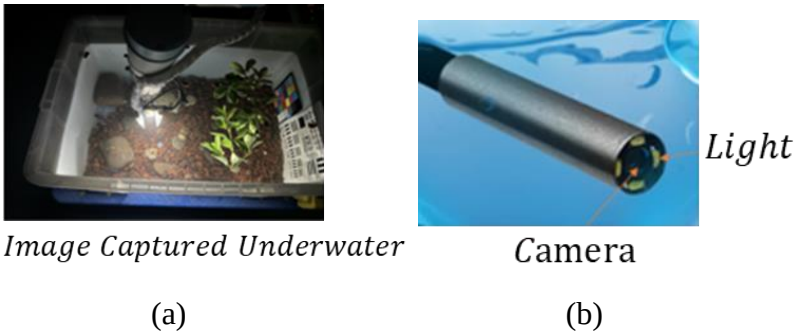


Figure6.1. Underwater robotic water tank experimental platform: (a) six-degree-of-freedom collaborative robot equipped with camera and light source for performing precise movements underwater to capture motion-degraded video sequences; (b) underwater clarity charts and color.

Compared to the public datasets, obtaining perfectly matched ground-truth sharp images in real-world water tank experiments could be challenging. Therefore, reference-based evaluation metrics, including the PSNR and SSIM indices, cannot be directly applied in such situations. This study adopted no-reference underwater image quality assessment metrics, the UIQM and UCIQE metrics, to assess the clarity, contrast, and color distortion of the output images utilizing only the images themselves, thereby eliminating the need for ground-truth references.

Visual Perception Experiments

Experimental Platform Design

This research involves underwater image processing and camera pose estimation and considers underwater imaging characteristics and motion blur noise, which requires the acquisition of underwater images under various environmental conditions.

In addition, a dedicated underwater image acquisition experimental platform was designed for experimental validation, as illustrated in **Figure 6.1.**

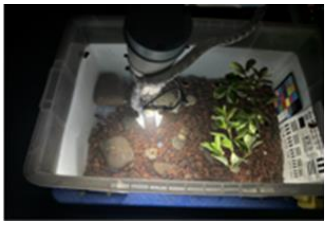
The basic configuration of the experimental platform comprises the following components: (1) a target image acquisition module comprising a camera-light source assembly with a 1-megapixel resolution (1280 × 720 pixels), depth of field of 30–80 mm, frame rate of 30 FPS, and a field of view of 80°; (2) a feature point coordinate system or scene representing the target object; (3) an image processing and pose estimation controller, which was a high-performance eight-core personal computer equipped with an AMD Ryzen 7 3700X CPU and 32 GB of RAM. To evaluate the performance of underwater image processing and pose estimation algorithms, the experimental platform also incorporated a six-degree-of-freedom robotic arm (UR5) which was substituted for an underwater robot to adjust either the target object's position or the camera's pose.

To simulate realistic underwater conditions, a custom-designed water tank experimental platform was also developed, as presented in **Figure6.2.** The tank was filled with stones of different sizes, aquatic plants, and sand, which allowed for simulating a low-texture seabed or riverbed environment. Suspended particulates, such as silica sand and organic materials, were introduced to simulate different turbidity levels (i.e., clear, low,

medium, and high turbidity). By adjusting both the turbidity and illumination conditions within the tank, a range of realistic underwater scenarios, including uneven lighting and suspended particulate matter, was recreated, thus enabling the validation of algorithmic robustness under diverse real-world conditions.

Stage 1: Image Restoration

Dark and Turbid Underwater Image



Scene Image Captured Underwater

Figure6.2. Scene photo obtained during image acquisition.

The experiments were conducted in a custom-designed water tank platform that simulated underwater environments. In the tank, stones, aquatic plants, and other obstacles were placed, and the water turbidity was adjusted by introducing a controlled amount of suspended particles, thereby creating varying levels of visual degradation. An Aubo-i5 collaborative robotic arm was deployed alongside the tank to perform precise linear and swinging motions; it was equipped with a monocular camera mounted at the end-effector to capture sequences of underwater images. By controlling the robot's motion speed and the turbidity level of the water, two distinct types of underwater degradation scenarios were generated: scenarios with moderately blurred image sequences and scenarios with severely blurred image sequences.

Raw underwater images were captured under various turbidity and illumination conditions. The proposed transmission optimization algorithm was applied to the collected

images and compared with several existing mainstream algorithms, including the AGCWD, MCRCR, WIGF, and Phaseformer algorithms. Performance evaluation was performed utilizing both quantitative measures (PSNR and SSIM) and qualitative visual assessments.

The specific steps of the experiment procedure were as follows.

1. Collection of original underwater images under different turbidity and illumination conditions
2. Restoration of collected images via the proposed transmission optimization algorithm
3. Comparison with the existing algorithms, including the AGCWD, MCRCR, WIGF, and Phaseformer algorithms
4. Calculation and analysis of PSNR and SSIM values, along with subjective visual assessments.

Objective Evaluation

The results obtained under moderate turbidity conditions were as follows (as presented in **Table 6.1-Table 6.2**.

- PSNR Metric: The WIGF method generally performed best, achieving up to 21.57 dB (Groups 2 and 6). The proposed algorithm achieved good results in the second group (20.53 dB) but indicated room for improvement in other groups;
- SSIM Metric: The proposed method demonstrated superior performance, particularly in Groups 2 (0.74) and 9 (0.73).

Table 6.1. PSNR results obtained under moderate turbidity conditions.

Method / Image	AGCWD	MCRCR	WIGF	Phaseformer	Proposed
1	13.69	13.11	17.34	15.01	13.16
2	17.05	14.46	21.57	16.09	20.52
3	11.84	15.95	17.39	15.29	13.96
4	9.56	8.21	16.72	13.91	10.38
5	14.32	12.84	18.54	15.88	15.82
6	11.13	10.71	21.57	17.25	11.48
7	10.99	8.98	20.88	15.98	9.58
8	12.44	11.35	15.32	10.03	12.91
9	14.05	13.63	15.66	9.89	14.48
10	15.16	13.52	19.39	15.13	17.39
11	14.35	12.04	17.16	14.73	14.38
12	13.88	13.23	19.06	15.84	16.24

Table 6.2. SSIM results obtained under moderate turbidity conditions.

Method / Image	AGCWD	MCRCR	WIGF	Phaseformer	Proposed
1	0.53	0.28	0.37	0.19	0.54
2	0.65	0.44	0.45	0.25	0.74
3	0.47	0.34	0.34	0.18	0.60
4	0.26	0.12	0.27	0.09	0.43
5	0.37	0.25	0.32	0.41	0.58
6	0.27	0.27	0.30	0.35	0.45
7	0.23	0.15	0.29	0.44	0.35
8	0.31	0.15	0.23	0.02	0.35
9	0.66	0.53	0.46	0.28	0.73
10	0.58	0.30	0.40	0.03	0.65
11	0.50	0.31	0.31	0.19	0.51
12	0.52	0.33	0.33	0.27	0.60

Table 6.3. PSNR results obtained under high turbidity condition.

Method / Image	AGCWD	MCRCR	WIGF	Phaseformer	Proposed
1	12.94	10.70	18.74	12.37	13.94
2	13.84	11.04	17.11	14.13	14.70
3	9.85	7.21	14.09	12.82	12.21
4	7.42	7.74	12.27	15.28	8.01
5	12.84	12.06	17.51	14.35	13.67
6	9.56	8.34	18.89	16.21	9.09
7	8.64	8.07	17.35	12.84	7.33
8	12.75	9.81	16.31	15.04	14.14
9	14.47	11.12	16.88	12.56	17.09
10	13.42	12.27	17.66	14.16	16.34
11	13.23	13.06	16.53	14.28	14.50
12	12.00	10.12	17.02	15.18	13.47

Table 6.4. SSIM results obtained under high turbidity conditions.

Method / Image	AGCWD	MCRCR	WIGF	Phaseformer	Proposed
1	0.41	0.15	0.33	−0.15	0.56
2	0.65	0.40	0.60	−0.01	0.74
3	0.55	0.30	0.47	0.13	0.66
4	0.30	0.15	0.28	0.07	0.43
5	0.27	0.16	0.30	0.25	0.54
6	0.17	0.10	0.28	0.27	0.37
7	0.13	0.06	0.24	0.21	0.29
8	0.26	0.11	0.25	−0.10	0.42
9	0.67	0.42	0.51	0.21	0.76
10	0.55	0.29	0.43	−0.03	0.69
11	0.45	0.28	0.31	0.06	0.56
12	0.47	0.29	0.33	0.15	0.59

The results obtained under high turbidity conditions were as follows (as is presented in **Table 6.3** and **Table 6.4**)

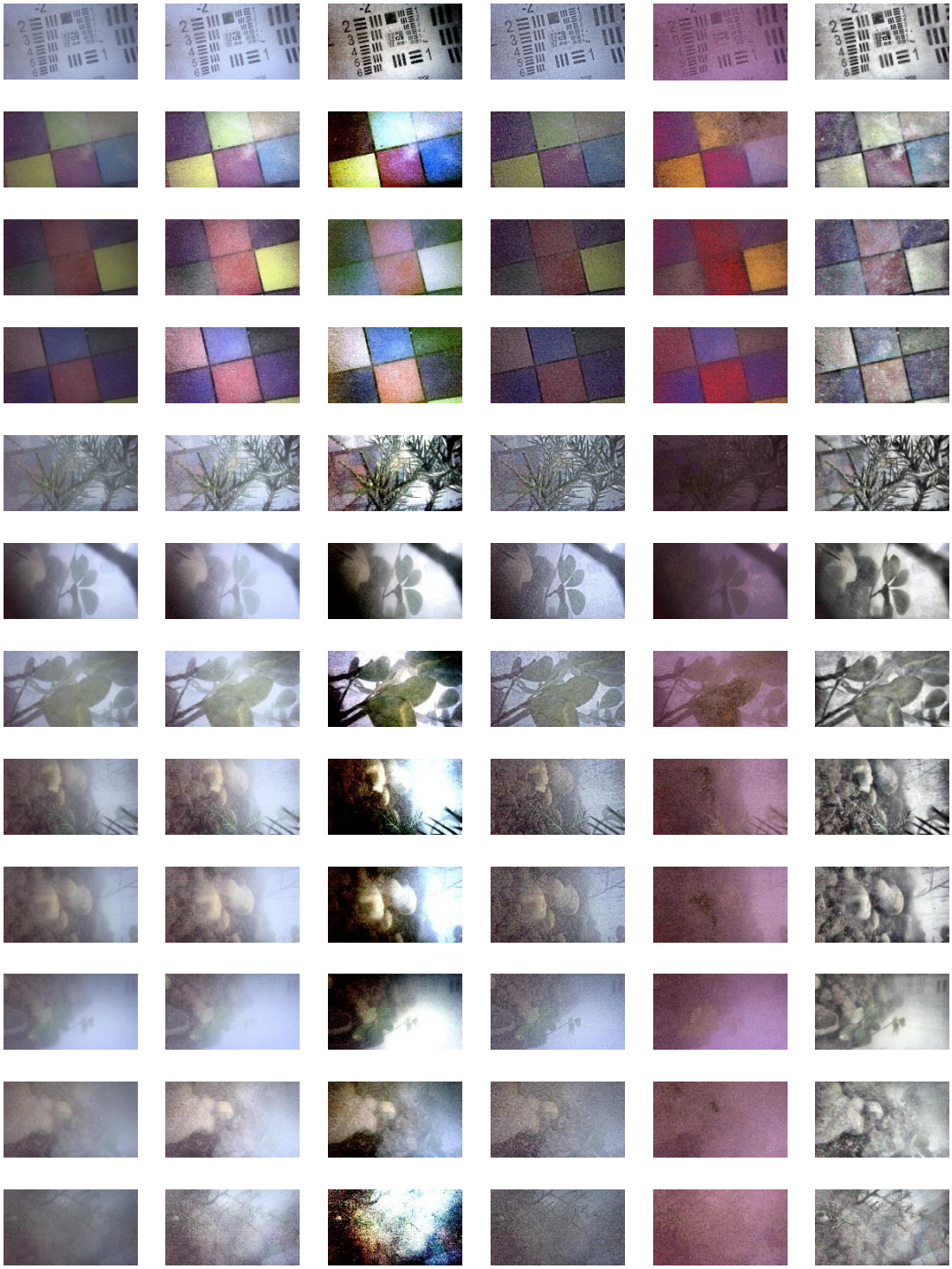
- PSNR Metric: The WIGF algorithm exhibited superior performance in most images, achieving 18.74 dB in Group 1. The proposed method yielded stable results, peaking at 17.09 dB in Group 9.
- SSIM Metric: The proposed algorithm clearly surpassed the other algorithms, particularly in Groups 2 (0.74) and 9 (0.76), demonstrating a strong structural detail preservation ability under severe conditions.

Overall, the proposed method excelled in terms of structural similarity (SSIM) but slightly lags behind the WIGF method regarding the detail recovery performance (PSNR).

Subjective Evaluation

The results obtained under moderate turbidity conditions are presented in **Figure6.3.**

- Calibration Chart: The proposed method significantly improved clarity, accurately restoring fine details and text. In contrast, the other methods exhibited evident color distortion or excessive noise.
- Color Blocks: The proposed method yielded natural colors with superior detail representation, outperforming the AGCWD and UWCNN methods.
- Plants: The proposed method demonstrated excellent clarity in leaf edges and textures compared to blurred or color-distorted output by the other algorithms.
- Mixed Rocks and Plants: The proposed method maintained optimal contrast and color fidelity.
- Rocks: The proposed method could clearly restore images, achieving effective noise reduction.
- Highly Turbid Objects: The proposed method demonstrated strong detail recovery capability and natural visuals.



(a) Raw (b) AGCWD (c) MSRCR (d) WGIF (e) Phaseformer (g) Proposed

Figure6.3. Experimental results obtained under moderate turbidity conditions.

The results obtained under high turbidity conditions are presented in **Figure6.4**, and they were as follows.

- Calibration Chart: The proposed method could recover clearer details with appropriate brightness, overcoming the dimness in the AGCWD and WIGF methods and the significant color bias in the UWCNN method. The MCRCR method generated severe noise.
- Color Blocks: The proposed method achieved accurate color restoration with clear boundaries, superior to the noise-heavy results of the MCRCR method and color-distorted images generated by the UWCNN method.
- Plant Images (Rows 3 and 4): The proposed method achieved superior recovery of leaf textures and edges, whereas the other methods' results were either excessively blurry or heavily distorted.
- Mixed Rocks and Plants: The proposed method achieved better edge definition and color accuracy than the other algorithms, which suffered from noise and color distortion.
- Extreme Turbidity Conditions: The proposed method consistently demonstrated clear detail visibility, significantly reduced noise, and superior visual quality, outperforming the other methods.

The superior performance of the proposed algorithm could be attributed to precise transmission estimation and effective removal of scattered light interference, which helped to enhance the overall image clarity, detail representation, and color authenticity.

(a) Raw (b) AGCWD (c) MSRCR (d) WGIF (e) Phaseformer (g) Proposed

Figure6.4. Experimental results obtained under high turbidity conditions.

The results obtained under moderate turbidity conditions are presented in **Figure6.3.**

- Calibration Chart: The proposed method significantly improved clarity, accurately restoring fine details and text. In contrast, the other methods exhibited evident color distortion or excessive noise.
- Color Blocks: The proposed method yielded natural colors with superior detail representation, outperforming the AGCWD and UWCNN methods.
- Plants: The proposed method demonstrated excellent clarity in leaf edges and textures compared to blurred or color-distorted output by the other algorithms.
- Mixed Rocks and Plants: The proposed method maintained optimal contrast and color fidelity.
- Rocks: The proposed method could clearly restore images, achieving effective noise reduction.
- Highly Turbid Objects: The proposed method demonstrated strong detail recovery capability and natural visuals.

The results obtained under high turbidity conditions are presented in **Figure6.4**, and they were as follows.

- Calibration Chart: The proposed method could recover clearer details with appropriate brightness, overcoming the dimness in the AGCWD and WIGF methods and the significant color bias in the UWCNN method. The MCRCR method generated severe noise.

- Color Blocks: The proposed method achieved accurate color restoration with clear boundaries, superior to the noise-heavy results of the MCRCR method and color-distorted images generated by the UWCNN method.
- Plant Images (Rows 3 and 4): The proposed method achieved superior recovery of leaf textures and edges, whereas the other methods' results were either excessively blurry or heavily distorted.
- Mixed Rocks and Plants: The proposed method achieved better edge definition and color accuracy than the other algorithms, which suffered from noise and color distortion.
- Extreme Turbidity Conditions: The proposed method consistently demonstrated clear detail visibility, significantly reduced noise, and superior visual quality, outperforming the other methods.

The superior performance of the proposed algorithm could be attributed to precise transmission estimation and effective removal of scattered light interference, which helped to enhance the overall image clarity, detail representation, and color authenticity.

Conclusions

This experiment comprehensively verifies the effectiveness of the proposed transmission optimization-based underwater image restoration method, particularly highlighting its superior performance in structural preservation and visual realism. Future studies could further enhance the proposed algorithm's detail recovery capability to comprehensively improve underwater image quality.

Stage 2: Motion Blur Parameter Estimation

As mentioned before, to simulate real-world underwater low-texture environments, this study placed stones of different sizes, aquatic plants, and sand in a custom-built water tank to approximate seabed or riverbed conditions.

In addition, by introducing specific particulate matter into the water and adjusting the intensity of artificial illumination, a range of turbidity levels and lighting conditions was achieved.

The specific settings utilized in the data acquisition procedure were as follows.

- Collaborative Robot and Camera Setup: An Aubo-i5 collaborative robotic arm, equipped with a monocular camera mounted at its end-effector, was employed. The camera performed linear or swinging motions outside the water tank, which were synchronized with the robotic arm's movements.
- Speed Configuration: The robot’s motion speed was gradually increased from 20% to its maximum speed of 100%, which was approximately 0.1 m/s. Images were captured at each speed value to generate a diverse set of samples exhibiting varying degrees of motion blur, ranging from mild to severe blurring.
- Data Acquisition: For each speed setting and turbidity condition, multiple frames of underwater blurred images were captured to construct a comprehensive and diverse testing dataset.

The above-presented procedure enabled a systematic evaluation of the proposed algorithm’s performance under different levels of motion-induced blur and varying

underwater optical characteristics, allowing for a profound assessment of its adaptability and robustness.

This experiment sought to validate the performance of the proposed motion blur parameter estimation algorithm based on cepstral morphological features in complex underwater environments, particularly under dark and turbid conditions. This experiment aimed to assess the proposed algorithm's stability and accuracy in realistic underwater scenarios and provide reliable technical support for underwater robotic visual perception.

As illustrated in **Figure 6.5 and Figure6.6**, different speeds were applied in the water tank environment. For instance, a linear motion was utilized to induce image motion blur, resulting in the corresponding motion-blurred image sequences. Subsequently, the inverse spectral morphological feature method was employed to estimate the blur angle and blur length; 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% were set at 0.1 m/s.

Figure6.5. Finishing line (in a race).

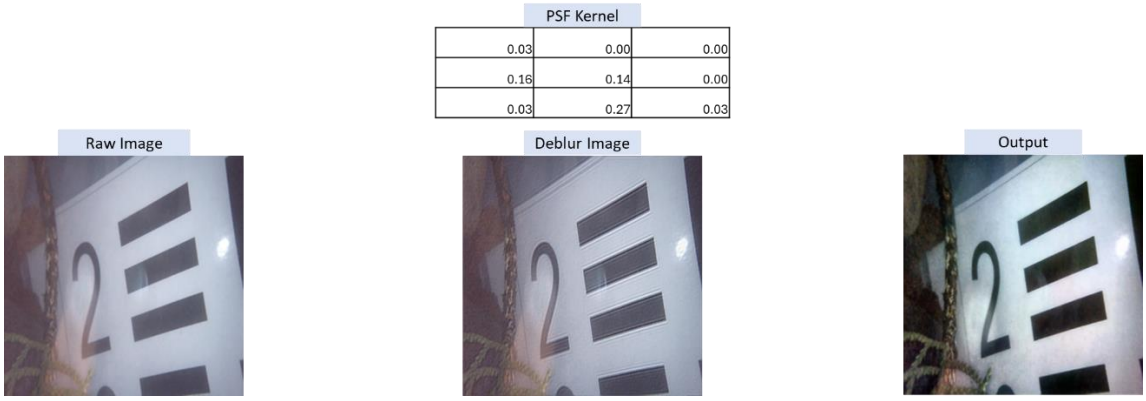


Figure6.6. Demo of motion deblur.

Experimental Methods

In this experiment, linear motion blur was generated by incrementally increasing the speed from 20% to 100%, setting a baseline speed of 0.1 m/s. Next, the captured motion-blurred image sequences were processed by different deblurring algorithms, including the UFPNET, MODEL-BASED, TRANSATT-NET, and Phaseformer algorithms (as demonstrated in **Figure 6.5**). The effectiveness and stability of each algorithm were evaluated utilizing two underwater image quality metrics, the UIQM and UCIQE.

Experimental Results

The experimental results were analyzed utilizing two metrics: UIQM and UCIQE. A detailed analysis of these two metrics is presented in the following part of this section.

UIQM Results Analysis: The UIQM results indicated significant performance differences between the algorithms as the motion speed changed, as presented in **Table 6.5**. The TRANSATT-NET algorithm consistently exhibited superior overall performance, achieving a peak UIQM value of 5.75 at the 80% speed, considerably outperforming the other algorithms. The UFPNET and MODEL-BASED algorithms demonstrated relatively

stable UIQM performance across various speed values but were consistently outperformed by the TRANSATT-NET algorithm.

UCIQE Results Analysis: Both the TRANSATT-NET and Phaseformer algorithms demonstrated high stability and accuracy within the 20%–40% speed range, with a maximum UCIQE value of 0.60. As speed exceeds 50%, a slight performance decline was observed for all algorithms, as presented in **Table 6.6**. Nevertheless, the TRANSATT-NET algorithm maintained a superior performance compared to the other algorithms.

Overall, the proposed cepstral morphology-based blur parameter estimation algorithm achieved a significantly better performance in estimating blur length compared to the other blur angle estimation methods in underwater conditions, demonstrating commendable stability and accuracy.

Table 6.5. UIQM results comparison.

Speed(%)	UFPNET	MODEL-BASED	TRANSATT-NET	Phaseformer
20	3.59	4.13	5.55	5.08
30	3.65	4.22	5.02	5.00
40	3.60	4.16	4.87	4.88
50	3.65	4.18	4.66	4.75
60	3.59	4.13	4.81	4.85
70	3.63	4.19	5.00	5.03
80	3.58	4.12	5.75	4.85
90	3.59	4.13	5.04	4.91
100	3.58	4.13	4.80	4.79

Table 6.6. UCIQE results comparison.

Speed(%)	UFPNET	MODEL-BASED	TRANSATT-NET	Phaseformer
20	0.41	0.42	0.59	0.59
30	0.40	0.41	0.60	0.59
40	0.40	0.41	0.59	0.58
50	0.41	0.42	0.58	0.57
60	0.40	0.41	0.59	0.58
70	0.40	0.41	0.59	0.58
80	0.41	0.42	0.59	0.58
90	0.40	0.41	0.58	0.58
100	0.41	0.42	0.58	0.58

Conclusions

This experiment confirms that the proposed motion blur parameter estimation algorithm, based on cepstral morphological features, can achieve stable and reliable blur estimation in complex underwater environments, particularly excelling in blur length estimation tasks. The findings of this study could provide a robust theoretical and experimental foundation for underwater robotic visual navigation and perception, validating the proposed algorithm's adaptability and practical potential for real-world underwater applications.

Discussion

The experimental validation, which included three sequential perception stages (i.e., the image restoration, motion blur parameter estimation, and camera pose estimation stages), provides a comprehensive assessment of the performance of the proposed underwater visual localization method in challenging dark and turbid environments.

The specific conclusions made based on the different stages' results are as follows.

Stage 1: The image restoration results demonstrated that dual transmittance estimation, guided by boundary constraints and local contrast priors, could effectively mitigate visibility degradation caused by light scattering and wavelength-dependent attenuation. The adaptive ambient light correction could effectively address color bias, producing visually natural and structurally enhanced outputs. Quantitatively, the proposed approach outperformed five state-of-the-art enhancement methods in terms of the UIQM, PSNR, and SSIM metrics, achieving average PSNR and SSIM values of 18.82 dB and 0.59, respectively; these results confirmed the proposed method's ability to preserve both color fidelity and structural integrity even in severely turbid scenarios;

Stage 2: The motion blur parameter estimation process focused on the frequent motion-induced degradation found in dynamic underwater scenes. The proposed cepstrum-morphology-based estimation method, which incorporated the Hanning window and least squares ellipse fitting in the frequency domain, was demonstrated to be particularly effective for small-scale blurs (typically less than 5 px), which were challenging for conventional Radon- and Hough-based estimators. The results of comparative experiments confirmed a superior accuracy of the proposed method in estimating both blur angle and length, particularly in low-texture and low-contrast underwater images, where competing methods could often fail because of feature ambiguity.

Across the two stages, a clear complementary effect was observed; namely, the image restoration module significantly improved contrast and texture visibility, and the motion deblurring step ensured sharper feature contours, allowing for more reliable

correspondence matching. Therefore, integrating physically informed priors (transmittance and blur kernel estimation) with learning-guided or geometry-guided models can ensure that each module reinforces the next.

Collectively, the findings demonstrate that visual degradation in underwater environments, which are considered a fundamental barrier to monocular vision-based localization, can be effectively mitigated through a systematic hybrid approach, including enhancement, blur compensation, and pose modeling. The proposed tri-stage approach offers a viable and lightweight alternative to conventional sensor-heavy SLAM systems, with direct implications for embedded deployment in AUVs, robotic manipulators, and inspection systems.

Future studies could focus on combining the IMU and sonar signals into a visual backbone to handle scale drift and severe occlusions, as well as on extending the proposed system to multi-frame and video-based contexts, thus enabling temporally consistent pose inference under dynamic environmental conditions.

Conclusion

This chapter performs a comprehensive experimental validation of the proposed image-based underwater localization method, which includes three progressive modules: image restoration, motion blur parameter estimation, and camera pose estimation. The results of the experiments conducted in simulated scenarios and real-world underwater robotic environments verify the effectiveness, robustness, and inter-module synergy of the proposed method.

In Stage 1 of the proposed method, the dual transmittance estimation algorithm, enhanced by adaptive ambient light correction and guided filtering, ensures superior performance in enhancing visual quality under turbid conditions. An extensive evaluation conducted with the UIQM, PSNR, and SSIM metrics also confirms that the proposed method can significantly outperform the state-of-the-art enhancement techniques, accurately restoring both perceptual clarity and structural detail essential for subsequent perception tasks.

In Stage 2 of the proposed method, the cepstrum-morphology-based blur estimation method helps to accurately recover motion blur parameters even in low-texture underwater scenes. In addition, the frequency-domain morphological analysis with least squares ellipse fitting provides reliable estimates of blur angle and length for small-scale degradation, enabling effective compensation of image distortion before localization.

Consequently, the staged experiments validate that the method of integrating physics-guided restoration, motion-aware enhancement, and geometry-driven localization can offer a robust and interpretable solution for underwater navigation. The modular design ensures flexibility and the tight coupling between vision modules enables reliable perception even in dark and visually degraded underwater environments.

The results lay a solid foundation for future real-time deployment of embedded robotic systems and for extending the pipeline to support full visual-inertial navigation and long-term autonomous underwater operations.

CHAPTER SEVEN

CONCLUSION AND FUTURE WORK

Conclusion

This dissertation presents a comprehensive exploration of robust image-based localization methods for AUVs operating in dark and turbid environments, addressing the intertwined challenges of severe optical degradation, motion-induced blur, and limited environmental features.

First, a thorough investigation in terms of underwater optical characteristics is conducted. A simplified underwater imaging model based on the Beer–Lambert law is designed and refined to consider complex illumination conditions. This model can capture the essential physical degradation mechanisms, such as wavelength-dependent absorption and scattering, and provide a solid theoretical foundation for subsequent image enhancement and localization tasks.

Building upon this, an innovative hybrid underwater image enhancement method is developed by combining dual transmission estimation and transformer-based feature fusion techniques. In addition, by integrating boundary-constrained and local-contrast methods coupled with adaptive ambient light estimation, the proposed method can effectively compensate for wavelength-dependent attenuation and color distortions. The enhanced images generated by the proposed method demonstrate significantly improved structural clarity and color fidelity across multiple benchmark datasets compared to conventional enhancement and recent deep learning-based methods.

Further, a cepstrum morphological feature-based motion blur parameter estimation method is introduced to further address dynamic image degradation, particularly motion blur induced by erratic underwater flows and robotic movements. This method combines the frequency domain analysis with the least squares ellipse fitting approach to accurately estimate blur angles and scales, which has been proven to be especially effective for small-scale blurs, which denote a common yet previously under-addressed challenge in underwater imaging.

Recognizing that reliable pose estimation remains pivotal for underwater navigation, this dissertation introduces a single-image position measurement system based on an advanced parallel perspective error propagation model. This system ensures robust initial pose estimation and incorporates line segment error modeling to enhance resistance against proximity-induced noise and feature scarcity, demonstrating superior stability and accuracy in both simulated and real-world data tests.

Finally, the proposed methods are systematically verified by controlled underwater experiments. The visual perception and motion blur estimation are integrated into a cohesive experimental platform, and results consistently demonstrate the effectiveness and generalizability of the proposed methods under varying turbidity levels and illumination complexities.

In summary, this dissertation advances the field of underwater visual perception by providing an end-to-end solution tailored for dark, turbid, and dynamically challenging environments. The proposed solution leverages physical imaging modeling to realize image enhancement, blur correction, and pose estimation. The proposed methods fill

currently critical gaps in the fields of underwater image restoration and localization and lay a strong foundation for future developments in intelligent autonomous underwater systems.

Future Studies

Although substantial advancements have been achieved by the proposed methods, several promising directions remain for future research.

First, the underwater optical imaging model could be further developed by incorporating additional complex phenomena, such as temporal variations in water turbidity, non-uniform and multi-source illumination patterns, and bioluminescent interference. In addition, advanced numerical simulations and data-driven approaches could facilitate more comprehensive and accurate predictions of underwater image degradation.

Second, although the proposed transformer-based hybrid enhancement framework achieves excellent performance under varying conditions, computational efficiency and real-time applicability on embedded platforms remain critical challenges. Therefore, future research could explore utilizing lightweight transformer architectures and semi-supervised or unsupervised training paradigms to further improve generalization capabilities in diverse and data-scarce underwater environments.

Third, addressing non-uniform motion blur induced by complex robotic maneuvers remains a pertinent problem in the field. Hence, future studies could extend the cepstrum morphology-based blur estimation approach by integrating spatially varying blur models,

possibly coupled with IMU data, to ensure more accurate and reliable deblurring results under complex underwater motion dynamics.

Further, while the monocular pose estimation method achieves good performance, integrating multi-sensor fusion strategies (e.g., combining visual data with sonar, depth, and IMU sensors) could be a valuable research direction. Moreover, investigating tightly coupled sensor fusion frameworks could further enhance localization accuracy and robustness, particularly in environments with sparse visual features or severe optical degradation.

Furthermore, large-scale field experiments could be conducted in diverse, open-water environments to comprehensively assess and validate the developed methods. Therefore, future research should aim to conduct extensive, long-duration deployments and real-world evaluations involving complex operational scenarios and mission profiles. This could significantly enhance the scalability, adaptability, and reliability of underwater robotic perception systems.

Finally, pursuing the aforementioned promising research directions could advance the field toward more intelligent, robust, and practically deployable autonomous underwater navigation and manipulation systems, thus further extending the capabilities and potential applications of AUVs in complex aquatic environments.

LIST OF REFERENCES

Akkaynak, D., & Treibitz, T. (2019). Sea-thru: A method for removing water from underwater images. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Alexandris, C., Papageorgas, P., & Piromalis, D. (2024). Positioning Systems for Unmanned Underwater Vehicles: A Comprehensive Review. *Applied Sciences*, *14*(21), 9671.

Anwar, S., & Li, C. (2020). Diving deeper into underwater image enhancement: A survey. *Signal Processing: Image Communication*, *89*, 115978.

Ballard, R. (2014). Why we must explore the sea. *Smithsonian Magazine*.

Bautista, N., Gutierrez, H., Inness, J., & Rakoczy, J. (1934). Precision landing of a quadcopter drone by smartphone video guidance sensor in a GPS-denied environment. *Sensors (Basel)*, *23*(4). https://doi.org/10.3390/s23041934

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In *European conference on computer vision* (pp. 213-229). Cham: Springer International Publishing.

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., ... & Gao, W. (2021). Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12299-12310).

Chen, L., Zhang, J., Lin, S., Fang, F., & Ren, J. S. (2021). Blind deblurring for saturated images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6308-6316).

Chen, K., Snavely, N., & Makadia, A. (2021). Wide-baseline relative camera pose estimation with directional learning. In *2021 IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)* (pp. 3257–3267).

Chen, S., Li, X., Wang, Z., & Prisacariu, V. A. (2022). DFNet: Enhance absolute pose regression with direct feature matching. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.), *Comput. Vision – ECCV 2022* (pp. 1–17). Springer Nature Switzerland.

Ding, Z. (2023). *Research on degradation mechanism of underwater optical image quality* [Master's thesis, Xi'an Institute of Optics and Precision Mechanics of CAS]. CAS Dissertation Database. CSTR:35001.37.02.33139.20220009 (in Chinese)

Cotta, J. L. S., Gutierrez, H., Bertaska, I. R., Inness, J. P., & Rakoczy, J. (2024). High-altitude precision landing by smartphone video guidance sensor and sensor fusion. *Drones*, *8*(2), 37. https://doi.org/10.3390/drones8020037

Dementhon, D. F., & Davis, L. S. (1995). Model-based object pose in 25 lines of code. *Int. J. Comput. Vision*, *15*(1), 123–141. https://doi.org/10.1007/BF01450852

Drews, P., Nascimento, E., Moraes, F., Botelho, S., & Campos, M. (2013). Transmission estimation in underwater single images. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 825-830).

Fang, Z., Wu, F., Dong, W., Li, X., Wu, J., & Shi, G. (2023). Self-supervised non-uniform kernel estimation with flow-based motion prior for blind image deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18105-18114).

Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T., & Freeman, W. T. (2006). Removing camera shake from a single photograph. In *Acm Siggraph 2006 Papers* (pp. 787-794).

Fu, J., Zhu, K., Mohsan, S. A. H., & Li, Y. (2023). Channel model and signal-detection algorithm for the combined effects of turbulence and link misalignment in underwater optical massive mimo systems. *Journal of Marine Science and Engineering*, *11*(3), 547.

Galdran, A., Pardo, D., Picón, A., & Alvarez-Gila, A. (2015). Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*, *26*, 132-145.

Gao, H., Tao, X., Shen, X., & Jia, J. (2019). Dynamic Scene Deblurring with Parameter Selective Sharing and Nested Skip Connections. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Gong, X., Lv, Y., Xu, X., Wang, Y., & Li, M. (2021). Pose estimation of omnidirectional camera with improved EPnP algorithm. *Sensors (Basel)*, *21*(12), 4008. https://doi.org/10.3390/s21124008

Guo, Y., Li, H., & Zhuang, P. (2019). Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE Journal of Oceanic Engineering*, *45*(3), 862–870.

He, K., Sun, J., & Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(12), 2341–2353.

Henderson, J. (2013). Mapping submerged archaeological sites using stereo-vision photogrammetry. *International Journal of Nautical Archaeology*, *42*(2), 243–256.

Hmam, H., & Kim, J. (2010). Optimal non-iterative pose estimation via convex relaxation. *Image Vision Comput*, *28*(11), 1515–1523. https://doi.org/10.1016/j.imavis.2010.03.005

Horaud, R., Dornaika, F., & Lamiroy, B. (1997). Object pose: The link between weak perspective, paraperspective, and full perspective. *Int. J. Comput. Vision*, *22*(2), 173–189. https://doi.org/10.1023/A:1007940112931

Hore, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. *SSIM." 2010 20th International Conference on Pattern Recognition*.

Huang, S.-C., Cheng, F.-C., & Chiu, Y.-S. (2012). Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Transactions on Image Processing*, *22*(3), 1032–1041.

Igbinenikaro, O. P., Adekoya, O. O., & Etukudoh, E. A. (2024). A comparative review of subsea navigation technologies in offshore engineering projects. *International Journal of Frontiers in Engineering and Technology Research*, *6*(02), 19–34.

Islam, M. J., Xia, Y., & Sattar, J. (2020). Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, *5*(2), 3227–3234.

Jaffe, J. S. (1990). Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, *15*(2), 101–111.

Ji, H., & Liu, C. (2008). Motion blur identification from image gradients. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jia, J. (2007). Single image motion deblurring using transparency. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jiang, Q., Pan, J., Yang, J., & Dai, Y. (2020). Learning Blind Motion Deblurring with Neural Motion Estimation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 10213–10222. https://doi.org/10.1109/CVPR42600.2020.01023

Jin, Z., & Yang, X. (2010). Analysis of a new variational model for multiplicative noise removal. *Journal of Mathematical Analysis and Applications*, *362*(2), 415–426.

Johannsson, H., Kaess, M., Englot, B., Hover, F., & Leonard, J. (2010, October). Imaging sonar-aided navigation for autonomous underwater harbor surveillance. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4396-4403). IEEE.

Kalaitzakis, M. (2021). Fiducial markers for pose estimation: Overview, applications and experimental comparison of the artag, apriltag, aruco and stag markers. *J. Intell. Robot. Syst*, *101*(4), 71. https://doi.org/10.1007/s10846-020-01307-9

Ke, K. (2022). Fusion-based restoration method for uneven illumination images in deep sea. *International Conference on Optical and Photonic Engineering, 12550*.

Khan, M. D. R. (2025). Phaseformer: Phase-based attention mechanism for underwater image restoration and beyond. *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV*.

Khan, S. (2022). Transformers in vision: A survey. *ACM computing surveys (CSUR*, *54*(10s), 1–41.

Kupyn, O., Martyniuk, T., Wu, J., & Wang, Z. (2018). DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kupyn, O., Martyniuk, T., Wu, J., & Wang, Z. (2019). DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. *IEEE International Conference on Computer Vision (ICCV)*.

Lee, C.-M. (2005). Underwater navigation system based on inertial sensor and doppler velocity log using indirect feedback kalman filter. *International Journal of Offshore and Polar Engineering*, *15*(02).

Lee, Z. (1994). *Visible-infrared remote sensing model and applications for ocean waters*. University of South Florida.

Leonard, J. J., & Durrant-Whyte, H. F. (2012). *Directed sonar sensing for mobile robot navigation* (Vol. 175). Springer Science & Business Media.

Lepetit, V., Moreno-Noguer, F., & Fua, P. (2009). EPnP: An accurate O(n) solution to the PnP problem. *Int. J. Comput. Vision*, *81*(2), 155–166. https://doi.org/10.1007/s11263-008-0152-6

Levin, A., Fergus, R., Durand, F., & Freeman, W. T. (2007). Deconvolution using natural image priors. *ACM Transactions on Graphics (TOG)*, *26*(3), 10.

Levin, A., Fergus, R., Durand, F., & Freeman, W. T. (2009). Understanding and evaluating blind deconvolution algorithms. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2009.5206818

Levin, A., Weiss, Y., Durand, F., & Freeman, W. T. (2011). Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(12), 2354–2367.

Li, C. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, *29*, 4376–4389.
Li, C. (2021). Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, *30*, 4985–5000.
Li, C., Anwar, S., & Porikli, F. (2020). Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, *98*, 107038.
Li, C., Guo, J., Pang, Y., Chen, S., & Wang, J. (2016). *Single underwater image restoration by blue-green channels dehazing and red channel correction*. IEEE ICASSP.
Li, S., Xu, C., & Xie, M. (2012). A robust O(n) solution to the perspective-n-point problem. *IEEE Trans. Pattern Anal. Mach. Intell*, *34*(7), 1444–1450. https://doi.org/10.1109/TPAMI.2012.41
Li, X. U., Guiming, L. U., & Zhenguang, Q. I. U. (2022). Adaptive Retinex Algorithm Based on Detail Selection Used in Underwater Image Enhancement. *Journal of Computer Engineering & Applications*, *58*(11).
Li, Z. (2014). Weighted guided image filtering. *IEEE Transactions on Image Processing*, *24*(1), 120–129.
Li, Z., Dong, J., Xiong, W., Jiang, Q., & Pan, J. (2023). Real-world local motion deblurring dataset and benchmark. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 1035–1045.
Liang, J. (2021). Swinir: Image restoration using swin transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
Liu, Q. (2023). Prior Guided Multiscale Dynamic Deblurring Network for Diffraction Image Restoration in Droplet Measurement. *IEEE Transactions on Instrumentation and Measurement*, *73*, 1–14.
Liu, R. (2022). Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, *31*, 4922–4936.
Liu, S., Xu, H., Lin, Y., & Gao, L. (1889). Visual navigation for recovering an AUV by another AUV in shallow water. *Sensors (Basel*, *19*(8). https://doi.org/10.3390/s19081889
Lu, Z. (2014). Iterative reweighted minimization methods for lp regularized unconstrained nonlinear programming. *Math. Program*, *147*(1), 277–307. https://doi.org/10.1007/s10107-013-0722-4
Lucy, L. B. (1974). An iterative technique for the rectification of observed distributions. *Astronomical Journal*, *79*, 745–754.
Mallios, A., Vidal, E., Campos, R., & Carreras, M. (2017). Underwater caves sonar data set. *International Journal of Robotics Research*, *36*, 1247–1251.
McGlamery, B. L. (1980). A computer model for underwater camera systems. *Ocean Optics, VI. Vol. 208*.
Nah, S., Kim, T. H., & Lee, K. M. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
Nie, Y., & He, Z. Y. (2014). Underwater imaging and real-time optical image processing under illumination by light sources with different wavelengths. *Acta Opt. Sin.*, *34*(7), 0710002.

Panetta, K., Gao, C., & Agaian, S. (2015). Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, *41*(3), 541–551.
Park, J., Kim, M., & Chang, S. (2011). Estimation of motion blur parameters using cepstrum analysis. *IEEE International Symposium on Consumer Electronics (ISCE)*.
Qian, J., Kong, B., & Yang, J. (2023). Underwater image clarification based on double-opponency light estimation and red channel prior. *IEEE Access*, *11*, 64383–64396.
Qing, S. (2012). Remote sensing retrieval of total absorption coefficient in the Bohai Sea. *Chinese Journal of Oceanology and Limnology*, *30*(5), 806–813.
Qu, Y., & Hou, W. (2019). Attitude accuracy analysis of PnP based on error propagation theory. *Opt. Precis. Eng*, *27*(2), 479–487. https://doi.org/10.3788/OPE.20192702.0479
Rahman, S., Rahman, M. M., Abdullah-Al-Wadud, M., Al-Quaderi, G. D., & Shoyaib, M. (2016). An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, *2016*(1), 35.
Rahman, Z., Jobson, D. J., & Woodell, G. A. (1996). Multi-scale retinex for color image enhancement. *Proceedings of 3rd IEEE International Conference on Image Processing*, *3*.
Rigby, P., Pizarro, O., & Williams, S. B. (2006). Towards geo-referenced AUV navigation through fusion of USBL and DVL measurements. *OCEANS*.
Rim, J., Nah, S., Son, S., Lee, G., & Lee, K. M. (2020). Real-World Blur Dataset for Learning and Benchmarking Deblurring Algorithms. *European Conference on Computer Vision (ECCV)*, 184–201. https://doi.org/10.1007/978-3-030-58598-3_11
Sarlin, P. E. (2021). Back to the feature: Learning robust camera localization from pixels to pose. In *2021 IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)* (pp. 3246–3256).
Shan, Q., Jia, J., & Agarwala, A. (2008). High-quality motion deblurring from a single image. *ACM Transactions on Graphics (TOG*, *27*(3), 73.
Sharma, P., Bisht, I., & Sur, A. (2023). Wavelength-based attributed deep neural network for underwater image restoration. *ACM Transactions on Multimedia Computing, Communications and Applications*, *19*(1), 1–23.
Shen, Y. (2021). Underwater optical imaging: Key technologies and applications review. *IEEE Access*, *9*, 85500–85514.
Shi, M., Xu, T., & Liang, J. (2013). PSF estimation via gradient cepstrum analysis for single blurred image. *Acta Physica Sinica*, *62*(229), 238.
Snyder, J. (2010). Doppler Velocity Log (DVL) navigation for observation-class ROVs. In *OCEANS 2010 MTS/IEEE SEATTLE* (pp. 1–9).
Strudel, R. (2021). Segmenter: Transformer for semantic segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
Su, L. (2022). Heuristic EPnP-based pose estimation for underground machine tracking. *Symmetry*, *14*(2), 385. https://doi.org/10.3390/sym14020385
Sun, J. (2015). Learning a convolutional neural network for non-uniform motion blur removal. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
Sun, J. (2022). OnePose: One-shot object pose estimation without CAD models. In *2022 IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)* (pp. 6815–6824).
Swinehart, D. F. (1962). The beer-lambert law. *Journal of Chemical Education*, *39*(7), 333.

Tao, X., Gao, H., Shen, X., Wang, J., & Jia, J. (2018). Scale-recurrent network for deep image deblurring. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Uzougbo, N. S., Ikegwu, C. G., & Adewusi, A. O. (2024). Regulatory frameworks for decentralized finance (DEFI): challenges and opportunities. *GSC Advanced Research and Reviews*, *19*(02), 116–129.

Vaccarella, A. (2013). *Multimodal sensors management in computer and robot assisted surgery*.

Vakhitov, A., Colomina, L. F., Agudo, A., & Moreno-Noguer, F. (2021). Uncertainty-aware camera pose estimation from points and lines. In *2021 IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)* (pp. 4657–4666).

Wang, H., Sun, S., & Ren, P. (2023). Meta underwater camera: A smart protocol for underwater image enhancement. *ISPRS Journal of Photogrammetry and Remote Sensing*, *195*, 462–481.

Wang, Y. (2019). An experimental-based review of image enhancement and image restoration methods for underwater imaging. *IEEE Access*, *7*, 140233–140251.

Wang, Z. (2004). Image quality assessment: Form error visibility to structural similarity. *IEEE Trans. Image Process*, *13*(4), 604–606.

Wang, Z. (2022). Uformer: A general U-shaped transformer for image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR*.

Wiener, N. (1964). *Extrapolation, interpolation, and smoothing of stationary time series*. The MIT Press.

Xiang, Y., Zhou, H., Li, C., Sun, F., Li, Z., & Xie, Y. (2024). Application of deep learning in blind motion deblurring: Current status and future prospects. *arXiv preprint arXiv:2401.05055*.

Xu, L., & Jia, J. (2010). Two-phase kernel estimation for robust motion deblurring. *IEEE European Conference on Computer Vision*, 157–170.

Yang, M., & Sowmya, A. (2015). An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, *24*(12), 6062–6071.

Yu, G., Cai, R., Su, J., Hou, M., & Deng, R. (2023). U-YOLOv7: A network for underwater organism detection. *Ecological Informatics*, *75*, 102108.

Zamir, S. W., Arora, A., & Khan, S. (2022). Restormer: Efficient transformer for high-resolution image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, W. (2023). Underwater image enhancement combining dual color space and contrast learning. *Optik*, *284*, 170926.

Zhang, Y. L., Wang, Y. M., & Huang, A. P. (2018). Influence of suspended particles based on Mie theory on underwater laser transmission. *Chin. J. Lasers*, *45*(05), 152–160.

Zhang, Z. X. (2012). *Study on the distribution characters of PAHs in particulate matter in typical cities of Jilin province*. Jilin University.

Zhao, J., Hu, Y., & Tian, M. (2021). Pose estimation of excavator manipulator based on monocular vision marker system. *Sensors (Basel)*, *21*(13), 4478. https://doi.org/10.3390/s21134478

Zhao, X., Jin, T., & Qu, S. (2015). Deriving inherent optical properties from background color and underwater image enhancement. *Ocean Engineering*, *94*, 163-172.

Zheng, Y., Kuang, Y., Sugimoto, S., Åström, K., & Okutomi, M. (2013). Revisiting the PnP problem: A fast, general and optimal solution. In *2013 IEEE Int. Conf. Comput* (pp. 2344–2351).

Zhou, J. (2022). Underwater image enhancement method with light scattering characteristics. *Computers and Electrical Engineering*, *100*, 107898.

Zhuang, P. (2022). Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE Transactions on Image Processing*, *31*, 5442–5455.

Zuiderveld, K. J. (1994). Contrast limited adaptive histogram equalization. *Graphics Gems*, *4*(1), 474–485.

VITA

Ning Hu was born in Beijing, China. He holds a master's degree in mechanical engineering from New York University. His research interests include automation in construction and robotics and sensing.