# Doctoral Dissertation

# Applications of Light Transport in Direct and Scattered Components

Mayuka Tsuji

Jan 30, 2024

Graduate School of Science and Technology
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Mayuka Tsuji

Thesis Committee:

   Professor Yasuhiro Mukaigawa    (Supervisor)
   Professor Hirokazu Kato      (Co-supervisor)
   Associate Professor Takuya Funatomi  (Co-supervisor)
   Assistant Professor Yuki Fujimura   (Co-supervisor)
   Assistant Professor Kazuya Kitano   (Co-supervisor)

# Applications of Light Transport in Direct and Scattered Components*

## Mayuka Tsuji

### Abstract

In computer vision, applying complex information processing to images has become common. However, in the real world, various optical phenomena exist, such as refraction, attenuation, and mutual reflection of light. Cameras project this information onto a two-dimensional image, losing much information about optical phenomena in a three-dimensional scene during the transformation process. Therefore, it is essential to focus on not only the image itself but also the optical phenomena in the image generation process. This approach, known as physics-based vision, centers around the concept of light transport from the light source to the object and its capture by the camera. The purpose of this paper is to explore new solution techniques that incorporate the idea of light transport for computer vision applications. Specifically, two applications have been pioneered. The first focuses on light transport derived from the direct component of light, leveraging geometric considerations to measure light from objects in a specific region and applying this to efficient image processing for touch sensing. The second focuses on the scattering component of the light transport. By solving the inverse problem of the Kubelka-Munk model, which models the scattering of light in the particle layer, we can estimate how long the light transport was scattered in the particle layer. This was applied to the analysis of pigments in an ancient mural painting on a heterogeneous substrate. These applications are challenging problems to solve by simply applying image processing algorithms to images captured by RGB cameras, and a solution that incorporates the concept of optical transport would advance the frontiers of computer vision.

**Keywords:**

light transport, physics-based vision, projector-camera system, spectral image, Kubelka-Munk model

# Contents

# List of Figures

# List of Tables

# 1 Introduction

Computer vision is a field of study that explores how computers can understand images. Various algorithms have been established to enable computers to perceive the real world as our vision does.

A critical aspect of computer vision is the acknowledgment that while the real world is three-dimensional, our vision and cameras project it into two dimensions. For example, in the real world, light undergoes various optical phenomena such as refraction, Mie scattering, Rayleigh scattering, mirror reflection, diffuse reflection, Fresnel reflection, and interreflection. However, when captured by a camera and integrated into two-dimensional information, these optical processes are lost, reducing to mere signal values of pixel brightness. This projection process inevitably reduces some information from the real world. Therefore, more than merely applying advanced information processing to images is needed.
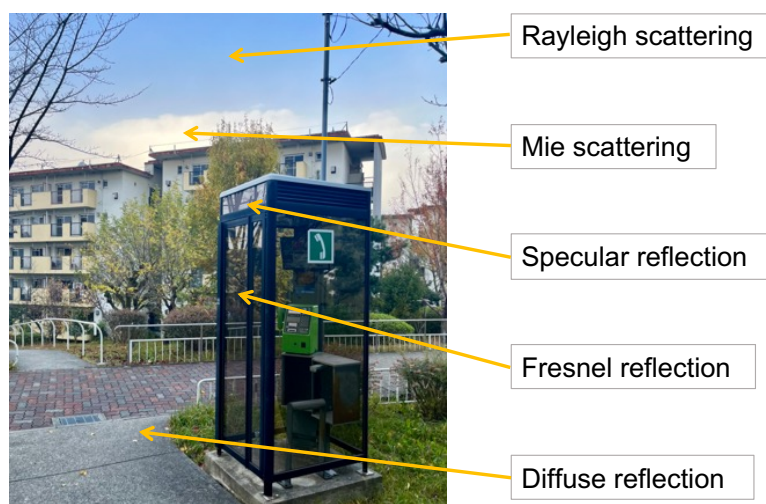


Figure 1.1: Some of the diverse optical phenomena in the scene.
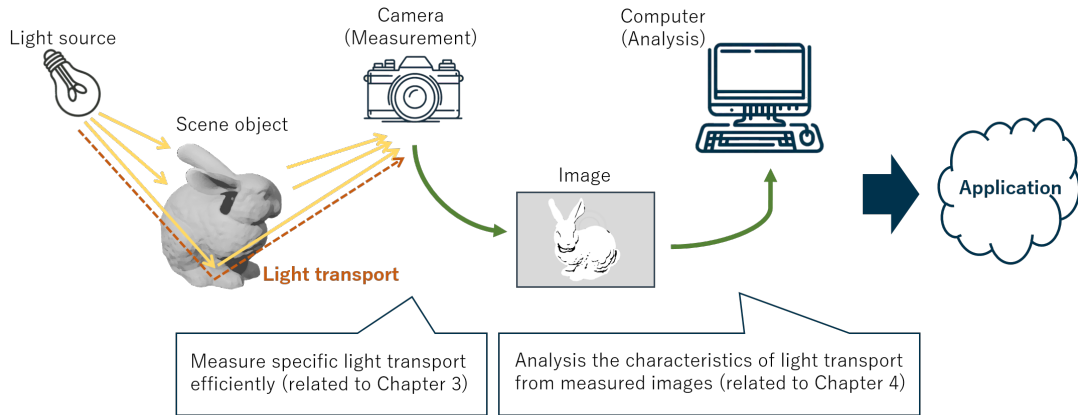
Figure 1.2: This thesis proposes computer vision applications considering light transport.

In response to this, the concept of Physics-based vision, driven by Shafer [2], Kanade [3], Ikeuchi [4, 5], Narasimhan [6, 7], and others, aims to recognize the real world with richer parameters by designing algorithms that consider real-world optical phenomena. In this context, the behavior of **Light Transport** — how light path from a source reflects off objects and reaches optical measuring instruments — becomes a vital concept.

Computer vision tasks can be broadly categorized into three stages: measurement, analysis, and application (Figure 1.2). The measurement phase involves capturing images, considering the physical properties of light and camera optics. Analysis entails interpreting these captured images to derive meaningful information while the application uses this information for various practical purposes.

This paper aims to explore new solution techniques, incorporating the idea of optical transport for computer vision applications. In particular, we focus on the optical path length of light transport. Optical path lengths can be considered on the macro-scale of objects on the order of tens of centimeters and the micro-scale of objects on the order of millimeters to address broad real-world problems effectively, and we will consider both of these scales. On the macro scale, our research zeroes in on object-position detection. Given its ubiquity and intuitiveness in human-computer interaction, We chose finger detection for touch sensing as our case study. For the latter scale, we focus on particle layer analysis, and in particular, we select pigment distribution analysis of cultural properties, which is

an active field.

Conventional finger detection for touch sensing is often performed using camera images or special infrared devices. When computer vision processing is applied to camera images, the camera generally measures the entire scene that fits within its angle of view, and the finger must be detected from an image that contains much extra information. In contrast, we achieved finger detection for touch sensing with more straightforward computer processing by measuring only the light transport reflected in a specific area where a touching finger may be present. Then, we applied a projector-camera system that efficiently measured only the "direct light" reflected from a specific area of optical path length.

Pigment distribution analysis refers to the study of which pigments are applied and at what thickness. Since it is difficult to measure the thickness of the particle layer directly, the thickness of the pigment is estimated from the measured data. Since the light incident on the particle layer causes wavelength-dependent scattering and absorption, the thickness of the pigment can be analyzed by estimating the optical path length that caused the scattering out of the light transport. This is a practical approach for cultural properties that are often difficult to measure from a conservation standpoint because it allows analysis based on previously measured data. We focused on the Kubelka-Munk model, which models light scattering in the particle layer. We solved the ill-posed inverse problem of this physical model to obtain the thickness of the pigment.

In other words, referring to Figure 1.2, finger detection for touch sensing, which handles macro-scale optical path lengths, focuses on the measurement part, and it is important how to measure the desired light transport efficiently. In the case of pigment distribution analysis, which deals with micro-scale optical path lengths, emphasis is placed on the analysis part, and it is essential to be able to recover the light transport in the scene from the measurement data. It is challenging to solve these problems simply by applying general image processing algorithms to images taken by an RGB camera, and a solution that incorporates the light transport concept will advance the frontier of computer vision.

The structure of this dissertation is organized as follows: Chapter 2 reviews the essential physical phenomena in light transport and aspects of light transport acquisition, positioning the research within this framework.

Chapter 3 delves into the study of finger detection for touch sensing by measuring light transport from a specific distance. The chapter details developing and implementing a projector-camera system to efficiently measure direct light reflection from a specific area. The chapter also discusses the implementation algorithm of this system for touch sensing and demonstrates how this approach simplifies the process of locating spatial objects using basic image processing techniques.

Chapter 4 transitions to applications focusing on the analytical aspects of scattered-light transport. It centers on the Kubelka-Munk model, a pivotal tool in modeling light scattering in particle layers. The chapter outlines the innovative approach to solving the inverse problem associated with this model, leading to groundbreaking methods for analyzing pigment distribution in ancient tomb murals.

Finally, Chapter 5 concludes the dissertation and discusses future research directions.

# 2 Theory of Light Transport

## 2.1 Physical Phenomena in Light Transport

Light transport — the path that light takes from being emitted or reflected by objects in a scene to its arrival at a camera's sensor — involves a variety of optical phenomena. This section reviews them in general. Please note that the explanations here focus on how light travels in straight lines and interacts with different surfaces and mediums. This review will omit discussions on the wave-like properties of light, such as diffraction and interference, which involve more complex aspects of light behavior.

### A. Reflection

When light emitted from a source reaches the surface of an object, part or all of it is reflected at the surface. Reflection is commonly categorized into two primary types: diffuse reflection and specular reflection.

Diffuse reflection occurs when light strikes a rough surface, scattering it in multiple directions. An ideal diffuse reflection, known as Lambertian reflection, distributes light uniformly in all directions. Specular reflection occurs when light hits a smooth surface. The surface reflects the light in a single direction while maintaining the equality of the angle of incidence and the angle of reflection.

Object surfaces exhibit both properties of diffuse and specular reflections. The dominant type of reflection influences the characteristic appearance of the object. Generally, object color analysis and shape measurements such as photometric stereo [8] ignore specular reflection components and assume only diffuse reflection components. The specular reflection component, known for its higher intensity and strong directionality compared to the diffuse component, presents significant challenges in both measurement and analysis. Therefore, studies have focused on

removing the effects of specular reflection when diffuse reflection is dominant [9, 10, 11]. Conversely, in situations where specular reflections are dominant, such as metals, special measurement techniques are often established to account for directional reflections [12, 13, 14, 15].

The consideration of light reflection in Computer Vision (CV) applications presents a multitude of diverse approaches. For instance, in the realm of industrial quality inspection, the analysis of light reflection is employed to assess the quality of products [16, 17]. Furthermore, environmental recognition in robotics heavily relies on understanding light reflection. Robots use this information to navigate, interact with their surroundings, and perform tasks more precisely and safely. Another significant application is found in the digitalization of cultural heritage. The shape and color of cultural heritage are digitized based on the reflection data of objects.

From this point onwards, this thesis assumes objects primarily exhibit diffuse reflection.

## B. Attenuation and Scattering

Light attenuation is the process by which light diminishes in intensity as it passes through mediums like air, water, or glass. For example, in a foggy environment, the light is weakened by scattering and absorption by fog particles. As a result, the light appears less bright, and the scene is less visible than in a fog-free environment. Because light attenuation, wildly scattering, complicates light transport, it has often been a challenge in CV.

The fog mentioned earlier is a typical example of forward scattering, where light or waves scatter in a direction close to that of the incident light after colliding with particles. This causes the visibility to become hazy and blurs the outlines of objects. On the other hand, there is also a phenomenon known as backward scattering, which can be observed in layers of particles such as paint. Light that does not cause specular reflection on the surface of an object penetrates inside. Within, the light undergoes wavelength-dependent absorption and scattering, resulting in diffusely reflected light with different characteristics from the incident light. In addition, translucent objects, such as skin and marble, experience internal sub-surface scattering [18, 19].

6

Examples of CV applications that take into account light attenuation include eliminating the effects of lousy weather [20, 21, 22], underwater imaging [23, 24, 25], and the design of media coatings [26, 27, 28].

## C. Refraction

Refraction is a phenomenon where the direction of light changes as it passes through different media, such as from air to water. A typical example of this phenomenon is the apparent bending of a straw when placed in water. This bending effect is due to the change in the light's direction, or refraction, which varies depending on the refractive index of the medium. The refractive index is a measure of how much a medium can bend light.

Refraction is also an essential principle in the context of cameras. The lenses in a camera utilize refraction to gather and focus light. As light passes through the curved lens of optical glass, its path is bent due to refraction. This bending of light by the lens allows the camera to capture a focused image of the scene. Different lenses have varying refractive properties, which photographers and videographers use to control how images are captured, including the focus, depth of field, and field of view.

In addition to refraction, when light encounters the boundary between two media with different refractive indices, Fresnel reflection occurs [29]. This type of reflection is particularly noticeable on transparent materials like glass and water surfaces. For instance, the reflections seen on the surface of a body of water or a glass window are examples of Fresnel reflection. This phenomenon arises due to the partial reflection of light at the boundary, which is influenced by factors such as the angle of incidence and the contrast in refractive indices of the two media.

Examples of applications that take into account light refraction include the removal of glass reflection [30, 31], correction of distortion in wide-angle lenses [32, 33], and estimation of the celestial body's apparent position [34, 35].

## D. Luminescence and Radiation

In some scenes, objects and organisms emit light energy, contributing to the broader concept of radiation. This includes thermal radiation, where all objects

emit heat, predominantly in the far-infrared spectrum, vital for applications like thermal imaging.

Bioluminescence and chemiluminescence are examples of light emission in nature and chemical processes, respectively. Bioluminescence occurs in deep-sea organisms, fireflies, and fungi, producing light through internal chemical reactions. Chemiluminescence is similar but occurs in various non-biological chemical reactions.

Phosphorescence is a captivating phenomenon in which materials absorb energy and re-emit it as light over time. It is intriguing that these materials continue to glow even after the energy source is removed due to their slower energy release.

Fluorescence is critical in CV, especially in medical imaging. It happens when substances absorb specific light wavelengths and re-emit light at longer wavelengths as they return to a stable state. This is used in cancer diagnostics, where irradiating cancer cells with specific light makes their proteins fluoresce, aiding in distinguishing them from healthy tissue [36, 37].

## 2.2 Light Transport Acquisition

### E. Active and Passive

Two primary measurement techniques are employed in image generation and analysis: active measurement and passive measurement.

Active measurement involves an integrated control of illumination within the measurement system. This technique captures reflected or altered light to form images. A typical example of active measurement is the use of structured light with projectors [38] or laser pulses in LiDAR (Light Detection and Ranging) systems [39]. This approach is beneficial in scenarios where precise depth information and detailed surface characteristics are required, as in 3D mapping and autonomous vehicle navigation.

Passive measurement utilizes external light sources and captures the light reflected from objects to form images. This method is similar to how the human eye operates, relying on ambient or existing light conditions without emitting controlled light from the measurement system. Standard cameras and most visual

imaging systems employ passive measurement techniques. Passive measurement is the preferred method in conditions where maintaining the subject's natural state is crucial or where active illumination could be intrusive.

Active measurement has a profound connection with the aspect of geometry in CV. Active measurement effectively enhances objects' geometric information, such as distance, shape, and surface characteristics. Meanwhile, an understanding of geometry aids in interpreting active measurement data, enabling more effective data collection. Passive measurement is more closely related to photometry than to geometry. Passive measurement allows for capturing the natural behavior of light, which is beneficial for understanding the characteristics of object surfaces and applications like color restoration and optical properties.

## F. Direct and Global

In addition to the classification based on the reflection from the object's surface, there are other ways to categorize reflections. In any given scene, multiple objects usually exist, and light interactions between objects occur. This results in two primary types of light transport: direct components and global components.

Light transport of direct components refers to the component of light that travels from the light source and reflects off the target object directly. This component is crucial as it defines characteristics for the primary illumination and visual of an object.

On the other hand, light transport of global components originates from the light that reflects off various surfaces within the scene, acting as indirect light sources. This light component creates subtle nuances in illumination, contributing to the realism and depth of the scene. However, global light complicates scene understanding due to its multifaceted interactions and varied paths before reaching the viewer or camera.

Given the complexity introduced by global light, studies focus on isolating the direct light component [40, 41, 1].

Also, in the study of light behavior, there are two primary perspectives: a local approach, which focuses on the behavior of light near object surfaces, and a global approach, which considers the entire scene comprehensively. The local approach, known as local illumination, models the behavior of light at the specific points

where it directly strikes objects from a light source. This method primarily deals with the direct interaction of light with surfaces, often simplifying or ignoring the complex interplay of light with the rest of the environment. On the other hand, global illumination takes a holistic view of the scene, incorporating both direct and indirect lighting in its modeling. Indirect lighting includes light that has reflected off other surfaces or has been diffused, contributing to more nuanced and realistic scene representations. Although global illumination is crucial for achieving realistic scene analysis and reproduction, it demands considerable computational resources.

## G. Geometry and Photometry

CV considering light transport can be broadly divided into two main aspects: geometry and photometry. Each aspect plays a pivotal role in capturing and analyzing images, influencing the development of various technologies and applications within the field.

Geometry in CV primarily focuses on the geometric characteristics of cameras, light sources, and objects when generating images. This involves understanding and modeling objects' spatial relationships and orientations in a three-dimensional space. The geometric aspect is crucial for object recognition, scene reconstruction, and stereoscopic imaging tasks. It helps determine objects' position, shape, and size within an environment, facilitating an understanding of the scene structure and the interaction of different elements within it.

Photometry in CV centers on image generation's optical properties, such as brightness and reflection characteristics. This aspect deals with how light interacts with objects and the environment to create an image's visual appearance. The photometric analysis includes understanding the distribution of light and color within an image, the way surfaces reflect light (diffuse and specular reflections), and how these properties affect the perception of the image. Photometry is essential for enhancing image quality, color correction, and understanding the illumination of scenes.

## 2.3 Overview

In general, the above should be considered when thinking about light transport, and depending on which optical phenomena to focus on and which aspects to pay attention to, there is a wide range of applications.

Chapter 3 will focus on reflected light, particularly in light transport in specific direct components. This chapter will focus on geometries that include active lighting.

Chapter 4 deals with applications regarding scattered light, exceptionally light transport in the particle layer below the object's surface. This chapter assumes a collimated light source, i.e., passive lighting, and focuses on photometry.

# 3 Application of Light Transport in Direct Component

This chapter explores applications considering light transport in macro-scale optical path lengths on the order of tens of centimeters. In the study of macro-scale optical pathlengths, special devices such as ToF cameras and LiDARs are often used to acquire the spatial position of objects, and critical applications in computer vision include automated driving technology [42] and NLoS, which recovers hidden objects that are not visible from cameras [43]. Focusing on spatial position detection, we consider finger detection for touch sensing, the most intuitive and common form of human-computer interaction, with a standard camera instead of the above special devices. Among touch sensing, we consider touch sensing for a projected image, which does not have a sensing mechanism on the touch plane itself (Figure 3.1).

Previous research in touch sensing with the camera focused on extracting fingertip positions from camera images through object segmentation methods. Nonetheless, touch sensing with cameras encounters two primary challenges. The first is the interference of the projected image with finger localization in the camera image. When the finger touches the projected screen, the overlapping projected image can alter its apparent color and contour. The second challenge is simultaneous touch detection and fingertip localization from a single camera image, as typical cameras reduce a three-dimensional scene to a two-dimensional plane.

The key observation here is that in scenes where a person is touching the screen with their finger, elements other than the touching finger are not necessary information for touch sensing (Figure 3.2). In applications where the optical path length of optical transmission is considered, the global light is usually measured
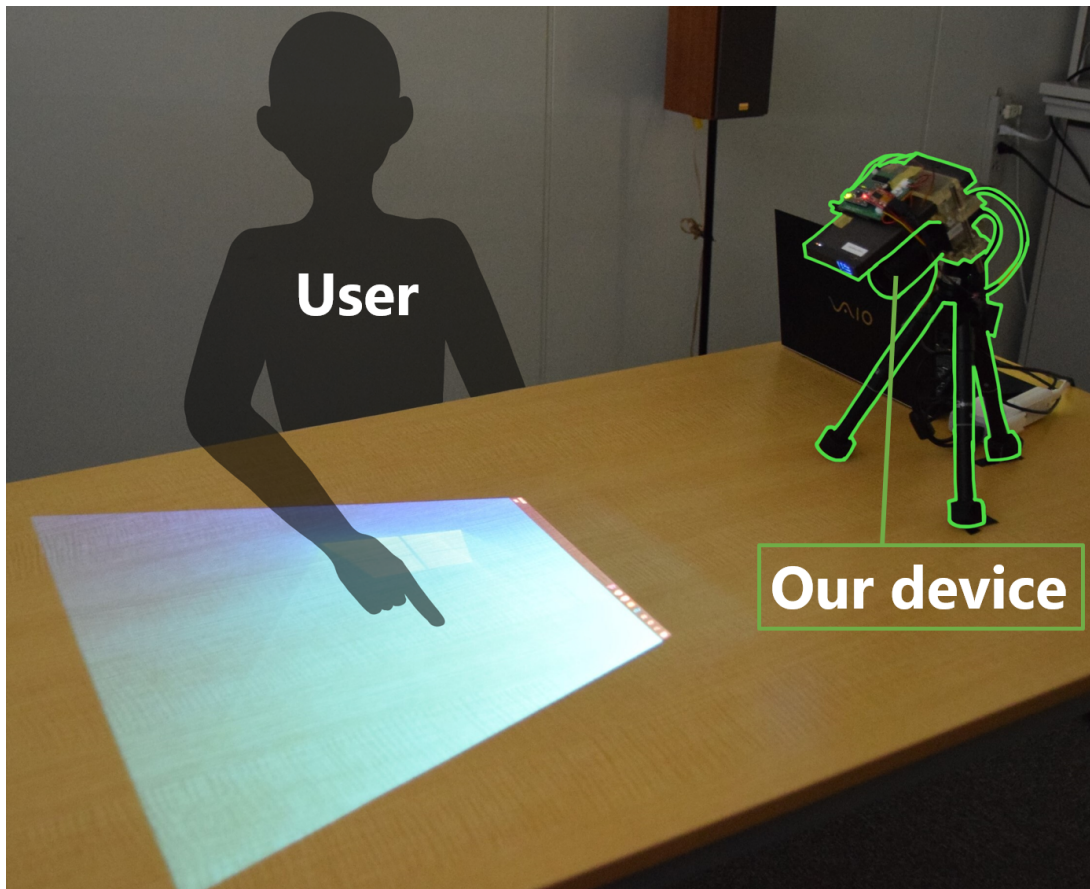
Figure 3.1: Application of touch sensing.

simultaneously. If the optical path length of the indirect light coincides with the optical path length of the space of the purpose, an error is introduced in the estimation. Therefore, instead of conventionally capturing the entire scene with a camera, we try to acquire only the light transports of direct components originating from a specific region necessary for touch sensing to eliminate excess information and make touch sensing more feasible with a simpler algorithm. Controlling the space also eliminates indirect light when obtaining the desired optical path length.

In order to obtain the light transport of direct components in a specific area, active lighting (described in Chapter 2), which controls the light source, is essential. We mainly focus on a research segment to explore how to efficiently transmit
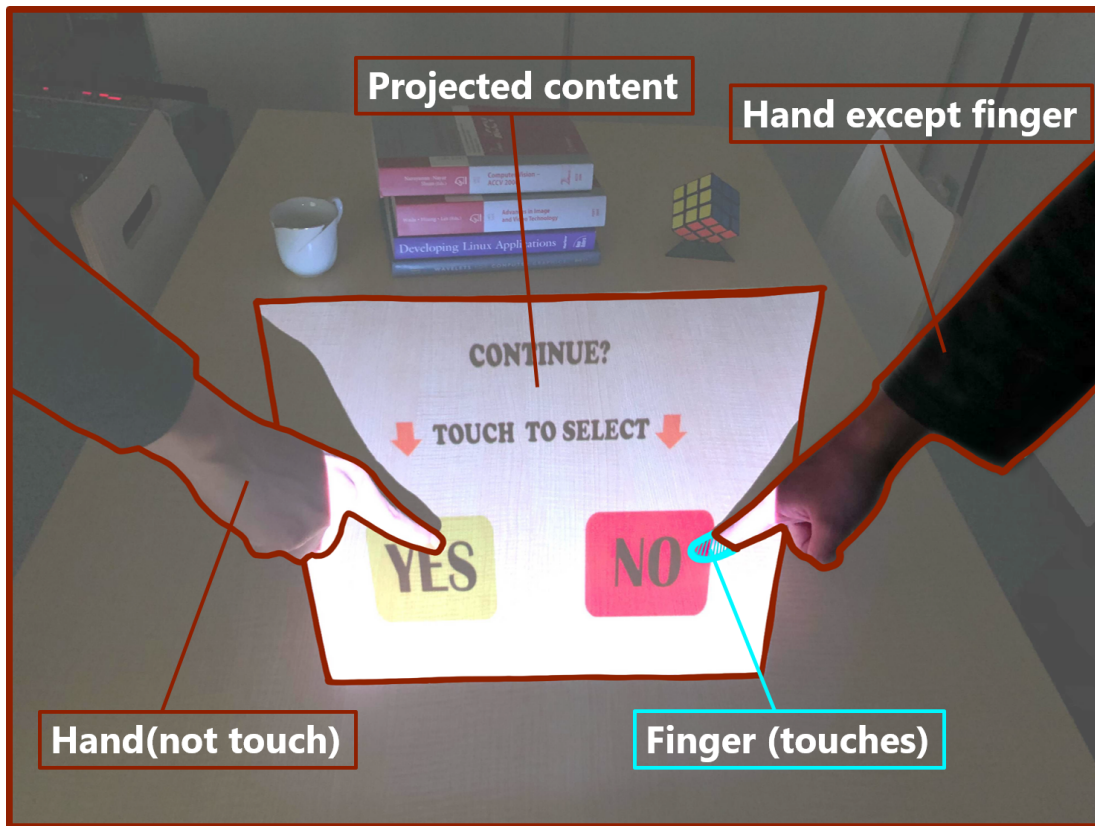
Figure 3.2: There is a lot of information that is not needed to detect a finger.

light from the source to the scene or effectively measure the light that travels from the scene to the camera. Notably, research focuses on controlling the light source and the camera to measure only the reflected light in a specific space, thereby maximizing energy efficiency, called disparity gating [44].

In particular, our study focused on a setup of disparity gating that aligns a synchronized projector-camera in a vertical configuration; Slope Disparity Gating [45]. Slope Disparity Gating is a system that controls various parameters such as imaging distance, the thickness of the imaging volume, and the tilt of the imaging plane by altering the camera's hardware parameters. This system allows for the selective imaging of specific areas within a scene while excluding and disregarding other areas.

## 3.1 Disparity Gating with Projector-camera System

### 3.1.1 Acquisition of Efficient Light Transport

Disparity gating is a concept in computer vision that focuses on capturing images from specific light transport [44]. Imagine you have a camera that can take pictures not of everything it sees but only of objects at a certain distance. This is what gated imaging achieves, like having a selective focus that can ignore unwanted background or foreground clutter. This idea is used in scenes with backscatter, such as in foggy or dusty environments [46, 47, 1], or for excluding indirect light and focusing solely on direct light transport [40, 48, 49]. Disparity gating generally uses laser light for active lighting. In this chapter, we focus on gated imaging using a projector-camera system.

To understand the projector-camera system for disparity gating, we first look at the stereo vision for triangulating (Figure 3.3a). It is difficult to know the geometric characteristics of the scene with a single camera. Therefore, stereo vision involves using two cameras, similar to our two eyes, to perceive a 3D scene. The relative position of the two cameras delimit the area in which an object is located (epipolar geometry), and is mainly used to determine the spatial position of an object. However, this passive sensing technique depends on the ambient light source.

If replacing one of these cameras with a projector (active light source); this is the basis of a projector-camera system (Figure 3.3b). Knowing the exact spatial relationship between the camera and the light source and controlling the light to a specific location facilitates selectively capturing light that travels a particular path from the source to the camera.

O'Toole *et al.* [1] devised a disparity gating technique based on a projector-camera system, where the camera and laser scanning projector are aligned horizontally and synchronized. As illustrated in Fig 3.4, in the method conceived by O'Toole *et al.*, the laser scanning projector does not project the entire image at once but sequentially emits light along horizontal lines. Similarly, the camera's rolling shutter does not expose the entire sensor simultaneously but exposes one

(a) Stereo vision  (b) Projector-camera system
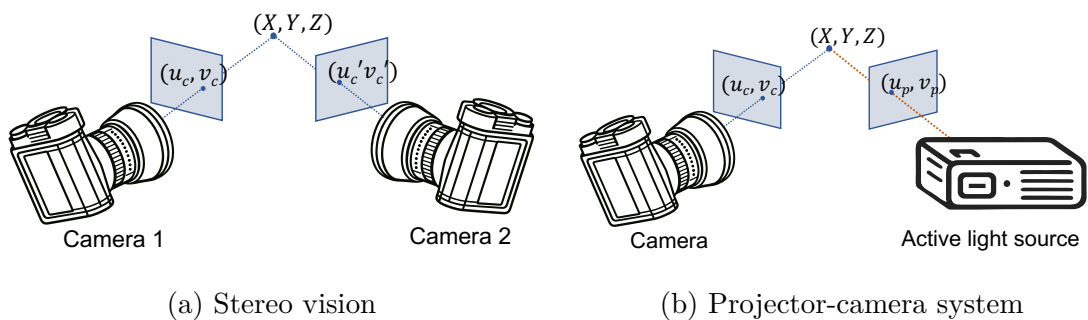
Figure 3.3: Understanding projector-camera systems from the stereo camera.
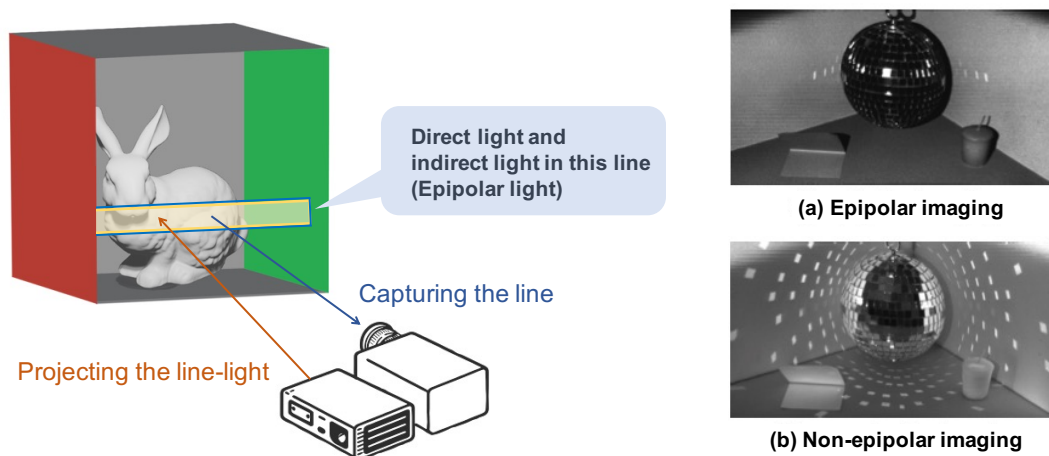


Figure 3.4: Separation of direct light and global light with the projector-camera system by O'toole *et al.* [1].

row at a time for photography. Aligning the camera and projector in parallel simplifies the epipolar geometry, allowing this synchronization to be efficiently executed.

The imaging obtained through this horizontal projection and capture comprises direct reflections from objects on the line of projector light and minimal indirect light components occurring on this line (called epipolar light), effectively eliminating the scene's indirect light. By effectively synchronizing the movements of the projector and camera, O'Toole *et al.*demonstrated the live selectively capturing mostly direct-light component (epipolar imaging in Fig 3.4(a)) and global light component(non-epipolar imaging in Fig 3.4(b)).

### 3.1.2 Slope Disparity Gating

Slope Disparity Gating [45] refined the O'Toole method for a novel disparity gating that captures only the reflected light emanating from objects within a precisely defined region of specific depth, tilt, and thickness. Crucially, this advancement was facilitated by leveraging the hardware parameters to accurately specify the targeted region's geometric characteristics. Their research elucidated that the region of focus is characterized as a sloped, near-planar surface.

Illustrated in Figure 3.5 is an overview of Slope Disparity Gating. As depicted, this technology is adept at isolating and capturing only the objects located at the intersection (highlighted in red) of the scanning planes of a laser-scanning projector (indicated in yellow) and a rolling shutter camera (illustrated in blue). Furthermore, these scanning lines perform repeated sweeping motions vertically, with the intersecting lines collectively forming a plane (the surface represented in red, termed the imaging plane). This plane's depth and tilt can be meticulously adjusted by modulating the synchronous timing and scanning velocities of both the projector and the camera, as shown in Figure 3.6. Specifically, the depth of the imaging plane changes due to the difference in synchronization timing between the camera and the projector, the tilt changes by changing the pixel clock of the camera, and the thickness changes by changing the exposure time of the camera.

Among the many attributes of Slope Disparity Gating is its capacity for real-time optical capture of the designated region and its inherent robustness against global light interference. Ueda *et al.* [45] demonstrated applications of privacy protection through real-time image masking (Figure 3.6) and clear imaging in scattering media like fog. Chandran *et al.* [44] extend the application to include such as light transport-based green screen compositing, and light transport probing techniques like novel disparity-dependent relighting.

## 3.2 Application Overview

We considered that the projector in touch sensing to the projection image could also be used as a light source device in active lighting of SDG.

To delineate the area of light transport measurement with SDG, we meticulously calibrated the camera parameters, including delay, pixel clock, and expo-
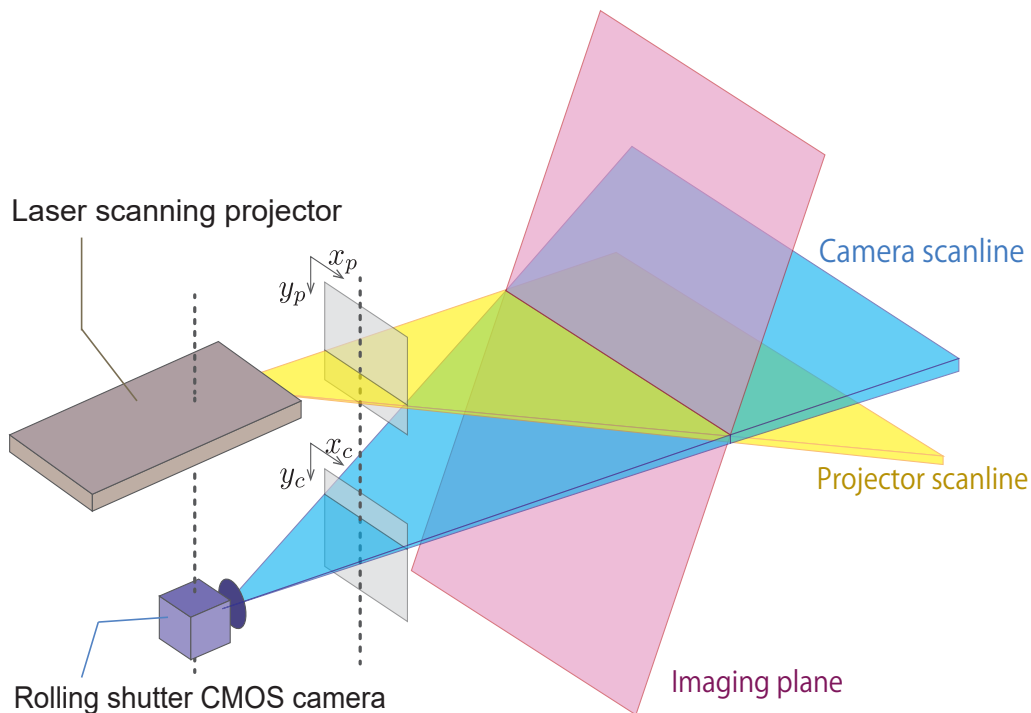
Figure 3.5: The concept of Slope Disparity Gating which allows a single plane in 3D space to be captured by the camera.

sure time. Specifically, as illustrated in Figure 3.7, we targeted a region just above the screen plane—co-planar and several centimeters thick—constituting the zone of interactive interest. We adopted the calibration method proposed by Ueda *et al.* [45] to ascertain the optimal values for the delay, pixel clock, and exposure time.

This calibration enables the camera to detect fingertips upon contact with the plane, obviating the need for complex image recognition techniques to track such interactions. Furthermore, as our capturing region excludes the projected screen area, surface texture or roughness does not compromise touch sensing. Our system necessitates a nearly planar surface devoid of clutter or objects that might intrude into the optically captured region above the surface.

A distinct divergence between Slope Disparity Gating and our system lies in

| Standard imaging plane | Depth control | Tilt control | Thickness control |



(a) Regular photo     (b) Selectively imaging only the table     (c) Image masking the table
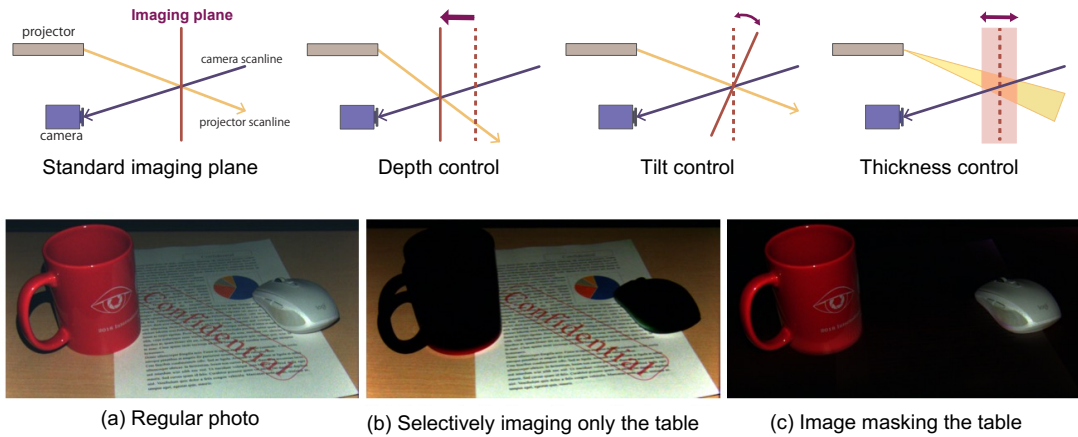
Figure 3.6: Slope Disparity Gating [45] can control the imaging area by adjusting parameters.

the projector's functional role. Within the SDG framework, the projector serves purely as a programmable illumination source. Conversely, the projector fulfills a dual role in our touch-input sensing paradigm: it operates as a programmable light source and an interface display unit.

One notable feature of our system is that the projector can simultaneously present different visuals to the human eye and the camera. The users can naturally see the projector's projected image, whereas the camera captures only the region slightly above the screen for touch sensing. In other words, the projector simultaneously takes the role of a display-interface device for users and the role of the light source for the camera to illuminate the touch-sensing area selectively.

### 3.2.1 Algorithm

The algorithm 1 delineates the system flowchart. Initially, employing the principle of Slope Disparity Gating, we measure the light transport from a specified area slightly above the touch surface, as illustrated in Figure 3.7. The algorithm integrates both touch detection and fingertip localization. The subsequent segments of this section detail how touch detection and fingertip localization are accomplished through straightforward image processing techniques.

Our imaging approach is designed to capture only direct light from a limited

---
**Algorithm 1** Touch Detection and Localization
---
    Start

    Capture image using SDG

    **if** Finger touch detected **then**

        Calculate coordinates on-camera image

        Apply homography transformation

        Register touch event

    **else**

        No touch event

    **end if**

    End
---

area. Conversely, if no object within the imaging area could reflect the light, the light will remain unmeasured, and the image will appear black. We envisage an ideal scenario where no occluding objects obstruct the surface, rendering it a free space accessible exclusively by fingers. Theoretically, only finger parts reflect the projector's light in such a scenario, manifesting non-zero brightness in the captured image. Thus, the presence of any luminance value in the image indicates the existence of a finger, while a zero luminance value signifies its absence.

However, when capturing with SDG, some pixels in the space where the object does not exist may have small pixel values due to the influence of ambient light. This ambient light is relatively weak because it is cut off significantly by the SDG mechanism and because the exposure time of the photographing system is extremely short. A threshold was applied to the image to remove the effect of the tiny pixel values. Note that this threshold depends on the exposure time, gain, and environment light. After applying the threshold, our system calculates the total luminance value of the image and carries out touch detection. If the total luminance value is non-zero, the finger touches touched; if it is zero, the image is not touched.

The center of luminance gravity in the entire image represents the finger position. As shown in Equation 3.1, we can represent the position of a finger in the camera image as the coordinates of the center of gravity $(x_c, y_c)$ of the image, which is weighted by the pixel values $m_{i,j}$ at the coordinates $(i, j)$ of the camera

image.

$$(x_c, y_c) = \frac{\sum\limits_{i,j} m_{i,j}(i,j)}{\sum\limits_{i,j} m_{i,j}} \tag{3.1}$$

Given that our touch-sensing methodology relies on camera imagery, it is imperative to translate the fingertip coordinates on the screen into coordinates derived from the captured image. Assuming a flat screen, we employ a homography matrix for this coordinate transformation. Let $C = (x_c y_c 1)^\top$ and $H$ denote the homogeneous coordinate of the finger as captured by the camera and the homography matrix, respectively. Then, the fingertip coordinates in the screen system $C' = (x_s y_s 1)^\top$ can be expressed as per Equation 3.2.

$$\begin{pmatrix} x_s \\ y_s \\ 1 \end{pmatrix} = H \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix} \tag{3.2}$$

To deduce the homography matrix, we collect $N (\geq 4)$ corresponding points between the camera and screen coordinates. We project N points covering the projection area uniformly, then instruct the user to touch these points sequentially. Subsequently, we estimate the corresponding finger position from the captured image. This process facilitates the collection of coordinate pairs $\{C_n, C'_n\}_{n=1}^N$ for homography matrix estimation. The image utilized in our method captures the region slightly above the touch surface; hence, the derived coordinates do not precisely indicate the fingertip position. This homographic transformation also serves to correct the fingertip position.

The image processing employed, as exemplified in Equations 3.1 and 3.2, is intentionally simplistic. This simplicity is feasible since fingertip segmentation is unnecessary 一the fingertip is directly sensed by its intersection with the red region.

### 3.2.2 Positioning and Contribution in Touch Sensing

This section examines how our mechanism can be positioned in touch-sensing research. Touch operation has been actively studied as a means of interaction
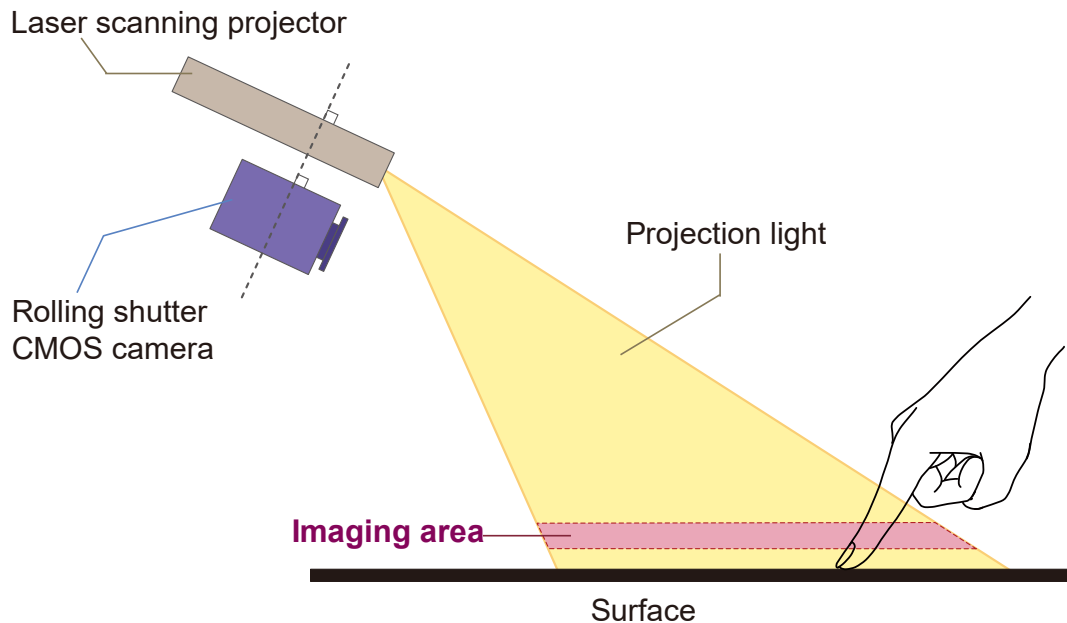
Figure 3.7: Imaging of a selected region.

due to its intuitive nature. We first discuss various technologies that enable touch sensing and compare them with our proposed solution.

## H. Direct Sensing

The most common method of touch sensing is detecting force on the touch surface itself. Capacitive touchscreens, widely used in daily life, especially in smartphones, have been actively studied in the work of Lee *et al.* [50], MTMR [51], and Smartskin [52]. In addition to capacitance, vibrational sensing has been effectively used to detect touch. Skinput [53] employs an armband sensor to detect vibrations transmitted through the body, allowing users to manipulate images projected onto their skin by touch.

Other methods to add sensing functions to the touch surface include using grids of wired [54] or acoustic sensors [55]. However, these devices have fixed screen sizes and are unsuitable for dynamic resizing. Moreover, large displays with these sensors remain expensive and physically challenging to install.

In contrast, our method requires no specialized surface electronics and can scale

to large display sizes effortlessly.

## I. Finger Input Devices

Another approach involves mounting special sensors on the object or finger used for input. This is effective when mounting sensors on the touch surface itself is difficult. Escritoire [56] utilizes an ultrasonic pen to enable touch control of a projector image. Sensetable [57] employs a physical, mouse-like device with an embedded electromagnetic sensor for interacting with the projector image.

However, research has recently increasingly focused on enabling intuitive hand operation rather than using an input device attached to the screen. For tasks not requiring fine strokes, like drawing, bare-handed operation offers more flexibility and natural interaction for the user. 3DTouch [58] utilizes finger-wearable devices with an optical laser sensor and a 9-DOF inertial measurement unit. Shi *et al.* [59] attached an accelerometer to the tip of a fingernail and employed deep learning to recognize vibration patterns for touch operation.

Similarly to these works, our system also enables intuitive hand operation with our virtual display.

## J. Passive Camera Sensing

Another research thrust utilizes passive sensing with camera images, allowing for the implementation of cost-effective systems. Letessier and Bérard [60] detected fingertips on display using a single camera image, although their method did not determine whether a finger actually touched the surface. Since a single camera cannot provide depth information, some studies have employed multiple cameras to perform stereo matching [61, 62]. Alternatively, Marshall *et al.* proposed using the color changes of the fingernail when touching [63] as a discriminator to sense depth from a single camera image. However, many of these studies require robust differentiation between fingers and the projected image on the display and can be confounded by visual clutter in the virtual scene.

## K. Active Sensing

A final avenue of research in touch sensing employs active sensing with active light sources. Many active sensing methods utilize an additional infrared light source to perform fingertip depth sensing without interfering with the projected content. Such depth information is highly effective for touch sensing, as demonstrated by Wilson and Benko [64], DIRECT [65], Omnitouch [66], and Dante Vision [67]. The use of LiDAR as an infrared light source has been explored in projects like LaserWall [68], SurfaceSight [69], and the Short-Throw Interactive Display Engine [70]. As an exceptional study not reliant on infrared light sources, Touchscreen Everywhere [71] utilized specially structured illumination from a projector. Our method employs visible light from a projected image to achieve active sensing.

We summarize the comparison between existing and our methods in Table 3.1. Our approach enables touch sensing with a single camera, eliminating the need for additional infrared devices. This advantageously positions our method as cost-effective. Minimal processing of the projected image is required, as the region of interest is selectively captured optically, resulting in low computational costs. Furthermore, since our method does not rely on machine learning, no training data is needed. It is also unnecessary to attach any devices to the finger or to require specialized display electronics for the surface itself. All these features collectively render our system an effective solution for implementing touch sensing on projected screens.

Our contributions to touch sensing are manifold. The hardware setup is simple, comprising a projector, camera, and microcontroller, without additional cameras, depth sensors, or light sources. The projector serves a dual purpose: it projects images onto the surface. It acts as the light source for imaging the area just above the screen in synchronization with the camera.

An advantage of our method is privacy preservation, as it captures only light from the space above the projected surface, ensuring that privacy-sensitive details outside the region of interest, such as human faces, are not recorded.

A salient feature of our system is the projector's ability to concurrently present distinct visuals to the human eye and the camera. While users perceive the image projected by the projector conventionally, the camera is exclusively attuned to the

Table 3.1: Comparison of our method with several existing methods with respect to various desirable properties for a touch sensing system.

| Method | Additional infrared light source | Camera device | Special projection | Learning data | Unique input device | Unique input plane |
|---|---|---|---|---|---|---|
| Capacitive touchscreen [50] [51] | not needed | not used | not needed | not needed | not needed | required |
| Skinput [53] | not needed | not used | not needed | not needed | not needed | required |
| DiamondTouch [54] | not needed | not used | not needed | not needed | not needed | required |
| Escritoire [56] | not needed | not used | multiple project | not needed | required | not needed |
| Sensetable [68] | not needed | not used | not needed | not needed | required | required |
| Ready, Steady, Touch! [59] | not needed | not used | not needed | required | required | not needed |
| Katz *et al.* [61] | not needed | multiple camera | not needed | not needed | not needed | not needed |
| TouchLight [62] | required | multiple camera | not needed | not needed | not needed | required |
| Touchscreen Everywhere [71] | not needed | single camera | Structured light | not needed | not needed | not needed |
| Pressing the Flesh [63] | not needed | single camera | not needed | not needed | not needed | not needed |
| Wilson *et al.* [62] | required | multiple kinect | not needed | not needed | not needed | not needed |
| DIRECT [65] | required | kinect | not needed | not needed | not needed | not needed |
| Omnitouch [66] | required | kinect | not needed | not needed | not needed | not needed |
| Dante vision [67] | required | thermal & kinect | not needed | not needed | not needed | not needed |
| SurfaceSight [69] | required | not used(LiDAR) | not needed | required | not needed | not needed |
| Our method | not needed | single camera | not needed | not needed | not needed | not needed |

region just above the screen to facilitate touch sensing. In essence, the projector performs dual functions: it acts as a display interface for user interaction and a light source for the camera, selectively illuminating the touch-sensing area.

Also, our system geometrically determines if a finger has touched the screen surface from a single camera image. It captures only the region slightly above the screen, making it robust against visual interference from the projected image.

### 3.2.3 Advantages in Hand Detection

To verify the robustness of this touch sensing system to video, we examine the case in which the projected contents include hands-image. When preparing the image to be projected from the projector, we used a realistic image, including a hand quote from unsplash, a stock photography service. As a typical example of hand detection, we used Handtrack [72].

We captured a scene where a user interacts with projected content, including fake hands, to test Handtrack' s hand recognition. Handtrack detected the projected fake hand but could not detect the actual hand, as shown in Figure 3.8(b). This result shows the brittleness of Handtrack in distinguishing between real and fake hands.

In contrast, our method eliminates the projected content optically, as shown

in Figure 3.8(c). Therefore, compared to conventional methods using general camera imaging, our system can make touch-sensing performance robust to the projected scene content.
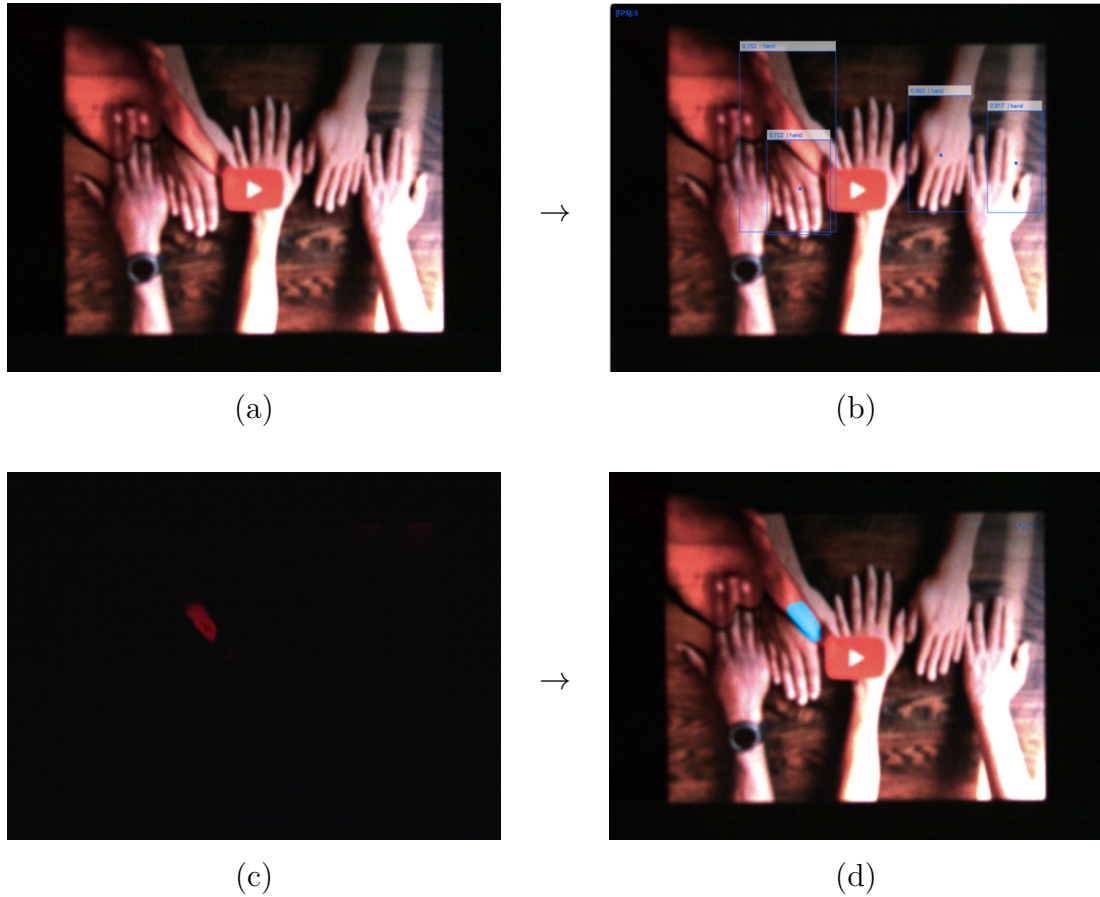


Figure 3.8: Robustness to scene content in our study: (a) the scene that finger touches a projected image of many synthetic hands, (b) the result of hand detection by HandTrack [72] (detecting only fake hand), (c) the image taken with the selective capturing technique in our system, (d) visualization of the detected finger in the original image.
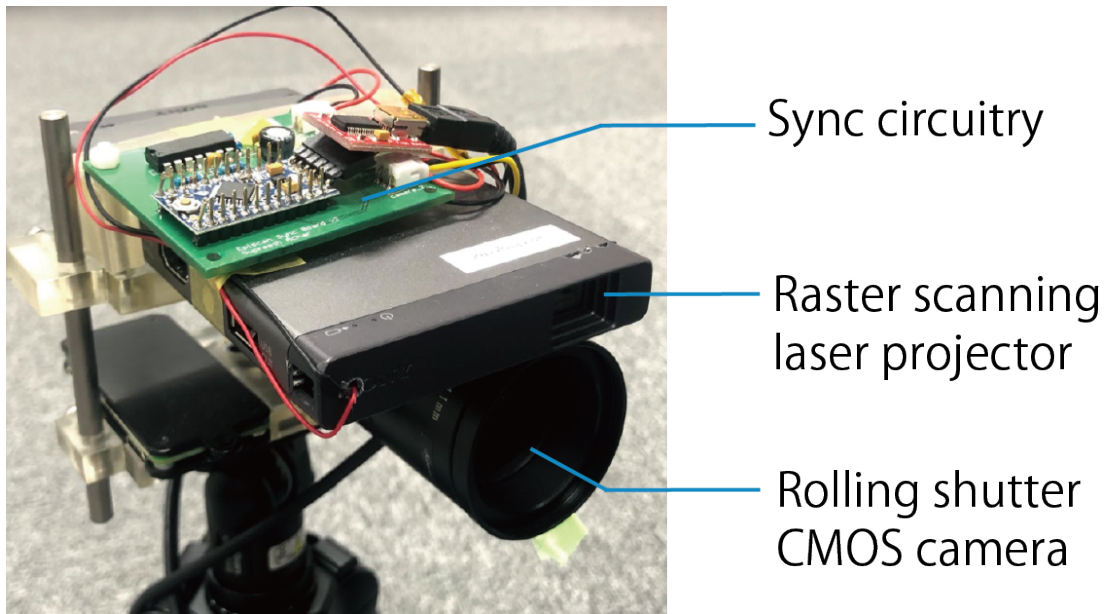
Figure 3.9: Slope disparity gating hardware prototype with raster-scanning projector aligned with the rolling-shutter camera along with synchronization electronics via an Arduino.

## 3.3 Experiments

### 3.3.1 Capturing Setup

This research delineates a prototype touch-sensing system, as depicted in Figure 3.9. The system incorporates a Sony MP-CL1A laser scanning projector, which boasts a resolution of $1280 \times 720$ and a refresh rate of 60Hz. Complementing the projector, an IDS UI-3250CP-C-HQ RGB camera is employed. This camera is distinctive in its support for both rolling and global shutter modes; However, the rolling shutter mode is specifically utilized to apply slope disparity gating. The camera's resolution stands at $1600 \times 1200$. A pivotal aspect of the hardware design is the spatial arrangement: the projector and camera are situated in proximity to each other to optimize spatial compactness. This configuration is achieved by securely mounting both devices onto a 3D-printed platform, which is then supported by a tripod.

Given the stationary nature of the relationship between the screen and the

system apparatus, manual adjustments were made to the aperture, shutter speed, and focus (The setting screen is as Fig 3.10). The aperture is fully opened to enhance the brightness of the pixel values corresponding to the finger, a necessary adaptation since Slope Disparity Gating, by its nature of rejecting ambient light and capturing only photons from the region of interest, tends to produce darker images. In this specific setup, the F-number is set to 1.6. It is pertinent to note that a fully open aperture results in a shallow depth of field. Considering that the target capture area is slightly above the projected screen, the camera is focused explicitly on the screen itself.

Synchronization between the camera and the projector is achieved through an Arduino, which is programmed to receive and analyze signals emanating from the MEMS mirror of the projector. This setup ensures the alignment of the camera's rolling shutter with these signals. The methodology for this synchronization draws upon the seminal work of Kubo *et al.* [49]. On the software front, the camera and projector are interfaced with a laptop, which runs C# code designed to implement the touch-sensing algorithm.

### 3.3.2 Touch Event

To demonstrate the efficacy of our prototype in realizing touch sensing, we configured a simple interactive system. This system operates such that when a finger touches the desk surface, which doubles as the projected screen (as shown in Figure 3.12a), a program displays the calculated touch position as a red circle.

Touch-sensing accuracy is verified by determining whether the red circle corresponds to the actual touch position. As depicted in Figure 3.12b, the center of luminance gravity calculation accurately indicates the finger's contact point. The finger's coordinates on the camera image, as obtained in Figure 3.12b, are transformed to the coordinates on the projected image using homography transformation. These coordinates are then displayed as a red circle on the projected image. Figure 3.12c shows the red circle, marking the touch position, projected precisely onto the fingertip. This demonstrates that our touch-sensing system can successfully identify the touch position of the fingertip.

To evaluate the accuracy of our touch-sensing system, we devised a task where the user is required to touch the center of a crosshair located at known coor-

dinates. The task involved a target image featuring 11 crosshairs, each with a line width of 5 pixels, displayed on a $1280 \times 720$ pixel canvas. This canvas size matches the resolution of our projector, as illustrated in Figure 3.13. Users were instructed to touch the intersection point of each of the 11 crosshairs using their bare hands.

In each trial, we determined the discrepancy between the actual position of the crosshairs and the calculated touch position of the finger, aggregating the results across 110 cumulative touch trials. The findings revealed an average horizontal error of 16.8 pixels, directed towards the side where the user stood, and an average vertical error of 4.2 pixels towards the projector's side. It's important to note that these errors are measured in the coordinates of the crosshair image. When translated to the scale of the entire projected image, as shown in Figure 3.14, it becomes evident that the size of the finger's pad contributes to this error. In essence, the error range observed in our method is within a margin that users would typically find acceptable.

### 3.3.3 Stress Testing

In our experiments, we assessed the impact of negative factors, such as low light conditions and camera defocus, on the touch-sensing performance.

We first measured the difference in touch sensing accuracy under two lighting conditions: 710 lux, which simulates standard daily lighting, and 5 lux, representing near-darkness. The results showed an average horizontal error of 0 pixels toward the side where the user stood and an average vertical error of 4.8 pixels under low-light conditions. This demonstrates that our device maintains robust performance irrespective of lighting variations.

Conversely, we observed a significant decrease in touch-sensing accuracy when the camera was out of focus. This decline can be attributed to the difficulty in controlling the captured area's thickness due to the shallower depth of field. Additionally, when the camera is out of focus, it tends to capture parts of the hand beyond the fingertip, adversely affecting the calculation of touch coordinates.

The system also affects the color of the projector image. If the image is completely black, almost no light intensity hits the finger from the projector, and touch sensing cannot be performed in this case. This is partly dependent on de-

vices such as the brightness of the projector output and camera sensitivity, but when the ambient light threshold is set with a luminance value of 20 in this setup, the color of the projector image is at the boundary where the illuminance of the projected screen is around 100 lux. It has been shown that the image projection is limited in this case.

### 3.3.4 Usefulness Considerations

Support for pinch-in and other multi-finger operations and double-tapping is needed to increase the convenience of touch-operated applications.

The current method of finding the luminance center of gravity cannot handle operations with multiple fingers, so an algorithm that detects clusters of luminance values in a closed region and finds the luminance center of gravity for each is needed. Such an implementation is a future work.

For the double-tap method, we confirmed that if the measurement area of the light transport is ideally set up, each touch can be detected with two double-taps within one second. However, there is a limitation when the thickness of the measurement area is large because the finger may enter the measurement area even if the finger is lifted off the plane between double taps.

## 3.4 Discussion

In this research, we have successfully developed a touch-sensing system utilizing slope disparity gating to capture direct light in a selected area. A critical aspect of our study involves the selective capturing of a region slightly above the projected screen, which is crucial for effective sensing. To achieve this, we synchronized a projector with a rolling-shutter camera, explicitly targeting the desired region for capture. By deliberately avoiding the capture of the screen surface initially, we have devised a touch-sensing mechanism that remains unaffected by the visual content projected on the screen. Furthermore, by eliminating the capture of scenes irrelevant to sense, we have developed a touch-sensing algorithm capable of performing both touch detection and fingertip localization from a singular camera image. Our method employs straightforward image processing techniques and uses the projector as a source of active illumination for touch sensing and as

a medium for displaying user content, thereby negating the necessity for multiple cameras and additional infrared devices.

There are several limitations inherent to our proposed approach. The system only facilitates simultaneous sensing on multiple planes by altering hardware parameters, which limits its applicability in dynamic tracking applications. Additionally, our single-camera setup cannot surmount issues of occlusion. Therefore, the projector's projection position and angle must be diagonally down in front of the user so that the user's body does not cause occlusion. The system requires initial calibration to align with the surface of interest. If the depth of field of the camera is too shallow or if the camera is not focused on the intended surface, controlling the thickness of the capture area becomes challenging, leading to diminished accuracy. In our experiments, false positives did not occur when the imaging area was correctly set in an environment devoid of objects other than fingers on the screen. However, the presence of other objects, such as a cup placed on the projected screen, may lead to false positives. Practically, our system is best suited for surfaces free of clutter or occluding objects, ensuring complete visibility for the user.
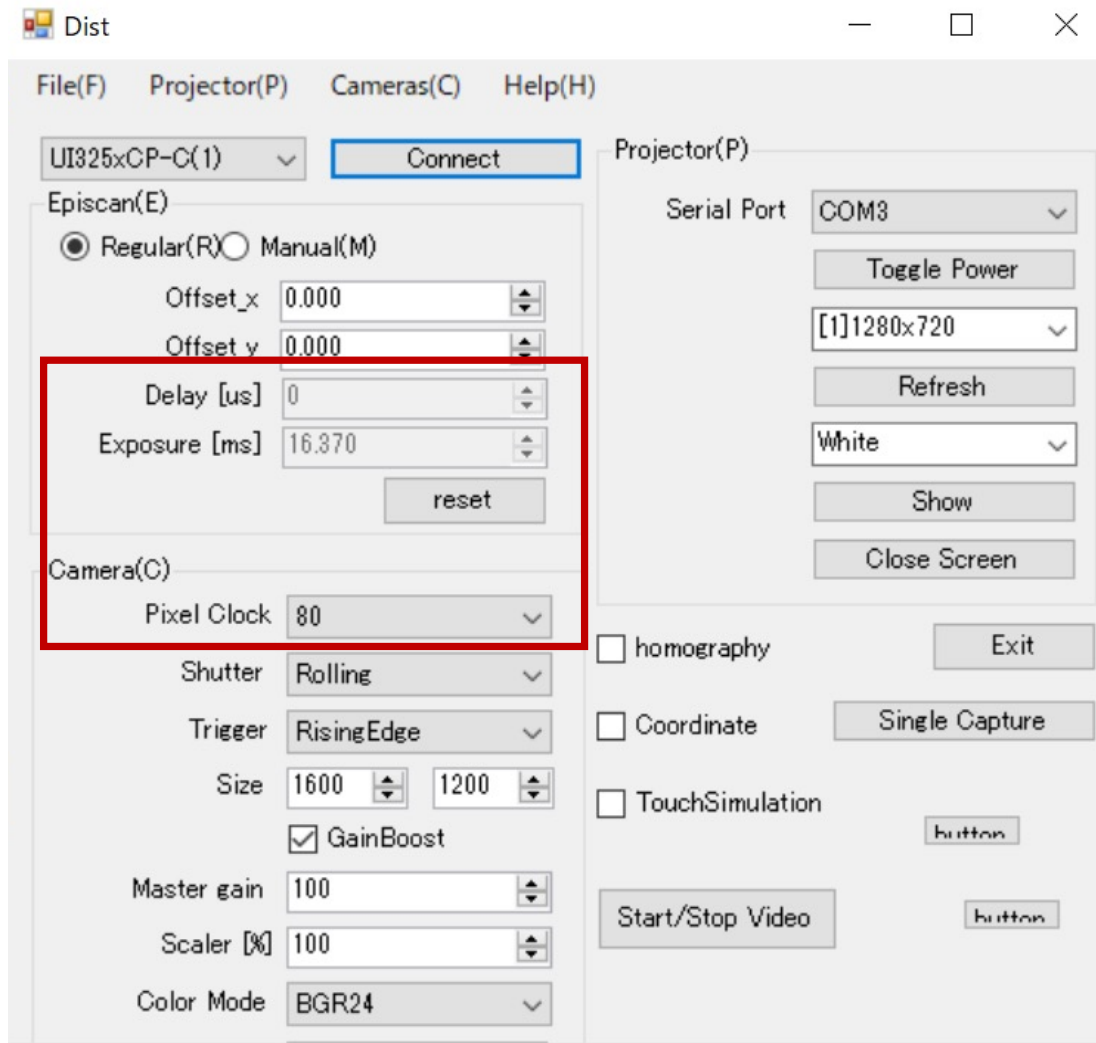
Figure 3.10: Setting screen for parameters for Slope Disparity Gating.

(a)  (b)  (c)  (d)
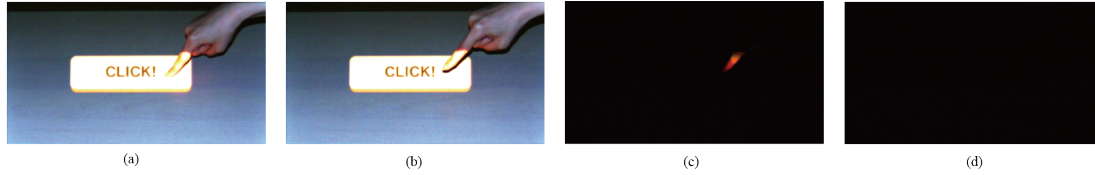
Figure 3.11: A comparison between a normal capturing system and the Slope Disparity Gating (SDG) system, in which a finger is close to the projected screen. (a) finger touches the projected screen. (b) the finger is hovering. (c) the exact same scene as (a), but captured with SDG. (d) the exact same scene as (b), but captured with SDG. In the (a) and (b) cases, both images look like the finger is touching, but in the (c) and (d) cases, grounding can be determined by whether or not the finger is in the picture.



(a)  (b)  (c)
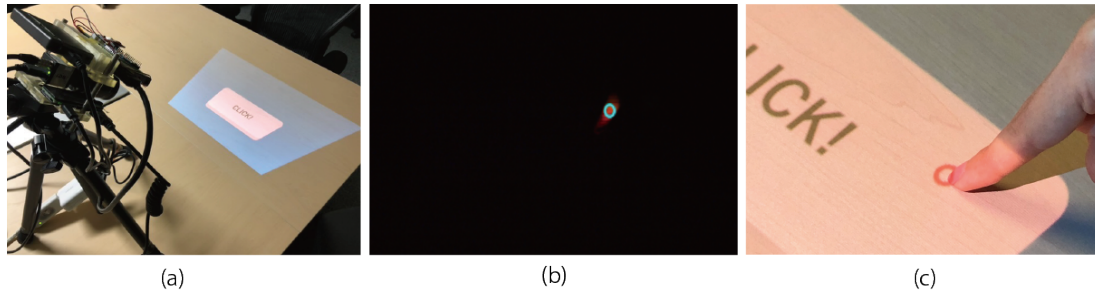
Figure 3.12: Touch sensing implementation: (a) Adjusting the angle of the hardware to project the image onto the desk surface, (b) calculating the luminance center of gravity of the SDG image when a finger touches the surface, (c) calculating the touch position by applying the homography transformation shown as a red circle.
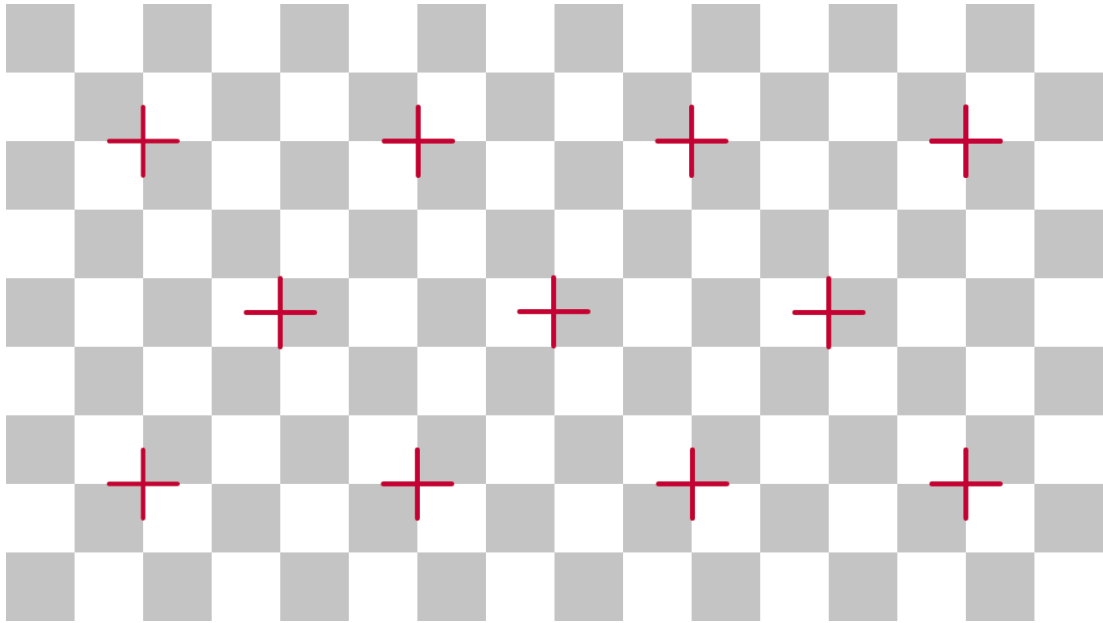
Figure 3.13: Target image of crosshairs projected for touch accuracy evaluation.



● ⋯ range of error

(a)          (b)          (c)
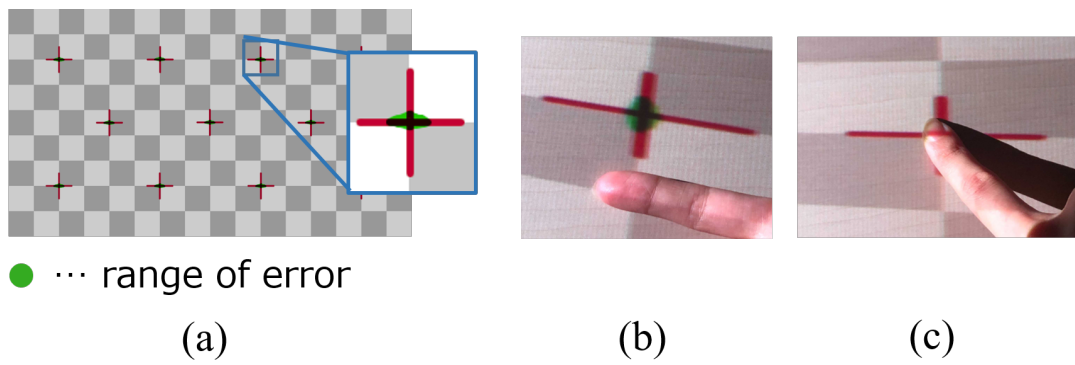
Figure 3.14: Adapting the error range to the crosshair image.

# 4 Application of Light Transport in Scattered Component

This chapter explores applications considering light transport in micro-scale optical path lengths on the order of millimeters. An example of a microscopic light path length is an object such as a translucent object or a particle layer that complicates light transport due to internal scattering. This can cause distortions in the measurement of light transport path lengths or determine the object's properties.

This thesis focuses on pigments, which are representative examples of particle layers. In cultural property science, studies are conducted to determine pigment distribution, which pigments are applied, and how thick they are. Since it is difficult to physically measure the distribution, especially concerning thickness, the photometric aspect of scattering by pigment particles is used as a clue to estimate the light transport.

One approach that can be useful in this case is to solve the inverse problem of the physical model. Using a physical model that describes the behavior of the light transport, i.e., a physical model that connects the parameters in the scene to the measured luminance values, it is possible to estimate the physical phenomenon by solving an inverse problem from the measurements. In this case, the thickness of the pigment can be analyzed by estimating the optical path length of the light transport that caused the scattering.

We focused on the Kubelka-Munk theory [73] as a physical model, which deals with optical properties of coatings such as pigment layers and is mainly applied to color reproduction and opacity of coatings. We focused on the fact that the Kubelka-Munk theory has two parameters: the scattering coefficient, which is specific to pigments, and the pigment thickness (strictly speaking, the optical

path length).

Solving the inverse problem of the Kubelka-Munk model is an ill-posed problem when determining pigment distribution, i.e., pigment type and thickness. In addition, the Kubelka-Munk model is pixel-by-pixel and does not consider spatial continuity. Therefore, we have incorporated deep learning of neural representations that consider spatial continuity as a solution method.

As a specific application, we have explored pigment distribution in tumulus murals because this study solves the ill-posed problem. The substrates of tumulus murals are rock, and the patterns make the inversion of the Kubelka-Munk model particularly challenging. To the best of our knowledge, this is the first time we have tackled pigment distribution in the substrate in such a heterogeneous.

## 4.1 Scattering and Absorption in Particle Layer

When light passes through particle matter, there are two main phenomena: scattering and absorption.

Light scattering occurs when light interacts with material, resulting in the deviation of light waves from their original path. This phenomenon occurs when light encounters small particles or irregularities within a material. The size, shape, and refractive index of the particles significantly influence the nature of scattering.

- **Rayleigh Scattering:** This type of scattering happens when the particles are much smaller than the wavelength of light [74]. It is responsible for the sky's blue color and the sunsets' red hues.

- **Mie Scattering:** Mie scattering occurs when the size of the particles is comparable to or slightly larger than the wavelength of light. It plays a crucial role in phenomena like the whiteness of clouds [75].

- **Geometrical Optics Scattering:** When the particles are significantly larger than the wavelength of light, scattering follows the rules of geometrical optics [76].

In this study, the particle size is sufficiently larger than the wavelength because the target is the pigment in the mural paintings of ancient tombs. In other words, geometric scattering is the target.

Light absorption is the process where a material absorbs the energy of light and converts it into another form of energy, usually heat. Absorption depends on the molecular and electronic structure of the material and determines the material's 'transparency' to specific wavelengths of light. The absorption coefficient measures how efficiently a material absorbs light at a particular wavelength. A higher coefficient indicates stronger absorption at that wavelength.

In real materials, scattering and absorption occur simultaneously and often influence each other. For example, scattering by particles can elongate the path of light, increasing the likelihood of absorption within the material. This interplay determines the color, transparency, and reflective properties of materials.

## 4.2 The Kubelka-Munk Model

The Kubelka-Munk model [73] is important in explaining the light behavior of light-scattering and light-absorbing particles. It is a proven framework in the field of color science and optical properties of materials. It is crucial to understand light's interaction with layered media such as paints, coatings, and biological tissues.

The Kubelka-Munk model is predicated on light attenuation due to scattering and absorption within the particle layer, considering two fluxes: incident and reflected directions. The two-flux attenuation model parameters include a scattering coefficient $S$ and an absorption coefficient $K$, each of which indicates the degree to which light gets attenuated by either absorption or scattering. We use $i$ and $j$ to represent the diffuse light flux in the top-to-bottom (incident) and bottom-to-top (reflected) directions within the pigment layer, respectively (Fig 4.1). We express the change in $i$ and $j$ for an infinitesimal thickness $dx$ as follows.

$$di = (S + K)idx - Sjdx, \qquad (4.1)$$

$$dj = -(S + K)jdx + Sidx. \qquad (4.2)$$
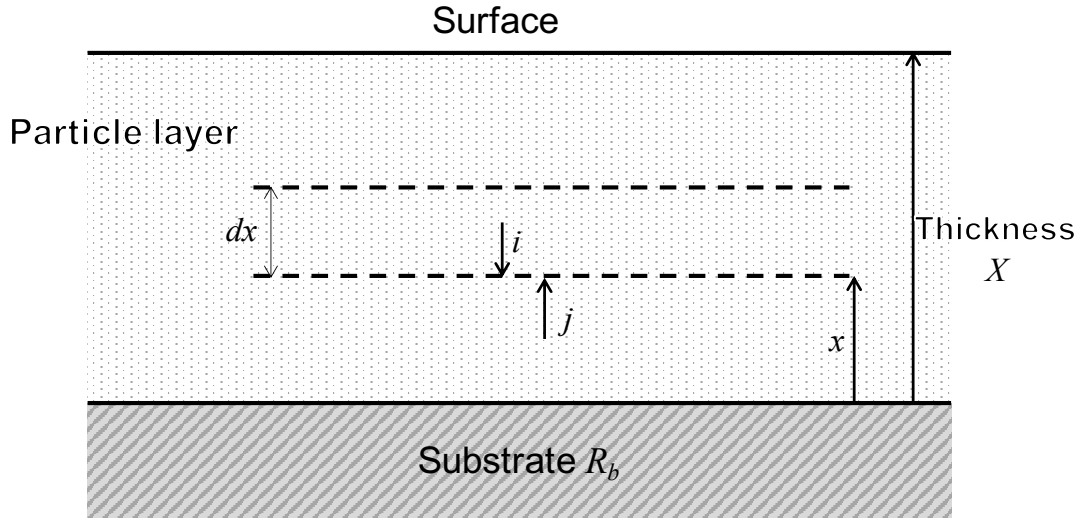
Figure 4.1: Schematic of Kubelka-Munk model.

The differential equations can be solved by transforming and adding Equations 4.1 and 4.2.

$$\frac{di}{i} = (S+K)dx - S\frac{j}{i}dx, \tag{4.3}$$

$$\frac{dj}{j} = -(S+K)dx + S\frac{i}{j}dx. \tag{4.4}$$

(4.4)-(4.3)

$$\frac{dj}{j} - \frac{di}{i} = d\ln\frac{j}{i} = -2(S+K)dx + S\left(\frac{i}{j} + \frac{j}{i}\right)dx. \tag{4.5}$$

Here, let us define $\frac{i}{s} = h$ and transform Equation 4.5, then

$$\int \frac{dh}{\frac{h^2}{2}\left(\frac{S+K}{S}\right)h + 1} = -S\int dx. \tag{4.6}$$

Consider a particle layer with thickness $x = X$, where the surface reflectance is denoted as $R_{km}$. At $x = 0$, that is, for the substrate reflectance, we denote it as $R_b$. Further, let us define $\frac{S+K}{S} = a$. Then, the final equation can be expressed as

$$\ln \frac{(R_{km} - a - \sqrt{a^2 - 1})(R_b - a + \sqrt{a^2 - 1})}{(R_b - a - \sqrt{a^2 - 1})(R_{km} - a + \sqrt{a^2 - 1})} = 2SX\sqrt{a^2 - 1}. \tag{4.7}$$

As stated by Kubelka [77], we can obtain the scattering coefficient $S$ and absorption coefficient $K$ for each pigment from measurements. Moreover, we can compute $R_\infty$ using these coefficients by Equation 4.7.

$$R_\infty = 1 + \frac{K}{S} - \sqrt{\left(\frac{K}{S}\right)^2 + 2\frac{K}{S}} \tag{4.8}$$

Also, we can determine the reflectance $R_{km}$ of layered-surface objects:

$$R_{km} = \frac{\frac{1}{R_\infty}(R_b - R_\infty) - R_\infty(R_b - \frac{1}{R_\infty})e^{SX(\frac{1}{R_\infty} - R_\infty)}}{(R_b - R_\infty) - (R_b - \frac{1}{R_\infty})e^{SX(\frac{1}{R_\infty} - R_\infty)}} \tag{4.9}$$

Note that the thickness $X$ of the pigment layer represents the relative optical path length, and as such, it doesn't have any units.

## 4.3 Application Overview

This thesis applies this physics model to a decorated ancient tumulus mural painting.

Decorated ancient tumulus, the burial mounds of ancient rulers, are recognized by their patterned decorations that adorn the rock grave chambers. These patterns offer valuable archaeological insights into the religious beliefs, art, and societal progression during that period. However, despite their significance, preservation concerns often keep these tombs closed to the public. Consequently, a critical task arises: mapping these patterns and subsequently digitizing them for broader accessibility.

Measuring ancient tomb murals presents significant challenges due to their large-scale and time-consuming nature. Additionally, these murals are highly susceptible to degradation in the presence of humidity, making repeated and prolonged measurements a risky endeavor. Given these constraints, shifting away

from geometry-based approaches to analysis becomes crucial. Instead of utilizing already acquired data, a photometric approach offers a more viable and less invasive means of analyzing these precious historical artifacts.

This thesis analyzes the extent of pigment residues present in ancient tomb murals. Due to the passage of time, these tombs are likely to have experienced significant degradation and loss, raising concerns about the future feasibility of archaeological research based on the original artifacts. Moreover, in the event of disasters causing severe damage to these murals, preserving research data about their original state becomes crucial for potential restoration efforts.

Analyzing the spatial distribution of pigments becomes essential to consider the optical path length of the attenuated light transport, commonly referred to as pigment mapping. When analyzing the patterns within paintings, distinguishing the pigment from the substrate is essential. This process involves viewing the painting as a multilayered surface object comprising pigment and substrate layers. Two optical models have proven particularly effective: Lambert-Beer's law and the Kubelka-Munk model (KM model). The latter, introducing more complex assumptions than the former, offers a more precise approximation of this layered structure.

Understanding the pigment distribution in tomb murals is challenging due to the heterogeneous patterns inherent in the rock substrate. The pigments used in these murals resemble semi-transparent optical models, similar to watercolor paints, indicating that the influence of the substrate surface is non-negligible. The substrate pattern's complexity significantly amplifies the task of pigment separation.

This study proposes a method for pigment mapping in tomb murals, even with heterogeneous rock patterns (Figure 1). Specifically, we analyze which pigments are of what thickness. Our approach employs neural representation and physics-based unsupervised learning. The inputs are the spectral reflectance data at a specific position of a tomb mural image and its position, while the outputs are pigment thickness, pigment class, and substrate class. For loss calculation, we estimate the spectral reflectance from these outputs in line with the KM model and calculate its error from the input. That enables the model to make inferences based on the physics model.
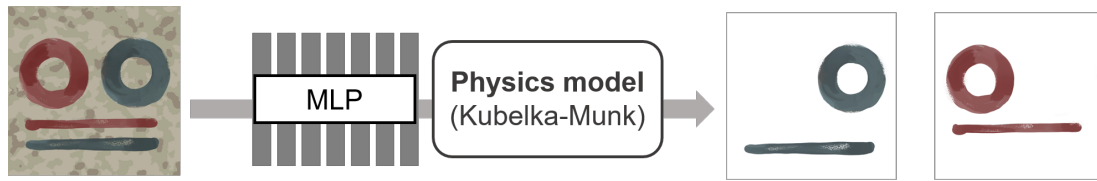
Figure 4.2: Performing pigment mapping using the Kubelka-Munk model and deep learning.

## 4.3.1 Positioning and Contribution in Pigment Analysis

## L. Pigment Analysis Using KM Model

We can divide the variables of the KM model into two categories: pigment layer parameters and a substrate parameter. Prior studies often assumed that the substrate parameter was either negligible or constant. For example, in the case of opaque paintings, such as oil or acrylic paintings, the pigment layer is so thick that it completely obscures the canvas substrate [78, 79, 80, 81, 82, 83]. The substrate parameter is supposed to be a constant value in [84, 85] because the substrate is often homogeneous paper (white) in the watercolors and printings. However, when we assume a non-negligible heterogeneous substrate, the problem becomes more complicated; we need to estimate the substrate parameter on each pixel while analyzing pigments.

## M. Other Pigment Analysis

The spider model [86] is an effective physics model for separating pigments from the substrates. This model is primarily based on Lambert-Beer's law, a one-fux model focusing on light attenuation in a singular direction, shown in Figure 2(a). However, it is crucial to note that Lambert-Beer's law does not consider the non-negligible scattering caused by pigment particles as a parameter. While the Spider model is reliant on RGB images, hyperspectral data for pigment mapping has been widely reported [87, 88, 89, 90, 91]. Hyperspectral data provides more detailed spectral information, which could potentially enhance the accuracy of pigment mapping. In recent years, with the rise of deep learning, research has emerged that combines hyperspectral data and deep learning to perform pigment

Figure 4.3: Unlike previous studies, this thesis conducts pigment mapping on heterogeneous substrates.

mapping of cultural heritage [92, 93, 94].

## 4.4 Neural Pigment Representation

### 4.4.1 Problem Setting

**Pigment layer:** Existing studies [95] suggest that specific types of pigments have been identified, and the pigments used in tombs are not mixed. In most cases, there are one to three pigment classes, with Ozuka-tumulus, which has six colors, being the rarest. This introduces the need to undertake a two-pronged estimation process. Firstly, we must identify which pigment is used at each point on the mural. Secondly, we need to estimate the thickness of the pigment application. Within the context of the KM model variables, this study aims to classify the values of $S$ and $R_\infty$, and estimate the value of $X$.

**Substrate layer:** In the spectral image of decorated ancient tumulus taken in a narrow stone chamber, the substrate shows heterogeneous. The substrate layer, comprised of various minerals, allows us to theoretically classify the substrate's spectrum based on the number of minerals in the rock. This study uses K-means clustering on the spectral data from parts of the substrate where no pigment was applied. This method will assist in determining the precise color of the rock at each location on the mural.

### 4.4.2 Overview of MLP

Figure 4.4 provides an overview of our model. To accomplish pigment mapping on a heterogeneous substrate, we utilize a network architecture similar to NeRF [96],
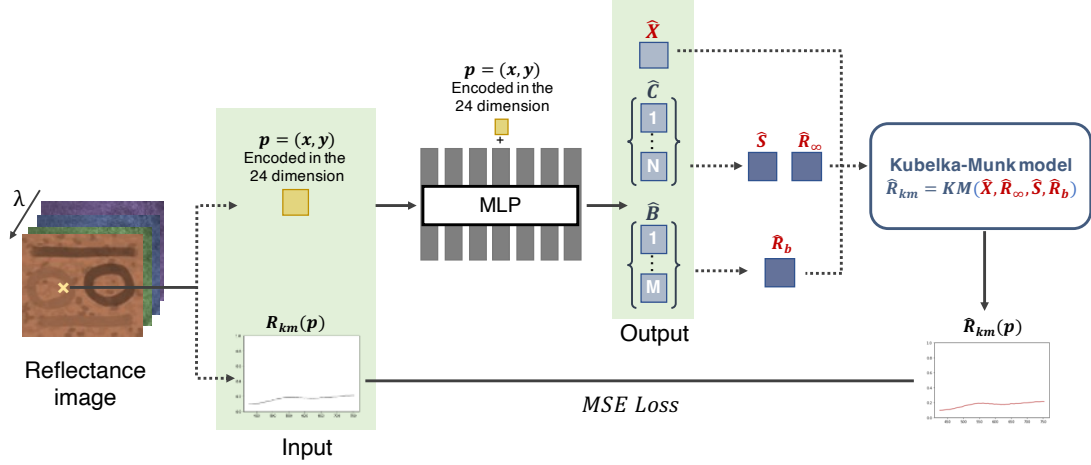
Figure 4.4: Overview of our deep learning model: We input the position of a reflectance image and its corresponding reflectance ($R_{km}$ in the KM model) into the MLP. The MLP's output is used to determine the variables of the KM model. As the KM model is differentiable, we can conduct learning by minimizing the error between the estimated reflectance ($\hat{R}_{km}$) and the input reflectance ($R_{km}$) used in the process.

wherein the network input comprises coordinates from the input image to account for spatial continuity.

We use the 2D position of the hyperspectral image $\boldsymbol{p} = (x, y)$, which are normalized to $[-1, 1]$, and the spectral reflectance at the position $R_{km}(\boldsymbol{p}) \in \mathbb{R}^\Lambda$ as the network inputs, where $\Lambda$ represents the number of bands. Our network generates the following outputs: the estimated thickness of the pigment layer $\hat{X} \in \mathbb{R}$, the probability of each pigment class $C^n \in [0, 1], (n = 1, 2, \ldots, N)$, and the probability of each substrate class $B^m \in [0, 1], (m = 1, 2, \ldots, M)$. A model that estimates optical parameters by inputting coordinates is called a neural representation, and this model is based on this idea. Neural representation allows the model to maintain spatial continuity in its estimation.

Please note that $\sum_{n=1}^{N} C^n = 1$ and $\sum_{m=1}^{M} B^m = 1$. For ensuring non-negative output, we use a ReLU as the activation function for $\hat{X}$, and for achieving maximally sparse outputs, we apply Sparsemax [97] to $C^n$ and $B^m$.

Based on the above, the MLP model $F_\Theta$ can be defined as follows:

43

$$F_\Theta : (\boldsymbol{p}, R_{km}(\boldsymbol{p})) \mapsto (\hat{X}, C^n, B^m), \tag{4.10}$$

where $\Theta$ represents the parameters of the network.

By using this defined deep learning model $F_\Theta$, we can estimate the parameters of the pigment simultaneously accounting for the substrate.

### 4.4.3 Positional Encoding

According to NeRF [96], mapping inputs to a higher-dimensional space with a high-frequency function before inputting them into the network improves the fit for high-frequency data. We adhere to this principle by embedding $\boldsymbol{p}$ within a high-frequency function as

$$PE(\boldsymbol{p}) = (\sin(2^0\pi\boldsymbol{p}), \cos(2^0\pi\boldsymbol{p}), \cdots, \sin(2^{L-1}\pi\boldsymbol{p}), \cos(2^{L-1}\pi\boldsymbol{p})). \tag{4.11}$$

This formula enables embedding a 2-dimensional vector $\boldsymbol{p}$ into a $2 \times 2L$ space. For this study, we expand to 24 dimensions with $L = 6$.

### 4.4.4 Hidden Layer

We combine the two kinds of inputs into a single vector and pass this combined vector through seven fully connected layers with ReLU activation (comprising 300 perceptrons). Additionally, we implement a skip connection linking the input $\boldsymbol{p}$, embedded in a $2 \times 2L$ space, to the activation of the fourth layer, as suggested by NeRF [96] and DeepSDF [98].

### 4.4.5 Loss using the Kubelka-Munk Model

We calculate our loss as the mean squared error between the estimated $\hat{R}_{km}$ and the actual $R_{km}$ used as input, following the method by Shitomi *et al.* [85]. We determine the estimated $\hat{R}_{km}$ using the model outputs. During the learning step, the model computes $S$, $R_\infty$, and $R_b$ as a weighted sum with the probability of each pigment $C^n$ to maintain differentiability:

$$
\begin{align}
S(C) &= \sum_{n=1}^{N} C^n S^n \tag{4.12} \\
R_\infty(C) &= \sum_{n=1}^{N} C^n R_\infty{}^n \tag{4.13} \\
R_b(B) &= \sum_{m=1}^{M} B^m R_b{}^m, \tag{4.14}
\end{align}
$$

where $S^n \in \mathbb{R}^\Lambda$ and $R_\infty{}^n \in \mathbb{R}^\Lambda$ are the scattering coefficient and the reflectance of $n$-th pigment class, and $R_b{}^m \in \mathbb{R}^\Lambda$ is the spectral reflectance of $m$-th substrate class. Substituting Eqs. (4.12)–(4.14) to Eq. (4.9), we obtain $\hat{R}_{km}$ as

$$
\hat{R}_{km} = KM(\hat{X}, S(\hat{C}), R_\infty(\hat{C}), R_b(\hat{B})). \tag{4.15}
$$

The mean squared error function is

$$
\mathcal{L} = \frac{1}{\Lambda} \sum_{\boldsymbol{p} \in \mathcal{R}} \|\hat{R}_{km}(F_\Theta(\boldsymbol{p}, R_{km}(\boldsymbol{p}))) - R_{km}(\boldsymbol{p})\|_2^2, \tag{4.16}
$$

where $\mathcal{R}$ represents the set of coordinates in each batch. By back-propagating the error through the KM model, we can calculate the gradients of the weights.

### 4.4.6 Pigment Mapping

In the testing phase, we define pigment mapping as the product of the pigment class and its thickness, calculated for each region. We select the pigment class as $argmax\, C^n$. Even though the restriction of $\sum_{n=1}^{N} C^n = 1$ leads to estimating the pigment class in the area without pigment, this doesn't cause any issues if the thickness estimation correctly results in zero.

## 4.5 Experiments

### 4.5.1 Dataset and Settings

We used a spectral image of the Mezurashiduka tomb, located in western Japan. The tomb's visual representation combines two kinds of pigments ($N = 2$), red (Bengara) and gray (Hekikaimatsu), applied over a granite-rock surface (Figure 4.6(a)). This tombs' dataset is a collection of spectral images with 81 bands,

ranging from 400nm to 720nm. With an image resolution of $420 \times 700$, we used 294,000 pixels for training. Since granite is composed mainly of six minerals, the substrate class was set as $M = 6$ in this study.

In real data, it is difficult to prepare pigments identical to those used in ancient murals. Therefore, $S$ and $R_\infty$ values for each pigment were determined by annotating the pixels where the pigment is applied, and subsequently optimizing using the KM model.

We used the reflectance of annotated pixels for the initial value for $R_\infty$. Also, we used the Adam optimizer to update $S$ and $R_\infty$ to minimize the discrepancy between the measured and modeled reflectance. The optimized $S$ and $R_\infty$ are shown in Figure 4.5(a) and (b), respectively.

For substrate, $R_b$ was obtained by annotating regions without pigment, clustering approximately 3,000 reflectance data points (seen in Figure 4.5(c)) using K-means. We adopt the centroid as the reflectance of each substrate class (Figure 4.5(d)).

In training sessions, we employed the Optuna [99] hyperparameter optimization framework. The parameters we focused on were the batch size and the learning rate. The batch size was selected from 512, 1024, 2048, and 4096. On the other hand, the learning rate was searched within a range from $1 \times 10^{-4}$ to $5 \times 10^{-4}$. The number of training sessions was set to 500, with CosineAnnealingLR completing one cycle every 25 iterations.

## 4.5.2 Results

We applied our proposed model to actual spectral data obtained from tumulus murals, assessing its ability to estimate. For validation purposes, we referred to the restoration sketches produced by Hakkou Kusaka, a noted Nihonga artist (Figure 4.6(e)). Furthermore, Figure 4.6(d) and (h) represent the segmentation of Hakkou's reference image using the K-means method, displaying the regions of each pigment. Commissioned by the Agency for Cultural Affairs between 1953 and 1955, Kusaka created these sketches. Although his artwork is based on subjectivity, it is a valid benchmark against which we can evaluate the accuracy of our model estimates.

The estimation results are included in the Figure 4.6 (b) and (f). It can be

---

**Algorithm 2** Optimization of Kubelka-Munk Model Parameters $S$ and $R_\infty$

---

**Procedure** UpdateLearningRate($lr_S$, $lr_K$, $epoch$)

**if** $(epoch + 1) \mod 100 == 0$ **then**

    $lr_S \leftarrow lr_S \times \frac{1}{10}$

    $lr_K \leftarrow lr_K \times \frac{1}{10}$

**end if**

**Return** $lr_S$, $lr_K$

**End Procedure**


**Procedure** OptimizeParams($specs$, $init\_R_\infty$, $init\_K$, $lr_S$, $lr_K$, $R_b$)

$thickness \leftarrow$ Constant Tensor($x$)

$R_{base} \leftarrow$ Constant Tensor($R_b$)

$ratio \leftarrow$ Constant Tensor of Ones(Size of $specs$)

**for** $epoch \leftarrow 1$ **to** $1000$ **do**

    $K \leftarrow$ Variable($init\_K$)

    $R_\infty \leftarrow$ Variable($init\_R_\infty$)

    $S \leftarrow \frac{2 \times R_\infty \times K}{(1 - R_\infty)^2}$

    $S_{mix} \leftarrow ratio \times S$

    $R_{\infty\_mix} \leftarrow ratio \times R_\infty$

    $inv\_R_\infty \leftarrow \frac{1}{R_{\infty\_mix} + \epsilon}$

    $temp \leftarrow (R_{base} - inv\_R_\infty) \times \exp(S_{mix} \times thickness \times (inv\_R_\infty - R_{\infty\_mix}))$

    $upR \leftarrow inv\_R_\infty \times (R_{base} - R_{\infty\_mix}) - R_{\infty\_mix} \times temp$

    $downR \leftarrow (R_{base} - R_{\infty\_mix}) - temp$

    $rendered\_R \leftarrow \frac{upR}{downR}$

    $normalized\_rendered\_R \leftarrow \frac{rendered\_R}{\sum(rendered\_R, axis=1)}$

    $loss \leftarrow \sum((normalized\_rendered\_R - norm\_specs)^2)$

    **if** $is\_nan(loss)$ **then**

        **Print** Debug Information

        **Break**

    **end if**

    $grad\_K, grad\_R_\infty \leftarrow$ Gradient($loss$, $K$, $R_\infty$)

    $lr_S, lr_K \leftarrow$ UpdateLearningRate($lr_S$, $lr_K$, $epoch$)

    $init\_K \leftarrow init\_K - lr_S \times grad\_K$

    $init\_R_\infty \leftarrow init\_R_\infty - lr_K \times grad\_R_\infty$

    $init\_S \leftarrow \frac{2 \times init\_R_\infty \times init\_K}{(1 - init\_R_\infty)^2}$

**end for**

**Return** $init\_K$, $init\_R_\infty$, $init\_S$    47

**End Procedure**

---

seen that our method captures the tendency of pigment distribution well.

Our model estimates that pigment is applied even in areas not painted on the reference image, and there are three possible interpretations for this. The first factor is that the model might have detected the presence of pigment that has deteriorated and washed away over time. The second one could be a drawback associated with the fact that our research is based on spectral information. Depending on the combination of the substrate and the pigment, it is possible that a spectrum similar to another substrate class could have been reproduced. The third factor pertains to the influence of the data used for estimation. This includes the error during the optimization of $S$ and $R_\infty$, the decreased expressiveness due to the classification of the substrate layer, the impact of illumination variations in the spectral image, and the noise when capturing a spectral image. It is reasonable to view our estimation results as a combination of the three phenomena mentioned.

Figure 4.7 demonstrates the outcome of applying the Difference of Gaussians (DoG) to our estimation results. Given that red pigments are often utilized for delineation, their corresponding values were set to accentuate the edges. Conversely, the grey pigments, frequently used for area-filling expressions, had their values adjusted to encourage smoothing. These processed images illustrate the potential for estimating mural patterns even when past reference images are not available. Also, referring to Hakkou's inferred figure, the results of this study provide a clue as to how the tumulus wall paintings deteriorated.

### 4.5.3 Comparison with Homogeneous Substrate

The substrate of the rock on which the murals are painted is inherently heterogeneous. However, for the sake of comparison, we also examined the case where we assume the substrate to be homogeneous. When treating the substrate as homogeneous, we adopted the average value of the substrate reflectance (the blue-line in Figure 4.5(c)) as the substrate spectrum.

The assumption of the homogeneous substrate simplified our modeling process, removing the necessity to classify $R_b$, which then becomes a constant. The results are shown in Figure 4.6(c) and (g). Our analysis reveals that when we assume the substrate to be constant, the estimation results exhibit more errors than when

considering the substrate to be heterogeneous. The estimation of red pigment is getting worse.

Furthermore, we have discovered that the results can deteriorate depending on how the images in the dataset are cropped under the assumption of a homogeneous substrate. The frequency of pattern changes in the dataset is associated with the size of the image area. Specifically, the smaller the image area, the higher the frequency of these pattern changes, especially when cropping occurs in regions with substantial pattern variations. Figure 4.8 shows estimation results when cropping a $200 \times 200$ area from the original dataset. The assumption of a homogeneous substrate clearly reduces the accuracy.

This is due to the fact that assuming a constant substrate reduces the expression of the model and due to physics-based estimation. Figure 4.8(b) is a plot of the reflectance data at the coordinates indicated by the '$\times$' marker in the right-middle of Figure 4.8. Despite the clear inaccuracies in pigment estimation assuming a homogeneous substrate, the reconstructed reflectance $\hat{R}_{km}$ is found to be similar to the original reflectance $R_{km}$. Such errors are unavoidable as long as reflectance is the basis for estimation, but the results indicate that assuming a heterogeneous substrate and increasing the representational capacity of the model can partially mitigate these issues. These observations underscore the importance of accurately capturing the substrate's heterogeneity in our analyses.

## 4.6 Ablation Study

We conducted a simulation experiment to validate our pigment mapping method on heterogeneous substrates. Due to the challenge of obtaining a sample with a known thickness, we initially synthesized a spectral image for verification purposes.

### 4.6.1 Dataset and Settings

Figure 4.9(a) displays an RGB visualization of the synthesized spectral data, encompassing 65 bands within the 430–760nm range. With an image resolution of $200 \times 200$, we utilized 40,000 pixels for training. The experiment was configured with $N = 2$ pigment classes and $M = 3$ substrate classes.

The spectral data was synthesized based on actual measurements. For pigments, we used red (Bengara) and green (Terre Verte), both frequently found in ancient tomb paintings. We derived the $S$ and $R_\infty$ values (Figure 4.9(c),(d)) or these pigments using the method detailed in [77], which involved measuring the reflectance of each pigment painted on both a white and a black substrate. Moreover, we determined the $X$ value (Figure 4.9(b)) by solving the inverse problem of the KM model from paint applied to a single-color substrate with constant substrate parameters. This process was executed using the technique described in [85], in which we analyzed the thickness of the pattern painted on white paper. For the substrate, we measured the spectral reflectance of a bare rock surface. Using k-means clustering, we identified patterns within the substrate. The spectral component of each class, representing the centroid of its respective cluster in k-means, was used as the reflectance of each class (Figure 4.9(d)). With the obtained $S$, $R_\infty$, $R_b$, and $X$ values for each region, we synthesized a spectral image using the KM model.

For our training regimen, we used the Adam optimizer with CosineAnnealingLR serving as a scheduler to adjust the learning rate. The learning rate started at $5 \times 10^{-4}$ and gradually decreased to $5 \times 10^{-7}$ during the optimization process. The training was conducted over 1,200 iterations, with CosineAnnealingLR completing one cycle every 25 iterations.

### 4.6.2 Pre-experiment

Figure 4.10 displays the estimates produced by our pigment mapping. The pigment classification performs well, slightly underestimating the thickness but nonetheless accurately capturing the original thickness distribution. We computed the SSIM and PSNR between the ground truth and estimated values for each pigment mapping. Our model achieved SSIM= 0.998, PSNR= 48.6 for the red pigment, and SSIM= 0.997, PSNR= 46.9 for the green pigment.

On the other hand, when the substrate was assumed to be homogeneous ($M = 1$), the effects of the substrate's pattern were reflected, causing significant errors in the estimation. In this case, SSIM= 0.278, PSNR= 21.1 for the red pigment, and SSIM= 0.359, PSNR= 23.1 for the green pigment.

Figure 4.11 presents the results of the substrate estimation. In Figure 4.11(c),

we use white to denote pixels where the class has been correctly estimated and red to denote pixels where the class has been incorrectly estimated. This model succeeds in classifying the class correctly for all pixels.

### 4.6.3 Results

Our model has $R_{km}(\boldsymbol{p})$, $PE(\boldsymbol{p})$ as inputs to the first layer, and $PE(\boldsymbol{p})$ as an input to the hidden layer. We have conducted an ablation study to validate our design choices. Table 4.1 shows each ablation design.

The accuracy was compared using SSIM and PSNR. Table 4.2 shows the result of the ablation study. Our method performed the best score for both SSIM and PSNR. This result shows it is essential to input both reflectance $R_{km}(\boldsymbol{p})$ and its position $PE(\boldsymbol{p})$. Furthermore, by comparing *Ours* with *Abs-2* and *Abs-3*, the effectiveness of including coordinates twice has been confirmed as presented in previous studies [96, 98].

Figure 4.12 compares the original and *Abs-4* errors. *Abs-4* is a pixel-by-pixel method with no position input, and the substrate pattern tends to show up more as an error. The position input allows spatial continuity to be considered, and the substrate pattern's influence can be eliminated more. These results indicate that a model based on a neural representation is valid.

Table 4.1: Design of ablation study.

| Input Layer | $R_{km}(\boldsymbol{p})$ First | $PE(\boldsymbol{p})$ First | $PE(\boldsymbol{p})$ Hidden |
|---|---|---|---|
| *Abs-1* | — | ✓ | ✓ |
| *Abs-2* | ✓ | ✓ | — |
| *Abs-3* | ✓ | — | ✓ |
| *Abs-4* | ✓ | — | — |
| *Ours* | ✓ | ✓ | ✓ |

Table 4.2: Result of ablation study.

| | Red pigment map | | Green pigment map | |
|---|---|---|---|---|
| | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ |
| *Abs-1* | 0.897 | 26.1 | 0.767 | 20.9 |
| *Abs-2* | 0.981 | 42.5 | <u>0.986</u> | <u>43.8</u> |
| *Abs-3* | <u>0.988</u> | <u>43.1</u> | 0.923 | 32.8 |
| *Abs-4* | 0.785 | 19.8 | 0.778 | 22.0 |
| *Ours* | **0.998** | **48.6** | **0.997** | **46.9** |

## 4.7 Discussion

This chapter explores applications considering light transport in micro-scale optical path lengths on the order of tens of centimeters. This chapter realizes the application of attenuated-light transport for the pigment mapping for an ancient tomb mural. The approach focuses on the Kubelka-Munk model, which models the attenuation of light in the particle layer to obtain the spatial distribution of pigments. Unsupervised learning is used to simultaneously estimate pigment class, thickness, and substrate class to account for unknown substrate patterns. The experimental results indicate the efficacy of concurrently estimating both the substrate and pigments without disregarding the fact that the substrate is heterogeneous. Moreover, as the ablation study shows, by inputting coordinates and estimation by neural representation, the model takes spatial continuity into account and estimates with less influence on the substrate pattern.

However, this study presents several limitations and future works. While many murals are depicted on relatively smooth rocks, some are located on rocks with significant irregularities. Our current estimation does not account for the three-dimensional structure of murals, potentially resulting in errors when mapping pigments on uneven surfaces, as our future work. Furthermore, we have based our real data on directly obtained pigment parameters from the mural. That is under the assumption that the parameters of the pigments are enough painted at some pixels. Addressing the spectral changes in the mural sections where the pigments have significantly washed out or faded, or have oxidized (undergone
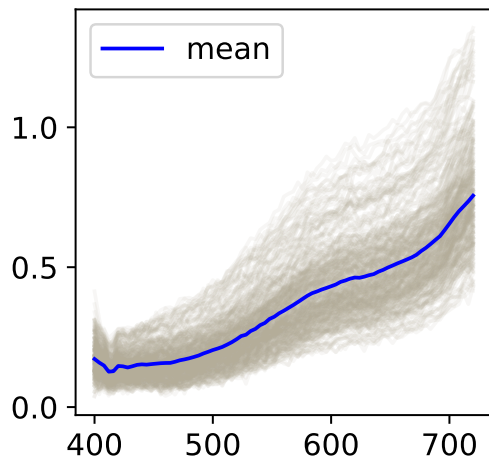
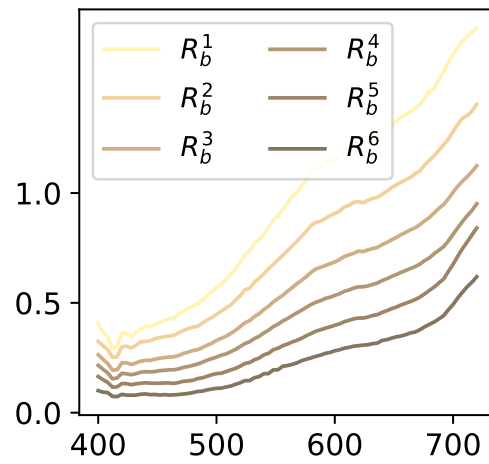chemical changes), is also part of our future work.
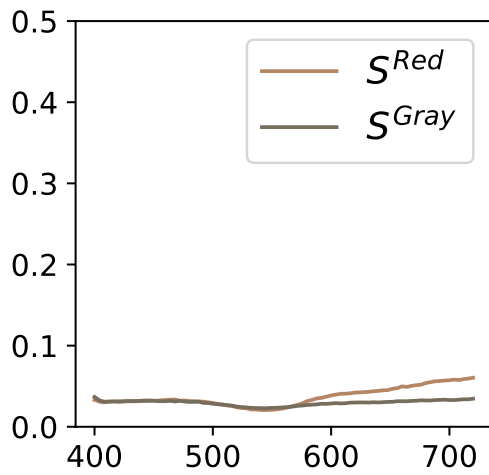
(a) Scattering coefficient
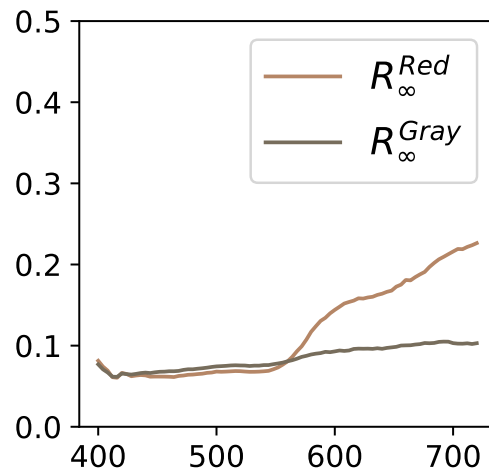
(b) $R_\infty$

(c) Substrates data

(d) Substrates class

Figure 4.5: Optical parameters of Mezurashizuka tomb.
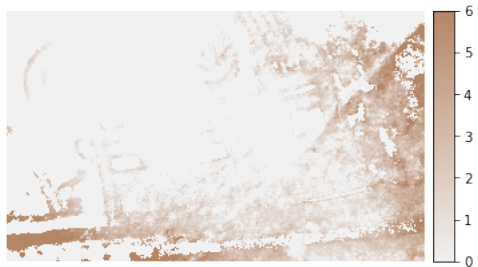
(a) Overview of dataset

(b) Reference image*

(c) Red map ($M = 6$)

(d) Gray map ($M = 6$)

(e) Red map ($M = 1$)

(f) Gray map ($M = 1$)

(g) Reference red map*

(h) Reference gray map*

Figure 4.6: Estimation results for the Mezurashizuka tomb, presented under both a heterogeneous ($M = 6$) and homogeneous ($M = 1$) assumption. The entries marked with an (*) are derived from the restoration reproductions carried out by Hakkou Kusaka in the 1950s. Please note that these reproductions were a subjective task and should not be strictly considered ground truth.
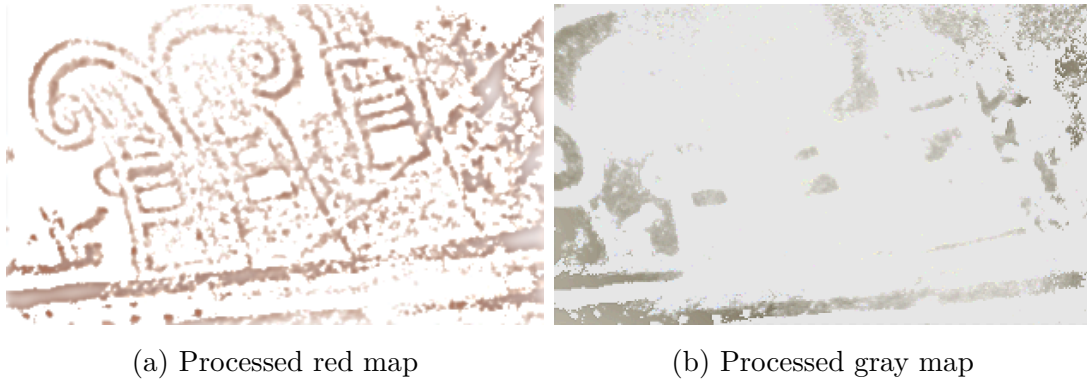
(a) Processed red map

(b) Processed gray map

Figure 4.7: Results of applying the Difference of Gaussian method to our pigment mapping estimation, which was derived from Figure 4.6(b) and (f).



(b) Reflectance of the position indicated by yellow '×' in right-middle images

Red

Gray

$R_{km}$

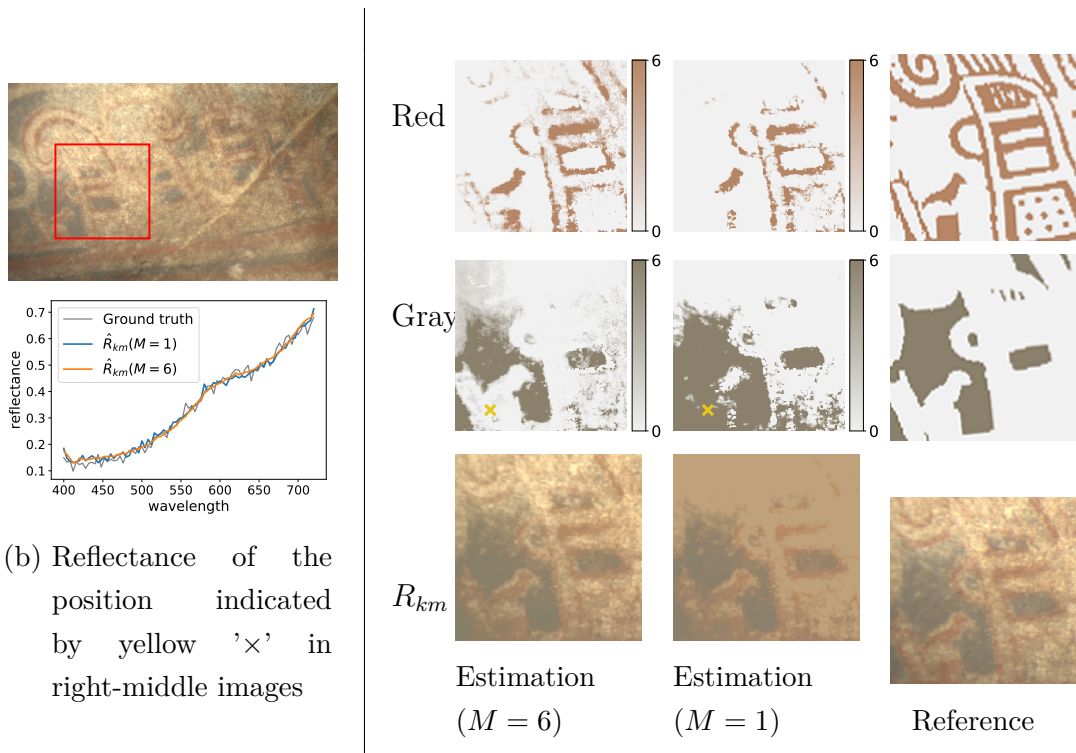Estimation $(M = 6)$

Estimation $(M = 1)$
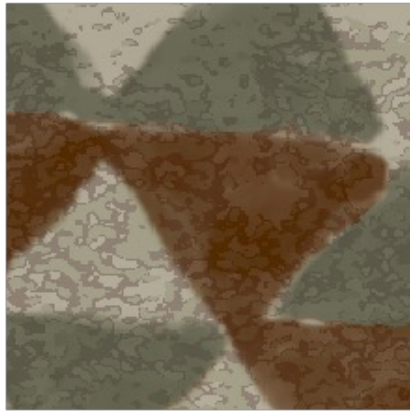
Reference

Figure 4.8: Estimation results in a $200 \times 200$ region. We are comparing the case where the substrate is assumed to be heterogeneous (substrate class $M = 6$) and the case where it is assumed to be homogeneous ($M = 1$).

56

(a) Synthesized dataset
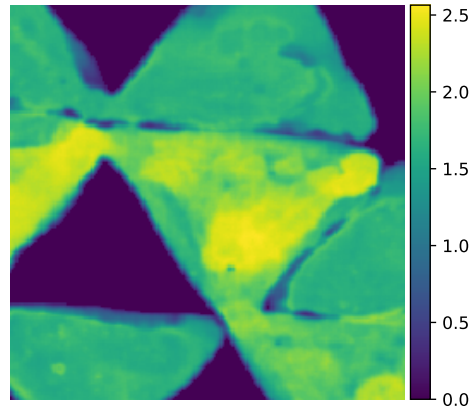


(b) Pigment thickness
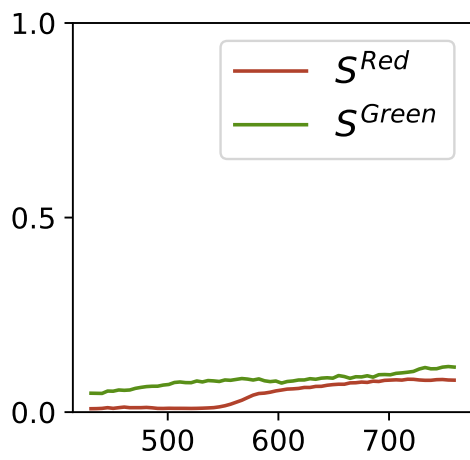


(c) Scattering coefficient
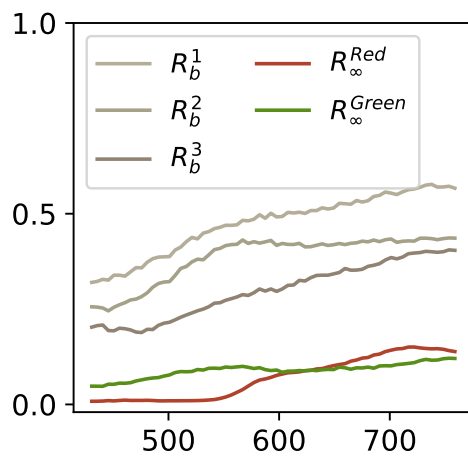


(d) Substrate and $R_\infty$

Figure 4.9: Overviews of synthesized data.
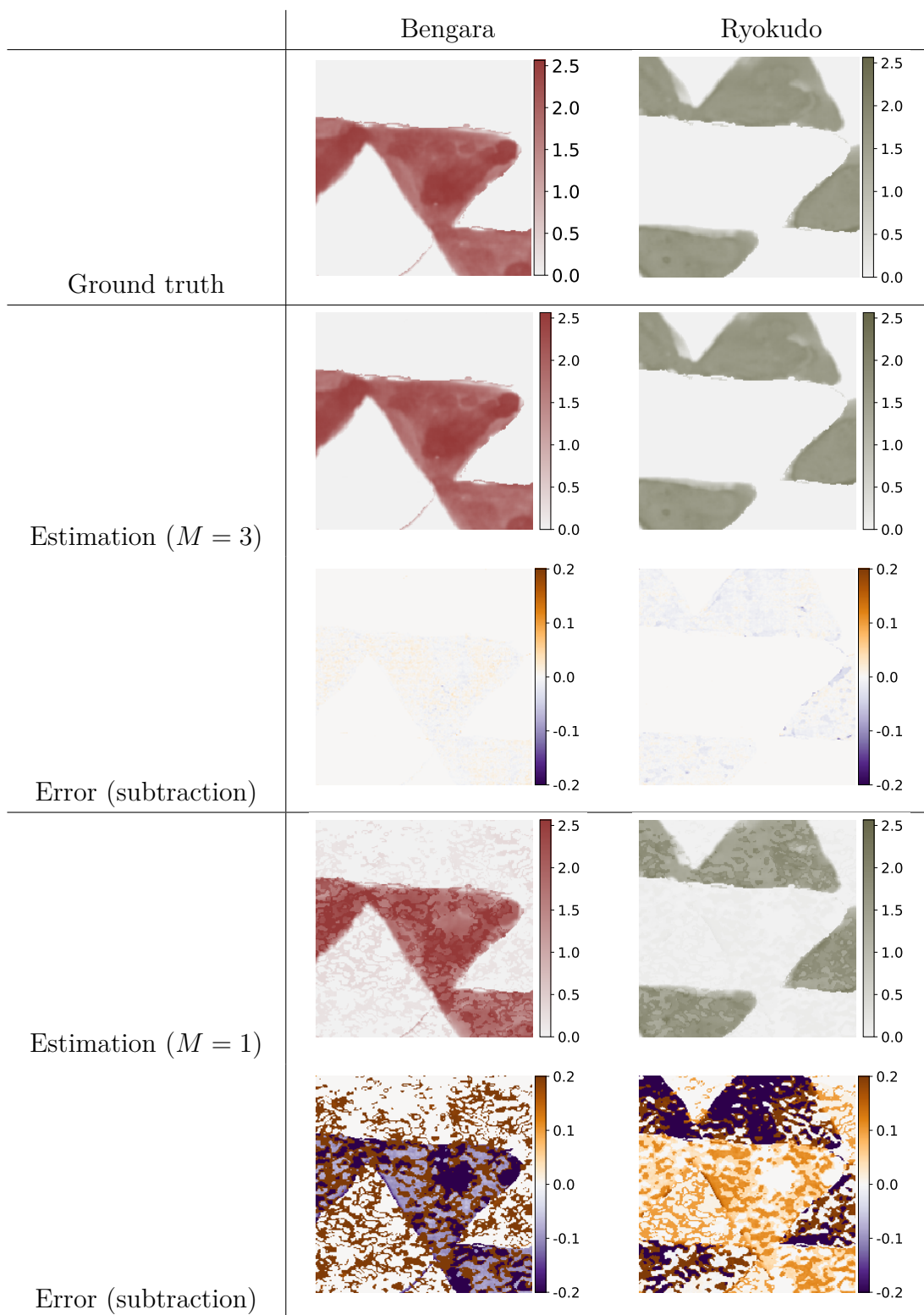
Figure 4.10: Estimation results of simulation data, presented under both a heterogeneous ($M = 3$) and homogeneous ($M = 1$) assumption.

(a) Ground truth

(b) $R_b$ estimation
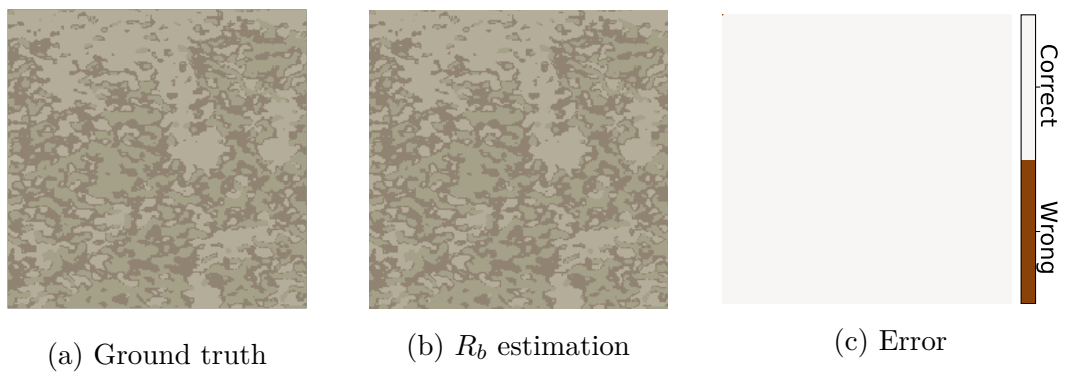
(c) Error

Figure 4.11: Substrate estimation results of simulation data.



(a) Error of *Ours*

(b) Error of *Abs-4*

(c) Substrate

Figure 4.12: Comparison of pigment mapping error between *Ours* and *Abs-4*. *Abs-4* is affected by the substrate pattern.

# 5 Conclusion

In this paper, we pioneer a computer vision application that takes optical transport into account. We have considered light transport in two scales: macroscopic light path length and microscopic light path length. On the former scale, in Chapter 3, we adopted the application of touch sensing to projector images among object position detection and incorporated a measurement m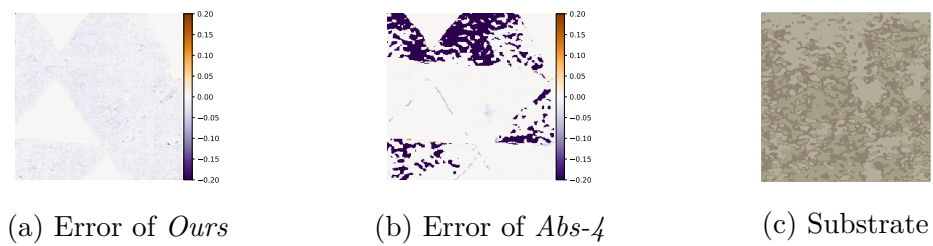ethod that eliminates indirect light as much as possible to obtain light transport in a specific space. In the latter scale, in Chapter 4, among the particle layer analysis, we adopted pigment distribution estimation in heterogeneous substrates and obtained how much of the light transport was the optical path length of the particle layer by solving an inverse problem of the physical model.

In Chapter 3, we focused on Disparity Gating with a projector-camera system that efficiently acquires the desired light transport. In this paper, we specifically apply Slope disparity gating, which acquires direct light in a specific area, to touch sensing on a virtual screen. In other words, the idea is to utilize the projector of the projector-camera system as a visual projection interface and perform camera-based finger position detection. In this case, since the scene contains a lot of unnecessary information for touch sensing, capturing only the direct light from the region necessary for touch sensing through Slope Disparity Gating is more effective than capturing the entire scene.

In Chapter 4, we focused on the light scattering in the particle layer and utilized the Kubelka-Munk model to effectively model it. By solving the inverse problem of the Kubelka-Munk model, we were able to estimate the thickness of the particle layer, which represents the optical path length of the attenuated light transport. We applied this approach to estimate the spatial distribution of pigments in ancient tomb murals. From a preservation perspective, it is more effective to analyze existing data rather than conduct new measurements on an-

cient tomb murals. Additionally, tomb murals are painted on heterogeneous substrates, making the problem more complex compared to previous studies on pigment spatial distribution in oil paintings or watercolors. In contrast, a solution based on inverse problems incorporating deep learning proved to be effective.

**Future Work:**

When considering macroscopic and microscopic optical path lengths, we have chosen applications that are being actively studied in each, but the exploration is still limited. Future work will be to expand these applications.

This thesis assumes diffuse reflection of light, but in reality, it may become necessary to consider specular reflection components as well. For example, touch sensing when a projector is projected onto a marble table in a kitchen is not considered. Additionally, assumptions about minerals used in ancient tomb murals or glossy pigments are ignored. Incorporating the light transport of specular reflection is a highly challenging task and will be considered as future work.

Furthermore, although we have been able to explore applications, there is still room for improvement in their implementation. In Chapter 3, we can implement faster algorithms to accommodate faster finger movements, and since measurements based on infrared are less accurate, there is a need to further improve the estimation accuracy. In Chapter 4, although we have a general understanding of the pigment distribution in ancient tomb murals, we have not yet achieved a perfect estimation. Future work will involve enhancing estimation accuracy by improving deep learning models or utilizing more complex models other than Kubelka-Munk.

# Acknowledgements

First and foremost, I would like to express my profound gratitude to Prof. Yasuhiro Mukaigawa, my principal advisor, for his continuous support, invaluable guidance, patience, and immense knowledge. His expertise and insights have been instrumental to the completion of this thesis.

I am deeply indebted to Prof. Hirokazu Kato, Prof. Takuya Funatomi, Prof. Yuki Fujimura, Prof. Kazuya Kitano, and Prof. Hiroyuki Kubo for their invaluable feedback, encouragement, and enlightening discussions throughout the course of this research.

I also appreciate Dr. Katsushi Ikeuchi, Dr. Tetsuro Morimoto, Prof. Takeshi Oishi, Dr. Jun Takamatsu, and Dr. Mawo Kamakura for their continuous feedback.

I would also like to extend my sincere gratitude to Prof. Suren Jayasuriya and his laboratory members for providing me with the internship opportunity.

Last but not least, I would like to thank my friends and my family, who have been my emotional anchor, offering much-needed breaks, support, and distractions during these challenging times.

# Bibliography

[1] Matthew O'Toole, Supreeth Achar, Srinivasa G Narasimhan, and Kiriakos N Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Transactions on Graphics*, 34(4):1–13, 2015.

[2] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.

[3] Takeo Kanade and Katsushi Ikeuchi. Introduction to the special issue on physical modeling in computer vision. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(07):609–610, 1991.

[4] Katsushi Ikeuchi and Berthold KP Horn. Numerical shape from shading and occluding boundaries. *Artificial intelligence*, 17(1-3):141–184, 1981.

[5] Katsushi Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):661–669, 1981.

[6] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International journal of computer vision*, 48:233–254, 2002.

[7] Srinivasa G. Narasimhan and Shree K. Nayar. Contrast restoration of weather degraded images. *IEEE transactions on pattern analysis and machine intelligence*, 25(6):713–724, 2003.

[8] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980.

[9] Shree K Nayar, X-S Fang, and Terrance Boult. Removal of specularities using color and polarization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 583–590. IEEE, 1993.

[10] Stephen Lin and Heung-Yeung Shum. Separation of diffuse and specular reflection in color images. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

[11] Shinji Umeyama and Guy Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):639–647, 2004.

[12] Stefan Rahmann and Nikos Canterakis. Reconstruction of specular surfaces using polarization imaging. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

[13] Thomas Bonfort, Peter Sturm, and Pau Gargallo. General specular surface triangulation. In *Asian Conference on Computer Vision*, pages 872–881. Springer, 2006.

[14] Sören Kammel and Fernando Puente Leon. Deflectometric measurement of specular surfaces. *IEEE Transactions on Instrumentation and Measurement*, 57(4):763–769, 2008.

[15] Zhan Song, Hualie Jiang, Haibo Lin, and Suming Tang. A high dynamic range structured light means for the 3d measurement of specular surface. *Optics and Lasers in Engineering*, 95:8–16, 2017.

[16] Pankaj B Pathare, Umezuruike Linus Opara, and Fahad Al-Julanda Al-Said. Colour measurement and analysis in fresh and processed foods: a review. *Food and bioprocess technology*, 6:36–60, 2013.

[17] Jingtian Guan, Ji Li, Xiao Yang, Xiaobo Chen, and Juntong Xi. Defect detection method for specular surfaces based on deflectometry and deep learning. *Optical Engineering*, 61(6):061407–061407, 2022.

[18] Henrik Wann Jensen, Stephen R Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 319–326. 2023.

[19] Pat Hanrahan and Wolfgang Krueger. Reflection from layered surfaces due to subsurface scattering. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, page 165‑174. Association for Computing Machinery, 1993.

[20] Srinivasa G Narasimhan and Shree K Nayar. Interactive (de) weathering of an image using physical models. In *IEEE Workshop on color and photometric Methods in computer Vision*, volume 6, page 1. France, 2003.

[21] Takeshi Muraji, Kenichiro Tanaka, Takuya Funatomi, and Yasuhiro Mukaigawa. Depth from phasor distortions in fog. *Optics express*, 27(13):18858–18868, 2019.

[22] Daiki Kijima, Takahiro Kushida, Hiromu Kitajima, Kenichiro Tanaka, Hiroyuki Kubo, Takuya Funatomi, and Yasuhiro Mukaigawa. Time-of-flight imaging in fog using multiple time-gated exposures. *Optics Express*, 29(5):6453–6467, 2021.

[23] Cosmin Ancuti, Codruta O Ancuti, Christophe De Vleeschouwer, Rafael Garcia, and Alan C Bovik. Multi-scale underwater descattering. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 4202–4207. IEEE, 2016.

[24] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1682–1691, 2019.

[25] Peixian Zhuang, Chongyi Li, and Jiamin Wu. Bayesian retinex underwater image enhancement. *Engineering Applications of Artificial Intelligence*, 101:104171, 2021.

[26] Marta Klanjšek Gunde and Zorica Crnjak Orel. Absorption and scattering of light by pigment particles in solar-absorbing paints. *Applied Optics*, 39(4):622–628, 2000.

[27] Bhrigu Rishi Mishra, Nithin Jo Varghese, and Karthik Sasihithlu. Semi-analytical technique for the design of disordered coatings with tailored optical properties. *Optics Express*, 31(6):10201–10216, 2023.

[28] Roy S Berns and Mahnaz Mohammadi. Single-constant simplification of kubelka-munk turbid-media theory for paint systems―a review. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 32(3):201–207, 2007.

[29] Deane B Judd. Fresnel reflection of diffusely incident light. *J. Res. Natl. Bur. Stand*, 29(5):329, 1942.

[30] YiChang Shih, Dilip Krishnan, Fredo Durand, and William T Freeman. Reflection removal using ghosting cues. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3193–3201, 2015.

[31] Zheng Dong, Ke Xu, Yin Yang, Hujun Bao, Weiwei Xu, and Rynson WH Lau. Location-aware single image reflection removal. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5017–5026, 2021.

[32] Juyang Weng, Paul Cohen, Marc Herniou, et al. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on pattern analysis and machine intelligence*, 14(10):965–980, 1992.

[33] Valérie Lavigne and Benoit Ricard. Fast risley prisms camera steering system: calibration and image distortions correction through the use of a three-dimensional refraction model. *Optical Engineering*, 46(4):043201–043201, 2007.

[34] Waldemar H Lehn and Siebren van der Werf. Atmospheric refraction: a history. *Applied Optics*, 44(27):5624–5636, 2005.

[35] Daniel Short, David Voelz, Jose Barraza, and Ivan Dragulin. Atmospheric refraction: applied image analysis and experimental example for index profile with curvature. In *Atmospheric Propagation XIII*, volume 9833, pages 93–100. SPIE, 2016.

[36] Yasuteru Urano, Daisuke Asanuma, Yukihiro Hama, Yoshinori Koyama, Tristan Barrett, Mako Kamiya, Tetsuo Nagano, Toshiaki Watanabe, Akira Hasegawa, Peter L Choyke, et al. Selective molecular imaging of viable cancer cells with ph-activatable fluorescence probes. *Nature medicine*, 15(1):104–109, 2009.

[37] Alyssa B Chinen, Chenxia M Guan, Jennifer R Ferrer, Stacey N Barnaby, Timothy J Merkel, and Chad A Mirkin. Nanoparticle probes for the detection of cancer biomarkers, cells, and tissues by fluorescence. *Chemical reviews*, 115(19):10530–10574, 2015.

[38] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Advances in optics and photonics*, 3(2):128–160, 2011.

[39] You Li and Javier Ibanez-Guzman. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Processing Magazine*, 37(4):50–61, 2020.

[40] Shree K Nayar, Gurunandan Krishnan, Michael D Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. In *ACM SIGGRAPH 2006 Papers*, pages 935–944. 2006.

[41] Matthew O'Toole, John Mather, and Kiriakos N Kutulakos. 3d shape and indirect appearance by structured light transport. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3246–3253, 2014.

[42] Zhangjing Wang, Yu Wu, and Qingqing Niu. Multi-sensor fusion in automated driving: A survey. *Ieee Access*, 8:2847–2868, 2019.

[43] Tomohiro Maeda, Guy Satat, Tristan Swedish, Lagnojita Sinha, and Ramesh Raskar. Recent advances in imaging around corners. *arXiv preprint arXiv:1910.05613*, 2019.

[44] Sreenithy Chandran, Hiroyuki Kubo, Tomoki Ueda, Takuya Funatomi, Yasuhiro Mukaigawa, and Suren Jayasuriya. Slope disparity gating: System

and applications. *IEEE Transactions on Computational Imaging*, 8:317–332, 2022.

[45] Tomoki Ueda, Hiroyuki Kubo, Suren Jayasuriya, Takuya Funatomi, and Yasuhiro Mukaigawa. Slope disparity gating using a synchronized projector-camera system. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2019.

[46] Ove K Steinvall, Hakan Olsson, Göran Bolander, Christina Anna Groenwall, and Dietmar Letalick. Gated viewing for target detection and target recognition. In *Laser Radar Technology and Applications IV*, volume 3707, pages 432–448. SPIE, 1999.

[47] Jian Wang, Joseph Bartels, William Whittaker, Aswin C Sankaranarayanan, and Srinivasa G Narasimhan. Programmable triangulation light curtains. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 19–34, 2018.

[48] Matthew O'Toole, Ramesh Raskar, and Kiriakos N Kutulakos. Primal-dual coding to probe light transport. *ACM Trans. Graph.*, 31(4):39–1, 2012.

[49] Hiroyuki Kubo, Suren Jayasuriya, Takafumi Iwaguchi, Takuya Funatomi, Yasuhiro Mukaigawa, and Srinivasa G Narasimhan. Programmable non-epipolar indirect light transport: Capture and analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(4):2421–2436, 2019.

[50] SK Lee, William Buxton, and Kenneth C Smith. A multi-touch three dimensional touch-sensitive tablet. *Acm Sigchi Bulletin*, 16(4):21–25, 1985.

[51] Dong Wei, Steven Zhiying Zhou, and Du Xie. Mtmr: A conceptual interior design framework integrating mixed reality with the multi-touch tabletop interface. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pages 279–280. IEEE, 2010.

[52] Jun Rekimoto. Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120, 2002.

[53] Chris Harrison, Desney Tan, and Dan Morris. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–462, 2010.

[54] Paul Dietz and Darren Leigh. Diamondtouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226, 2001.

[55] Robert Xiao, Greg Lew, James Marsanico, Divya Hariharan, Scott Hudson, and Chris Harrison. Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 67–76, 2014.

[56] Mark Ashdown and Peter Robinson. Escritoire: A personal projected display. *IEEE MultiMedia*, 12(01):34–42, 2005.

[57] James Patten, Hiroshi Ishii, Jim Hines, and Gian Pangaro. Sensetable: a wireless object tracking platform for tangible user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 253–260, 2001.

[58] Anh Nguyen and Amy Banic. 3dtouch: A wearable 3d input device for 3d applications. In *2015 IEEE Virtual Reality (VR)*, pages 373–373, 2015.

[59] Yilei Shi, Haimo Zhang, Kaixing Zhao, Jiashuo Cao, Mengmeng Sun, and Suranga Nanayakkara. Ready, steady, touch! sensing physical contact with a finger-mounted imu. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):1–25, 2020.

[60] Julien Letessier and François Bérard. Visual tracking of bare fingers for interactive surfaces. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 119–122, 2004.

[61] Itai Katz, Kevin Gabayan, and Hamid Aghajan. A multi-touch surface using multiple cameras. In *Advanced Concepts for Intelligent Vision Systems: 9th International Conference, ACIVS 2007, Delft, The Netherlands, August 28-31, 2007. Proceedings 9*, pages 97–108. Springer, 2007.

[62] Andrew D Wilson. Touchlight: an imaging touch screen and display for gesture-based interaction. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 69–76, 2004.

[63] Joe Marshall, Tony Pridmore, Mike Pound, Steve Benford, and Boriana Koleva. Pressing the flesh: Sensing multiple touch and finger pressure on arbitrary surfaces. In *Pervasive Computing: 6th International Conference, Pervasive 2008 Sydney, Australia, May 19-22, 2008 Proceedings 6*, pages 38–55. Springer, 2008.

[64] Andrew D Wilson and Hrvoje Benko. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 273–282, 2010.

[65] Robert Xiao, Scott Hudson, and Chris Harrison. Direct: Making touch tracking on ordinary surfaces practical with hybrid depth-infrared sensing. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*, pages 85–94, 2016.

[66] Chris Harrison, Hrvoje Benko, and Andrew D Wilson. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 441–450, 2011.

[67] Elliot N Saba, Eric C Larson, and Shwetak N Patel. Dante vision: In-air and touch gesture sensing for natural surface interaction with combined depth and thermal cameras. In *2012 IEEE International Conference on Emerging Signal Processing Applications*, pages 167–170. IEEE, 2012.

[68] Joseph A Paradiso, Kai-yuh Hsiao, Joshua Strickon, Joshua Lifton, and Ari Adler. Sensor systems for interactive surfaces. *IBM Systems Journal*, 39(3.4):892–914, 2000.

[69] Gierad Laput and Chris Harrison. Surfacesight: a new spin on touch, user, and object sensing for iot experiences. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[70] Microvision. Product brief:short-throw interactive display module & starter kit. http://www.microvision.com/wp-content/uploads/2020/01/DB0140010_MV-2407sti-421_Product_Brief.pdf, 2020. (Accessed on June 2021).

[71] Jingwen Dai and Chi-Kit Ronald Chung. Touchscreen everywhere: On transferring a normal planar surface to a touch-sensitive display. *IEEE transactions on cybernetics*, 44(8):1383–1396, 2013.

[72] Dibia Victor. Handtrack: A library for prototyping real-time hand tracking-interfaces using convolutional neural networks. *GitHub repository*, 2017.

[73] Paul Kubelka and Franz Munk. An article on optics of paint layers. *Z. Tech. Phys*, 12(593-601):259–274, 1931.

[74] John W Strutt. Xv. on the light from the sky, its polarization and colour. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 41(271):107–120, 1871.

[75] Hendrik Christoffel Hulst and Hendrik C van de Hulst. *Light scattering by small particles*. Courier Corporation, 1981.

[76] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[77] Paul Kubelka. New contributions to the optics of intensely light-scattering materials. part i. *Josa*, 38(5):448–457, 1948.

[78] Yonghui Zhao, Roy S Berns, Lawrence A Taplin, and James Coddington. An investigation of multispectral imaging for the mapping of pigments in paintings. In *Computer image analysis in the study of art*, volume 6810, pages 65–73. SPIE, 2008.

[79] Daniel W Dichter. Kubelka-munk model of full-gamut oil colour mixing. *Journal of the International Colour Association*, 32:70–78, 2023.

[80] Kensuke Fukumoto, Norimichi Tsumura, and Roy Berns. Estimating pigment concentrations from spectral images using an encoder-decoder neural network. *Journal of Imaging Science and Technology*, 64(3):30502–1, 2020.

[81] Eric Kirchner, Ivo van Der Lans, Frank Ligterink, Muriel Geldof, Art Ness Proano Gaibor, Ella Hendriks, Koen Janssens, and John Delaney. Digitally reconstructing van gogh's field with irises near arles. part 2: pigment concentration maps. *Color Research & Application*, 43(2):158–176, 2018.

[82] Neda Rohani, Emeline Pouyet, Marc Walton, Oliver Cossairt, and Aggelos K Katsaggelos. Nonlinear unmixing of hyperspectral datasets for the study of painted works of art. *Angewandte Chemie*, 130(34):11076–11080, 2018.

[83] Abu Md Niamul Taufique and David W Messinger. Hyperspectral pigment analysis of cultural heritage artifacts using the opaque form of kubelka-munk theory. In *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imagery XXV*, volume 10986, pages 297–307. SPIE, 2019.

[84] Elad Aharoni-Mack, Yakov Shambik, and Dani Lischinski. Pigment-based recoloring of watercolor paintings. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*, pages 1–11, 2017.

[85] Ryuta Shitomi, Mayuka Tsuji, Yuki Fujimura, Takuya Funatomi, Yasuhiro Mukaigawa, Tetsuro Morimoto, Takeshi Oishi, Jun Takamatsu, and Katsushi Ikeuchi. Unsupervised learning with a physics-based autoencoder for estimating the thickness and mixing ratio of pigments. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 40(1):116–128, 2023.

[86] Tetsuro Morimoto, Robby T Tan, Rei Kawakami, and Katsushi Ikeuchi. Estimating optical properties of layered surfaces using the spider model. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 207–214. IEEE, 2010.

[87] John K Delaney, Jason G Zeibel, Mathieu Thoury, ROY Littleton, Michael Palmer, Kathryn M Morales, E René de La Rie, and ANN Hoenigswald. Visible and infrared imaging spectroscopy of picasso's harlequin musician: mapping and identification of artist materials in situ. *Applied spectroscopy*, 64(6):584–594, 2010.

[88] Hilda Deborah, Sony George, and Jon Yngve Hardeberg. Pigment mapping of the scream (1893) based on hyperspectral imaging. In *Image and Signal Processing: 6th International Conference, ICISP 2014, Cherbourg, France, June 30–July 2, 2014. Proceedings 6*, pages 247–256. Springer, 2014.

[89] Antonino Cosentino. Identification of pigments by multispectral imaging; a flowchart method. *Heritage Science*, 2(1):8, 2014.

[90] Tania Kleynhans, David W Messinger, and John K Delaney. Towards automatic classification of diffuse reflectance image cubes from paintings collected with hyperspectral cameras. *Microchemical Journal*, 157:104934, 2020.

[91] Roxanne Radpour, Glenn A Gates, Ioanna Kakoulli, and John K Delaney. Identification and mapping of ancient pigments in a roman egyptian funerary portrait by application of reflectance and luminescence imaging spectroscopy. *Heritage Science*, 10(1):1–16, 2022.

[92] Bartosz Grabowski, Wojciech Masarczyk, Przemysław Głomb, and Agata Mendys. Automatic pigment identification from hyperspectral data. *Journal of Cultural Heritage*, 31:1–12, 2018.

[93] Tania Kleynhans, Catherine M Schmidt Patterson, Kathryn A Dooley, David W Messinger, and John K Delaney. An alternative approach to mapping pigments in paintings with hyperspectral reflectance image cubes using artificial intelligence. *Heritage Science*, 8(1):1–16, 2020.

[94] Ailin Chen, Rui Jesus, and Márcia Vilarigues. Identification and visualization of pure and mixed paint pigments in heritage artwork using machine learning algorithms. *SN Computer Science*, 4(2):115, 2022.

[95] Katsushi Ikeuchi, Tetsuro Morimoto, Mawo Kamakura, Nobuaki Kuchitsu, Kazutaka Kawano, and Tomoo Ikeda. Kyushu decorative tumuli project: from e-heritage to cyber-archaeology. *International Journal of Computer Vision*, 130(7):1609–1626, 2022.

[96] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radi-

ance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[97] Andre Martins and Ramon Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *International conference on machine learning*, pages 1614–1623. PMLR, 2016.

[98] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

[99] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631, 2019.

# Publication List

## Journal Paper (Reviewed)

1. <u>Mayuka Tsuji</u>, Hiroyuki Kubo, Suren Jayasuriya, Takuya Funatomi and Yasuhiro Mukaigawa, Touch sensing for a projected screen using slope disparity gating, IEEE Access, July. 2021

2. Ryuta Shitomi, <u>Mayuka Tsuji</u>, Yuki Fujimura, Takuya Funatomi, Yasuhiro Mukaigawa, Tetsuro Morimoto, Takeshi Oishi, Jun Takamatsu and Katsushi Ikeuchi, Unsupervised learning with a physics-based autoencoder for estimating the thickness and mixing ratio of pigments, JOSA A, Dec. 2022

## International Conference (Reviewed)

1. <u>Mayuka Tsuji</u>, Yuki Fujimura, Takuya Funatomi, Yasuhiro Mukaigawa, Tetsuro Morimoto, Takeshi Oishi, Jun Takamatsu and Katsushi Ikeuchi, Pigment Mapping for Tomb Murals using Neural Representation and Physics-based Model, 4th ICCV Workshop on e-Heritage, Paris, France, Oct. 2023

## Domestic Conference

1. <u>辻茉佑香</u>, 渡辺佳宏, 石川武典, 喜島大揮, 橋本敬志, 久保尋之, 波長選択イメージングによる化粧品の塗布領域の可視化, 映像表現・芸術科学フォーラム 2020 (Expensive Japan 2020), Mar. 2020

2. <u>辻茉佑香</u>, 久保尋之, 舩冨卓哉, 向川康博, Slope Disparity Gating を用いたプロジェクション映像とのタッチセンシングの実現, 第 26 回画像センシングシンポジウム (SSII2020), June. 2020

## Award

1. 第 26 回画像センシングシンポジウム (SSII2020) オーディエンス賞, 画像センシング技術研究会, June. 2020

2. MIRU2020 若手プログラムプレゼンテーション賞, 情報処理学会コンピュータビジョンとイメージメディア (CVIM) 研究会, Aug. 2020