Robust 3D reconstruction in noisy environments

Shirsendu Sukanta Halder

CMU-RI-TR-21-40

August 12, 2021

The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA

Thesis Committee:
Ioannis Gkioulekas, *Advisor*
Matthew P. O'Toole, *Advisor*
Aswin C Sankaranarayanan
Alankar Shashikant Kotwal

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Robotics.

To my parents

Abstract

Automated inspection in industrial manufacturing can minimize the total production cost of a part. Current inspection solutions often involve measuring a part manually, which interrupts the machining process. We present two non-contact real-time systems which integrate visual inspection in-line with CNC (computer numerical control) machines and ensure dimensional model generation of parts with high accuracy. We first present a camera-projector scanning system that uses photometric stereo and structured light scanning to reconstruct the shape of objects in the presence of specular chip-like noise and high-speed object revolution. We obtain reconstruction accuracies down to 0.5 mm for objects with complex reflectance on a representative CNC lathe. For rotationally symmetric objects, we also propose a novel shape from silhouette system which uses principles from light transport theory to efficiently image transmissive paths through a scattering medium. The system enables in-line and highly accurate geometric reconstructions down to 60 μm on CNC lathe machines in the presence of scattering fluid and specular metallic shavings. Both systems are compact and cost-effective alternatives to the current use of CMMs (coordinate measuring machines) for manual inspection of machined parts.

Acknowledgments

Firstly, I would like to start by thanking my advisors Ioannis Gkioulekas and Matthew O'Toole for their support and guidance throughout my time at the Robotics Institute. I am deeply grateful for their belief in me and for giving me the freedom to work on an innovative problem involving research and hardware. Their profound knowledge, patience, and strong commitment kept me motivated throughout my thesis. Apart from being my advisors, they are amazing teachers who taught me courses that were invaluable in developing the theoretical understanding required for this project. There are innumerable things that I have learned from them in these past two years that shaped me up into a better researcher. I could not have asked for better advisors.

I would like to thank Aswin Sankaranarayanan for being a member of my thesis committee. His valuable feedback and suggestions have been very useful in this work. I am indebted to the people at Lockheed Martin Corporation who believed in us and this project. I would especially like to thank Alexander Gutierrez and Amitabh Vyas for being the project supervisors for this project. Their direction and insightful discussions during our biweekly presentations were invaluable in shaping this work into its complete form. I would also like to thank Dean Warner, Justin Solis, and Samuel Marriott-Green for their practical ideas and advice incorporated throughout this project. My MSR thesis would not have been possible if not for Timothy Angert, the RI machine shop manager. Tim was kind to lend us his CNC lathe machine and the space around it for our experiments without any hesitation. Tim was extremely helpful in setting up our imaging setup on the CNC lathe machine by building custom stages and machining objects. He also trained me to use CNC machines safely.

My time at the Robotics Institute would have been dull without my awesome lab-mates. I am thankful to Dorian Chan for going through my presentation slides, project documents, and thesis and providing important suggestions. I would also like to thank Adithya Pediredla and Ben Attal who were monumental in helping me set up my hardware equipment in my living room when COVID shut down the campus! My lab life as well as my personal life in Pittsburgh would be incomplete without Alankar Kotwal. He has been a friend, guide, mentor, confidant, and most importantly my table-tennis partner. Our daily discussions on research, courses, life, the universe, and everything were very engaging and helped me grow as a

person. He is an amazing researcher and a more amazing person and I wish for him nothing but the best.

I would like to acknowledge my friends in MSR, MSCV, and MRSD for gracing my life with their presence: Divam Gupta, Siddharth Agarwal, Harsh Sharma, Karthik Arcot, Mayank Singhal, Rohan Rao, Anand Bhoraskar, Arpit Jangid, Prakhar Kulshreshtha, Sanil Pande. My CMU life would be uneventful if not for our assignment sessions, get-togethers, and outings. I would especially like to thank my project partners: Rohit Jena, Ankur Deka, Mosam Dabhi, and Vinay Sameer Raja Kadi for imparting their knowledge during our course project sessions. I appreciate the love and support I received from my friends during good and bad times: Romel Baral, Sachit Malik, Sarthak Munshi, Divyansh Arora, Nishit Mengar, Nikhil Angad Bakshi, Anmol Jagetia. Pittsburgh is incomplete without you!

I am deeply indebted to my friends before graduate school for contributing in shaping me into who I am: Sanatan Sharma, Asutosh Palai, Ratnesh Madaan, Utkarsh Bajpai, Rounak Banik, Anushrut Gupta, Shivani Sharma, Rahil Goyal, Shiva Soni, Chetan Maheshwari, Anil Jatain, Ayush Daruka, Shivani Khokhani, Arpit Gaur, Divya Govekar, and Nithin Nair. Lastly, I would like to express deep gratitude to my family and my parents for their unconditional love, endless support, and unfathomable belief.

Funding

This research has been supported by the Master Research Grant by Lockheed Martin Corporation MRA19-001-RPS002.

Contents

When this dissertation is viewed as a PDF, the page header is a link to this Table of Contents.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

CNC machining is ubiquitous in today's manufacturing industries like aerospace and aircraft, automotive, and electronics. CNC stands for computer numerical control, and it is a subtractive manufacturing process that typically uses computerized controls and machine tools to remove layers of material from a raw stock piece (e.g., a block of aluminium metal).

It is impossible to manufacture parts on CNC machines like routers or lathes to exact specifications due to imprecision in the machining process. This makes inspection an integral part of the manufacturing process. Inspection is the process by which a fabricated product is compared with the specifications defined at the design stage. Automatic inspection for CNC machines is preferred over manual inspection due to the following reasons:

1. Automatic inspection is devoid of manual errors which arise due to inconsistency in product evaluation by humans.
2. Automatic inspection makes inspection possible in dangerous environments.
3. Automatic inspection consumes less time which decreases the production cost.

Automatic inspection methods are broadly classified into two categories: non-contact (Aguilar et al. [2], Ahn and Schultes [3], Brosed et al. [11], Connolly [15], Kostamovaara et al. [39], Li et al. [41], Reinhart [57], Smutny et al. [61], Usamentiaga

et al. [68]) and contact methods (Johnson [34], Vermeulen et al. [69], Zhou et al. [79]). Contact based inspection techniques are normally carried out using a coordinate measuring machine (CMM). A list of inspection methods is presented in Fig. 1.1. The final detailed inspection is performed using any one of the contact/non-contact

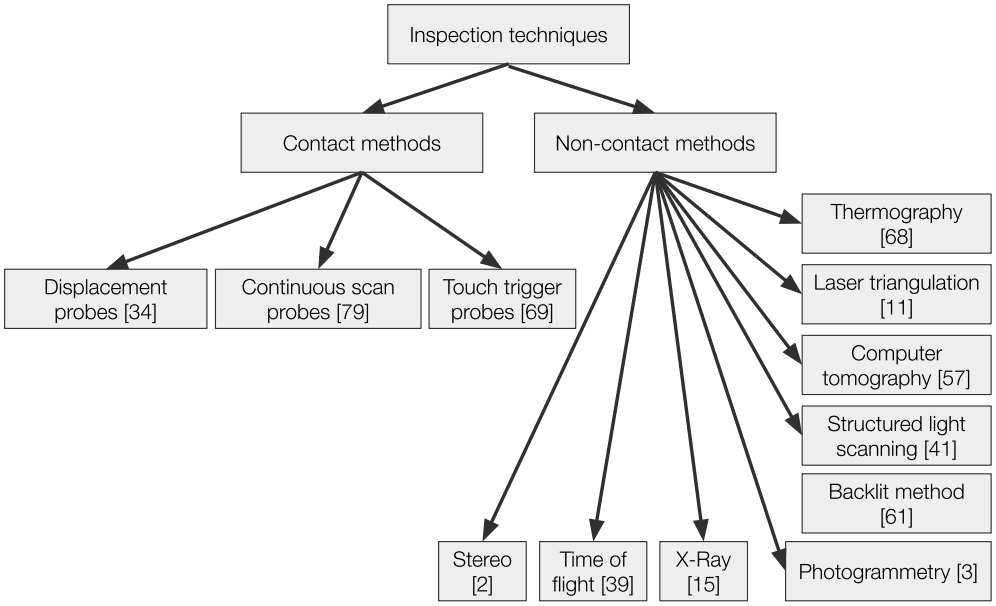


Figure 1.1: Automatic inspection methods classification.

methods mentioned above. If a defect is detected in this final step, the part needs to be reworked or a whole part needs to be re-manufactured again from raw stock, increasing the production cost. Therefore, an important step in the production pipeline is to ensure real-time precision while parts get manufactured.

Contact based systems cannot be used for real-time inspection due to their bulkiness and extremely slow data captures. They also interrupt the machining process by introducing manual work as the manufactured part has to be accurately placed. Current non-contact systems address these limitations but they are less accurate than CMMs and break down in non-ideal environments for the following reasons:

1. CNC machines often machine metallic objects that exhibit complex surface reflectance properties that are non-ideal for vision based setups.
2. The shearing of the raw stock surface generates metallic shavings introducing,

sparse random noise in the environment. These metal shavings might also damage the measurement equipment.

3. The friction between the cutting tool and the part generates a prohibitive amount of heat which threatens the tool and the raw stock. They also introduce unnecessary roughness on the object's surface. Coolant fluids are used as a lubricant to dissipate heat and to produce smoother parts. However, the light scattering properties associated with these fluids hinder vision-based imaging setups.
4. CNC machines tend to experience severe mechanical vibrations which might be undesirable for measurement tools.

1.2 Contributions

This thesis serves as an attempt to bridge the gap between high accuracy dimensional estimates and efficient in-line parts inspection. Specifically, we explore imaging-based non-contact methods that ensure in-line high-quality parts inspection on CNC lathe machines in noisy environments.

- We first explore the challenges of using structured light scanning methods and photometric stereo in CNC environments. We demonstrate the capability to measure objects up to an accuracy of $\frac{1}{50}^{\text{th}}$ of an inch (0.5 mm). We perform our experiments in a lab-based setting where we mitigate the effect of specular chip-like noise using a matrix decomposition method. We also translate this system on a novel CNC machining environment and demonstrate reconstructions of shiny metallic objects rotating at high revolutions per minute (RPM).
- We propose a novel shape from silhouette system that can reconstruct the geometry of objects on a CNC lathe machine through dense scattering medium and in the presence of sparse specular noise in the form of metallic shavings. For descattering, we incorporate multiple clues from light-transport theory to selectively separate transmissive paths from scattered paths. This ensures high-contrast, non-blurry silhouetted images of objects. We also introduce a spatio-temporal processing method to mitgate degradations in our image measurements due to the presence of fluid, specular chips, high-speed rotation,

and ambient light. This enables reconstruction accuracies up to 60 µm for representative metallic objects. We also test our system in a lab-based setting in the presence of a dense scattering medium with pathlength up to 5 cm to demonstrate the descattering capabilities of our setup.

1.3 Outline

This thesis is organized as follows:

- Chapter 2 introduces background literature relevant to our works.
- Chapter 3 introduces our structured light and photometric stereo setup. It provides the theory for the imaging setup, followed by simulation experiments acting as a proof of concept, and experimental results in a controlled lab setting and CNC machining environment.
- Chapter 4 proposes our ballistic shape from the silhouette scanning method. It explains the light behaviour in a scattering medium, followed by the theory for our setup, and concludes with experimental evaluations in different scanning environments.
- Chapter 5 summarizes our work on robust 3D reconstruction and discusses the limitations of our methods.

Chapter 2

Background

This section provides a brief review of background literature in 3D reconstruction, including: (i) photometric stereo, (ii) structured light scanning, (iii) depth and normal fusion, and (iv) light path separation. The first three components focus on 3D shape estimation methods that are relevant to Chapter 3, and the fourth is relevant to Chapter 4. Component one focuses on estimating surface orientation (i.e., normals) using photometric stereo for both Lambertian and non-Lambertian objects. Component two explores structured light scanning methods for depth estimation. Component three surveys algorithms that combine depth and normal estimates for improving 3D reconstructions. The final component focuses on light path separation methods, which is used in our proposed setup for 3D reconstruction of rotationally symmetric objects.

Photometric stereo: The pioneering work of Woodham [74] introduces estimation of normals of a surface using three or more images. However, this work assumes that objects are always Lambertian which is never the case in real life. A lot of work has been proposed in computer vision literature for shape estimation of objects having arbitrary reflectance. Hertzmann and Seitz [22] proposed example-based photometric stereo where they use a reference object to impart normal directions to a target object using orientation-consistency. Mallick et al. [45] and Zickler et al. [80] eliminate the requirement of a reference object by using the dichromatic reflectance model and colour space consistency to separate the diffuse and specular components. Goldman

et al. [18] capture the shape as well as the spatially-varying BRDFs using an optimization framework based on the observation that for general objects, the pixels can be represented as a combination of one or two different materials. Barsky and Petrou [7] use a 4-source colour photometric stereo procedure that uses spectral information as additional visual cues for detecting the specular highlights and separating local gradient information in an object.

Other methods exploit the symmetries in BRDFs to reconstruct the geometry of surfaces. Zickler et al. [81] use Helmholtz reciprocity principle for stereopsis which says that the pixel information of a surface point does not change if the ray direction and viewing direction are interchanged. Alldrin and Kriegman [5] explore the symmetry of BRDFs across the plane spanned by the normal and the viewing direction and apply this constraint for photometric stereo. Holroyd et al. [25] reconstruct complex isotropic and anisotropic object shapes using the analysis that BRDFs are symmetric across the surface normal, normal-tangent plane, and the normal-bitangent plane. They capture at least 100 images with varying light directions in the upper hemisphere and optimize a distance function to get the per-pixel normals and tangents. Lu et al. [43] extend the 1D/2D halfway symmetric property of BRDFs to uncalibrated lighting situations using an elevated angle optimization method.

Robust photometric stereo for varying BRDFs have been explored under dictionary-based data-driven frameworks in Alldrin et al. [4], Hui and Sankaranarayanan [29], and Wagenmaker et al. [70]. Another direction of work exploits the rank-3 constraints of the normals. Wu et al. [75] use low-rank matrix recovery and completion to decompose corrupted images into diffuse and error components. They perform photometric stereo using this diffuse component which is non-susceptible to errors caused by specularities and shadows. Ikehata et al. [30] use Bayesian approximations to separate diffuse and non-Lambertian components. Photometric stereo has also been combined with other shape estimation methods to increase generalization and accuracy.

Structured light scanning: Structured light is another widely-used approach for active scanning based shape acquisition. Structured light systems provide ambient illumination independence over photometric stereo methods. Photometric stereo methods preserve high-frequency details and have low-frequency biases while struc-

tured light techniques are robust in low-frequency depths but lack high-frequency details. Active triangulation used by structured light systems involves projecting a light pattern on the scene and using the distorted geometric relations of the object to reconstruct the shape. A prominent work using laser line scans for 3D shape recovery is the digital Michaelangelo project by Levoy et al. [40] where they reconstruct the statue of David merging 2 billion polygons and 7000 images. Using laser lines are accurate but very time-consuming due to the huge number of patterns required to find correspondences for triangulation. A lot of research in the last couple of decades focused on proposing different coding patterns for time-multiplexing with varying spatial frequency. Posdamer and Altschuler [56] pioneered the direction of structured light by proposing a method that involved projecting a sequence of m binary patterns to encode 2^m different columns. Binary codes are noisy especially near the edges of the projected patterns and this was resolved by binary gray code patterns as proposed in Inokuchi [31]. An important property of gray code patterns are that any two consecutive codes will differ by only one bit which makes it more robust to correspondence errors. Caspi et al. [13] introduced scanning with n-ary codes where m^n stripes could be coded using m patterns. Structured light scanning has also been explored in a spatial neighbourhood framework using coloured De Bruijn patterns by Boyer and Kak [10].

Another group of patterns known as phase-shifting have continuous values instead of binary codes. Due to the non-discrete nature of phase-shifting patterns, correspondences are obtained at a very high sub-pixel resolution. The standard N-step phase-shifting algorithm was proposed by Srinivasan et al. [63] where a set of sinusoidal wave patterns of a certain frequency are projected into the scene. These patterns are then encoded using a phase unwrapping method. Since, there are 3 unknowns in the solution equation, a lot of the methods in literature use 3-step phase shift patterns (Bruning et al. [12]). Another set of works use 4-step shifted sinusoidal patterns for better correspondence retrieval as explored in Abdelsalam et al. [1] and [64]). The conventional 3-step shifting algorithm is not robust and suffers from incorrect phase estimations and non-linearity error. These are corrected using a double 3-shift algorithm proposed in works by Huang et al. [27] and Huang and Zhang [26] where they project 3-shift patterns of two different frequencies producing phase-maps that cancel each other's errors on averaging. Other works include the

Hariharan 5-step phase shifting method (Hariharan et al. [21]), triangular phase shift patterns (Jia et al. [33]), and trapezoidal phase shift patterns (Huang et al. [28]). Phase-shift patterns offer very high resolution but are limited spatially due to their repetitive nature. To overcome this limitation, several works use hybrid binary patterns along with phase-shift patterns. Gray code patterns give a coarse map for the whole scenario and phase-shift patterns are used for refining correspondence maps as presented in Zhang [78] and Wang et al. [73]. For our experiments, we use gray code patterns with double three-step phase shifting patterns. Recently, a couple of works Chen et al. [14], Mirdehghan et al. [47] propose automatic optimal generation of structured light patterns using response inputs when a set of coded patterns are emitted into the scene. Mirdehghan et al. [47] optimize the system using a MLE framework while Chen et al. [14] use stochastic gradient-descent for optimization.

Depth and normal fusion: Efficiently combining normal estimates and 3D locations offer us exact 3D reconstructions in terms of depth absolute accuracy as well as intricate surface details. Song and Chung [62] illuminate the scene with a coloured pseudorandom rhombic grid of patterns. Using the grid points between patterns as the feature points, they estimate the position as well as the orientation and combine them to get a better surface estimate. Lu et al. [44] propose a novel 3D imaging framework that combines coarse depth maps with fine normal maps. They capture the depth equipped with a gray code structured light system. A macro lens setup captures a focal stack of the object to extend the depth of field and then performs photometric stereo on this focal stack to get the high-resolution normal map. They fuse the normals and depth with a multi-resolution patch-based surface reconstruction method. Another work by Haque et al. [20] combines photometric stereo with depths from an infra-red camera. They reconstruct the surface geometry using an adaptive weighing optimization algorithm that balances the reliability of the normal and depth map. Zhang et al. [77] extend the fusion of normals and depth to dynamic scenes by using a novel reconstruction method that considers the depth discontinuities due to motion. Nehab et al. [52] propose a linear framework for combining orientation and positional information that proceeds in two separate steps. The reconstruction algorithm approximates a surface by minimizing a combination of position error and normal error. The position error constrains the surface to match

the dense depth map while the normal error causes the tangent to be orthogonal as they are reprojected onto the normals. Joshi and Kriegman [35] extend Nehab et al. [52] by adding a smoothness term on the second derivative of the estimated mesh. A novel total generalized variation approach by Antensteiner et al. [6] improves the surface reconstruction by minimizing the gradient distance of the surface orientation.

Light path separation: Nayar et al. [51] first demonstrated that with a set of high frequency three-shift sinusoidal patterns, the scene can be decomposed into a direct illumination component and a global illumination component. Gu et al. [19] extend this work for recovering N direct components for N different light sources using multiplexed illumination. The key idea of modulated multiplexing is to use a unique frequency for each light source. Seeing through diffuse media by capturing select light path has been explored through transient imaging techniques. These works use sophisticated hardware equipment like streak cameras in Wang et al. [72], continuous-wave time-of-flight cameras in Kadambi et al. [36], and single-photon avalanche diodes in Boccolini et al. [9]. Another direction of work, especially involving biological microscopy, uses optical coherence tomography for controlling light propagation in a scattering medium (Nasr et al. [50], Popoff et al. [55]). Descattering by scattered path rejection has been explored under a polarization setting in Gilbert and Pernicka [17] and Treibitz and Schechner [67]. A number of light probing techniques have also been presented for separating light paths using a projector-camera system. O'Toole et al. [53] establish that image formation in a projector-camera system is dominated by epipolar and non-epipolar paths. They propose a novel technique called Structured Light Transport that combines light transport with stereo geometry in the optical domain to selectively image direct and global illumination components in real-time. In another work by O'Toole et al. [54], a low-light projector and a rolling shutter camera is used to reconstruct the 3D shape of objects through smoke. This is done through structured light epipolar-only probing that rejects non-epipolar scattered lights. Mukaigawa et al. [49] present a new analytical method for light transport in a scattering medium. They use 1-D high-frequency patterns to separate single and multi-bounce light paths. They further decompose the complex multi-bounce light field into individual bounce parts from the light transport equation. In a series of works by Tanaka et al. [65, 66], an infra-red parallel high-frequency illumination

measurement system is used to separate transmissive and scattered light paths. They demonstrate their results for descattering as well as recovering inner slices in translucent objects. Our proposed shape from silhouette setup in Chapter 4 efficiently combines light path probing techniques with descattering cues from light transport theory to image ballistic photons for scanning objects through a scattering medium.

Chapter 3

Combining surface orientations and range estimates for 3D geometry

In this chapter, we use existing 3D geometry methods for shape reconstruction in noisy environments and CNC machining environments. This chapter provides insight into the performance of previously-proposed algorithms for complex reflectance objects in non-ideal situations like high-speed revolution and ambient noise. Sec. 3.1 focuses on normal estimation methods using photometric stereo. Sec. 3.2 explores depth estimation methods using different types of time-multiplexing structured light coding strategies followed by a description of a simple linear framework for combining depth and normal estimates into a surface mesh. Sec. 3.3 explores a matrix factorization method to mitigate specular sparse noise exhibited in noisy machining environments for photometric stereo and structured light scanning. Sec. 3.4 shows some simulated scanning and reconstruction results for objects with different reflectances. Sec. 3.5 describes our hardware implementation, calibration methods for our system, followed by qualitative and quantitative results for 3D geometry reconstruction of objects with different BRDFs (Bi-directional Reflectance Distribution Functions) in a controlled environment and machining environment.

3.1 Surface orientation estimation

3.1.1 Photometric stereo for normal maps

A common way of representing the surface geometry of an object is through a per-pixel orientation map. Photometric stereo is a technique used to recover the surface structure using multiple images under different illumination captured from the same viewpoint. It exploits image pixel intensity changes arising due to unique illumination positions for static objects to generate these orientation measures. For any Lambertian object, the image captured can be represented as:

$$\mathbf{I} = \rho \, (\mathbf{N} \cdot \mathbf{L}) \tag{3.1}$$

where $\mathbf{N}$ represents the normals, $\mathbf{L}$ are the positions of the light sources in 3D, and ρ is the diffuse albedo of the object. To recover the normals, at least 3 images under different illumination are required. We solve this problem using a least-squares method to estimate the pseudo-normals, which can be decomposed into unit normals and albedos. The positions of the light sources can be estimated through a calibration method that detects specularity of a mirror-like chrome sphere. In real-life, objects are never perfectly Lambertian which is why photometric stereo fails in such cases. In the next sub-section, we explore a more generalized photometric stereo method for estimating normals for objects having spatially varying BRDFs.

3.1.2 Improving normal maps for complex isotropic and anisotropic objects

One of the key areas in computer graphics involves describing the interaction of light and objects realistically. This interaction of light with different materials can be described by the BRDF. A BRDF, $f(\omega_i, \omega_o)$, is defined as the ratio of radiance reflected from a surface in the direction ω_o corresponding to a unit incoming irradiance from direction ω_i. Both ω_i, ω_o are characterized by their azimuthal and zenith angles $[\theta_i, \phi_i]$ and $[\theta_o, \phi_o]$ respectively. It is defined as:

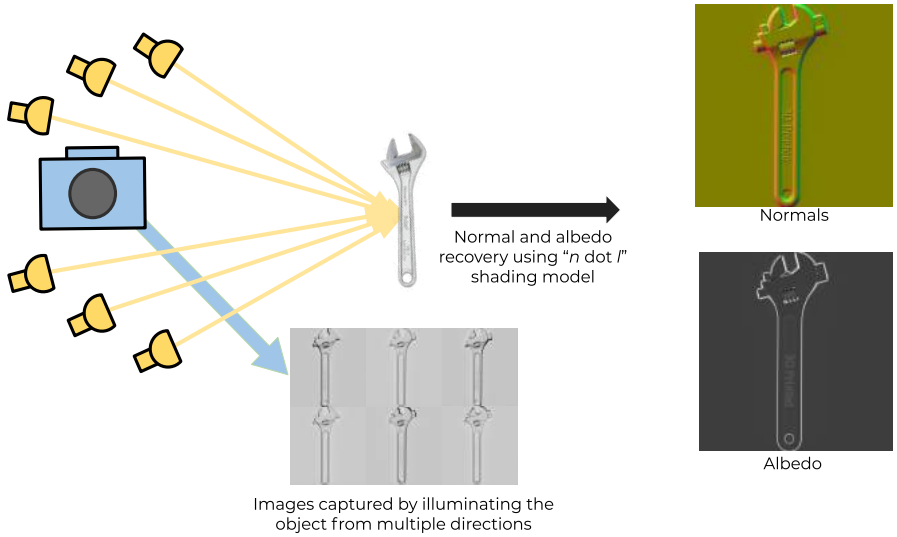


Figure 3.1: Photometric stereo used for estimating per-pixel normals and albedos.

$$f(\omega_i, \omega_o) = \frac{dL_r(\omega_o)}{dE_i(\omega_i)} = \frac{dL_r(\omega_o)}{L_i(\omega_i)\cos\theta_i d\omega_i} \tag{3.2}$$

$$E_i = \text{Irradiance of the surface in direction } (\theta_i, \phi_i)$$
$$L_r = \text{Radiance of the surface in direction } (\theta_o, \phi_o)$$

where n is the normal of the surface as shown below. An elaborate discussion about BRDF and reflectance models can be found in [58, 71].

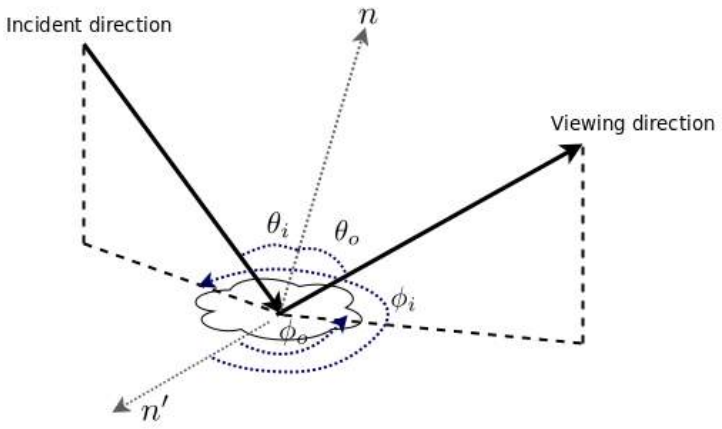


Figure 3.2: Bi-directional Reflectance Distribution Function (BRDF) representation.

In CNC manufacturing settings, objects are metallic, thus rendering Lambertian assumptions in photometric stereo inaccurate. To obtain consistent surface orientations of real-world objects with varying BRDFs, we use the technique described in Holroyd et al. [25]. The algorithm leverages reflective symmetries observed in 2D BRDF slices, parametrized using the halfway vector. The halfway vector is described using the light position and the viewing position as $h = \frac{l+v}{|l+v|}$. We lit the scene with a set of 150 different light positions in the visible hemisphere and construct the 2D BRDF slices $\beta_v(\theta_h, \phi_h)$ for every single pixel (Fig. 3.3). Here, $\theta \in [0, 90]$ represents the polar angle while $\phi \in [0, 180]$ is the azimuth angle.

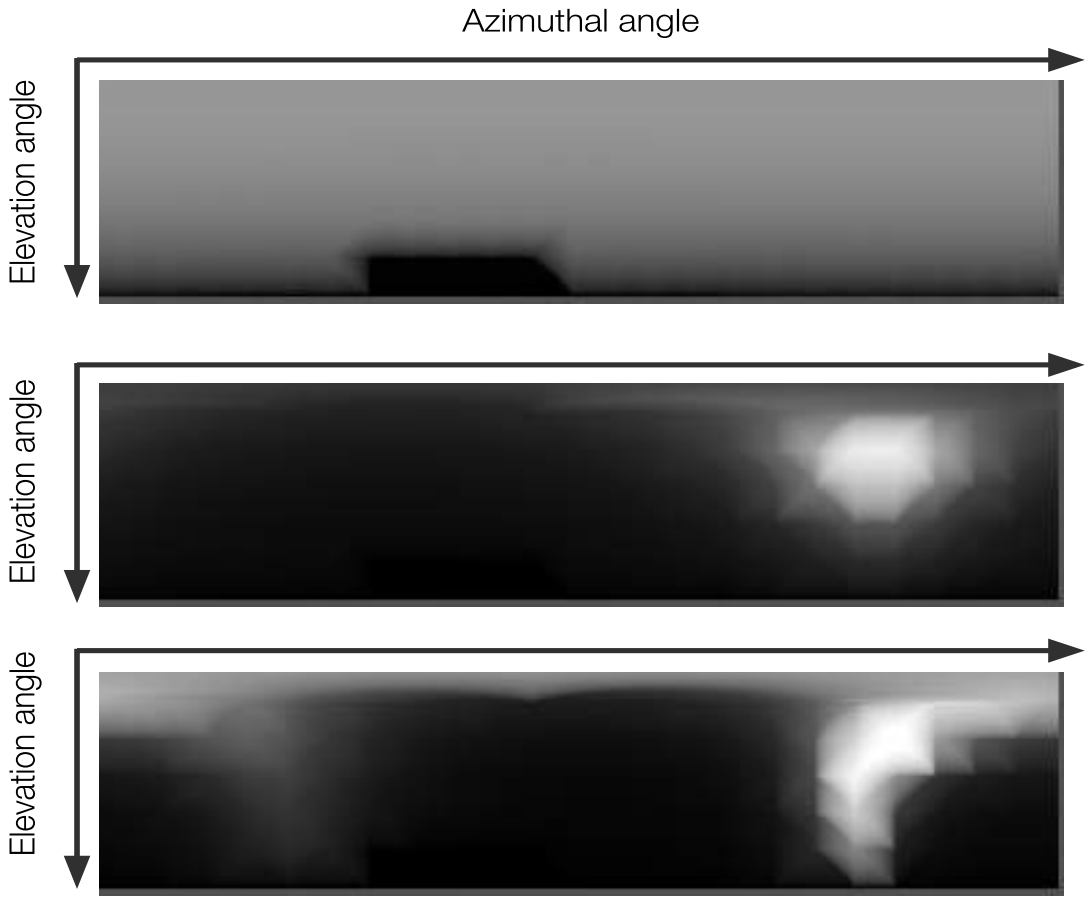


Figure 3.3: 2D BRDF slices of diffuse, specular isotropic, and anistotropic BRDF.

We exploit symmetries across the normal, normal-tangent, and the normal-bitangent planes as follows:

1. Reflection of h across the normal-tangent **nt** plane
   $\beta_v(\theta_h, \phi_h) = \beta_v(\theta_h, -\phi_h)$
2. Reflection of h across the normal-bitangent **nb** plane
   $\beta_v(\theta_h, \phi_h) = \beta_v(\theta_h, -\phi_h + \pi)$
3. Reflection of h across the surface normal
   $\beta_v(\theta_h, \phi_h) = \beta_v(\theta_h, \phi_h + \pi)$

Using the generated BRDF slices and the symmetry functions, we optimize a symmetry distance (Eq. 3.3) that uses the Singer and Nelder [59] simplex algorithm:

$$SD(\mathbf{n}, \mathbf{t}) = SD_{\tau_{\mathbf{nt}}}(\mathbf{n}, \mathbf{t}) + SD_{\tau_{\mathbf{nb}}}(\mathbf{n}, \mathbf{b}) + SD_{\tau_{\mathbf{n}}}(\mathbf{n}) \tag{3.3}$$

$$SD_{\tau}(\mathbf{n}, \mathbf{t}) = \frac{\int_{\Omega_\tau} ||(\mathbf{n} \cdot \tau(\mathbf{l}))I_1 - (\mathbf{n} \cdot \mathbf{l})I_{\tau(\mathbf{l})}||^2 d\omega_{\mathbf{l}}}{\int_{\Omega_\tau} ||(\mathbf{n} \cdot \tau(\mathbf{l}))I_{\mathbf{l}}||^2 d\omega_{\mathbf{l}}} \tag{3.4}$$

Here, $\mathbf{b} = \mathbf{n} \times \mathbf{t}$, $\tau(\mathbf{l})$ describes any arbitrary transformation on the lightning positions $\mathbf{l}$, and Ω_τ is the domain of integration. Light positions $\mathbf{l}$ and their transforms $\tau(\mathbf{l})$ that are not in the upper hemisphere or whose halfway vectors $h_{\mathbf{n}}, h_{\tau(\mathbf{l})} \notin [0, \theta_{d_{max}}]$ are culled away.

3.2 Range scanning using structured light

3.2.1 Multiplexing structured light line scanning using gray codes

Normal maps can be integrated using a normal integration procedure to derive the object depths. However, these integrated normal maps are very sensitive to small noise or deviations. In addition, they do not offer accurate relative depth estimates and suffer from low-frequency bias. Therefore, we explore explicit depth estimation algorithms using structured light methods. A schematic representation of the general structured light scanning setup is shown in Fig. 3.4

A simple way of using structured light scanning is to use a rectified camera-projector setup to project single lines sequentially on the scene and compute the

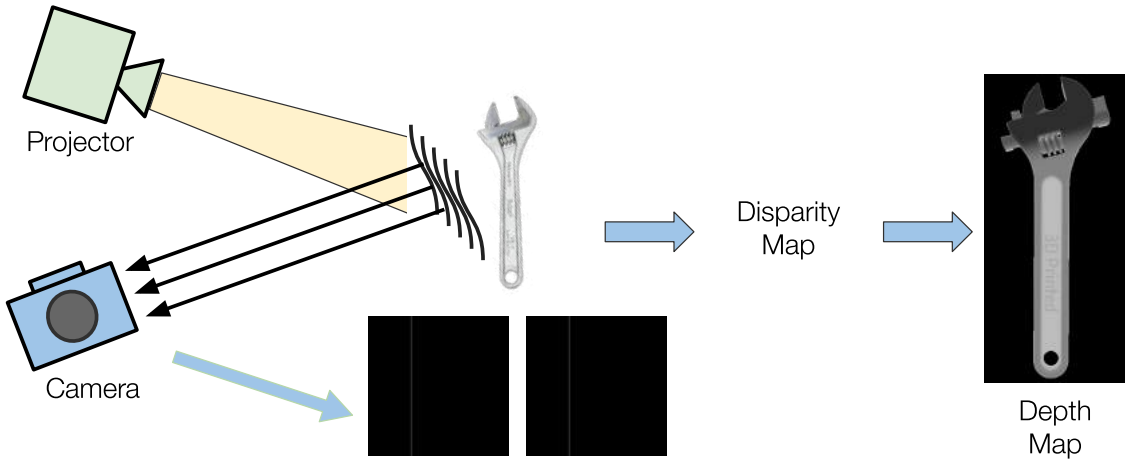


Figure 3.4: Schematic of a general structured light line scanning setup.

disparity map. In practical situations, single line structured light scanning is forbidden due to two factors: the number of images required for line scanning is too high (equivalent to the resolution of the projector in the horizontal direction) and rectifying a camera projector system is inconvenient. Therefore, practical scanning systems use time-multiplexed spatial binary patterns for object scanning. These encoded binary patterns projected on the objects are captured by the camera and then decoded to find the correspondence map. For our experiments, we use gray code patterns which are a type of sequential binary codes that vary only at a 1-bit position. They are chosen for their error correction capabilities in case of discontinuities around spatial pattern edges due to projector hardware limitations. The binary images are created by converting the indices of each column into its corresponding gray code and then traversing the bits from the most significant bit (MSB) to the least significant bit (LSB). Some example gray code patterns are shown in Fig. 3.5. These encoded patterns are used for getting the horizontal and the vertical correspondence maps. The object shape is reconstructed using 3D triangulation from the distortions of these correspondence maps.

Figure 3.5: Generated binary gray code patterns used for structured light scanning.

3.2.2 Improving reconstruction resolution using phase shifting fringe projection

Gray code patterns guarantee uniqueness but have limited resolution and smoothness due to their discrete nature causing harsh quantization effects, especially at pattern edges. This creates rough surface meshes with redundant repetitive high-frequency artifacts. To get rid of such uneven correspondence retrievals, we use the phase-shifting profilometry (PSP) technique that exploits higher spatial resolution periodic patterns to give high-resolution distortion estimates of the object. Instead of projecting binary patterns, we project continuous sinusoidal fringe patterns on the scene and recover the correspondence indices using phase unwrapping methods. The drawback of using a phase-shift pattern method is the limited depth range it offers, causing inherent ambiguities arising due to pattern repetition. As gray code scanning methods are efficient in recovering the approximate depth of objects, we combine gray code scanning with phase-shifting profilometry. The first gives us a coarse absolute correspondence map devoid of ambiguities, while the latter refines this correspondence map resolution at sub-pixel levels. A standard phase-shift projector pattern can be defined with the intensity values such as:

$$I_p(x^p, y^p) = a^p + b^p \cos(2\pi f_0^p x^p - 2\pi n/N) \tag{3.5}$$

where (x^p, y^p) is the projector pixel, a is the mean value of the pattern, b is the amplitude, n is the phase shift index, N is the total number of patterns, and f_0 is the frequency of the pattern repetition. The patterns after projecting and getting captured can be represented as:

$$I(x,y) = A(x,y) + B(x,y)\cos[\phi(x,y) - 2\pi n/N] \tag{3.6}$$

Here we have 3 unknowns: $A(x,y), B(x,y)$, and $\phi(x,y)$ and hence we need at least three phase-shifted patterns. We use the double three-phase shift algorithm by Huang et al. [27] as they reduce errors caused due to projector luminance nonlinearity. The only overhead is the requirement of 3 more patterns for both vertical and horizontal directions. The first set has phase-shifts $0, \frac{2\pi}{3}$, and $-\frac{2\pi}{3}$ and the second set has phase-shifts $\frac{\pi}{3}, \pi$, and $\frac{5\pi}{3}$. The patterns used are shown in Fig. 3.6. The intensities of the three sinusoidal fringe pattern images for the first set can be represented as:

$$I_0 = A + B\cos(\phi) \tag{3.7}$$

$$I_1 = A + B\cos(\phi - \frac{2\pi}{3}) \tag{3.8}$$

$$I_2 = A + B\cos(\phi + \frac{2\pi}{3}) \tag{3.9}$$

We can estimate them as:

$$\phi = \arctan\frac{\sqrt{3}(I_1 - I_2)}{2I_0 - I_1 - I_2} \tag{3.10}$$

$$A = \frac{I_0 + I_1 + I_2}{3} \tag{3.11}$$

$$B = \frac{1}{3}\sqrt{3(I_1 - I_2)^2 + (2I_0 - I_1 - I_2)} \tag{3.12}$$

Similarly, we can estimate the phase map for the second set using the same formula and average the phase maps to cancel the errors arising from using a single phase-shift pattern method.

3.2.3 3D mesh from normals and depths

Mesh formation using only the depth information is not accurate due to the loss of high-frequency details. To retain the high-frequency information of the objects as well as the accurate range information, we use the method proposed by Nehab et al. [52] for all our experiments to combine orientation and position estimates. This

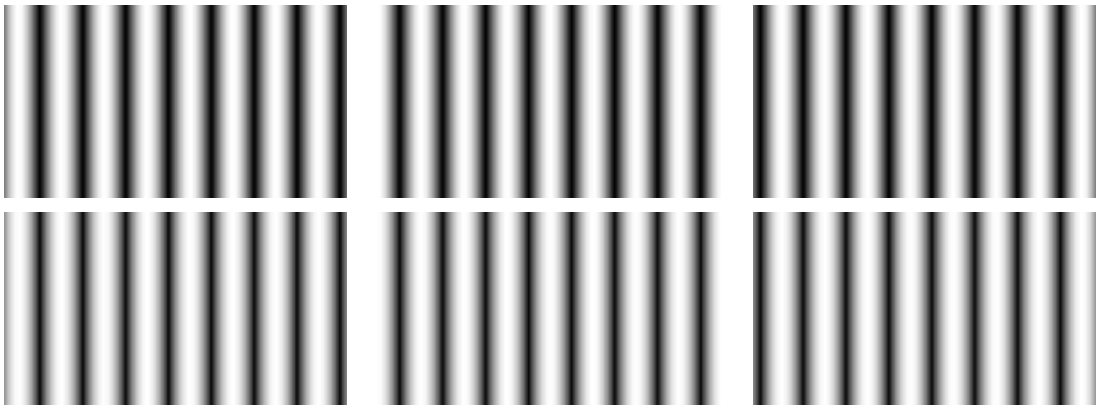


Figure 3.6: Generated double three-phase shift patterns used for phase-shift profilometry.

process improves the quality of geometric information by approximating a surface that uses both normal and depth under a two-stage linear framework.

The first step is eliminating the low-frequency bias in the measured normal maps from the photometric stereo N_m using the position map. This is done by first constructing the normal map from the position map using gradients N_p and then smoothing both normal maps by the same amount using a Gaussian filter, giving us SN_m and SN_p. A rotation field R is defined which has a one-to-one correspondence from SN_m to N_m, that captures only the high-frequency details in N_m. Finally, the corrected normal map N_c is computed as $N_c = R \cdot SN_p$. The second step involves minimizing the sum of two errors: the normal error E_n and the position error E_p. Using perspective projection, we define the position $P(x, y)$ as:

$$P(x,y) = \begin{bmatrix} -\frac{x}{f_x}Z(x,y) & -\frac{y}{f_y}Z(x,y) & Z(x,y) \end{bmatrix}^T \tag{3.13}$$

where f_x and f_y are the focal length in the x and y direction in pixels. The position

error between the measured position P_{m_i} and the optimized position P_i is defined as:

$$E_p = \sum_i ||P_i - P_{m_i}||^2 \tag{3.14}$$

$$= \mu_i^2 (Z_i - Z_{m_i})^2 \tag{3.15}$$

$$\mu_i^2 = \left(\frac{x}{f_x}\right)^2 + \left(\frac{y}{f_y}\right)^2 + 1 \tag{3.16}$$

For the normal error, we first construct the tangents T_x and T_y corresponding to x and y directions using partial derivatives as follows:

$$T_x = \frac{\partial P}{\partial x} = \begin{bmatrix} -\frac{1}{f_x}(x\frac{\partial Z}{\partial x} + Z) & -\frac{1}{f_y} y \frac{\partial Z}{\partial x} & \frac{\partial Z}{\partial x} \end{bmatrix}^T \tag{3.17}$$

$$T_y = \frac{\partial P}{\partial y} = \begin{bmatrix} -\frac{1}{f_x} x \frac{\partial Z}{\partial y} & -\frac{1}{f_y}(y\frac{\partial Z}{\partial y} + Z) & \frac{\partial Z}{\partial y} \end{bmatrix}^T \tag{3.18}$$

Using the tangents and corrected normals N_c, we define the normal error E_n which is minimized when the corrected normals are orthogonal to the corresponding tangents:

$$E_n = \sum_i |T_x(P_i) \cdot N_{c_i}|^2 + |T_y(P_i) \cdot N_{c_i}|^2 \tag{3.19}$$

The entire minimization process is formulated into a large sparse linear system that is solved by least-squares optimization process:

$$\begin{bmatrix} \lambda\mu I \\ \dots \\ (1-\lambda) N_c \cdot T \end{bmatrix} Z = \begin{bmatrix} \lambda\mu Z_m \\ \dots \\ 0 \end{bmatrix} \tag{3.20}$$

Here, λ controls the weight of the influence of normals and positions. We solve this system using the conjugated gradient method for sparse systems [23]. The depth map is retrieved from any structured light method and we use any of the above-mentioned photometric stereo methods for normal maps. Using these optimized normals and positions, we create an oriented point cloud that is used for an accurate surface mesh reconstruction.

The creation of meshes from oriented point clouds are done using screened Poisson surface reconstruction [37]. This method constructs the surface by defining a con-

tinuous 3D vector field $\vec{V} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ and finding a scalar function whose gradients matches the vector field. This scalar function is known as the indicator function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ which has a value of 1 inside the model and 0 outside the model. A weighted offset energy term is also incorporated that penalizes the objective function if its value deviates from zero at the point cloud samples. The objective function is defined as:

$$E(\chi) = \int_p ||\vec{V}(p) - \Delta f(p)||^2 dp + \alpha \frac{\text{Area}(P)}{\sum_{p \in P} w(p)} \sum_{p \in P} w(p) f^2(p) \tag{3.21}$$

Here, P is the set of oriented points, w describes the weight of every point, and α is the parameter that expresses the importance of the additional screened term. The weight $w(p)$ is set to be 1 for every point. The above equation is solved using a discretized multi-grid octree. The weighting parameter for the two energy terms is adaptively adjusted across different octree depths. We use the Dirichlet boundary conditions for all the reconstructions which has a value of $-\frac{1}{2}$ along the boundaries of the integration domain.

3.3 Robust image decompostion for specular sparse noise elimination

One of the adversities of manufacturing objects in a CNC machine is the presence of random metal shavings which corrupt the captured images in two ways: 1) blocking light rays from the light source creating shadows and 2) introducing specular non-uniform spots captured by the camera. Generally, these artifacts tend to be sparse and are present in both photometric stereo and structured light scanning.

It is a well-studied concept that Lambertian objects, under a fixed viewpoint, when illuminated from at least 3 different directions, can be used to determine the surface orientation [8]. We use the approach described in Wu et al. [75] that uses this 3-rank constraint to cast the photometric stereo problem into a low-rank matrix decomposition problem. This decomposition separates the diffuse component from the noisy corrupted image measurements. We represent the corrupted measurements in photometric stereo as a combination of the low-rank approximation matrix from

the $\mathbf{N} \cdot \mathbf{L}$ Lambertian model and the sparse noise matrix as follows:

$$\mathbf{I} = \rho \, (\mathbf{N} \cdot \mathbf{L}) + \mathbf{E} \tag{3.22}$$

$$\mathbf{I} = \mathbf{M} + \mathbf{E} \tag{3.23}$$

Enforcing the low-rank constraints, the optimization problem becomes:

$$\underset{\mathbf{M},\mathbf{E}}{\text{argmin}} \quad \text{rank}(\mathbf{M}) + \lambda ||\mathbf{E}||_0 \tag{3.24}$$

Here, $||\mathbf{E}_0||$ is the ℓ_0-norm denoting the number of non-zero entries in the matrix. This optimization is an NP-hard problem and using a convex relaxation method known as principal component pursuit (PCP), this problem has the same optimal solution as:

$$\underset{\mathbf{M},\mathbf{E}}{\text{argmin}} \quad (||\mathbf{L}||_* + \lambda ||\mathbf{E}||_1) \tag{3.25}$$

Here, $||\mathbf{L}||_*$ is the nuclear norm and $||\mathbf{E}||_1$ represents the ℓ_1-norm.

We extend the same idea to structured light gray code scanning, where for a single binary pattern, we capture multiple images of the object. The set of single-patterned images are decomposed into low-rank and corrupted measurements where we expected the rank of the low-rank matrix to be 1. We solve both these optimization problems using Augmented Lagrangian Multiplier by Lin et al. [42].

3.4 Simulation experiments

The purpose of simulated experiments is to gauge the performance of our algorithms in an ideal noise-free situation. We perform experimental evaluations on a physically accurate simulation platform called Mitsuba [32] using the pipeline shown in Fig. 3.7. The scene usually consists of an environment map describing the surroundings of the object, a camera, a controllable light projector, light sources, and the object's geometry. To render scenes realistically, we use the library of BRDFs provided

by Mitsuba that is based on the microfacet model by Cook and Torrance [16]. In the microfacet model theory, the surface of an object is represented as microscopic surfaces that behave like Fresnel mirrors when light rays are cast. The two main components of the microfacet model are the distribution of the facets and a function describing the masking/shadowing properties of the facets. Mitsuba provides multiple tune-able parameters like material type, material properties, and roughness extent in the tangent-bitangent directions. This enables easy reproduction of objects with versatile surface appearances. For all our simulation experiments, we use the wrench object that has 14,799 vertices and 29,312 faces. The centre of the object is always kept 10 cm away from the camera centre in the z-direction. A mesh skeleton and a rendering of the wrench in an environment is shown in Fig. 3.8

For simulation experiments, we use structured light line scanning as a baseline as it guarantees more accuracy than multiplexed patterns like gray codes or PSP. The resolution of the projector is 1024×768 and therefore, we project 768 spatially-varying line patterns on the scene. We use a rectified camera-projector system and generate a disparity map by finding correspondences on every row of the captured image. The depth map is then generated from the disparity map by using the formula $\mathbf{D} = \frac{bf}{\mathbf{d}}$, where $\mathbf{d}$ is the disparity map, b is the distance between the camera and the projector centre in the x-dimension, also known as baseline, and f is the focal length of the camera.

We show the qualitative and comparative quantitative results for estimating normal maps and depth maps, followed by the optimization process in Nehab et al. [52], and finally creating a 3D surface mesh. In the figures, rc-α-β means a parametric rough conductor BRDF (material type) and $\alpha, \beta \in [0, 1]$ specifies the anisotropic roughness parameters along the tangent and bi-tangent directions. If $\alpha = \beta$, the object is isotropic, else it is anisotropic.

The surface orientation maps using the two photometric stereo methods for objects with different BRDFs are shown in Fig. 3.9. Normals in all result figures are visualized as $(\mathbf{N} + 1)/2$ to prevent the clipping of negative values. From the normal map estimation results, it is clear that photometric stereo performs the best when the object is diffuse and gives erroneous results in cases of specular isotropic and anisotropic BRDFs. Also, the method mentioned in 3.1.2 gives us more accurate normal map estimation as shown in the middle row of 3.9. The normals estimated

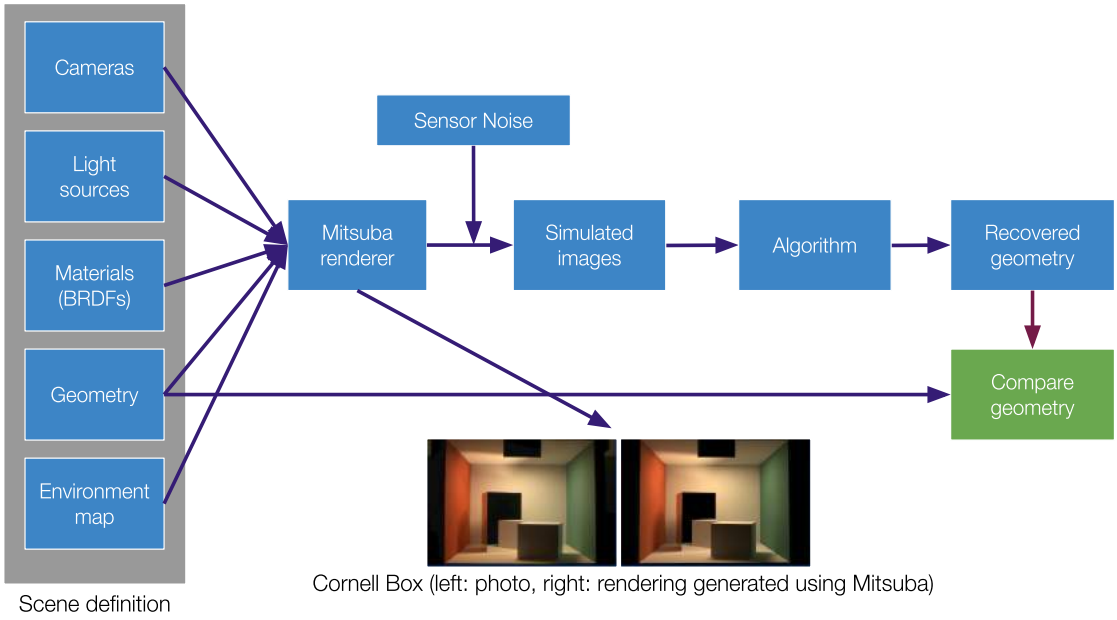


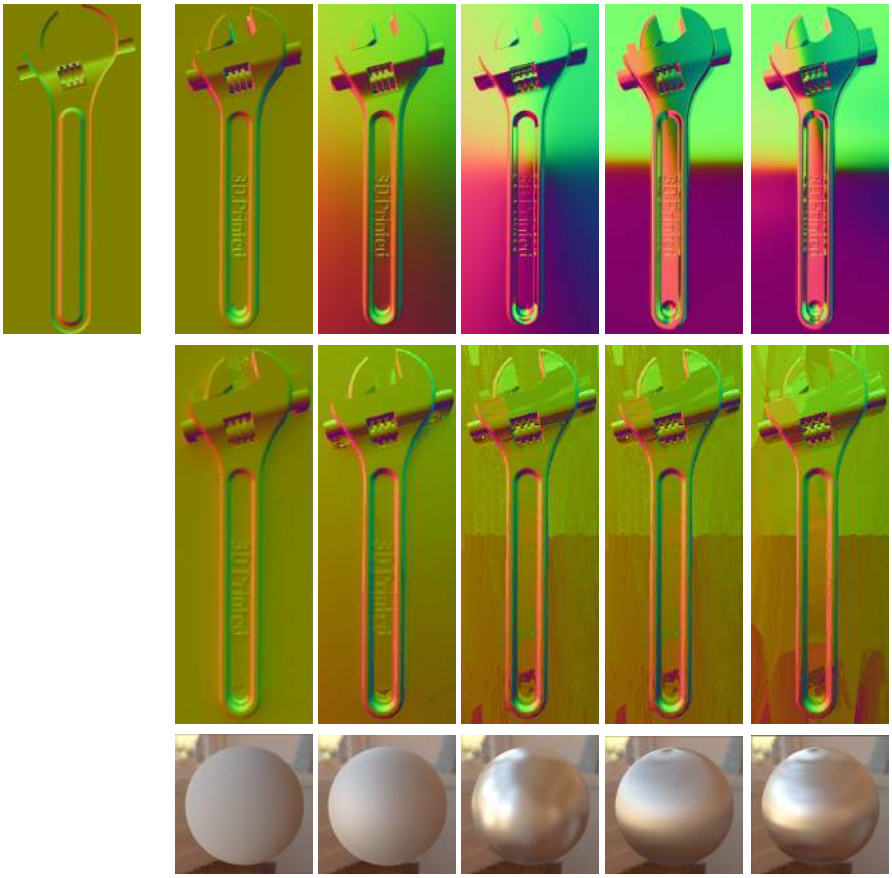
Figure 3.7: Rendering pipeline used for simulation experiments.

Figure 3.8: Left: Mesh skeleton of wrench object. Right: Rendering of the wrench mesh using Mitsuba.

using this method is more uniform, compared to photometric stereo, even in cases of complex BRDFs.

BRDF	**Diffuse**	**rc-0.5-0.5**	**rc-0.1-0.1**	**rc-0.5-0.1**	**rc-0.2-0.05**
Error	0.718	0.723	0.781	0.81	0.815

Table 3.1: Mean depth map reconstruction error (in mm) versus BRDF.



Ground Truth Diffuse rc-0.5-0.5 rc-0.1-0.1 rc-0.5-0.1 rc-0.2-0.05

Figure 3.9: Simulation experimental results for estimating normal maps. Top Row: Normal maps from photometric stereo. Middle row: Improving normal maps using the method discussed in Section 3.1.2. Bottom Row: Reference rendering of a sphere for the corresponding BRDF.

Ground Truth Diffuse rc-0.5-0.5 rc-0.1-0.1 rc-0.5-0.1 rc-0.2-0.05

Figure 3.10: Simulation experimental results for estimating depth maps using structured light line scanning. Bottom Row: Reference rendering of a sphere for the corresponding BRDF.

The reconstructed mesh is compared with the reference mesh using Hausdorff measure, which is defined as the average distance between two meshes obtained by sampling points on the reconstructed mesh and computing the closest distance from the reference mesh.

We use the depth and normal maps from the previous results to reconstruct a 3D surface mesh. Specifically, we compare three different types of algorithms for mesh error comparisons:

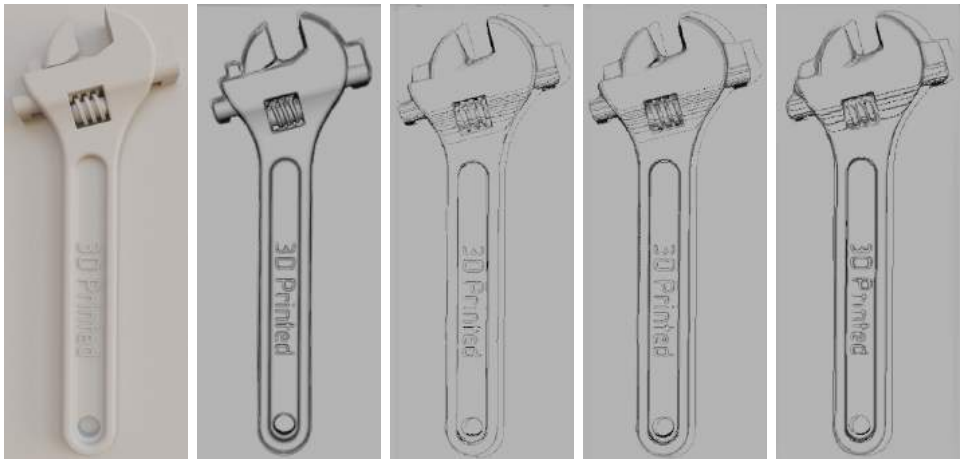
Algo 1: Depth from structured light scanning + normals from fitting planes to neighbourhood points

Algo 2: Depth from structured light scanning + normals from photometric stereo

Algo 3: Depth from structured light scanning + normals from the method discussed in Section 3.1.2

The first algorithm estimates per-point normal for each point in the point cloud by fitting planes using 10 neighbour points. The second algorithm uses the normal map from photometric stereo while the final algorithm uses the generalized photometric stereo method. The dimension of the object is 10 cm in height, 4 cm in width, and

1 cm in depth. We provide the qualitative results for the wrench object for all the 5 BRDFs showing the mesh reconstruction results in Figs. [3.11, 3.12, 3.13, 3.14, 3.15]. A summary of the error metrics is shown in Tab. 3.2 for all the 3 listed algorithmic processes. We conclude from these simulation experiments that as the object reflectance moves from Lambertian to specular isotropic and anisotropic, the reconstruction error for all the algorithms increase.

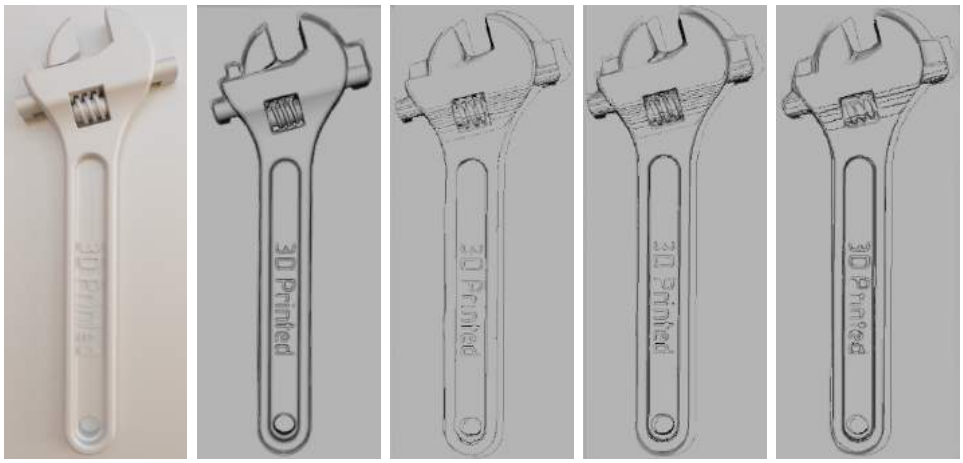


(a) Rendering (b) Expected (c) Algo 1 (d) Algo 2 (e) Algo 3

Figure 3.11: 3D mesh reconstruction comparison for diffuse BRDF.

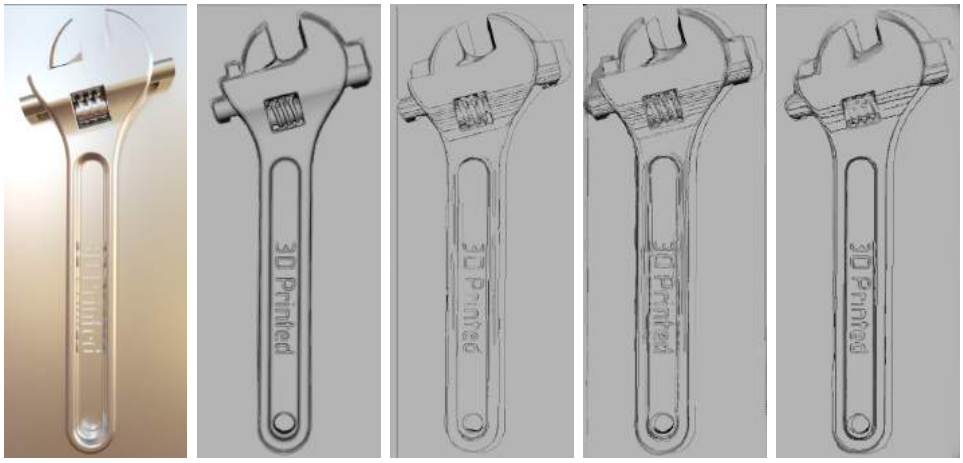
Method \ **BRDF**	$Diffuse$	$rc-0.5-0.5$	$rc-0.1-0.1$	$rc-0.5-0.1$	$rc-0.2-0.05$
Depth + Plane-fit normals	0.31	0.315	0.32	0.32	0.315
Depth + Photometric stereo normals	0.286	0.288	0.293	0.29	0.295
Depth + Holroyd et al. [25] normals	0.219	0.22	0.223	0.221	0.232

Table 3.2: Simulation results: 3D mesh geometry error in mm versus BRDF for our comparative algorithms.



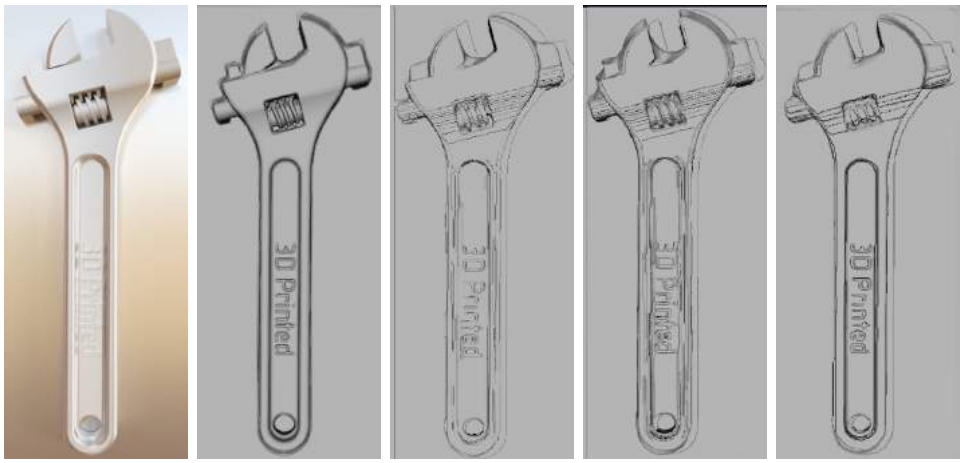
(a) Rendering (b) Expected (c) Algo 1 (d) Algo 2 (e) Algo 3

Figure 3.12: 3D mesh reconstruction comparison for isotropic specular BRDF with roughness parameter = 0.5 in both directions.



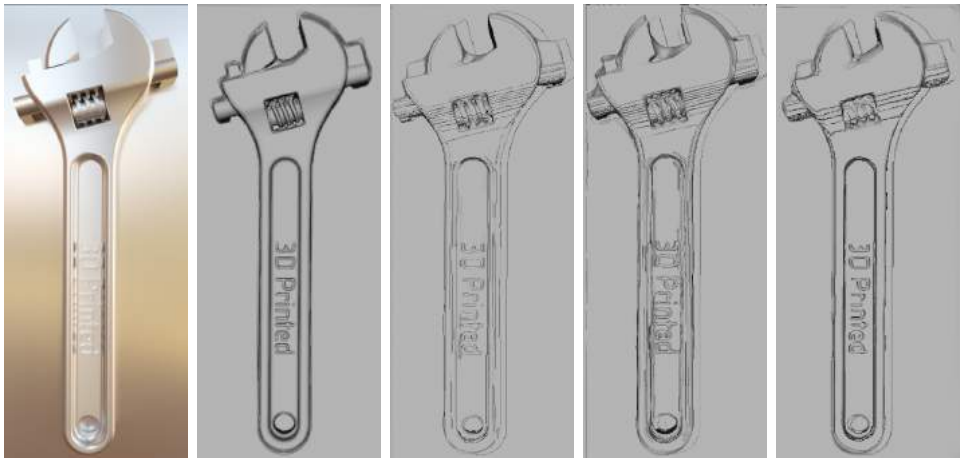
(a) Rendering (b) Expected (c) Algo 1 (d) Algo 2 (e) Algo 3

Figure 3.13: 3D mesh reconstruction comparison for isotropic specular BRDF with roughness parameter = 0.1 in both directions.



(a) Rendering (b) Expected (c) Algo 1 (d) Algo 2 (e) Algo 3

Figure 3.14: 3D mesh reconstruction comparison for anisotropic BRDF with roughness parameters = 0.5 and 0.1 in tangent and bi-tangent directions respectively.



(a) Rendering (b) Expected (c) Algo 1 (d) Algo 2 (e) Algo 3

Figure 3.15: 3D mesh reconstruction comparison for anisotropic BRDF with roughness parameters = 0.2 and 0.05 in tangent and bi-tangent directions respectively.

3.5 Experiments

In this section, we first show experiments using our practical setup for photometric stereo and structured light scanning. We reconstruct the 3D shape of multiple objects and perform quantitative analysis for our reconstructed 3D geometries. We then deal with specular noisy environments using the low-rank matrix decomposition method (Sec. 3.3) for robust 3D reconstructions. Finally, we perform experiments using our setup on a representative CNC lathe.

3.5.1 3D geometry reconstruction

Hardware setup: Our hardware setup consists of the following: Allied Vision Prosilica GX-1910 camera, Edmund Optics 8.5 mm/F1.3 lens, DLP LightCrafter 4500 projector by Texas Instruments, Lowel ViP Pro-Light, and a chrome sphere ball for photometric stereo light position calibration. For structured light scanning, we use a calibrated projector-camera pair. The chrome sphere ball and the collimated light source are required especially for photometric stereo. Our setup is shown in Fig. 3.16. We use three objects: a planar white box, a pink spherical ball, and a white horse. The objects in our experiments are always placed in front of a black surface for simple background-foreground segmentation.

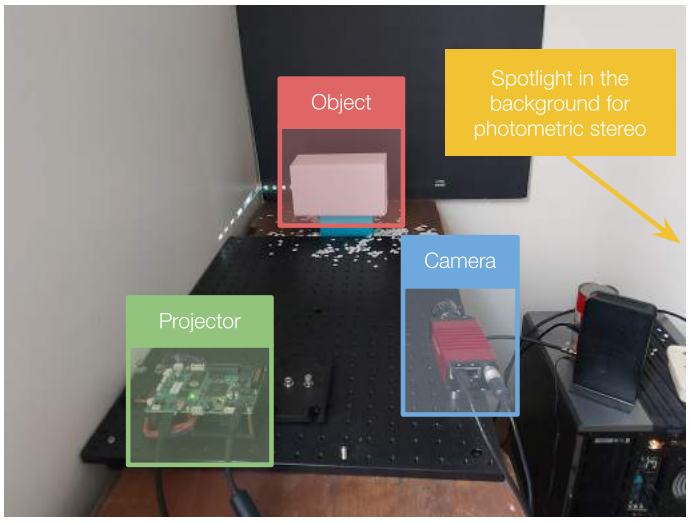


Figure 3.16: Hardware setup used for photometric stereo and structured light scanning.

System calibration: We describe different types of calibration required for photometric stereo, gray code scanning, and phase-shifting patterns. These calibration processes are essential for increasing the accuracy of a scanning system.

1. *Light direction calibration for photometric stereo*

   For photometric stereo, we calibrate the light positions using the toolbox provided by Xiong [76]. A sphere chrome ball is placed in the scene and the points of the boundary of the sphere are manually selected. Then, a circle is fitted to the manually extracted positions (Fig. 3.17). For every lightning position used for capturing the image, we estimate the 3D direction from the specular highlight present in the 2D circle image of the chrome sphere as shown in Fig. 3.18.

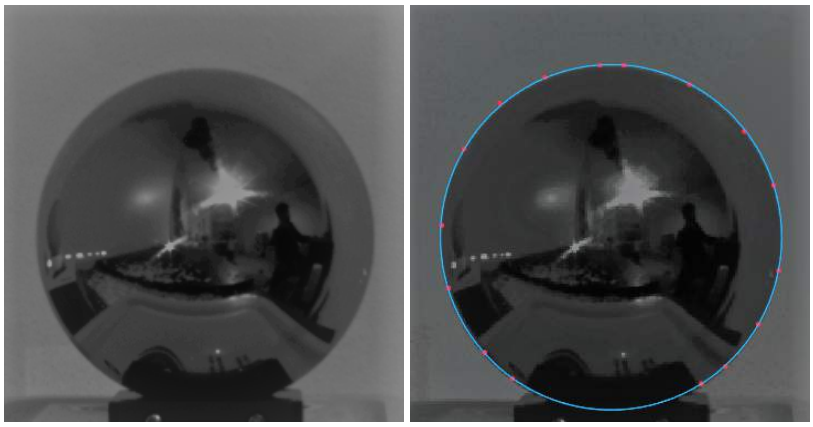


Figure 3.17: Left: Image of the chrome sphere in the scene. Right: Fitted circle on the sphere using manually selected points.

2. *Projector-camera calibration for structured light scanning*

   High-precision in gray code structured light systems can only be obtained through accurate calibration of the camera and the projector. We calibrate our projector-camera system using the method described by Moreno and Taubin [48] that models the projector as an inverse camera using the general pinhole camera model. They take into account the radial and the tangential distortions of both the camera and the projector. A plane checkerboard pattern is used as a calibration object for computing the local homographies as corners are easy to identify. The homography is a 3×3 matrix that relates the camera's

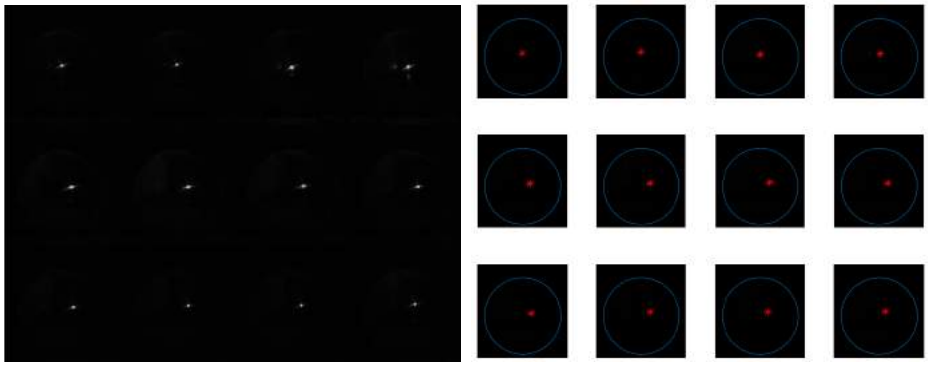


Figure 3.18: Left: Specular highlights in the chrome sphere for various lighting positions. Right: Estimated lighting positions in red.

reference frame with the projector's reference frame.

The camera-projector calibration steps are mentioned in Algorithm 1.

Algorithm 1 Camera projector stereo calibration.

1: **procedure** CALIBRATECAMERAPROJECTOR(img)
2: Generate gray code patterns for the projector resolution.
3: Capture multiple sequences S of the gray code pattern with different positions of the checkboard pattern (sample captured checkerboard images in Fig. 3.19)
4: Calibrate the camera first using the all-on images i of the sequences S
5: Compute the correspondences of the known points in the calibration object using all the gray code sequences S. An example correspondence map for the x and y co-ordinates are shown in Fig. 3.20
6: Calibrate the projector using the same process to calibrate the camera
7: **end procedure**

3. *Projector radiometric calibration for phase shifting*

Phase shifting profilometry (PSP) are very commonly used methods to get sub-pixel resolution accuracy for the correspondence maps between a camera projector system which results in smoother point clouds and surface meshes. The performance of PSP systems are often impaired due to the nonlinear luminance, commonly known as gamma correction. To increase the accuracy of the reconstruction we calibrate the projector compensating for the intensity nonlinearity using the method described in Hoang et al. [24]. The process involves generating two sets of PSP patterns with arbitrary gamma encoding

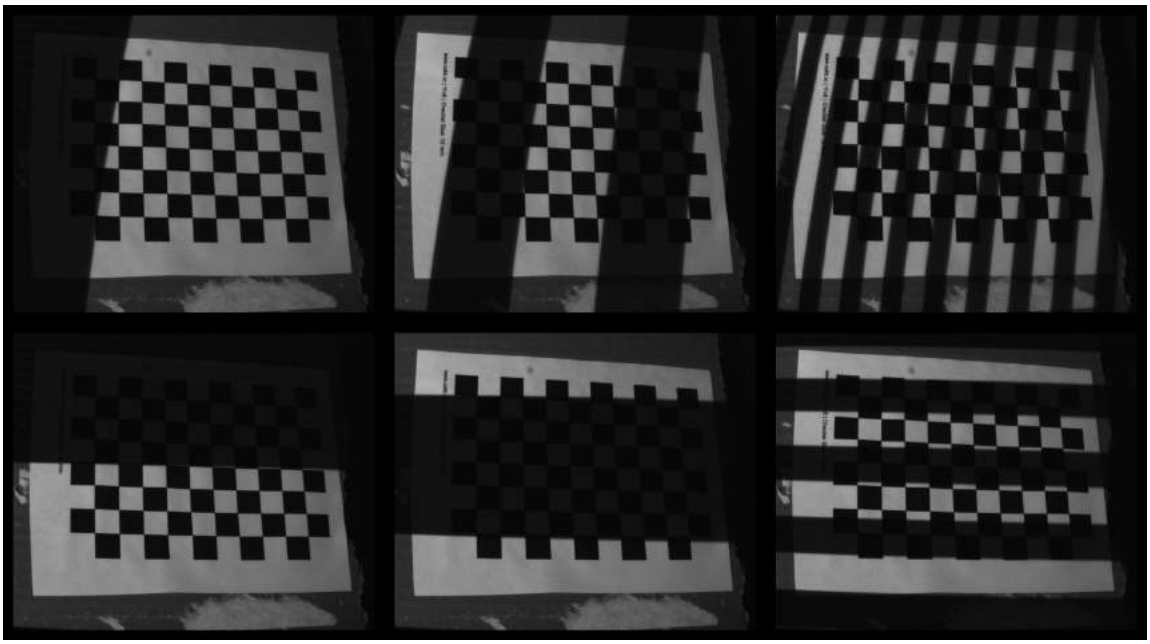
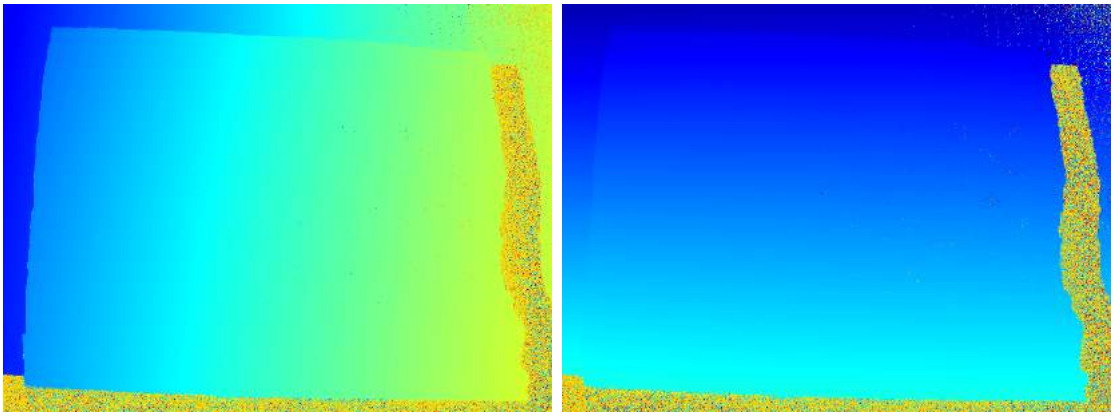


Figure 3.19: Captured gray code patterns projected on the checkerboard pattern.



(a) Horizontal correspondences (b) Vertical correspondences

Figure 3.20: Camera projector dense point correspondences for checkerboard gray code sequence.

values. The patterns are projected on a planar surface and images are captured. The total gamma γ_t of the scanning system can be described with the actual projector gamma γ_p along with two other gamma parameters γ_a, γ_b arising due to the complexity of the projector-camera system. Estimating interme-

diate gamma values γ_a, γ_b through the phase-shift scanning and then using $\gamma_p = (1-\gamma_b)\gamma_a$ gives the gamma correction value of the projector-camera system.

Normal map estimation: The captured images for estimating the normal map using photometric stereo along with the estimated normal maps for three objects (box, ball, and horse) are presented in Figs. [3.21, 3.22, 3.23] respectively.

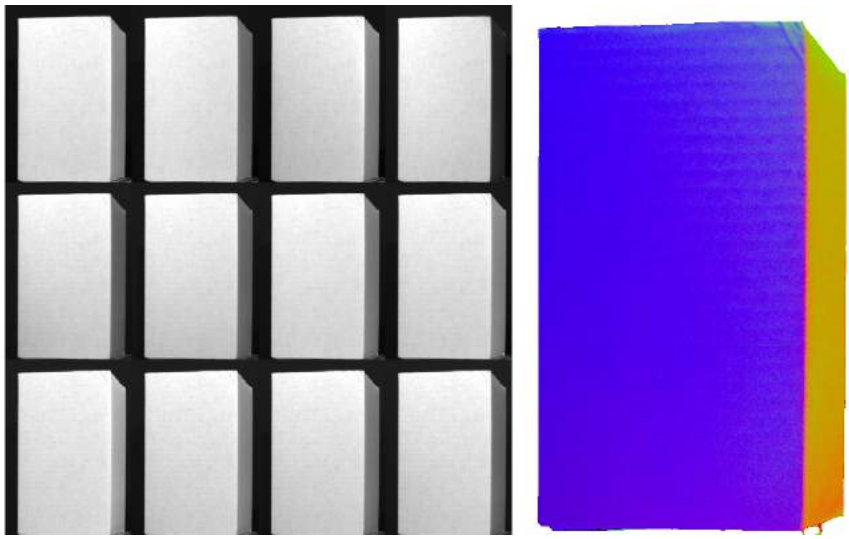


Figure 3.21: Left: Captured images used for photometric stereo. Right: Recovered per-pixel normal map for planar box.

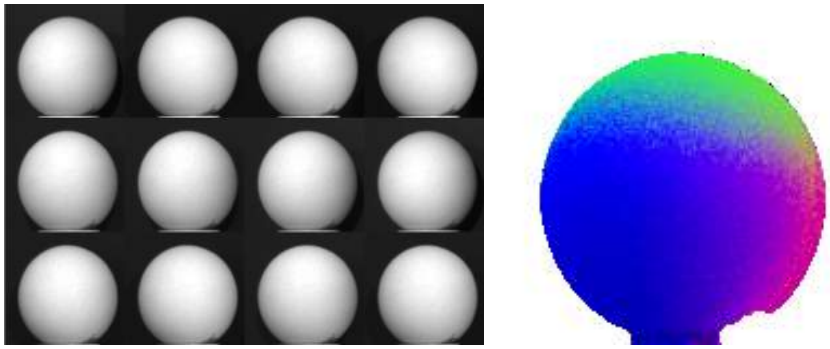


Figure 3.22: Left: Captured images used for photometric stereo. Right: Recovered per-pixel normal map for spherical ball.

Gray code structured light: We project 44 gray code patterns into the scene, capture, and then decode them to get correspondence maps. These correspondence

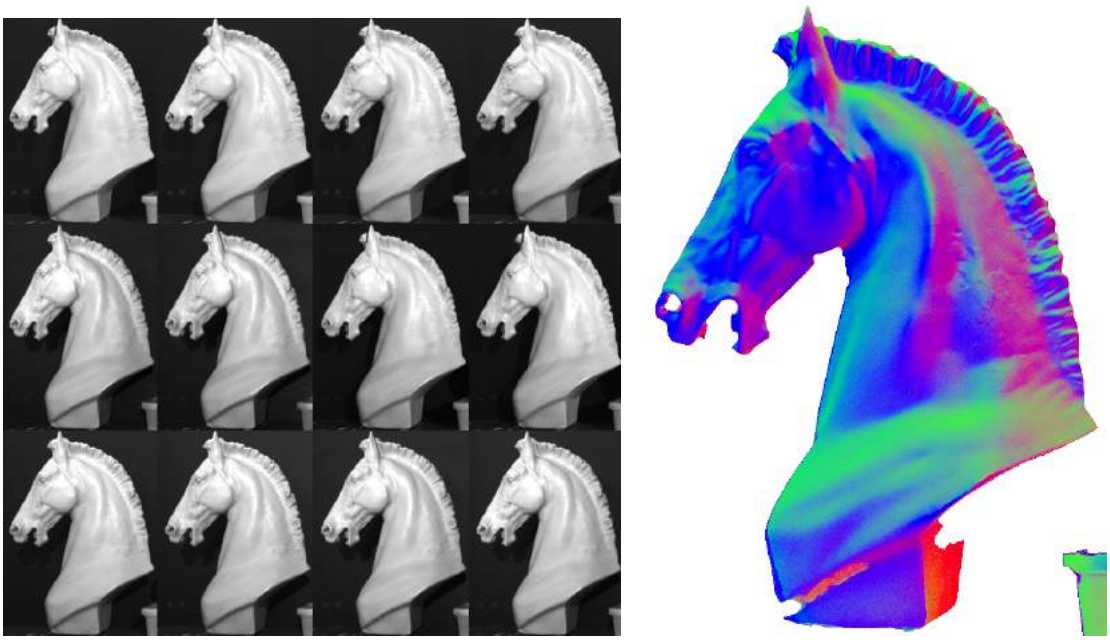
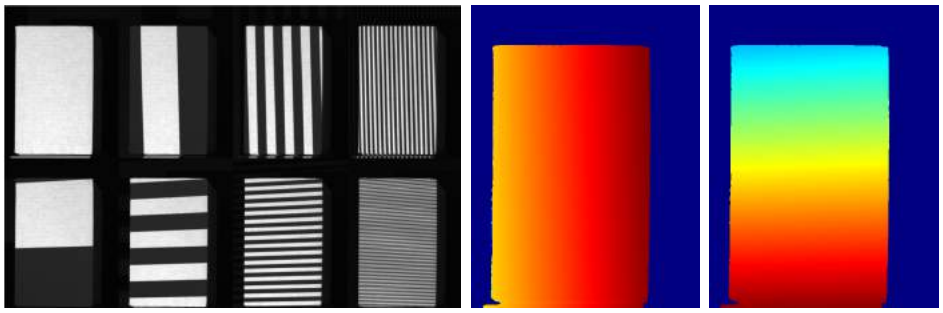


Figure 3.23: Left: Captured images used for photometric stereo. Right: Recovered per-pixel normal map for horse.

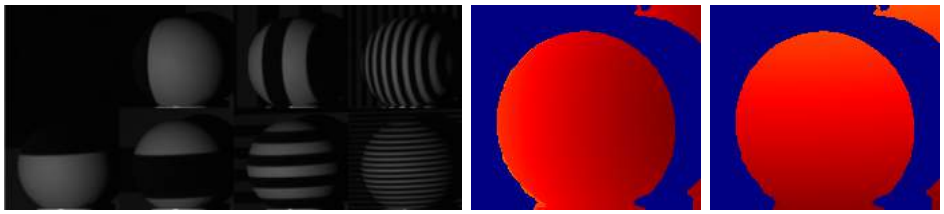
maps relate the projector pixels with the camera. The gray code patterns projected on all the objects along with their correspondence maps visualized using pseudo-color are shown in Figs. [3.24, 3.25, 3.26].



(a) Sample captured images (b) Horizontal (c) Vertical

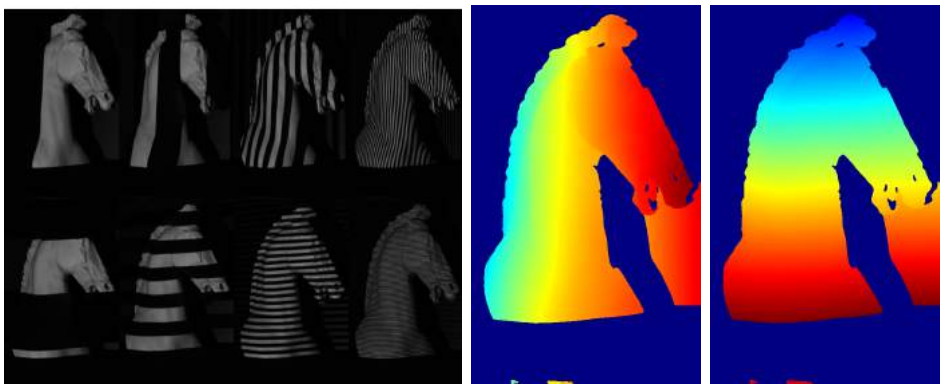
Figure 3.24: Captured images and correspondences used for gray code scanning for planar box.

Gray code + Phase-shifting patterns: For the hybrid gray code-PSP pattern



(a) Sample captured images (b) Horizontal (c) Vertical

Figure 3.25: Captured images and correspondences used for gray code scanning for spherical ball.

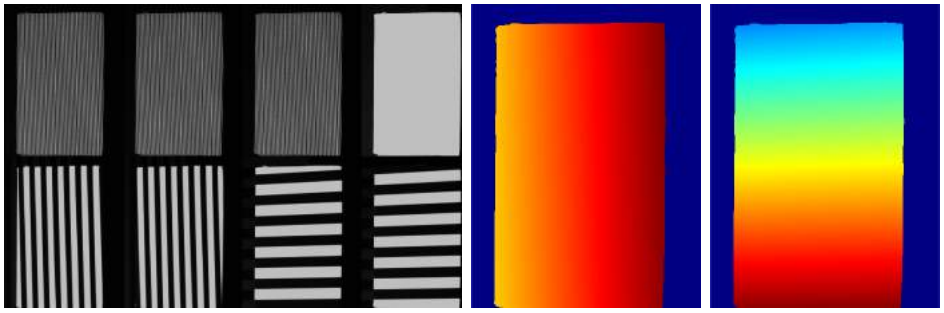


(a) Sample captured images (b) Horizontal (c) Vertical

Figure 3.26: Captured images and correspondences used for gray code scanning for horse.

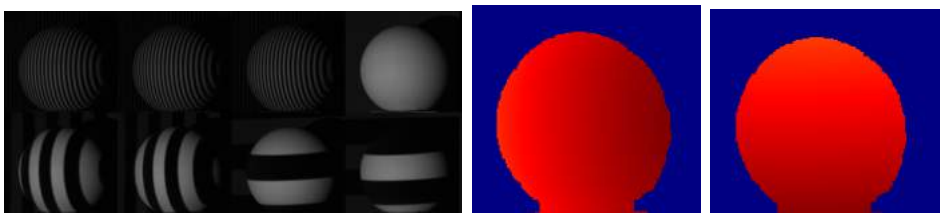
scanning process, we first project the 44 gray codes to compute the pixel-wise indices and then use 2 different frequencies of sinusoidal patterns with 3 phase shifts. The sinusoidal patterns are projected for horizontal as well as vertical directions, to refine these correspondence indices, thus making a total of 54 images. The hybrid gray code-PSP patterns projected on all the objects along with their correspondence maps are shown in Figs. [3.27, 3.28, 3.29].

3D shape from normals and depths: The estimated correspondences between the camera and projector matrix from gray code and the hybrid scanning process are used to construct the point cloud of the object, using the intrinsic and extrinsic camera projector matrices. Normals maps are used from the photometric stereo process. Finally, we input the point cloud and the per-pixel normal map (oriented



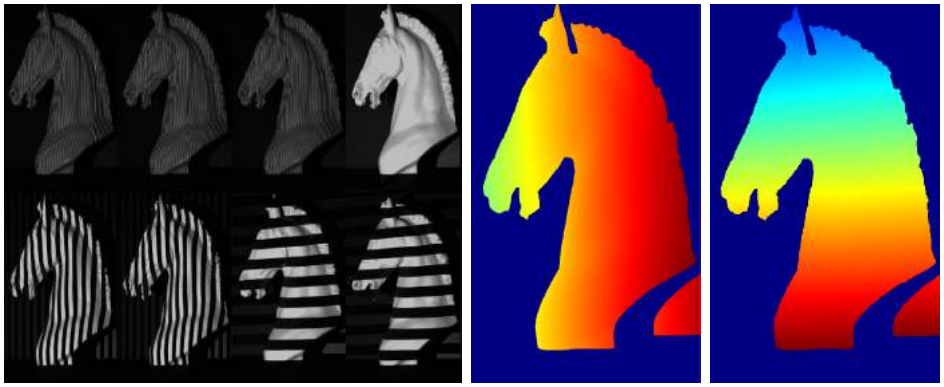
(a) Sample captured images (b) Horizontal (c) Vertical

Figure 3.27: Captured images and correspondences used for gray code + PSP patterns for planar box.



(a) Sample captured images (b) Horizontal (c) Vertical

Figure 3.28: Captured images and correspondences used for gray code + PSP patterns for spherical ball.



(a) Sample captured images (b) Horizontal (c) Vertical

Figure 3.29: Captured images and correspondences used for gray code + PSP patterns for horse.

point cloud) for the surface reconstruction process for multiple objects.

We observe that the point clouds generated using the gray code-PSP patterns generate more accurate and denser point clouds compared to gray codes scanning. This is due to the discretization of the correspondence maps by the encoding-decoding process in gray code pattern. Also, the 3D surface mesh reconstructions in case of using the hybrid patterns are more smooth and devoid of repetitive artifacts. Comparative point cloud and mesh reconstruction figures for all three objects are shown in Figs. [3.30, 3.31].

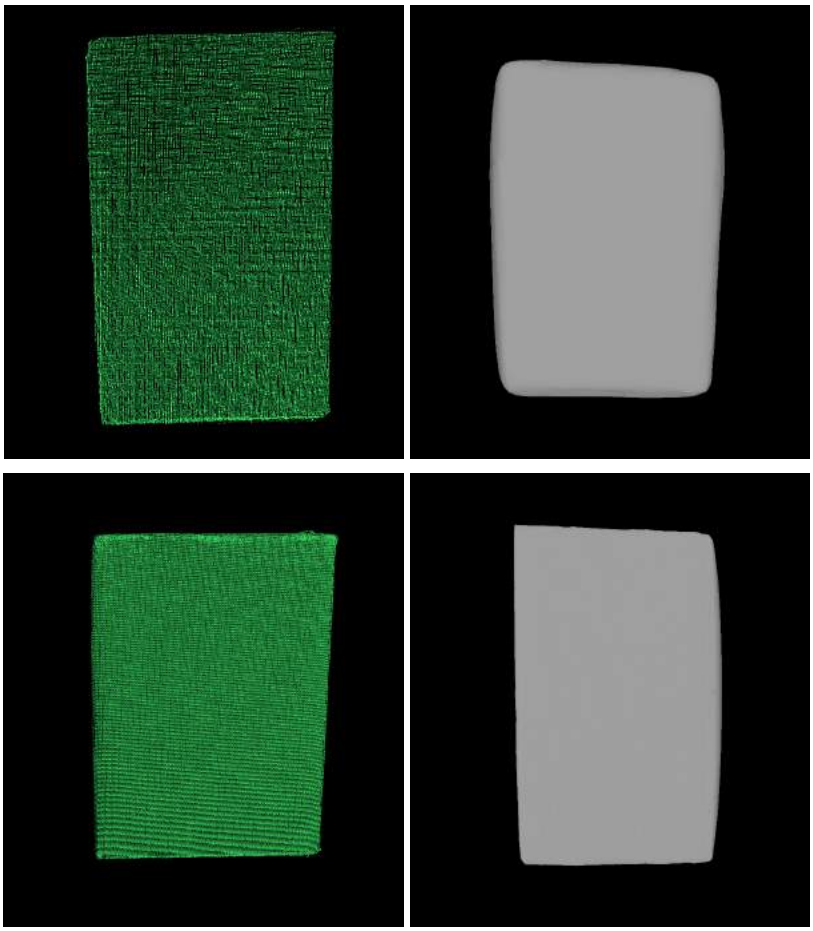


Figure 3.30: Top: Recovered point cloud and surface mesh using gray codes. Bottom: Recovered point cloud and surface mesh using gray codes + PSP for planar box.

Quantitative evaluations: We use the planar box and spherical ball for our

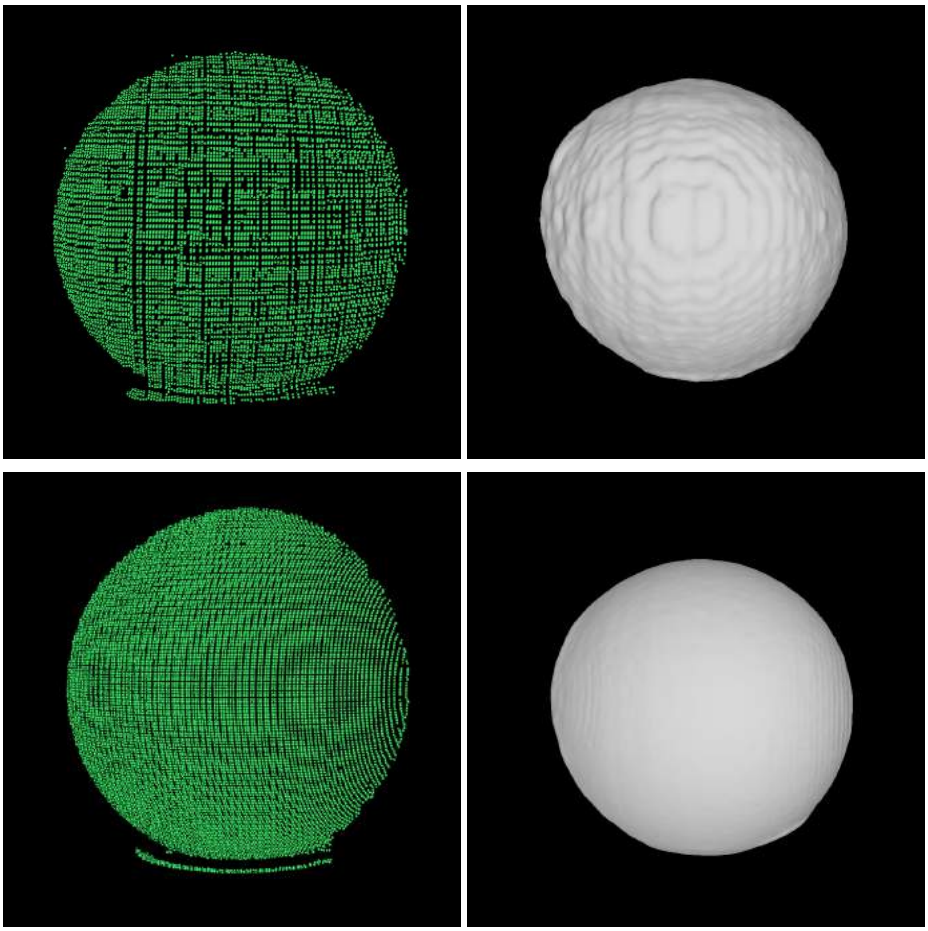


Figure 3.31: Top: Recovered point cloud and surface mesh using gray codes. Bottom: Recovered point cloud and surface mesh using gray codes + PSP for spherical ball.

reconstruction error evaluation. For the box, we fit a parametric plane to the recovered point cloud. The average error estimate is the mean distance of samples in the point cloud with the planar model. For the ball, we measure the diameter first and then create a parametric sphere for ground truth. We calculate the distance of every point from this expected spherical model. The average error values in mm are presented in Table 3.3. From the table, we infer that using a combination of gray codes and PSP patterns offer an improvement over using just gray code patterns. In addition to improving the visual quality of the mesh, it is also capable of estimating surface geometries more accurately.

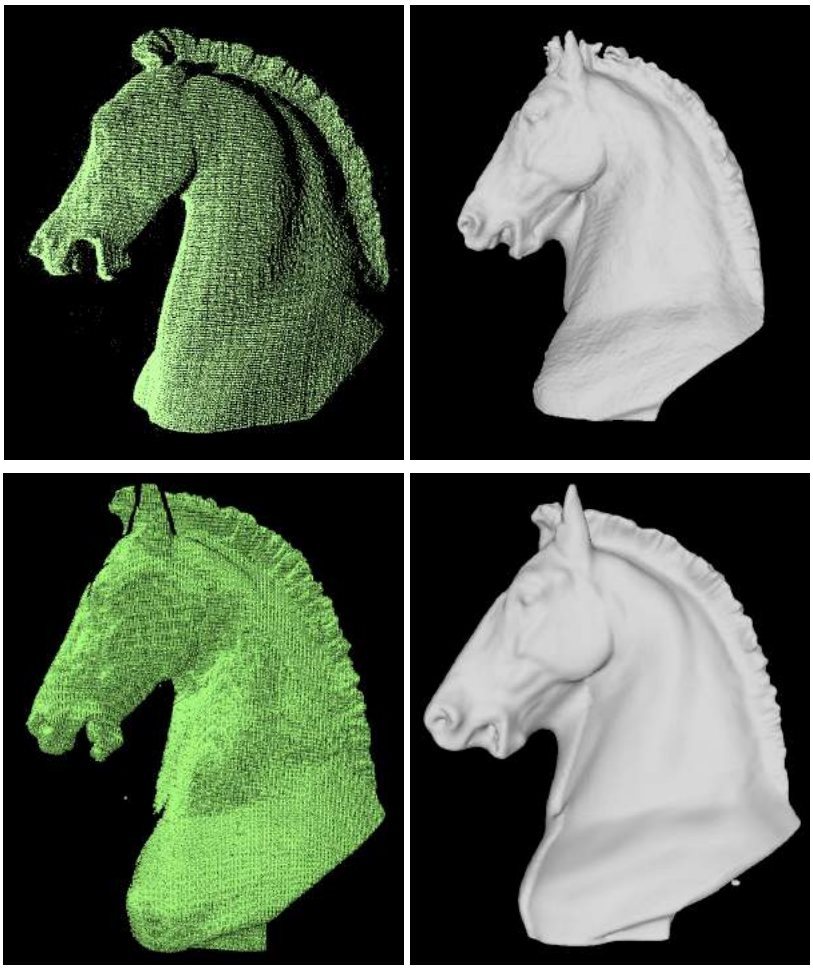


Figure 3.32: Top: Recovered point cloud and surface mesh using gray codes. Bottom: Recovered point cloud and surface mesh using gray codes + PSP for horse.

Table 3.3: Comparative error analysis using our practical setup for 3D range scanning.

Object	**Method**	**Error in mm**
Box	Photometric stereo + gray code scanning	0.513
	Photometric stereo + PSP + gray code scanning	0.320
Ball	Photometric stereo + gray code scanning	0.421
	Photometric stereo + PSP + gray code scanning	0.260

3.5.2 Robust 3D reconstruction in noisy environment

Hardware setup modifications and scanning environment: Our previous 3D scanning system consisted of a collimated spotlight required for photometric stereo. This light required manual movement for obtaining different illumination directions. For convenience, we replace the light source with a monitor display. This enables automatic light illumination direction manipulation by selective pixel multiplexing of the display. The light direction calibration process for photometric stereo is the one as described in Section 3.5.1. Our modified setup is shown in Fig. 3.33. For creating a sparse specular noisy scanning environment, we scatter mirror-like confetti particles around the viewing area of the camera and projector. These particles emulate sparse shadows and shiny spots on and around the object like in a real machining environment. The effect of confetti on our image captures is shown in Fig. 3.34.

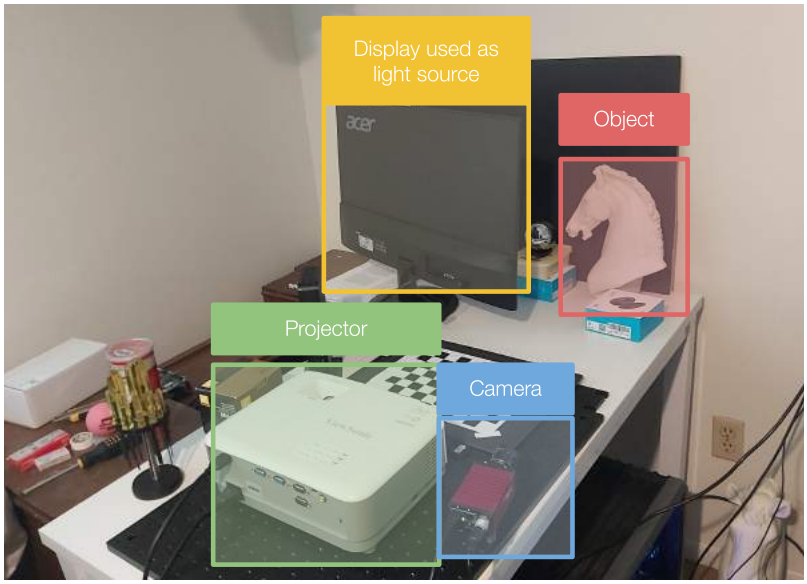
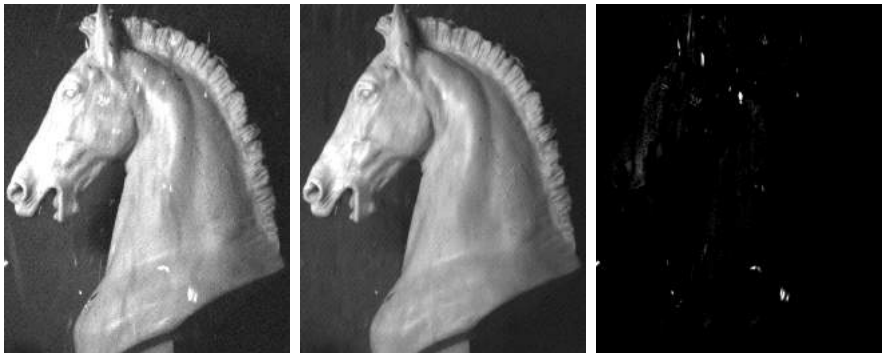


Figure 3.33: Our modified setup used for specular sparse noise elimination experiments.

Normal map improvement: We use the diffuse horse object mentioned above for all our noise elimination experiments. We capture 11 images for photometric stereo. The decomposed diffuse and noisy components from the corrupted measurements are presented in Fig. 3.35. The normal map estimation comparison using the noisy

Figure 3.34: Left: Image frame showing the effect of confetti while scanning with our setup. Right: Setup image taken after scanning the object in the presence of specular noise-generating confetti particles.

corrupted and the low-rank recovered set of images are shown in Fig. 3.36. We deduce that the algorithm gives us better estimates of the surface orientation on using the low-rank decomposed diffuse images.



(a) I (corrupted) (b) L (diffuse) (c) E (sparse)

Figure 3.35: Low-rank matrix completion and recovery for photometric stereo.

Depth map improvement: We use gray code scanning for our robust depth estimation experiments as sinusoidal phase-shift patterns are intensity dependent. This makes them sensitive to surface reflectance and more vulnerable to correspondence

(a) Expected normals (b) Normals using corrupted images (c) Normals using low-rank images

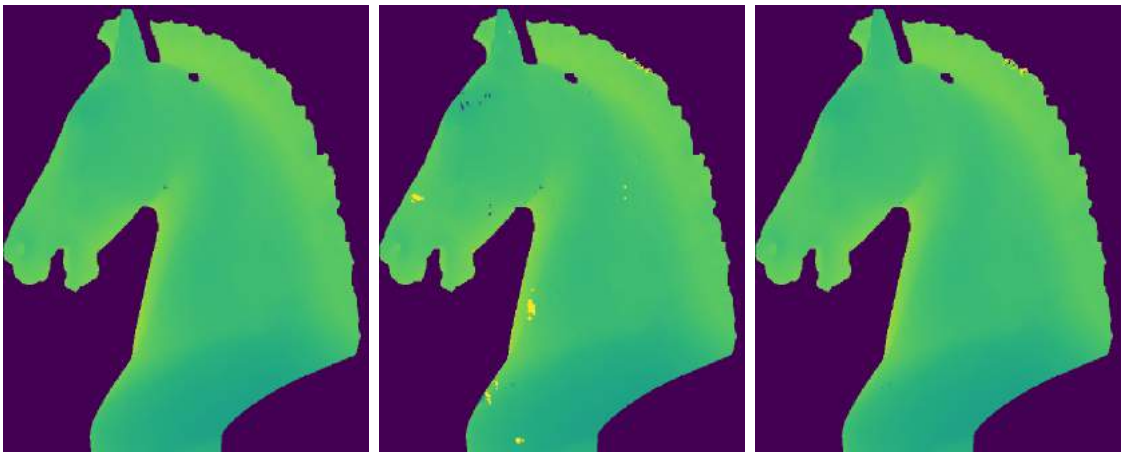
Figure 3.36: Photometric stereo normal improvement using corrupted and low-rank recovered non-corrupted images.

errors. We capture 6 images for every single gray code pattern. The normal gray code scanning requires 44 images which takes a capture time of 0.7 seconds while the low-rank matrix decomposition method requires 264 images with a capture time of 4.2 seconds. The sparse noise removal process for gray code scanning is shown in Fig. 3.37.

(a) Multiple captures of a single gray code pattern (b) Recovered non-corrupted image

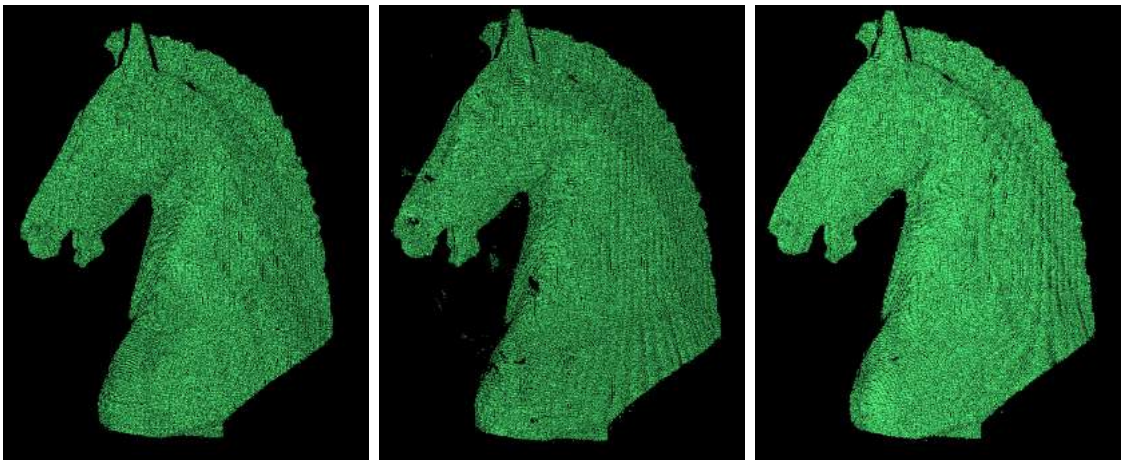
Figure 3.37: Low-rank matrix completion and recovery for gray code structured light scanning.

We observe improvement for the depth map estimation using low-rank matrix decomposition in the case of gray code scanning as shown in Fig. 3.38. We reconstruct the 3D point clouds using triangulation for the corrupted and non-corrupted cases



(a) Expected depth map (b) Depth map using corrupted images (c) Depth map using low-rank images

Figure 3.38: Depth map improvement using corrupted and low-rank recovered non-corrupted images.



(a) Expected point cloud (b) Using corrupted images (c) Using low-rank images

Figure 3.39: Point cloud reconstruction improvement using corrupted and low-rank recovered non-corrupted images.

as shown in Fig. 3.39. The corrupted point cloud has holes and more outlier point locations due to shadows and complex reflectivity generated from the specular particles. The low-rank recovered point cloud closely conforms with the expected ground truth point cloud.

Quantitative evaluations: The mesh errors are computed by comparing our reconstructed mesh with a ground truth mesh in the absence of specular confettis. We compare the errors relative to a reference mesh as finding the real mesh shape required precise equipment which was infeasible for our situation.

The normals error in degrees and depth error in mm is shown in Table 3.4. Using the normal and depth maps obtained for both scenarios (with and without low-rank completion captures), we reconstruct the surface mesh for 4 different scenarios:

1. Photometric stereo images and gray code scanning images (Noisy)
2. Low-rank completion of photometric stero images and gray code scanning images (Stereo LR only)
3. Photometric stereo images and low-rank completion of gray code scanning images (Gray code LR only)
4. Low-rank completion of photometric stereo images and low-rank completion of gray code scanning images (Both LR)

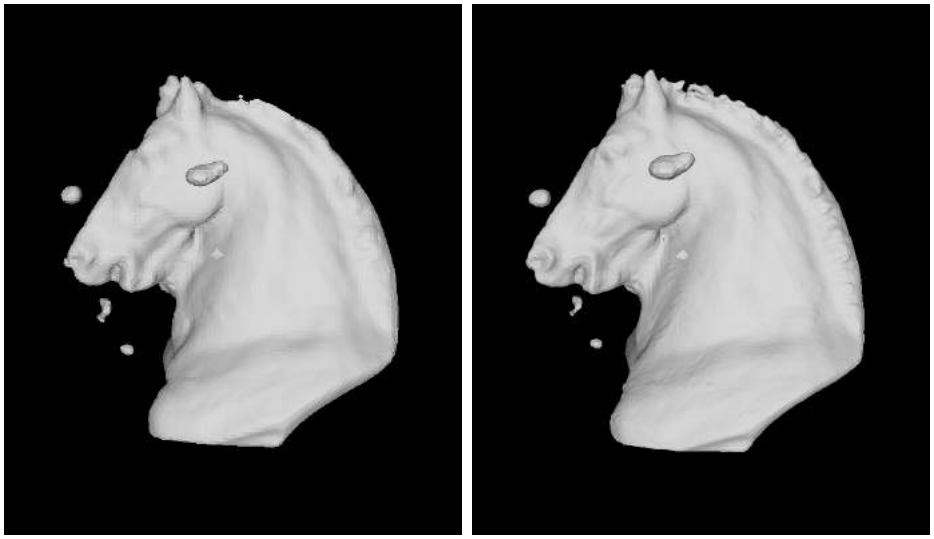
The mesh reconstructions for all the four cases are shown in Fig. 3.40 and the errors are reported in Table 3.5. The effect of not using low-rank decomposition lowers the reconstruction errors as visually inferred from the mesh reconstructions. In the cases where gray code low rank decomposition is not used, these sparse artifacts manifest as arbitrary blobs around the actual object mesh.

Table 3.4: Comparison of normal and depth errors for corrupted captures vs low-rank non-corrupted captures.

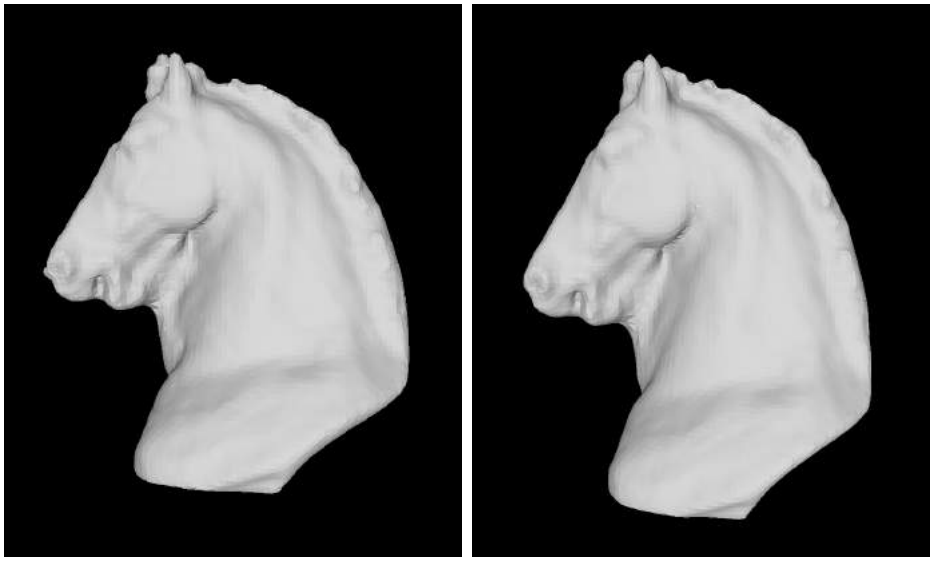
Error type	**Normal captures (corrupted)**	**Low-rank captures**
Normal (in degrees)	15.82	10.72
Depth (in mm)	2.63	1.79

Table 3.5: Comparison of mesh reconstruction errors (in mm) for the 4 scenarios mentioned in Section 3.5.2.

Method	**Noisy**	**Stereo LR only**	**Gray code LR only**	**Both LR**
Avg error	4.89	4.83	0.383	0.312



(a) Noisy (b) Stereo LR only



(c) Gray code LR only (d) Both LR

Figure 3.40: 3D mesh reconstruction errors in noisy environments for the 4 scenarios mentioned in Section 3.5.2.

3.5.3 3D scanning on a representative CNC lathe

Machining environment and setup modifications: We perform all CNC machining environment experiments on a Haas Toolroom Lathe (TL)-1. The machining environment and the lathe system is shown in Fig. 3.41. The raw stock that needs

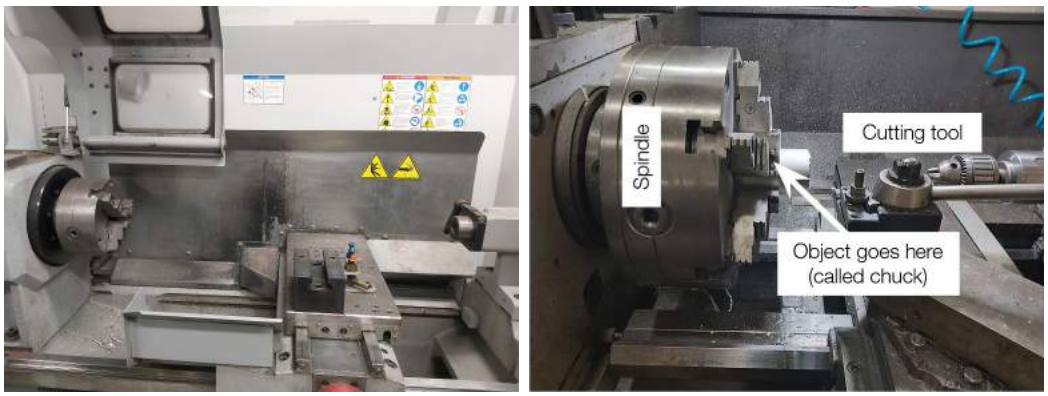


Figure 3.41: Machining environment for our scanning experiments.

to be manufactured is held tightly by the three metallic jaws known as chuck. The spindle is the part of the machine that spins at very high revolutions per min (RPM). We observed that putting in a raw stock and revolving it produces chatter at the edges due to manual setting errors. To prevent these edge chatters, we do a manual cut of the raw stock into a smooth cylinder such that the axis of this cylinder is centre-aligned with the spindle. To make sure, these edge movements are at a minimum, we use a touch probe system whose deviations determine the smoothness of this cylinder. We carry this process in an iterative fashion, till we are below a roughness tolerance for the object surface. We build a stage on one side of the lathe system to place our 3D reconstruction setup. Our scanning schematic and actual setup on the lathe machine is shown in Fig. 3.42. Due to the space constraints on the CNC lathe machine, we use a mirror in front of the projector as it has a minimum working distance for the light to be in focus. We also incorporate polarizers in our setup to mitigate the effect of the complex light interaction with the specular/anisotropic object which interferes with accurate shape reconstruction. The effectiveness of using polarizers for 3D point cloud reconstruction using structured light scanning is shown in Fig. 3.43. The algorithm gives us wrong estimates of the point cloud due to

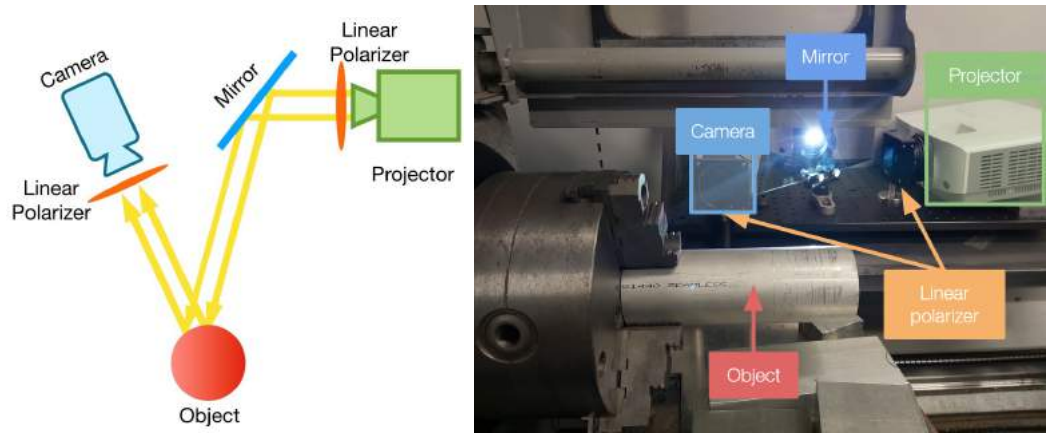
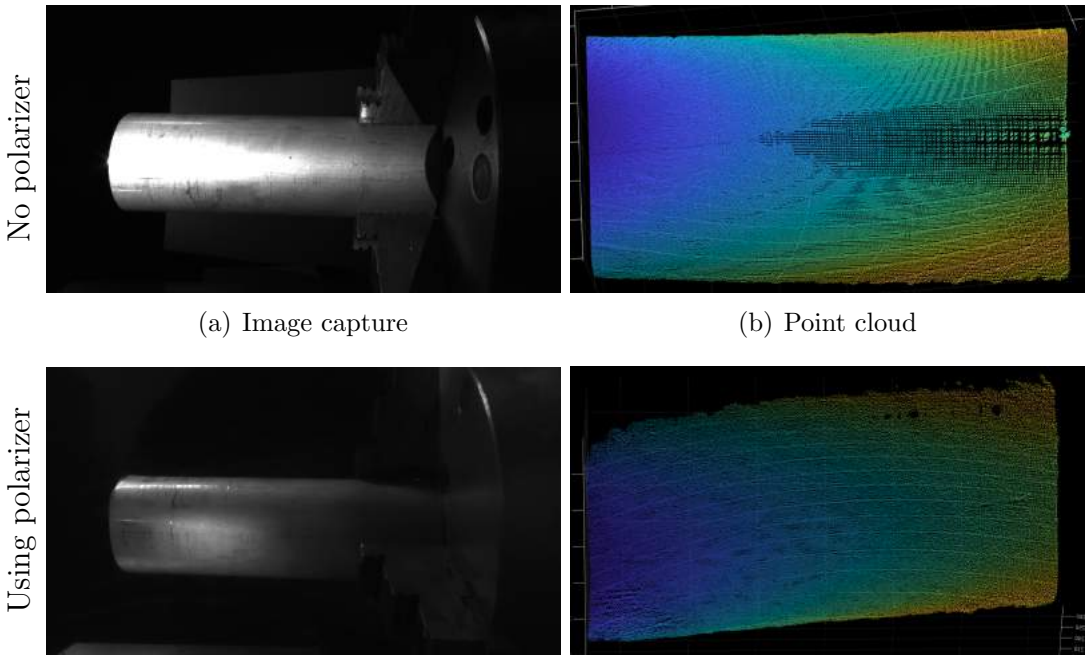


Figure 3.42: 3D scanning setup on a representative CNC lathe machine.



(a) Image capture (b) Point cloud

(c) Image capture (d) Point cloud

Figure 3.43: Advantage of using polarizers for imaging objects with complex isotropic and anisotropic BRDFs.

saturated specular highlights near the front middle area of the object when there is no polarizer. Using a polarizer, results in smoother point clouds and also reduces the errors in camera projector correspondences resulting in more accurate 3D location

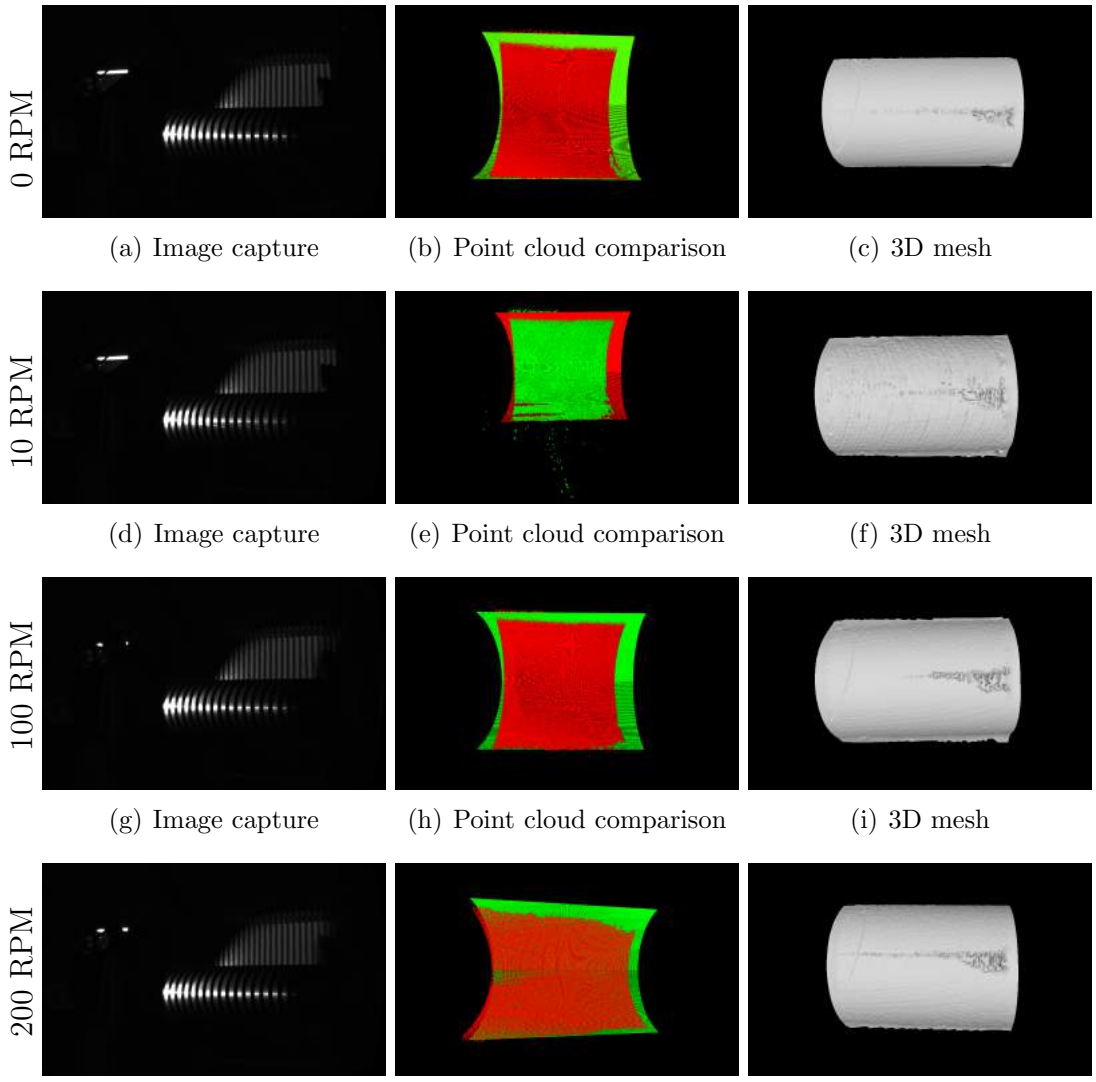
estimates.

Results: We calibrate our camera projector system for 3D stereo reconstructions using the procedure mentioned in Section 3.5.1. We use a metallic aluminium cylinder that has been manufactured on the lathe for stable centre alignment with the spindle axis. We generate the mesh from the oriented set of point clouds under varying revolutions per minute (RPM) of the object. For the reference ground truth mesh structure, we measure the object at 10 different positions along its length using a digital micrometer and average the distance. We fabricate the reference cylindrical point cloud using this ground truth distance for our error comparisons.

We scan the object under 0, 10, 100, and 200 RPM for our quantitative evaluations. From Fig. 3.44 and Table 3.6, we conclude that the object reconstruction has the highest accuracy when it is static. When the object is revolving at 10 RPM, the textures/scratches on the object manifest as line-like artifacts in the image captures. These cause 3D location outliers as well as line-like artifacts as seen in the point cloud image (zoom in). These artifacts are prominently visible in the mesh reconstruction for 10 RPM. However, when the object rotates at 100 RPM, these artifacts tend to average out due to the high speed. This makes captures under 100 RPM more robust, decreasing the mesh reconstruction error and producing a better surface mesh. In 200 RPM, we observe more edge vibrations compared to 100 RPM due to the higher capacity at which the machine operates. This results in some wrong correspondence estimates near the edges thus increasing the reconstruction error compared to 100 RPM.

Table 3.6: Comparison of mesh reconstruction errors (in mm) for different RPM of object on CNC lathe.

Revolution speed	**0 RPM**	**10 RPM**	**100 RPM**	**200 RPM**
Avg rec. error	0.515	0.574	0.52	0.548



(a) Image capture (b) Point cloud comparison (c) 3D mesh

(d) Image capture (e) Point cloud comparison (f) 3D mesh

(g) Image capture (h) Point cloud comparison (i) 3D mesh

(j) Image capture (k) Point cloud comparison (l) 3D mesh

Figure 3.44: Structured light scanning sample captured image, point cloud comparison with reference point cloud (in red), and the reconstructed 3D mesh for part under 0, 10, 100, 200 RPM.

3.6 Summary

We present an active scanning imaging setup that combines photometric stereo with structured light scanning. We use the estimated normals and depths to reconstruct point clouds and 3D meshes. We use our setup to:

1. reconstruct 3D geometries objects of different BRDFs in a desktop environment;
2. make our 3D reconstruction robust by mitigating sparse specular corruptions - ubiquitous in a machining environment - through a low-rank matrix recovery-completion algorithm;
3. reconstruct accurate 3D shapes of representative specular and anisotropic objects on a CNC lathe under high-speed revolutions.

This chapter addresses the performance of structured light scanning and photometric stereo in lab-based noisy environments and machining environments. The high reconstruction error (0.5 mm) and long capture times required are unfavourable for accurate real-time scanning on CNC machines. We resolve these issues in the next chapter where we propose a novel scanning setup having rapid acquisition times and improved 3D reconstruction accuracy.

Chapter 4

Ballistic shape from silhouette for 3D reconstruction in scattering medium

The presence of coolant fluids is essential in an active machining environment. They are used for dissipating the heat generated from the friction between the cutting tool and the high-speed revolving raw stock. Coolant fluids also help in the smooth machining of objects by lubricating the cutting tip as well as preventing them from rusting. However, they possess an inherent disadvantage for vision-based scanning systems due to their scattering properties. In this section, we deal with the problem of scanning and reconstructing the shape of objects in the presence of different scattering mediums and specular metal shavings. We propose a shape from silhouette ballistic scanning setup that efficiently images high contrast and sharp silhouette images by selective light path captures. In addition, we propose an algorithm that is used to reconstruct the point cloud and mesh from a single silhouette image, enabling rapid in-line error feedback. We use our scanning system to accurately reconstruct the 3D shape of rotating objects on a representative CNC lathe machine down to 60 μm. We also introduce a spatio-temporal processing technique that mitigates the effect of scattering fluids and sparse specular shavings, making our scanning setup robust in an active machining environment.

This chapter is arranged as follows: Sec. 4.1 describes general light transport

theory in scattering medium. Sec. 4.2 forms the core theory behind our novel ballistic shape from silhouette scanning setup. Sec. 4.3 delves into the hardware details of our novel scanning system. Sec. 4.4 contains algorithm designs for single-image 3D shape reconstruction and spatio-temporal processing. Sec. 4.5 shows our descattering and 3D geometry reconstruction results in controlled lab settings and noisy machining environments.

4.1 Light interaction in scattering medium

Conventional vision based systems are often inefficient in imaging through scattering media like murky water, fog, and diffusing liquids. An extensive amount of work has been done in hazy image restoration through removal of the back-scattering component known as airlight [60]. A common assumption in such methods is the linear propagation of the light paths from distant light sources which is violated in near-field illumination. Light undergoes multiple travel paths in a complex scattering medium (Fig. 4.1). A part of the photons may lose their energy as well as coherence properties completely. These photons get absorbed by the particles of the medium after undergoing multiple scattering events. A few of the photons will undergo specular reflection following Snell's law. The energy content of such photons is completely preserved with a reversal in the propagation direction about the normal. Following, we have multiple scattered photons, that prohibit the clear imaging of objects inside or behind the scattering medium. These photons are called diffusive photons and are the principal reason for brightness attenuation, blurred boundaries, and contrast loss in captured images. The degradation effects of the scattering medium on images increase exponentially with distance, making it challenging to capture internal object details. A fraction of the photons might undergo very few scattering events. These photons are called snake photons and may also be helpful in imaging the objects behind the scattering medium. Finally, we have photons that take the shortest path and pass straight through without scattering. They are called ballistic or transmissive photons and they exit the medium at the same angle that they entered. They also have their coherence properties preserved. The amount of ballistic photons exiting the medium is dependent on the concentration and the wavelength of the light source. It is crucial in our application to reduce the effects of scattering paths and amplify

the effects of these transmissive photons for gaining the ability to reconstruct object geometry accurately.

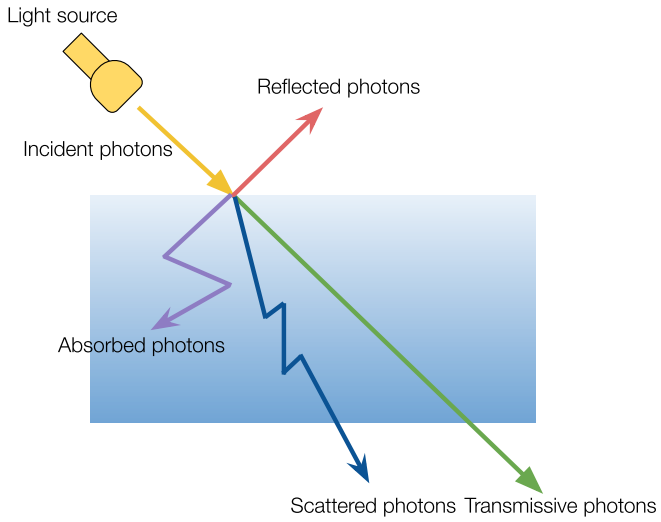


Figure 4.1: Interaction of light with scattering medium.

4.2 Seeing through scattering medium

Transmissive paths are vital in revealing the structure of an internal object present in a diffuse medium. Multiple strategies derived from light propagation theory can be used for isolating transmissive photons from scattered paths. The first strategy is using polarization that constrains the electric field wave oscillation orientations to a single plane. When incident light paths travel through a medium in a straight line, they do not change the polarization direction. This can be used to separate the scattered light paths that encounter a change in their polarization properties due to collision with multiple particles inside the medium. The second strategy is to use angular cues. When photons travel from the light source to the camera in a straight line, they do not undergo a change in the direction at the other end of the scattering medium. However, this direction is modified in the case of scattered light paths due to multiple interactions with the scattering medium particles. The third strategy exploits the positional differences between the incident and the emitted photon paths. A collimated beam of light in air will originate from the opposite

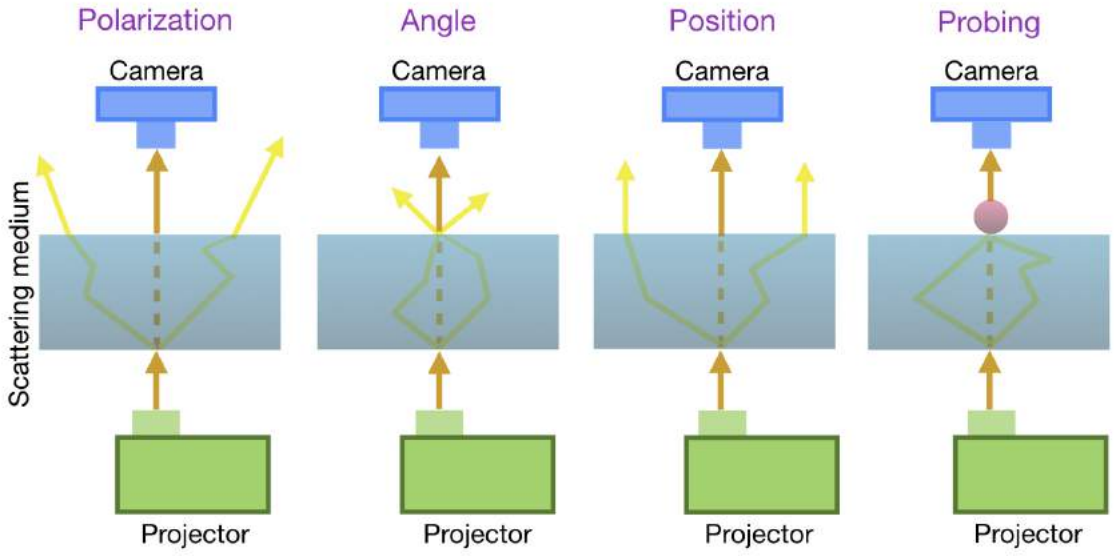


Figure 4.2: Different cues for imaging using only transmissive light paths.

end at the same position. However, in a scattering medium, photons on the other side might emerge in straight lines but will observe a shift in positions compared to the ballistic photons. Thus, through careful alignment of the light source and the sensor, we can use the position as a cue, such that the camera only captures the photon beam that is straight to the light source. The final strategy is using light probing techniques for rejecting scattered paths. A probing strategy that can be used for capturing only ballistic paths is aligning the illumination light plane and the corresponding sensor rows such that only one row is exposed in the camera at one time. This idea is similar to the epipolar scanning strategy [54] where they only expose the camera rows that lie the epipolar plane. A minor cue that can be used for descattering is wavelength that suggests using longer wavelength light sources to minimize the scattering effect. These cues are summarized in Fig. 4.2

Our objective is to capture the ballistic photons that travel in a straight line giving us sharp high-contrast silhouette images. Theoretically, we could use a collimated point light beam with a 2D micron-scale resolution translation stage to light up the scene one pixel at a time and activate only the corresponding camera pixel on the opposite side. Additionally, we could measure the distance between the light source

and the sensor and compute the light travel time for time-gating the sensor pixel. This will only deliver the rays that travel in straight lines from the source pixel to the sensor pixel. However, due to hardware limitations and complexity it is not feasible. Our designed imaging setup is a more generalized version of this idea and uses positional and angular information, combined with ballistic light path probing strategies. The incorporation of multiple cues helps to disambiguate scattered light from ballistic light.

Our setup is essentially a projector-camera system. Our projector emits light planes into the scene, which gets modified by the scattering medium containing the object and is received on the sensor end. A prototype diagram of our setup is shown in Fig. 4.3. We align our projector-camera in a confocal fashion. The projector and camera are placed on the opposite sides of the scattering medium. This satisfies positional cues, as in air, light planes transmitted from the source will enter the sensor directly on the other side. The choice of camera and imaging lens, as well as the illumination source, play an integral role in our setup. For satisfying angular cues, we use a telecentric projector along with a telecentric imaging lens that captures only the rays parallel to the optical line. Telecentricity is a property of imaging optics that makes the principal rays emerging from a lens parallel to the optical axis. This is done by placing an aperture at the focal point of the lens to block all off-axis light rays. Telecentric lenses offer an orthographic view of the object eliminating perspective error, thus negating any angular deviation of the field-of-view of the object. Another advantage of telecentric lenses is the larger useable depth-of-field due to symmetrical blur outside the working distance area. In this case, the edges tend to retain their centre of mass making it possible to estimate the exact location due to the absence of parallax error. Telecentric lenses combined with telecentric illumination offer enhanced performance of the system. Diffuse sources tend to create incorrect edges due to the ray bending that causes overlaps of bright pixels in the background with the dark pixels of the silhouette edges. The characteristic combination of telecentric lenses having zero parallax error along with minimal lens distortion values (0.1%) and telecentric light sources emitting only parallel rays give us non-overlapping transmissive paths for accurate silhouettes.

We use our telecentric lens with a rolling shutter CMOS sensor. A camera equipped with a rolling shutter does not expose the whole sensor to the scene at once. Instead,

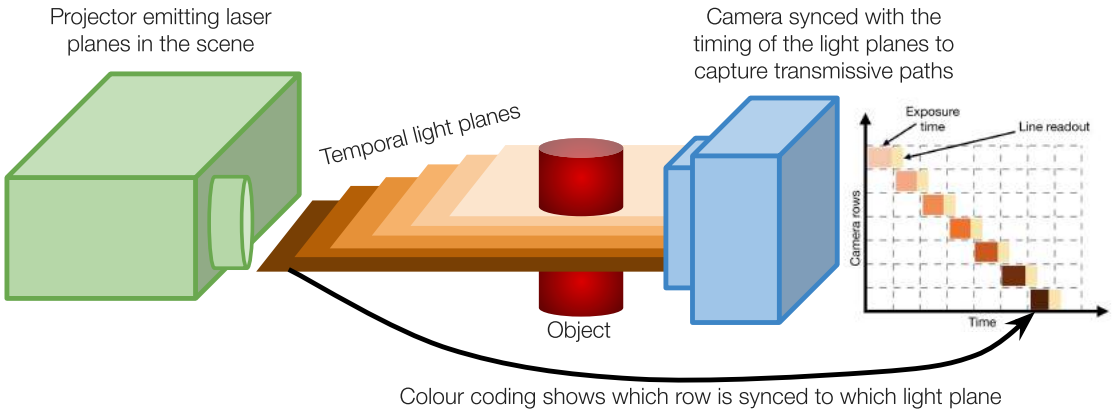


Figure 4.3: Proposed scanning setup concept for ballistic path captures.

it scans the scene from top to bottom by exposing only the rows of the image pixels to the scene. We use this space-time sampling of the rolling shutter sensors for our benefit by controlling the pixel clock timings of the row readouts. As we have the camera and projector placed opposite each other satisfying the positional strategy, we sync the positions of the light planes according to the exposed row positions of the rolling shutter camera. We are using the ballistic probing cue because the exposed camera row and the corresponding light plane lie on a single plane at any given time, having a one-to-one correspondence. A CMOS sensor with telecentric imaging and telecentric illumination lowers the exposure time while maintaining a high signal-to-noise ratio because almost all the rays travelling from the source are delivered directly to the camera sensor with minimal energy loss. An advantage of low exposure times are rapid image acquisition. This is extremely advantageous as CNC lathes operate at very high RPMs when they manufacture parts, thus modifying the object part in milliseconds. We do not incorporate the polarization cues in our setup due to the loss of light requiring more exposure time.

4.3 Hardware setup implementation

The schematic diagram of our ballistic shape from silhouette setup is shown in Fig. 4.4. It consists of broadly three components: the illumination side, the imaging side, and electronic equipment required to sync the illumination and the rolling shutter

camera.

The illumination part is the orthographic projector which projects horizontal light planes on the medium containing the object. The light source is a photodiode Lasos 532 nm 50 mW that gets focused on a fibre collimator lens (Thorlabs PAF1P-11A). The tip of this optical fibre acts as a point light source and is kept at a focal distance away from a Nikon 105 mm f/2.8 AF Micro lens. This, essentially, behaves as a collimated source of light. The beam diameter is approximately 13 mm in size and is dependent on the focal length of the collimating lens used. We pass this beam of light pass through a Thorlabs plano-convex 75 mm cylindrical lens (LJ1703RM) that makes the beam stay collimated in the horizontal direction but diverging in the vertical direction. This single-axis collimated light beam travels to the galvo mirror and gets reflected into the Thorlabs telecentric scan lens (LSM03-VIS). The laser line generated from this scan lens is around 6 mm limiting our scanning area. To increase our scanning field-of-view we place a 0.16X SilverTL$^{\text{TM}}$ telecentric lens at the working distance of this scan lens that increases the length of this laser line offering us a scanning area of 40 mm $\times$ 30 mm ($H \times W$).

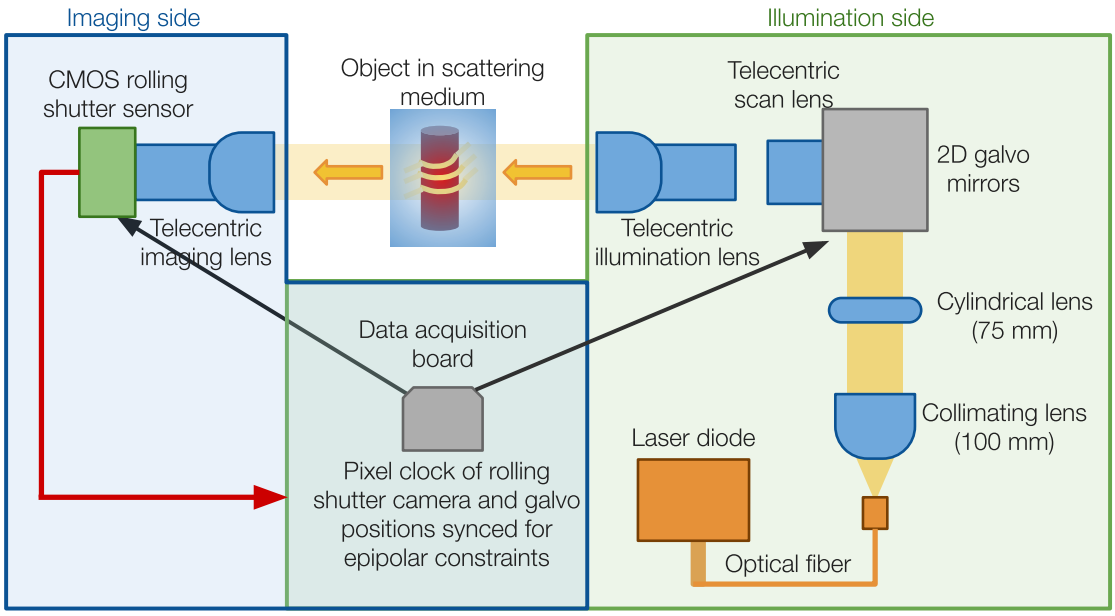


Figure 4.4: Schematic of our shape from silhouette setup for seeing though scattering medium.

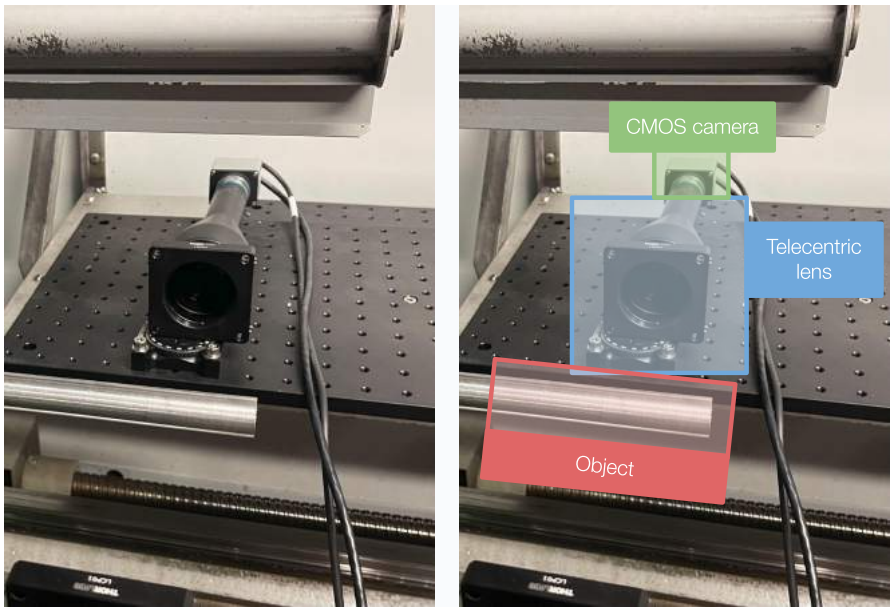


Figure 4.5: Imaging side of our practical hardware setup.

The light plane interacts with the object in the scattering medium and is captured on the other side of the object using another 0.16X SilverTL™ telecentric lens. We use an IDS UI-5240SE-M-GL monochromatic rolling shutter CMOS sensor camera. The pixel clock of the camera determines the speed of the rolling shutter as it moves from the top to bottom scanning the whole field of view. The angular positions of the galvo mirrors are also aligned to match the corresponding positions of the camera rows. Therefore, at any given time, and in no participating medium, the light from a single laser plane will only get transmitted to the exposed camera row while getting blocked anywhere else on the camera sensor. The exposed camera row and the light plane are coplanar. We use a National Instruments (NI) data acquisition system (USB-6343-BNC) for synchronising the galvo mirrors and the camera rows together that operate after receiving a manual trigger signal. The scene gets illuminated from the top-most galvo position to the bottom-most galvo position. The one-to-one correspondence between the laser plane and the camera row allows only the transmissive photons from the scene, giving us a clearer silhouette image. The actual illumination side of our proposed setup is shown in Fig. 4.6 and the imaging side is shown in Fig. 4.5. The laser line generated by our illumination system is shown in Fig. 4.7

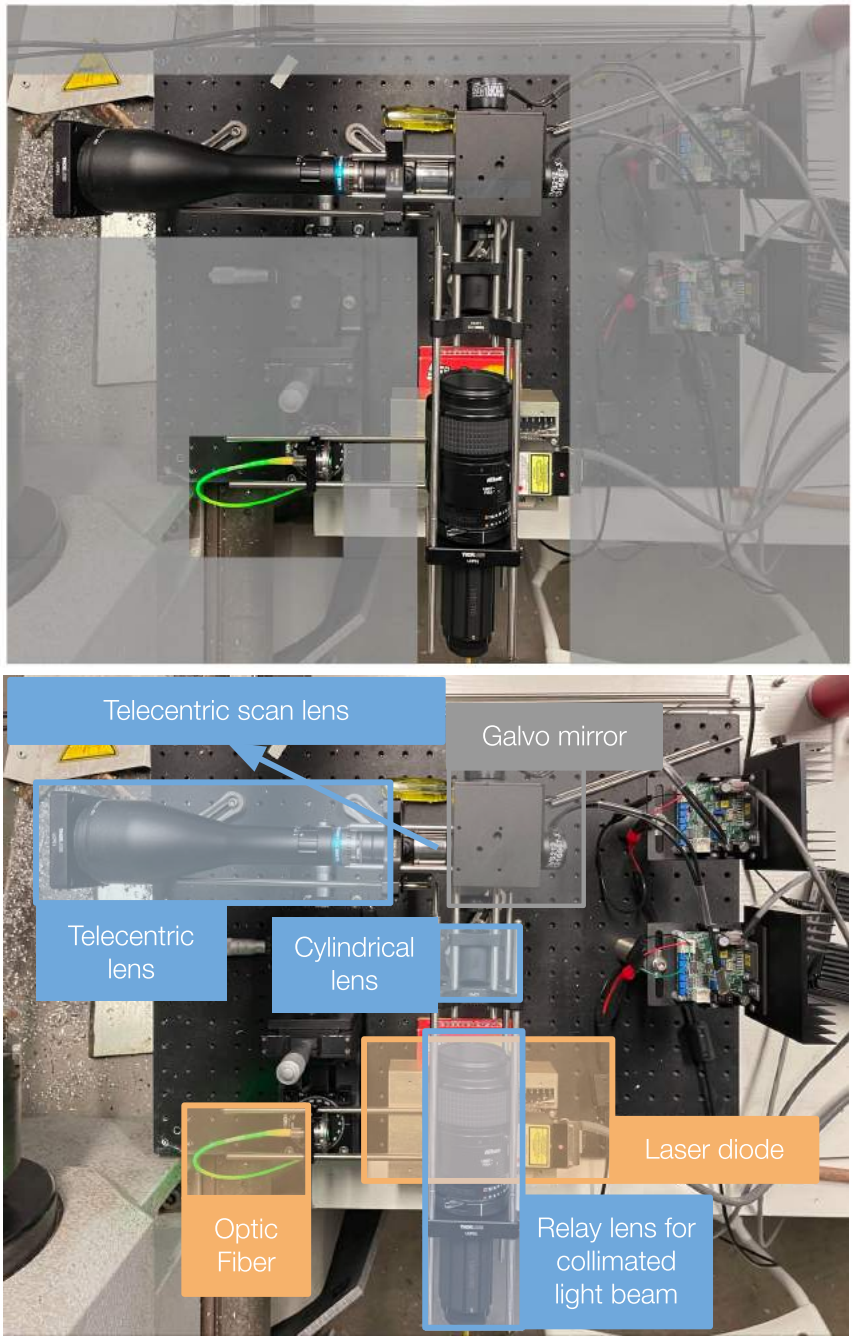


Figure 4.6: Illumination side of our hardware prototype.

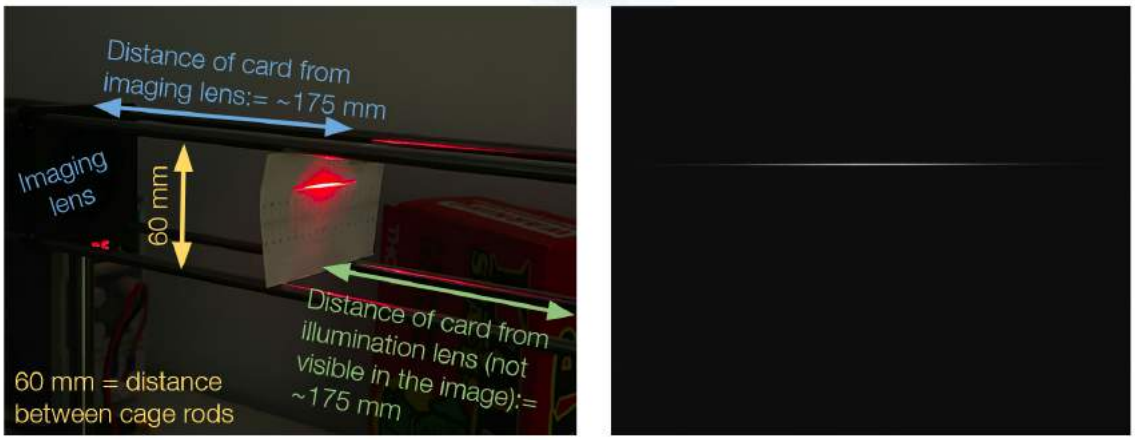


Figure 4.7: Laser line generated and captured using our setup.

4.4 Shape reconstruction from silhouette images

In this section, we first describe an algorithm to reconstruct the 3D geometry of a rotationally symmetric object from a single silhouette image in a non-machining environment. Next, we propose a spatio-temporal silhouette processing method for fluid and specular artifacts alleviation in a CNC lathe machining environment. This spatio-temporal processing combined with the reconstruction algorithm is used to reconstruct accurate object geometries for parts rotated at high RPM in the presence of dynamic sprayed fluids and metallic specular shavings.

4.4.1 3D shape from single image

In lab settings with static objects, we generally need a single silhouette image of rotationally symmetric objects to reconstruct the 3D shape accurately. Algorithm 2 describes the steps taken to generate a point cloud and the surface mesh.

4.4.2 Spatio-temporal processing for silhouette images

When scattering coolant fluids sprays objects revolving at high RPM, they get dispersed over a wide area. Similarly, the presence of metallic shavings makes imaging systems vulnerable to incorrect image measurements due to obscurity or specularities. In addition, both these adversities possess a threat of damaging the optical and

Algorithm 2 Algorithm for converting single captured image to point cloud and surface mesh.

procedure CREATEPOINTCLOUD(img)
    Threshold the captured silhouette image I with value $t = \frac{2}{N}\sum_{N}^{i=1} I(i)$
    Apply morphological snakes with an initial circular structural element of 100 pixels placed at the centre of the image (Marquez-Neila et al. [46]) to get a binary image
    Find the edges from the binary image using Sobel filter
    Using Hough transform, fit a line to the left and the right edge
    Compute the minor axis of the silhouette region
    D = LIST; PC = LIST; z = 1; ▷ Initialize diameter and point cloud list
    for $p \in P$ **do** ▷ P is the set of points in the left edge
        Using point $p = (x, y)$ and the minor axis direction $d = (d_1, d_2)$, traverse along the ray to find a point $q = p + t \cdot d$ on the right edge
        $d = p - q = \frac{\text{distance in pixels} \times \text{pixel size}}{1000 \times \text{magnification of lens}}$ ▷ Euclidean distance (mm)
        D.append(d)
        θ = linspace($[0 : \frac{2\pi}{N} : 2\pi]$ ▷ Circle polar form angle sampling
        $x = d\cos(\theta)/2$
        $y = d\sin(\theta)/2$
        $xyz = [x; y; z]$ ▷ Create circle point cloud for every pixel
        PC.append(xyz)
        z = z + 1 ▷ Increment z (height of cylinder)
    end for
    Average edge estimation error: $E = rms(D - R)$ ▷ R = reference measured distance
    Estimate normals (N) of point cloud (XYZ) at every point by fitting 10 neighbourhood pixels forming a plane
    Input (N, PC) $\rightarrow$ 3D mesh ▷ Using [38]
end procedure

electronic equipment. To avoid that, we shield our setup with clear acrylic sheets on the imaging and illumination side.

The presence of fluids and chips introduces many noticeable adversities in our scene. Fig. 4.8 shows the different kinds of degradation caused by the presence of fluids and specular shavings. Firstly, due to the friction of the object and surface tension of the fluid, drops tend to cling to the object edges. The adherence of these coolant fluid drops introduce black blob-like structures on object edges in images (red in left image of Fig. 4.8). This is because these drops act as spherical lenses that modify the straight light paths from the projector. They produce incorrect diameter estimation at the edge pixels, resulting in wrong point cloud location estimates and increased reconstruction errors. Secondly, some fluid particles get deflected from the object due to high RPM and get attached to the acrylic sheets. This hampers camera visibility producing bokeh-like patterns due to unfocused drops (yellow in left image of Fig. 4.8). Thirdly, the drops on the acrylic sheets also appear as random dark patches in the background (blue in left image of Fig. 4.8). The specular chips appear as sparse random gray artifacts in the camera image captures. This can be due to the metallic particle itself or due to the shadow cast by the particle on the object (red in right image of Fig. 4.8). We observe that both in the case of adhering drops and specular shavings, the positions of these degradations vary spatially with time.

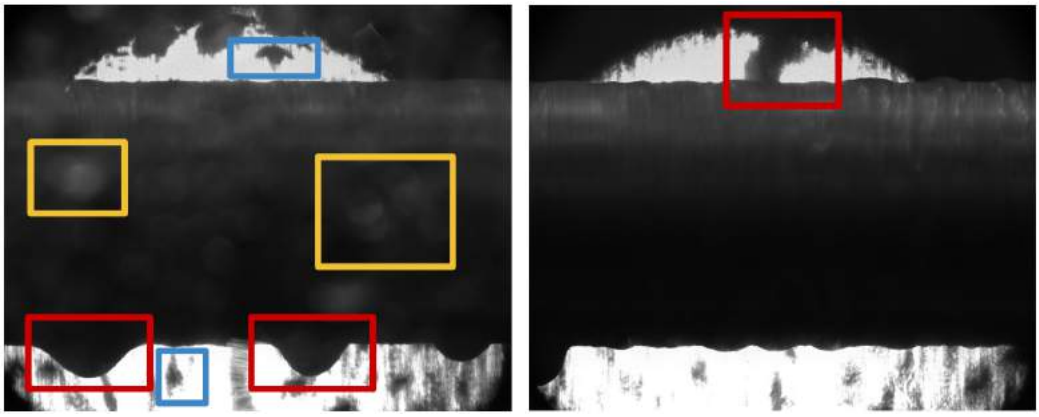


Figure 4.8: Left: Degradations caused due to the presence of scattering fluids. Right: Degradations caused due to the presence of specular shavings.

We benefit from the time-varying characteristics of these degradations and use a spatio-temporal processing to mitigate the effects of scattering fluids and specular

metal shavings. In a noisy machining environment, single image capture of the object contains bumpy edges, bokeh circular patterns due to unfocused drops, and sparse random dark patches due to specular chips. Instead of capturing a single image through a manual trigger, we capture a sequence of images through multiple VSYNC trigger signals that are synchronized with the capture of the images using the data acquisition board. The spatio-temporal processing algorithm uses the maximum intensity value among all the sequence images for every pixel location. As the object stays axially fixed, the uneven bumps on object edges created by the fluid get negated. The spatio-temporal processing also gets rid of the random dark patches created by the metal specular shavings as they change their position in every frame. Due to the benefit of our proposed telecentric setup, photons that get refracted due to the drops on the protective sheets get rejected. This does not introduce any bokeh-like halo structures in our ballistic image captures. Using the processed image from our spatio-temporal frames, we compute the gradient image. For every row in the gradient image, we select the top-2 column values of the gradient magnitudes. These two indices represent the two edge pixels of our object for that corresponding row. Our proposed spatio-temporal processing algorithm used for enhancing the camera vision system for accurate silhouette image is presented in Algorithm 3.

Algorithm 3 Temporal-spatial silhouette image extraction.

procedure GENERATESILHOUETTEIMAGE(Sequence of images)
  Given I number of images in a sequence with dimensions $H \times W$
  Initialize final image F = 0
  for $i \in I$ **do** ▷ Select one image from the sequence
    for $h \in H$ **do**
      for $w \in W$ **do** F(i, j) = **max**(F(i, j), i(i, j))
      end for
    end for
  end for
  Find the x and y gradients of the final processed image
  Calculate the magnitude of the gradient images: $I_{xy} = sqrt((I_x)^2 + (I_y)^2)$
  Traverse every column and find the top 2 gradient value indices i_1, i_2
  if $i_1 < H/2$ **then** ▷ H is the image height
    Assign i_1 to top edge index
  else
    Assign i_1 to bottom edge index
  end if
  Fit line to top and bottom edges using Hough transform
  Find slope of top line c
  if $c \neq 0$ **then**
    $rotateImg(F, c)$ such that silhouette image is horizontally aligned
  end if
  D = LIST; PC = LIST; z = 1; ▷ Initialize diameter and point cloud list
  for $p \in P$ **do** ▷ P is the set of points in top edge
    Traverse vertically to find a point q on the right edge
    $d = p - q = \frac{\text{distance in pixels} \times \text{pixel size}}{1000 \times \text{magnification of lens}}$ ▷ Euclidean distance (mm)
    D.append(d)
    θ = linspace($[0 : \frac{2\pi}{N} : 2\pi]$ ▷ Circle polar form angle sampling
    $x = d\cos(\theta)/2$
    $y = d\sin(\theta)/2$
    $xyz = [x; y; z]$ ▷ Create circle point cloud for every pixel
    PC.append(xyz)
    z = z + 1 ▷ Increment z (height of cylinder)
  end for
  Average edge estimation error: $E = rms(D - R)$ ▷ R = reference measured distance
  Estimate normals (N) of point cloud (XYZ) at every point by fitting 10 neighbourhood pixels forming a plane
  Input (N, PC) → 3D mesh ▷ Using [38]
end procedure

4.5 Experiments

This section delves into our results demonstrating the effectiveness of our telecentric scanning setup for imaging through scattering medium and high accuracy reconstruction in a lab setting and in a CNC machining environment.

4.5.1 Reconstruction algorithm

We show an example of our reconstruction algorithm described in Section 4.4 which requires a single silhouette image. We use a Thorlabs cage rod (left most image Fig. 4.9) for as our experimental object which is radially symmetric and has a radius of 6 mm. We show the captured silhouette image along with the linear edge fits in Fig. 4.9.

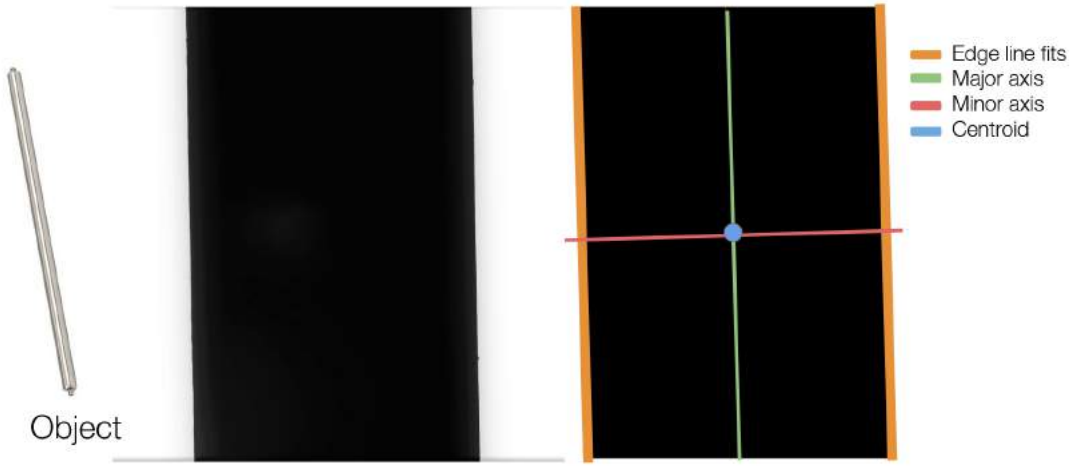
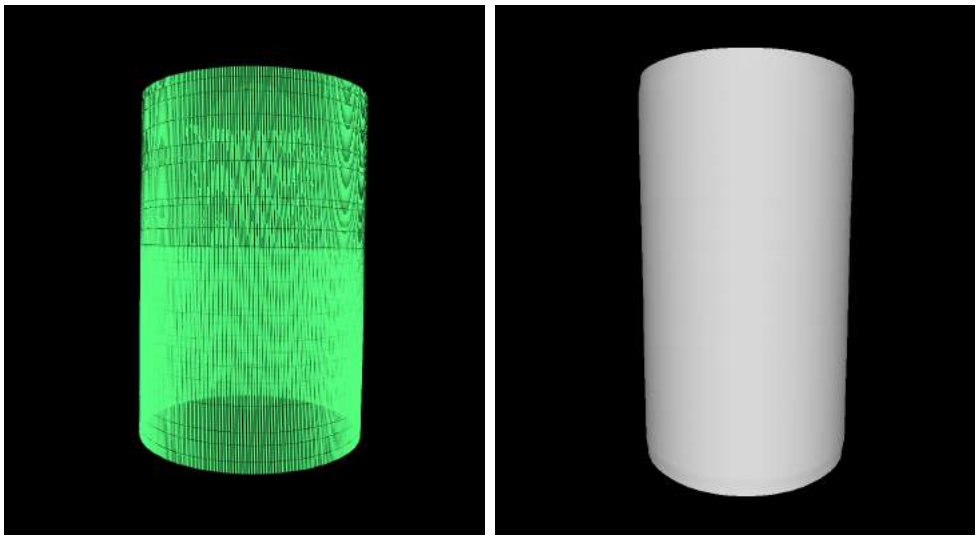


Figure 4.9: Object, silhouette image along with estimated edge fits, major axis, and minor axis used for point cloud and mesh formation.

The reconstructed point cloud along with the mesh is presented in Fig. 4.10

4.5.2 Descattering

We use objects of various shapes in different types of liquids as well consider different path lengths of scattering medium for our experimental evaluations.



(a) Point cloud (b) Surface mesh

Figure 4.10: Reconstructed 3D point cloud and surface mesh using our algorithm.

Qualitative results: We first establish the capability of our scanning method using a Thorlabs 1/4"-20 screw. We use a small glass container having a path length of 10 mm and mix together shampoo and water in various concentrations. We compare our method with a shape from silhouette capture method where all the rows are exposed at the same time. This deteriorates the camera vision by allowing diffuse photons in the camera sensor. We call this reference method 'telecentric capture' throughout this thesis as it still uses the positional and angular cues from Sec. 4.2, as captured with our telecentric scanning setup. However, it does not incorporate ballistic probing in contrast to our method. Our ballistic shape from silhouette scanning method captures only the transmissive paths by exposing the camera rows one at a time in synchronization with the illumination line. Fig. 4.11 shows a comparative performance of our scanning method with the telecentric method. We present the results in 3 different concentrations of shampoo. In lower concentrations, both methods give us relatively clear silhouette images. In the middle image, we can observe that our method gives sharper edges especially in the bottom area of the container where there is more concentrated liquid. For higher concentrations in the third image, the telecentric scanning silhouette gets completely blurred due to the scattering light paths while our method is still able to produce a coherent silhouette

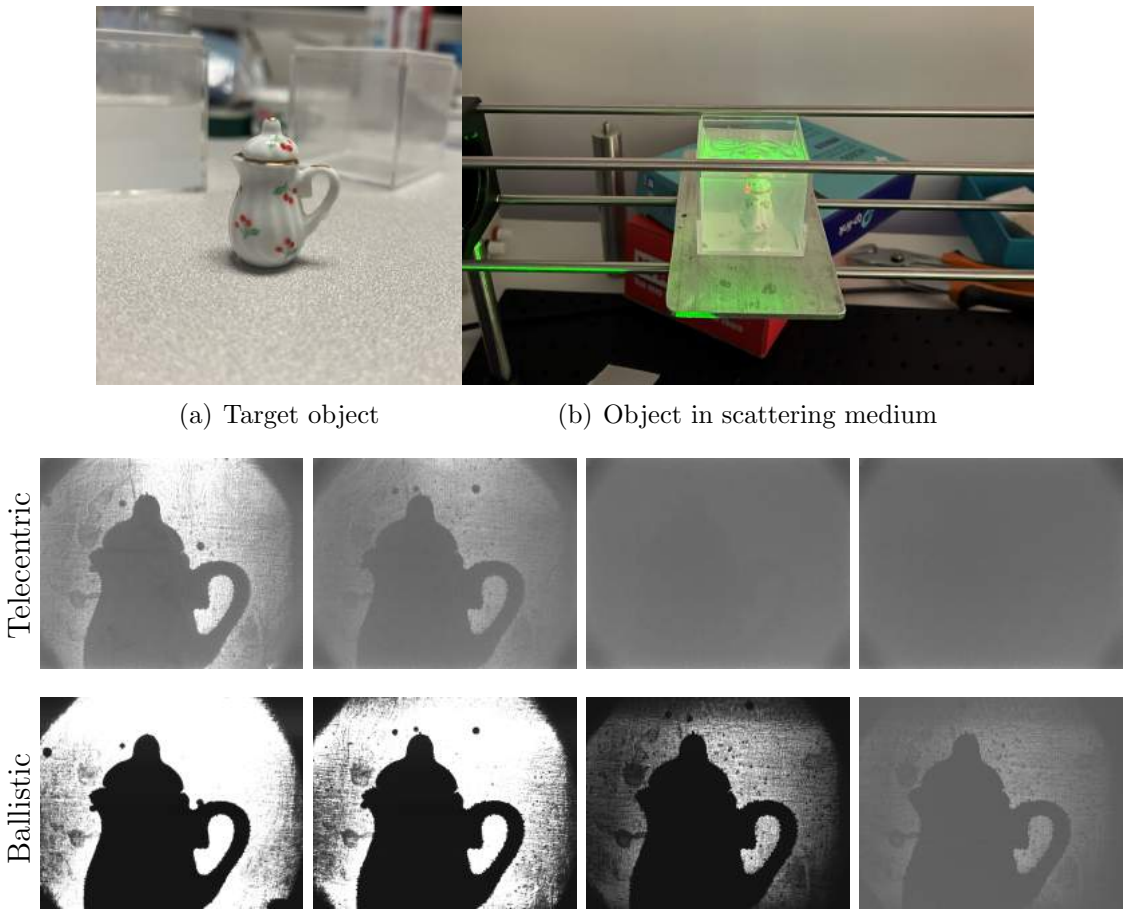
image.

(a) Target object (b) Object in scattering medium

Figure 4.11: Silhouette capture results for a screw using telecentric and our ballistic scanning method in different concentrations.

We also apply our scanning methods to a different scene. We use a different pot object as well as a liquid container cube having a path length of 50 mm. As the diffusion of light photons is path-length dependent, we observe more degraded vision in this scenario. We also change our scattering liquid from shampoo to a representative coolant fluid (TRIM^{TM} MicroSol^{TM} 585XT) used for general CNC machining. We

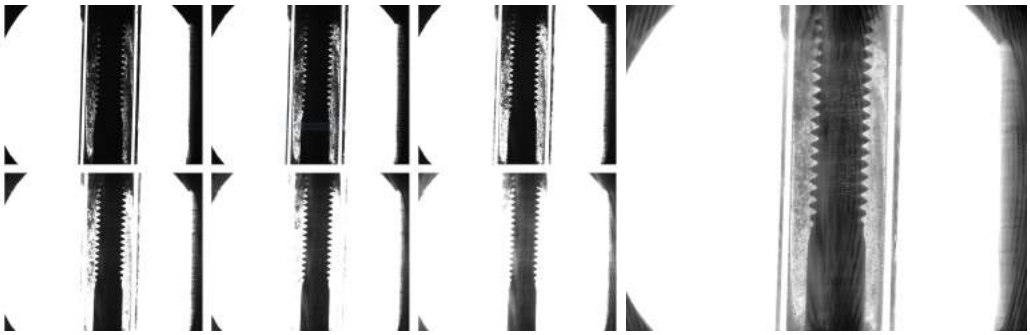
capture silhouette image measurements at varying concentrations of the fluid and demonstrate the performance of both our methods under different conditions in Fig. 4.12. We see from the results that in very dilute concentrations both the methods can image accurate silhouette images though the telecentric silhouette images lack contrast. However, in higher concentrations, our time-gated ballistic scanning method is able to produce sharp silhouettes distinct from the background. Another noticeable fact is the appearance of the flowery patterns in the first telecentric capture image. Our ballistic scanning method rejects any surface textures present on the object due to ambient noise rejection and gives us a dark silhouette image.



(a) Target object (b) Object in scattering medium

Figure 4.12: Silhouette capture results for the pot using telecentric and our ballistic scanning method in different concentrations.

HDR silhouette imaging in non-homogeneous fluids: In cases where we have non-homogeneous liquid concentrations as in Fig. 4.11, the silhouette visibility varies spatially. We can increase the camera exposure to capture the darker regions but they will saturate the brighter regions causing a loss of information. Generally, cameras have this limitation of a constrained exposure range which make pixels over-saturated or under-exposed. For such conditions, we can capture multiple low dynamic range (LDR) images of the object and combine them to form a high dynamic range (HDR) image for better visibility. We capture 6 such low dynamic range images and then use a simple weighing function to compost all these images together into an HDR image which retains all the details of the silhouette as shown in Fig. 4.13.



(a) LDR images (b) HDR image

Figure 4.13: Improving silhouette image with HDR imaging.

Quantitative results: For computing diameter estimation errors of rotationally symmetric objects, we use a Thorlabs cylindrical optical post which is 25.302 mm wide. We place the post in the previously used container having a path length of 50 mm. The target object is shown in Fig. 4.14. We fill the container with water and incrementally add 0.3 ml of the concentrated coolant fluid TRIM™ MicroSol™ 585XT. We use the process described in Algorithm 2 to generate the 3D point cloud. We use morphological snakes [46] for binarizing our silhouette images as due to the large path length, the light interaction with the scattering medium creates random textured patterns. This makes it hard to separate the background region from the foreground silhouette using simple thresholding operations. We show some results of the captured images using our telecentric and ballistic scanning technique (first and

third rows of Fig. 4.16). We show the results for some sample concentrations in rows

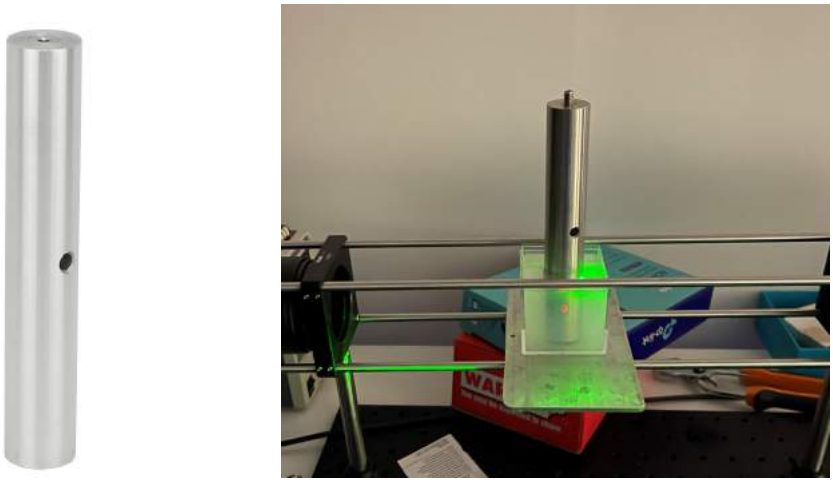


Figure 4.14: Target object for quantitative evaluations.

two and four of Fig. 4.16, which contain the estimated foreground silhouette (green edges) separated from the background (pink overlay) using the morphological snakes method. The concentration in the images is denoted as $a - b$ where a = amount of concentrated coolant fluid in ml and b = amount of water in ml. We can see that at higher concentrations $(1.8 - 90)$ of the fluid, the morphological snake segmentation process breaks down in the telecentric scanning method but scanning with our method provides robust silhouettes. This is because our ballistic scanning method offers better contrast and more pronounced edges by rejecting scattering paths, which is efficient for active contour morphological snakes. We use root mean square as our error metric. We compute this by taking the difference between the estimated edge distance at every row and the measured object edge. Table 4.1 summarizes our distance error estimates with increasing concentration. For the telecentric capture scenario, the edge estimation error increases with increasing fluid concentration till *Conc* 5. The ∞ in *Conc* 6 onwards indicate that the morphological snake method was unable to get a binary image due to image degradation in a scattering medium. For the ballistic scanning, we obtain high-quality silhouette images for the first four concentrations, thus having low reconstruction errors. After *Conc* 3, we observe a steady increase in the edge reconstruction error till the point where we are not able to extract a silhouette image for the ballistic capture. We also fit lines to our estimated edges as

shown in the left image of Fig. 4.15. We observe that the quality of the silhouette edge fits is dependent on the contrast between the background and the foreground. To verify that dependency, we selected patches of the background and the foreground (right image of Fig. 4.15), compute their mean, and calculate the difference in contrast for different concentrations. Table 4.2 summarizes the effect of concentration on the silhouette-background contrast. This reinforces the visual inference from the qualitative results that contrast between foreground and background decrease with increasing concentration making it harder to extract accurate edges.

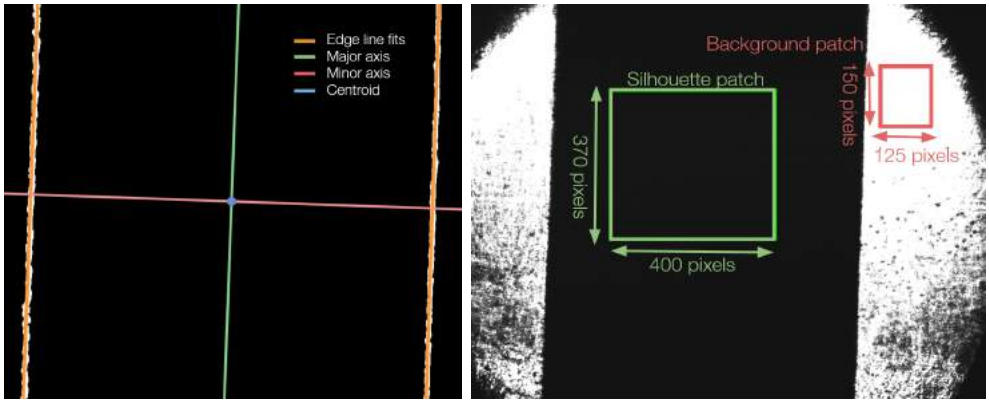


Figure 4.15: Left: Example of estimated edge fits using the background segmentation process using morphological snakes. Right: Silhouette and background patch selection for contrast comparison.

Method \ Concentration	Conc 1 0.3-90	Conc 2 0.6-90	Conc 3 0.9-90	Conc 4 1.2-90	Conc 5 1.5-90	Conc 6 1.8-90	Conc 7 2.1-90
Telecentric	0.203	0.24	0.278	0.747	6.786	∞	∞
Ballistic	0.094	0.081	0.068	0.0952	0.248	1.67	3.11

Table 4.1: Effect of fluid concentration on diameter estimation error (in mm).

Method \ Concentration	Conc 1 0.3-90	Conc 2 0.6-90	Conc 3 0.9-90	Conc 4 1.2-90	Conc 5 1.5-90	Conc 6 1.8-90	Conc 7 2.1-90
Telecentric	164	127	23	11	8	7	4
Ballistic	233	231	232	155	30	20	6

Table 4.2: Effect of fluid concentration on silhouette-background contrast.

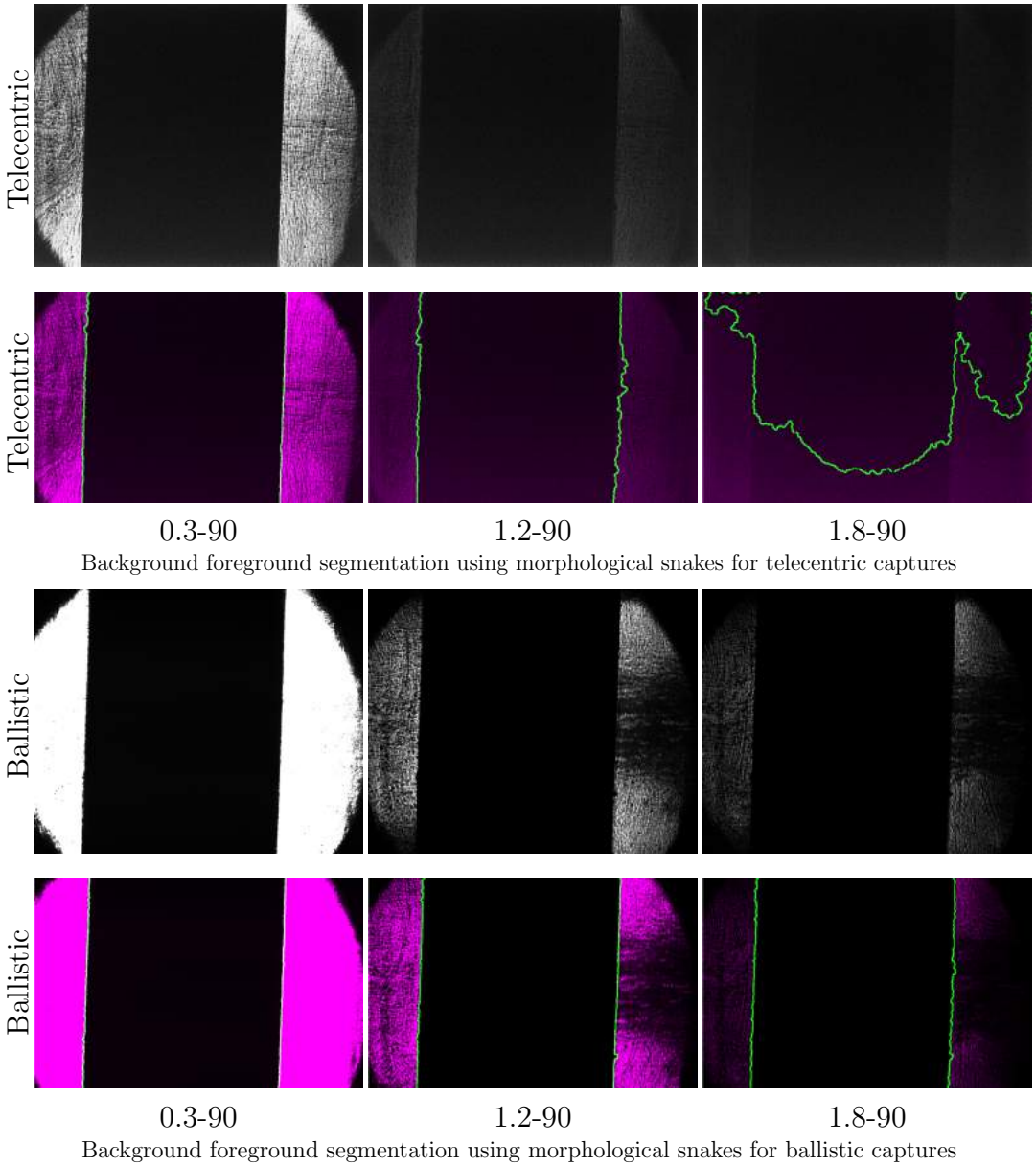


Figure 4.16: Silhouette extraction (in green) for post object in different coolant fluid concentrations for diameter error evaluations. Concentration of fluid is represented as 1—90 which means that the object was placed in a fluid having 1 ml of concentrated fluid dissolved in 90 ml water.

4.5.3 Scanning on a CNC lathe

In this subsection, we use our proposed setup to image objects on a lathe machine when they are rotating at a very high speed in the presence of sprayed coolant fluids and metal specular shavings. We perform experiments to demonstrate the effectiveness of our setup in generating accurate 3D reconstructions of objects from a single image frame in line with the manufacturing process. We discuss the description of the noisy machining environment and modifications made to our setup to enable efficient scanning, followed by our results using the spatio-temporal processing method as described in Sec. 4.4. A detailed error analysis follows to gauge the reconstructed point cloud and mesh errors.

Machining enviroment and setup modifications: We translate our shape from silhouette ballistic scanning setup on the Haas TL-1 CNC machine by placing the imaging side on the custom made stage. The illumination side and the synchronization electronics are placed on a table on the open side of the lathe. Our ballistic scanning setup on the CNC lathe is shown in Fig. 4.17.

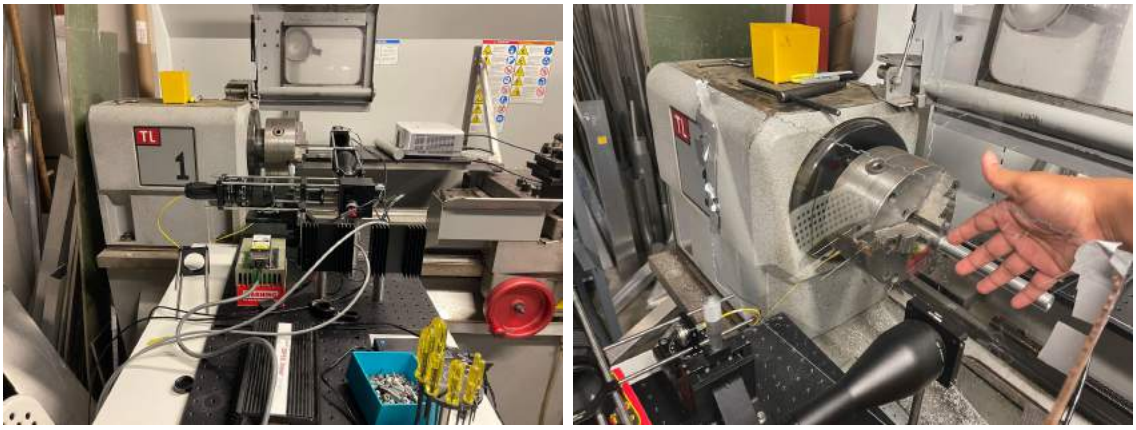


Figure 4.17: Left: Our shape from silhouette ballistic scanning setup on a Haas TL-1 CNC lathe. Right: Transparent acrylic sheets used to prevent our setup from fluids.

We mix the fluid TRIM^{TM} MicroSol^{TM} 585XT in a 10:1 ratio (10 parts water in 1 part coolant) as used for actual machining purposes. To imitate an actual noisy environment, we spray the coolant fluid continuously over the object as it rotates

along the spindle (Fig. 4.18). For our later experiments, we sprinkle specular metal shavings in our imaging field of view manually, along with spraying the scattering fluid. For safety purposes, the highest speed that we operate our CNC lathe is 500 RPM.

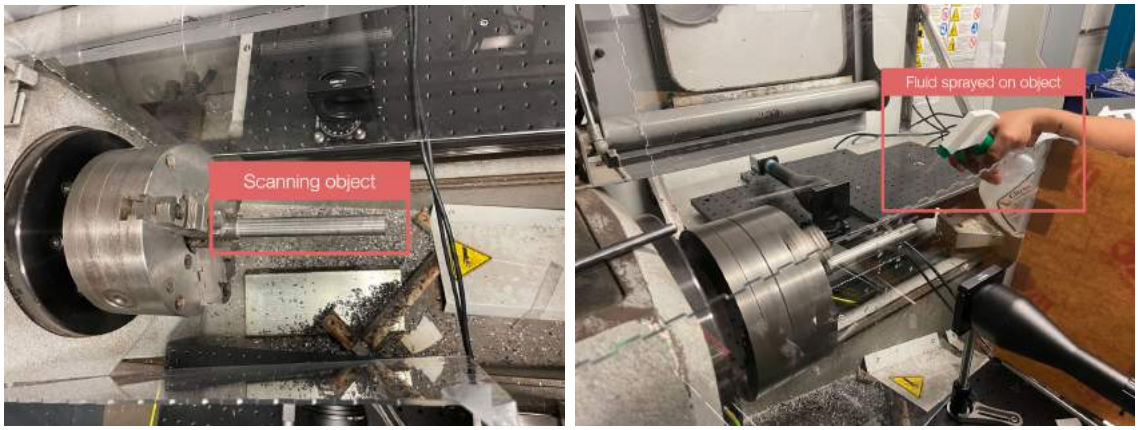
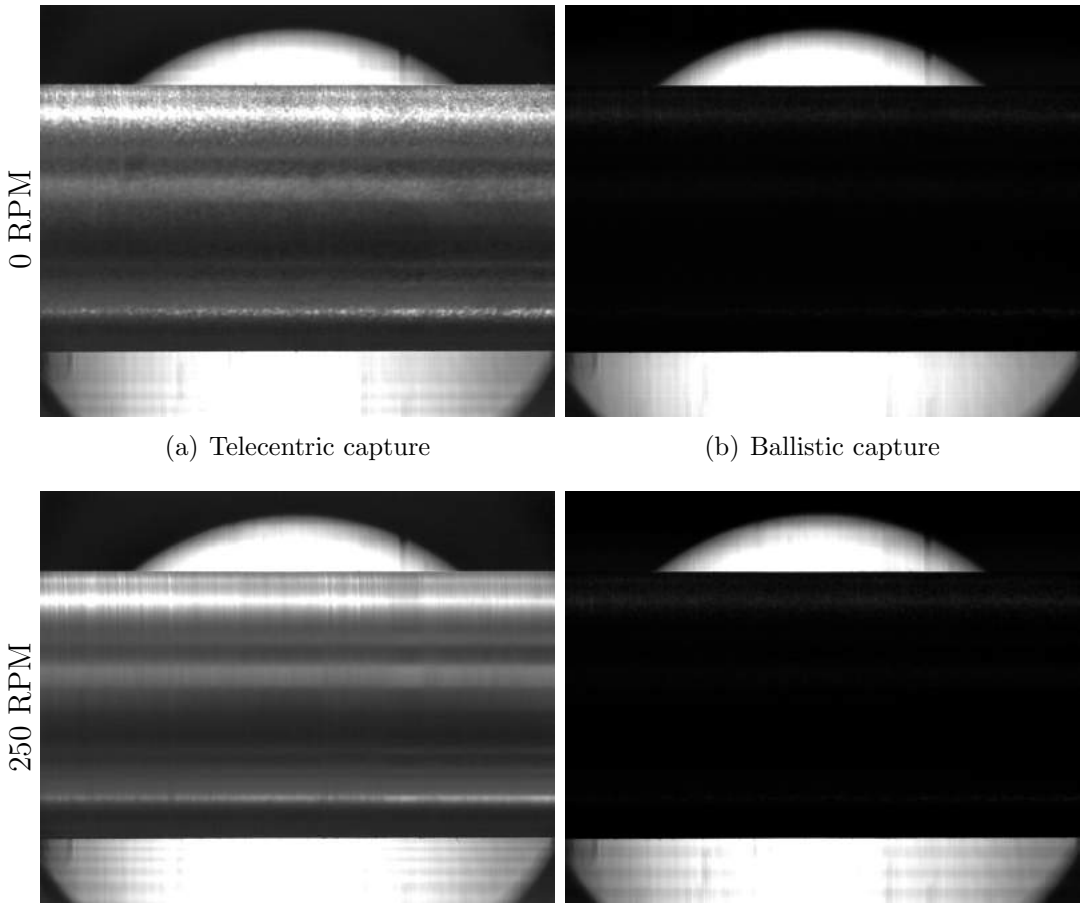


Figure 4.18: Left: Manufactured object used for scanning. Right: Fluid sprayed on the object.

Ambient noise removal: Our telecentric scanning setup is capable of taking high-quality silhouette images in the presence of scattering medium from our experiments in the previous subsection. The silhouette images captured are independent of the type of BRDF the object exhibits. In a controlled dark environment, the presence of ambient light can be neglected. This is not the same in a machining environment where there will be a substantial presence of ambient light sources for safety protocols. One of the very important advantages of our setup is its ability to reject ambient light. These ambient light paths interact with our object having a complex reflectance creating directional specularity when imaged under a telecentric capture setting. Due to the ballistic scanning methodology and extremely low exposure times, our proposed setup is able to reject these ambient light paths and textures on the object effectively by capturing only the straight light paths. We show examples of this property in Fig. 4.19 where the object rotates at 0 RPM and 250 RPM. The object in the telecentric capture appears to have a band like an appearance due to its material property while the silhouette image captured using our ballistic scanning method is completely black which enables efficient segmentation using simple thresholding operations.



(a) Telecentric capture (b) Ballistic capture

(c) Telecentric capture (d) Ballistic capture

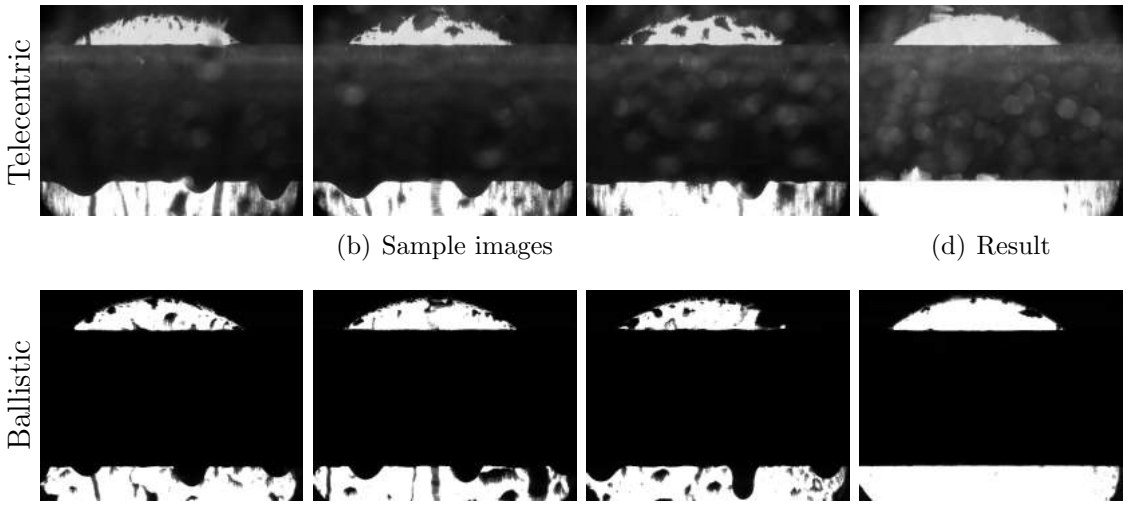
Figure 4.19: Ambient light rejection using our ballistic shape from silhouette scanning method in a machining environment.

Fluid effect mitigation using spatio-temporal scanning: We use a cylindrical object already manufactured on our lathe to align it accurately with the spindle axis. This prevents edge chatter when the part is rotated at high RPM. To ensure smoothness of the object surface, we use a gauge probe and ensure minimum deviation in its measurements.

The field-of-view is 40 mm $\times$ 30 mm, and therefore, the complete object in both directions is not visible to the camera. We ensure accurate reference measurements by using markers on the manufactured object. By doing this, we measure the scanning object only within the section which is visible to the camera. We take step-by-step measurements at ten different positions along the longitudinal section of the object using a digital micrometre. We average these measurements to obtain our ground truth diameter measurement and construct a reference point cloud.

The capture time for each frame is approximately 82 ms. For reference, at our highest speed (500 RPM), each rotation around the axis takes 120 ms. We capture continuous frames of the object when it is rotated in the lathe spindle. We spray the coolant liquid continuously to ensure the presence of fluid in every frame. The presence of bumps due to liquid drops is prevalent in the sample capture images for the telecentric and ballistic case (Fig. 4.20). However, we can see the obscurity effect caused due to drops adhering onto the protective acrylic sheets only in the telecentric capture scenario. Using our spatio-temporal processing in the telecentric capture setting improves the silhouette edges, but it also increases the obscurity effect due to unfocused drops. This is because the drops cause higher-intensity bokeh regions which are amplified when we take the maximum value across image sequences. Also, using the spatio-temporal processing method in telecentric scanning *reduces* the occupied area of the final processed silhouette image through pixel intensity saturation at edges. This occurs due to the complex interaction of light paths with clinging drops at the edges that do not get rejected. Our ballistic scanning method captured the image measurements devoid of such drop obscurities and artifacts. We get a very pronounced silhouette image devoid of any edge pixel saturations.

We perform two different types of quantitative analysis using the spatio-temporal processing method. We first construct the processed final image from the set of 200 sequential frames, then compute the magnitude of gradient image for every pixel. The edge pixels of the silhouette image are estimated by taking the top-2 maximum



(b) Sample images (d) Result

(f) Sample images (h) Result

Figure 4.20: Results using spatio-temporal image processing for robust silhouette images in telecentric capture vs our ballistic scanning.

values of the gradient image. This is shown in Fig. 4.21. We see that in the case of telecentric capture, the final processed image gets corrupted due to its inability to reject non-transmissive paths. These corruptions are in form of intensity saturation near the edges and bokeh-like artifacts on the object silhouette. In the ballistic case, we get sharp silhouettes with straight edges. This is because our ballistic scanning system is robust to the degradation effects of scattering medium and ambient noise. For computing the edge or diameter reconstruction error, we compute the root mean square distance of the two edge pixels at every vertical location in the image. The edge errors for both our scanning methods for different RPMs is presented in Table 4.3. This corroborates with our visual inferences as we obtain reconstruction errors in the ballistic scanning case down to 60 µm . In case of our telecentric scanning, the reconstruction error is at least 2× more and can go up to 400 µm. We also show the relation between edge reconstruction errors and number of images used for the spatio-temporal processing in Fig. 4.22 and Tab. 4.4. For this, we use only 400 RPM image measurements and a maximum of up to 500 sequential frames. In case of telecentric scanning, the error is the maximum for 1 image and keeps on decreasing till 200 images but again starts increasing. This can be attributed to the fact that a larger number of images cause more *reduced* edges at multiple pixels due

to saturation from drop refraction. Thus the area occupied by the object silhouette becomes smaller from the edges as we increase the number of images. Our ballistic scanning does not exhibit such behaviour and decreases expectedly with an increasing number of images. This is because more images make the silhouette estimate tighter thus reducing the error.

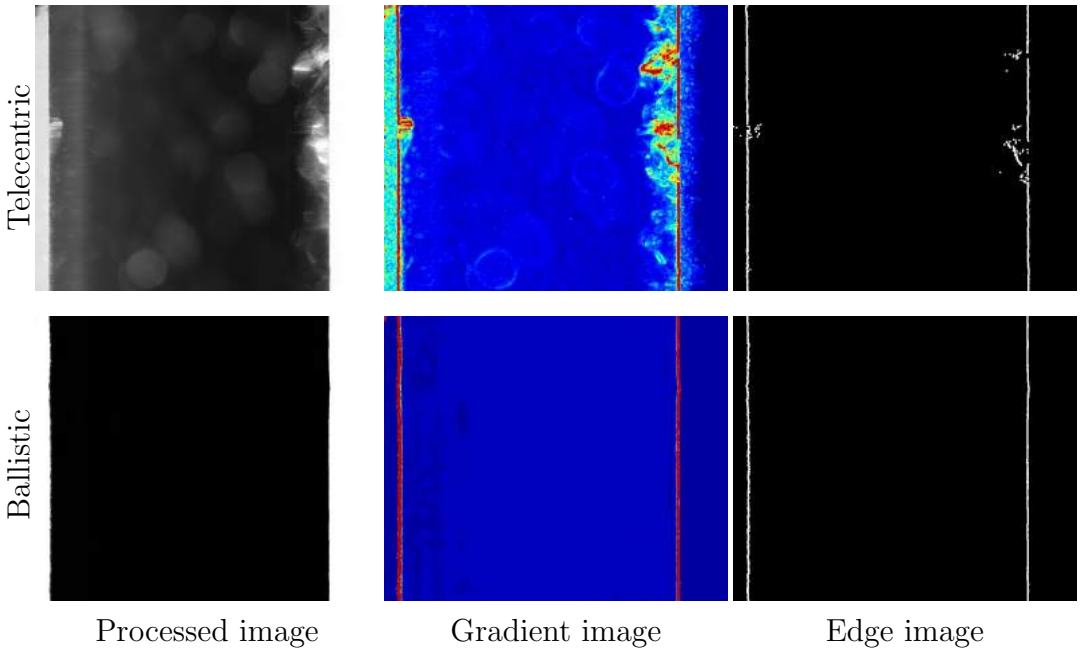


Figure 4.21: Processed image using our spatio-temporal method, gradient magnitude images (red spots show the highest magnitude pixels), and edge images for telecentric and our ballistic scanning.

Method \ Revolution speed	**50**	**100**	**200**	**300**	**400**	**500**
Telecentric	154	176	384	256	119	221
Ballistic	68	80	69	72	81	59

Table 4.3: Effect of different RPMs on edge reconstruction error (in μm) using spatio-temporal processing method using 200 frames.

Method \ Revolution speed	1	5	10	25	50	100	200	300	400	500
Telecentric	1444	941	884	622	346	181	119	163	174	190
Ballistic	669	333	310	267	222	152	81	73	67	57

Table 4.4: Effect of number of images used for spatio-temporal processing method on edge reconstruction error (in μm).

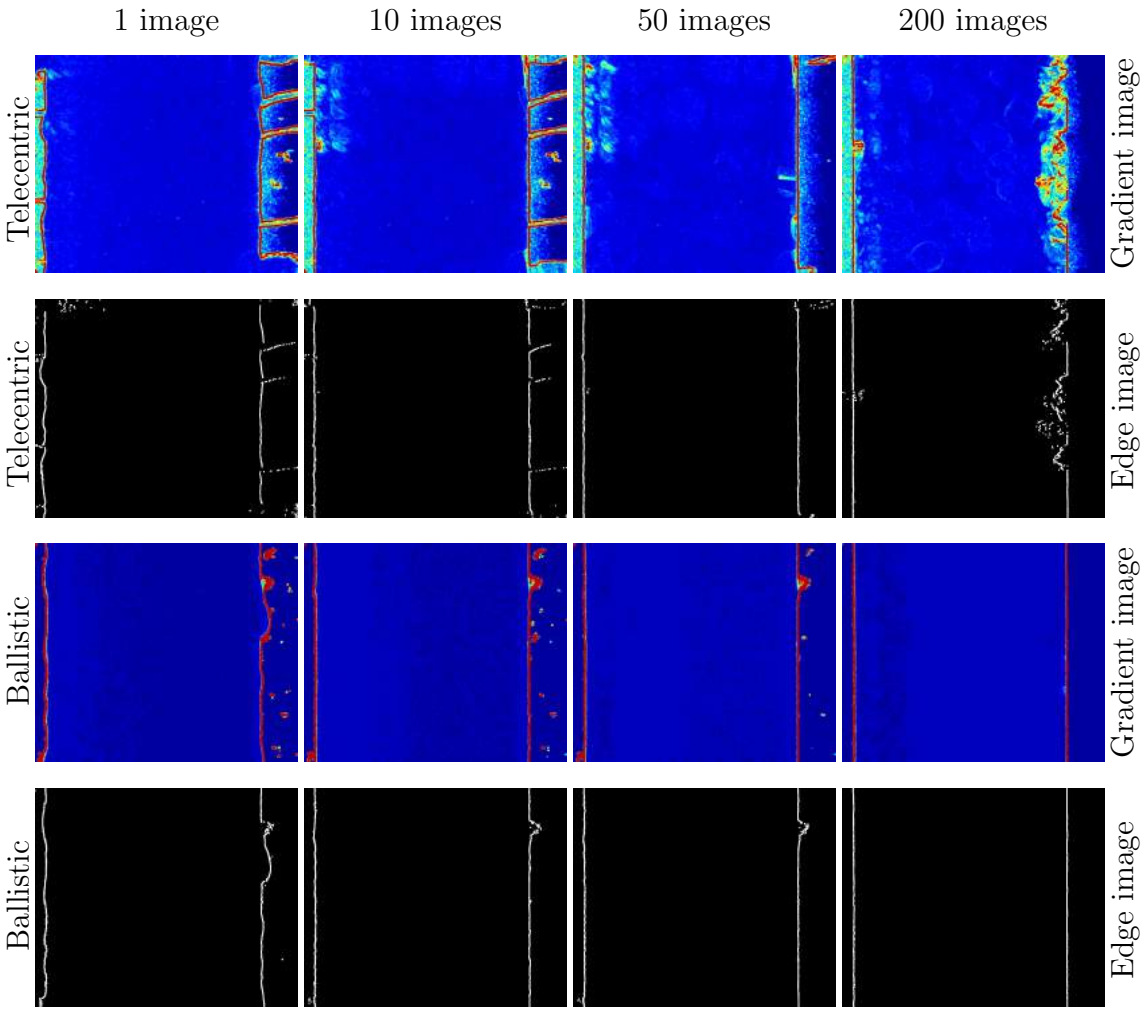


Figure 4.22: Gradient magnitude images (red spots show the highest magnitude pixels) and edge images for telecentric and our ballistic scanning for spatio-temporal method using 1, 10, 50, and 500 sequential images.

Scanning in presence of fluid and specular chips: To demonstrate the effectiveness of our scanning setup, we also perform experiments in the presence of scattering coolant fluid and specular shavings. We spray the fluid as in the previous experiments and manually pour metallic chips around the scanning field-of-view of our setup as shown in Fig. 4.23. For these experiments, we only capture 200 frames. The specular shavings manifest in our images as sparse and dark artifacts. The effect of chips in the telecentric and ballistic capture is shown in Fig. 4.24 in red boxes. The processed image using the spatio-temporal method, gradient image, and object edges are shown in Fig. 4.25. We perform the same quantitative evaluations as presented above: comparison with different RPMs and the effect of the number of images used for spatio-temporal processing on the edge reconstruction error. These are shown in Tables 4.5 and 4.6. The insertion of specular chips increases the edge reconstruction error in both the telecentric and ballistic scanning methods. However, for each setting, our ballistic scanning gives more accurate reconstructions.

Figure 4.23: Fluid and specular chips being introduced in our scanning environment.

Method \ Revolution speed	50	100	200	300	400	500
Telecentric	271	132	163	190	203	160
Ballistic	79	87	100	103	83	106

Table 4.5: Edge reconstruction error (in μm) using spatio-temporal processing method in the presence of specular shavings and coolant fluid for different RPMs.

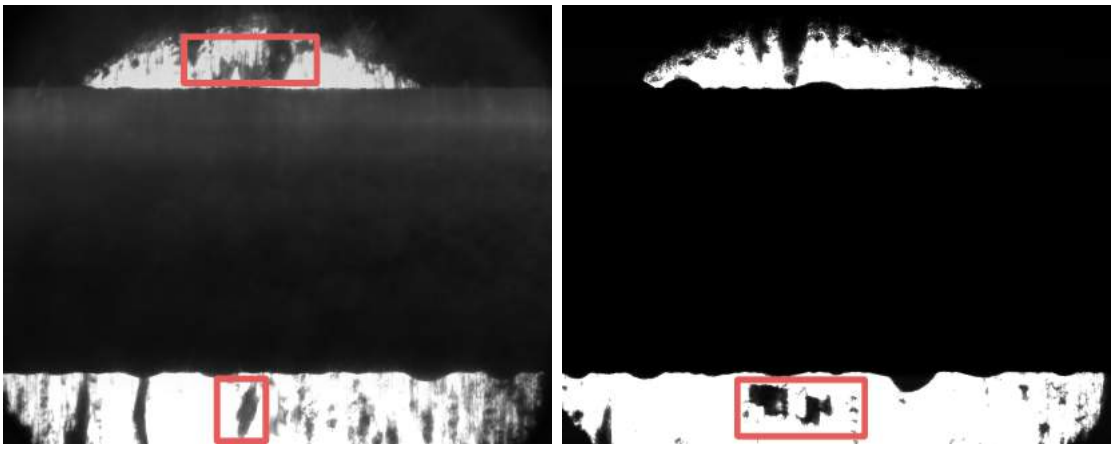


Figure 4.24: Effect of fluids and metal chips on telecentric and ballistic captures.

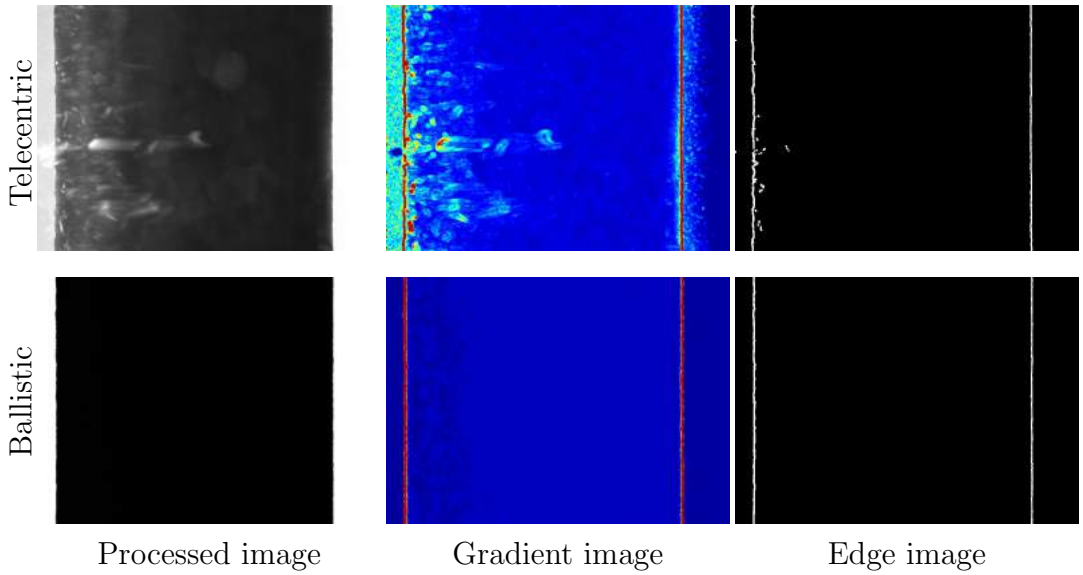


Figure 4.25: Final processed image using our spatio-temporal method, gradient magnitude images (red spots show the highest magnitude pixels), and edge images for telecentric and our ballistic scanning in presence of both coolant fluid and specular shavings.

Method \ **Revolution speed**	**1**	**5**	**10**	**25**	**50**	**100**	**200**
Telecentric	1123	908	885	850	776	663	203
Ballistic	460	244	247	184	155	111	83

Table 4.6: Effect of number of images used for spatio-temporal processing method on edge reconstruction error (in µm).

4.6 Summary

We propose a novel scanning setup that uses silhouette images to reconstruct objects in scattering media. It descatters the silhouette image by selectively imaging only transmissive light paths through an ballistic-only scanning method. Using this setup we:

1. Image through dense scattering medium having pathlength as large as 5 cm
2. Reconstruct the 3D geometry of rotationally symmetric objects with high accuracy
3. Improve reconstruction accuracy for parts on a CNC machine by alleviating the effect of scattering coolant fluids and specular shavings using a spatio-temporal processing method

Our setup is capable of acquiring silhouette images rapidly and obtain 3D reconstruction accuracy down to 60 µm. Using our spatio-temporal processing, we alleviate the degradations caused in the scanning environment due to the presence of scattering coolant fluids and specular metal shavings. Our proposed setup and scanning algorithm provides a fast, cheap, and convenient alternative to the use of contact-based CMMs used for accurate inspection.

Chapter 5

Conclusion and future directions

We explore non-contact vision-based setups for imaging under CNC noisy environments, and propose a novel silhouette-based imaging system for in-line scanning of rotationally-symmetric parts manufactured on a lathe.

In Chapter 3, we use structured light scanning and photometric stereo setup to scan objects with different BRDFs on desktop as well as on a CNC lathe. We also use rank constraints of captured images to perform a low-rank decomposition that mitigates the effect of specular chip-like noise. We get reconstruction accuracy down to 0.5 mm. One of the limitations of using this setup on an in-line CNC manufacturing unit is the time required to capture the images for 3D reconstruction that is approximately 6 seconds. Typically, a CNC lathe operates at over 1000 revolutions per minute which equals 16 revolutions per second. Thus, the object is bound to get modified throughout the projection-capture process involving multiple structured light patterns. This temporal modification of object shape in consecutive image captures will lead to incorrect 3D geometry estimates under a manufacturing setting. For future work, we can explore single shot structured light methods that encode colour and geometrical features for pattern design. Other disadvantages of this setup are the limited reconstruction accuracy and inability to scan in scattering medium. To increase the accuracy, we can increase the magnification of our imaging and illumination optics. However, this requires designing a 2D translation stage and capturing multiple sets of images for different locations of the object and stitching them to form a complete point cloud. For scanning in scattering medium, we can use

high-frequency patterns for separating transmissive paths from scattering paths.

In Chapter 4, we propose a shape from silhouette setup which uses ballistic probing along with angular and positional strategies to see through a scattering medium. We demonstrate efficient descattering of static objects in varying concentrations of different fluids. We also use high dynamic range imaging to improve visibility in non-homogeneous density distributions of fluids. We extend our proposed setup on a CNC lathe and use a spatio-temporal scanning method to get rid of artifacts created due to the presence of scattering fluid and specular shavings. This setup provides us with 3D reconstruction accuracy down to 60 μm for rotationally symmetric metallic objects under high-speed revolutions. A drawback of using this setup is the limited scanning field-of-view which can be increased or decreased by interchanging the set of imaging and illumination lenses. There is an inherent trade-off between magnification and reconstruction error where a higher magnification lens with a lower scanning field-of-view will reduce the reconstruction error while a lower magnification lens with a higher scanning field-of-view will increase the reconstruction error. A more complicated method for increasing the field of view is to use a 2D translation stage as mentioned above and capture multiple frames for different positions of the object. However, this requires precise engineering and accurate attachment of spatial image captures. Another disadvantage of using this setup along with our spatio-temporal processing method is that we can use only rotationally symmetric objects. For future work, we can explore reconstruction methods using aligned visual hulls for non-symmetric objects. Another direction of work involves exploring probing methods for rejecting scattered light paths. Transmitted light rays travel the shortest path length and take the least time to reach the camera because they travel in straight lines. However, scattered light rays collide with particles in the medium, taking a comparatively longer time to reach the sensor. We can use time-gated probing such that we expose the sensor for a short amount of time that is sufficient to isolate the ballistic photons from the non-ballistic ones.

Bibliography

[1] DG Abdelsalam, Baoli Yao, Peng Gao, Junwei Min, and Rongli Guo. Single-shot parallel four-step phase shifting using on-axis fizeau interferometry. *Applied optics*, 51(20):4891–4895, 2012. 2

[2] Juan-José Aguilar, F Torres, and MA Lope. Stereo vision for 3d measurement: accuracy analysis, calibration and industrial applications. *Measurement*, 18(4): 193–200, 1996. 1.1

[3] Sung Joon Ahn and Markus Schultes. A new circular coded target for the automation of photogrammetric 3d-surface measurements. *Optical 3-D measurement techniques IV*, pages 225–234, 1997. 1.1

[4] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. 2

[5] Neil G Alldrin and David J Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007. 2

[6] Doris Antensteiner, Svorad Štolc, and Thomas Pock. A review of depth and normal fusion algorithms. *Sensors*, 18(2):431, 2018. 2

[7] Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, 2003. 2

[8] Peter N Belhumeur and David J Kriegman. What is the set of images of an object under all possible illumination conditions? *International journal of computer vision*, 28(3):245–260, 1998. 3.3

[9] Alessandro Boccolini, Francesco Tonolini, Jonathan Leach, Robert Henderson, and Daniele Faccio. Imaging inside highly diffusive media with a space and time-resolving single-photon sensor. In *Imaging Systems and Applications*, pages ITu3E–2. Optical Society of America, 2017. 2

[10] Kim L Boyer and Avinash C Kak. Color-encoded structured light for rapid active ranging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1): 14–28, 1987. 2

[11] Francisco Javier Brosed, Juan José Aguilar, David Guillomía, and Jorge Santolaria. 3d geometrical inspection of complex geometry parts using a novel laser triangulation sensor and a robot. *Sensors*, 11(1):90–110, 2011. 1.1

[12] J Herriot Bruning, Donald R Herriott, JE Gallagher, DP Rosenfeld, AD White, and DJ Brangaccio. Digital wavefront measuring interferometer for testing optical surfaces and lenses. *Applied optics*, 13(11):2693–2703, 1974. 2

[13] Dalit Caspi, Nahum Kiryati, and Joseph Shamir. Range imaging with adaptive color structured light. *IEEE Transactions on Pattern analysis and machine intelligence*, 20(5):470–480, 1998. 2

[14] Wenzheng Chen, Parsa Mirdehghan, Sanja Fidler, and Kiriakos N Kutulakos. Auto-tuning structured light by optical stochastic gradient descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5970–5980, 2020. 2

[15] Christine Connolly. X-ray systems for security and industrial inspection. *Sensor Review*, 2008. 1.1

[16] Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)*, 1(1):7–24, 1982. 3.4

[17] Gary D Gilbert and John C Pernicka. Improvement of underwater visibility by reduction of backscatter with a circular polarization technique. *Applied Optics*, 6(4):741–746, 1967. 2

[18] Dan B Goldman, Brian Curless, Aaron Hertzmann, and Steven M Seitz. Shape and spatially-varying brdfs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1060–1071, 2009. 2

[19] Jinwei Gu, Toshihiro Kobayashi, Mohit Gupta, and Shree K Nayar. Multiplexed illumination for scene recovery in the presence of global illumination. In *2011 International Conference on Computer Vision*, pages 691–698. IEEE, 2011. 2

[20] Mohammadul Haque, Avishek Chatterjee, Venu Madhav Govindu, et al. High quality photometric reconstruction using a depth camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2275–2282, 2014. 2

[21] Parameswaran Hariharan, Bozenko F Oreb, and Tomoaki Eiju. Digital phase-shifting interferometry: a simple error-compensating phase calculation algorithm. *Applied optics*, 26(13):2504–2506, 1987. 2

[22] Aaron Hertzmann and Steven M Seitz. Example-based photometric stereo:

Shape reconstruction with general, varying brdfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, 2005. 2

[23] Magnus R Hestenes, Eduard Stiefel, et al. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49(6):409–436, 1952. 3.2.3

[24] Thang Hoang, Bing Pan, Dung Nguyen, and Zhaoyang Wang. Generic gamma correction for accuracy enhancement in fringe-projection profilometry. *Optics letters*, 35(12):1992–1994, 2010. 3

[25] Michael Holroyd, Jason Lawrence, Greg Humphreys, and Todd Zickler. A photometric approach for estimating normals and tangents. *ACM Transactions on Graphics (TOG)*, 27(5):1–9, 2008. 2, 3.1.2, 3.4

[26] Peisen S Huang and Song Zhang. Fast three-step phase-shifting algorithm. *Applied optics*, 45(21):5086–5091, 2006. 2

[27] Peisen S Huang, Qingying J Hu, and Fu-Pen Chiang. Double three-step phase-shifting algorithm. *Applied optics*, 41(22):4503–4509, 2002. 2, 3.2.2

[28] Peisen S Huang, Song Zhang, and Fu-Pen Chiang. Trapezoidal phase-shifting method for three-dimensional shape measurement. *Optical Engineering*, 44(12): 123601, 2005. 2

[29] Zhuo Hui and Aswin C Sankaranarayanan. A dictionary-based approach for estimating shape and spatially-varying reflectance. In *2015 IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2015. 2

[30] Satoshi Ikehata, David Wipf, Yasuyuki Matsushita, and Kiyoharu Aizawa. Robust photometric stereo using sparse regression. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 318–325. IEEE, 2012. 2

[31] Seiji Inokuchi. Range imaging system for 3-d object recognition. *ICPR, 1984*, pages 806–808, 1984. 2

[32] Wenzel Jakob. Mitsuba renderer. `https://www.mitsuba-renderer.org/index_old.html`, 2014. 3.4

[33] Peirong Jia, Jonathan Kofman, and Chad E English. Two-step triangular-pattern phase-shifting method for three-dimensional object-shape measurement. *Optical Engineering*, 46(8):083201, 2007. 2

[34] Mark Johnson. Fiber displacement sensors for metrology and control. *Optical Engineering*, 24(6):961–0, 1985. 1.1

[35] Neel Joshi and David J Kriegman. Shape from varying illumination and viewpoint. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–7. IEEE, 2007. 2

[36] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher

Barsi, Adrian Dorrington, and Ramesh Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Transactions on Graphics (ToG)*, 32(6):1–10, 2013. 2

[37] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 3.2.3

[38] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, volume 7, 2006. 21, 37

[39] Juha Tapio Kostamovaara, Kari E Maatta, and Risto A Myllylae. Pulsed time-of-flight laser range-finding techniques for industrial applications. In *Optics, Illumination, and Image Sensing for Machine Vision VI*, volume 1614, pages 283–295. International Society for Optics and Photonics, 1992. 1.1

[40] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, et al. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 131–144, 2000. 2

[41] Yuan Li, Qing Lin Wang, You Fu Li, De Xu, and Min Tan. On-line visual measurement and inspection of weld bead using structured light. In *2008 IEEE instrumentation and measurement technology conference*, pages 2038–2043. IEEE, 2008. 1.1

[42] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010. 3.3

[43] Feng Lu, Xiaowu Chen, Imari Sato, and Yoichi Sato. Symps: Brdf symmetry guided photometric stereo for shape and light source estimation. *IEEE transactions on pattern analysis and machine intelligence*, 40(1):221–234, 2017. 2

[44] Zheng Lu, Yu-Wing Tai, Fanbo Deng, Moshe Ben-Ezra, and Michael S Brown. A 3d imaging framework based on high-resolution photometric-stereo and low-resolution depth. *International journal of computer vision*, 102(1):18–32, 2013. 2

[45] Satya P Mallick, Todd E Zickler, David J Kriegman, and Peter N Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 619–626. Ieee, 2005. 2

[46] Pablo Marquez-Neila, Luis Baumela, and Luis Alvarez. A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):2–17, 2013. 3, 4.5.2

[47] Parsa Mirdehghan, Wenzheng Chen, and Kiriakos N Kutulakos. Optimal structured light a la carte. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6248–6257, 2018. 2

[48] Daniel Moreno and Gabriel Taubin. Simple, accurate, and robust projector-camera calibration. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pages 464–471. IEEE, 2012. 2

[49] Yasuhiro Mukaigawa, Ramesh Raskar, and Yasushi Yagi. Analysis of scattering light transport in translucent media. *IPSJ Transactions on Computer Vision and Applications*, 3:122–133, 2011. 2

[50] Magued B Nasr, Bahaa EA Saleh, Alexander V Sergienko, and Malvin C Teich. Demonstration of dispersion-canceled quantum-optical coherence tomography. *Physical review letters*, 91(8):083601, 2003. 2

[51] 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. 2

[52] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM transactions on graphics (TOG)*, 24(3):536–543, 2005. 2, 3.2.3, 3.4

[53] 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. 2

[54] Matthew O'Toole, Supreeth Achar, Srinivasa G Narasimhan, and Kiriakos N Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Transactions on Graphics (ToG)*, 34(4):1–13, 2015. 2, 4.2

[55] Sébastien M Popoff, Geoffroy Lerosey, R Carminati, Mathias Fink, Albert Claude Boccara, and Sylvain Gigan. Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media. *Physical review letters*, 104(10):100601, 2010. 2

[56] Jeffrey L Posdamer and MD Altschuler. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1):1–17, 1982. 2

[57] Christof Reinhart. Industrial computer tomography–a universal inspection tool. In *17th world conference on nondestructive testing*, pages 25–28. Citeseer, 2008. 1.1

[58] Christophe Schlick. A survey of shading and reflectance models. In *computer graphics forum*, volume 13, pages 121–131. Wiley Online Library, 1994. 3.1.2

[59] Saša Singer and John Nelder. Nelder-mead algorithm. *Scholarpedia*, 4(7):2928, 2009. 3.1.2

[60] Dilbag Singh and Vijay Kumar. Comprehensive survey on haze removal techniques. *Multimedia Tools and Applications*, 77(8):9595–9620, 2018. 4.1

[61] V Smutny, Václav Hlavác, and Petr Palatka. High precision measurements of small backlit objects in mechanical engineering. In *IECON'98. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (Cat. No. 98CH36200)*, volume 2, pages 1226–1229. IEEE, 1998. 1.1

[62] Zhan Song and Ronald Chung. Determining both surface position and orientation in structured-light-based sensing. *IEEE transactions on pattern analysis and machine intelligence*, 32(10):1770–1780, 2009. 2

[63] Venugopal Srinivasan, Hsin-Chu Liu, and Maurice Halioua. Automated phase-measuring profilometry of 3-d diffuse objects. *Applied optics*, 23(18):3105–3108, 1984. 2

[64] Tatsuki Tahara, Ryota Mori, Yasuhiko Arai, and Yasuhiro Takaki. Four-step phase-shifting digital holography simultaneously sensing dual-wavelength information using a monochromatic image sensor. *Journal of Optics*, 17(12):125707, 2015. 2

[65] Kenichiro Tanaka, Yasuhiro Mukaigawa, Yasuyuki Matsushita, and Yasushi Yagi. Descattering of transmissive observation using parallel high-frequency illumination. In *IEEE international conference on computational photography (ICCP)*, pages 1–8. IEEE, 2013. 2

[66] Kenichiro Tanaka, Yasuhiro Mukaigawa, Hiroyuki Kubo, Yasuyuki Matsushita, and Yasushi Yagi. Recovering inner slices of translucent objects by multi-frequency illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5464–5472, 2015. 2

[67] Tali Treibitz and Yoav Y Schechner. Active polarization descattering. *IEEE transactions on pattern analysis and machine intelligence*, 31(3):385–399, 2008. 2

[68] Ruben Usamentiaga, Yacine Mokhtari, Clemente Ibarra-Castanedo, Matthieu Klein, Marc Genest, and Xavier Maldague. Automated dynamic inspection using active infrared thermography. *IEEE Transactions on Industrial Informatics*, 14 (12):5648–5657, 2018. 1.1

[69] Marcus Martinus Petrus A Vermeulen, PCJN Rosielle, and PHJ Schellekens. Design of a high-precision 3d-coordinate measuring machine. *Cirp Annals*, 47 (1):447–450, 1998. 1.1

[70] Andrew John Wagenmaker, Brian E Moore, and Raj Rao Nadakuditi. Robust photometric stereo via dictionary learning. *IEEE Transactions on Computational Imaging*, 5(2):212–227, 2018. 2

[71] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. *Rendering techniques*, 2007:18th, 2007. 3.1.2

[72] QZ Wang, X Liang, L Wang, PP Ho, and RR Alfano. Fourier spatial filter acts as a temporal gate for light propagating through a turbid medium. *Optics letters*, 20(13):1498–1500, 1995. 2

[73] Yajun Wang, Song Zhang, and James H Oliver. 3d shape measurement technique for multiple rapidly moving objects. *Optics express*, 19(9):8539–8545, 2011. 2

[74] Robert J Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Image understanding systems and industrial applications I*, volume 155, pages 136–143. International Society for Optics and Photonics, 1979. 2

[75] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Asian Conference on Computer Vision*, pages 703–717. Springer, 2010. 2, 3.3

[76] Ying Xiong. A matlab toolbox for photometric stereo. `https://github.com/yxiong/PSBox`, 2014. 1

[77] Qing Zhang, Mao Ye, Ruigang Yang, Yasuyuki Matsushita, Bennett Wilburn, and Huimin Yu. Edge-preserving photometric stereo via depth fusion. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2479. IEEE, 2012. 2

[78] Song Zhang. Flexible 3d shape measurement using projector defocusing: extended measurement range. *Optics letters*, 35(7):934–936, 2010. 2

[79] Zi Zhou, Yang Zhang, and Kai Tang. Sweep scan path planning for efficient freeform surface inspection on five-axis cmm. *Computer-Aided Design*, 77:1–17, 2016. 1.1

[80] Todd Zickler, Satya P Mallick, David J Kriegman, and Peter N Belhumeur. Color subspaces as photometric invariants. *International Journal of Computer Vision*, 79(1):13–30, 2008. 2

[81] Todd E Zickler, Peter N Belhumeur, and David J Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision*, 49(2):215–227, 2002. 2