DOCTORAL PROGRAM IN

INFORMATION ENGINEERING AND COMPUTER SCIENCE

Doctoral candidate

Ali Karami

Cycle	**35**
Thesis	**Image-based 3D metrology of non-collaborative surfaces**
Advisor	**Fabio Remondino** (3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation)
Co-advisor	**Fabio Menna** (3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation)

1. **List of publications**

Published articles

1) Karami, A; Menna, F; Remondino, F. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences; Gottingen Vol. XLIII-B2-2021, Gottingen: Copernicus GmbH. (2021): 519-526. DOI:10.5194/isprs-archives-XLIII-B2-2021-519-2021.

2) Karami, A., Menna, F., Remondino, F. and Varshosaz, M., 2022. Exploiting light directionality for image-based 3d reconstruction of non-collaborative surfaces. The Photogrammetric Record, 37(177), pp.111-138.

3) Karami, A., Battisti, R., Menna, F. and Remondino, F., 2022. 3D digitization of transparent and glass surfaces: state of the art and analysis of some methods. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

4) Karami, A., Menna, F. and Remondino, F., 2022. Combining Photogrammetry and Photometric Stereo to Achieve Precise and Complete 3D Reconstruction. Sensors, 22(21), p.8172.

5) Morelli, L., Karami, A., Menna, F. and Remondino, F., 2022. Orientation of Images with Low Contrast Textures and Transparent Objects. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, pp.77-84.

6) Karami, A., Varshosaz, M., Menna, F., Remondino, F. and Luhmann, T., 2023. Fft-Based Filtering Approach to FUSE Photogrammetry and Photometric Stereo 3d Data. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 14, pp.363-370.

To be submitted:

1) G. Mazzacca, A. Karami, S. Rigon, and F. Remondino, 2023. NeRF FOR HERITAGE 3D RECONSTRUCTION. CIPA conference, Florence, Italy.

2) A. Karami, S. Rigon, and F. Remondino, 2023. A critical analysis of image-based 3d reconstruction with nerf. Journal of Remote Sensing, Special Issue "Photogrammetry Meets AI"

3) Image fusion using photometric stereo lighting system, suggested journal: Measurement

2. Research/study activities

Attended Conferences:

1) ISPRS conference, 2021, Nice, France (Remote)
2) ISPRS conference, 2022, Nice, France (In presence)
3) O3DM workshop, 2022, Würzburg, Germany (In presence)
4) 3D-ARCH conference, conference, 2022, Mantova, Italy, Germany (In presence)

Research abroad:

- Visiting researcher for two months at Institut for Applied Photogrammetry and Geoinformatics (IAPG), Oldenburg, Germany, 2022.
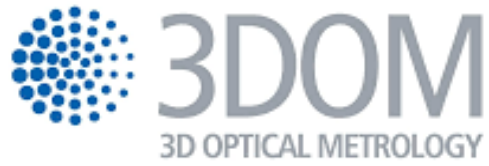
Image-based 3D metrology of non-collaborative surfaces

Ali Karami

3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation (FBK)

Department of Information Engineering and Computer Science, University of Trento

Supervisor (s):

Fabio Remondino

Fabio Menna

Thesis for the Degree of Doctor of Computer Science

January 2023

Abstract

Image-based 3D reconstruction has been employed in industrial metrology for micro measurements and quality control purposes. However, generating a highly-detailed and reliable 3D reconstruction of non-collaborative surfaces (textureless, shiny, and transparent) is still an open issue. This thesis presents various methodologies to successfully generate a highly-detailed and reliable 3D reconstruction of non-collaborative objects using the proposed photometric stereo image acquisition system. The first proposed method employs geometric construction to integrate photogrammetry and photometric stereo in order to overcome each technique's limitations and to leverage each technique's strengths in order to reconstruct an accurate and high-resolution topography of non-collaborative surfaces. This method uses accurate photogrammetric 3D measurements to rectify the global shape deviation of photometric stereo meanwhile uses photometric stereo to recover the high detailed topography of the object. The second method combines the high spatial frequencies of photometric stereo depth map with the low frequencies of photogrammetric depth map in frequency domain to produce accurate low frequencies while retaining high frequencies. For the third approach, we utilize light directionality to improve texture quality by leveraging shade and shadow phenomena using the proposed image-capturing system that employs several light sources for highlighting roughness and microstructures on the surface. And finally, we present two methods that effectively orient images by leveraging the low-contrast textures highlighted on object surfaces (roughness and 3D microstructures) using proper lighting system. Various objects with different

surface characteristics including textureless, reflective, and transparent are used to evaluate

different proposed approaches. To assess the accuracy of each approach, a comprehensive

comparison between reference data and generated 3D points is provided.

Dedication

I would like to dedicate this thesis to all **the brave women of Iran and**

Kurdistan, heroes of the year 2022, who are risking their lives every night

to fight for their freedom, who stand up for their basic rights in a braver

manner than any man in the world. And dedicated to everyone else who died

on the battlefield for their freedom!

Women, Life, freedom

ژن ژیان ئازادی

Acknowledgments

Table of Contents

List of Table

List of Figures

Symbols

LED	Light-Emitting Diode
3D	Three Dimensional
MVPS	Multi-view Photometric Stereo
MVS	Multi-view Stereo
PS	Photometric Stereo
SfM	Structure From Motions
SfP	Shape from Polarisation
SfS	Shape from Silhouette
FFT	Fast Fourier Transform
GLCM	Gray-Level Co-Occurrence Matrix
NCC	Normalized Cross-Correlation
SVD	Singular Value Decomposition
RANSAC	Random Sample Consensus
BRDF	Bidirectional Reflectance Distribution Function
SFSE	Single Fiber Scanning Endoscope
RGB-D	Red, Green, Blue, Depth
CMM	Coordinate-Measuring Machine
IR	Infrared Radiation
LWIR	Long Wavelength Infrared

fps	frames per second
CO2	Carbon Dioxide
LTFtF	Laser Tracking Frame to Frame
CNN	Convolutional Neural Network
6DoF	Six Degrees of Freedom
RMSE	Root Mean Square Error
ISO	International Standards Organization
ReLU	Rectified Linear Unit
SIFT	Scale Invariant Feature Transform
AI	Artificial Intelligence
NN	Nearest Neighbor
NNR	Nearest Neighbor Ratio
GSD	Ground Sample Distance
ICP	Iterative Closest Point
MAE	Mean Absolute Error
STD	Standard Deviation
MRE	Mean Reprojection Error
PCS	Photometric Consistency Score
GT	Ground Truth
CCD	Charge-Coupled Device
RTI	Reflectance Transformation Imaging

CHAPTER I

Introduction

1.1 Background

3D reconstruction in computer vision and photogrammetry is the process of recovering the shape and appearance of a real-world object starting with data acquisition and ending with 3D visualization on a computer. It is an interesting and long-running task in computer vision and photogrammetric communities. There has, for long, been a demand for accurate and reliable 3D measurements in various application fields (Sansoni et al., 2009; Luhmann et al., 2010; Yang et al., 2023). For example, in industrial applications, 3D measurements are used for quality inspection of welds (Rodríguez-Martín et al., 2015), checking the strength of materials (Shmueli et al., 2015), reverse engineering of complex and free-form objects (Carbone et al., 2001), or measuring the 3D dimension of complex surfaces (Hosseininaveh et al., 2015). In almost all of the applications mentioned above, high geometric accuracy, and high-resolution details of the 3D reconstruction along with low cost, portability, and flexibility of the method are required (Luhmann et al., 2019; Karami et al., 2022c).

Generally, existing approaches for 3D reconstruction are classified into contact and non-contact methods (Luhmann et al., 2019). Contact-based methods usually use some physical equipment such as coordinate measuring machines, or calipers to measure the 3D shape of an object. Even though precise geometrical 3D measurements are possible and suited for many applications, they have some drawbacks. For example, the process of acquiring data is extremely time-consuming; sparse 3D data is reconstructed from the object; some parts of the object might remain unmeasured due to object shape complexity and the system's restrictions in measuring; in other situations, such as with antique, soft,

or deformable objects, these methods are not appropriate due to the risk of damaging the object during the measurement process; these equipment are required to be protected from temperature variation and vibration in a controlled-environment.

In contrast, 3D reconstruction is achieved using non-contact methods that can overcome the drawback of contact-based techniques. These approaches, according to the type of sensor used, are divided into active and passive. Active methods project structured light, or lasers on the object and detect them using a camera to generate 3D shape of an object while passive methods rely on other sources of energy (e.g. natural sunlight). Current active approaches rely on costly technologies such as laser scanning arms, structured light systems, or confocal white light; in addition to this, they may be impractical in some cases such as deformable objects. Passive image-based approaches (such as photogrammetry, photometric stereo, shape from shading, shape from texture, shape from specularity, shape from contour, or shape from 2D edge) reconstruct 3D shape of an object from 2D images using different mathematical models.

While active methods have advantages in terms of accuracy and robustness, they are often more expensive, less flexible, and require more complex equipment setups. Passive image-based methods, such as photogrammetry and photometric stereo, are becoming increasingly popular due to their cost-effectiveness, portability, and flexibility in a wide range of applications, including industrial inspection (Barazzetti et al., 2012; Karami et al., 2022c) and quality control (Sansoni et al., 2009; Rodríguez-Martín et al., 2015; Luhmann et al., 2019), where highly detailed micro-topography of surfaces is required.

However, passive image-based methods have limitations, including sensitivity to lighting conditions, and the textural properties of the surface, which can result in failure. Despite these challenges, image-based methods are gaining more attention recently because of advancements in image processing algorithms and hardware capabilities. These advancements have led to improved accuracy, robustness, and performance of passive image-based methods, making them a viable option for 3D reconstruction of non-collaborative objects. Additionally, with the increasing availability and affordability of high-quality cameras and software, image-based methods have become more accessible to researchers and industries.

1.2   Objects with non-collaborative surfaces

The term "non-collaborative" offers a nuanced and specific characterization of objects that present unique obstacles for 3D digitization using active or passive sensors. While other descriptors such as "challenging" or "difficult" may imply a broad category of objects, the term "non-collaborative" is more precise and refers specifically to objects that do not lend themselves well to the 3D digitization process. These non-collaborative objects are typically those that lack texture or contrast, are made of reflective or metallic materials, or are translucent or transparent. For example, reflective or metallic objects can create reflections or shadows that can obscure key features or details. Similarly, translucent or transparent objects may be difficult to capture accurately because the light is not always reflected in the same way. The use of the term "non-collaborative" emphasizes the fact that these objects do not actively participate in the digitization process. They require specialized

approaches and techniques that go beyond standard 3D digitization procedures using active or passive sensors. By using this term, we can accurately describe the specific challenges involved in capturing these objects digitally, highlighting the need for unique and customized solutions.

a) textureless

b) reflective
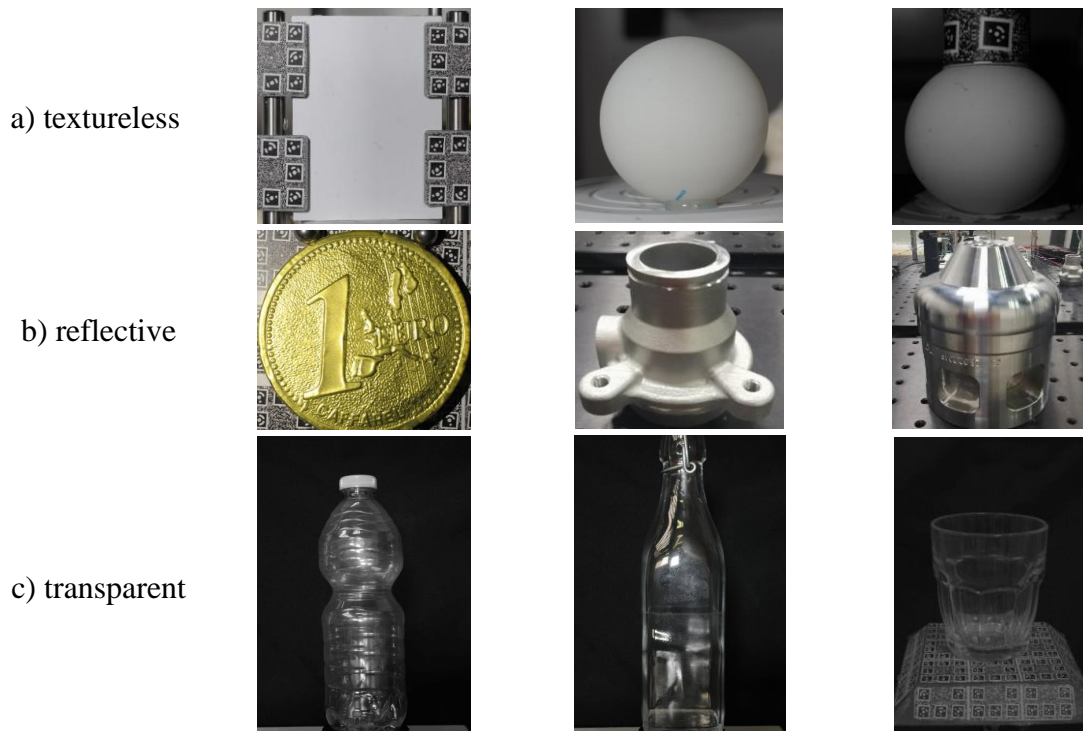
c) transparent

Figure 1-1. Various objects featuring non-collaborative surfaces.

Examples of different objects featuring non-collaborative surfaces, including textureless, metallic, reflective, and transparent.

Figure 1-1 show some example of such non-collaborative objects including textureless surfaces (Figure 1-1a) that cause problems or failures of image-based approaches, glossy and reflective surfaces (Figure 1-1b) that cause problems for all 3D measurement technologies, and transparent objects (Figure 1-1c) that do not diffusely

reflect the incoming light and do not have a texture of their own needed for image matching tasks. Instead, because of refraction and specular reflections, their appearance depends on the object's shape, surrounding background, and lighting conditions with light traveling through the surface, distorting or changing the path of the light in the process.

1.3   Challenges with photogrammetry and photometric stereo

Among image-based 3D reconstruction approaches, photometric stereo and photogrammetry are two established techniques in image-based 3D reconstruction that have gained attention from researchers. These methods are known for their cost-effectiveness, portability, and versatility in various applications, such as industrial inspection and quality control, where a detailed micro-topography of surfaces is necessary.

Photogrammetry can generate a geometrically accurate and dense model of a real-world object from a series of images of an object or a scene taken from various viewpoints under the assumption of known materials and lighting conditions (Karami et al. 2022c; Luhmann et al., 2019). However, it is still challenging to achieve high-accuracy 3D measurement of non-collaborative objects (Figure 1-1) due to the sensitivity of photogrammetry to the textural properties of the surface (Ahmadabadian et al., 2019; Santoši et al., 2019; Karami et al., 2022a). For example, when the surface of an object is featureless or displays repetitive patterns, methods based on feature extraction face difficulties in finding a sufficient number of corresponding image points that are needed for image orientation (Hosseininaveh et al., 2015; Karami et al., 2021). In case of polished and shiny surfaces, such as industrial and metallic components (Figure 1-1b), the incoming

light is subject to the law of reflection and is observed as specular reflection. Such reflections, present in captured images, are undesirable and dense image-matching procedures produce noisy results on high-reflective and poorly textured objects (see Figure 1-2a). In transparent objects, the ability to diffusely reflect light is very limited, and in addition, they are almost textureless. Due to refraction and mirror-like reflections, a part of the surface recorded textures of such objects is not invariant to the camera viewpoint being also dependent on the object's shape, surrounding environment, and illumination conditions.
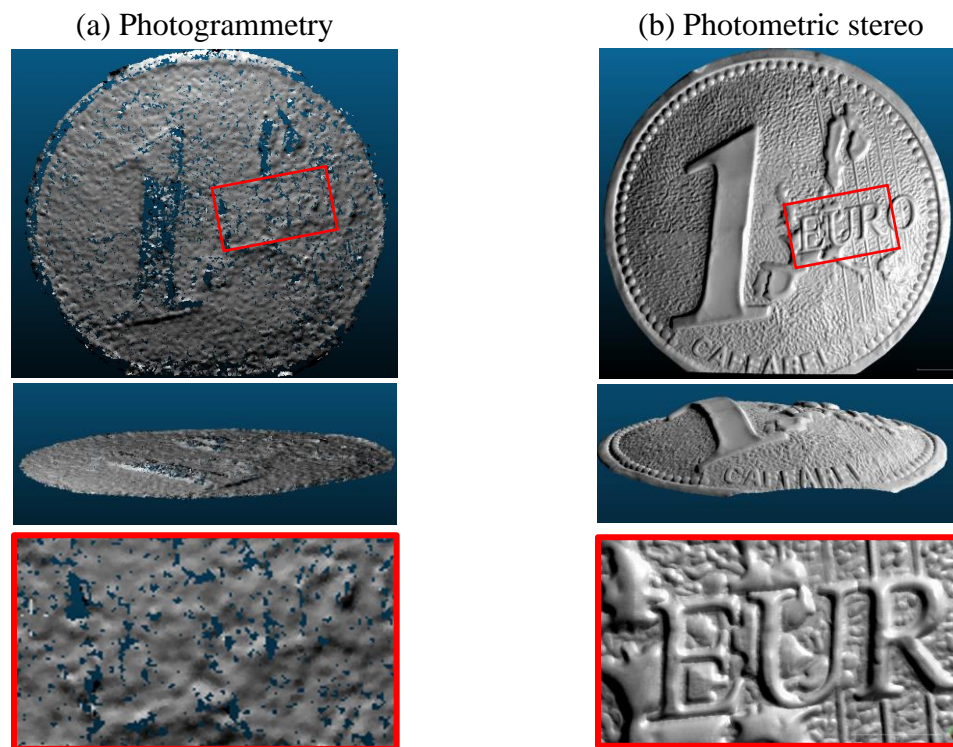
Figure 1-2. Visual comparison between photogrammetry and photometric stereo.

*Visual comparison between photogrammetry and photometric stereo in terms of low and high-frequency information retrieved by the two techniques. (**a**) accurate low-frequency information but noisy 3D details derived with photogrammetry; (**b**) high-details but deformed global shape derived with photometric stereo.*

Therefore, standard image-based procedures like those implemented in SfM applications become ineffective, leading to significant errors and, most frequently, failures in the image matching and orientation process (Hosseininaveh et al., 2015; Wu et al., 2018; Karami et al., 2022c).

Photometric stereo, on the other hand, is an effective method able to retrieve surface normal using a set of images captured under various lighting conditions (Woodham, 1980) and applying the gradient field (Scherr, 2017; Antensteiner et al., 2018; Li et al., 2020) to directly compute object depth from surface normals. This technique can recover a very detailed topography of objects even with texture-less or shiny surfaces (Li et al., 2020; Jiddi et al., 2020). Indeed, as the photometric stereo technique requires images captured under multiple light directions, the problem of specular reflection is partially mitigated. However, a global deformation of the recovered 3D shape is generally present (see Figure 1-2b) due to unfulfilled assumptions and to simplifications made to the mathematical model on how light interacts with the object surface (Shi et al., 2018; Ren et al., 2021; Karami et al., 2022c), in particular:

- The surface of the object should have an ideal diffuse reflection with no shadow and specularities on the surface.

- Light rays arriving at the surface should be parallel to each other.

- Camera uses an orthogonal projection.

Furthermore, 3D data generated using a photometric stereo are produced up to a scale factor, and accurate scaling is not as straightforward as other techniques—such as photogrammetry.

1.4 Aims and objectives of the research

Given the difficulties discussed in Section 1.3, the objectives of this research are as follows:

- Needs for an automatic data acquisition system suitable for data collection and method integration.

- Aim at high precision and detailed 3D reconstruction of challenging objects including textureless, reflective, and refractive objects.

- Aim at different method integration to overcome the constraints of one method by leveraging the strengths of the other.

- Evaluate different methodologies on challenging objects.

1.5 Contributions

The major contributions presented in this thesis are:

- Development of an image acquisition system based on the near-field photometric stereo lighting system suitable for integrating photogrammetry measurements and photometric stereo (Section 3).

- Development of a simple yet effective method for calibrating the geometry of the lighting system and the camera interior and exterior orientation parameters using some coded targets embedded in the scene – also useful to address scaling and transformation issues (Section 3.2).

- A novel algorithm for removing specular reflections and shadows, as well as determining lighting direction and illumination attenuation at each surface point, using the accurate geometry of the lighting system and the object's sparse 3D shape (Section 4.2).

- Development of three alternative approaches that take advantage of photogrammetric 3D measurements to correct the global shape deviation of photometric stereo depth caused by simplified assumptions such as orthogonal projection, perfect diffuse reflection, or unknown error resources (Sections 4.2.7 - 4.2.9).

- Development of an FFT-based filtering approach to fuse the high spatial frequencies of photometric stereo with low frequencies from photogrammetry (Section 4.3).

- Development of a novel procedure that leverages the PS image acquisition system with multiple light sources to highlight roughness and microstructures of non-collaborative surfaces which are not visible under diffuse lighting direction. These roughness are then used as a sort of chiaroscuro texture in image orientation and multi-view stereo (MVS) algorithms to ensure effective matching procedures (Section 4.4).

- Development of five different image-based fusion methods (average, median, albedo, GLCM-based, and deep learning-based) to select out the best grazing-angle regions and fuse them into a single, highly detailed, shadow- and highlight-free

image using the advantage of known geometry of the lighting system and the approximate 3D shape of the object (Section 4.4.4).

- Development of an object-based fusion procedure by combining all the individual point clouds generated at various grazing angles in object space to generate a reliable, accurate, and complete 3D reconstruction of the non-collaborative surface (Section 4.4.5).

- Development of structure-from-motion (SfM) pipeline for image matching and image orientation of refractive objects through leveraging the low contrast textures present on the surface of transparent objects. To take full advantage of the geometrical content of the patches, the normalized cross-correlation (NCC) must be run on the gradient map of the grayscale image since applying NCC on grayscale images of transparent objects is not robust enough (Section 4.5).

1.6 Overview and organization

The rest of the thesis is organized as follows:

- Chapter 2 reviews the related works for 3D reconstruction of non-collaborative objects featuring textureless, reflective, and refractive surfaces.

- In chapter 3, we propose an automatic image acquisition system used for collecting data.

- The various developed and investigated methodologies are introduced for generating 3D reconstruction of non-collaborative objects in Chapter 4.

- Chapter 5 presents the results achieved with the developed solutions for data acquisition, and reports 3D reconstruction of non-collaborative surfaces and essential analyses using the proposed algorithms.

- Finally, conclusions are drawn and presented together with future research plans.

CHAPTER II

Literature

2.1 Introduction

In this chapter, we review the related works for 3D reconstruction of non-collaborative objects featuring textureless, reflective (metallic and shiny objects), and refractive surfaces (transparent and translucent) as presented in Figure 2-1.
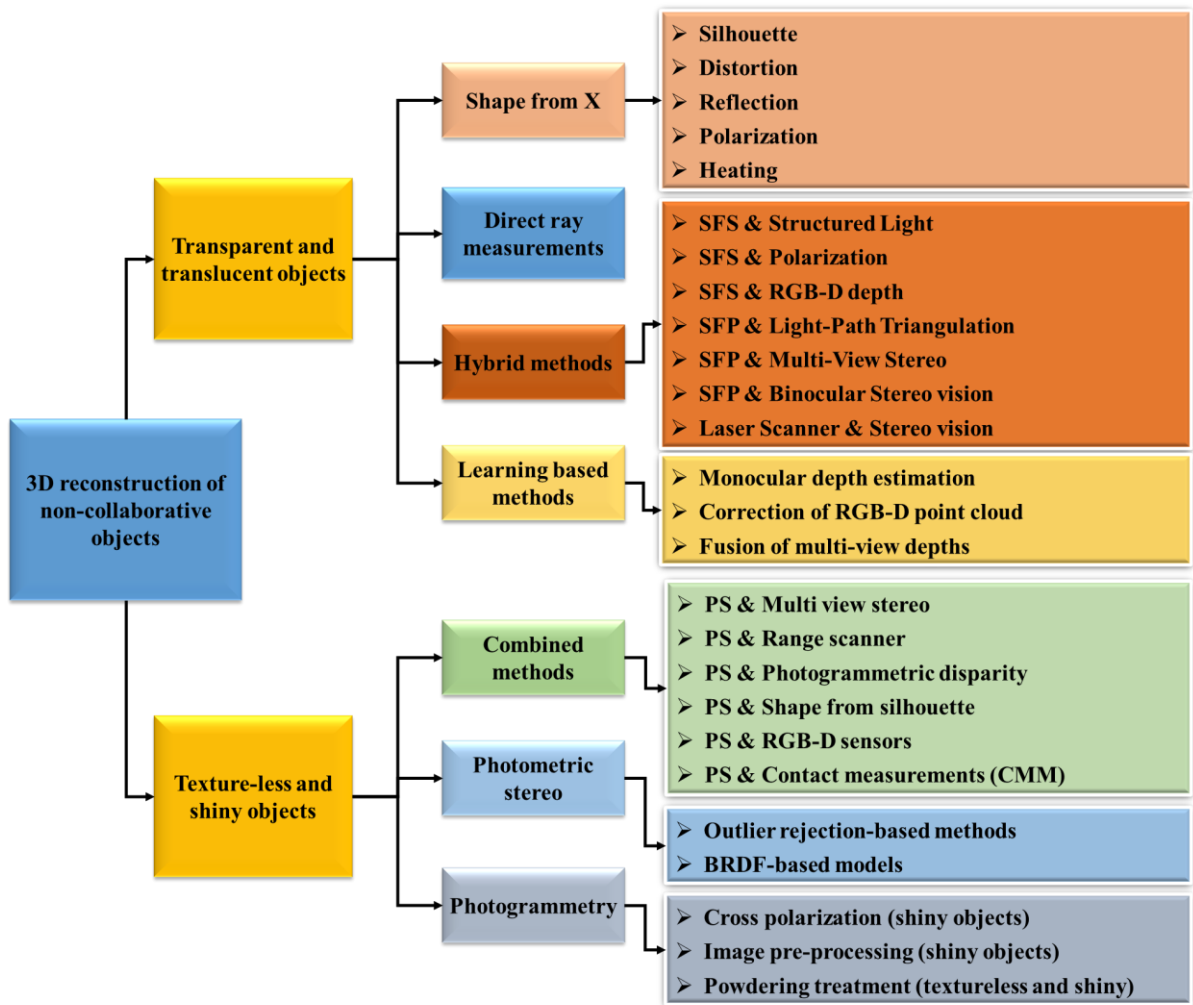
Figure 2-1. General taxonomy of 3D digitization of non-collaborative surfaces.

We have divided non-collaborative objects into two groups based on their surface characteristics and the need for different treatment to reconstruct their 3D models. The first group includes transparent and translucent objects with similar surface properties (i.e., light refraction and scattering) that may cause image distortions. We use four different approaches, including shape-from-X, direct ray measurements, hybrid, and learning-based techniques, to correct these distortions.

The second group comprises textureless and metallic objects that do not allow light to pass through, resulting in specular reflections instead of image distortions. We primarily use three different methods, such as photogrammetry, photometric stereo, and hybrid techniques, to treat this group of objects.

2.2 Texture-less and shiny objects

In this Section, we summarize the research works related to the 3D measurement of non-collaborative surfaces including textureless and shiny objects into three different categories: photogrammetry, photometric stereo, and combined methods.

2.2.1 Photogrammetry

Photogrammetry has historically and widely been regarded as one of the most effective techniques for 3D modeling of well-textured objects. Photogrammetry allows to

15

reconstruct the 3D shape of the object accurately and reliably compared to photometric stereo. However, regions with poorly texture or repetitive patterns are difficult to reconstruct since all reconstruction methods of this kind require matching correspondences in various images (Santoši et al., 2019; Hosseininaveh et al., 2015). Over the years, various photogrammetric methods have been developed to deal with the 3D reconstruction of such non-collaborative objects. In the case of textureless Lambertian objects, several solutions for enhancing the surface texture are suggested with, for example, the projection of known patterns (Menna et al., 2017; Mousavi et al., 2018), random (Hosseininaveh et al., 2015; Ahmadabadian et al., 2019) or synthetic (Santoši et al., 2019; Hafeez et al., 2022) ones onto the object surface. For example, Ahmadabadian et al. (2019) established a relatively inexpensive automated image acquisition system used for 3D modeling of textureless objects that works by projecting a random pattern onto the examined object. Menna et al. (2017) have developed a similar automatic workflow based on the know pattern projection such as structured-light pattern for 3D digitization of heritage artifacts. Methods based on the pattern projection improve the surface texture and, as a result, the accuracy of the final 3D reconstruction when dealing with only Lambertian surfaces (Mousavi et al., 2018). However, these methods have problems when dealing with highly reflective surfaces with heavy specular reflection or interreflection (Ahmadabadian et al., 2019; Mousavi et al., 2018). In the case of reflective objects, cross polarisation (Nicolae et al., 2014; Menna et

16

al., 2016) and image pre-processing approaches (Wallis, 1976; shih et al., 2015; Gaiani et al., 2016; Calantropio et al., 2020) have also been employed to decrease specular reflections. However, these procedures may smooth off surface roughness or vary the texture from one view to the next, affecting negatively the results (Karami et al., 2022b). Another common approach is to spray the surface with a thin layer of white or colored powder (Lin et al., 2017; Palousek et al., 2015; Pereira et al., 2019) can also be used as a common solution. However, powdering the object surface might be impractical when the surface topography of an object is needed at high spatial resolution since the added layer increases the total object volume and can smooth out local information. Besides, surface treatment is impossible in the case of delicate cultural heritage assets, or real-time 3D surface inspection (Karami et al., 2022a; Lin et al., 2017; Palousek et al., 2015; Pereira et al., 2019).

2.2.2 Photometric stereo

Photometric stereo is a technique for estimating an object's surface normal using illumination changes, which was first proposed by Woodham (1980). Over the years, many techniques (Rostami et al., 2012; Antensteiner et al., 2018; Li et al., 2020) have been developed to extract the geometry of objects from surface normals using the gradient field. However, the classical photometric stereo approaches work with perfectly diffuse

17

(Lambertian) surfaces, which is often an improper assumption for many objects such as metallic, glossy, and shiny. Therefore, the performance of such techniques degrades on real-world objects, which frequently exhibit non-Lambertian reflectance such as interreflection and specular reflection (Shi et al., 2018; Li et al., 2020; Ren et al., 2021). To address these issues, different approaches have been developed over the years. The first group of approaches classifies and removes the specular highlights when dealing with non-Lambertian surfaces. For example, earlier approaches (Solomon, and Ikeuchi, 1996; Barsky, and Petrou, 2003) used three illumination directions out of four at each surface point in which the surface seems more Lambertian to approximate the direction of the surface normal. Following this, several algorithms were proposed based on RANSAC (Sunkavalli et al., 2010), graph cuts (Quéau et al., 2017), maximum-likelihood estimation (Peng et al., 2017), using robust SVD (Cho et al., 2018), or Markov random field (Chandraker et al., 2007) to extract Lambertian images in a more stable form. However, more input images are also needed for statistical analysis. Moreover, their output negatively affects complex objects with interreflection and speculative reflection due to the large number of outliers in non-Lambertian surfaces(Chen et al., 2018; Shi et al., 2018).

Instead of discarding specular reflection as outliers, the second group of investigations modeled the behavior of the light using a nonlinear analytic Bidirectional

18

Reflectance Distribution Function (BRDF[1]). In this regard, different BRDF models were developed such as the Ward model (Chung, and Jia, 2008), the Torrance-Sparrow model (Georghiades, 2003), the specular spike (Yeung et al., 2014), bivariate BRDF (Otani et al., 2019), symmetry-based approach (Lu et al., 2017), spatially-varying BRDF (Boss et al., 2020) and etc. Unlike the previous group, they have the benefit of using more available data. The downside to such methods is that analytical models vary considerably from one object to the next and each is confined to a specific material class. Such approaches also require a complex case-by-case analysis of different content classes in theory (Chen et al., 2018; Shi et al., 2018).

Photometric stereo-based methods, unlike photogrammetry techniques, can reconstruct a very detailed surface's topography even with non-collaborative objects (Zheng et al., 2019; Karami et al. 2021). However, owing to some mathematical assumptions, such as parallel light direction and orthogonal projection of the sensor, global deformation of the reconstructed 3D shape typically exists (Fan et al., 2017; Shi et al., 2018; Karami et al. 2021; Ren et al., 2021). The global shape deviation can vary depending on the surface properties and dimensions of the object. For instance, the generated 3D

---

[1] . BRDF is a mathematical model that describes the way light is reflected off a surface in different directions. It takes into account the incoming light direction, the surface normal, and the viewing direction, among other factors, to calculate the amount of light that is reflected in each direction. BRDF is typically used in physically-based rendering to generate realistic images of 3D models under different lighting conditions.

reconstruction can be deformed with a maximum shape deviation of about 13mm on a Lambertian flat object with 340*270mm dimensions when ignoring the assumptions mentioned above (Fan et al., 2017; Karami et al. 2021).

2.2.3 Combined methods

Various researchers combined photometric stereo with other techniques such as structured light or photogrammetry. In the developed methods, high-frequency spatial information is recovered from photometric stereo, whereas the other techniques are applied to retrieve low-frequency information. For example, Smithwick and Seibel (2002) proposed a Single Fiber Scanning Endoscope (SFSE) system for generating dense range maps and 3D measurements based on the fusion of photogrammetric and photometric stereo methods, providing precise volume measurements for dosage, risk estimate, and healing progress analyses. Nehab et al. (2005) combined 3D reconstruction generated from a range scanner with photometric normals to improve the accuracy and level of detail. Hernandez et al. (2008) used a multi-view geometric constraint from shape from silhouette (SFS) to mitigate photometric stereo's low-frequency surface distortion. Although this method is simple and flexible, it works only with particular parametric BRDF models (Kaya et al., 2020a). Several works (Peng et al., 2017; Zollhöfer et al., 2018; Bylow et al., 2019) combined photometric stereo with RGB-D sensors to derive the 3D details from

Photometric stereo while improving the low-frequency information using RGB-D data. Later, Park et al. (2013, 2016) then after suggested an uncalibrated multi-view photometric stereo (MVPS) approach for recovering precise 3D reconstruction of the object utilizing a coarse mesh with a 2D displacement map. However, the approach is unable to reconstruct objects with a wide range of surface reflectance characteristics as well as textureless surfaces (Li et al., 2020). Logothetis et al. (2019) Proposed a new MVPS approach capable of modeling objects with complex geometry where occlusions and/ or cast shadows may occur. More recently, Ren et al. (2020, 2021) integrated photometric stereo with sparse 3D points generated using contact measurements (CMM) to correct the global distortion caused by photometric stereo. The use of expensive technology restricts the method to special laboratories and projects with particular metrological demands, despite the fact that these systems may achieve high precision performances. Li et al. (2020) developed an MVPS approach which uses a sparse 3D point to improve the geometry of the depth map generated by photometric stereo. However, this procedure includes explicit geometric modeling stages such as multi-view depth propagation, iso-depth contour estimation, and/or tracing contours, which must be processed and completed properly in order to obtain a 3D reconstruction of the surface making it more difficult, time-consuming, and challenging. Furthermore, they used a turntable to rotate the object while keeping the camera and light sources fixed in order to capture multi-view images, which means that

21

the light sources are not constant from one view to the other. This could change the object texture from one view to the other resulting in noise or false matching during the image orientation and dense matching process. Recently, a few works have investigated the use of different learning-based approaches (Kaya et al., 2022a; Kaya et al., 2022b) to fuse photometric stereo and MVS for effectively utilizing their complementary strengths. Although these approaches are simple and easy to use, they are much less precise than traditional integration methods, making them unsuited for industrial applications where 3D measurement precision and reliability are required. Furthermore, training such algorithms necessitates large datasets labeled for a unique object type, making generalization to real-world objects problematic.

2.3 Transparent objects

In this Section, we provide an overview of research works related to the 3D measurement of transparent surfaces summarizing them into four different categories: shape-from-X, direct ray measurements, hybrid, and learning-based approaches.

2.3.1 Shape from X

Several approaches known as Shape from X techniques have been developed for extracting shape information from 2D images, where X could be distortion, Silhouette, reflection, polarization, heating, and so on.

Shape from distortion, also known as **Deflectometry**, is one of the earliest methods specifically developed for transparent objects. This technique recovers the 3D shape of an object by analyzing the distortion of a known pattern placed behind or near the surface. This approach has been investigated for long to reconstruct either mirror-like surfaces (Tarini et al., 2005), liquids (Murase, 1990; Gao et al., 2022), or solid refractive surfaces (Ben-Ezra and Nayar, 2003; Wetzstein et al., 2011; Tanaka et al., 2016; Kim et al., 2017). The 3D reconstruction of refractive surfaces is more complex than the corresponding specular, or textureless surfaces because the ray path depends on the refractive index in addition to the dependence on the surface normal (Wu et al., 2018; Lyu et al., 2020). These approaches are also limited to the recovery of a single refractive surface or the reconstruction of parametric surface with simple geometry and therefore are not generalizable if not with approximation to a wider range of object categories (Wu et al., 2018; Lyu et al., 2020).

Shape from Silhouette (SFS) is a well-known 3D reconstruction method applied to a wider range of object categories. This method reconstructs the 3D shape of an object

using a sequence of images taken from different views, where the silhouette of the object is the sole relevant feature of the image. Depending on the geometric projection of the imaging system (e.g.: telecentric, central perspective) the silhouette of the object at each station (image) can be seen as the base of a prismatic /conic volume in three-dimensional space. The silhouette itself represents the locus of tangent points on the straight line departing from the perspective center of the camera (for a central perspective). By intersecting the pyramidal volumes, which is also known as Space Carving, a 3D reconstruction of an object can be generated. This method was first presented by Baumgart in 1974. Since then, various versions of the SFS have been proposed. For example, Martin and Aggarwal (1983) used volumetric descriptions to represent the reconstructed shape. Following this, Potmesil (1987) used an octree data structure to speed up the 3D reconstruction process. Szeliski (1993) built a non-invasive 3D digitizer using a turntable and a single camera with SFS as the reconstruction method.

SFS can recover the 3D shape of an object regardless of the object's property and shape as long as the region of the object in each image is distinguishable from the background (Karami et al., 2022a). However, the accuracy of SFS is directly depending on the silhouette boundary binarization, which can be done using automated or user-defined global thresholding of an image. In many cases, it might be difficult to determine the optimum threshold for distinguishing transparent objects from the background. As a result,

the silhouette of an object may be reduced or increased, making the resulting 3D model smaller or larger than the real size of the object or making it noisier. Moreover, another primary issue with SFS is that concavities on an object's surface remain unseen, finding it unsuitable for reconstructing the inside of a hole or concave areas (Karami et al., 2022a). To deal with this issue, Zuo et al. (2015) incorporate internal occluding contours into traditional SFS methods to recover the concavities on an object's surface. Wu et al. (2018) and Lyu et al. (2020) started with an initial 3D shape reconstruction generated from traditional SFS, and then gradually optimizes the model.

Shape from reflection/refraction is also another approach introduced for the first time by Morris and Kutulakos (2007) to recover the 3D shape of transparent objects. This approach usually describes the behavior of rays as they pass through a refractive object by controlling the background behind the refractive object itself (Morris and Kutulakos, 2007; Yeung et al. 2015; Han et al., 2021).

However, this method may be challenging and inefficient when it comes to collecting data. Moreover, it is necessary to manually rotate a spotlight around the hemisphere to illuminate the object and a reference sphere from various angles. Following a similar idea, Yeung et al. (2015) used a more convenient data collection method to obtain the specular reflection information on the surface of a transparent object and applies the graph cut theory to recover and optimize the normal vectors, consequently the depth map.

25

Although the results are insufficiently precise for industrial inspection, they are promising for 3D computer graphics animation.

Iwabuchi et al. (2011) also presented a similar method based on inverse ray-tracing. This method uses multiple CCD cameras placed around a transparent object with simple geometry and can recover the shape and refraction index of the object. Chari and Sturm, (2013) proposed a method that combines both geometric and radiometric information to do reconstruction. The position and direction for each light-path were recovered and combined with light radiance at the beginning and end of each light-path. More recently, Han et al. (2021) employed a single camera that was set in place with a refractive object in front of a checkerboard background. The approach required two images with the background pattern placed in two different known locations. However, the approach required a change in refractive index, necessitating immersion of the object in water, which is a significant disadvantage for industrial purposes.

Shape from Polarization (SFP) Miyazaki et al., (2002), Huynh et al. (2010), and Sun et al. (2020) recover the 3D shape of an object from polarization information of the reflected light. The basic principle is that after capturing the polarization information such as the intensity, degree of polarization, and polarization phase angle, the surface normal can be recovered by analyzing the relationship between the surface normal and the polarization image formation model. This method has been applied on different object

types with various reflection properties such as dielectrics (Huynh et al., 2010), black (Miyazaki et al., 2016), and transparent (Miyazaki et al., 2002; Huynh et al., 2010; Sun et al., 2020) objects. This method is also quite robust and stable in different lighting conditions such as indoors, outdoors, or under patterned illumination as long as incident light is unpolarized (Durou et al., 2020). These methods calculate surface normals, which must afterward be converted into a height map. However, the results are highly vulnerable to noise since they depend solely on the weak shape cue supplied by polarization and do not ensure integrability (Durou et al., 2020). The ambiguity in polarization analysis is also one of the main issues for this approach. To resolve the azimuth and zenith angle ambiguity, for example, Miyazaki et al. (2002) used the polarization degree in the far-infrared wavelength for estimating the surface orientation instead of the visible wavelength. Stolz et al. (2012) proposed a multispectral method for determining the optimal zenith angle. More recently, ambiguities in this approach are adjusted by combining with other approaches in which rough geometric information is provided such as Multi-View Stereo (Miyazaki et al., 2004), light-path triangulation (Xu et al., 2017), etc. (Durou et al., 2020; Karami et al., 2022a).

Shape from heating is another technique for 3D reconstruction of transparent objects (Eren et al., 2009) that, unlike the previously described approaches, ignores the refractive properties of the object. Laser range scanning of transparent objects is possible

using an IR laser rather than visible light since long-wave and thermal infrared spectrum is not refracted by glass. This technique is based on the principle of infrared thermal imaging, in which the infrared source heats up the object, and then the IR-sensitive sensor detects and records the geometric surface information of the object. Aubreton et al. (2013) also demonstrated a very similar approach for high-specular objects utilizing high-power lasers. Since these approaches utilized single laser spots as activating light sources, their measurement areas and acquisition speed are restricted owing to the time required for scanning. There are additional limitations in spatial resolution and precision because of the size of the laser dots. To overcome these restrictions, Brahm et al. (2016) developed a stereo-vision configuration consisting of two uncooled long-wave infrared (LWIR) cameras to detect the emitted heat radiation from an object induced by a pattern projection unit generated by a CO_2 laser. More recently, Landmann et al. (2019) demonstrated real-time 3D thermographs with a 30-frames per second frame rate (fps). This technique is well suited to applications where the geometry or temperature distribution of the objects is rapidly changing. Landmann et al. (2021) developed a simplified and robust projection approach based on a focused single thermal fringe that can rapidly scan across the object's surface. Higher intensities were obtained using such focused single thermal fringe compared to multi-fringe projection, which increased acquisition speed while improving measurement accuracy.

28

2.3.2 Direct-ray measurements

Direct ray measurement techniques, which detect light rays directly, have for long been utilized for refractive surface 3D reconstruction. Kutulakos et al. (2008) published foundational work on measuring the geometry of refractive objects using light-ray correspondences. By mapping the light rays which reach and depart from the object, the geometry of transparent objects characterized by depths and surface normal can be determined. As shown in Figure 2-2, The projection of a point is defined by the 3D path(s) that light would take to reach the camera, given an arbitrary 3D point p, a known viewpoint c, and a known image plane. As expressed by Kutulakos et al. (2008), refractive surface reconstruction problems are expressed as N-K-M triangulation, where N represents view-points required for reconstruction, K represents refractive surface points on a piecewise linear path, and M represents the number of calibrated reference points along the ray exiting the refractive object.
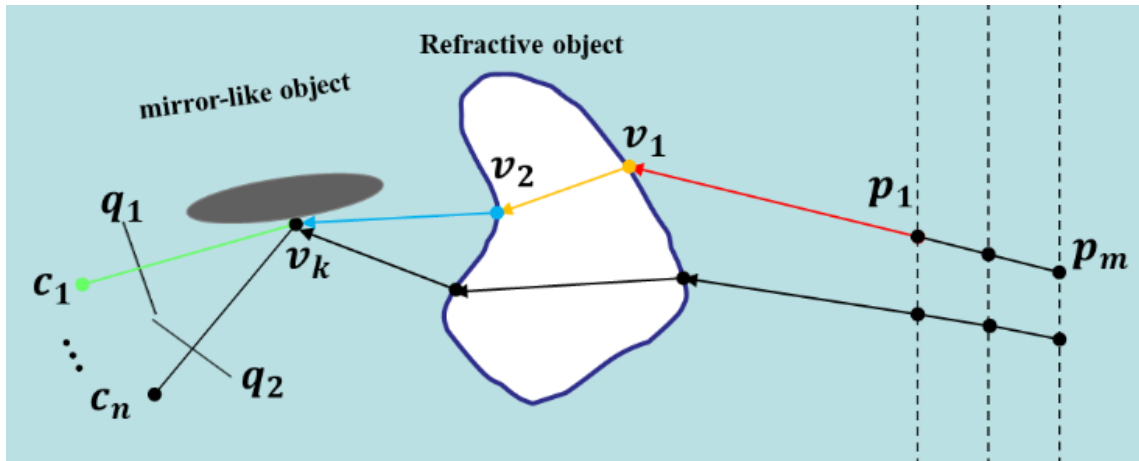
Figure 2-2. The geometry of N-K-M triangulation.

The geometry of N-K-M triangulation expressed by Kutulakos et al. (2008). To reach point q on the image plane, the light path from p crosses three surfaces, including refractive and mirror-like ones, passing from three vertices, v1, v2, and v3, which form four segments. The objective of light-path triangulation is to estimate the normals and coordinates of the vertices using the known coordinates of c, q, and p.

However, methods based on light-path triangulation are known to have collinearity ambiguities as the 3D surface point can be located anywhere along the optical ray that passes through the pixel. To remove the ambiguity, Tsai et al. (2015) assumed that the light rays are refracted twice. They recovered the geometry of a transparent object with only one monocular image using a monitor controlling the background image without even needing to immerse the object in the water.

Morris and Kutulakos (2011) employed stereo/multiple cameras to record the refractive surface, relying on a cross-view normal consistency constraint: the normals computed using the pixel-point correspondences obtained from multiple viewpoints must be consistent. Alternatively, some studies have been conducted to estimate ray-ray correspondences utilizing specific devices such as Bokode (Ye et al., 2012) and light field probes (Wetzstein et al., 2011; Tsai et al., 2021) by capturing the incident rays released from the background and the exiting rays traveling to the camera. Although 3D results appear to be highly promising, the high cost of such devices is an important downside. In addition, one of the main common shortcomings of the aforementioned approaches is that they provide only normals but noisy depths. To provide the boundary condition for the integration of normal, they need to presume a planer surface near the boundary (Ye et al., 2012; Karami et al., 2022a) or approximate the border using noisy depths (Morris and Kutulakos, 2011; Wetzstein et al., 2011). To address the restrictions mentioned above, Qian et al. (2017) propose a position-normal consistency based on a global optimization method to restore depth maps of the surface from front and back. Similarly, Kim et al. (2017) proposed a method based on optimizing the object's shape and refractive index to minimize the disparity between observed and simulated transmission/refraction rays traveling through an object. It cannot, however, be applied to any non-symmetric objects. Following that, Wu et al. (2018) expanded this technique and provided the non-intrusive method to

reconstruct the whole geometry of a transparent object; nevertheless, the results are always over-smoothed due to their independent optimization and multi-view fusion of recovered point clouds. Lately, Lyu et al. (2020) expanded this work by optimizing directly the surface mesh generated from the SFS method using differentiable rendering algorithms. However, these approaches rely on feature correspondence across several views to discover similar features for triangulation, requiring more assumptions and constraints making it insufficient for actual industrial applications that must struggle with a wide range of circumstances or environments.

2.3.3 Hybrid approaches

This group of methods includes combinations of different approaches. The primary goal of combining two techniques is to overcome the constraints of one method by leveraging the strengths of the other, allowing complete and precise 3D reconstruction of optically non-cooperative objects to be generated. For instance, SFS is considered a more suitable and practical approach to reconstruct the 3D shape of transparent objects regardless of object's property and shape. However, the concavities on an object's surface remain unseen. Therefore, some works (Kampel et al., 2002; Tosovic, 2002) have been conducted to correct the problem of SFS by combining a structured light method.

Narayan et al. (2015) merged the silhouette information and depth images on the 2D image domain, which can improve 3D reconstruction for concave and transparent objects with interactive segmentation. Ji et al. (2017) also combined silhouette information and depth from an RGB-D sensor to retrieve the missing surface of transparent objects. First, they seek the 3D region from multiple views that includes the transparent object using incorrect depth led by transparent materials. The 3D shape was then retrieved inside these noisy areas using SFS technology.

Another solution developed to deal with transparent surfaces is to combine SFP with other approaches such as light-path triangulation (Xu et al., 2017), conventional raytracing (Miyazaki et al., 2007), and Multi-View Stereo (Miyazaki et al., 2004). For instance, Miyazaki et al. (2007) developed a polarization raytracing approach, which combines traditional raytracing (calculates the path of light rays) with SFP (calculates the polarization state of the light). Starting with an initial shape of the transparent object, by modifying the shape, the difference between the input polarization data and the rendered polarization data obtained by polarization raytracing was minimized.

More recently, He et al., (2022) developed a pipeline based on the fusion of the laser tracking frame to frame (LTFtF) method and stereo vision to distinguish and extract the reflected laser lines on the front surface from several laser reflection candidates caused by the refraction of the transparent objects.

33

2.3.4 Learning-based methods

Recently, many researchers have used (machine or deep) learning-based approaches to solve the problem of measuring 3D transparent objects. These approaches could be categorized into three groups as follows.

2.3.4.1 **Multi-view 3D reconstruction**

Li et al. (2020) suggested a physically-based network for generating the 3D geometry of transparent objects using multiple images acquired from different viewpoints while also taking into account light transport patterns. More similar to Lyu et al. (2020), this method (Li et al., 2020) optimizes surface normals corresponding to a back-projected ray from both sides of the object using an in-network differentiable rendering layer, given the visual hull construction as an initial 3D reconstruction. Despite the fact that their method is less restrictive than previous ones (Wu et al., 2018; Lyu et al., 2020) that utilized multi-view images, it still requires the environment map and the object's refraction index. It is also difficult to be used in real-time applications because of the time-consuming optimization procedure. Furthermore, these data-driven algorithms rely on training synthetic images since getting a significant quantity of real image training data is difficult (Lyu et al., 2020).

Recently, Mildenhall et al. (2020) proposed NeRF (Neural Radiance Fields) which is a 3D scene representation technique for implicit 3D reconstruction. The original NeRF method uses a neural network to learn a representation of the 3D shape of an object from 2D images. Although NeRF provides an alternative solution for 3D reconstruction of transparent objects compared to traditional photogrammetry methods and can produce promising results in situations where photogrammetry may fail to deliver accurate results, it still faces several limitations (Zhang et al., 2021; Barron et al., 2022; Guo et al., 2022; Yu et al., 2022). Some of the main issues from a 3D metrological perspective that need to be considered include the 3D mesh resolution, requiring significant amounts of computing power and memory, requiring a large number of input images with small baselines. However, in recent years, researchers have proposed several modifications and extensions to the original NeRF method to improve its performance for various scenarios. For exmaples, researchers have focused on improving the resolution of the generated mesh in different way including model acceleration (Müller et al., 2022), compression (Chen et al., 2022), and relighting (Verbin et al., 2022). Some works (Jain et al., 2021; Yu et al., 2022; Niemeyer et al., 2022) have aimed to reduce the number of input images. To improve the accuracy of 3D reconstruction in the presence of noise, previous studies have also incorporated various priors including semantic similarity (Jain et al., 2021), depth

smoothness (Niemeyer et al., 2022), surface smoothness (Zhang et al., 2021), Manhattan world assumptions (Guo et al., 2022), and monocular geometric priors (Yu et al., 2022).

2.3.4.2 **Depth completion (from partial RGB-D depths)**

These approaches use different learning-based methods to fill in missing depths (where transparent objects are) acquired with an RGB-D sensor (Figure 2-3). Sajjan et al. (2020) presented a deep learning approach (named ClearGrasp) for predicting the 3D geometry of transparent objects partly surveyed with an RGB-D sensor. Deep networks are used to identify masks, occlusion borders, and surface normals given RGB images, and then the initial depth is optimized using the network predictions. The optimization, however, needs transparent objects having interaction boundaries with non-transparent objects. Otherwise, the depth of the transparent region remains unpredictable. Figure 2-3 shows an example of a depth completion using the method of Sajjan et al. (2020): the missing parts of the scene (where both transparent objects are located) are predicted and the new point cloud is more complete.

Zhu et al. (2021) proposed another learning-based technique which uses a local implicit neural representation built on ray-voxel pairs that can generalize to unseen objects and fill in missing depth on given noisy depth maps.

a) Input Image   b) Input incomplete RGB-D depth   c) Depth completion
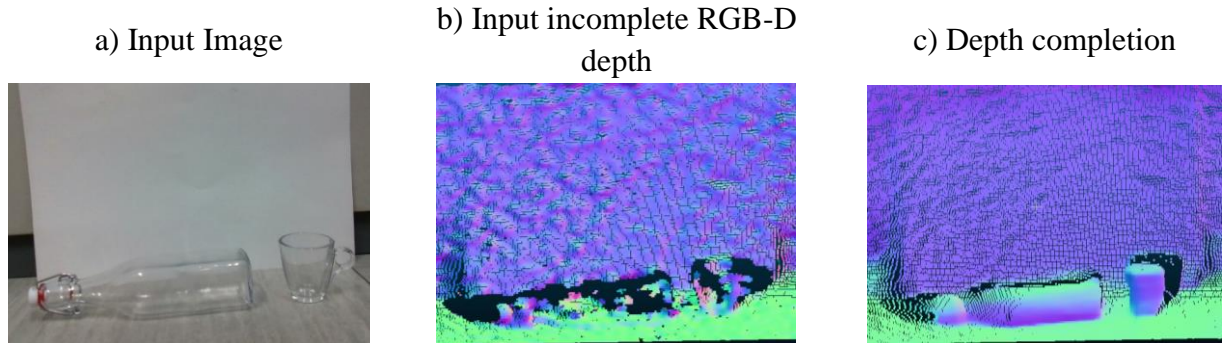
Figure 2-3. Depth completion using a learning-based method

Depth completion using a learning-based method (Sajjan et al., 2020): a depth map of a scene with two transparent objects (glass bottle and tea cup) placed at the scene. Given an RGB input image (a) with an uncompleted depth map (b), the missing areas from input depth were predicted (c).

2.3.4.3 **Monocular shape prediction**

This group of approaches requires only a single image as input in order to predict the 3D shape of transparent objects. Stets et al. (2019) proposed a deep convolutional neural network (CNN) method for determining depths and normals of a transparent object using a single image obtained under an arbitrary environment map. More recently, Eppel et al. (2022) presented a method for predicting 3D points of transparent objects straight from an image taken from unknown source using an advanced neural net that is independent of camera parameters. In this method, each pixel in the predicted map is assigned with the X, Y, Z coordinates of a point rather than the distances to that point. To

train the net, 50k transparent container images containing 13k different objects, 500 different environments, and 1450 material textures were utilized. A total of 104 real-world transparent images of various containers with depth maps were also utilized. Instead of using absolute XYZ coordinates to calculate the training loss, the distance between pairs of points inside the 3D model was utilized, making the loss function translation invariant.

Unlike previous methods, this approach does not require camera parameters and can work with images from unknown cameras. The method was designed for specific manipulation applications of transparent chemical bins but with specific re-training operations, it could be generalized to other objects.

a) bottle of water   b) glass bottle   c) transparent teacup
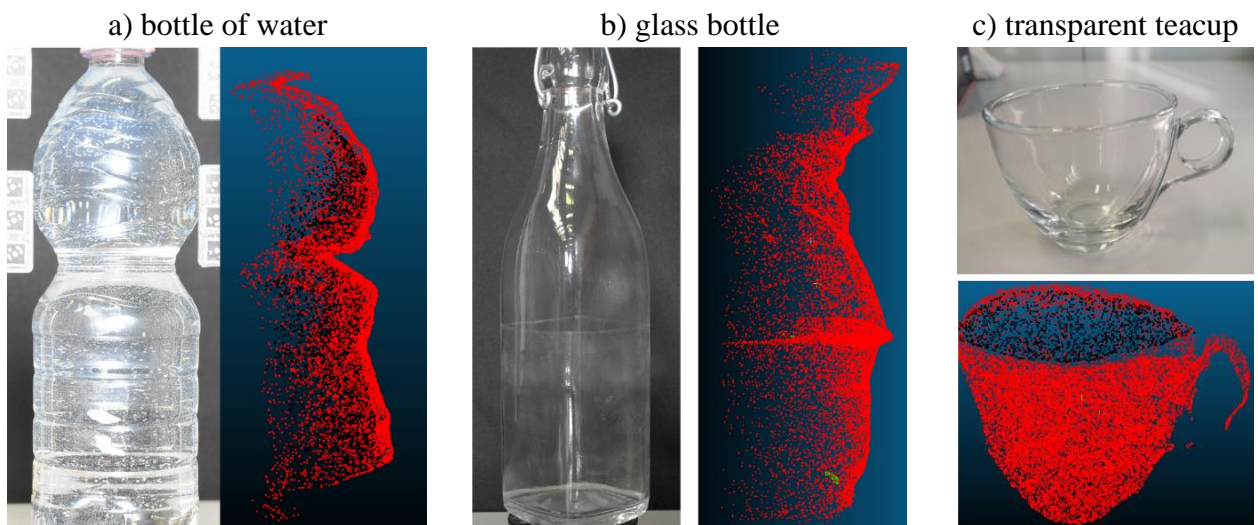
Figure 2-4. Monocular shape prediction

Learning-based 3D reconstruction of three transparent objects (bottle of water, glass bottle, and teacup) from a single image based on Eppel et al. (2022).

Figure 2-4 shows some results obtained using the method presented in Eppel et al. (2022). It can be seen that the predicted 3D shape is only an approximate 3D shape with also anisotropic scaling issues remaining unsolved.

2.4 Summary

In this chapter, we presented a general overview of 3D digitization methods for non-cooperative objects featuring textureless, reflection, and refraction. First, we reviewed the related investigations for 3D reconstruction of textureless and reflection surfaces using photogrammetry, photometric stereo, and the combined methods. Then, the most relevant research works for 3D reconstruction of transparent objects were reviewed, summarizing them into four categories, including shape-from-X, direct ray measurements, hybrid, and learning-based.

CHAPTER III

Data capturing

3.1 Introduction

In this chapter, we aim to propose our automatic image acquisition system which is used for collecting data suitable for the integration of both photogrammetry and photometric stereo approaches. The general overview of the proposed image acquisition system with its calibration steps is summarized in Figure 3-1. In order to prepare for the data processing that will be discussed in Chapter 4, an automatic image acquisition system is developed to capture multiple images under varying illuminations and from different camera stations (camera positions). To this end, two image acquisition systems (Single and multi-synchronized cameras) along with their system calibration process are presented.

3.2 Proposed data acquisition system

In this section, two automatic and semi-automatic image acquisition systems based on the near-field photometric stereo lighting system are presented, which are suitable for integrating photogrammetry measurements and photometric stereo. Table 3-1 shows a summary of the specified features for each system.
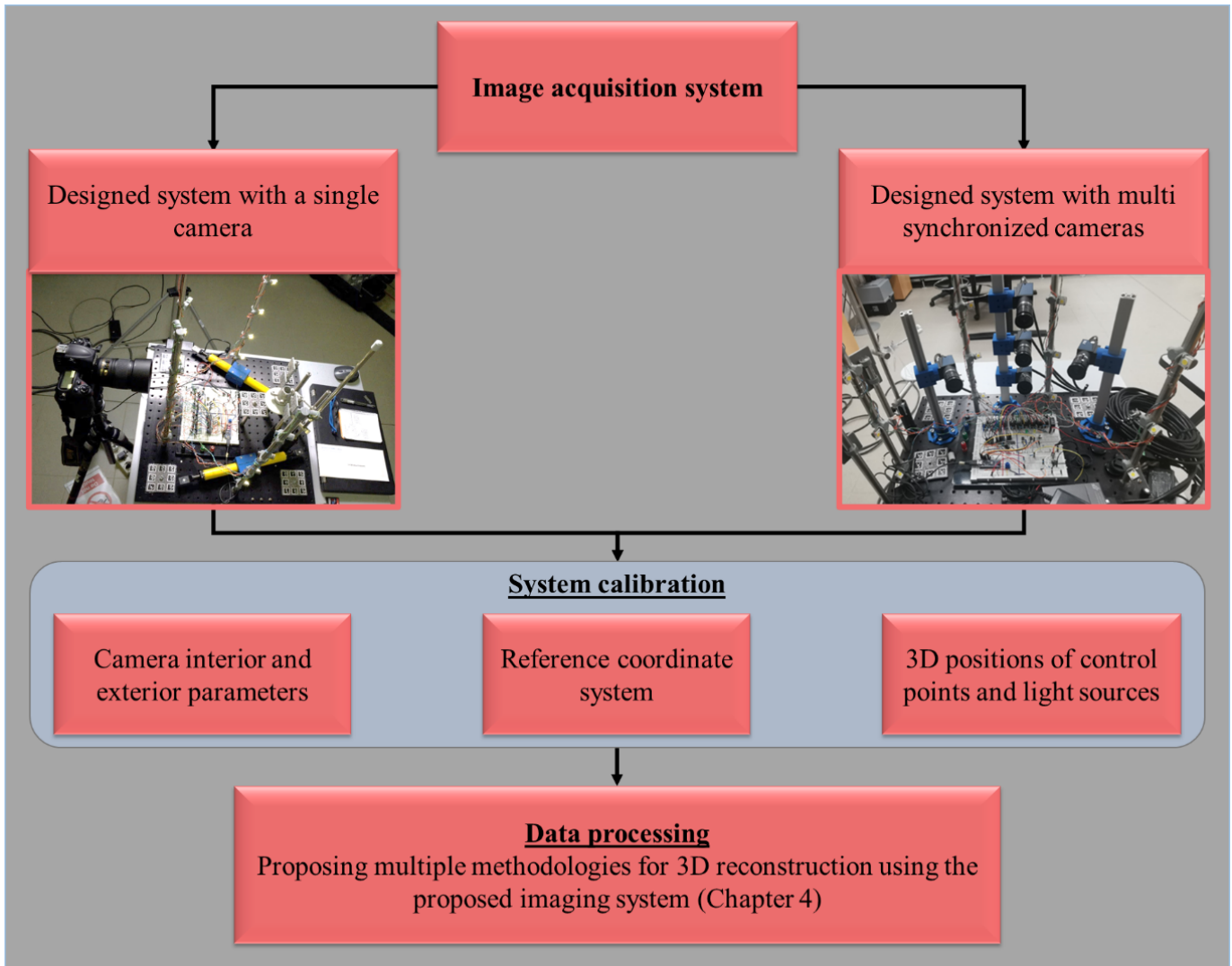
Figure 3-1. The general overview of the capturing system with its calibration steps.

Table 3-1. A summary of the system specifications for both developed system.

	Camera model	Resolution	Distance from object	LEDs	Focal length	GSD	Arduino model
Single system	Nikon D3X Nikon D750	24 Mpx	~400 mm	20	60 mm 120 mm	0.04 mm 0.02 mm	Arduino Nano- Atmel Atmega328
Multi-synchronized system	five synchronized GigE MER-1520-7GC	12Mpx	~400 mm	20	8 mm	0.06 mm	Arduino Nano- Atmel Atmega328

3.2.1 Designed system with a single camera

The proposed image acquisition system (Figure 3-2) is a modular system built using off-the-shelf and customized 3D-printed parts mounted on an optical breadboard. The hardware of the system is composed of four main segments: 1) a single digital camera; 2) multiple dimmable LED lights mounted on vertical poles; 3) a turntable; 4) a microcontroller (Arduino) with electronic circuitry to allow synchronization between the camera and LEDs.

The camera (Nikon D3X, 24 Mpx) is placed on a tripod with adjustable height, at nearly 50 cm from the objects. The lighting system consists of 20 high-power LEDs attached on four vertical poles. The poles are symmetrically placed on the breadboard, close to the object table.

43

The labels in the figure read:

LEDs

Camera

Registering
coded targets

Object on
the table

Microcontroller

Scale bar

Referencing
coded targets

Digital camera

Object

Arduino

Lighting system
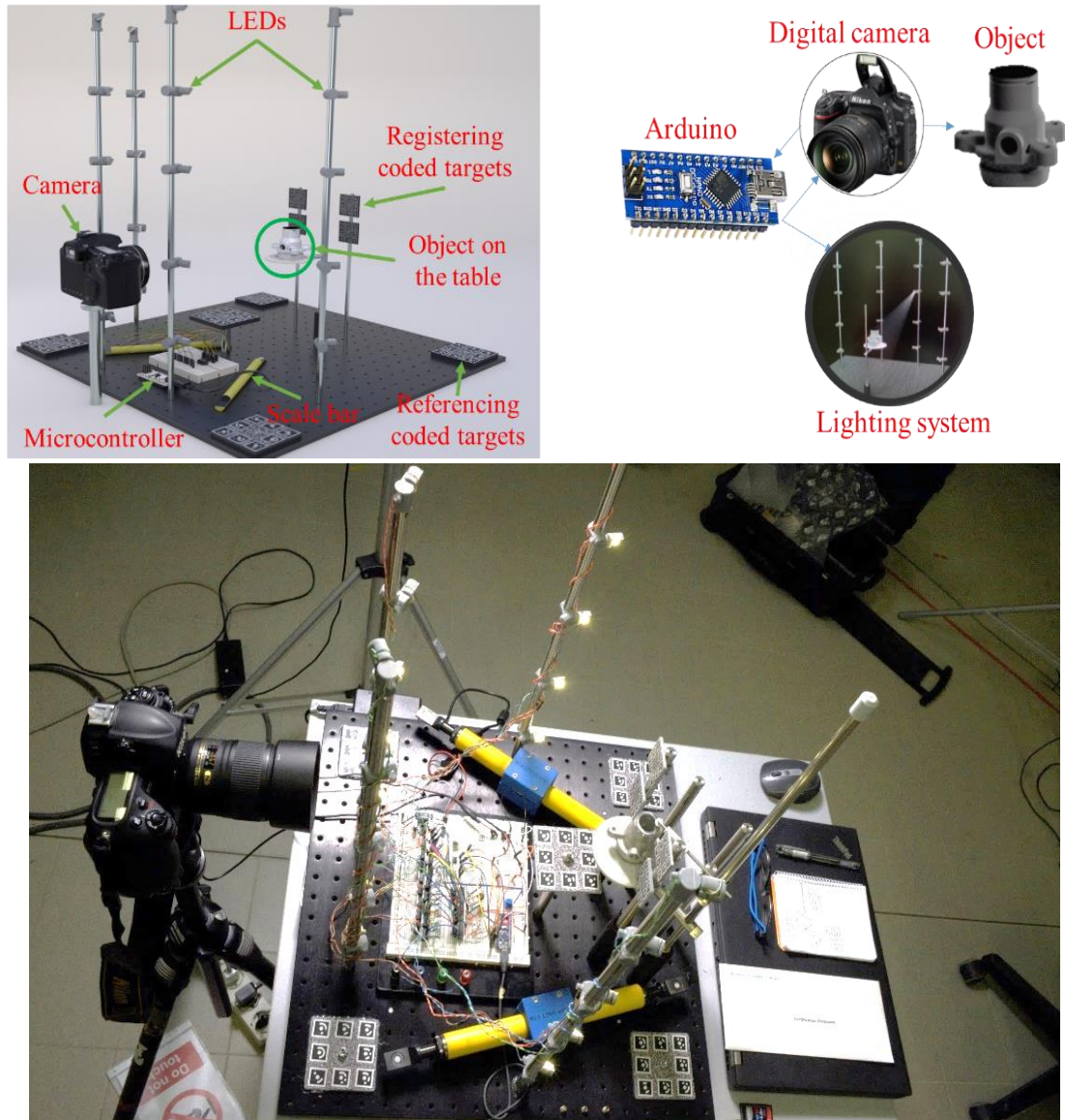
Figure 3-2. The implemented version of the proposed single image acquisition system.

On each pole, five LEDs are mounted at a distance of 10 cm from each other. The distance from each light to the object is approximately 40 cm. The system is highly modular with the orientation and position of each component adjustable to accommodate different object sizes.

The light sources are positioned in locations allowing various parts of the object to be illuminated under different grazing angles. This is to make sure roughness and micro-structures on the object surface can be properly highlighted. We use an Arduino Nano microcontroller to synchronize the camera, lights, and turntable.

To define a reference coordinate system, rigid square plates containing eight coded targets were fixed at the corners of the optical breadboard on the main plane in a horizontal position. A fifth plate was placed horizontally in the middle of the board near the object under inspection. Placing the targets in clusters over rigid square plates has several advantages. Repositioning each plate is flexible and using a single plate for camera pose estimation is possible, which significantly simplifies the image orientation. Also, by using the optical breadboard and standard studs, the lighting system can be easily reconfigured and optimized for objects of different sizes. In addition to these targets, four smaller (35x35 mm^2) target plates are mounted around the object in a vertical position. These, positioned at a distance within the camera's depth of field, are used as control points to register the

generated dense clouds to the calibrated coordinate system. Once the object is placed on the turntable and the camera parameters are set, images are acquired for each LED light.

Before starting the image acquisition, the system needs to be calibrated. The goal is to establish a geometric reference frame to which individual 3D surfaces can be registered to. The reference coordinate system is meant to remain stable and unchanged over time. Some coded targets are therefore fixed on the optical breadboard (Figure 3-3a) which are arranged in a cluster of 8 on each plate.

This allows computing the 6 degrees of freedom (6DoF) of the target plate. Targets are also placed vertically in the middle of the board, i.e., in front of the object being inspected, and are measured during the calibration stage. In this way, whenever an image is taken, the camera pose can be registered to the fixed reference system of the optical breadboard. During the calibration process, the coded targets and the light source coordinates are also calculated. To this purpose, two certified Brunson scale bars with a length of 350 ± 0.013 mm are also placed on the board to scale the measurements (Figure 3-3b). To perform the system calibration, all LEDs are switched on, and the whole system is imaged from 52 different positions (Figure 3-3c). Taking the images from many stations ensures that all coded targets and LEDs are imaged with a reliable camera network geometry.

46
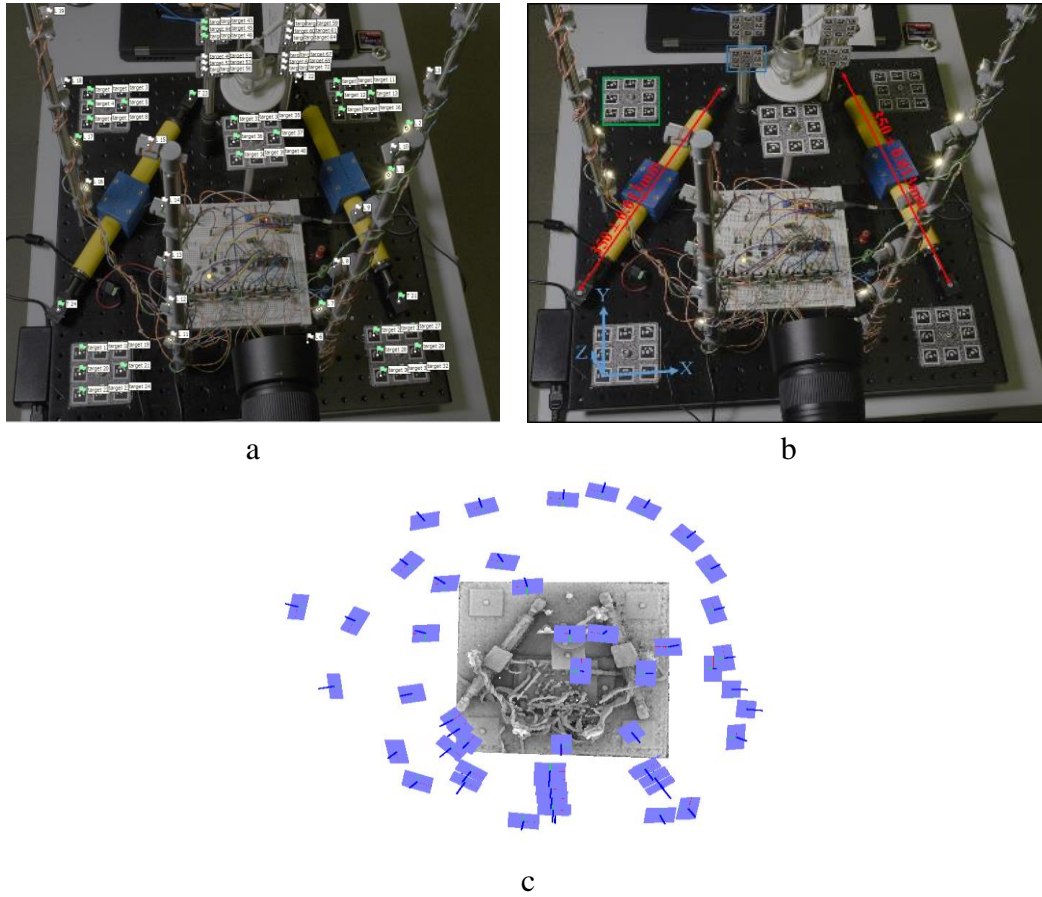
Figure 3-3. The calibration set-up of the proposed single acquisition system.

The calibration set-up of the proposed data acquisition system: a) detected coded targets and circular targets (LEDs). b) reference bars and coded targets useful to accurately calibrate the system. c) image network with recovered camera poses and 3D tie points (images taken from 52 stations).

Thereafter, the images are processed in AgiSoft Metashape V1.6.3 Build 10732 which is able to measure the image coordinates of the coded targets automatically. With the origin of the system fixed to the left bottom target (Figure 3-3c), the images are oriented, and the interior as well as exterior orientation parameters of the camera, are computed. Overall, a total number of 72 coded targets along with four scale-bar endpoints are used to perform a self-calibration of the system. The 20 LED center points are instead triangulated after the self-calibrating bundle adjustment. The results of this process include the adjusted internal camera parameters, the external orientation parameters of the camera in all 52 stations, and the 3D coordinates of all measured target points and LEDs centers. Table 3-2 reports the estimated Root Mean Squares (RMS) of image residuals (px) averaged over all points on all images during the calibration process.

Table 3-2. The estimated RMS of image residuals for scaling.

	Coded targets	LED Lights	scale-bar endpoints	Tie points
Number of the points	72	20	4	30,556
Error (pix)	0.1	0.3	0.17	0.7

The estimated RMS of image residuals (px) during the calibration process for all coded targets, light sources, scale bars, and observed tie points.

Regarding the radiometric calibration, we take two different kinds of light attenuation into consideration. The first is radial intensity attenuation, which occurs due to

the decrease in light energy that is proportional to the inverse squared distance between the light source and the surface point. This means that the intensity of the light decreases as it travels away from the source. In addition to radial attenuation, which refers to the decrease in light energy with distance from the source, angular intensity attenuation is another important factor to consider. Many light sources are directional, meaning they are brightest along a principal direction and become less bright at angles further from the main direction. This behavior can be observed and effectively simulated by multiplying the light intensity with the cosine of the angle between the light direction and the surface normal, also known as the attenuation coefficient. More details regarding intensity attenuation are expressed in Section 4.2.4.

3.2.2 Designed system with multi-synchronized cameras

The proposed PS image acquisition system with multi-synchronized cameras is presented in Figure 3-4. The capturing system is built to enable acquiring images under different view and illumination-conditions, explicitly allowing various parts of the object to be illuminated under different grazing angles (Karami et al., 2021) which is latter used to improve image orientation and dense 3D reconstruction of non-collaborative surfaces.

The system is similar to the one represented in Section 3.2.1, with the exception that instead of utilizing a single DSLR camera and manually moving to each station, we employ five synchronized industrial cameras mounting on three vertical poles at a distance of roughly 400mm from the object. A microcontroller (Arduino) with electronic circuitry is used to manage the synchronization and control the GigE cameras and LEDs.

An array of five synchronized GigE MER-1520-7GC cameras (See Figure 3-4) are positioned in a plus sign arrangement in front of the object. The cameras are distributed equally and mounted on three vertical poles placed on the optical breadboard, with each camera approximately fourteen degrees away from its neighbor(s). Table 3-3 lists the general specifications for GigE MER-1520-7GC cameras. Each camera has a LCM-10MP-08MM-F2.8-1.5-ND1 lens with a narrow field of view oriented towards the object placed on a rotating table. The lens features a variety of focal lengths to ensure that the appropriate working distance is met for each object. The aperture and focus are both adjustable and can be locked in place with a screw.
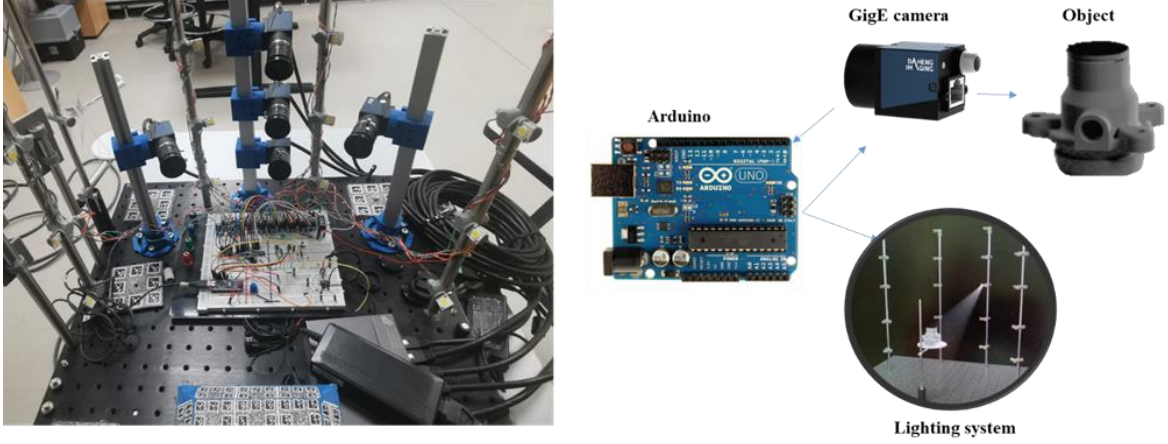
Figure 3-4. The proposed photometric stereo multi-view image acquisition system.

Table 3-3. The general specifications for GigE MER-1520-7GC camera.

Resolution	4608×3288
Pixel Size	1.4μm×1.4μm
Frame Rate	7fps@4608×3288
Dimensions	29mm×29mm×29mm

The cameras also are equipped with a standard GigE interface, and support Power over Ethernet (PoE, compatible of IEEE802.3af standard) making it easy to install and transmit power and data through a CAT6 network cable using a Power over Ethernet switch.

Once the object is placed on the rotating table, the camera parameters, i.e. distance to the object, focal length, F-Stop, and ISO, are manually set by an operator and kept constant.

51

(a)   (b)   (c)

(d)   (e)
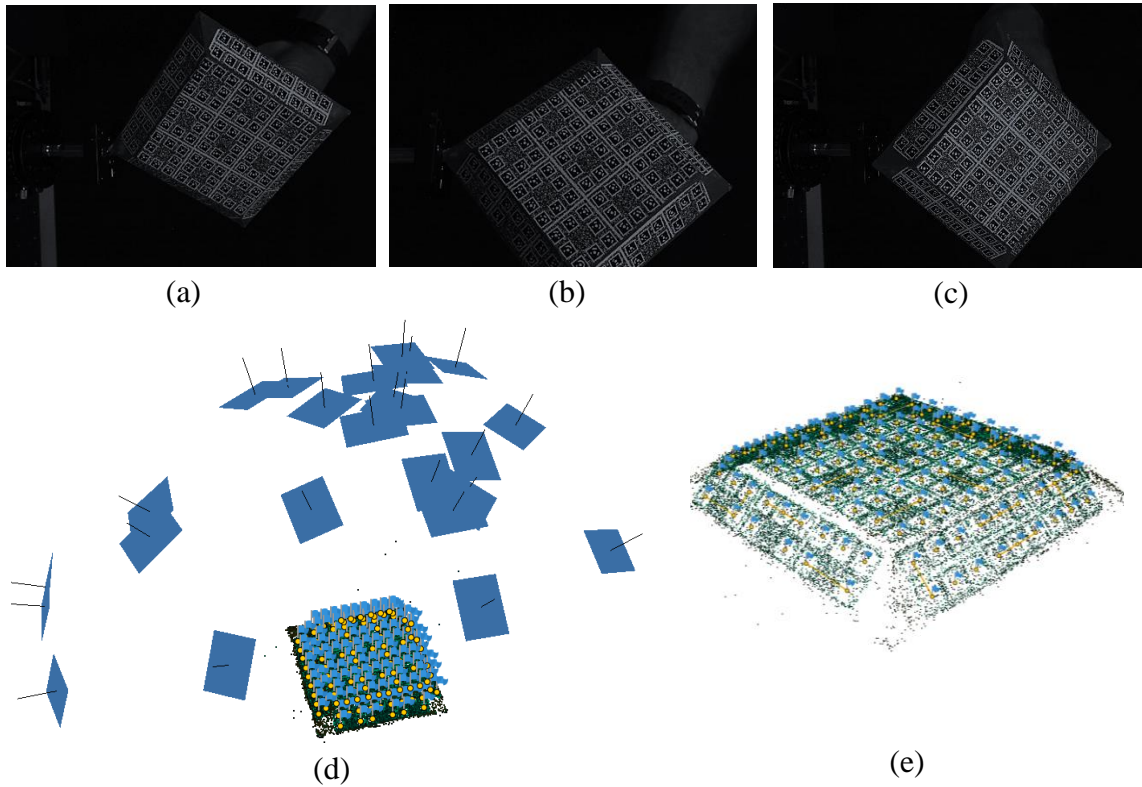
Figure 3-5. Self-calibration for multi synchronized GigE camera.
Three samples of the plate with embedded coded targets that moved/rotated in front of GigE cameras (a-c). The structure of imaging network (d) with sparse generated points with detected targets (e). the distance between targets are known (12.5mm) which is used to scale the model.

To begin image acquisition, turn on the first LED, take five synchronized images, turn off the LED, turn on the second LED, and capture the second image. This procedure

is carried out again for the remaining LEDs (twenty LEDs). Images are obtained within five seconds.

To calibrate the GigE cameras, we applied a self-calibration approach. To do this, we used a plate with embedded coded targets that moved/rotated in front of GigE cameras as shown in Figure 3-5. Instead of moving the cameras to acquire multi-view images for self-calibration, we freely moved/rotated the plate around while synchronized GigE cameras captured images. For each camera, 25 images were captured. Then after, the images were processed in AgiSoft Metashape V1.7.6 Build 13779 same procedure was done for single cameras in the previous Section 3.2.1.

3.3 Effect of lighting system on the generated 3D reconstruction

This experiment aimed to evaluate the adaptability and robustness of the proposed method to changes in the lighting system, specifically, to determine the extent to which the method's outcomes were affected by shifts in the lighting system's position. To achieve this, we employed Monte Carlo simulation, a methodology that can model various outcomes in complex processes that involve random variables. By using this method, the researchers were able to identify the impact of risk and uncertainty in the 3D reconstruction process when the lighting system was perturbed.

53

The simulation involved randomly perturbing each light position within a limited range to establish a statistical distribution of light positions that served as input for the proposed method's integrated algorithm. The algorithm generated a 3D reconstruction for each set of input parameters (i.e., perturbed light positions), and the Root Mean Squared Error (RMSE) of each reconstruction was calculated using a cloud-to-cloud comparison with reference data. This process was repeated 200 times, with each iteration randomly perturbing the light locations and yielding a different RMSE outcome.

Two scenarios were used in the simulation. In the first scenario, the position of the light sources was randomly disturbed within a range of ±5mm, and in the second scenario, within a range of ±10mm. We used object J as the test object for the simulation. The results of these experiments are shown in Figure 3-6 and Figure 3-7.
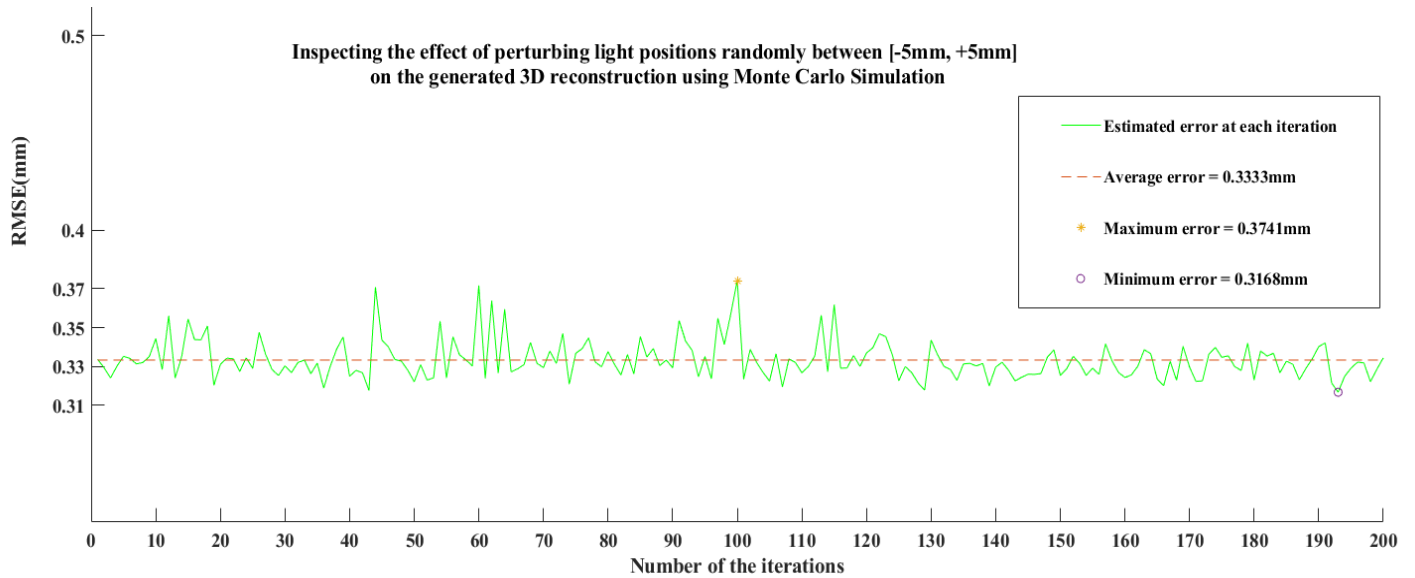
Figure 3-6. The result of perturbing the light locations.
The estimated RMS on the generated 3D reconstruction of object J as a result of perturbing the light locations randomly in the range of ±5mm.

Figure 3-6 and Figure 3-7 show the results of the Monte Carlo simulation for modeling the effect of perturbing light positions on the generated 3D reconstruction using the proposed algorithm. The average error in the first scenario, where the coordinate of light sources were perturbed with ±5mm, was 0.33mm with an uncertainty of ± 0.027mm. While in the second scenario, with perturbation of ±10mm, the average estimated RMSE was 0.35mm with an uncertainty of ± 0.05mm. This demonstrates the proposed integrated algorithm's flexibility and robustness to the perturbation of the lighting system.

55
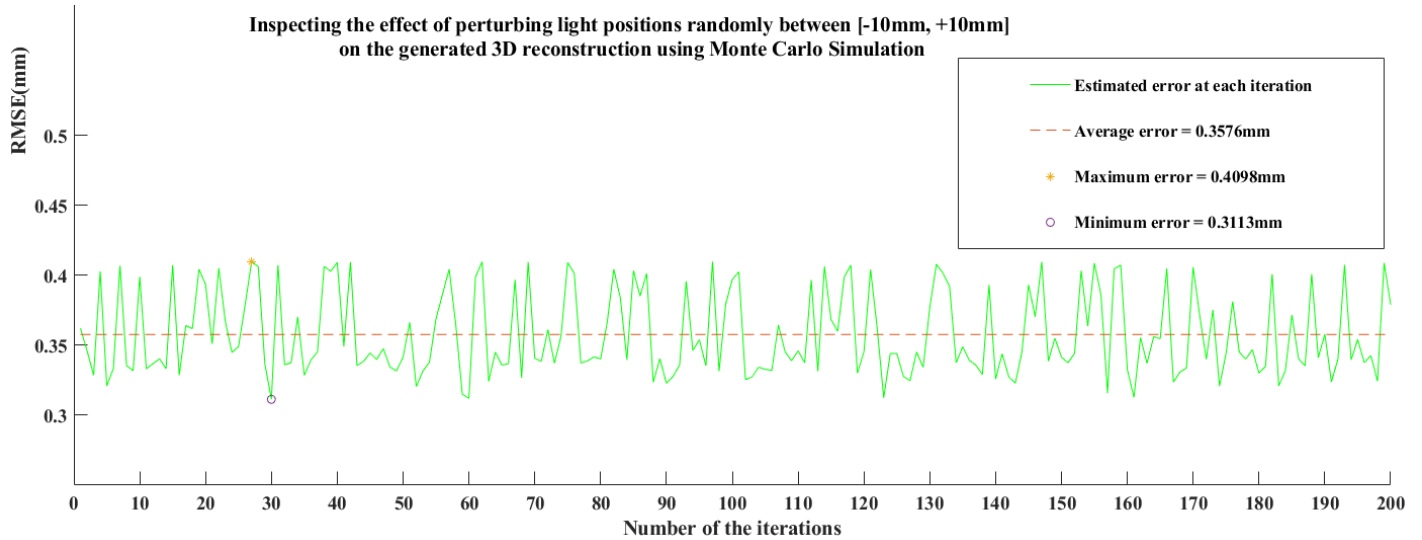
Figure 3-7. The result of perturbing the light locations.

The estimated RMSE on the generated 3D reconstruction of object J as a result of perturbing the light locations randomly in the range of ±10mm.

3.4   Summary

In this chapter, we proposed different image capture systems. First, we proposed an image acquisition system using a single camera. Then, we developed the same system with multi-synchronized cameras for fully-automatic data collection. Their calibration processes were also explained for each system.

56

CHAPTER IV

Methodology

4.1 Introduction

This thesis proposes different pipelines for 3D reconstruction of non-collaborative surfaces (textureless, shiny, and transparent) using our proposed PS data acquisition system. The general overview of the proposed methodology is summarized in Figure 4-1.
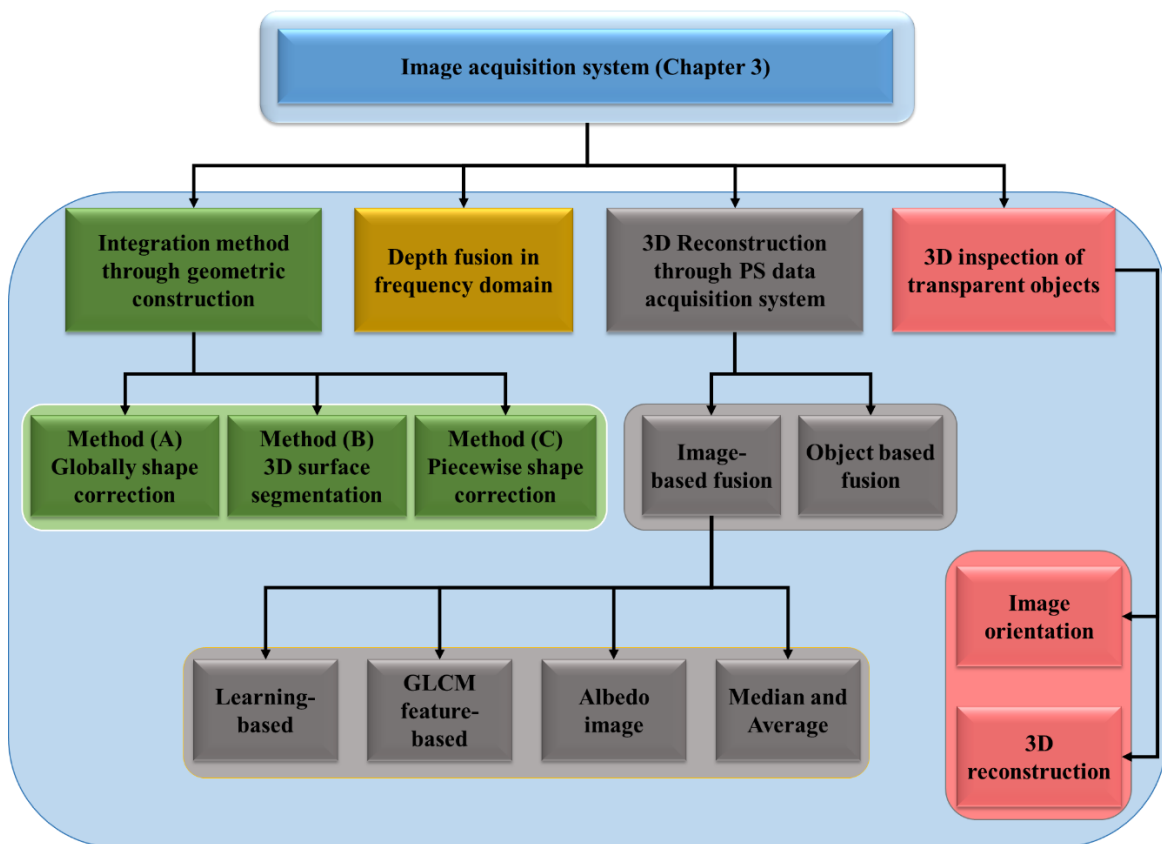
Figure 4-1. The general overview of the proposed method.

The first step is to provide an automatic image acquisition system to capture images under different illuminations and from different camera stations (camera positions) to satisfy the input requirement for each method. To this end, two image acquisition systems (Single and multi-synchronized cameras) along with their system calibration are proposed and discussed in Section 3. Following that, four different strategies are presented utilizing the proposed image acquisition system. The first proposed solution utilizes geometric construction to integrate photogrammetry and photometric stereo, leveraging the advantages of both techniques while overcoming their inherent limitations. (Section 4.2). This method uses photogrammetric 3D measurements to rectify the global shape deviation of photometric stereo depth induced by assumptions such as orthogonal projection, ideal diffuse reflection, or unknown error resources. The second method (Section 4.3) aims to provide detailed and precise 3D reconstructions of non-collaborative surfaces by fusing photogrammetry with photometric stereo depth maps in the frequency domain. To that aim, both photogrammetry and photometric stereo depth maps are transformed to the frequency domain in order to merge the high spatial frequencies of photometric stereo with the low frequencies of photogrammetry, resulting in accurate low frequencies while keeping high frequencies. Section 4.4 discusses how to utilize light directionality to improve texture quality by leveraging shade and shadow phenomena using the proposed image-capturing

system that employs several light sources to highlight roughness and microstructures on the surface. Finally, in Section 4.5, we first provide two methods that effectively orient images by leveraging the low-contrast textures highlighted on object surfaces (roughness and 3D microstructures). Second, a dense reconstruction of the testing objects is produced using various approaches given the oriented images.

The methodology underlying each approach is described in depth in the following sections.

4.2   Proposed integrated method

In this section, we propose a method for the 3D reconstruction of non-collaborative surfaces which combines photogrammetry and photometric stereo taking advantage of both methods and overcoming their own limits. The proposed method is summarized in Figure 4-2. The first step is to provide an automatic image acquisition system to capture images under different illuminations and from different camera stations (camera positions) to satisfy the input requirements of the integrated method (Section 4.2). A 3D point cloud is then generated with a photogrammetry pipeline (Karami et al., 2022b). The photogrammetric 3D shape measurements and the calibrated light positions are utilized to compute light direction and intensity attenuation (radial and angular) at each surface point.

The initial surface normal is then recovered given the light direction and the corresponding intensities at each surface point. Following this, the object's regions with shadow and specular reflection are detected and removed on the captured images depending on the angle between the light direction and the initial surface normal at each surface point. After removing outliers from images, the surface normal is updated given the light direction and the corresponding intensities at each surface point. A depth map is afterward generated from the integration of the surface normal. The scale factor is computed using corresponding 3D points between photogram-metric 3D reconstruction and photometric stereo depth maps. At the final stage, to further mitigate the shape deformation errors, three different approaches are proposed as follows:

1. **Method A:** it corrects the shape deviation by applying polynomial adjustment globally on the whole object.

2. **Method B:** it segments the object based on the normal and curvature then apply shape correction procedure on each segment separately.

3. **Method C:** it splits the object into small patches and then applies shape correction procedure on each patch separately.
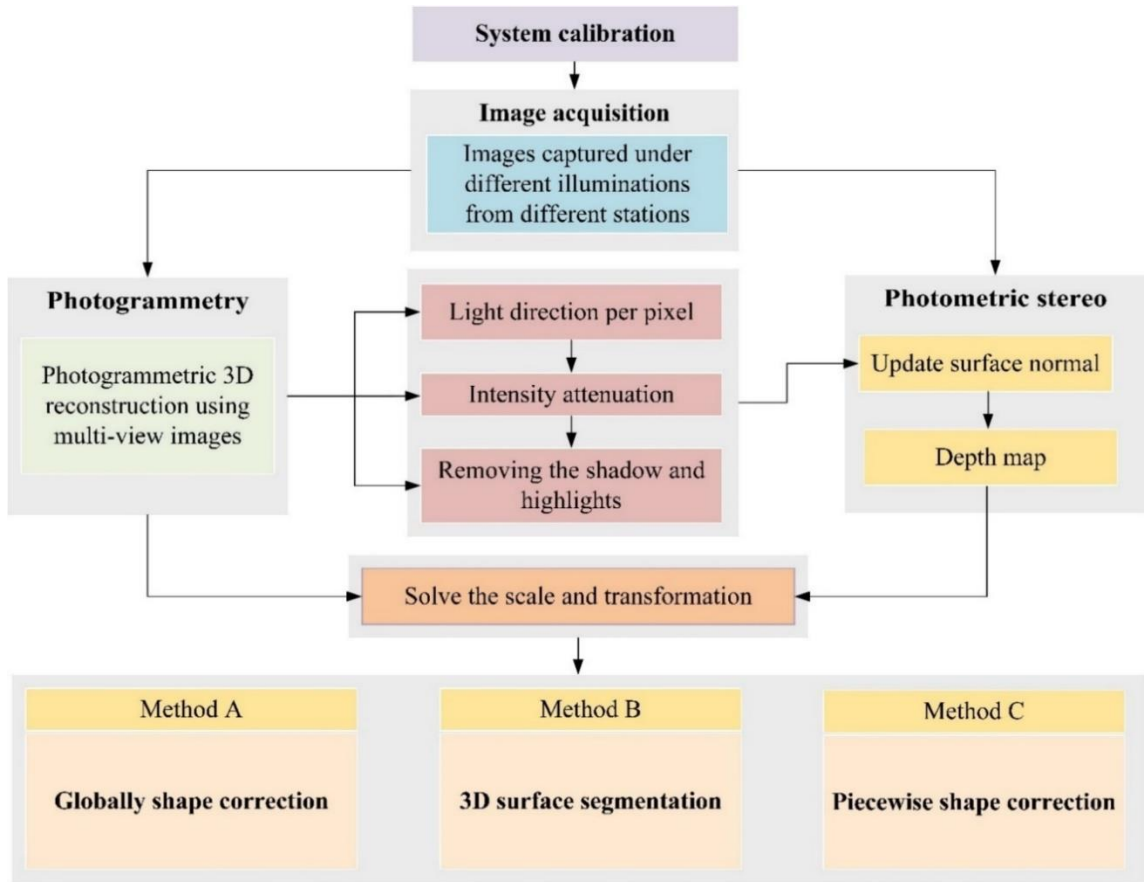
Figure 4-2. The proposed integration method.

The general overview of three methods to integrate Photogrammetry and photometric stereo.

4.2.1 Basic photometric stereo

Photometric stereo is a method to recover surface normal from multiple images that are taken under different lighting directions. The mathematical form of photometric stereo, for each surface point (X), is expressed in Equation (4-1).

$$I_i(X) = l_i(k\hat{n}) \qquad (i = 1,2,\dots,t) \qquad \textit{(4-1)}$$

Where I_i is the intensity observed in the i-th image, l_i is normalized light direction of the i-th source, k is the Lambertain reflection albedo, and $\hat{n}$ is the normalized surface normal at each surface point (X), which is unknown.

If $t > 3$, l can be inversed as $l^{-1}.I = k\hat{n}$. But, when the vector of light direction (l) is non-square, the generalization of the inverse is computed by multiplying both sides of the Equation (1) with l^T as following.

$$l^T.I = l^T.k\hat{n}$$
$$(l^T l)^{-1}.l^T.I = k\hat{n} \qquad \textit{(4-2)}$$

If we consider $k\hat{n}$ as a vector, the length of this vector is k and $\hat{n}$ is the normalized direction of that vector. Therefore, the surface normal and albedo can be recovered using Equation (4-3).

In theory, at least three lighting directions (images that are taken under different illuminations) are required to recover normal at each point. However, practically, more than three images are used to minimize noises involved in the process.

$$k = ||k\hat{n}|$$
$$\hat{n} = \frac{k\hat{n}}{k}$$
(4-3)

After computing the surface normal, the surface from gradients technique (Scherr, 2017; Shi et al., 2018) is applied to generate a 3D shape of the object from the field of normal vectors. The depth map is specifically given as $z = f(p, q)$, and the normal of the surface points towards the gradient direction. Where the p and q values are obtained as $p = -\frac{n_x}{n_z}$ and $q = -\frac{n_y}{n_z}$.

4.2.2 Light direction per-pixel

Conventional photometric stereo assumes that the light rays coming from the source are parallel. Providing parallel illumination conditions is more complicated and inefficient to implement in a close-range lighting system. Furthermore, it is also clear that rays coming from the light sources are not parallel anymore, especially when the lighting system is near to the object. Therefore, this impact must be addressed for accurate measurement of the

normal surface as any change in light position negatively affect the normal and consequently the resulting 3D reconstruction. For these reasons, we explored a geometric model with punctiform light sources and light divergence as expressed in (Karami et al., 2021; Zollhöfer et al., 2018).

To compute a unique light direction at each surface point, the 3D shape of the object and the light positions must be known in the same reference system (camera coordinate system). For doing this, the light positions are accurately measured with respect to the camera coordinate system during the system calibration (Karami et al., 2021; MacDonald et al., 2015). Then the sparse 3D points ($P_s(X, Y, Z)$) are then back-projected to the camera coordinate system using collinearity equation and known interior and exterior orientation parameters to find their corresponding pixels on the image ((I(u, v))). Then, as expressed in Equation (4-4), the normalized light direction ($\widehat{v_{k,s}}$) for each surface point($P_s(X, Y, Z)$ is computed given the coordinate of the kth light source (l_k) (Karami et al., 2021; Fan et al., 2017).

$$\widehat{v_{k,s}} = \frac{(l_k - p_{s(X,Y,Z)})}{\|l_k - p_{s(X,Y,Z)}\|} \qquad (4\text{-}4)$$

Finally, the surface normal is computed at each surface point given $\widehat{v_{k,s}}$, and their associated image intensities (I(u,v)) as expressed in Equation (4-5).

65

$$\widehat{n_s} = \frac{(v_{k,s}{}^T . v_{k,s})^{-1} . v_{k,s}{}^T . I_k}{\|(v_{k,s}{}^T . v_{k,s})^{-1} . v_{k,s}{}^T . I_k\|} \qquad (4\text{-}5)$$

4.2.3 Backprojection

As we explained in the previous section, in order to compute a unique light direction for each surface point, we need to have the 3D coordinate of each light source as well as surface point in the camera coordinate system where the depth from photometric stereo is computed. The 3D points and the light positions, measured during the calibration process are transformed into the camera coordinate system using Equation (4-6).

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R * \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t \qquad (4\text{-}6)$$

Where (X, Y, Z) are the 3D coordinates of a surface point or light source in the local coordinate system defined during the calibration step, R and t are the rotation matrix and translation vector, respectively.

Successively, the 3D points are back-projected in the image coordinate system using the perspective transformation expressed in Equation (4-7), and then to the pixel coordinate system.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x/z \\ y/z \end{bmatrix} \qquad (4\text{-}7)$$

66

Generally, the lenses used in any project have some distortion, which could be modeled as radial distortion and tangential distortion (Brown, 1971). These distortions are modeled using Equation (4-8).

$$\begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} x'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + k_3 r^8) + (p_1(r^2 + 2x'^2) + 2p_2 x'y')(1 + p_3 r^2 + p_4 r^4) \\ y'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + k_4 r^8) + (p_2(r^2 + 2y'^2) + 2p_1 x'y')(1 + p_3 r^2 + p_4 r^4) \end{bmatrix} \qquad (4\text{-}8)$$

Where

$$r^2 = x'^2 + y'^2;$$

k_1, k_2, k_3, and k_4 are radial distortion coefficients;

p_1, p_2, p_3, p_4 are tangential distortion coefficients.

After estimating the lens's distortion, we can ultimately reach to the pixel coordinate system by utilizing the Equation (4-9).

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x . x'' + c_x + w.\,0.5 + B_1 . x'' + B_2 . y'' \\ f_y . y'' + c_y + h.\,0.5 \end{bmatrix} \qquad (4\text{-}9)$$

Where (u, v) is the coordinate of the image point in pixels corresponding to the reconstructed 3D points $(P_s(X, Y, Z))$, f_x, and f_y are the focal lengths in pixel unit, and (c_x, c_y) is the coordinate of the principal point. B_1, B_2 are affinity and non-orthogonality (skew) coefficients, respectively; w, h are image width and height in pixels.

67

4.2.4 Intensity attenuation

There are two different kinds of light attenuation that are needed to be taken into consideration when point light sources are used (see Figure 4-3). The first factor is caused by a decrease in light energy which is proportional to the inverse squared distance between the light source to the surface point (radial intensity attenuation). The second attenuation factor that we address is a realistic directional model of a light source (angular intensity attenuation).
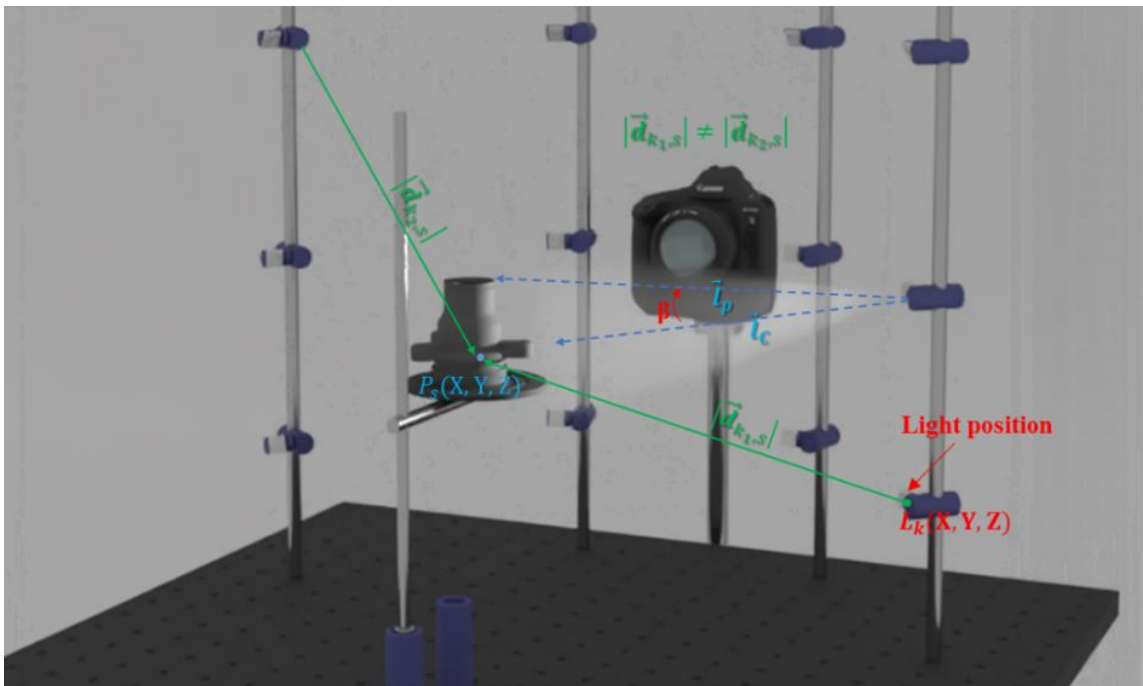
Figure 4-3. Radial and angular intensity attenuations.

4.2.4.1 **Radial intensity attenuation**

The intensity of the light decreases when the object moves away from the illumination source (Mecca et al., 2014). In theory, this behavior of the light is modeled using Equation (4-10).

$$F_{k,s}^R = \frac{1}{h_c + h_l|\vec{d}_{k,s}| + h_q(|\vec{d}_{k,s}|)^2}$$

(4-10)

Where $|\vec{d}_{k,s}|$ is the distance between the k-th light source and surface point $(P_s(X,Y,Z))$. h_c, h_l, and h_q are the attenuation coefficients. However, in practice, the first and the second terms are ignored since their values are negligible compared to the third term. In this paper we also considered h_q equal to one.

4.2.4.2 **Angular intensity attenuation**

The light intensity decreases as we move away from a light source, but also when the light moves angularly (β) further from the cone axis ($\vec{l}_c$) (Mecca et al., 2014). Equation (4-11) is a commonly-used approach to model such phenomenon:

$$cos(\beta)^\mu = (\vec{l}_c.\vec{l}_s)^\mu$$

(4-11)

Where β is the angle between the cone axis ($\vec{l}_c$) and the direction from the light position to the surface point ($\vec{l}_s$), and μ is the attenuation coefficient.

69

4.2.5 Shadow and specular reflection removal

The known camera geometry, lighting system, and approximate 3D shape of the object were used to automatically detect shadow and specular reflection and keep only the best-highlighted parts on each image where 3D microstructures and roughness are seen. After capturing multiple images from different stations, the images are automatically inspected, and those regions on each image that appear to result in inferior quality and/or noise (shadow and secularity) are excluded (Karami et al., 2022b). To do this, as shown in Figure 4-4, the incoming angle (θ_1) and reflected angle (θ_2) are estimated at each surface point P(x,y,z) corresponding to each pixel (i,j) given the light direction $l_{(x,y,z)}$, camera direction $r_{(x,y,z)}$ and normal $n_{(x,y,z)}$ using Equations (4-12) and (4-13):

$$\theta_1 = cos^{-1}\left(\frac{n_{(x,y,z)}\cdot l_{(x,y,z)}}{\|n\|\|l\|}\right) \qquad \textit{(4-12)}$$

$$\theta_2 = cos^{-1}\left(\frac{n_{(x,y,z)}\cdot r_{(x,y,z)}}{\|n\|\|r\|}\right) \qquad \textit{(4-13)}$$

Specular reflection occurs when an incoming light ray ($l_{(x,y,z)}$) reflects off of a surface point $P_{(x,y,z)}$ at an equal but opposite angle (θ_2) to its incoming angle θ_1. In our experiment, we consider a small amount of light scattering (e) around the reflection vector (r) as the specular reflection zone, which is tuned to ±2 degrees. This value can be changed

depending on the strength of the light source. By increasing this value we assure that a larger area around the reflection vector (r) is considered as an outlier.

A self-shadowed pixel is shaded by itself. Geometrically, the angle between the surface normal (n) and the light source direction (l) is more than 90 degrees ($\theta_1 > 90$). In this experiment, a pixel is considered as self-shadow if the computed incoming angle θ_1 is larger than 85 degrees. Once self-shadow and specular reflection are removed from images, the surface normal is updated for further processing.

$$\theta_1 \cong \theta_2 \mp e$$
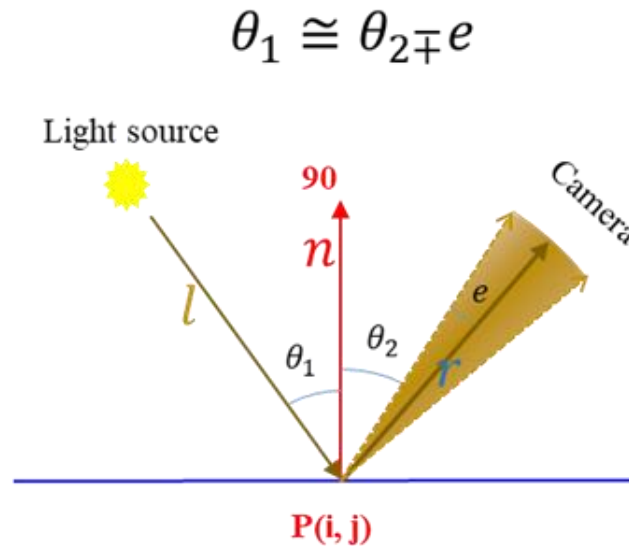
Figure 4-4. Shadow and specular reflection removal.

Removing shadow and specular reflection using the accurate geometry of the lighting system and object's sparse 3D shape.

71

4.2.6 Helmert transformation

The Helmert 3D transformation is one of the most often used transformation methods in geodetic applications. This transformation is defined with seven parameters, including three translations, three rotations, and a scale factor, which allow us to transform the photometric stereo depth map to a defined coordinate system from which photogrammetric 3D points are obtained. When this transformation is performed to a 3D point cloud, it rotates, transforms, and scales the point cloud with respect to the defined coordinate systems. The mathematical form of this transformation is expressed in Equation (4-14).

$$X_T = C + \mu RX \qquad\qquad (4\text{-}14)$$

Where X_T is the original 3D Points (3D reconstruction from photogrammetry), and X is the transformed 3D points (3D reconstruction from photometric stereo), C is the three translations (x_t, y_t, z_t) along the coordinate axes, R is the rotation matrix, and μ is the scale factor. The seven Helmert transformation parameters must be calculated using corresponding points from both datasets.

4.2.7 Global shape correction with polynomial model (Method A)

The 3D reconstruction computed using the photometric stereo principle shows residual deformation even after correcting the light directions and applying intensity attenuations because of the other mathematical simplifications and unknown error sources. Therefore, to further mitigate the residual errors we use a polynomial mapping from the 3D model obtained with the photometric stereo to the one obtained through photogrammetry. This technique was inspired by polynomial adjustment in aerial triangulation (Mikhail et al., 2001) with 20 coefficients. The 20 coefficients are computed using least square principles after forming the design matrix containing Equation (4-15) one for each common point in the photogrammetric (X,Y,Z) and photometric stereo (x,y,z) models.

$$X = x + a_1 + a_3 x - a_4 y + a_5(x^2 - y^2)2a_6xy + a_7(x^3 - 3xy^2) - a_8(3x^2y - y^3)$$
$$Y = y + a_2 + a_4 x + a_3 y + a_6(x^2 - y^2) + 2a_5xy + a_7(3x^2y - y^3) + a_8(x^3 - 3x^2y) \qquad (4\text{-}15)$$
$$Z = z + b_0 - 2b_2 x + 2b_1 y + c_1 x^2 + c_2 x^3 + c_3 x^4 + d_1 xy + d_2 x^2 y + d_3 x^3 y + d_4 x^4 y + e_1 y^2 + e_2 xy^2$$

It is worth noting that both the deformed model (from photometric stereo depth map) and reference data (photogrammetric 3D points) should be approximately scaled and aligned before computing the coefficients. The application of global polynomial adjustment (Method A) to complex-geometry surface objects can present a notable

73

disadvantage, as these adjustments have the potential to smooth out the edges and boundaries of the object. This is particularly evident in objects with complex shapes, such as those represented by Objects C and D. The reason for this is that polynomial adjustments are designed to fit a single global polynomial onto the entire object based on a set of 3D control points. When dealing with objects with complex shapes, this approach can prove challenging, and the resulting fit may not accurately represent the true edges and details of the object.

To address this issue and ensure that the edges of complex-geometry objects are preserved, two alternative approaches (Methods B and C) have been developed. These methods involve dividing the object into smaller patches and applying shape deformation adjustments to each patch separately. By breaking the object down in this way, it is possible to preserve the finer details and edges of the object while still achieving an accurate overall fit.

4.2.8 3D surface segmentation (Method B)

A 3D segmentation is performed to the point cloud in order to divide the object into many parts, and then applied transformation and shape correction to each segment separately (Figure 4-5).
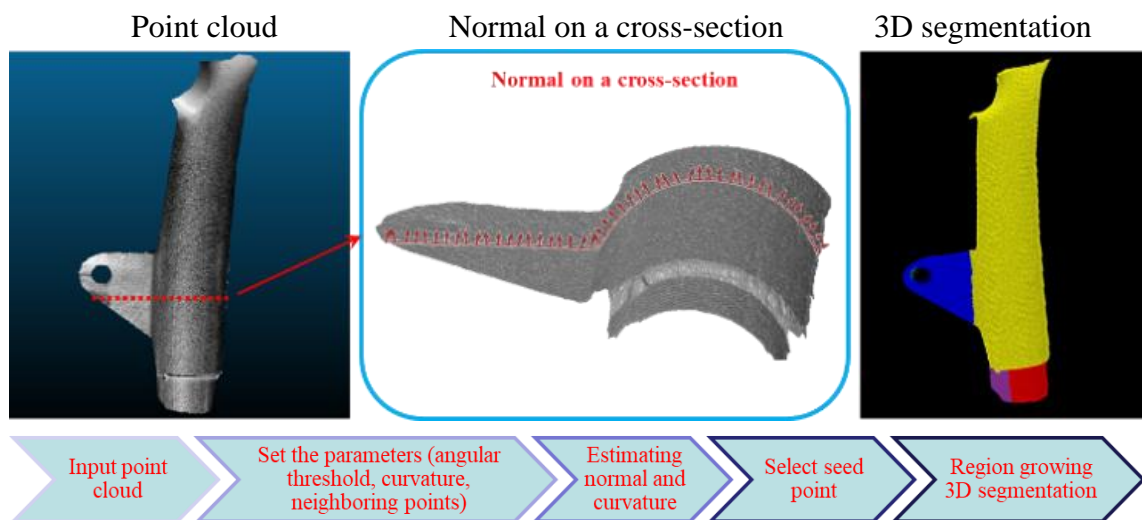
Figure 4-5. 3D Surface Segmentation (Method B)

Schematic view of the region growing 3D segmentation approach (Method B). Each color represents different segment.

To segment the point cloud, a region-growing approach (Murtiyoso, and Grussenmeyer, 2020) is adopted. This method uses local features (point cloud normal and curvature values) obtained from neighboring points to segment nearby points with similar

properties. After obtaining the k-NN for a point p, the salient local 3D features (e.g. point normal, curvature, etc.), are calculated for each point p. The point p with the minimum curvature value is chosen as the first seed point to begin the region-growing process. Therefore, the point might be selected in a smoother area on the object where the surface variation is lower. Following the selection of seed points, the region-growing segmentation starts and gradually expands by adding new points. Once the first segment is complete according to region-growing criteria (Murtiyoso, and Grussenmeyer, 2020), a new seed point is selected for the following segment.

4.2.9 Piecewise shape correction (Method C)

The object is divided into small patches (Pn) as illustrated in Figure 4-6, and then the shape correction is applied to each patch (Pn) individually. As a result, the global geometry of the surface in each patch becomes less complicated, with fewer edges and boundaries, hence, the object's deformation is more likely to be corrected more effectively. To this end, after splitting the object into small patches in the image space, we select the first patch and then using the corresponding points in both models (photometric stereo and photogrammetry), the patch from the deformed model is roughly transformed and fitted to

76

the course photogrammetric model using helmet transformation. Then, a polynomial adjustment is applied to the same patch to mitigate the deformation locally. This procedure should be performed for the next patch (P_2) and the remaining patches. Although this approach corrects the global deformation, the model's 3D details might be negatively affected near the patch's borders due to disconnectivity. To solve this, two constrains are considered: (i) each patch must have an overlapping area with its neighboring patches, (ii) there must be always some corresponding points in the overlapping area to stitch all the patches together.
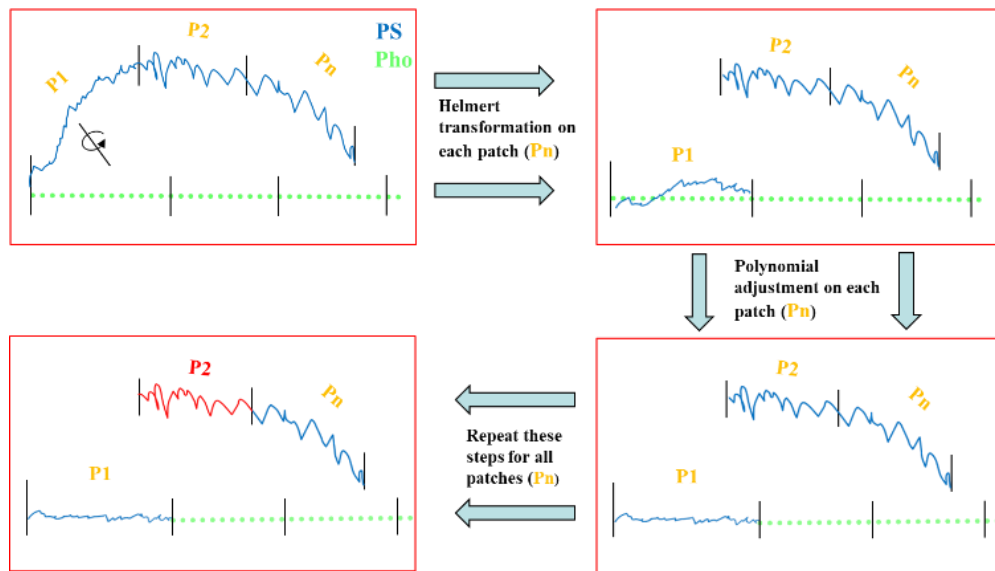
Figure 4-6. shape correction approach (Method C).

Schematic view of pricewise shape correction approach (Method C).

4.3 Fusion in frequency domain

In this section, we aim to combine photogrammetry and PS depth maps in frequency domain to get detailed and precise 3D reconstructions of non-collaborative (i.e. shiny, texture-less, translucent, etc.) surfaces.

In PS depth map, inaccurate low frequencies are normally present due to several not-fulfilled assumptions of the PS mathematical model, such as ideal diffuse reflection with no shadow nor specularities on the surface, parallel illumination direction, and orthogonal projection. But high-frequency information is preserved with high accuracy regardless of these assumptions. Photogrammetric depth, on the other hand, fails to exploit high frequencies when the assumption of ideal diffuse reflection with a well-textured surface is not satisfied, despite the fact that its low-frequency information is still reliable. Therefore, we propose an FFT-based method to fuse the high spatial frequencies of photometric stereo with low frequencies from photogrammetry in order to have accurate low frequencies and meanwhile keep high frequencies.

Figure 4-7 shows a general overview of 3D integrated depth maps using the proposed method. After collecting a stack of images obtained under different illuminations from different viewpoints, the depth maps using both photogrammetry and PS are generated (Karami et al., 2021, 2022b).
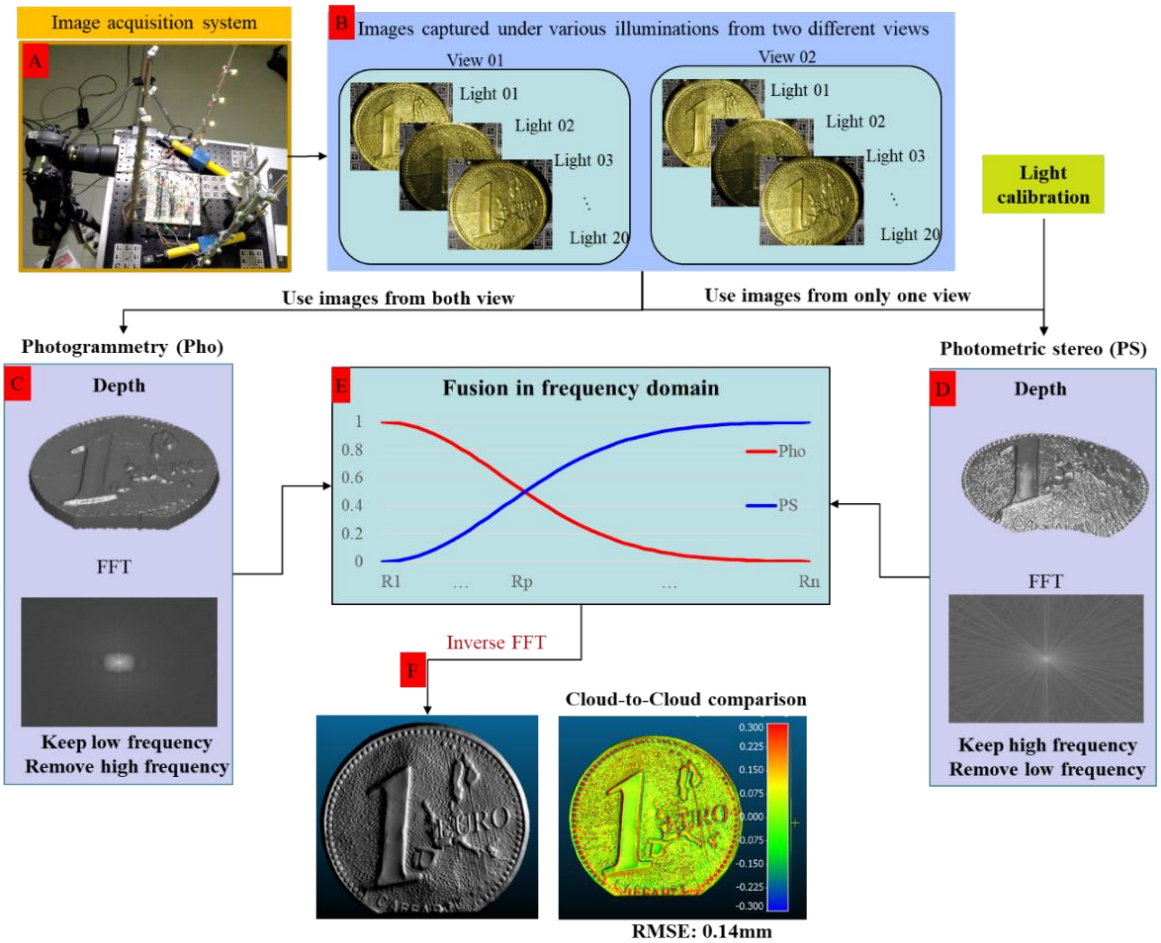
Figure 4-7. Fusion in frequency domain.

> *General steps to fuse photogrammetry and PS data using FFT filter. A) acquire the images, B) generate depth using both photogrammetry and PS, C) transfer them to the frequency domain, D) Filter low-frequency from PS and high-frequency from photogrammetry, E) fuse them using a non-linear blending approach, F) apply inverse FFT to convert the fused depth back to spatial domain and convert the depth map to a point cloud.*

79

Then, they are converted to frequency domain using Fourier transformation decomposing each 2D depth map into its sine and cosine components. In this way, it is more convenient to distinguish and modify low and high frequencies. In the frequency domain, we first create a weighting plane to assign value to each pixel, and then we use a non-linear interpolation to eliminate incorrect frequencies while fusing accurate low and high frequencies. Finally, the generated fused depth maps are transferred back to spatial domain using an inverse Fourier transformation and then converted to 3D point clouds.

4.3.1 Fourier transformation

Photogrammetric depth map (Figure 4-7C) is generated using image pair captured from different stations (Karami et al., 2021, 2022b). On the other side, the PS depth map (Figure 4-7D) is generated given the stack of 20 images taken only at one viewpoint and the known calibrated light direction (Karami et al., 2021, 2022b). Afterward, both depth maps are transferred to the frequency domain using FFT. For an image of size M×N, the two-dimensional Fourier Transform is expressed in Equation (4-16).

$$F(p,q) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} f(u,v) * e^{-j(2\pi/M)p_m}e^{-j(2\pi/N)q_n}$$

$$p = 0, 1, \ldots, M-1$$
$$q = 0, 1, \ldots, N-1$$

(4-16)

Where *f(v, u)* is a 2D image in the spatial domain and the exponential term is the basis function corresponding to each point F(p, q) in the frequency domain.

4.3.2 Weighting plane

Typically, the Fourier image is shifted in a way that F(0,0) is displayed in the center of the image. An image point's associated frequency rises with distance from the center. Thus, low-frequency power is concentrated in a small area at the center of the shifted frequency plane, while high-frequency power is displayed as one moves away from the center as shown in Figure 4-8.

Therefore, we define a Gaussian weighting plane (W) using Equation (4-17), which assigns each pixel in the frequency domain a weight between 0 and 1 based on its distance from the plane's center (Fourier image).
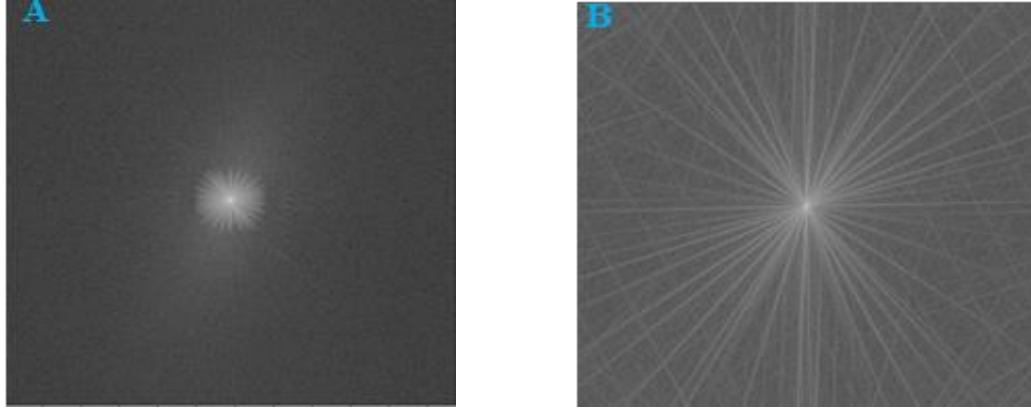
Figure 4-8. Depth transformation using FFT.

Two examples of photogrammetric (A) and PS (B) depth transformation in frequency domain using FFT for Object 03.

$$W = exp^{-\hat{R}^2/2*T} \tag{4-17}$$

$$\hat{R} = \frac{R - R_{min}}{R_{max} - R_{min}} \tag{4-18}$$

Where, $\hat{R}$ is the normalized radial from center of the image. This should be noted that Rp (Figure 4-7E) in the weighting plane, where allocated weight to both depths are equal, can be changed by modifying parameter T in Equation (4-17). This parameter must be set empirically depending on the dataset.

4.3.3 Depth fusion in frequency domain

To fuse both depths, a non-linear interpolation (Figure 4-7E) expressed in Equations (4-19) is used to generate a fused 2D depth out of the photogrammetry and PS depth maps.

$$FusdF = W * F_{Pho} + (1 - W) * F_{PS} * \frac{Total_{F_{Pho}}}{Total_{F_{PS}}}$$

$$Total_{F_{Pho}} = \sum_{i=1}^{m} \sum_{j=1}^{n} F_{Pho}(i,j)$$

$$Total_{F_{PS}} = \sum_{i=1}^{m} \sum_{j=1}^{n} F_{PS}(i,j)$$

(4-19)

Where, W is a Gaussian weighting plane defined in Equation (4-19); F_{Pho} and F_{PS} represent the transferred depth maps in frequency domain for both photogrammetry and PS, respectively; $Total_{F_{Pho}}$ and $Total_{F_{PS}}$ are also the total frequency power for both photogrammetry and PS, respectively.

With such interpolation, PS high frequencies are gradually replaced with photogrammetric low frequencies. As one moves away from the center, where the distribution of high-frequency power is, the weighting trend progressively turns inverted. The weight assigned to photogrammetric frequency (Pho) drops to 0 while the weight assigned to photometric stereo (PS) rises to 1.

83

4.3.4 Depth transformation to 3D point cloud

The final step after fusing both depths is to apply the inverse Fourier Transform to transfer the generated fused depth to spatial domain (Figure 4-7F). Successively, the depth is converted to 3D point cloud given the 2D depth map and an RGB image (Pan et al., 2016). To this end, as shown in Equation (4-20), the 3D vertex $v(p) = (x, y, z)$ of image point p in the camera's coordinate space is created by using a calibration matrix:

$$v(p) = D(p)K^{-1}[p, 1] \qquad \qquad (4\text{-}20)$$

Where $D(p)$ is the depth value, and K is the 3×3 intrinsic camera parameters, which is estimated during camera calibration. $v(p)$ is the 3D vertex corresponds to image point p(i,j).

To compute the normal vector $n(p)$ corresponding to each 3D vertex $v(p)$ through computing the cross product of the nearby re-projected points as expressed in Equation (4-21):

$$n(p) = \big(v(i + 1, j)\big) - v(i, j)) * (v(i, j + 1) - v(i, j) \qquad (4\text{-}21)$$

Finally, Exterior orientation matrix (rotation R and translation T) is applied to convert the 3D vertex and corresponding normal from camera coordinate system to global coordinate system using Equations (4-22) and (4-23).

$$V(p) = [R\ T]v(p) \hspace{3cm} \text{(4-22)}$$
$$N(p) = [R\ T]n(p) \hspace{3cm} \text{(4-23)}$$

4.4 3D Reconstruction through PS data acquisition system

In this section, we aim to present, for the first time, the effects of light directionality on the quality of 3D reconstruction of non-collaborative objects by photogrammetry. To this aim, we used our image-capturing system that utilizes multiple light sources to highlight roughness and microstructures of non-collaborative surfaces which are not visible under diffuse lighting direction. These roughness are then used as a sort of chiaroscuro texture in image orientation and multi-view stereo (MVS) algorithms to ensure effective matching procedures.

4.4.1 Effects of light directionality

To illustrate the effects of light direction on highlighting 3D microstructures, we designed and printed an object with a surface topography displaying sinusoidal waves in four distinct orientations (Figure 4-9). The surface was then illuminated with diffuse lighting and with light coming from different angles (i.e., left, bottom, diagonal). As shown in Figure 4-9, when light interacts with rough surfaces, the highlighted microstructures

vary according to the surface structure and direction of the light. For example, when light is coming from the left, parts B, C, and D are more noticeable than part A. On the other hand, light from the bottom highlights the micro-topography in part A and makes part D shadowless.

a) Diffuse lighting	b) Light coming from left	c) Light coming from bottom	d) Light coming diagonally
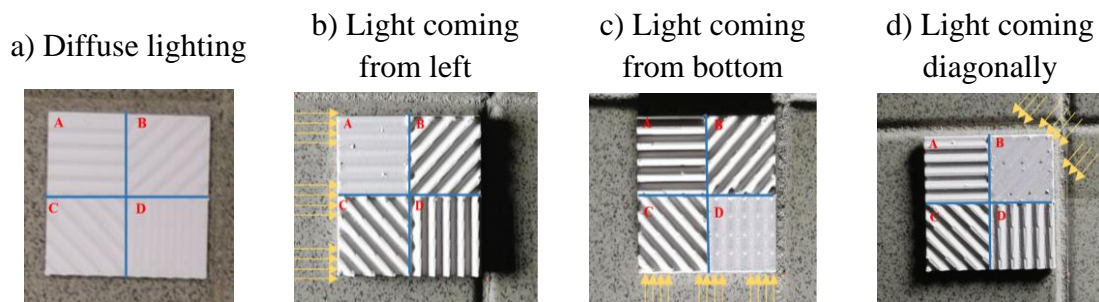
Figure 4-9. Generating different sinusoidal waves through different illumination.

The importance of light directions for highlighting microstructures on a surface: a) an object featuring different patterns (seen under diffuse light). b-d) the same object seen under different illumination directions (shown with the yellow arrows).

Therefore, the use of light directionality, obtained through a photometric stereo image acquisition protocol, can be used also for better highlighting microstructures on the surface. Thus, depending on the light direction, shadows and shading phenomena can produce different spatially varying chiaroscuro patterns. The patterns can be used as a kind

of texture for successful matching processes within image orientation and multi-view stereo (MVS) algorithms.

4.4.2 Grazing angle

As shown in Figure 4-10, the grazing angle (g) is the angle between the ray incident on a surface and the tangent surface at the point of incidence. In other words, it is defined as the 90-degree complement to the angle of incidence (α).
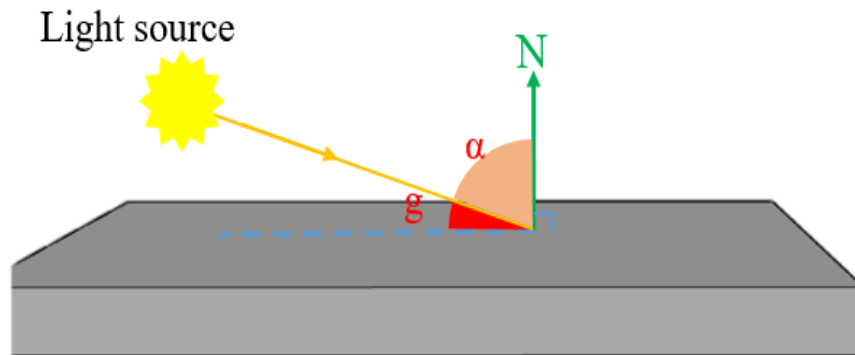
Figure 4-10. Schematic of the grazing incidence geometry.

When a light ray is incident on a surface at an angle, it interacts with the surface in various ways, depending on the nature of the surface and the angle of incidence. If the surface is perfectly smooth and flat, and the angle of incidence is less than the critical angle (which would be 90 degrees in this case), the light ray would be reflected at an equal angle

to the incident angle, according to the law of reflection. The law of reflection states that the angle of incidence is equal to the angle of reflection, and the reflected ray lies in the same plane as the incident ray and the normal to the surface. If the angle of incidence is greater than the critical angle, the light ray would be reflected back into the medium (create shadow on the image). This phenomenon is known as total internal reflection. If the surface is rough or irregular, the behavior of the light ray upon incidence would depend on the nature and orientation of the surface irregularities. In general, when a light ray is incident on a rough surface, it is scattered in all directions due to the varying angles of incidence at different points on the surface.

4.4.3 Selecting the best-highlighted regions

Once the multiple images from different stations are captured, the images are automatically inspected using the advantage of known geometry of lighting system and the approximate 3D shape of the object, and only those regions with the best and favorable grazing angles are selected as shown in Figure 4-11. Consequently, on each image, only parts in which surface roughness and microstructures are best highlighted remain, with the rest that seems to lead to lower quality and/or noise (shadow and secularity) disregarded.

To this end, the grazing angle (g), shown in Figure 4-11, is computed at each surface point ($P_{(i, j)}$) given the light direction (l) and normal (n) using Equation (4-24).

$$g = 90 - \cos^{-1}\left(\frac{n.l}{\|n\|\|l\|}\right) \qquad (4\text{-}24)$$
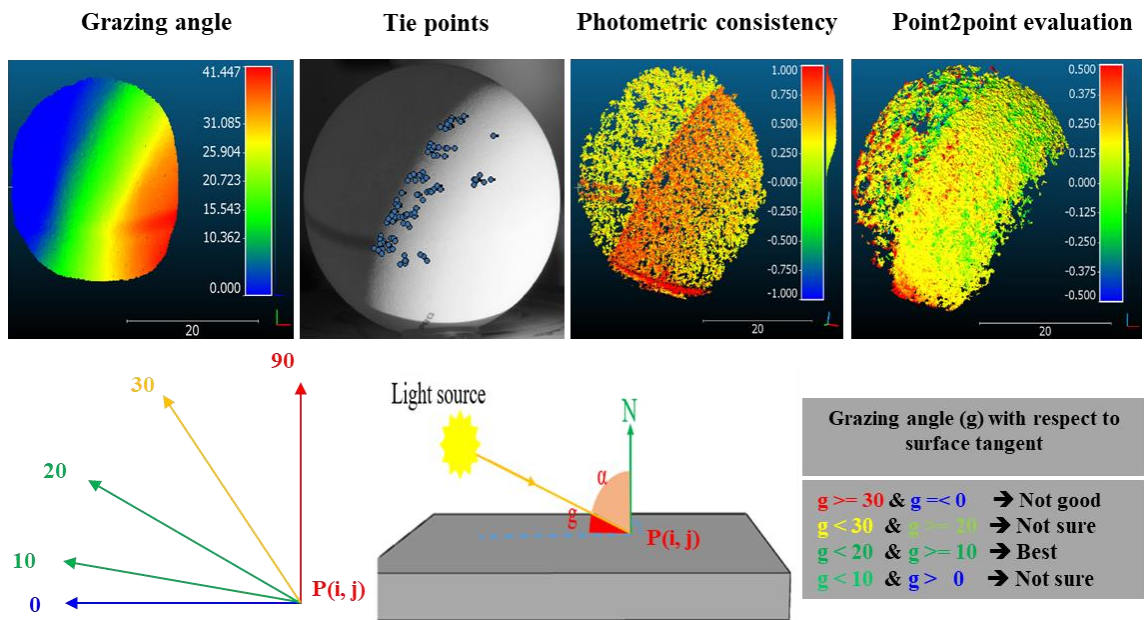
Figure 4-11. Selecting the best-highlighted region.

Synopsis of the presented analyses over a spherical object to show the effect of the grazing angles in relationship with point cloud accuracy analysis. The grazing angles between 10 to 20 degrees significantly enhance the texture content of the image due to the chiaroscuro effect created at a micro-scale level.

89

4.4.4 Image-based fusion

The technique of fusing relevant information from two or more images into a single image is known as image fusion which is more informative than the input images. The input images are fused at the pixel level followed by the information extraction. The idea of such fusion is to improve the signal-to-noise ratio, obtain an image free from shadows and highlights, and improve the image quality of the output image. In this regard, five different methods are proposed including average, median, albedo, GLCM-based, and deep learning based.

4.4.4.1 **Median and Average**

The first approach for Image fusion was to extract the median and average images from a stack of images taken under different illuminations. Since the original grazing angle images at any station are all captured from the same position, they all overlap. Therefore, for any pixel (i, j), there exist multiple intensities (equal to n the number of light sources), with lower intensity values representing shadow and higher intensity values referring to the highlighted points. The value of a pixel in the median or average images is defined by the average or median of its intensities. Pixels in the images generated in these methods cast diffuse illumination with reduced shadow and secularity effects. The main aim of using

these images was to provide a diffuse lighting condition where the intensity on the object

is independent to the viewer and constantly reminds from one station to another station.

4.4.4.2 **Albedo image**

To produce fused images with diffuse illumination, the authors considered a second

method: using photometric stereo to extract the albedo map for each viewpoint and then

using these maps for 3D reconstruction. As shown in Figure 4-12, if we consider $k_s\hat{n}_s$ as

a vector at each surface point (P_s) corresponding to $(I_{(u,v)})$, the length of this vector is

albedo (k_s) and $\hat{n}_s$ is the normalized direction of that vector, as expressed in Equation

(4-3). Figure 4-13 shows an example of extracted albedo image and normal map. Albedo

is a measure of the diffuse reflection observed when light interacts with a surface and is

determined by the ratio of radiosity to the irradiance observed by a surface. Obtaining the

albedo map from multiple images taken under varying illuminations using photometric

stereo allows for a resulting image that casts a minimal shadow and highlights effects,

similar to median and average images.
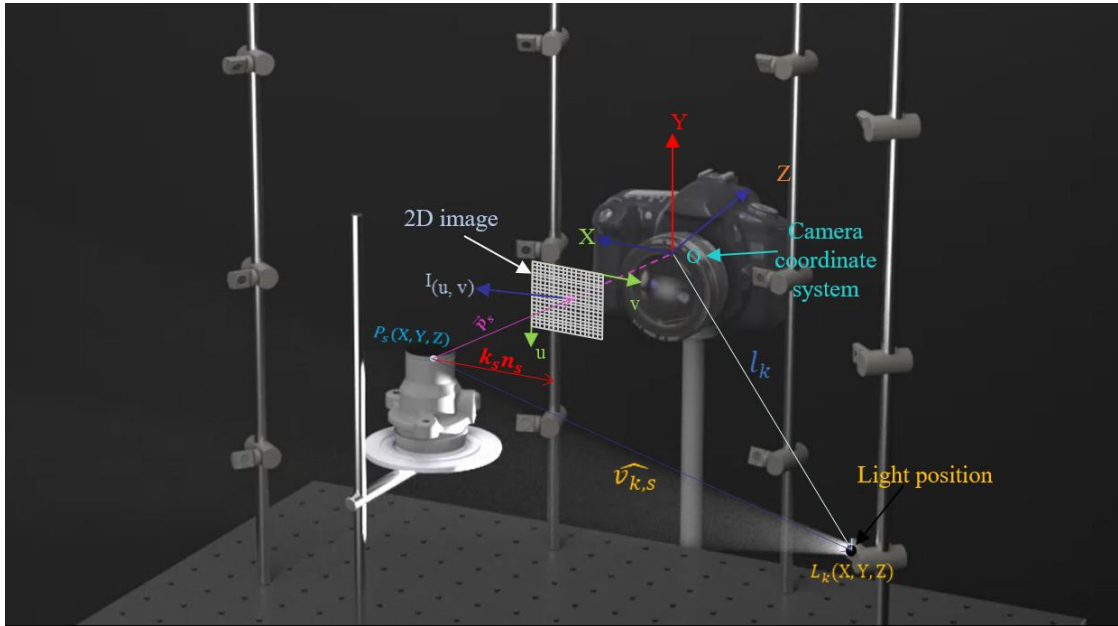
Figure 4-12. Visual representation of normal and albedo at each surface point.
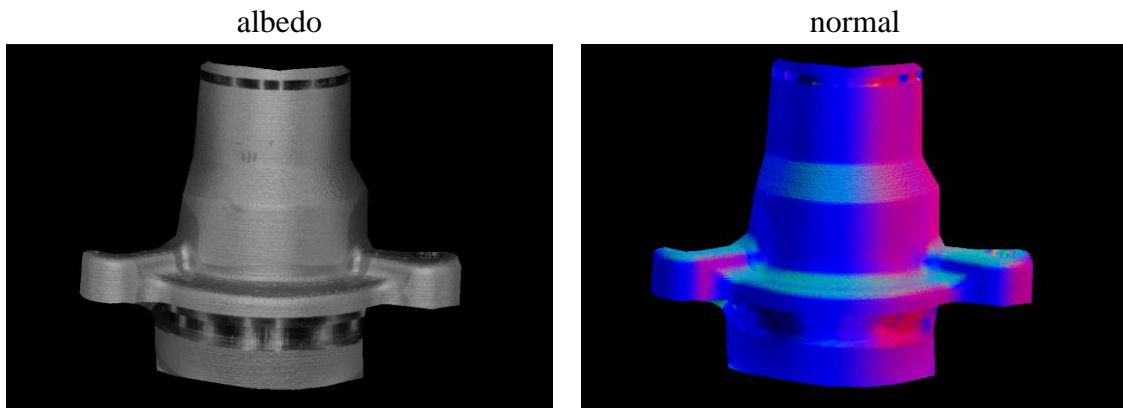
albedo   normal

Figure 4-13. A sample of extracted albedo and normal map

Additionally, the texture of the resulting albedo image remains constant and independent of the viewer's perspective, which is a necessary condition in photogrammetry procedures. This map is a unitless and non-dimensional value that varies between zero and one, indicating the surface whiteness in a grayscale image. A value of zero indicates that the pixel is black, representing a perfect reflector, while a value of one indicates that the surface is white, representing a perfect absorber. By utilizing this method, the authors were able to obtain a high-quality albedo map that accurately represents the diffuse reflection observed by the surface. This map can then be used in 3D reconstruction procedures to produce fused images with diffuse illumination, resulting in a final image with minimal shadow and highlight effects and constant texture independent of the viewer's perspective.

4.4.4.3 **Image fusion based on the GLCM features**

The third methodology to fuse images is to use GLCM (Gray Level Co-occurrence Matrix) feature extraction as shown in Figure 4-14. The main aim of this work is to investigate the use of the gray-level co-occurrence matrix technique as an absolute image quality metric to extract the region on the image where the roughness and micro-structures are highlighted better.
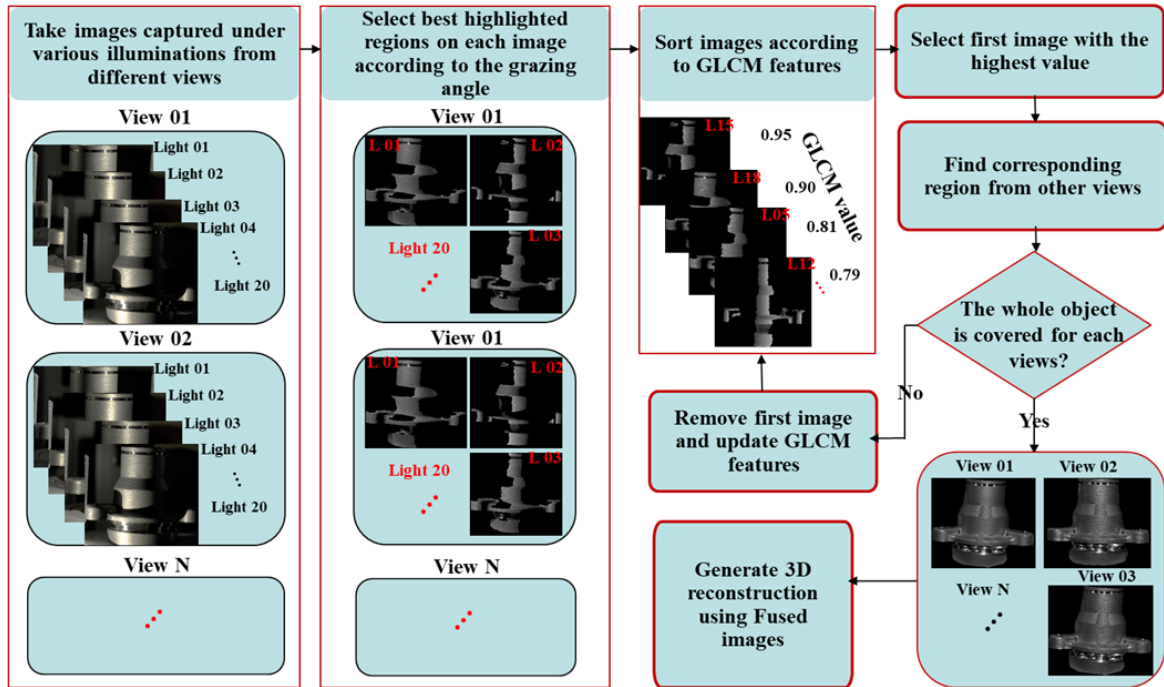
Figure 4-14.　Image fusion based on the GLCM features.

The general steps of fusing images using GLCM .(1) Take the images (2) Select the best highlighted region (3) Sort the images according to the GLCM values (4) create the fused images Using the grazing images valued by GLCM features (5) Generate 3D reconstruction.

The GLCM is a symmetric matrix, where each element (i,j) represents the frequency of occurrence of a pair of pixels with gray-level values i and j and a specific spatial relationship. The matrix is normalized to obtain probabilities, which represent the likelihood of a pair of pixels occurring at a specific distance and direction with specific

94

gray-level values. Once the GLCM is constructed, several statistical measures can be derived from it, which are used as texture features for image analysis. These features include contrast, dissimilarity, and homogeneity, which are used to characterize the spatial distribution of gray-level values in the image.

To do this, first, a small area (window size) is defined to be used for filling in the GLCM and measuring the texture. Then, a window is placed in the first position over top left of the image. Once the first calculation is completed, the window is shifted one pixel to the right and the procedure of calculating a new GLCM and a new texture measure is repeated.

Once the first row was done, the window is moved down one row and the procedure is repeated. This process is repeated until the entire image is covered. In this way, a new image is created with texture values. To measure the texture, three statistical texture measures (Haralick, 1979) including Contrast (Con), Dissimilarity (Dis), and Homogeneity (Hom) were used to summarize the normalized symmetrical GLCM in helpful ways.

$$Con = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \qquad (4\text{-}25)$$

$$Dis = \sum_{i,j=0}^{N-1} P_{i,j}(i-j) \qquad (4\text{-}26)$$

$$Hom = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i-j)^2} \qquad\qquad (4\text{-}27)$$

Where, $P_{i,j}$ is the probability of values i and j occurring in adjacent pixels in the image within the defined window. i and j are the labels of the columns and rows (respectively) of the GLCM.

The difference between contrast and dissimilarity is the assigned weight to GLCM values as they move away from the diagonal. Values on the GLCM diagonal show no contrast, and contrast increases away from the diagonal. In Con, the weight is the square of the distance, but in Dis, the weight is the absolute value of the distance.

Once the texture measure calculation was done for all the images in the first station (view), a weight $w_{(im)}$ ranging from 0 (worst) to 1 (best), is assigned to each grazing image. The images are sorted in descending. The higher the value, the more roughness, and microstructure appear on the images.

$$w_{(im)} = \frac{(Con + \ Dis + (1 - Hom))}{3} \qquad\qquad (4\text{-}28)$$

Then after, the first image with the highest weight is selected as the first candidate and used to create the fused image in the first station. To create the fused image in the next station, we must first find the corresponding region in the following station using interior

96

and exterior camera parameters and then fill this region with the same image candidate but from the second station. The same procedure is repeated for other stations. In this way, the object texture for the corresponding regions remains constant, avoiding false matches and artifacts in the final generated 3D dense point clouds.

However, using only one grazing image can't cover the entire object because some parts/regions of each image that were saturated or shadowed were removed during the previous stage (Section 4.4.3). Thus, to cover/fill the empty region on the objects, the second image candidate with the highest GLCM value is used. This procedure is repeated until the whole object is covered in each station. And finally, the 3D point cloud is reconstructed given fused images.

It should be noted that since the fused image is the combination of all images illuminated from different light sources, the appearance texture on each region to the neighboring one can be different. Furthermore, disconnectivity and brightness changes at the border of each region appear on the images making fused images unbalanced and chaos. Therefore to avoid this, we blend the candidated grazing images. To this end, we first use a smoothing weighting function based on the Euclidean distance to weight each image. The weighting value (w_i) that is assigned to each pixel is the function of the distance of that

pixel from the closest boundary, hence, it decreases gradually from one to zero as getting close to the borders (see Figure 4-15). Then considering this weighting map corresponding to each image (w_i), both images (I_1, I_2) are fused/ blended as expressed in Equation (4-29).

$$I_{(fused)} = \frac{w_1 * I_1 + w_2 * I_2}{w_1 + w_1} \tag{4-29}$$
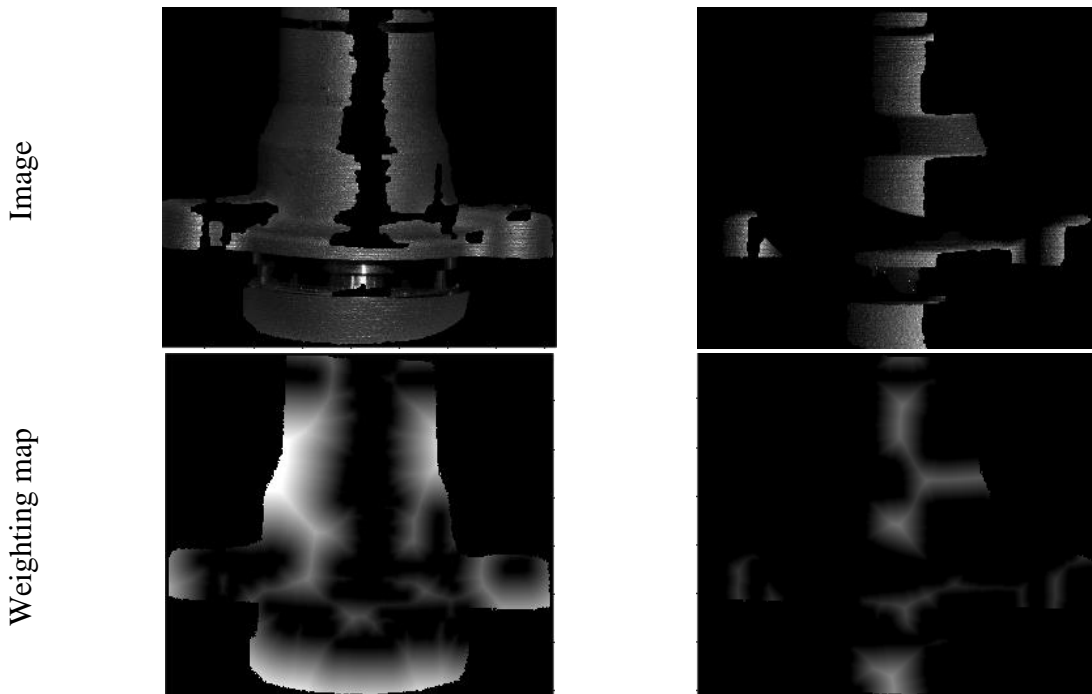
Figure 4-15. Weighting grazing images.

This is an example of weighting each image pixel according to the Euclidean distance between a pixel and the nearest nonzero pixel (border) on image.

4.4.4.4 Image fusion based on deep learning

In this Section, we use a deep-learning approach to fuse images. The first step is to select the best areas on each view and later apply the fusion using all the resulting views. The architecture of the adopted network is presented in Figure 4-16.
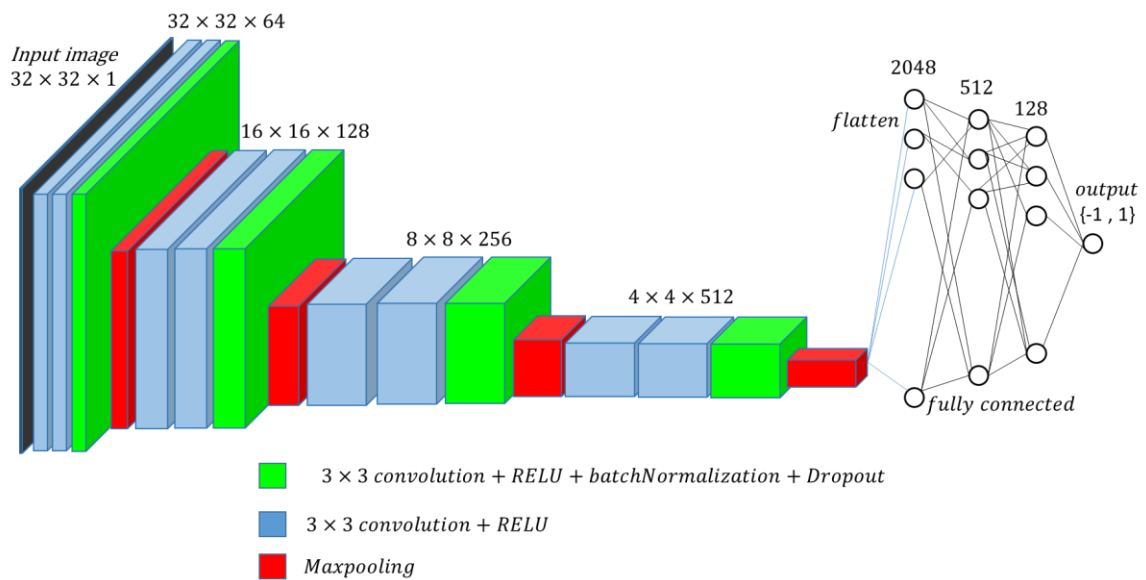
Figure 4-16. Image fusion based on deep learning.

The used CNN model is a standard CNN for classification composed of blocks, each block is composed of 2 or 3 layers of 2D convolution followed by a max-pooling layer. The last layer is flattened and connected to a fully connected network with one

output. All layers have a ReLU as an activation function except the last one (the output) where the hyperbolic tangent activation function is used.

To train the net, an automatic procedure is applied to avoid manually selecting the best and the worse part on each image taking the advantage of known geometry lighting system and approximate 3D shape of the objects. To this end, as shown in Figure 4-17, pixels on each image are segmented and classified into three different groups according to their corresponding grazing angles. The first category belongs to those pixels that are illuminated at the grazing angle larger than 30 degrees and below zero (mostly area in shadow and around specular reflection). According to the experiment presented by Karami et al. (2022a), pixels (area) on each image that are illuminated at a grazing angle ranging from 10 to 20 degrees are selected as the best-highlighted region where 3D microstructure and roughness are best highlighted (Second category). This angle can also be modified depending on the object. By decreasing or narrowing this angle, a smaller area is selected as the best-highlighted region while by increasing this angle, it is more likely to select/consider areas where the roughness and microstructure are not highlighted well. And the third category is comprised of areas where we are unsure whether roughness and microstructure are sufficiently highlighted or not, and it is situated between the first and second groups.

100

To train the model, we utilize pixels (regions) from the second group, then we apply the trained net to the second and third groups to obtain the fused image.
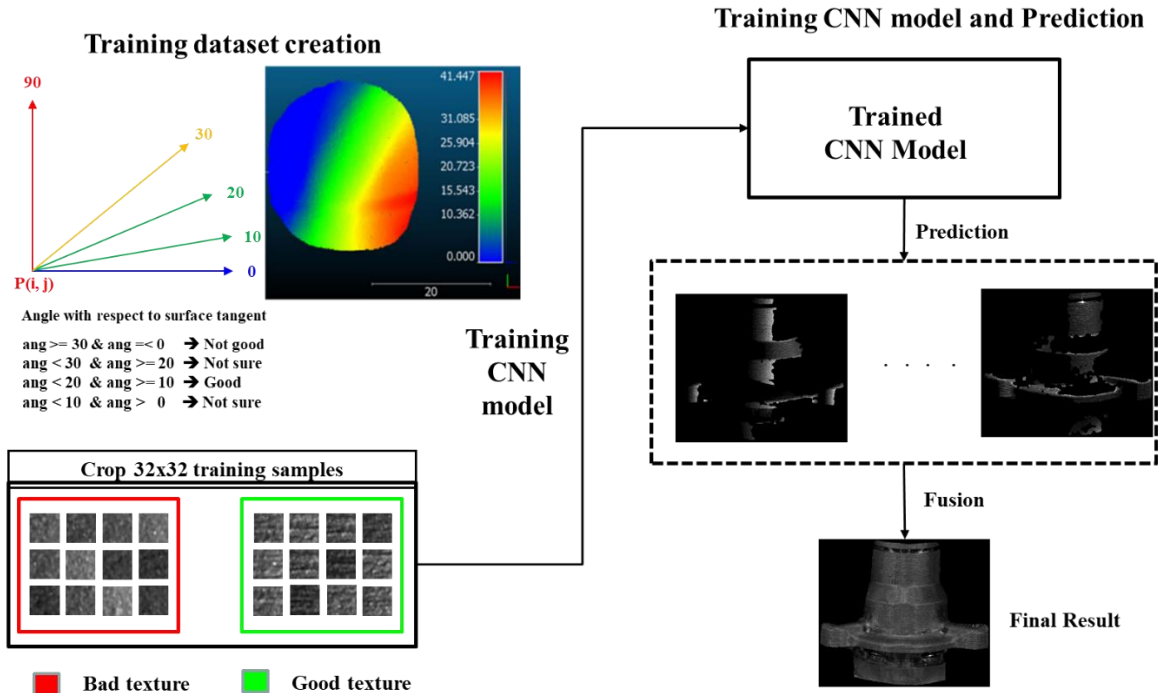
Figure 4-17. Training data generation for image fusion using the CNN model.

Each image (view) is fully scanned with a window of size 32x32 and an overlap of 8 pixels. Each crop of 32x32 is passed through the CNN to decide whether this selected area is good or no. Because of the overlapping, many pixels are selected many times and can have different labels (with some crops classified as good whereas others are not). The

final decision for such ambiguity is to classify each pixel as good if it was classified at least one time as good during the scanning/classification of the whole image.

4.4.5 Object-based fusion

In this method instead of fusing images in the image space, for each LED, n different point clouds are produced and combined to form the final model. As shown in Figure 4-18, for each LED, a stereo pair is captured using the proposed image acquisition prototype. Then the images are automatically inspected using the advantage of known geometry of lighting system and the 3D shape of the object, and only those regions with the best and favourable grazing angles are selected (For more detail please refer to Section 4.4.3). Consequently, on each image, only parts in which surface roughness and microstructures are best highlighted remain, with the rest that seems to lead to lower quality and/or noise (shadow and secularity) disregarded. However, this procedure highlights locally the surface's topography depending on the light direction, and some parts including shadow and specular reflection are removed as shown In Figure 4-18(b), the individual point clouds are incomplete. To address this issue and generate a complete 3D shape of the object, in each dataset, the individual point clouds generated under each grazing angle can

be merged into a unique 3D model shown in Figure 4-18(c). Since the selected point clouds are all oriented and registered within the same reference coordinate system, merging them to create the final point cloud is a simple process. Thus, in this way, a complete 3D object reconstruction can be achieved.
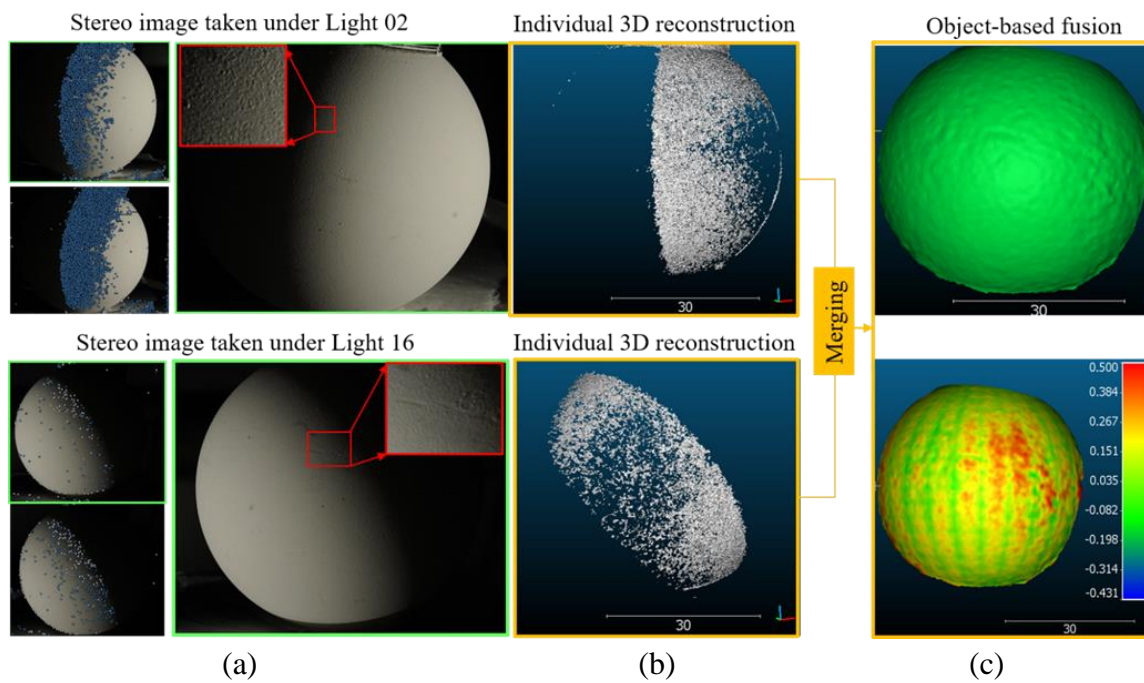
Figure 4-18. Object-based fusion.

Samples of stereo images taken under different grazing angles and a closer view that shows the highlighted roughness (a). Individual 3D reconstructions generated under favorable grazing angles using only a stereo pair (b). Result of the merged point cloud and the point-to-point comparison between the reference data and the proposed 3D reconstruction method (c).

The final point cloud might include outliers resulting from specular reflection and undesired shadows in the individual models, especially when the object has a complex shape. Therefore, in the proposed system, a statistical noise/outlier removal algorithm is applied. The noise removal algorithm can also help save memory and computational resources. The algorithm used in this paper is described in Carrilho et al. (2018). To detect the outliers, a plane is fitted to any surface point and its K surrounding points. The point is considered as an outlier if it is too far away from the fitted plane. This decision is made based on the standard deviation and the sum of the mean distance.

4.5 Image orientation and dense reconstruction of transparent objects

In this section, we first analyze the critical issues that cause image orientation failures of transparent objects, and then we propose two approaches that leverage the low-contrast textures present on object surfaces (roughness and 3D microstructures) to accurately orient images. Unlike conventional SfM-based methods that prioritize high-contrast textures, both approaches privilege tie point detection on low-contrast textures, discarding specular reflections and static tie points. For the first approach, local descriptors are extracted in those regions where roughness and micro-structures are better highlighted, applying the normalized cross-correlation (NCC) on the gradient map of the images to fully

exploit the geometrical content of the patches. The second approach builds on the first method adapting the classic SIFT pipeline and obtaining a faster and more reliable approach. Finally, given the oriented images, several methodologies are used to generate a dense reconstruction of the testing objects.

4.5.1 The main challenges of transparent surfaces

During our investigations, we analyzed the results of COLMAP (Schonberger and Frahm, 2016) and Agisoft Metashape[2], which failed at the image-matching stage, thus no camera parameters could be recovered as shown in Figure 4-19.

Figure 4-19. Incorrect image orientation in Metashape and COLMAP.

---

[2] . https://www.agisoft.com/

The reason is that SIFT-like descriptors prioritize image patches with high contrast, e.g. textures along the silhouette of the transparent object (Figure 4-20a), or reflections coming from surrounding objects and light sources (Figure 4-20b), and discard all the low-contrast textures.

(a)   (b)   (c)
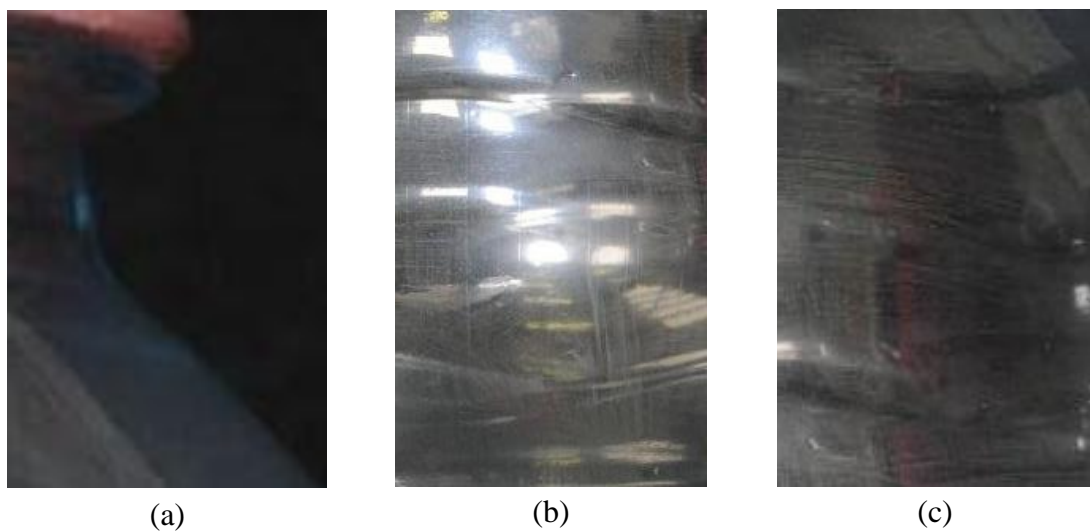
Figure 4-20. The challenges of transparent surfaces.

Silhouette regions (a) and specular reflections (b) appear almost constant in different images. Low contrast textures areas (c).

If the camera position and the surroundings remain constant and only the object rotates, the reflected textures remain quite steady from one image to the next if the object is a solid of revolution (or the surface of the surveyed object changes slowly). In this

106

situation, the descriptor finds matches in the same position across all images, assuming that the object stayed consistent. Conversely, if the object is not a solid of revolution, a patch that is first located on the silhouette in a successive image will be located inside the object with a completely different appearance. In both cases, the matching step of SfM generally fails. In addition, regular descriptors that prioritize high-contrast textures ignore low-contrast textures on extremely high-resolution images that roughness and microstructures (Figure 4-20c), even if they can be utilized for image matching.

4.5.2 Cross-correlation pipeline (first approach)

The detection is strongly related to the description which relies on the similarity of patches extracted on the gradient map. Therefore, the best candidate tie points should be those surrounded by discriminative regions, where there is enough gradient. Furthermore, we need to eliminate reflection regions since they usually do not move solidly with the object rotating on the turn table. Based on these assumptions, Figure 4-21 depicts the essential steps of the proposed detection pipeline with an example.

- A sequence of images is used as input. High-resolution images must be preferred to highlight surface roughness and details on the transparent surfaces.

- (2-3) Backgrounds are removed with Removal.AI (https://removal.ai/), a deep-learning tool for background removal, and converted to masks.

- (4) Masks are applied to process only image areas containing the transparent object, and especially discarding the coded targets for metric evaluations. The purpose of the paper is in fact to orient the images with only the texture of the object.
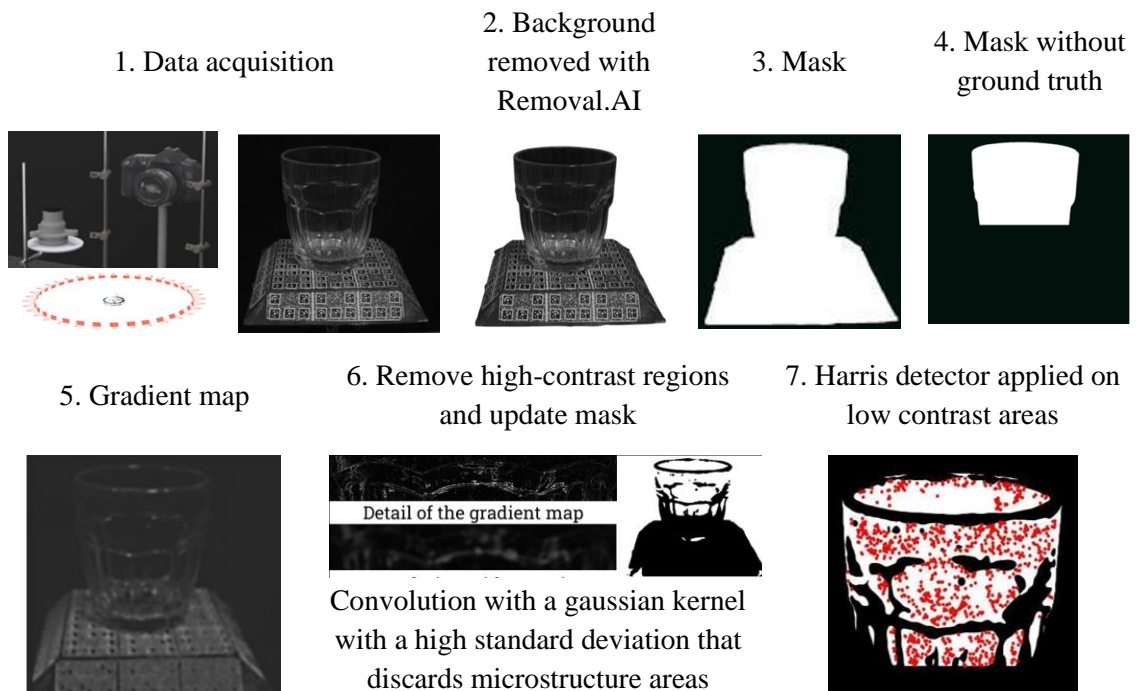
Figure 4-21. Proposed detection pipeline based on cross-correlation.

- (5-6) The gradient of the image is estimated to highlight and emphasize the geometrical content of the textures, then a gaussian kernel with a large standard deviation is applied to mask specular reflections and keep only those parts where roughness and microstructures are highlighted better.

- Finally, the Harris detector is applied to low-contrast areas to extract keypoints.

The description and matching step leverage the knowledge of the approximate epipolar lines (almost horizontal because of the acquisition network) to search for the best match of a candidate 33x33 pixel patch cropped around each keypoint (Figure 4-22a). Each patch is converted in its gradient map and compared with NCC within a rectangular searching window extracted along the a-priori known epipolar lines (Figure 4-22b). For NCC we used the formula expressed in Equation (4-30) without the window normalization since our datasets do not present scale changes.

$$S_{m1,m2} = \sum_{u=-w}^{w} \sum_{v=-w}^{w} \left[[A_{uv} - \underline{A}] \cdot [B_{uv} - \underline{B}] \right] / \left(\sigma(A) \, \sigma(B) \right) \qquad (4\text{-}30)$$

Where S is the score, $m1$ is the reference patch cropped around the keypoint, and $m2$ is one of all the possible patches inside the rectangular searching window. The patch size is $(2w + 1) \cdot (2w + 1)$, u and v are the local coordinate system with $(0, 0)$ located in

the center of the patch, A_{uv} and B_{uv} are the gradient intensity in position (u, v). $\underline{A}$ and $\underline{B}$ are the average intensity value of each patch, $\sigma(A)$ and $\sigma(B)$ are the standard deviations.

The output is a score map (Figure 4-22c) and the best match is extracted in the maximum (Figure 4-22d). For computational constraints, the matches of each image are calculated for only two images forward and two backward.

a) Patch around key point   b) Searching window on gradient map   c) NCC score   d) Patch with max score
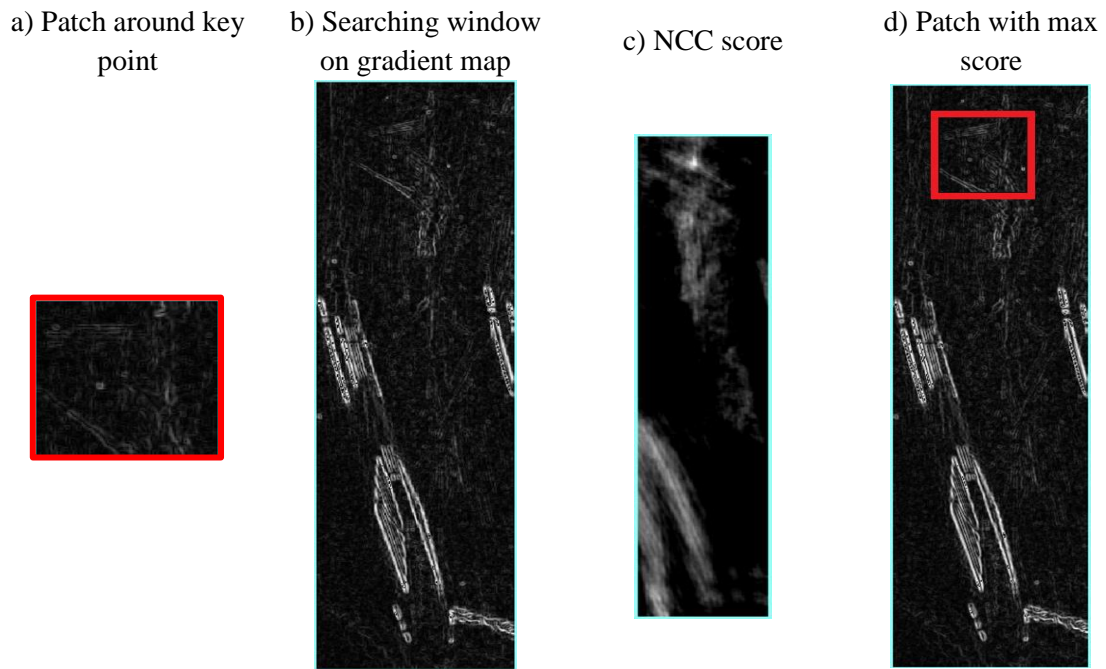
Figure 4-22. Description and matching with NCC.

(a) patches extracted around each candidate keypoint, (b) searching window on the previous or following gradient maps, (c) score map after NCC, (d) visualization of the patch around the maximum score.

110

The patches are limited to a 33x33 pixel size. The method has several hyper-parameters that have been chosen to balance computation times and final accuracy.

The raw matches are imported to be geometrically verified in COLMAP with RANSAC. All matches, both inliers, and outliers are extracted along approximate epipolar lines, therefore the RANSAC maximum error threshold was set to 1 pixel to be very restrictive and be able to extract more correct matches. Because of the simple, non-redundant acquisition network, a weak initialization might lead the solution to converge to a local minimum rather than the global maxima.

4.5.3 SIFT-based pipeline (second approach)

From the analyzes done in Section 4.5.1, it is possible to adapt the SIFT pipeline to work with low-contrast local features and discard high-contrast features for these types of datasets. To this end, this approach is based on the following changes to the standard SIFT pipeline. The adaptations and changes are mainly based on the rough knowledge of the camera motion and increasing the matching reliability.

- The pipeline was forced to extract local features also in the low-contrast areas by optimizing the "peak threshold". The contrast peak threshold was decreased from 0.066 to 0.026.

- using a not rotation invariant descriptor (the "upright" version) increases the number of absolute correct matches and the inlier ratio.

- The use of nearest neighbor strategy (NN) instead of the nearest neighbor ratio (NNR) to increase the number of matches.

- The usage of sequential matching instead of the brute-force approach. The brute-force matching without NNR can lead to consider as good matching images without overlap, causing a partial failure in the orientation. Sequential matching works in our case study with only a strip of circularly acquired images, but in other cases, a rough knowledge of the camera motion is enough to avoid failures.

- The elimination of static tie points checks the optical flow for each consecutive image pair needed to be larger than a certain threshold, e.g., 50 px (see Figure 4-23).

- Finally the usage of RANSAC as a global geometric constraint to filter outlier matches.
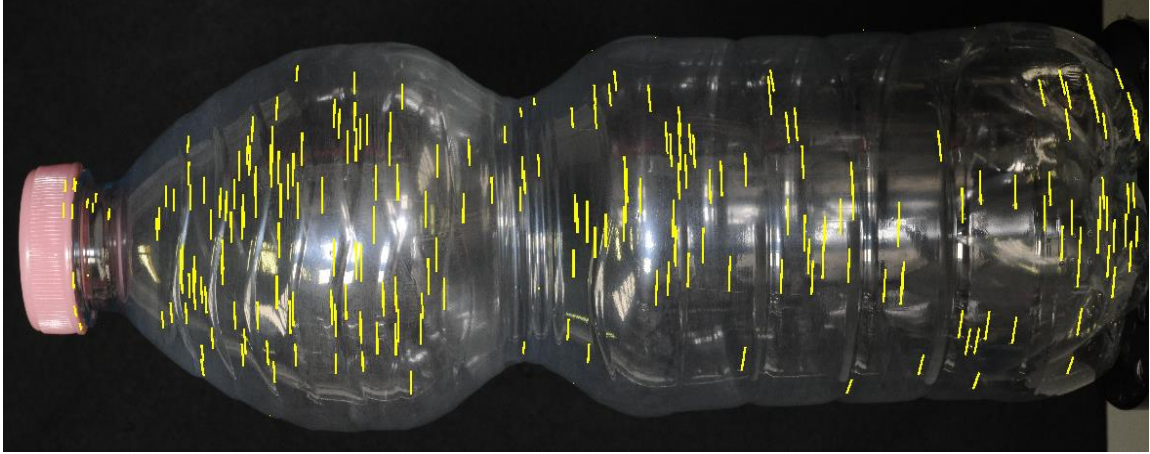
Figure 4-23. Ground-truth flux calculated from manually taken tie points.

The SIFT-based pipeline contains several hyper-parameters (see Table 4-1): the optimal value for the contrast threshold in the detection step, the threshold error for RANSAC, SIFT upright vs rotation invariant SIFT, the nearest neighborhood ratio (NNR) vs the nearest neighborhood (NN) approach, intersection vs union strategy for brute-force matching (Jin et al., 2021), RGB images vs the gradient map as input for the descriptor. To set these hyper-parameters, we selected a pair of images and determined the parameters combination that generates the higher number of correct matches with a high inlier ratio. To estimate the number of correct matches, we employed a reference fundamental matrix calculated from a set of manually selected tie points.

To further filter out the incorrect matches that lie along the epipolar lines, we also calculated the approximate flux from the manually taken tie points (yellow lines in Figure 4-23) and we interpolated the data to estimate the optical flow for all the areas covered by the bottle. This approximate ground-truth flux has been used to filter out the outliers along the epipolar lines, not filtered by the fundamental matrix. Table 4-1 reported the results for different sets of hyper-parameters in terms of absolute correct matches and inlier-ratio. The parameters to be evaluated are reported in bold, while the best parameters are highlighted in green, which are: 0.0026 for the contrast (peak) threshold, NN and not NNR with the intersection strategy for the brute-force matching, SIFT upright as a descriptor, and RGB images used as input for the descriptor.

Table 4-1. Fine-tuning of the hyper-parameters of the SIFT-based pipeline.

HYPER-PARAMETERS						MATCHES			RESULTS	
input	RootSIFT	NNR	NNR strategy	gradient peak	RANSAC error	matches after NN	static filtering	after RANSAC	correct matches	inlier ratio
RGB	upright	1.00	intersection	0.0026	**4 px**	5469	228	165	124	0.75
RGB	upright	1.00	intersection	0.0026	**2 px**	5469	228	155	122	0.79
RGB	upright	1.00	intersection	0.0026	**1 px**	5469	228	132	112	0.85
RGB	upright	1.00	intersection	**0.0036**	1 px	4565	205	117	97	0.83
RGB	upright	1.00	intersection	**0.0026**	1 px	5469	228	132	112	0.85
RGB	upright	1.00	intersection	**0.0016**	1 px	6764	249	143	118	0.83
RGB	upright	1.00	intersection	**0.0066**	1 px	2673	118	66	50	0.76
RGB	upright	**1.00**	intersection	0.0026	1 px	5469	228	132	112	0.85
RGB	upright	**0.90**	intersection	0.0026	1 px	1563	109	89	76	0.85
RGB	upright	**0.80**	intersection	0.0026	1 px	768	56	51	43	0.84
RGB	upright	1.00	**intersection**	0.0026	1 px	5469	228	132	112	0.85
RGB	upright	1.00	**union**	0.0026	1 px	25430	698	217	157	0.72
RGB	**upright**	1.00	intersection	0.0026	1 px	5469	228	132	112	0.85
RGB	**no rotation**	1.00	intersection	0.0026	1 px	5673	158	97	79	0.81
RGB	upright	1.00	intersection	0.0026	1 px	5469	228	132	112	0.85
gradient	upright	1.00	intersection	0.0026	1 px	4087	201	86	65	0.76

4.6 Summary

In this chapter, four different methodologies for 3D reconstruction of non-collaborative objects were developed utilizing the proposed multi-view photometric stereo data acquisition. The first method uses geometric construction to combine photogrammetry and photometric stereo, whilst the second approach employs FFT filtering to combine both techniques in order to generate a precise and high-detailed 3D reconstruction of non-collaborative objects. Then, we utilized the PS lighting system to highlight microstructure and roughness to enhance image orientation and dese reconstruction. Finally, we presented two different image orientation solutions for transparent objects.

CHAPTER V

Experiments and results

5.1 Introduction

In this chapter, the objective is to report experiments and results achieved for each proposed method, in particular: the findings for the proposed integrated technique (Section 5.4) and FFT-based fusion approach (Section 5.5), the results regarding the proposed pipeline obtained through our PS lighting system to improve the quality and reliability of dense reconstruction of non-collaborative objects (Section 5.6) and, finally, the outcomes for the image orientation and 3D reconstruction of transparent objects (Section 5.7). The objects used in our experiments are presented in Section 4.2 along with their ground truth in Section 5.3.

5.2 Testing objects

In this thesis, various objects with different surface characteristics including textureless, reflective, and transparent are used to evaluate different proposed approaches.

5.2.1 Textureless objects

Figure 5-1 shows various texture-less objects used to evaluate the proposed pipeline featuring complex geometry, poor texture, and/or diffuse that scatter incident illumination

117

evenly in all directions, with almost no specular highlights in the images. The first object

is a white paper (Object A), which has been glued to a flat aluminum surface. Object B is

an industrial sphere with a known radius of 25 mm±0.0134 mm which was provided by

IAPG – Jade University (Oldenburg, Germany). Object C is a standard ping pong ball with

a radius of 20mm with an uncertainty measurement of 0.1mm. The surface of these two

objects is exceptionally smooth and featureless. Finally, object D is a flat wood with a

slightly better texture than the others. Moreover, its surface is quite rough, which helps us

better understand the grazing angle's ability to improve the dense matching process by

highlighting microstructures (Section 5.6).

Object A	Object B	Object C	Object D

110*80mm	50*50 mm	39*39mm	70*70mm
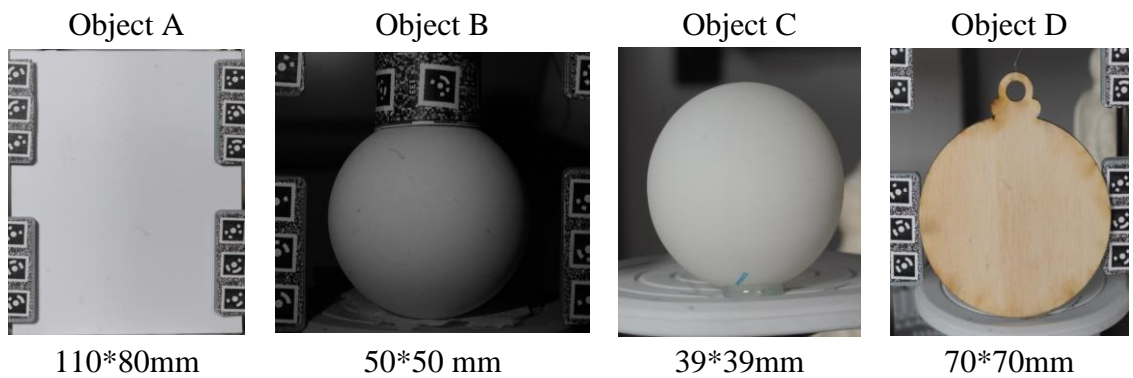
Figure 5-1. Textureless Objects.

Example of surveyed objects with non-collaborative surfaces featuring textureless.

5.2.2 Shiny and metallic objects

Various metallic and reflective objects (Figure 5-2), with complicated geometry, are used for the assessment. Objects E, F, and G are metallic with reflectivity while featuring a geometrically complex shape. Objects H and I are also two reflective objects with a curved and flat shapes. The phenomenon of interreflection makes it more complex compared to the other objects.

Object E	Object F	Object G	Object H
50*50mm	140*50mm	120*70mm	92*92mm
Object I	Object J	Object K	Object L
200*160mm	100*100mm	26*26mm	26*26mm
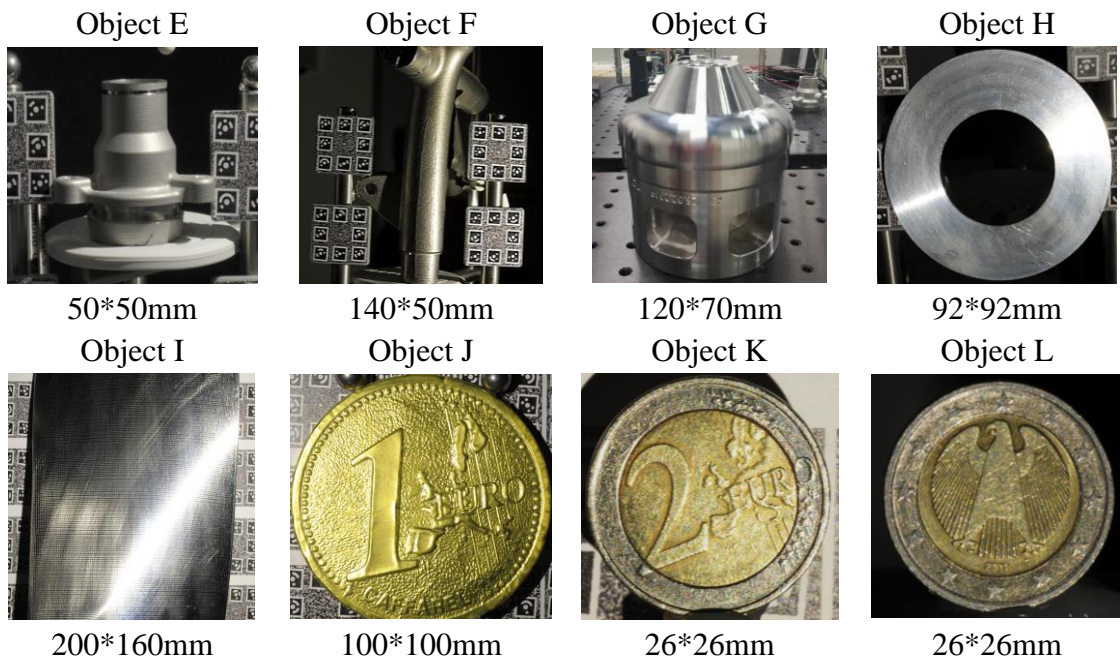
Figure 5-2. Metallic and reflective objects.

Object J is a gold foiled surface shaped like a Euro coin and objects K and L are the back and front sides of a two-euro coin. The surface of objects is very reflective, with very detailed structures. These objects are good examples to emphasize the proposed method's capability for recovering microstructures on the surface while keeping the low-frequency information.

5.2.3 Transparent objects

To evaluate the accuracy of the proposed pipeline for image orientation and 3D reconstruction of transparent objects, various objects shown in Figure 5-3 were used.

Object M	Object N	Object O
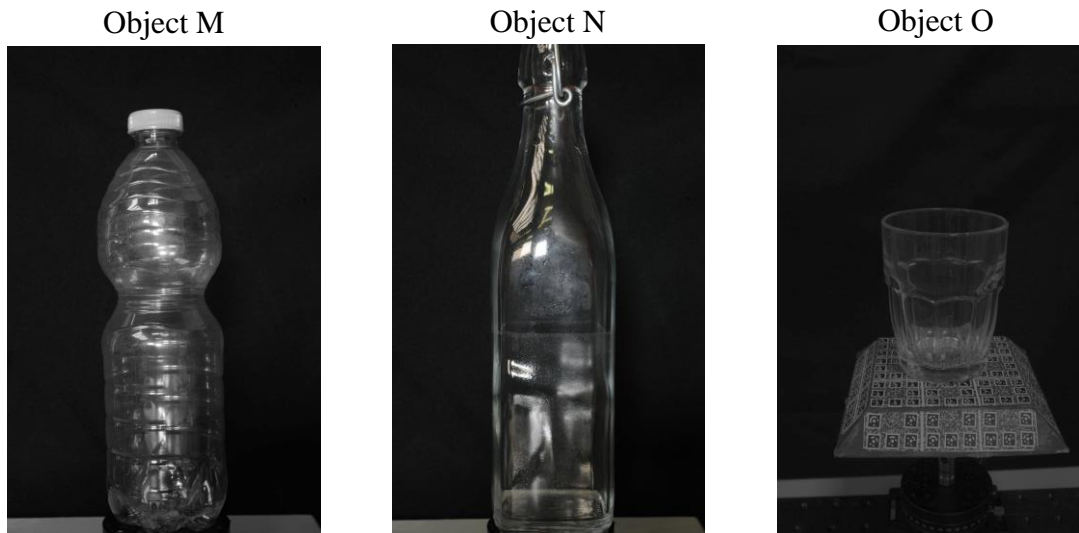
Figure 5-3. Transparent objects.

For each object, a set of images was captured using our proposed image acquisition system (Chapter 3) from 36 different stations under constant lighting, directed to enhance the appearance of the low-contrast texture. To provide a metric assessment for the proposed method performance, a set of photogrammetric coded targets with known relative distances were printed and mounted on a rotating table.

5.3   Ground truth

5.3.1 Low-frequency ground truth

In order to provide reference data in the low-frequency domain (Figure 5-4), a Hexagon active scanner called AICON Primescan (Hexagon, 2020) with a nominal accuracy of 63µm was used to scan the objects E, F, G, and I. Also, an Evixscan 3D Fine Precision (Evixscan, 2022) with a special resolution of 20µm was used to scan object K.

Since the object's laser scanner 3D model was unavailable for object J, an additional photogrammetric 3D reconstruction is employed as reference data since its low-frequency information is still accurate. To generate this dataset 30 additional images were taken (Karami et al., 2021).

To generate reference data for objects M and N, their surfaces were covered with a thin layer of random colored powder to i) make it diffusely reflecting and remove refraction and ii) provide texture on the surface. After surface treatment, an additional photogrammetric 3D reconstruction was employed to generate a dense 3D reconstruction (considered as reference data).

The reference for objects A, B, C, D, and H was provided by a geometric constraint (Mohammadi et al., 2021; karami et al., 2022b). For example, for objects A, D, and H, a best-fit plane was used as the reference, whereas for objects B and C, a sphere with a known radius was used. Indeed, according to the international table tennis federation (https://www.ittf.com/), the maximum allowed manufacturing tolerance for a ping-pong ball (Object C) radius is ± 0.1 mm, which we assumed as its known uncertainty in this study.

The reference data generated using active/passive scanners were used to evaluate the accuracy of the point cloud, while the reference data provided by geometric constraint was used mainly to evaluate the precision at a local scale.

The generated 3D points were aligned to the reference data using the ICP technique (Besl and McKay, 1992) and the RMSEs of the Euclidean distances were measured and

compared to analyze low-frequency information. This geometric comparison allows estimating possible global deformations of the recovered 3D shapes.

Object E	Object F	Object G	Object I

Object J	Object K	Object M	Object N
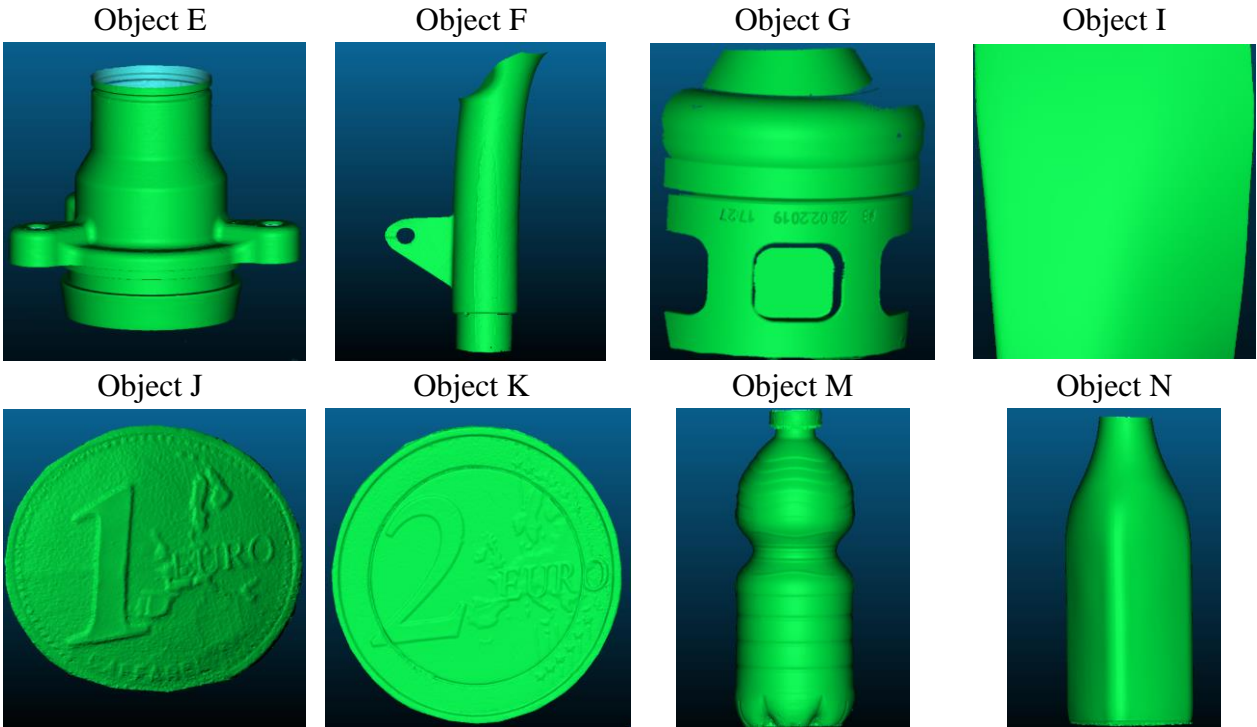
Figure 5-4. Generated reference data for low-frequency evaluation.

5.3.2 High-frequency ground truth

Objects K and L were used to evaluate the proposed methodology in the high-frequency domain. These objects can be a good example to emphasize the proposed fusion

method's capability for recovering microstructures on a surface while keeping the low-frequency information since its surface is very reflective, with very detailed structures. A small patch (2.5mm*3mm) on Object K (Figure 5-5a) was selected and measured with a non-contact 3D optical SENSOFAR scanner (Figure 5-5b) at 0.5µm resolution which was provided by IAPG – Jade University (Oldenburg, Germany). The selected patch (letter R) was completely scanned using 12 acquisitions with overlap (Figure 5-5c&d).
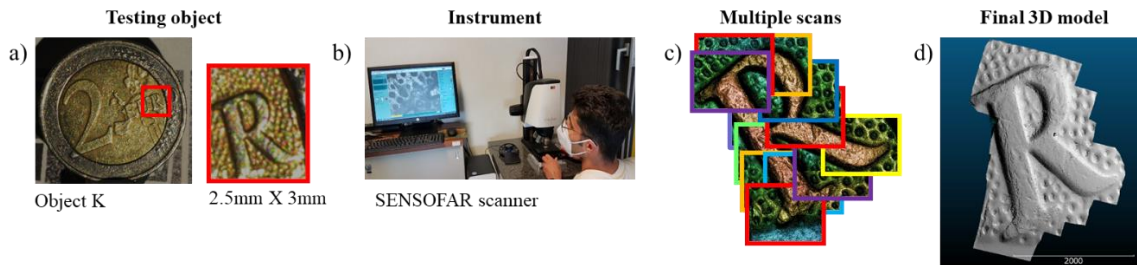
Figure 5-5. Reference data for high-frequency evaluation.

*(a) A small area (letter R with the dimension of 2.5 mm*3mm) on a two Euro coin (Object K) was selected and measured. (b) The SENSOFAR scanner with an optical resolution of 0.5µm and (c) the 12 overlapping scans to form the entire letter (R). (d) Final 3D model after alignment and co-registration of the scans.*

A contact-type profilometer (Mitutoyo, Surftest SJ-210 [mm]; R2µm; 0,75mN; Item number: 178-560-11D) was used to collect reference data for Object L. The profilometer has a diamond stylus of radius 2µm and is used to measure several profiles

(red lines in Figure 5-6) on the surface of object L to provide an accurate high-frequency profile.
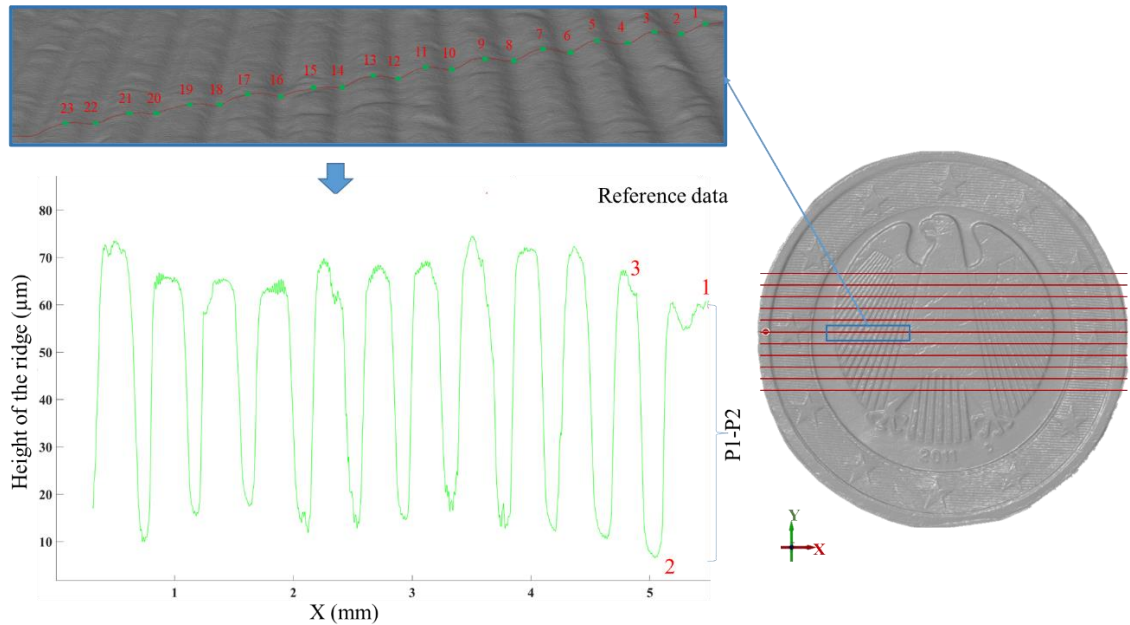
Figure 5-6. High-frequency sections scanned using a contact-type Mitutoyo profilometer.

5.4 Proposed integrated method

The proposed integrated method (Figure 4-2) was evaluated on the six non-collaborative objects (E, F, I, J, and K) presented in Figure 5-2. For each object, a set of images with a GSD of ≈20 µm are acquired from three stations using the proposed image

acquisition system (proposed in Chapter 3). From each station, multiple images are acquired under twenty different illuminations.

A 3D point cloud with a photogrammetric pipeline was generated using images taken from three different stations. The 3D coordinates of LEDs (as obtained during system calibration) and 3D object shape were then utilized to estimate light directions and intensity attenuations at each surface point. Regions with shadows and specular reflections were detected and masked out from the captured images, given the estimated light directions and the initial normal at each surface point. Following that, the surface normal was computed at each surface point given the light directions and intensities (only multiple images from the first station were used to generate surface normal). The depth map was then generated from the integration of the surface normal. Using the interior and exterior orientation camera parameters, the estimated depth map was transformed to the same coordinate system where the photogrammetric 3D point cloud was reconstructed. The scale factor was computed using corresponding points between photogrammetric 3D reconstruction and the refined photometric stereo depth map. Finally, the three different methods were proposed (Method A, Method B, and Method C) described in Section 4.2 to adjust the remaining global deformation of the estimated photometric stereo depth map. Figure 5-7 reported some examples of the 3D results obtained using the proposed integrated method with

126

respect to those achieved using a photometric stereo implemented by Xiong et al. (2014) and standard photogrammetry (AgiSoft Metashape V1.7.6 Build 13779).

Object	(**a**) Photometric stereo	(**b**) Photogrammetry	(**c**) Integration method
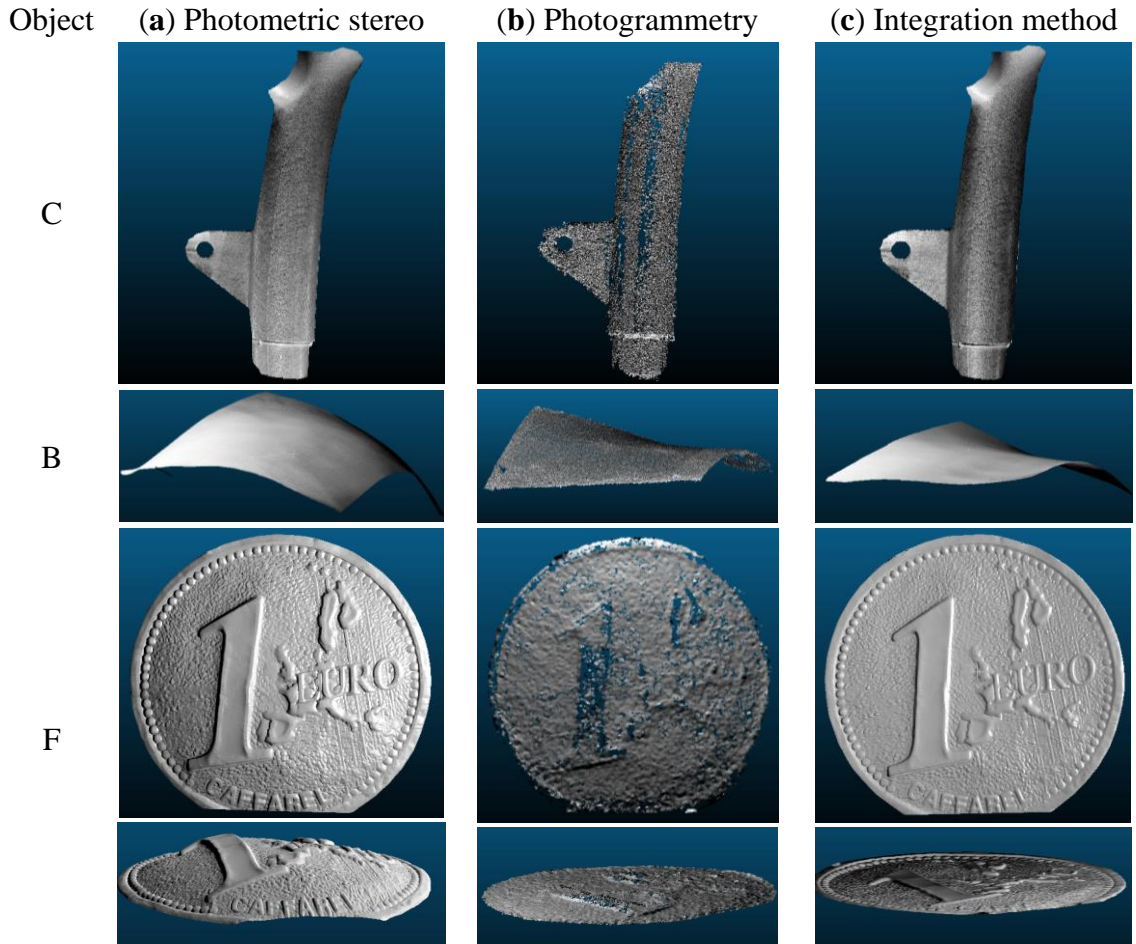
Figure 5-7. 3D point could generated using the proposed integrated method.

*3D reconstruction was generated using the basic photometric stereo (**a**), photogrammetry (**b**), and proposed integration method (**c**) on three different non-collaborative objects (F, I, and J).*

127

The obtained 3D results indicate the clear advantage of the proposed integration. The proposed approach took the advantages of photogrammetry and photometric stereo to generate a reliable and high-detail 3D reconstruction of the non-collaborative objects. Indeed, thanks to the inclusion of photogrammetric 3D measurement, the global shape deviation, caused by assumptions and unknown error resources, was greatly mitigated (Figure 5-7c). Photogrammetric 3D reconstruction (Figure 5-7b) provided accurate geometric information compared to 3D photometric stereo (Figure 5-7a), where the generated 3D reconstruction was globally deformed. The proposed integrated algorithm reduced the global shape deformation aided by photogrammetry while keeping the 3D details from the photometric stereo.

5.4.1 Low-frequency evaluation

In order to provide a comprehensive qualitative low-frequency evaluation, two different tests (cloud-to-cloud comparison and profiling) were accomplished using different non-collaborative objects.

5.4.1.1 **Cloud-to-cloud comparison**

To provide a cloud-to-cloud comparison, all the ground truth data were registered and transferred to a defined coordinate system (from which photogrammetric 3D points were obtained) using an Iterative Closest Point (ICP) technique. The RMSE of the Euclidean cloud-to-cloud distances between the 3D points on the reconstructed and reference models was then computed and compared in CloudCompare.
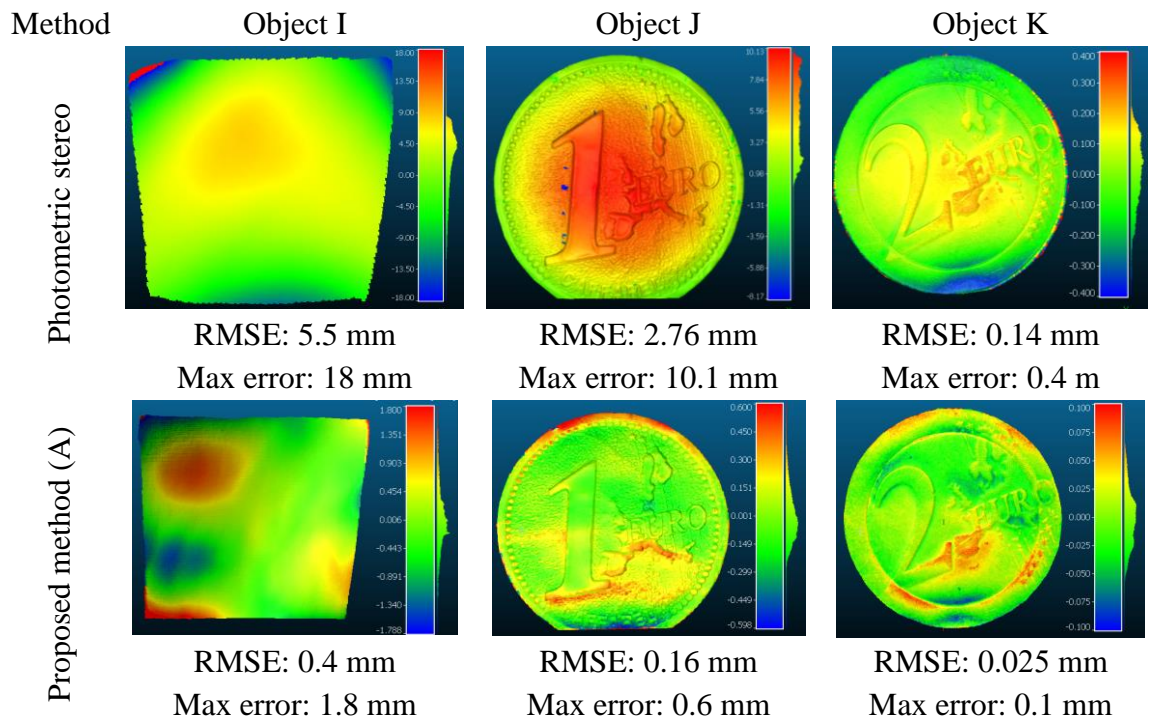
Figure 5-8. Cloud-to-cloud comparisons.

Cloud-to-cloud comparisons with reference data for basic photometric stereo and the proposed method (method A) on objects I, J, and K.

The cloud-to-cloud comparisons for the basic photometric stereo and the proposed approach (Method A) on three objects were represented in Figure 5-8.

It can be seen that the highest low-frequency error belongs to object I with an RMSE of 0.4 mm. This is because the object's size is quite large with a complex shape and a highly reflecting surface, making the 3D reconstruction challenging. However, there is a dramatic improvement compared to photometric stereo (RMSE of 5.5 mm). The low-frequency error for the rest of the objects is less than 0.2 mm proving that the proposed integration method can reduce the global shape deformation of 3D reconstructions.

The larger errors observed in the boundaries of object I, as shown in Figure 5-8, were likely due to the absence of control points in those areas. The lack of control points made it challenging to accurately predict and adjust the 3D model outside of the control point regions using the polynomial model. Furthermore, another downside of performing polynomial adjustment globally (Method A), as previously stated, was that the edges and boundaries of complex-geometry surface objects (e.g., objects E and F) can be smoothed out. Therefore, to preserve the edges of objects with complicated geometry, methods B and C were proposed. The proposed methods were tested on two objects with complex geometry (objects E and F). The comparative 3D results after final shape correction using three proposed methods were presented in Figure 5-9.

Object	Method A	Method B	Method C
E	RMSE: 0.54 mm	RMSE: 0.41 mm	RMSE: 0.06 mm
F	RMSE: 0.61 mm	RMSE: 0.35 mm	RMSE: 0.09 mm
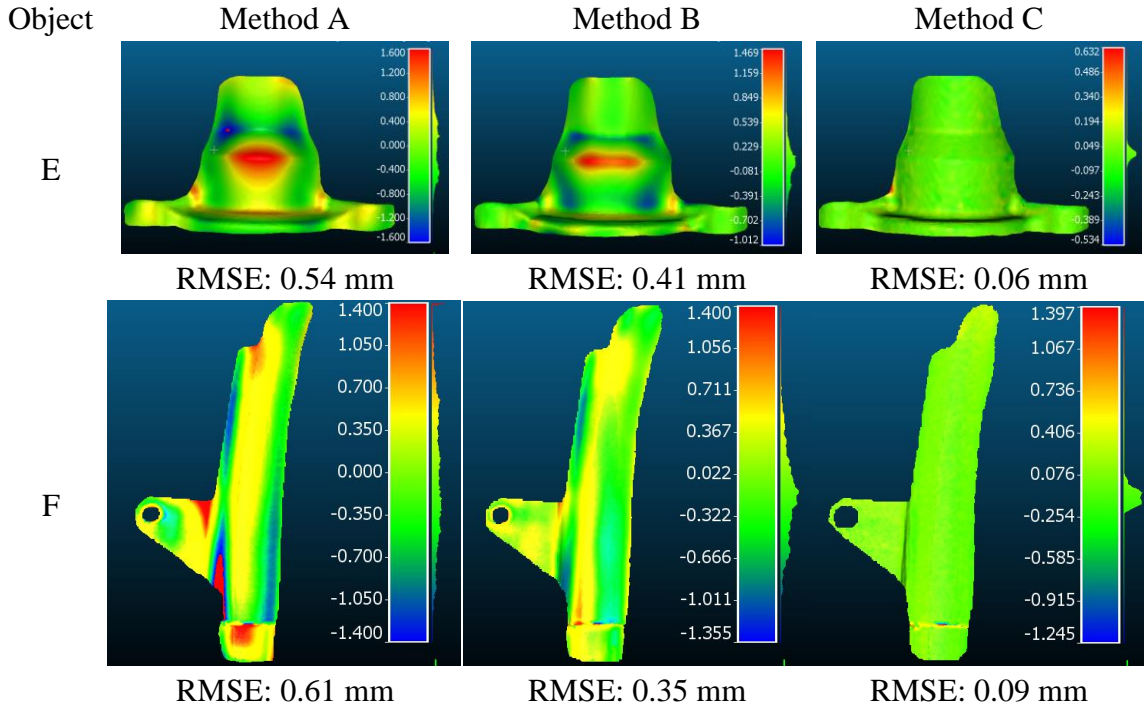
Figure 5-9. Cloud-to-cloud comparisons for the proposed methods on objects C and D.

As shown in Figure 5-9, the worst results for both objects were obtained by method A where the shape correction was applied globally at once on the object. This was due to the geometry of the objects' surfaces, which had many edges and borders; hence, applying such a polynomial can smooth out the edges and boundaries, negatively affecting the low-frequency information. However, these results were slightly improved for method B, where the object is first segmented and divided into small parts and then applied shape adjustment to each segment independently. For instance, the estimated RMSE in method A for object

C is 0.61 mm while this value for method B decreased to 0.35 mm. The disadvantage of method B is that the 3D segmentation of the object takes time. Furthermore, the segmentation outcome is constantly dependent on certain input parameters, making precise segmentation of the object problematic. The final result is directly depending on the segmentation part and therefore it is not always reliable.

The method that yielded the best results for both objects was method C, which involved dividing the object into smaller patches and applying shape deformation adjustment to each patch. The estimated RMSE was less than 0.09 mm. While this approach effectively corrected global deformation, the 3D details of the model near patch borders were dependent on patch size, an overlapping area with neighboring patches, and the number of control points used for stitching patches together. These parameters were selected manually through trial and error, but they could be further studied to automate the optimization of the final 3D model.

5.4.1.2 **Profiling**

Profiling, or the extraction of a cross-sectional profile, was another helpful criterion to evaluate the performance of the proposed method in low-frequency domain. A cross-sectional profile can display the linear route of the obtained 3D points on a perpendicular

plane which provides well-detailed geometric features of the profile. Object J was considered to be evaluated for this test. As shown in Figure 5-10 with different colors, four cross-sectional profiles were extracted and evaluated using photogrammetry, basic photometric stereo, the proposed method, and the algorithm implemented in (Peng et al., 2017). The extracted section in each dataset was geometrically compared against the photogrammetric dataset using the well-known formula of Root Mean Square Error (RMSE). It is worth mentioning that the photogrammetric 3D reconstruction was chosen as a point of reference due to its reliable low-frequency data, which could be used to validate other measurements in low-frequency domain. Additionally, the unavailability of the laser scanner 3D model for the object made the photogrammetric reconstruction the best available option for this purpose.

In Figure 5-10, the green line presents the photogrammetric cross-sectional profile while the red, blue, and magenta ones represent the basic photometric stereo, the proposed method, and the Peng approach, respectively. From Figure 5-10, it can be seen that the proposed cross-sectional profile (red line) shows smaller errors (RMSE of 0.09 mm) and is closer to the photogrammetric section (green line) compared to the other approaches.

The result of profiling (Y = 40mm) on object F
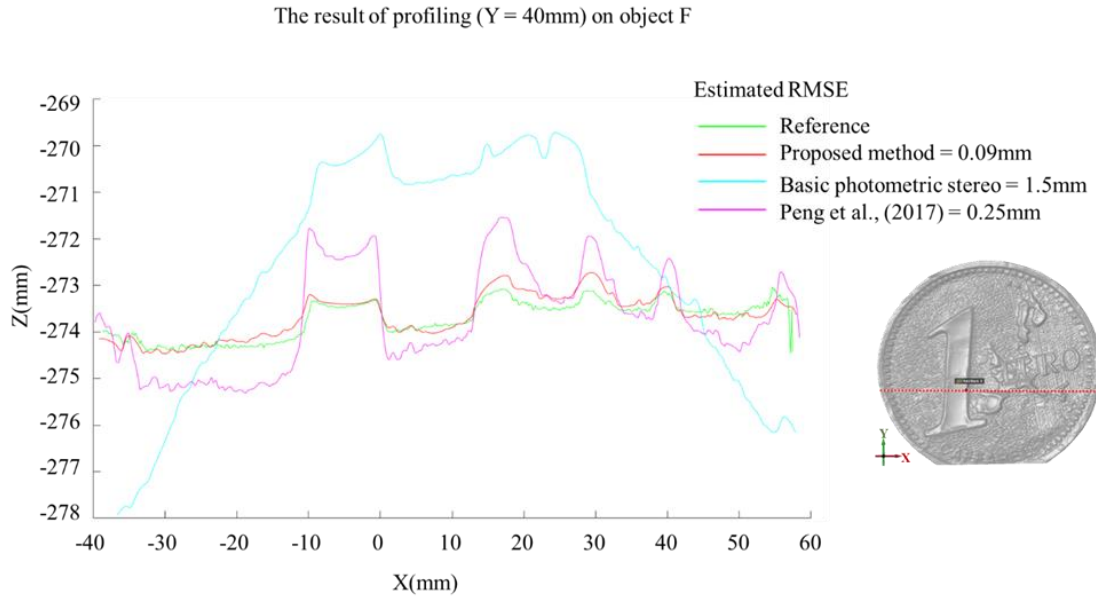
Figure 5-10. The comparison result of profiling for object J.

The green profile represents the reference data (photogrammetry), the red profile represents the proposed method, the blue represents basic photometric stereo, and the magenta profile represents the algorithm implemented by Peng et al., (2017).

5.4.2 High-frequency evaluation

In order to evaluate the accuracy of the reconstructed high-frequency information by the proposed methodology, the obtained 3D results are compared against reference data collected with a contact-type profilometer (Mitutoyo, Surftest SJ-210 [mm]; R2µm; 0,75

mN; Item number: 178-560-11D). The profilometer was used to measure a profile on the surface of object K to provide an accurate high-frequency profile. Then, the reference profile (the green profile in Figure 5-11) is compared to the same profile generated on the 3D data obtained by the proposed method (the red profile shown in Figure 14). The height of ridges on both extracted profiles is measured and compared. The results of this comparison are shown in Tables 4-1, 4-2, and 4-3.
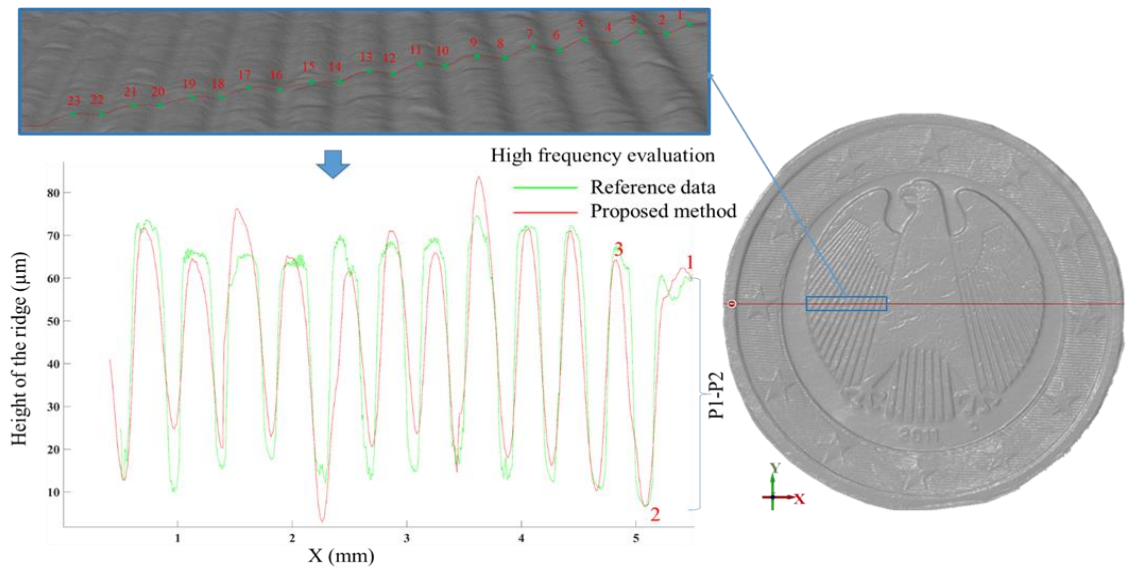
Figure 5-11. High-resolution evaluation using profiling.

High-resolution evaluation between the profiles measured with the profilometer in green and the proposed method in red for object E.

135

From the achieved results, it can be seen that the estimated ridge heights for the proposed method is quite close to the ground truth provided by the profilometer. For instance, a maximum residual of 15.36 µm is measured for ridge P14-P15 while the estimated RMSE and Mean Absolute Error (MAE) for all ridges are about 1.5 µm and 5.48 µm, respectively. The achieved results generate a highly-detailed 3D reconstruction of the surface topography, with a high level of agreement with the ground truth.

Table 5-1. The estimated residuals of the ridge height (µm) between the proposed method and reference data from point 1 to point 13.

	P1-P2	P2-P3	P3-P4	P4-P5	P5-P6	P6-P7	P7-P8	P8-P9	P9-P10	P10-P11	P11-P12	P12-P13
Reference	53.79	60	54.03	59.73	58.53	58.03	56.03	58.63	57.7	51.8	52.85	52.05
Proposed	55.683	57.523	53.52	60.4	54.46	54.63	53.11	62.07	65.59	51.35	41.55	46.61
Residual	1.893	−2.477	−0.51	0.67	−4.07	−3.4	−2.92	3.44	7.89	−0.45	−11.3	−5.44

Table 5-2. The estimated residuals of the ridge height (µm) between the proposed method and reference data from point 13 to point 23.

	P13-P14	P14-P15	P15-P16	P16-P17	P17-P18	P18-P19	P19-P20	P20-P21	P21-P22	P22-P23
Reference	55.56	56.56	56.8	52.77	46.97	47.7	49.8	49.4	55	63.4
Proposed	50.44	41	58.95	61.57	41.24	52.39	55	44.68	41.99	48.72
Residual	−5.12	−15.36	2.15	8.8	−5.73	4.69	5.2	−4.72	−13.01	−14.68

Table 5-3. The results of high-frequency evaluation (µm) for the proposed method.

Mean of Residuals	Maximum Residual	RMSE	MAE
−2.46	−15.36	1.5	5.48

5.5   Fusion in frequency domain

To evaluate the proposed FFT-based fusion method (Section 4.3), three non-collaborative objects (Figure 5-2 objects F, J, and K) were used. Two different experiments were carried out to evaluate the accuracy potential of the proposed fusion method. The first test was accomplished to evaluate the accuracy of low-frequencies obtained by the proposed method. The second test was performed to evaluate the potential of the proposed method to exploit the high frequencies.

5.5.1 Low frequencies evaluation

To evaluate the accuracy potential of the proposed FFT-based method in the low-frequency domain, 3D results achieved with photometric stereo (Figure 5-12a), photogrammetry (Figure 5-12b), and the proposed method (Figure 5-12c) were geometrically compared using available reference data.

After aligning the generated 3D points to the reference data, the RMSEs of the Euclidean distances were measured and compared to analyze low-frequency information. This geometric comparison allows estimating possible global deformations of the recovered 3D shapes. The results of this comparison was shown in Figure 5-12.

137

The negative values of the legend (towards blue color) in Figure 5-12 indicate that the generated 3D surface is below the reference surface, while the positive values (towards red color) show areas above the reference surface.

Object	a) Photometric stereo	b) Photogrammetry	c) Proposed fused method
F	RMSE: 2.4mm	RMSE: 0.07mm	RMSE: 0.09mm
I	RMSE: 5.4mm	--	RMSE: 0.14mm
K	RMSE: 0.1mm	RMSE: 0.02mm	RMSE: 0.04mm
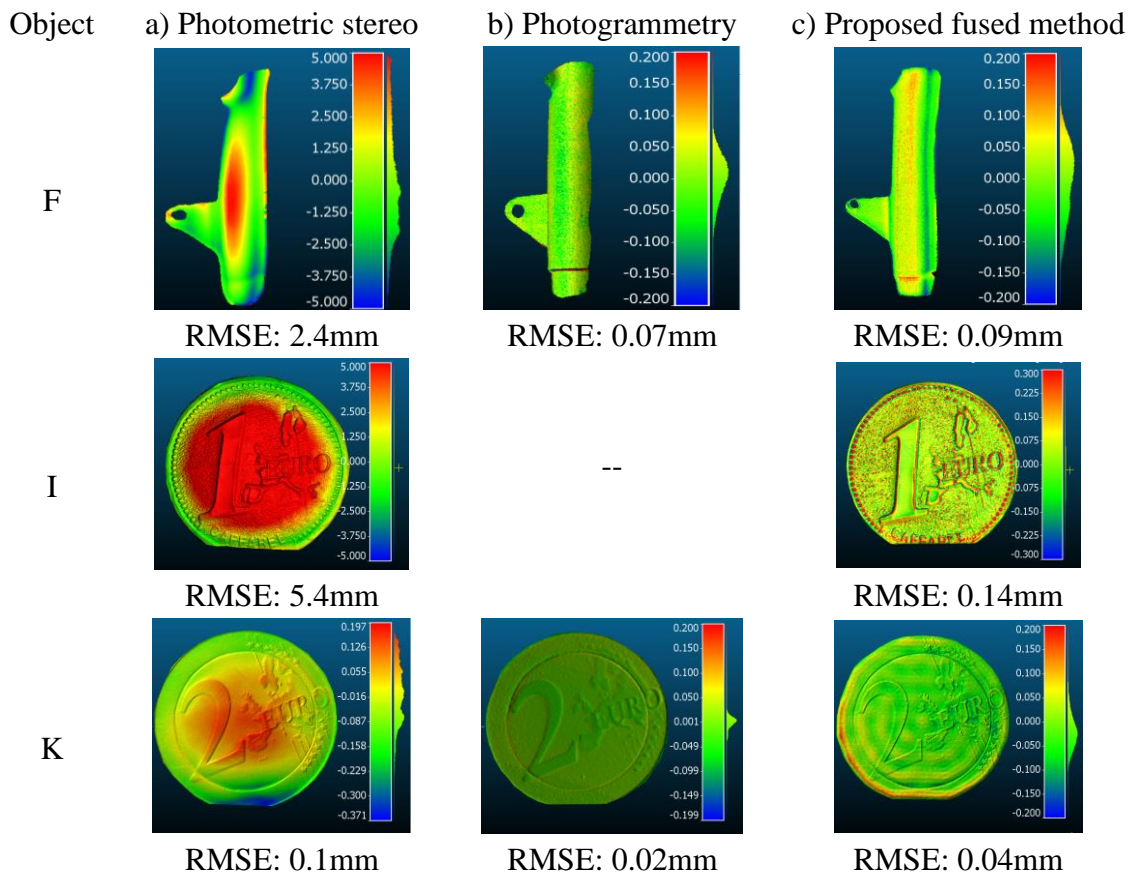
Figure 5-12. The cloud-to-cloud comparison For the FFT-based fusion method.

The result of the cloud-to-cloud comparison with reference data for photometric stereo, photogrammetry, and proposed method on three non-collaborative objects.

The quantitative analysis demonstrates that the proposed fusion method performs noticeably better than photometric stereo. For instance, the estimated RMSE of Euclidean distances for photometric stereo for objects F and J are 2.4mm and 5.76 mm, respectively. On the other hand, values for the proposed method decreases remarkably for both objects to 0.09mm and 0.14mm, respectively, which was quite close to the estimated RMSE for photogrammetry (0.07mm). This analysis demonstrates that the proposed integrated method maintained high frequencies while also improving low spatial frequencies.

5.5.2 High frequencies evaluation

Object K was a good example to emphasize the proposed fusion method's capability for recovering microstructures on a surface while keeping the low-frequency information since its surface is very reflective, with very detailed structures. A small patch (2.5mm*3mm) on Object K was selected and measured with a non-contact 3D optical SENSOFAR scanner (Figure 5-5). Then, the reference data was compared to the same patch/area on the 3D data generated by the proposed method, photometric stereo, and photogrammetry. To this end, the RMSE of the Euclidean cloud-to-cloud distances

139

between the 3D points on the reconstructed and reference models was computed and compared using CloudCompare software (Figure 5-13).

A) Photometric stereo   B) Photogrammetry   C) Proposed method

RMSE: 15µm   RMSE: 35µm   RMSE: 19µm
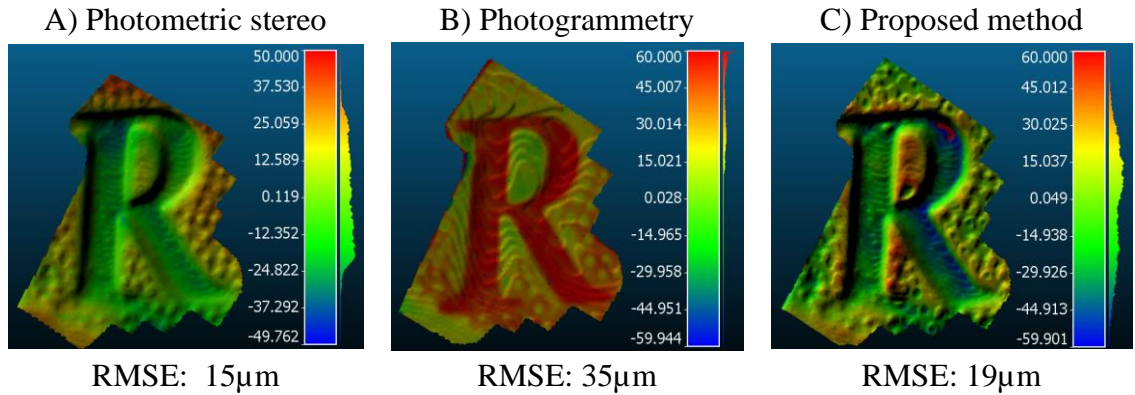
Figure 5-13. High-frequency evaluation For FFT-based fusion method.

*Evaluation results for the high-frequencies on Object K: cloud-to-cloud comparisons using reference data for photometric stereo, photogrammetry, and the proposed approach on the small 2.5mm * 3mm patch of the coin.*

From the achieved results, it can be seen that the estimated RMSE for the proposed method was less than 19 µm which was quite close to the photometric stereo with an RMSE of 15 µm. While the estimated RMSE for photogrammetry was almost two times higher (about 35µm) than the proposed method. The achieved 3D results indicated how the proposed method generates a high-detail 3D reconstruction of the surface topography quite

140

similar to photometric stereo while preserving low-frequency information (Figure 5-12) thanks to the fusion of photogrammetric and photometric stereo depth measurements.

5.6   3D reconstruction through PS data acquisition system

In this section, we aim to incorporate and quantify the effect of directional lighting, obtained through our PS lighting system, as a method to improve the quality and reliability of dense reconstruction of non-collaborative objects. The designed system employs several light sources that can better highlight roughness and microstructures on the surface (Section 4.4). The shadows and shading that occurred, due to the interaction of the grazing rays with microstructures, can produce spatially varying chiaroscuro patterns that increase the local contrast of the image thus facilitating image matching in orientation and multi-view stereo-dense reconstruction. Eight different shiny and textureless objects of varying complexity were used. The experiments, first, analyzed the effect of light directionality using different tests including density the number of extracted tie points, measuring photometric consistency of generated point cloud, and geometric evaluation of the point cloud generated under grazing angle illumination. Then, we reported the results of five different image fusion-based methods including average, median, albedo, GLCM, and learning-based to improve 3D reconstruction of non-collaborative using our PS image

acquisition system (Section 4.4.4). And finally, the 3D results of object-based fusion (Section 4.4.5) are presented.

5.6.1 Evaluating the effect of light directionality

The objective is to study and compare the quality of 3D surveys achieved under individual grazing angles against those obtained under conventional diffuse lighting. To this end, 20 stereo pairs (ground sample distance – GSD ≈ 38 μm) were acquired, each illuminated with one of the LEDs. Figure 5-14 shows some sample stereo pairs for objects C and E, respectively. The main reason to use only two stations (a stereo image) was to investigate to what extent dense image matching could be enhanced with the minimum number of required images. Otherwise, increasing the number of images (captured from different stations) might help to overcome the low signal-to-noise ratio and mitigate random and systematic errors. As the grazing angle images are acquired under different lighting conditions, the diffuse images are expected to mitigate shadow and specular effects, simulating a diffuse lighting condition needed to produce a reasonable dense cloud. To take the diffuse images, two large LED panels (about 50x50 cm2) were mounted on a mini adjustable tripod stand and positioned at about 1.5m distance from the object.

| | LED 01 | | LED 05 | | Diffuse images | |
| | Station 01 | Station 02 | Station 01 | Station 02 | Station 01 | Station 02 |
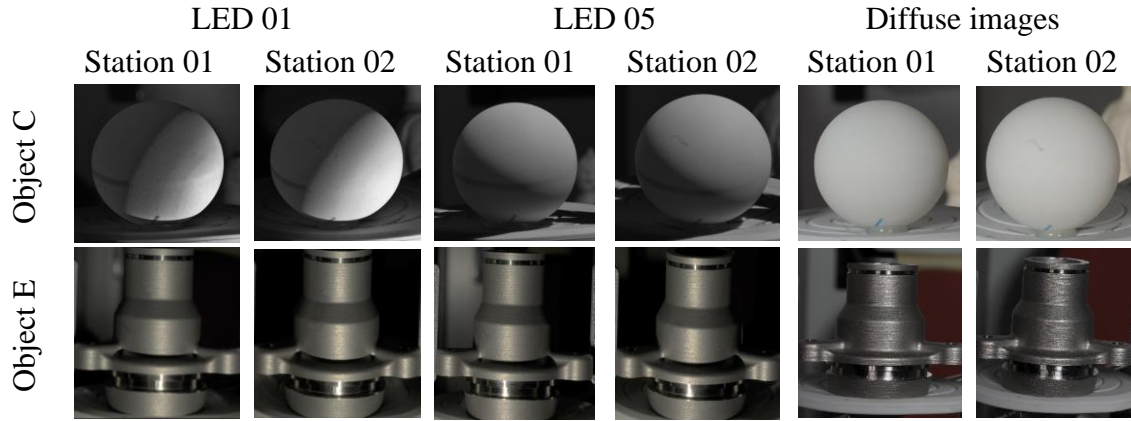
Figure 5-14. Sample grazing angles and diffuse images.

Sample grazing angles and diffuse images captured from objects C and E using the proposed image acquisition device. LED 01 and LED 05 are two examples of grazing angles.

In front of each panel, a flat light diffuser was installed to uniformly scatter the incoming light in all directions.

5.6.1.1 Density and distribution of tie points

Feature extraction is the first essential step in photogrammetric 3D reconstruction. The number of tie points extracted in diffuse and grazing angle images are hereafter compared. The goal is to assess the object texture quality in each of these images. Objects A, C, D, and E were used. For the 20 grazing angle stereo pairs, we used those illuminated

by LED lights 01, 05, 16, and 20, which, on average, provided less shadow and specular reflections on the test objects. To extract the tie points, the AgiSoft Metashape V1.6.3 Build 10732 was used with exactly the same settings for all of the objects. The number of tie points refers to the number of feature-based points extracted and matched during the image orientation process.

The results are shown in Figure 5-15, Table 5-4, Table 5-5, and

Table 5-6. The blue dots in Figure 5-15 represent tie points. As can be seen from Table 5-4 and Table 5-5, the number of tie points extracted in grazing angle images for textureless and shiny objects is significantly higher than that obtained by diffuse lighting.

Object	A	C	D	E
Grazing angle				
Diffuse lighting				
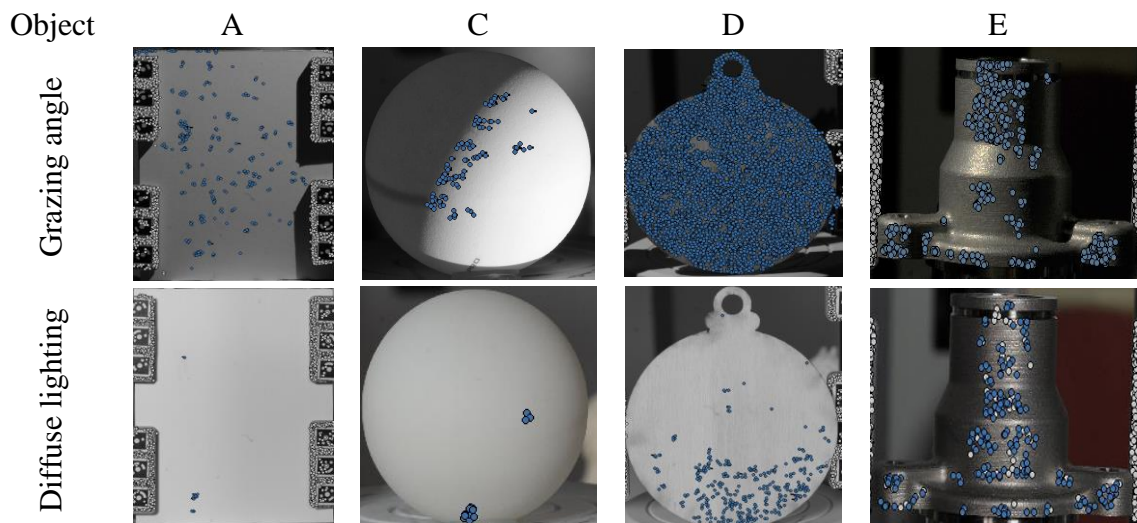
Figure 5-15. The visual impression of extracted tie points on four objects.

For example, Light 16 has the fewest tie points in grazing angle images for object D, with 4580 features, which is still far more than diffuse lighting images. The pattern holds true for the other objects as well.

Table 5-4. The extracted tie points on shiny objects.

Object	Diffuse lighting	Light 01	Light 05	Light 16	Light 20
E	259	825	581	154	569
F	1913	1875	2865	2064	1962
G	505	749	510	800	530
H	552	853	176	430	313

The extracted tie points on shiny objects using diffuse lighting, and four grazing angles. The best and worst results are represented by green and brown colors, respectively.

Table 5-5. The number of extracted tie points on textureless objects.

Object	Diffuse lighting	Light 01	Light 05	Light 16	Light 20
A	21	137	304	166	167
B	1061	1536	1298	1271	1710
C	19	106	26	28	32
D	365	4642	5422	4580	5840

Number of extracted tie points on textureless objects using diffuse lighting, and four grazing angles. The best and worst results are represented by green and brown colors, respectively.

Table 5-6. The estimated RMS of image residuals (pix).

Object	Diffuse lighting	Light 01	Light 05	Light 16	Light 20
A	1.1	0.2	0.2	0.18	0.2
B	1.0	0.4	0.45	0.4	0.48
C	0.6	0.3	0.6	0.2	0.61
D	0.51	0.19	0.2	0.19	0.19
E	0.3	0.11	0.13	0.12	0.13
F	0.21	0.2	0.2	0.12	0.15
G	0.2	0.2	0.19	0.16	0.18
H	0.3	0.2	0.3	0.15	0.29

The estimated RMS of image residuals (pix) averaged over all tie points on a stereo image. The best and worst results are represented by green and brown colors, respectively.

Table 5-6 shows that the estimated RMS of reprojection errors for grazing angles were substantially lower than diffuse images. This may highlight the fact that the image matching and orientation in the dataset created utilizing grazing angles is far superior to the other due to the higher contrast and stronger signal in the corresponding images. A higher reprojection error for generated tie points may indicate the presence of a problem in

the project, such as a low-quality surface texture or an insufficient number of tie points incorrectly detected.

Overall, from Table 5-4, Table 5-5, and

Table 5-6, it can be observed that there was at least one grazing angle image for each object, which resulted in significantly more tie points with lower error than diffuse images.

5.6.1.2 Photometric consistency

The Photometric Consistency Score (PCS) is defined by the normalized cross-correlation value of pixels across corresponding images in which the point can be seen (Furukawa and Ponce, 2009). In this experiment, PCS is used to assess the quality of extracted correspondences across multiple input images using the open-source PMVS2 software. The generated 3D points are projected into each visible image, and the similarity of image textures near their projections is estimated. PCS is a scalar value that can be associated with any 3D point and ranges from -1 (worst) to 1 (best). The higher the value, the more reliably the corresponding match is found across the different images.

Visual results for objects A, C, D, and E, are reported in Figure 5-16, with dark blue color for a PCS of -1 and dark red color for a PCS of +1.0. The point clouds created using grazing angle images present significantly higher PCS for all objects than diffuse images.

Table 5-7 confirms this, where the grazing angle images have not only contained more points, but the majority of the extracted points also have higher scores. For example, for object D, more than 800 thousand points were extracted using grazing angle images, with PCS values mostly bigger than 0.7. A similar pattern can be seen for other objects.
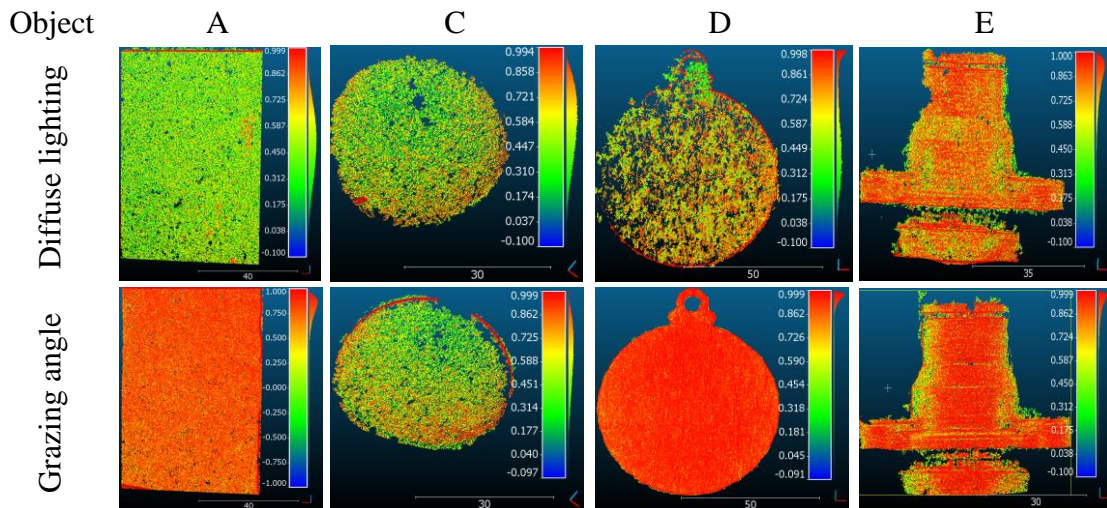

Figure 5-16. Comparison of photometric consistency.

Comparison of photometric consistency of points produced using different image types. The unitless photometric consistency values range between -1.0 (worst) to +1.0 (best).

148

It should be noted that, with the complex shape of object E, the total number of extracted points by grazing angle images is slightly less than that of diffuse lighting images although grazing angle images outperformed diffuse lighting images in terms of PCS. This could be due to occlusions, which prevent light from reaching the covered areas. Of course, increasing the number and distribution of light sources is one solution to this problem.

Table 5-7. Photometric consistency score of 3D points.

Object	Image	Consistency groups				Total points
		0.7 to 1	0.4 to 0.7	0 to +0.4	-1 to 0	
A	Diffuse lighting	68380	282800	170900	3287	525367
	Grazing angle	481700	132500	10410	0	624610
C	Diffuse lighting	29330	54150	24180	394	108054
	Grazing angle	42170	53020	16700	232	112122
D	Diffuse lighting	165000	67890	27260	1240	261390
	Grazing angle	851200	3861	0	0	855061
E	Diffuse lighting	205800	46720	10150	0	262670
	Grazing angle	232000	20160	0	0	252160

The results of photometric consistency on four different objects. The best and worst results are represented by green and brown colors, respectively.

5.6.1.3 Geometric evaluation of individual models

The acquired and produced images were used to create individual models used for evaluations. To evaluate the accuracy of the point clouds, we first register the generated

3D point cloud to the ground truth (GT). The Euclidean distances of each point from the ground truth surface are then considered as an error. We report the Root Mean Square of Errors (RMSE) according to the definition also given in Gruen and Beyer (2001). Figure 5-17 shows the results, with a negative (blue) or positive (red) value indicating whether the generated point is below or above the closest point on the reference mesh surface, respectively. Table 5-8 reports the estimated RMSE (mm) of the point-to-point comparison between the generated point clouds and the reference mesh for both shiny and textureless objects.

As illustrated in Figure 5-17, the grazing angle point cloud fit very well to the reference data, particularly for smooth and textureless objects (A and C), whereas diffuse lighting images produced very noisy results. Furthermore, the estimated RMSE of points reconstructed using grazing angle images ranges between 0.059 mm and 0.15 mm, depending on the surface type and the direction of the light source, as shown in Table 5-8. The grazing angle concept performed best on objects A and D with an RMSE of less than 0.1 mm. Objects C have slightly higher values (0.15 mm and 0.2 mm, respectively) than the other objects. This could be due to occlusions, which occur when light is unable to interact properly with an object's surface.

The diffuse lighting method produced a very noisy point cloud with an RMSE of 0.25 mm for object C and 0.17 mm for object A, while these values for grazing angle images, were less than 0.1 mm and 0.15 mm for objects A and C, respectively.

It also should be noted that when the object is extremely polished, smooth, and reflective (e.g., object G), the grazing angle model RMSE is slightly higher but still better than the others.
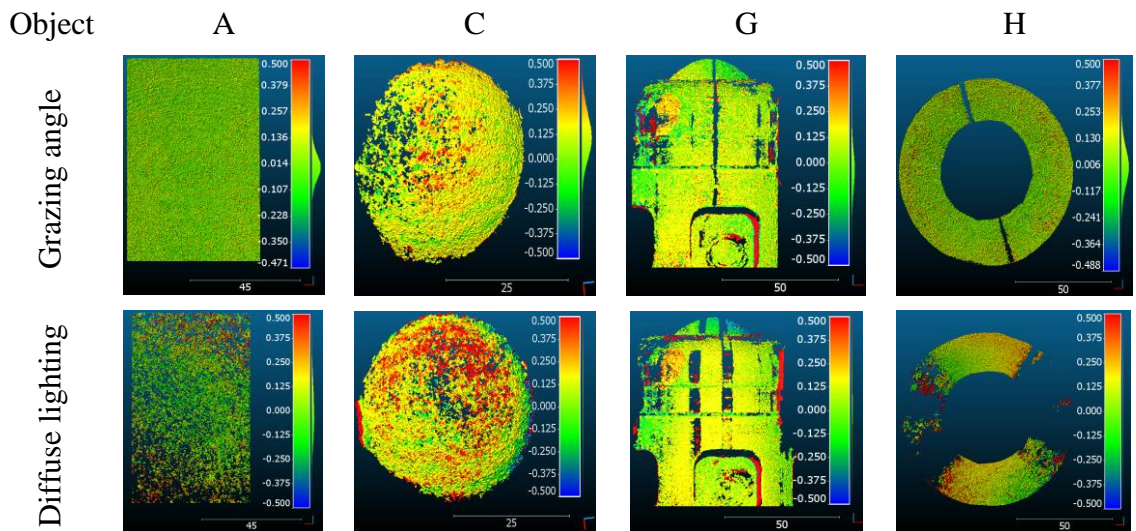

Figure 5-17. Cloud-to-cloud comparison for two textureless and two shiny objects.

Table 5-8. The results of point-to-point comparison.

Object	Diffuse lighting	Light 09	Light 11	Light 14	Light 16
A	0.178	0.09	0.1	0.072	0.059

C	0.25	0.15	0.13	0.14	0.12
D	0.12	0.097	0.106	0.094	0.079
E	0.165	0.13	0.125	0.13	0.126
F	0.24	0.08	0.094	0.097	0.097
G	0.185	0.14	0.14	0.14	0.145
H	0.22	0.11	0.11	0.12	0.12

The estimated RMSE (mm) of the point-to-point comparison for four grazing angles against diffuse images. The best and worst results are represented by green and brown colors, respectively.

5.6.2 3D reconstruction using image fusion-based methods

In this section, we aim to present the results of image-based fusion methods (average, median, albedo, GLCM, and deep learning) for improving 3D reconstruction using our proposed image acquisition system. As mentioned before, the idea of fusing multiple images, that are taken under different illuminations, is to improve the signal-to-noise ratio, obtain a fused image free from shadows and highlights, and improve the image quality. Each method was tested on two different objects (one pure textureless and one reflective object). We provided a cloud-to-cloud comparison for each method using reference data. Figure 5-18 shows the results, with a negative (blue) or positive (red) value indicating whether the generated point is below or above the closest point on the reference mesh surface, respectively. Max error is set based on the three-sigma limits. Table 5-9

reports the estimated RMSE, MAE, and STD of the point-to-point comparison between the generated point clouds and the reference mesh for each method on both shiny and textureless objects. From the estimated results, it can be seen that the method based on GLCM outperforms the other method regardless of the object's property. While the performance of other methods directly depends on the object's type therefore their performance can be different from one object to the other.
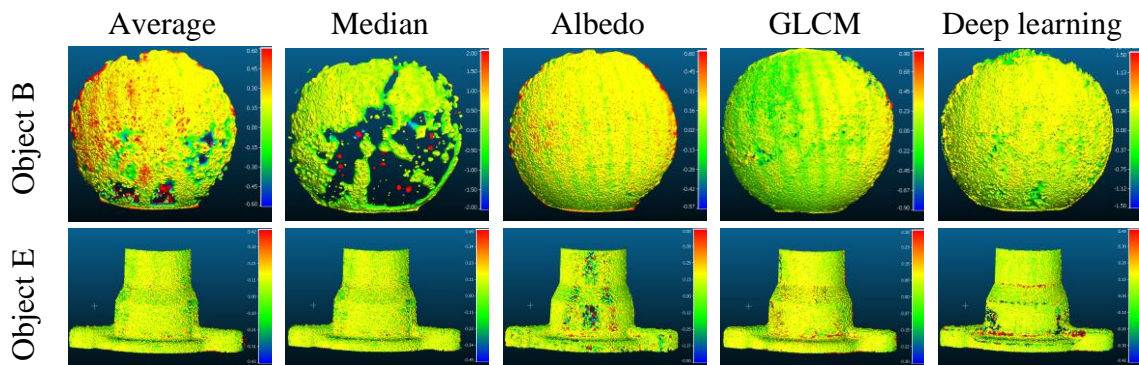
Figure 5-18. Comparison results of image fusion for objects E and B.

The visual representation of the point-to-point comparison for five image-based fusion approaches (Average Median, Albedo, GLCM, and Deep learning) on two non-collaborative objects (object B: texture-less and object E: shiny).

Table 5-9. Comparison results of image fusion for objects B and E.

	metric	Average	Median	Albedo	GLCM	Deep learning
Obj	RMSE	0.16	0.47	0.13	0.13	0.23

	MAE	0.23	0.34	0.16	0.13	0.26
	STD	0.22	0.58	0.11	0.16	0.3
	ME	0.16	-0.07	0.16	0.15	0.13
Object E	RMSE	0.07	0.05	0.1	0.05	0.08
	MAE	0.07	0.06	0.1	0.05	0.07
	STD	0.09	0.08	0.15	0.06	0.1
	ME	0	0	0	0	0

The estimated results (in millimeters) of the point-to-point comparison on two shiny (object E) and textureless (object B) objects for five different approaches (Average Median, Albedo, GLCM, and Deep learning). The best and worst results are represented by green and brown colours, respectively.

For instance, for object B (pure textureless), the GLCM and Albedo provided the best results in comparison with other methods with an estimated RMSE of 0.13mm and the worst results belonged to Median with RMSE of 0.47 mm. Whereas, for object E (metallic and shiny), the best results were achieved for the Median and GLCM methods with an estimated RMSE of less than 0.05 mm while the poorest results were obtained for Albedo with a measured RMSE of more than 0.1 mm. The results also show that GLCM is more robust and produces more reliable results than the other techniques. Furthermore, when the object has no texture (object B), the GLCM results are somewhat higher but still better than the others. This is because the surface of object E has higher roughness and 3D microstructures which can be better emphasized under grazing angles.

5.6.3 3D reconstruction using object-based fusion

In this method, instead of fusing images in the image space (Section 4.4.4), n different point clouds, generated under the best-selected grazing angles (Section 4.4.3), were combined in object space to form the final model (see details in Section 4.4.5). To evaluate this method, four different objects, two texture-less and two metallics, were used/ compared. For each dataset, at least three individual 3D point clouds were selected and merged in object space to shape the final model.

Object A   Object B

Texture-less objects
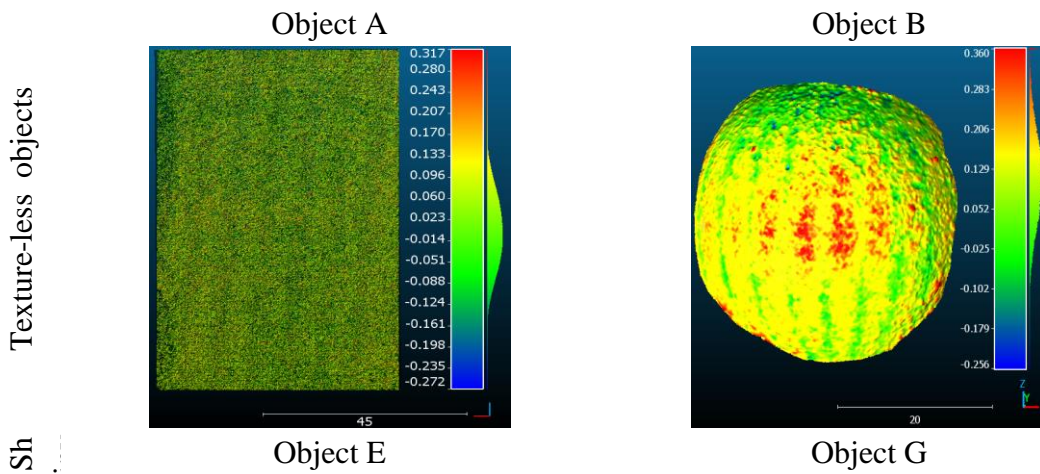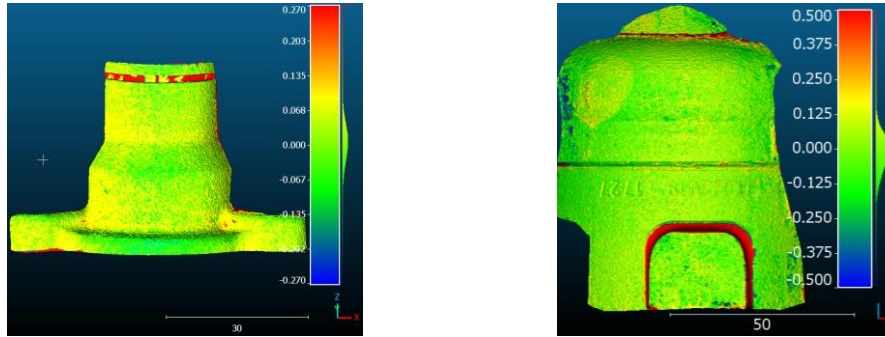
Object E   Object G

Figure 5-19. Comparison results of object-based fusion.

The visual representation of the point-to-point comparison of object-based fusion on four different non-collaborative objects (two texture-less and two shiny objects).

Table 5-10. Comparison results of object-based fusion on four objects, Unit: mm.

Object type	Texture-less		Shiny	
Object	Object A	Object B	Object E	Object G
RMSE	0.03	0.08	0.05	0.10
MAE	0.05	0.11	0.05	0.11
STD	0.06	0.09	0.06	0.14
ME	0	0.11	0.0	0.01
Max error	0.32	0.35	0.3	0.5

The results of a point-to-point comparison between the object-based fusion approach and reference data are represented in Figure 5-19. The estimated RMSE, MAE, and STD for each object are reported in Table 5-10.

This comparison aimed to understand how much the idea of merging different 3D reconstructions generated under grazing angles can improve the final 3D model instead of increasing the number of images. As shown in Table 5-10, the estimated error metrics (RMSE, MAE, and STD) for 3D reconstruction observed after merging was changeable between 0.03 mm to 0.14 mm depending on the surfaces' roughness. For example, the best performance for the method based on the merging 3D reconstruction in object space was obtained for Objects A and E with the RMSE, MAE, and STD less than 0.06mm, which has a relatively rougher surface than other objects (C, G).

Whereas the worst-performing results were achieved on Object G with an estimated error two times larger than objects A and E. This is because the surface of this object is high-reflective, textureless, and smoothed making it difficult to generate noiseless 3D reconstruction. This test can indicate that the rougher a surface is, the better 3D reconstruction is observed in general.

In another effort, this evaluation was extended to compare the proposed methods based on image fusion (Section 4.4.4) against the object-based fusion method (Section 4.4.5). The results of this comparison for object E are presented in Figure 5-20. From the estimated error metrics of the point-to-point comparison, it can be seen that GLCM (one

of the image-based fusion methods) and the method based on the integration of individual

3D reconstruction under the best grazing angles (object-based fusion) produce a better 3D

results compared to other methods (Average, Median, Albedo, and Deap learning). For

instance, the estimated STD for GLCM and object-based fusion methods was less than 0.06

mm while this value for other approaches were larger than 0.08 mm.
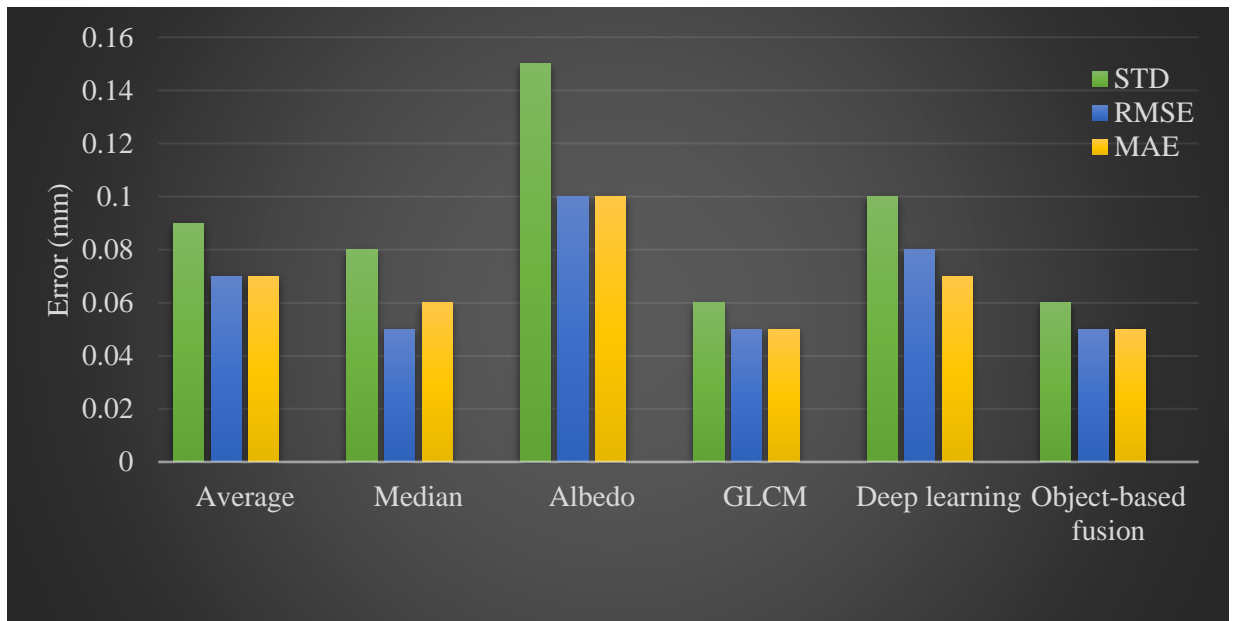
Figure 5-20. Comparison of image-based fusion against object-based fusion.

The comparison results obtained by five image-based methods (Average, Median, Albedo, GLCM, and Deep Learning) in comparison with the method based on the merging 3D reconstruction (object-based fusion).

5.7　Image orientation and 3D reconstruction of transparent surfaces

In this section, we first evaluate the performance of two suggested approaches for determining the orientation of transparent objects (cross-correlation and SIFT-based pipelines). Using three different pipelines (Shape from silhouette, OpenVMS, and COLMAP), a 3D dense cloud of the objects were then generated and evaluated given the image orientation.

5.7.1 Image orientation evaluation

To evaluate the accuracy of the two proposed pipelines, we tested the plastic bottle (object M), the glass bottle (object N), and the teacup (object O) shown in Figure 5-21. The visual representation of the generated sparse 3D reconstruction and an example image for each of the three transparent objects are presented in Figure 5-21. For each object, a set of images was captured from 36 different stations under constant lighting, directed to enhance the appearance of the low-contrast texture. A set of photogrammetric coded targets were printed and mounted on a rotating table with known relative distances to provide a metric assessment using different criteria such as mean reprojection error (MRE), and the mean error and standard deviation. The results are reported in Table 5-11
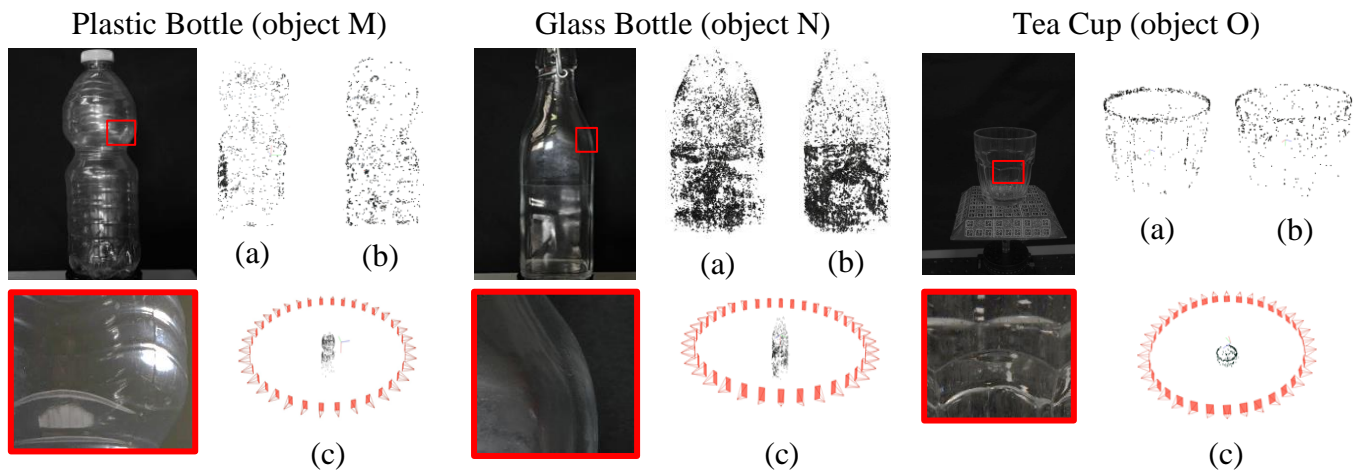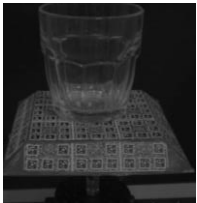
Figure 5-21. Visual representation of image orientation for three transparent objects

Sparse reconstruction (a and b) and image detail (red boxes) for the three transparent objects used in the accuracy evaluation. (a) Image matching using adapted SIFT descriptor, (b) image matching using Cross Correlation descriptor, and (c) recovered camera network.

For object O, both pipelines achieve an MRE of about 0.4 pixels. SIFT obtains an Std of 0.052 mm with Mean Absolute Error (MAE) of 0.048 mm, while the cross-correlation pipeline obtains an Std of 0.079 mm with MAE of 0.083 mm, twice larger than SIFT. For object N with GSD of 0.05 mm/px, the SIFT-based pipeline has an Std of 0.012

160

mm, MAE of 0.011 mm, and MRE of 0.581 px. The cross-correlation pipeline achieves a

similar MRE of 0.532 px, while a significantly worse Std and MAE of 0.068 mm and 0.106,

respectively.

Table 5-11. Quantitative image orientation evaluation.

DATASET	Camera Focal length Pixel size Distance	GSD [mm/px]	Adapted SIFT-BASED APPROACH			CROSS-CORRELATION		
			Mean Absolute Error [mm]	Std [mm]	Mean Reprojection Error [px]	Mean Absolute Error [mm]	Std [mm]	Mean Reprojection Error [px]
	NIKON D750 28 mm 5.98 μm 240 mm	0.05	0.011	0.012	0.581	0.106	0.068	0.532
	NIKON D750 28 mm 5.98 μm 240 mm	0.05	0.054	0.044	0.648	0.016	0.011	0.291
	NIKON D3X 60 mm 5.98 μm 300 mm	0.03	0.048	0.052	0.436	0.083	0.079	0.423

| Average | | 0.043 | 0.037 | 0.036 | 0.555 | 0.068 | 0.052 | 0.41 |

Accuracy evaluation for two proposed pipelines (adapted sift-based approach and cross-correlation) using different criteria of Mean Reprojection Error (MRE) in pixel, the residuals in mm (Mean Absolute Error (MAE) and standard deviation(std)). The green color represents the best results and the red color represents the worse results.

This is because it did not orient four images so in this case, we do not have a loop-closure with a decrease in accuracy. Failure to close the network may also be due to the lower number of local features used in the cross-correlation method compared to SIFT, which has been kept small for computational reasons. For object M the cross-correlation approach reaches a better result for all criteria compared to SIFT approach.

5.7.2 Dense cloud evaluation

After orienting the images, the dense cloud was generated using different pipelines, including OpenMVS (Moulon et al., 2013), COLMAP (Schonberger and Frahm, 2016), and Shape from Silhouette (SFS) to propose different possible solutions for 3D dense reconstruction. To generate reference data for each object, their surface was covered with a thin layer of random colored powder to i) make it diffusely reflecting and remove refraction and ii) provide texture on the surface. After surface treatment, an additional

photogrammetric 3D reconstruction was employed to generate a dense 3D reconstruction (considered as reference data).

To evaluate the accuracy potential of the suggested method in low-frequency domain, the 3D results achieved with each pipeline were geometrically compared against reference data (photogrammetry). To this end, the obtained 3D point clouds were registered to the reference data using an Iterative Closest Point (ICP) algorithm. The RMS of the shortest distance between the homologous points on the reconstructed and reference models was then calculated and compared. The results of the point-to-point comparison for both objects (M and N) are presented in Figure 5-22.

The quantitative analysis shown in Figure 5-22 demonstrates that the SFS can recover the 3D shape of an object regardless of whether it is textureless or transparent, as long as the region of the object in each image is distinguishable from the background. However, SFS failed to reconstruct concavities or holes on an object's surface making it difficult for geometrically complex objects. It is quite possible that the silhouette of an object would be trimmed or expanded, resulting in a 3D model that is smaller or bigger than the actual size of the object.

On the other hand, the 3D results using OpenMVS and COLMAP (Figure 5-22) directly depend on the high-resolution texture of the object surface, hence it completely

failed or generate noisy point clouds in the area where microstructure and roughness are not highlighted well (see red boxes).
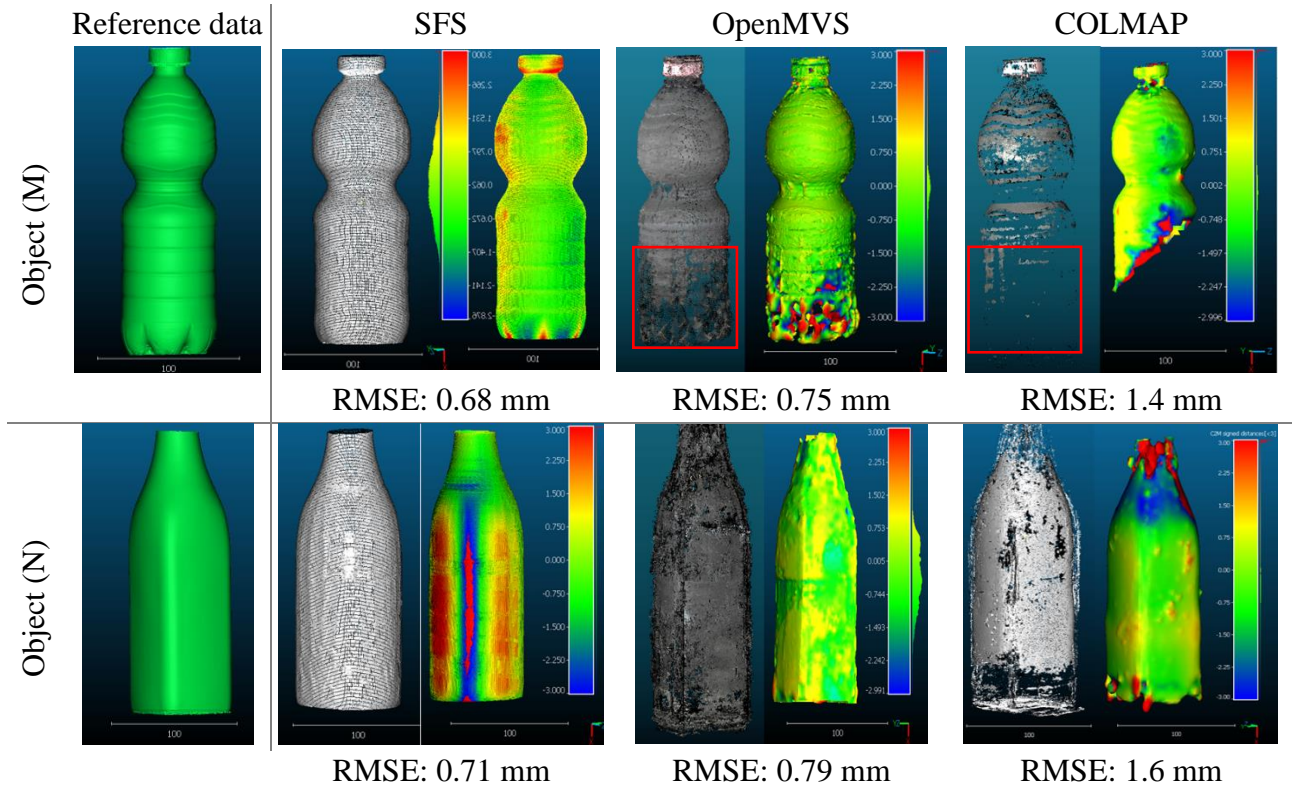
Figure 5-22. Results of the point-to-point comparison for transparent objects.

The comparison results for objects M and N using three different dense reconstruction pipelines.

5.8 Summary

This chapter demonstrates in depth the advantages of the proposed methods. First, we presented a variety of objects with distinct surface characteristics, such as textureless, reflective, and transparent, that were utilized to evaluate various proposed methods. Then, the procedure of gathering ground truth to quantitatively assess our proposed methodologies in the low and frequency domains. The quantitative results of the four strategies provided in Chapter 4 were then presented.

CHAPTER VI

Conclusions and future works

6.1 Conclusions

This thesis proposed various pipelines for 3D shape reconstruction of non-collaborative surfaces including textureless, shiny, and transparent, using our proposed PS data acquisition system. Unlike the most of existing approaches, the proposed solutions can generate an accurate and high-resolution topography of non-collaborative surfaces. In chapter 5 we demonstrated extensively the advantages of the proposed methods, which can be briefly summarized as follows:

- As a first step, we presented two automatic and semi-automatic image acquisition systems based on the near-field PS lighting system, as well as a straightforward but effective calibration procedure for measuring the lighting system's system geometry and the camera's internal and exterior orientation parameters. The designed system is suited for collecting data for photogrammetry and photometric stereo techniques, as well as merging photogrammetry and photometric stereo measurements (Chapter 3).

- We proposed an integrated method for correcting the global shape deviation of the PS depth map by utilizing the advantage of reliable photogrammetric 3D measurements. We utilized the measured camera geometry, lighting system, and approximate 3D geometry of the object to correct the light directions, compute intensity attenuations, and automatically detect and remove shadow and specular reflection (Section 4.25.4).

- Then, the integration strategy was extended through a combination of photogrammetry and PS depth maps in the frequency domain. We proposed an FFT-based approach for fusing the high spatial frequencies of photometric stereo with the low spatial frequencies of photogrammetry in order to obtain accurate low spatial frequencies while maintaining high spatial frequencies. Regarding this, we first built a weighting plane in the frequency domain to assign a value to each pixel, followed by a non-linear interpolation to eliminate wrong frequencies while fusing accurate low and high frequencies. The achieved 3D results indicated that the proposed FFT-based fusion recovered high-resolution details with an estimated RMSE below 20 µm quite close to photometric stereo results while inheriting the geometric information (low-frequency) from photogrammetry with the RMSE of less than 100 µm (Section 5.5).

- Moreover, we presented a pipeline to improve the quality and reliability of dense reconstruction of non-collaborative objects by employing the directional lighting obtained from our PS lighting system. Due to the interaction of grazing rays with microstructures, shadows and shading could produce spatially varying chiaroscuro patterns on the surface, which improve the local contrast of the image, hence improving image matching in orientation and multi-view stereo-dense reconstruction. We analyzed the effect of light directionality on the generated point clouds through

168

different criteria including the density of extracted tie points, photometric consistency score, and geometric cloud-to-cloud evaluation (Section 5.6.1).

- We proposed five different image fusion-based methods including average, median, albedo, GLCM, and learning-based to fuse grazing images into a unique image where the signal-to-noise ratio is improved, and the output image is free from shadows and highlights. The achieved results indicated that the method based on GLCM outperforms the other method regardless of the object's property. While the performance of other methods directly depends on the object's type therefore their performance can be different from one object to the other (Section 5.6.2).

- After that, instead of fusing images in image space, n different point clouds generated utilizing the best grazing angles, were merged in object space (3D space) to generate the final model. Since the selected point clouds were all orientated and registered within the same reference coordinate system, combining them to generate the final point cloud was a straightforward process. From the obtained results, it is clear that object-based fusion is more robust to objects with diverse surface properties and shape geometries than image-based fusion techniques (Section 5.6.3).

- Lastly, we proposed two principles that successfully orient images of transparent objects by exploiting the low-contrast textures emphasized on object surfaces

(roughness and 3D microstructures) with the aid of an appropriate lighting system (Section 5.7.1). Both methods prioritize tie point detection on low-contrast textures over high-contrast textures, hence rejecting specular reflections and static tie points. Unlike conventional SfM-based methods that prioritize high-contrast textures, both approaches privilege tie point detection on low-contrast textures instead of high-contrast textures, discarding specular reflections and static tie points. For the first approach, local descriptors are extracted in those regions where roughness and micro-structures are better highlighted. Also, the normalized cross-correlation (NCC) was applied on the gradient map of the images to fully exploit the geometrical content of the patches. The second approach builds on the first method by adapting the classic SIFT pipeline. Although both approaches may effectively orient images, the SIFT-based method is slightly quicker and more accurate.

6.2   Recommendations for future works

The data acquisition and techniques proposed in this thesis have several limits and drawbacks, which have led to the identification of the following areas as suggestions for further research.

- As possible future work, the LED number in the system could be increased in order to boost its flexibility and take advantage of light directionality for better surface inspection, allowing more images to be captured under a wider range of grazing angles, and making the system more flexible for object inspection of varying complexity.

- For some particular objects, like a 3D inspection of plants, using LEDs with infrared light can also be useful.

- Spatially varying BRDF will be also investigated for better surface rendering.

- RTI (Reflectance Transformation Imaging) is a technique used to capture and display the surface texture of an object. It involves taking multiple photographs of the object under different lighting conditions, and then combine them into a single image that can be interactively lit from different angles. RTI can reveal fine details in the surface texture of an object that might be difficult to see in a single photograph.

- The size and weight of the system restrict it to the laboratory and cannot be taken to the site where objects are located, thus, working on its mobility might be an alternative.

- Another alternative would be to install the cameras and light sources on adjustable arms in order to handle objects of various sizes.

- It would be possible to completely automate the system for mass 3D digitization in industrial sectors.

171

- Regarding the 3D reconstruction of transparent objects, deep learning descriptors could be investigated as an alternative to cross-correlation and SIFT method, considering more complex datasets featuring more than one strip or multi-camera acquisitions.

- Starting from the estimated image orientation, it would be important to evaluate the potential of Neural Radiance Fields (NeRF) for 3D reconstruction of non-collaborative objects (Morelli et al., 2022).

- Combining photometric stereo with nerf can be also interesting to generate a precise and high-resolution 3D shape for a transparent object.

- Regarding the radiometric calibration, the radiometric response of the imaging system can be calculated using for example a macbeth chart. This involves measuring the pixel values of the images and comparing them to the known reflectance or emissivity values of the calibration targets. Then, using the radiometric response values, generate a radiometric correction map that can be used to correct the pixel values of the actual images captured by the imaging system.

References

Ahmadabadian, A.H., Karami, A. and Yazdan, R., 2019. An automatic 3D reconstruction system for texture-less objects. *Robotics and Autonomous Systems*, *117*, pp.29-39.

Antensteiner, D., Štolc, S. and Pock, T., 2018. A review of depth and normal fusion algorithms. *Sensors*, *18*(2), p.431.

Aubreton, O., Bajard, A., Verney, B. and Truchetet, F., 2013. Infrared system for 3D scanning of metallic surfaces. *Machine Vision and Applications, 24(7),* pp.1513-1524.

Barazzetti, L., Gianinetto, M. and Scaioni, M., 2012. Automatic image-based 3D modeling for medical applications. In *2012 5th International Conference on BioMedical Engineering and Informatics* (pp. 228-232). ieee.

Barsky, S. and Petrou, M., 2003. 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), pp.1239-1252.

Baumga`rt, B.G.*, 1974.* Geometric Modeling for Computer Vision. *PhD Thesis, Stanford University.*

Besl, P.J. and McKay, N.D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2).

Boss, M., Jampani, V., Kim, K., Lensch, H. and Kautz, J., 2020. Two-shot spatially-varying brdf and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3982-3991).

Brahm, A., Rößler, C., Dietrich, P., Heist, S., Kühmstedt, P. and Notni, G., 2016. Non-destructive 3D shape measurement of transparent and black objects with thermal fringes. *In Dimensional Optical Metrology and Inspection for Practical Applications V (Vol. 9868, p. 98680C). International Society for Optics and Photonics.*

Brown, DC., 1971. Close-Range Camera Calibration. *Photogrammetric Engineering.* 37(8), 855–866.

Bylow, E., Maier, R., Kahl, F. and Olsson, C., 2019. Combining depth fusion and photometric stereo for fine-detailed 3d models. In *Scandinavian Conference on Image Analysis* (pp. 261-274). Springer, Cham.

Calantropio, A., Chiabrando, F., Seymour, B., Kovacs, E., Lo, E. and Rissolo, D., 2020. Image pre-processing strategies for enhancing photogrammetric 3d reconstruction of underwater shipwreck datasets. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *43*.

Carbone, V., Carocci, M., Savio, E., Sansoni, G. and De Chiffre, L., 2001. Combination of a vision system and a coordinate measuring machine for the reverse engineering of freeform surfaces. *The International Journal of Advanced Manufacturing Technology*, *17*(4), pp.263-271.

Chandraker, M., Agarwal, S. and Kriegman, D., 2007, June. Shadowcuts: Photometric stereo with shadows. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.

Chari, V. and Sturm, P., 2013. A theory of refractive photo-light-path triangulation. Proc. *CVPR, pp. 1438-1445.*

Chen, G., Han, K. and Wong, K.Y.K., 2018. PS-FCN: A flexible learning framework for photometric stereo. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3-18).

Chen, A., Xu, Z., Geiger, A., Yu, J. and Su, H., 2022. Tensorf: Tensorial radiance fields. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel,

October 23–27, 2022, Proceedings, Part XXXII (pp. 333-350). Cham: Springer Nature Switzerland.

Cho, D., Matsushita, Y., Tai, Y.W. and Kweon, I.S., 2018. Semi-calibrated photometric stereo. *IEEE transactions on pattern analysis and machine intelligence*, *42*(1), pp.232-245.

Chung, H.S. and Jia, J., 2008. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.

Durou, J.D., Falcone, M., Quéau, Y. and Tozza, S. eds., 2020. Advances in Photometric 3D-Reconstruction. *Springer International Publishing.*

Eppel, S., Xu, H., Wang, Y.R. and Aspuru-Guzik, A., 2021. Predicting 3D shapes, masks, and properties of materials, liquids, and objects inside transparent containers, using the TransProteus CGI dataset. *arXiv preprint arXiv:2109.07577.*

Eren, G., Aubreton, O., Meriaudeau, F., Secades, L.S., Fofi, D., Naskali, A.T., Truchetet, F. and Ercil, A., 2009. Scanning from heating: 3D shape estimation of transparent objects from local surface heating. *Optics Express, 17(14),* pp.11457-11468.

Evixscan. Evixscan 3D Fine Precision [Online]. Available: *https://evixscan3d.com* [last access: November 03, 2022].

Fan, H., Qi, L., Wang, N., Dong, J., Chen, Y. and Yu, H., 2017. Deviation correction method for close-range photometric stereo with nonuniform illumination. *Optical Engineering*, *56*(10), p.103102.

Furukawa, Y. and Ponce, J., 2009. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, *32*(8), pp.1362-1376.

Gaiani, M., Remondino, F., Apollonio, F.I. and Ballabeni, A., 2016. An advanced pre-processing pipeline to improve automated photogrammetric reconstructions of architectural scenes. *Remote sensing*, *8*(3), p.178.

175

Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D., Litany, O., Gojcic, Z. and Fidler, S., 2022. GET3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images. *arXiv preprint arXiv:2209.11163*.

Georghiades, A.S., 2003. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *Computer Vision, IEEE International Conference on* (Vol. 3, pp. 816-816). IEEE Computer Society.

Gruen, A. and Beyer, H.A., 2001. System calibration through self-calibration. *In Calibration and orientation of cameras in computer vision. Springer, Berlin, Heidelberg* (pp. 163-193).

Guo, H., Peng, S., Lin, H., Wang, Q., Zhang, G., Bao, H. and Zhou, X., 2022. Neural 3d scene reconstruction with the manhattan-world assumption. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5511-5520).

Hafeez, J., Lee, J., Kwon, S., Ha, S., Hur, G. and Lee, S., 2020. Evaluating feature extraction methods with synthetic noise patterns for image-based modelling of texture-less objects. *Remote Sensing*, *12*(23), p.3886.

Han, K., Wong, K.Y.K. and Liu, M., 2021. Dense Reconstruction of Transparent Objects by Altering Incident Light Paths Through Refraction. *arXiv e-prints, pp.arXiv-2105.*

He, K., Sui, C., Huang, T., Dai, R., Lyu, C. and Liu, Y.H., 2022. 3D Surface reconstruction of transparent objects using laser scanning with LTFtF method. *Optics and Lasers in Engineering, 148,* p.106774.

Hernandez, C., Vogiatzis, G. and Cipolla, R., 2008. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(3), pp.548-554.

Hexagon. AICON PrimeScan scanner [Online]. *Available: https://www.hexagonmi.com* [last access: November 03, 2022].

Hosseininaveh, A., Yazdan, R., Karami, A., Moradi, M. and Ghorbani, F., 2015. A low-cost and portable system for 3D reconstruction of texture-less objects. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(1), p.327.

Huynh, C.P., Robles-Kelly, A. and Hancock, E., 2010. Shape and refractive index recovery from single-view polarisation images. Proc. *CVPR,* pp. 1229-1236.

Iwabuchi, Y., Li, L., Baba, M. and Ohtani, K., 2011. 3D shape measurement of a transparent object with unknown refractive index by inverse ray tracing method. *In SICE Annual Conference 2011*, pp. 16-21.

Jain, A., Tancik, M. and Abbeel, P., 2021. Putting NeRF on a diet: Semantically consistent few-shot view synthesis. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 5885-5894).

Ji, Y., Xia, Q. and Zhang, Z., 2017. Fusing depth and silhouette for scanning transparent object with RGB-D sensor. *International Journal of Optics.*

Jiddi, S., Robert, P. and Marchand, E., 2020. Detecting specular reflections and cast shadows to estimate reflectance and illumination of dynamic indoor scenes. *IEEE transactions on visualization and computer graphics*.

Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K.M. and Trulls, E., 2021. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, *129*(2), pp.517-547.

Kampel, M., Tosovic, S. and Sablatnig, R., 2002. Octree-based fusion of shape from silhouette and shape from structured light. *In Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission,* pp. 754-757.

Karami, A., Menna, F. and Remondino, F., 2021. Investigating 3d Reconstruction Of Non-Collaborative Surfaces Through Photogrammetry And Photometric Stereo. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, pp.519-526.

Karami, A., Battisti, R., Menna, F. and Remondino, F., 2022a. 3d digitization of transparent and glass surfaces: state of the art and analysis of some methods. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, pp.695-702.

Karami, A., Menna, F., Remondino, F. and Varshosaz, M., 2022b. Exploiting Light Directionality for Image-Based 3D Reconstruction of Non-Collaborative Surfaces. *The Photogrammetric Record*, *37*(177), pp.111-138.

Karami, A., Menna, F. and Remondino, F., 2022c. Combining Photogrammetry and Photometric Stereo to Achieve Precise and Complete 3D Reconstruction. *Sensors*, *22*(21), p.8172.

Kaya, B., Kumar, S., Oliveira, C., Ferrari, V. and Van Gool, L., 2022a. Uncertainty-Aware Deep Multi-View Photometric Stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12601-12611).

Kaya, B., Kumar, S., Sarno, F., Ferrari, V. and Van Gool, L., 2022b. Neural radiance fields approach to deep multi-view photometric stereo. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 1965-1977).

Kim, J., Reshetouski, I. and Ghosh, A., 2017. Acquiring axially-symmetric transparent objects using single-view transmission imaging. Proc. *CVPR,* pp. 3559-3567.

Kutulakos, K.N. and Steger, E., 2008. A theory of refractive and specular 3D shape by light-path triangulation. *International Journal of Computer Vision,* 76(1), pp.13-29.

Landmann, M., Heist, S., Kühmstedt, P. and Notni, G., 2019. 3D shape from thermal patterns: investigation of projection parameters in simulation and experiment. *In Optical Measurement Systems for Industrial Inspection XI (Vol. 11056, p. 1105615). International Society for Optics and Photonics.*

Landmann, M., Speck, H., Schmieder, J.T., Heist, S. and Notni, G., 2021. Improvement of Thermal Fringe Projection for Fast and Accurate 3D Shape Measurement of Transparent Objects. *of Materials,* p.99.

Li, Z., Yeh, Y.Y. and Chandraker, M., 2020. Through the looking glass: neural 3D reconstruction of transparent shapes. Proc. *CVPR,* pp. 1262-1271.

Li, M., Zhou, Z., Wu, Z., Shi, B., Diao, C. and Tan, P., 2020. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *IEEE Transactions on Image Processing*, *29*, pp.4159-4173.

Lin, H., Gao, J., Zhang, G., Chen, X., He, Y. and Liu, Y., 2017. Review and comparison of high-dynamic range three-dimensional shape measurement techniques. *Journal of Sensors*, *2017*.

Logothetis, F., Mecca, R. and Cipolla, R., 2019. A differential volumetric approach to multi-view photometric stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1052-1061).

Lu, F., Chen, X., Sato, I. and Sato, Y., 2017. Symps: Brdf symmetry guided photometric stereo for shape and light source estimation. *IEEE transactions on pattern analysis and machine intelligence*, *40*(1), pp.221-234.

Luhmann, T., 2010. Close range photogrammetry for industrial applications. *ISPRS journal of photogrammetry and remote sensing*, *65*(6), pp.558-569.

Luhmann, T., Robson, S., Kyle, S. and Boehm, J., 2019. Close-range photogrammetry and 3D imaging. In *Close-Range Photogrammetry and 3D Imaging*. de Gruyter.

Lyu, J., Wu, B., Lischinski, D., Cohen-Or, D. and Huang, H., 2020. Differentiable refraction-tracing for mesh reconstruction of transparent objects. *ACM Transactions on Graphics, 39*(6), pp.1-13.

MacDonald, L.W., Ahmadabadian, A.H. and Robson, S., 2015. Determining the coordinates of lamps in an illumination dome. In *Videometrics, Range Imaging, and Applications XIII* (Vol. 9528, pp. 156-167). SPIE.

Martin, W.N. and Aggarwal, J.K., 1983. Volumetric descriptions of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. (2), pp.150-158.

Mecca, R., Wetzler, A., Bruckstein, A.M. and Kimmel, R., 2014. Near field photometric stereo with point light sources. *SIAM Journal on Imaging Sciences*, *7*(4), pp.2732-2770.

Menna, F., Nocerino, E., Remondino, F., Dellepiane, M., Callieri, M. and Scopigno, R., 2016. 3d digitization of an heritage masterpiece-a critical analysis on quality assessment. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *41*.

Menna, F., Nocerino, E., Morabito, D., Farella, E.M., Perini, M. and Remondino, F., 2017. An open source low-cost automatic system for image-based 3D digitization. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, p.155.

Mikhail, E.M., Bethel, J.S. and McGlone, J.C., 2001. Introduction to modern photogrammetry. *John Wiley & Sons*.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. Proc. ECCV.

Miyazaki, D., Saito, M., Sato, Y. and Ikeuchi, K., 2002. Determining surface orientations of transparent objects based on polarization degrees in visible and infrared wavelengths. *JOSA A,* 19(4), pp.687-694.

Miyazaki, D., Kagesawa, M. and Ikeuchi, K., 2004. Transparent surface modeling from a pair of polarization images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 26(1), pp.73-82.

Miyazaki, D. and Ikeuchi, K., 2007. Shape estimation of transparent objects by using inverse polarization ray tracing. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 29(11), pp. 2018-2030.

Miyazaki, D., Shigetomi, T., Baba, M., Furukawa, R., Hiura, S. and Asada, N., 2016. Surface normal estimation of black specular objects from multiview polarization images. *Optical Engineering,* 56(4), p.041303.
180

Mohammadi, M., Rashidi, M., Mousavi, V., Karami, A., Yu, Y. and Samali, B., 2021. Quality evaluation of digital twins generated based on uav photogrammetry and tls: bridge case study. *Remote Sensing*, *13*(17), p.3499.

Morelli, L., Karami, A., Menna, F. and Remondino, F., 2022. Orientation of Images with Low Contrast Textures and Transparent Objects. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*, pp.77-84.

Morris, N.J. and Kutulakos, K.N., 2007. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. Proc. *ICCV*, pp. 1-8.

Morris, N.J. and Kutulakos, K.N., 2011. Dynamic refraction stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8),* pp.1518-1531.

Moulon, P., Monasse, P. and Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. Proc. *ICCV*. pp. 3248-3255.

Mousavi, V., Khosravi, M., Ahmadi, M., Noori, N., Haghshenas, S., Hosseininaveh, A. and Varshosaz, M., 2018. The performance evaluation of multi-image 3D reconstruction software with different sensors. *Measurement*, *120*, pp.1-10.

Müller, T., Evans, A., Schied, C. and Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (ToG), 41(4), pp.1-15.

Murase, H., 1992. Surface shape reconstruction of an undulating transparent object. Proc. *ICCV*.

Murtiyoso, A. and Grussenmeyer, P., 2020. Virtual disassembling of historical edifices: Experiments and assessments of an automatic approach for classifying multi-scalar point clouds into architectural elements. *Sensors*, *20*(8), p.2161.

Narayan, K.S., Sha, J., Singh, A. and Abbeel, P., 2015. Range sensor and silhouette fusion for high-quality 3D Scanning. *Proc. IEEE ICRA*, pp. 3617-3624.

Nehab, D., Rusinkiewicz, S., Davis, J. and Ramamoorthi, R., 2005. Efficiently combining positions and normals for precise 3D geometry. *ACM transactions on graphics (TOG)*, *24*(3), pp.536-543.

Nicolae, C., Nocerino, E., Menna, F. and Remondino, F., 2014. Photogrammetry applied to problematic artefacts. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(5), p.451.

Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S., Geiger, A. and Radwan, N., 2022. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5480-5490).

Otani, H., Komuro, T., Yamamoto, S. and Tsumura, N., 2019. Bivariate BRDF Estimation Based on Compressed Sensing. In *Computer Graphics International Conference* (pp. 483-489). Springer, Cham.

Pan, H., Guan, T., Luo, Y., Duan, L., Tian, Y., Yi, L., Zhao, Y. and Yu, J., 2016. Dense 3D reconstruction combining depth and RGB information. *Neurocomputing*, 175, pp.644-651.

Palousek, D., Omasta, M., Koutny, D., Bednar, J., Koutecky, T. and Dokoupil, F., 2015. Effect of matte coating on 3D optical measurement accuracy. *Optical Materials*, *40*, pp.1-9.

Park, J., Sinha, S.N., Matsushita, Y., Tai, Y.W. and Kweon, I.S., 2013. Multiview photometric stereo using planar mesh parameterization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1161-1168).

Park, J., Sinha, S.N., Matsushita, Y., Tai, Y.W. and Kweon, I.S., 2016. Robust multiview photometric stereo using planar mesh parameterization. *IEEE transactions on pattern analysis and machine intelligence*, *39*(8), pp.1591-1604.

Peng, S., Haefner, B., Quéau, Y. and Cremers, D., 2017. Depth super-resolution meets uncalibrated photometric stereo. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 2961-2968).

Pereira, J.R.M., de Lima e Silva Penz, I. and da Silva, F.P., 2019. Effects of different coating materials on three-dimensional optical scanning accuracy. *Advances in Mechanical Engineering*, *11*(4), p.1687814019842416.

Potmesil, M., 1987. Generating octree models of 3D objects from their silhouettes in a sequence of images. *Computer Vision, Graphics, and Image Processing*, 40(1), pp.1-29.

Qian, Y., Gong, M. and Yang, Y.H., 2017. Stereo-based 3D reconstruction of dynamic fluid surfaces by global optimization. Proc. *CVPR,* pp. 1269-1278.

Quéau, Y., Wu, T., Lauze, F., Durou, J.D. and Cremers, D., 2017. A non-convex variational approach to photometric stereo under inaccurate lighting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 99-108).

Remondino, F. and El-Hakim, S., 2006. Image-based 3D modelling: a review. *The photogrammetric record*, *21*(115), pp.269-291.

Ren, M., Ren, J., Wang, X., Gao, F., Zhu, L. and Yao, Z., 2020. Multi-scale measurement of high-reflective surfaces by integrating near-field photometric stereo with touch trigger probe. *CIRP Annals*, *69*(1), pp.489-492.

Ren, J., Jian, Z., Wang, X., Mingjun, R., Zhu, L. and Jiang, X., 2021. Complex surface reconstruction based on fusion of surface normals and sparse depth measurement. *IEEE Transactions on Instrumentation and Measurement*, *70*, pp.1-13.

Rodríguez-Martín, M., Lagüela, S., González-Aguilera, D. and Rodríguez-Gonzálvez, P., 2015. Procedure for quality inspection of welds based on macro-photogrammetric three-dimensional reconstruction. *Optics & Laser Technology*, *73*, pp.54-62.

Rostami, M., Michailovich, O. and Wang, Z., 2012. Gradient-based surface reconstruction using compressed sensing. In *2012 19th IEEE International Conference on Image Processing* (pp. 913-916). IEEE.

Sajjan, S., Moore, M., Pan, M., Nagaraja, G., Lee, J., Zeng, A. and Song, S., 2020. Clear grasp: 3d shape estimation of transparent objects for manipulation. Proc. *IEEE ICRA*, pp. 3634-3642.

Sansoni, G., Trebeschi, M. and Docchio, F., 2009. State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, *9*(1), pp.568-601.

Santoši, Ž., Budak, I., Stojaković, V., Šokac, M. and Vukelić, Đ., 2019. Evaluation of synthetically generated patterns for image-based 3D reconstruction of texture-less objects. *Measurement*, *147*, p.106883.

Scherr, T., 2017. Gradient-based surface reconstruction and the application to wind waves (*Doctoral dissertation*).

Schonberger, J.L. and Frahm, J.M., 2016. Structure-from-motion revisited. Proc. *CVPR*, pp. 4104-4113.

Shi, B., Mo, Z., Wu, Z., Duan, D., Yeung, S.K. and Tan, P., 2018. A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(2), pp.271-284.

Shih, Y., Krishnan, D., Durand, F. and Freeman, W.T., 2015. Reflection removal using ghosting cues. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3193-3201).

Shmueli, J., Eder, M.A. and Tesauro, A., 2015. A versatile stereo photogrammetry based technique for measuring fracture mode displacements in structures. *Precision Engineering*, *39*, pp.38-46.

Smithwick, Q.Y. and Seibel, E.J., 2002. Depth enhancement using a scanning fiber optical endoscope. In *Optical Biopsy IV* (Vol. 4613, pp. 222-233). SPIE.

Solomon, F. and Ikeuchi, K., 1996. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(4), pp.449-454.

Stolz, C., Ferraton, M. and Meriaudeau, F., 2012. Shape from polarization: a method for solving zenithal angle ambiguity. *Optics Letters*, 37(20), pp.4218-4220.

Stolz, C., Kechiche, A.Z. and Aubreton, O., 2016. Short review of polarimetric imaging based method for 3D measurements. *In Optics, Photonics and Digital Technologies for Imaging Applications IV,* Vol. 9896, p. 98960P.

Sun, Z., Qiao, Y., Jiang, Z., Xu, X., Zhou, J. and Gong, X., 2020. An Accurate Fourier-Based Method for Three-Dimensional Reconstruction of Transparent Surfaces in the Shape-From-Polarization Method. *IEEE Access,* 8, pp.42097-42110.

Sunkavalli, K., Zickler, T. and Pfister, H., 2010, September. Visibility subspaces: Uncalibrated photometric stereo with shadows. In *European Conference on Computer Vision* (pp. 251-264). Springer, Berlin, Heidelberg.

Szeliski, R., 1993. Rapid octree construction from image sequences. *CVGIP: Image Understanding,* 58(1), pp.23-32

Tarini, M., Lensch, H.P., Goesele, M. and Seidel, H.P., 2005. 3D acquisition of mirroring objects using striped patterns. *Graphical Models,* 67(4), pp.233-259.

Tosovic, S., 2002. Adaptive 3D modeling of objects by combining shape from silhouette and shape from structured light. *Technical Report, PRIP, TU Wien, Austria.*

Tsai, C.Y., Veeraraghavan, A. and Sankaranarayanan, A.C., 2015. What does a single light-ray reveal about a transparent object? Proc. *IEEE ICIP*, pp. 606-610.

Tsai, D., Corke, P., Peynot, T. and Dansereau, D.G., 2021. Refractive Light-Field Features for Curved Transparent Objects in Structure from Motion. *IEEE Robotics and Automation Letters,* 6(4), pp. 6923-6930.

Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T. and Srinivasan, P.P., 2022. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5481-5490). IEEE.

Verein Deutscher Ingenieure; Verband Der Elektrotechnik. VDI/VDE 2634 BLATT 3, 2008. Optical 3D-Measuring Systems—Multiple View Systems Based on Area Scanning; *Engl. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik: Düsseldorf, Germany.*

Wallis, R.H., 1976. An approach for the space variant restoration and enhancement of images. In *Proc. Symposium on Current Mathematical Problems in Image Science, Nov*.

Wetzstein, G., Roodnick, D., Heidrich, W. and Raskar, R., 2011. Refractive shape from light field distortion. Proc. *ICCV,* pp. 1180-1186.

Woodham, R.J., 1980. Photometric method for determining surface orientation from multiple images. *Optical engineering*, *19*(1), pp.139-144.

Wu, B., Zhou, Y., Qian, Y., Gong, M. and Huang, H., 2018. Full 3D reconstruction of transparent objects. *arXiv preprint arXiv:*1805.03482.

Xiong, Y., Chakrabarti, A., Basri, R., Gortler, S.J., Jacobs, D.W. and Zickler, T., 2014. From shading to local shape. *IEEE transactions on pattern analysis and machine intelligence*, *37*(1), pp.67-79.

Xu, X., Qiao, Y. and Qiu, B., 2017. Reconstructing the surface of transparent objects by polarized light measurements. *Optics Express,* 25(21), pp.26296-26309.

Yang, R., Wang, Y., Liao, S. and Guo, P., 2023. Dpps: A Deep-Learning Based Point-Light Photometric Stereo Method for 3d Reconstruction of Metallic Surfaces. *Available at SSRN 4233893*.

Ye, J., Ji, Y., Li, F. and Yu, J., 2012. Angular domain reconstruction of dynamic 3d fluid surfaces. Proc. *CVPR*, pp. 310-317.

Yeung, S.K., Wu, T.P., Tang, C.K., Chan, T.F. and Osher, S.J., 2014. Normal estimation of a transparent object using a video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(4), pp.890-897.

Yeung, S.K., Wu, T.P., Tang, C.K., Chan, T.F. and Osher, S.J., 2015. Normal Estimation of a Transparent Object Using a Video. *IEEE Transactions on Pattern Analysis & Machine Intelligence,* 37(04), pp.890-897.

Yu, Z., Peng, S., Niemeyer, M., Sattler, T., Geiger, A., 2022. MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. Proc. NIPS.

Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T. and Barron, J.T., 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics (TOG), 40(6), pp.1-18.

Zheng, Q., Kumar, A., Shi, B. and Pan, G., 2019. Numerical reflectance compensation for non-lambertian photometric stereo. *IEEE Transactions on Image Processing*, *28*(7), pp.3177-3191.

Zhu, L., Mousavian, A., Xiang, Y., Mazhar, H., van Eenbergen, J., Debnath, S. and Fox, D., 2021. RGB-D Local Implicit Function for Depth Completion of Transparent Objects. Proc. *IEEE CVPR,* pp. 4649-4658.

Zollhöfer, M., Stotko, P., Görlitz, A., Theobalt, C., Nießner, M., Klein, R. and Kolb, A., 2018. State of the art on 3D reconstruction with RGB-D cameras. In *Computer graphics forum* (Vol. 37, No. 2, pp. 625-652).

Zuo, X., Du, C., Wang, S., Zheng, J. and Yang, R., 2015. Interactive visual hull refinement for specular and transparent object surface reconstruction. Proc. *IEEE CVPR,* pp. 2237-2245.