

Thermal Polarimetric Multi-view Stereo

Takahiro Kushida Kenichiro Tanaka
Ritsumeikan University
2-150 Iwakura-cho, Ibaraki, Osaka, Japan
{tkushida, ken-t}@fc.ritsumei.ac.jp

Abstract

This paper introduces a novel method for detailed 3D shape reconstruction utilizing thermal polarization cues. Unlike state-of-the-art methods, the proposed approach is independent of illumination and material properties. In this paper, we formulate a general theory of polarization observation and show that long-wave infrared (LWIR) polarimetric imaging is free from the ambiguities that affect visible polarization analyses. Subsequently, we propose a method for recovering detailed 3D shapes using multi-view thermal polarimetric images. Experimental results demonstrate that our approach effectively reconstructs fine details in transparent, translucent, and heterogeneous objects, outperforming existing techniques.

1. Introduction

3D shape reconstruction is a fundamental problem in computer vision, with applications spanning both academia and industry. Over the years, various 3D reconstruction methods using RGB cameras have been proposed, but many of these methods rely on strong assumptions about illumination and reflection models. For instance, multi-view stereo [19, 44, 45, 47] depends on surface texture to find correspondences, while structured light [16, 43], photometric stereo [10, 20, 41, 50], and shape from polarization [1, 2, 32, 34, 46, 49] require specific lighting conditions and assume opaque surfaces, limiting their applicability to a narrow range of materials.

Thermal imaging presents an attractive alternative for 3D reconstruction in challenging scenarios because it does not rely on environmental illumination; any object with heat emits long-wave infrared (LWIR) light, effectively serving as its own light source. Furthermore, most materials, except for those designed for thermal optics, are opaque in the LWIR spectrum, eliminating the need for many of the assumptions required by traditional reflection models. Building on this concept, several methods, such as thermal multi-view stereo [53] and absorption-based depth es-

timation [9, 14, 24, 35] have been proposed. Although these methods demonstrate potential applicability, their results also imply the limitation inherent in model-based approaches, especially with regard to accuracy.

Photometric cues in the LWIR spectrum are useful for recovering the accurate shape, similar to the visible-light spectrum. For example, thermal photometric stereo [48] and shape from heat conduction [36] have demonstrated detailed surface reconstruction. However, these methods require active heating and cooling of the object, making them both time-consuming and impractical. In this paper, we aim to reconstruct fine 3D shapes in a steady state using polarimetric cues, a type of photometric cue that can be obtained without monitoring the heating or cooling processes.

Table 1 presents a comparison of various 3D reconstruction approaches. Our approach is illumination-independent, material-independent, free from heating or cooling process, and highly accurate. Our contributions are twofold. First, we establish a unified theoretical framework for LWIR polarization, showing that, unlike visible-light polarization, it is not affected by specular-diffuse ambiguities. Second, we show that LWIR polarization serves as a powerful cue for detailed 3D shape reconstruction. Experimental results confirm that our method outperforms the existing approaches on heterogeneous materials.

2. Related Work

2.1. Shape from visible polarization

Shape from Polarization (SfP), which recovers surface normals from the polarization state of reflected light, has a long-standing history in computer vision [1, 2, 32, 34, 46, 49]. Recently, the advent of snapshot polarization cameras has significantly boosted this area of research [40].

SfP approaches continue to tackle ambiguities in surface normals that arise from the mixture of specular and diffuse polarizations [46]. One effective way to resolve these ambiguities is to use multi-view polarization images, which provide a richer set of constraints on the surface normals and lead to more accurate and robust surface reconstruc-

Method	Illumination	Material	Accuracy	Measurement
Visible Multi-view Stereo [19, 44, 45, 47]	Dependent	Textured only	High	Easy
Visible SL / PS [10, 16, 20, 41, 43, 50]	Dependent	Opaque only	High	Easy
Visible Polarization [1, 2, 32, 34, 46, 49]	Dependent	Dependent	High	Easy
Visible NeRFs [30, 51]	Dependent	Wide range of materials	Low	Easy
Thermal Multi-view Stereo [53]	Independent	Textured only	High	Easy
Thermal SL [11]	Independent	Diffuse reflective only	High	Easy
Thermal PS [48]	Independent	Wide range of materials	Moderate	Heating & cooling
Thermal NeRFs [28, 52]	Independent	Wide range of materials	Low	Easy
Heat Conduction [36]	Independent	Wide range of materials	Moderate	Heating
Thermal Polarization (Ours)	Independent	Wide range of materials	High	Easy

Table 1. Comparison table of shape reconstruction methods. While other methods have pros and cons, our approach is independent of illumination and material’s optical properties as well as accurate shape recovery is possible in steady state measurements.

tions [8]. Another approach involves using polarimetric BRDF models, which can handle complex material properties [3, 13, 21, 55]. Recent approaches incorporate neural scene representations and differentiable rendering, enabling the joint estimation of shape, environment maps, and reflectance [18, 27]. Nevertheless, SfP in the visible light spectrum remains fundamentally sensitive to the illumination conditions, such as the assumption of uniform, unpolarized ambient light, and to variations in material’s optical properties, including transparency, translucency, matte, or shiny finishes [46].

In this paper, we introduce the use of polarization in the LWIR spectrum. The LWIR light is emitted directly from the object and is independent of external illumination. Moreover, the polarization of LWIR emission is not mixed up with the diffuse and specular reflections or transmission effects. These properties make LWIR polarization advantageous for robust shape reconstruction.

2.2. LWIR imaging and physics-based vision

Beyond 3D reconstruction, thermal imaging is utilized for a variety of physics-based vision tasks. The analysis of heat conduction, for instance, has been leveraged for material classification by exploiting differences in thermal diffusivity and conductivity [42]. Reflection analysis of LWIR light has facilitated non-line-of-sight (NLoS) imaging, expanding scene understanding by examining how thermal rays propagate around the corner [23, 29]. Joint analyses of both visible and LWIR light transport have further enriched scene understanding [38]. We also advance the area of physics-based thermal vision through a shape reconstruction using LWIR polarization.

Thermal imaging is an effective approach for 3D shape reconstruction in some challenging scenarios. For example, it is effective for stealth observation as the thermal radiation does not rely on external illumination [6, 35, 52, 53]. Another example is that it is feasible to observe the shape of challenging materials in the visible light spectrum because many materials that appear transparent or translucent are opaque in the LWIR spectrum [48].

Existing 3D reconstruction techniques using thermal cameras can be broadly categorized into geometric and photometric approaches. Geometric methods [12, 31, 53] such as based on stereo/multi-view observations recover surface depth by matching corresponding points across thermal images, commonly assuming near-Lambertian emission or sufficient texture. Photometric approaches infer depth, shape, or surface normal from intensity cues, which include thermal photometric stereo [48], absorption-based method [9, 14, 24, 35], and heat conduction-based method [36]. While they show strong possibilities of shape reconstruction using a thermal camera, their accuracy is generally limited. There are hybrid approaches that combine visible-light and LWIR polarization [34], where the LWIR polarization is used only as an auxiliary cue to resolve the specular–diffuse ambiguity in visible polarization rather than directly contributing to shape recovery.

Recently, neural rendering methods inspired by Neural Radiance Fields (NeRF) have been explored in the thermal domain [6, 28, 52] as well. By jointly optimizing a volumetric scene representation and a rendering function to satisfy the rendered images and align observed images from multiple viewpoints, these thermal NeRF techniques recover both geometry and radiometric properties. We are also inspired by this neural representation and model the object’s surface by a neural implicit surface.

In this paper, we leverage the polarization state of thermal radiation to enrich the information available for 3D reconstruction. By incorporating the polarization cues, geometric rotation, and neural representation, our method enables more accurate shape estimation.

3. Polarimetric LWIR Observation

In this section, we present a general theory of polarization observation and discuss the advantages of thermal polarization over visible polarization.

3.1. Physics of Polarization

The polarization state of light, including both visible and LWIR spectra, changes when it is reflected from or trans-

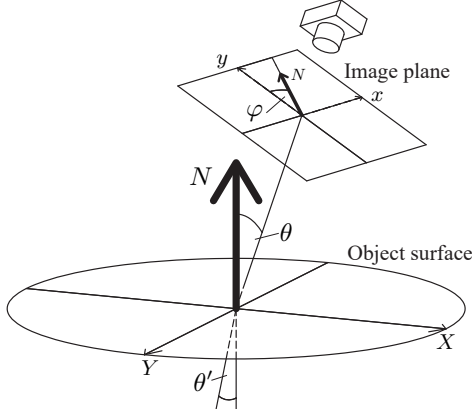


Figure 1. Coordinate systems and the zenith and projected azimuth angles of the surface normal. While Mueller calculus is performed in the object's surface coordinate system, the observed Stokes parameter and projected azimuth angle are represented in the image coordinate system.

mitted through an object's surface. The transport of polarized light can be described using the Stokes parameters $\mathbf{s} = [s_0, s_1, s_2, s_3]^T$ and the Mueller matrix \mathbf{M} as

$$\mathbf{s}_o = \mathbf{M} \mathbf{s}_i, \quad (1)$$

where \mathbf{s}_i and \mathbf{s}_o denote the Stokes parameters of the incoming and outgoing light, respectively, both defined in the surface coordinate system [7].

The observed scene is a combination of specular reflection, diffuse reflection, transmission, and emission. Consequently, the overall polarization state \mathbf{s} can be expressed as

$$\mathbf{s} = \mathbf{s}_s + \mathbf{s}_d + \mathbf{s}_t + \mathbf{s}_e, \quad (2)$$

where \mathbf{s}_s , \mathbf{s}_d , \mathbf{s}_t , and \mathbf{s}_e represent the Stokes vectors corresponding to the specular, diffuse, transmission, and emission components, respectively.

Specular polarization The change in the polarization state due to specular reflection [3, 22] is described by

$$\mathbf{s}_s = \mathbf{C}(\varphi) \mathbf{F}^R(\theta) \mathbf{s}_i, \quad (3)$$

where \mathbf{s}_i denotes the Stokes vector of the incident light in the object's surface coordinate system, \mathbf{s}_s denotes the corresponding Stokes vector in the image coordinate system, $\mathbf{C}(\varphi)$ is a rotation matrix that transforms coordinates from the image to the object's surface, φ is the projected azimuth angle (*i.e.* the angle between the vertical axis of the image plane and the surface normal projected into the image plane), $\mathbf{F}^R(\theta)$ is the Mueller matrix for the specular component, and θ is the zenith angle of the reflection (*i.e.* the angle between the surface normal and the viewing direction). The

surface coordinate system, image coordinate system, zenith angle, and projected azimuth angle are illus in Fig. 1. The rotation matrix $\mathbf{C}(\varphi)$ is given by

$$\mathbf{C}(\varphi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\varphi & -\sin 2\varphi & 0 \\ 0 & \sin 2\varphi & \cos 2\varphi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

The Mueller matrix $\mathbf{F}^R(\theta)$ is derived from Fresnel's law and expressed as

$$\mathbf{F}^R(\theta) = \begin{bmatrix} r_+ & r_- & 0 & 0 \\ r_- & r_+ & 0 & 0 \\ 0 & 0 & r_\times \cos \delta & r_\times \sin \delta \\ 0 & 0 & -r_\times \sin \delta & r_\times \cos \delta \end{bmatrix}, \quad (5)$$

where $r_\pm = \frac{r_s \pm r_p}{2}$, $r_\times = \sqrt{r_p r_s}$, θ' denotes the refractive angle. Here, $\cos \delta$ equals -1 when θ is less than Brewster's angle and 1 otherwise, with the sign of $\sin \delta$ reversed accordingly [3]. The Fresnel reflectances for s- and p-polarization, r_s and r_p , are given by

$$r_s(\theta) = \left(\frac{\cos \theta - \eta \cos \theta'}{\cos \theta + \eta \cos \theta'} \right)^2, r_p(\theta) = \left(\frac{\cos \theta' - \eta \cos \theta}{\cos \theta' + \eta \cos \theta} \right)^2. \quad (6)$$

According to Snell's law, the refractive angle is given as

$$\theta' = \sin^{-1} \left(\frac{1}{\eta} \sin \theta \right), \quad (7)$$

where η represents the refractive index of the object.

Diffuse polarization The change in the polarization state due to diffuse reflection or subsurface scattering [1, 22] is described by

$$\mathbf{s}_d = \mathbf{C}(\varphi) \mathbf{F}^T(\theta) \int_{\Omega} \mathbf{D} \mathbf{F}^T(\theta) \mathbf{s}_{i,\omega} d\omega, \quad (8)$$

where $\mathbf{s}_{i,\omega}$ denotes the Stokes vector of incident light coming from the direction ω . The matrix \mathbf{D} , which describes the depolarization, is defined as

$$\mathbf{D} = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (9)$$

where ρ denotes the proportion of the depolarization. The Mueller matrix for transmittance $\mathbf{F}^T(\theta)$ is given by

$$\mathbf{F}^T(\theta) = \begin{bmatrix} t_+ & t_- & 0 & 0 \\ t_- & t_+ & 0 & 0 \\ 0 & 0 & t_\times \cos \delta & t_\times \sin \delta \\ 0 & 0 & -t_\times \sin \delta & t_\times \cos \delta \end{bmatrix}, \quad (10)$$

where $t_{\pm} = \frac{t_s \pm t_p}{2}$, $t_{\times} = \sqrt{t_p t_s}$. t_p and t_s denote the Fresnel transmittance coefficients for s- and p- polarization, respectively, and are expressed as

$$t_s(\theta) = \left(\frac{2 \cos \theta}{\cos \theta + \eta \cos \theta'} \right)^2, t_p(\theta) = \left(\frac{2 \cos \theta}{\cos \theta' + \eta \cos \theta} \right)^2. \quad (11)$$

Transmission polarization The polarization state also changes due to light transmission. According to the Fresnel's law, the transmission polarization is described by

$$\mathbf{s}_t = \mathbf{C}(\varphi) \mathbf{F}^T(\theta) \mathbf{s}_i, \quad (12)$$

where \mathbf{s}_i denotes the Stokes vector of the light coming from under the surface.

Emission polarization The polarization state of emitted light is described by

$$\mathbf{s}_e = \mathbf{C}(\varphi) \mathbf{E}(\theta) \mathbf{s}_c, \quad (13)$$

where \mathbf{s}_c denotes the Stokes vector of the emitted light including thermal radiation as well as any chemical or electrical emissions and $\mathbf{E}(\theta)$ is the Mueller matrix corresponding to the emission component. Since there are many types of light emission, the elements of the emission Mueller matrix are not specified explicitly in general.

3.2. LWIR Polarization Observation

Polarization of visible light Since objects other than light sources do not emit light in the visible spectrum, the emission component is generally ignorable. Therefore, the observed polarization can be modeled as

$$\mathbf{s} = \mathbf{s}_d + \mathbf{s}_s + \mathbf{s}_t. \quad (14)$$

The magnitude of each component varies in complex ways depending on the object's material properties and the illumination environment. This complexity makes polarization analyses more challenging, leading existing visible polarization studies to rely on various assumptions to facilitate the analysis.

Polarization of LWIR light Most objects are known to be opaque in the LWIR spectrum; therefore, the transmission component can be safely ignored. Moreover, an interesting property of LWIR light polarization is that the relative contributions of each component can be controlled by the object's temperature. In a typical environments, where no objects are extremely hot in the surroundings (*i.e.*, where there are no significant LWIR light sources in the surroundings) the magnitude of reflection components is small compared to that of the emission. In such cases, the LWIR polarization observation can be modeled as

$$\mathbf{s} = \mathbf{s}_e. \quad (15)$$

Since LWIR emission primarily arises from black body radiation, Eq. (13) can be rewritten as

$$\mathbf{s}_e = \mathbf{C}(\varphi) \mathbf{E}(\theta) \mathbf{s}_B(T), \quad (16)$$

where $\mathbf{E}(\theta)$ denotes the Mueller matrix for black body radiation, and $\mathbf{s}_B(T)$ represents the Stokes vector of the radiation, which can be expressed as

$$\mathbf{s}_B(T) = \begin{bmatrix} B(T) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (17)$$

where B denotes the black body radiation. According to Planck's law, B is given by

$$B(T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1}, \quad (18)$$

where c is the speed of light, k is the Boltzmann constant, h is the Planck constant, T is the object's absolute temperature, and λ is the observation wavelength. Equation (17) demonstrates that the object's temperature T controls the power of emission.

The Mueller matrix of black body radiation $\mathbf{E}(\theta)$ [26, 33] is given by

$$\mathbf{E}(\theta) = \begin{bmatrix} \varepsilon_+ & 0 & 0 & 0 \\ \varepsilon_- & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (19)$$

where $\varepsilon_{\pm} = \frac{\varepsilon_s \pm \varepsilon_p}{2}$. ε_p and ε_s denote the emissivities for s- and p-polarization, respectively. According to the energy conservation law, also known as Kirchhoff's law, the sum of reflectance, transmittance, and emissivity is equal to 1. Since the transmittance is negligible, as discussed above, we obtain

$$\varepsilon_s(\theta) = 1 - r_s(\theta) \quad (20)$$

$$\varepsilon_p(\theta) = 1 - r_p(\theta). \quad (21)$$

In summary, in the LWIR spectrum, the Stokes observation \mathbf{s} can be analytically represented without ambiguity, thereby simplifying the analysis of the polarization state, which is a significant advantage for shape reconstruction.

Polarization to surface normal: LWIR polarization provides cues for an object's surface normal, which can be determined analytically. The degree of linear polarization (DoLP) is related to the zenith angle θ as follows:

$$\text{DoLP} = \frac{\sqrt{s_1^2 + s_2^2}}{s_0} \quad (22)$$

$$= \frac{\left(\eta - \frac{1}{\eta} \right)^2 \sin^2 \theta}{2 + 2\eta^2 - \left(\eta + \frac{1}{\eta} \right) \sin^2 \theta + 4 \cos \theta \sqrt{\eta^2 - \sin^2 \theta}}. \quad (23)$$

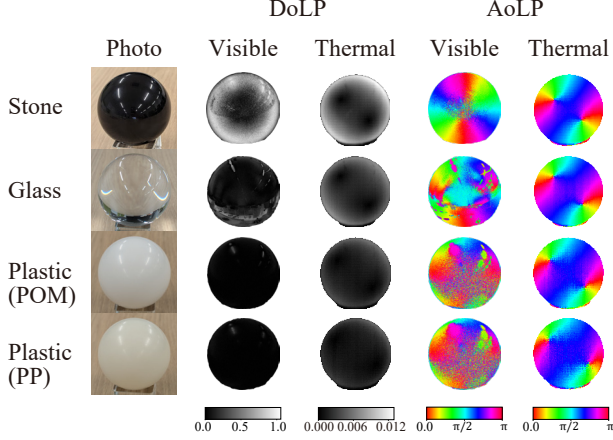


Figure 2. Visible and thermal polarization observations of spheres made from various materials under natural room lighting. While visible polarization is unstable due to variations in a material’s optical properties, thermal polarimetric cues, especially AoLP images, remain consistent across different materials.

The angle of linear polarization (AoLP) is equivalent to the azimuth angle φ as follows:

$$\text{AoLP} = \frac{1}{2} \tan^{-1} \left(\frac{s_2}{s_1} \right) = \varphi. \quad (24)$$

These correspondences enable the reconstruction of the surface normal from polarization observations. Especially, AoLP serves as a robust cue because it directly reflects the azimuth angle of the surface normal and is independent of material properties, whereas DoLP is material-dependent due to the inclusion of the refractive index term.

3.3. Examples of Visible vs. LWIR Polarization

In the visible light spectrum, specular, diffuse, and transmission polarization components are often mixed [2, 25, 46], resulting in complicated observation. Fig. 2 shows the polarimetric observations of spheres made from various materials; in the visible spectrum, unwanted polarization observations frequently appear. For example, because black stone exhibits strong specular polarization, the DoLP image is heavily affected by reflections from the surrounding environment. In the case of transparent glass sphere, the observation is notably contaminated by background textures. For translucent plastic materials, both DoLP and AoLP images are affected by surface texture and subsurface scattering. These unstable observations significantly complicate subsequent shape recovery.

In contrast, in the LWIR spectrum, polarimetric observations remain consistent and independent of both material properties and environmental reflections, particularly in AoLP images. This consistency is a key advantage of thermal polarization cues, as it provides robustness against vari-

ations in materials and surrounding reflections.

4. Thermal Polarimetric Multi-view Stereo

Our goal is to recover the 3D shape using thermal polarimetric cues. As discussed in the previous section, the AoLP observation is invariant to material properties. To leverage this strong cue, we only use the AoLP images for material-independent 3D shape reconstruction.

Our approach employs multi-view thermal AoLP images to estimate the Signed Distance Function (SDF), which is implicitly represented by a Multi-Layer Perceptron (MLP) optimized via differentiable rendering. To effectively incorporate the surface normal information present in AoLP images, we adopt the concept of tangent space constraint (TSC) proposed in a multi-view azimuth stereo (MVAS) [5].

Shape representation

The SDF is implicitly represented by an MLP $f(\mathbf{x}; \boldsymbol{\theta}) : \mathbb{R}^3 \times \mathbb{R}^d \rightarrow \mathbb{R}$, where $\mathbf{x} \in \mathbb{R}^3$ denotes a point in space and $\boldsymbol{\theta} \in \mathbb{R}^d$ are the learnable parameters of the MLP. The zero-level set of the SDF defines the object’s surface \mathcal{M} as [37]

$$\mathcal{M}(\boldsymbol{\theta}) = \{\mathbf{x} | f(\mathbf{x}; \boldsymbol{\theta}) = 0\}. \quad (25)$$

The MLP is optimized within the framework of implicit differentiable renderer (IDR) [51].

Loss function

We adopt the loss function proposed in MVAS [5], which is defined as,

$$\mathcal{L} = \mathcal{L}_{\text{TSC}} + \lambda_1 \mathcal{L}_{\text{silhouette}} + \lambda_2 \mathcal{L}_{\text{Eikonal}}, \quad (26)$$

where \mathcal{L}_{TSC} is the tangent space consistency loss, $\mathcal{L}_{\text{silhouette}}$ is the silhouette loss and $\mathcal{L}_{\text{Eikonal}}$ is the Eikonal regularization. λ_1 and λ_2 are the weights assigned to the respective loss terms.

Tangent space consistency loss [5] The AoLP captured from different camera positions enforces a strong constraint on surface normal estimation. For a unit normal vector $\mathbf{n} \in \mathcal{S}^2 \in \mathbb{R}^3$ at a surface point $\mathbf{x} \in \mathbb{R}^3$ and the projected azimuth angle φ at the corresponding pixel, the following relationship holds:

$$\mathbf{r}_1^\top \mathbf{n} \cos \varphi - \mathbf{r}_2^\top \mathbf{n} \sin \varphi = 0, \quad (27)$$

where $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]^\top$ denotes the rotation matrix of the camera pose. This equation can be rearranged as

$$\mathbf{n}^\top \underbrace{(\mathbf{r}_1 \cos \varphi - \mathbf{r}_2 \sin \varphi)}_{\mathbf{t}(\varphi)} = 0, \quad (28)$$

where $t(\varphi)$ is called a projected tangent vector and the above equation is called tangent space constraint. The TSC loss is the sum of the square error of the tangent space constraint and is defined as

$$\mathcal{L}_{\text{TSC}} = \frac{1}{P} \sum_{\mathbf{x} \in \mathbf{X}} \frac{\sum_{i=1}^N \Phi_i (\mathbf{n}^\top(\mathbf{x}; \boldsymbol{\theta}) \mathbf{t}_i(\mathbf{x}))^2}{\sum_{i=1}^N \Phi_i} \quad (29)$$

where P denotes the number of pixel samples, Φ_i is a binary indicator of the visibility of point \mathbf{x} from the i -th view, $\mathbf{n}^\top(\mathbf{x}; \boldsymbol{\theta})$ is the estimated surface normal at the point \mathbf{x} given network parameters $\boldsymbol{\theta}$, and $\mathbf{t}_i(\mathbf{x})$ is the tangent vector corresponding to the i -th view's pixel onto which the point \mathbf{x} is projected.

Silhouette loss [5, 51] The silhouette loss constrains the visual hull of the shape and is defined as

$$\mathcal{L}_{\text{silhouette}} = \frac{1}{\alpha P} \sum_{\mathbf{x} \in \tilde{\mathbf{X}}} \Psi(O(\Pi(\mathbf{x})), \sigma(\alpha f^*)), \quad (30)$$

where Ψ denotes the cross entropy and O is a binary mask indicating whether a pixel lies inside the object silhouette (as provided by the input). P is the number of pixel samples, Π is the projection from the world coordinate to the image coordinate, $\tilde{\mathbf{X}}$ represents the pixels with no ray-surface interactions or outside the silhouette, f^* represents the minimum SDF distance between the ray that does not have intersections and the object surface, σ is a sigmoid function, and α controls the sharpness of the sigmoid function.

Eikonal regularization [17] The Eikonal loss regularizes the gradient of SDF, enforcing the SDF to maintain its property of having a gradient close to 1:

$$\mathcal{L}_{\text{Eikonal}} = \mathbb{E}_{\mathbf{x}} ((\|\mathbf{n}(\mathbf{x}; \boldsymbol{\theta})\|_2 - 1)^2). \quad (31)$$

$\mathbb{E}_{\mathbf{x}}(\cdot)$ denotes the expected value computed over sample points \mathbf{x} that are randomly sampled within the bounding box.

5. Experiments

5.1. Setup

Implementation Following MVAS [5], we use an 8-layer MLP with softplus activations, combined with positional encoding (dimension 10). The input 3D position and encoded features are skip-connected to the 4th layer of the network. The optimization is performed using the Adam optimizer with a learning rate of 0.001. The weights for the loss terms are set to $\lambda_1 = 50$ (silhouette loss) and $\lambda_2 = 0.1$ (Eikonal loss). Training is conducted for 50 epochs with a batch size of 4096 pixels, and the silhouette loss weight is halved every 10 epochs. Training takes approximately 2 hours on a single RTX A6000 GPU.

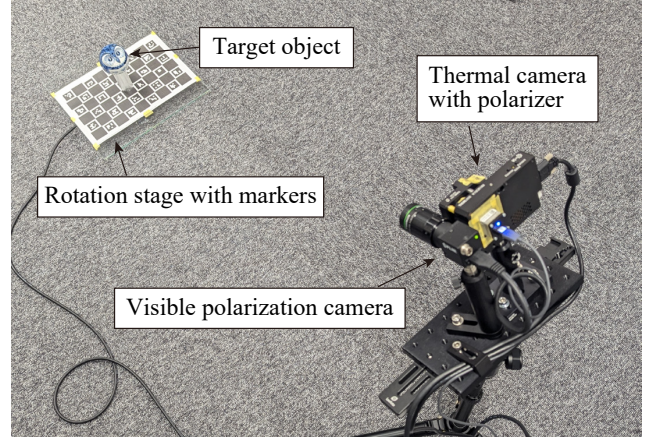


Figure 3. Camera system and setup. The system consists of a rotation stage to rotate the target and a thermal camera with a wire-grid polarizer. A visible polarization camera is used for camera pose estimation and for comparison with the visible MVAS.

Camera system Figure 3 shows our camera system and experimental setup. Our thermal polarization camera system consists of a thermal camera (FLIR Boson 320) with 15 mm lens and a wire-grid polarizer (Thorlabs WP25M-IRC) mounted on a motorized rotation mount (Thorlabs K10CR2). The polarizer is installed between the image sensor and imaging lens to mitigate the narcissus reflections. Thermal polarization images are captured by rotating the polarizer using the rotation mount. We captured four polarization images at 0, 45, 90, and 135 degrees. A monochrome visible-light polarization camera (FLIR Blackfly S BFS-U3-51S5P) is also placed adjacent to the thermal camera. The visible-light polarization camera is used both to capture images for a baseline method and to get the poses of both visible and thermal cameras.

Camera poses We employed a marker-based approach to estimate the camera poses. We put Aruco markers [15] on the rotation stage to estimate the pose of visible-light camera. Since printed markers are not visible in the thermal images, the poses of the thermal camera were obtained by transforming the camera poses of the visible-light camera. For stereo calibration between the visible-light and thermal cameras, we used a calibration target that can be observed by both cameras. The calibration board consists of a white-painted aluminum pegboard placed in front of a black surface heater. By heating the back panel, the circle grid pattern became visible to both cameras. The calibration parameters were estimated using Zhang’s method [54].

Dataset We prepared a real-world dataset with ground-truth shapes. We captured 7 objects with various shapes and materials. The target objects include transparent and

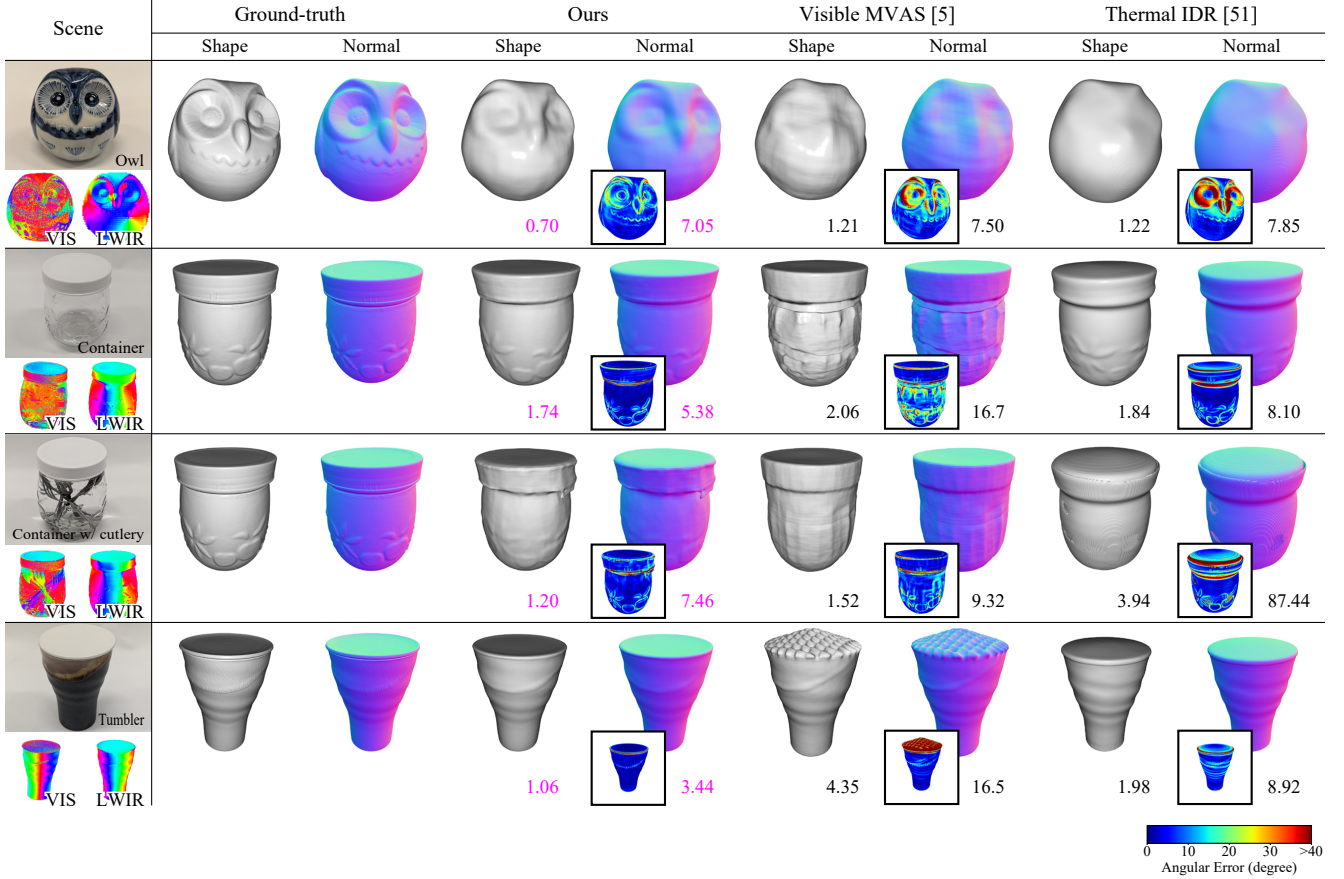


Figure 4. Experimental results. AoLP images in both visible and LWIR spectra are shown below the scene photograph. Each method’s result includes the recovered shape, surface normals, and angular error maps. The Chamfer distance is displayed below the estimated shape and the mean angular error is displayed to the right of the angular error map. For both metrics, lower values indicate better performance. The smallest errors are shown in magenta. It demonstrates that the proposed method outperforms the other methods.

low-reflective objects, which pose challenges for existing reconstruction methods. The objects were placed on a rotating table and captured from about 20–30 views by rotating the table. To robustly capture thermal polarization, the objects were kept warm during the measurements. Masks for thermal images were generated by thresholding pixel values, while masks for visible images were generated by using Segment Anything Model 2 [39]. Ground-truth shapes were obtained using a structured light 3D scanner (EinScan-SP V2). To facilitate the ground-truth measurement, the surfaces of the objects were painted white to make them opaque.

Baselines We compare our proposed method with two state-of-the-art methods: MVAS [5] based on visible-light polarization cues and IDR [51], which is one of the neural implicit surface reconstruction methods, with thermal image input. For MVAS with visible-light polarization images, there exists $\pm \frac{\pi}{2}$ ambiguity in surface normals due to

the mixture of specular and diffuse polarization. Therefore, TSC loss is modified to account for the $\pm \frac{\pi}{2}$ ambiguity, as originally proposed in MVAS. Since our method does not exhibit this ambiguity, the modified loss function is applied only to visible light observation.

Evaluation metrics We use two metrics to evaluate both the surface points and surface normals. Chamfer distance is the metric for evaluating the geometric error, which calculates the point cloud distance between the estimate and the ground-truth. The mean angular error is the metric for surface normal, which calculates the angular error of normal maps. For both metrics, a smaller value indicates a better result. Prior to computing these metrics, the estimated and the ground-truth meshes are aligned using ICP algorithm [4].

5.2. Results

Figure 4 shows the results. The target objects are an owl made of ceramic with a clear coating, a glass container with

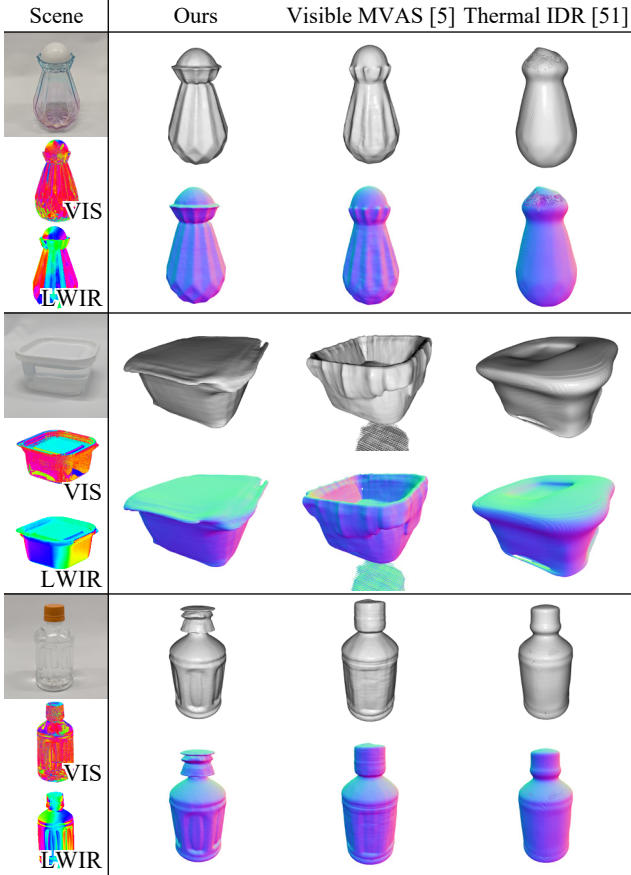


Figure 5. Qualitative results for other objects. AoLP images are shown below the scene photos. For each method, the upper image is the estimated shape and the lower is the corresponding normal map. It is shown that our method successfully reconstructs fine details of the object, particularly the concave parts in the transparent vase and bottle, while other methods fail to reconstruct such details.

and without interiors, and a black tumbler, some of which are challenging objects for the existing approaches, especially a container with cutlery. In AoLP observations, while visible light is affected by a mixture of reflections and transmission components, thermal observations remain consistent and align with the azimuth angle of the surface normal. This consistency leads to a clear difference in detailed 3D reconstruction results between visible polarization and thermal polarization.

In the owl’s result, while the detailed shape such as the eyes and the beak is missing in both visible MVAS and thermal IDR, they are successfully recovered in our result. Similarly, the relief pattern on the glass container is clearly visible in our results. For a black tumbler with a white lid shown in the fourth row, the AoLP observation on the lid in the visible-light spectrum is too noisy. In such a case, the surface normal of the lid is not reconstructible, resulting in

a wavy surface. In contrast, our method recovers a reasonable shape due to the presence of a unique AoLP on the lid in LWIR polarization. Overall, visible MVAS tends to reconstruct with wavy artifacts as AoLP is noisy, and thermal IDR tends to recover swollen shapes as it does not consider the surface normal.

To quantitatively compare the results, the estimated shapes are evaluated using Chamfer distance, and the surface normals are evaluated by mean angular error. Our method outperforms the other methods in both Chamfer distance and the mean angular error for all objects.

Figure 5 shows additional qualitative results. Our results are visually better than the other methods. The concave details of the vase and bottle are especially notable, as they are precisely reconstructed in our method while other results appear blurred. In addition, the surface of the container’s lid is only reconstructed by our method while there is a big hole in the results of the others. Some artifacts are seen in thin parts in our result, particularly in the neck of the bottle. This artifact comes from the low spatial resolution of the thermal camera, which can be reduced by using a higher-resolution camera or increasing camera views.

6. Conclusion

In this paper, we propose a 3D shape reconstruction using multi-view thermal polarimetric observations. We theoretically show that thermal polarimetric cues, especially AoLP images, are independent of material properties and illumination environment. Based on this property, we demonstrate a detailed 3D shape recovery using multi-view thermal AoLP images and show that our approach is applicable to heterogeneous objects including transparent and low-reflective objects.

Although the proposed method performs well, some challenges remain in expanding the applicable scenes. Our approach assumes the sufficient power of the emission component compared to the reflection component. This assumption is not valid when the hot objects are placed near the target object or when the target’s temperature is lower than the environment. Moreover, our method is less effective on metallic surfaces or rough surfaces, where the polarization signal can become unstable or noisy. In such cases, decomposing the emission and reflection components [48] before applying the shape recovery or integrating additional cues may be a good option. Improving the observation system is also an important factor since our capturing system needs some time to rotate the polarization filter. A single-shot polarized LWIR camera could accelerate the thermal polarization analysis in the near future, as in the visible-light spectrum.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP24K20802, JP22K18420, JP24K02965.

References

- [1] G.A. Atkinson and E.R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE TIP*, 15(6):1653–1664, 2006. 1, 2, 3
- [2] Yunhao Ba, Alex Gilbert, Franklin Wang, Jinfa Yang, Rui Chen, Yiqin Wang, Lei Yan, Boxin Shi, and Achuta Kadambi. Deep Shape from Polarization. In *ECCV*, pages 554–571, 2020. 1, 2, 5
- [3] Seung-Hwan Baek, Daniel S. Jeon, Xin Tong, and Min H. Kim. Simultaneous acquisition of polarimetric SVBRDF and normals. *ACM TOG*, 37(6):1–15, 2018. 2, 3
- [4] P.J. Besl and Neil D. McKay. A method for registration of 3-D shapes. *IEEE TPAMI*, 14(2):239–256, 1992. 7
- [5] Xu Cao, Hiroaki Santo, Fumio Okura, and Yasuyuki Matsushita. Multi-View Azimuth Stereo via Tangent Space Consistency. In *CVPR*, pages 825–834, 2023. 5, 6, 7
- [6] Qian Chen, Shihao Shu, and Xiangzhi Bai. Thermal3D-GS: Physics-Induced 3D Gaussians for Thermal Infrared Novel-View Synthesis. In *ECCV*, pages 253–269, 2025. 2
- [7] Charly Collin, Sumanta Pattanaik, Patrick LiKamWa, and Kadi Bouatouch. Computation of polarized subsurface BRDF for rendering. In *Proceedings of Graphics Interface 2014*, pages 201–208. Canadian Information Processing Society, 2014. 3
- [8] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric Multi-view Stereo. In *CVPR*, pages 369–378, 2017. 2
- [9] Unay Dorken Gallastegi, Hoover Rueda-Chacón, Martin J. Stevens, and Vivek K Goyal. Absorption-based hyperspectral thermal ranging: Performance analyses, optimization, and simulations. *Optics Express*, 32(1):151, 2024. 1, 2
- [10] Jean-Denis Durou, Maurizio Falcone, Yvain Quéau, and Silvia Tozza. A Comprehensive Introduction to Photometric 3D-Reconstruction. In *Advances in Photometric 3D-Reconstruction*, pages 1–29. Springer International Publishing, 2020. 1, 2
- [11] Jack Erdozain, Kazuto Ichimaru, Tomohiro Maeda, Hiroshi Kawasaki, Ramesh Raskar, and Achuta Kadambi. 3d Imaging For Thermal Cameras Using Structured Light. In *ICIP*, pages 2795–2799, 2020. 2
- [12] Gonen Eren, Olivier Aubreton, Fabrice Meriaudeau, L.A. Sanchez Secades, David Fofi, A. Teoman Naskali, Frederic Truchetet, and Aytul Ercil. Scanning from heating: 3D shape estimation of transparent objects from local surface heating. *Optics Express*, 17(14):11457, 2009. 2
- [13] Yoshiki Fukao, Ryo Kawahara, Shohei Nobuhara, and Ko Nishino. Polarimetric Normal Stereo. In *CVPR*, pages 682–690, 2021. 2
- [14] Unay Dorken Gallastegi, Hoover Rueda-Chacón, Martin J. Stevens, and Vivek K Goyal. Absorption-Based, Passive Range Imaging from Hyperspectral Thermal Measurements. *IEEE TPAMI*, pages 1–18, 2025. 1, 2
- [15] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *PR*, 47(6):2280–2292, 2014. 6
- [16] Jason Geng. Structured-light 3D surface imaging: A tutorial. *Advances in Optics and Photonics*, 3(2):128, 2011. 1, 2
- [17] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit Geometric Regularization for Learning Shapes, 2020. 6
- [18] Yufei Han, Heng Guo, Koki Fukai, Hiroaki Santo, Boxin Shi, Fumio Okura, Zhanyu Ma, and Yunpeng Jia. NeRSP: Neural 3D Reconstruction for Reflective Objects with Sparse Polarized Images. In *CVPR*, pages 11821–11830, 2024. 2
- [19] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd ed edition, 2003. 1, 2
- [20] A. Hertzmann and S.M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE TPAMI*, 27(8):1254–1264, 2005. 1, 2
- [21] Inseung Hwang, Daniel S. Jeon, Adolfo Muñoz, Diego Gutierrez, Xin Tong, and Min H. Kim. Sparse ellipsometry: Portable acquisition of polarimetric SVBRDF and shape with unstructured flash photography. *ACM TOG*, 41(4):1–14, 2022. 2
- [22] Tomoki Ichikawa, Matthew Purri, Ryo Kawahara, Shohei Nobuhara, Kristin Dana, and Ko Nishino. Shape from Sky: Polarimetric Normal Recovery Under The Sky. In *CVPR*, pages 14827–14836, 2021. 3
- [23] Masaki Kaga, Takahiro Kushida, Tsuyoshi Takatani, Kenichiro Tanaka, Takuya Funatomi, and Yasuhiro Mukaigawa. Thermal non-line-of-sight imaging from specular and diffuse reflections. *IPSP Transactions on Computer Vision and Applications*, 11(1):8, 2019. 2
- [24] Takahiro Kushida, Ryutaro Nakamura, Hiroaki Matsuda, Wenhao Chen, and Kenichiro Tanaka. Affine transform representation for reducing calibration cost on absorption-based LWIR depth sensing. *Scientific Reports*, 14(1):26429, 2024. 1, 2
- [25] Chenyang Lei, Chenyang Qi, Jiaxin Xie, Na Fan, Vladlen Koltun, and Qifeng Chen. Shape from Polarization for Complex Scenes in the Wild. In *CVPR*, pages 12622–12631, 2022. 5
- [26] Daniel A. LeMaster and Michael T. Eismann. Passive Polarimetric Imaging. In *Multi-Dimensional Imaging*, pages 391–428. Wiley, 1 edition, 2014. 4
- [27] Chenhao Li, Taishi Ono, Takeshi Uemori, Hajime Mihara, Alexander Gatto, Hajime Nagahara, and Yusuke Moriuchi. NeISF: Neural Incident Stokes Field for Geometry and Material Estimation. In *CVPR*, pages 21434–21445, 2024. 2
- [28] Yvette Y. Lin, Xin-Yi Pan, Sara Fridovich-Keil, and Gordon Wetzstein. ThermalNeRF: Thermal Radiance Fields. In *ICCP*, pages 1–12, 2024. 2
- [29] Tomohiro Maeda, Yiqin Wang, Ramesh Raskar, and Achuta Kadambi. Thermal Non-Line-of-Sight Imaging. In *ICCP*, pages 1–11, 2019. 2
- [30] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf:

- Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2
- [31] Seiichi Mita, Xu Yuquan, Kazuhisa Ishimaru, and Sakiko Nishino. Robust 3D Perception for any Environment and any Weather Condition using Thermal Stereo. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 2569–2574, 2019. 2
- [32] Miyazaki, Tan, Hara, and Ikeuchi. Polarization-based inverse rendering from a single view. In *ICCV*, pages 982–987 vol.2, 2003. 1, 2
- [33] Daisuke Miyazaki, Megumi Saito, Yoichi Sato, and Katsushi Ikeuchi. Determining surface orientations of transparent objects based on polarization degrees in visible and infrared wavelengths. *J. Opt. Soc. Am. A*, 19(4):687, 2002. 4
- [34] D. Miyazaki, M. Kagesawa, and K. Ikeuchi. Transparent surface modeling from a pair of polarization images. *IEEE TPAMI*, 26(1):73–82, 2004. 1, 2
- [35] Yasuto Nagase, Takahiro Kushida, Kenichiro Tanaka, Takuya Funatomi, and Yasuhiro Mukaigawa. Shape from Thermal Radiation: Passive Ranging Using Multi-spectral LWIR Measurements. In *CVPR*, pages 12651–12661, 2022. 1, 2
- [36] Sriram Narayanan, Mani Ramanagopal, Mark Sheinin, Aswin C. Sankaranarayanan, and Srinivasa G. Narasimhan. Shape from heat conduction. In *ECCV*, pages 426–444, 2025. 1, 2
- [37] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In *CVPR*, pages 165–174, 2019. 5
- [38] Mani Ramanagopal, Sriram Narayanan, Aswin C. Sankaranarayanan, and Srinivasa G. Narasimhan. A Theory of Joint Light and Heat Transport for Lambertian Scenes. In *CVPR*, pages 11924–11933, 2024. 2
- [39] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. 7
- [40] Joaquin Rodriguez, Lew-Fock-Chong Lew-Yan-Voon, Renato Martins, and Olivier Morel. Pola4All: Survey of polarimetric applications and an open-source toolkit to analyze polarization. *Journal of Electronic Imaging*, 33(01), 2024. 1
- [41] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep Photometric Stereo Networks for Determining Surface Normal and Reflectances. *IEEE TPAMI*, 44(1):114–128, 2022. 1, 2
- [42] Philip Saponaro, Scott Sorensen, Abhishek Kolagunda, and Chandra Kambhamettu. Material classification with thermal imagery. In *CVPR*, pages 4649–4656, 2015. 2
- [43] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR*, pages I–195–I–202, 2003. 1, 2
- [44] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-Motion Revisited. In *CVPR*, pages 4104–4113, 2016. 1, 2
- [45] Johannes L. Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise View Selection for Unstructured Multi-View Stereo. In *ECCV*, pages 501–518, 2016. 1, 2
- [46] Boxin Shi, Jinfa Yang, Jinwei Chen, Ruihua Zhang, and Rui Chen. Recent Progress in Shape from Polarization. In *Advances in Photometric 3D-Reconstruction*, pages 177–203. Springer International Publishing, 2020. 1, 2, 5
- [47] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer International Publishing, 2022. 1, 2
- [48] Kenichiro Tanaka, Nobuhiro Ikeya, Tsuyoshi Takatani, Hiroyuki Kubo, Takuya Funatomi, Vijay Ravi, Achuta Kadambi, and Yasuhiro Mukaigawa. Time-Resolved Far Infrared Light Transport Decomposition for Thermal Photometric Stereo. *IEEE TPAMI*, 43(6):2075–2085, 2021. 1, 2, 8
- [49] L.B. Wolff and T.E. Boulton. Constraining object features using a polarization reflectance model. *IEEE TPAMI*, 13(7): 635–657, 1991. 1, 2
- [50] Robert J. Woodham. Photometric Method For Determining Surface Orientation From Multiple Images. *Optical Engineering*, 19(1), 1980. 1, 2
- [51] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *NeurIPS*, 33, 2020. 2, 5, 6, 7
- [52] Tianxiang Ye, Qi Wu, Junyuan Deng, Guoqing Liu, Liu Liu, Songpengcheng Xia, Liang Pang, Wenxian Yu, and Ling Pei. Thermal-NeRF: Neural Radiance Fields from an Infrared Camera. In *IROS*, pages 1046–1053, 2024. 2
- [53] Jincheng Zhang, Andrew R Willis, and Kevin Brink. Photometric Correction for Infrared Sensors. In *CVPRW*, pages 431–439, 2023. 1, 2
- [54] Z. Zhang. A flexible new technique for camera calibration. *IEEE TPAMI*, 22(11):1330–1334, 2000. 6
- [55] Jinyu Zhao, Yusuke Monno, and Masatoshi Okutomi. Polarimetric Multi-View Inverse Rendering. *IEEE TPAMI*, 45(7):8798–8812, 2023. 2