# Shape from Thermal Radiation: Passive Ranging Using Multi-spectral LWIR Measurements

Yasuto Nagase[1*] Takahiro Kushida[1*] Kenichiro Tanaka[2]
Takuya Funatomi[1] Yasuhiro Mukaigawa[1]
[1]Nara Institute of Science and Technology (NAIST), Japan [2]Ritsumeikan University, Japan
{nagase.yasuto.nx0, kushida.takahiro.kh3, funatomi, mukaigawa}@is.naist.jp
ken-t@fc.ritsumei.ac.jp

## Abstract

*In this paper, we propose a new cue of depth sensing using thermal radiation. Our method realizes passive, texture independent, far range, and dark scene applicability, which can broaden the depth sensing subjects. A key observation is that thermal radiation is attenuated by the air and is wavelength dependent. By modeling the wavelength-dependent attenuation by the air and building a multi-spectral LWIR measurement system, we can jointly estimate the depth, temperature, and emissivity of the target. We analytically show the capability of the thermal radiation cue and show the effectiveness of the method in real-world scenes using an imaging system with a few bandpass filters.*

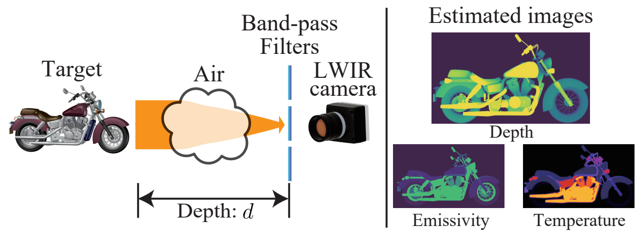



Figure 1. Overview of the shape from thermal radiation. Our method jointly estimates the depth, temperature, and emissivity images from multi-spectral LWIR images. A key observation is that thermal radiation from the target object is attenuated through the air and is wavelength dependent.

## 1. Introduction

Depth sensing is an important technology, as shown by its wide range of applications. Depth sensing techniques refer to physics-based cues used to recover the scene depth, known as "shape from X." Historically, a great number of cues such as triangulation/disparity [23], time-of-flight [21], polarization [46], shading [25], silhouettes [9], focus/defocus [44], and texture [60] have been proposed. These "X" cues play an important role in depth sensing in various scenarios. However, existing depth sensing techniques require either clear visibility of the texture or an active light source to illuminate the surface, and there is no approach to passively measure depth in dark environments.

In this paper, we propose an unprecedented depth sensing modality that can be used in dark and passive scenarios. As a key to achieving this property, we use the attenuation of *long wavelength infrared* (LWIR) radiation through the air. As all objects emit thermal radiation according to their temperature, we can see the objects without any active light sources. Besides, since the attenuation of LWIR is much larger than that of visible light and the amount of attenuation decays exponentially with respect to the depth, it is possible to recover the depth.

Because the observed intensity depends on both the object's temperature and the attenuation, we have to separate these factors. Fortunately, attenuation through the air is wavelength dependent. Thus, we use multi-spectral observation in the LWIR range to jointly estimate the depth and temperature of the target object as shown in Fig. 1.

This paper focuses on the possibility of depth sensing by only using passive thermal radiation. The contributions of this study are twofold.

- We propose a novel cue for "shape from X" techniques. The shape from thermal radiation uses the attenuation of thermal radiation by the air. This is new to computer vision and computational imaging areas and broadens the field of research.
- To the best of our knowledge, the proposed method is the first attempt to realize passive, texture-less, far range, and can be used for dark scenes. This property proves that air absorption cues in LWIR are useful for depth sensing.

*Equal contribution.

## 2. Related Work

**Three-dimensional imaging** Three-dimensional (3D) imaging has been studied over the decades and remains an active research topic in the computer vision area. Here, we briefly review the existing shape from X approaches.

Triangulation-based approaches use the disparity between two viewing/lighting positions. They include stereo cameras [6, 45, 52, 54, 65], multi-view stereo [17, 23, 59], structure from motion [57], and structured light [11, 19, 37, 49, 67]. Stereo reconstruction using thermal cameras is also proposed [38, 56]. The accuracy of triangulation depends on the baseline; hence, it is not suitable for far-range imaging.

Time-of-flight [21] is also a major 3D imaging technique and known as the principle of Lidar and the time-of-flight camera. They use a pulse or periodically modulated light and measure the round-trip time of light. It is effective for texture-less targets; however, either strong active illumination or a point scanning mechanism is required.

Monocular depth estimation methods that use deep learning [14, 35, 66] estimate the scene depth from a single RGB image. This approach is up-to-scale and easy to deceive using, for example, a printed photo placed in front of the camera. Recently, to mitigate this problem, physics-based assistance for monocular depth estimation is proposed [7, 10, 20, 42, 53, 62]. Another extension of learning-based depth estimation is to use thermal imaging to apply dark and night environments [29, 33], which are pure learning-based.

Depth from focus/defocus [24, 44, 63] is another well-known approach of monocular depth estimation, which recovers the depth from the amount of blur. These methods require textures on the target to measure the focus and are not applied for dark scenes.

Photometric information is a rich cue for 3D imaging. Several methods, for example, shape-from-shading [64], photometric stereo [1, 50, 61], shape from polarization [5, 40, 46], shape from water absorption [2, 3, 30], light fall-off stereo [32], and attenuation by fog or haze [31, 43] have been proposed. Although these methods can recover a high-quality shape, an active or pre-calibrated light source is required.

Table 1 shows a comparison of major depth sensing approaches. Our method first uses multi-spectral LWIR observations to estimate scene depths based on air absorption of thermal radiation. Our approach is fundamentally passive, texture independent, and very far range depth sensing. One of the advantages of the approach is that it is fully passive and applicable to a dark scene.

**Computational thermal imaging** The unique property of thermal radiation is very useful for computer vision and computational imaging problems. For example, the polarization cue of thermal radiation is used to resolve the ambiguity of shape from polarization [39]. The transient feature of the thermal heating and cooling cycle is also used for material classification [51]. LWIR observation is useful for computational imaging problems, such as non-line-of-sight imaging [28, 34] and lensless imaging [47]. Our method falls into this computational thermal imaging category as a new 3D imaging.

| Method | Dark scene | Far range | Texture independence |
|---|---|---|---|
| Time-of-flight | Light source | No | Yes |
| Lidar | Light source | Yes | Yes |
| Structured light | Light source | Baseline | Yes |
| Multi-camera | No | Baseline | No |
| Focus/Defocus | No | No | No |
| **Ours** | Yes | Yes | Yes |

Table 1. Comparison of depth sensing approaches. Fundamental features are compared. Our method is adaptable to a dark scene, capable of sensing a far-range target, and texture independent.

The properties of thermal radiation have also been used for 3D sensing problems. Tanaka *et al.* [55] illuminated an object from various positions and recovered the surface normal of various materials using the shape-from-shading approach. Erdozain *et al.* [15] developed an LWIR projector and applied structured light algorithms. The shape of transparent and metallic objects can be measured using triangulation by scanning the heating point [4, 8, 16] because the objects are opaque in the LWIR region. Although these approaches are effective in active illumination scenarios, our method is the first attempt at passive LWIR 3D imaging.

**Thermography** The study of thermography has focused on the accurate measurement of temperature [12, 58]. Because thermal radiation through the air is attenuated according to its distance from the target object, the temperature of a distant object must be corrected by atmosphere transmittance databases [13, 18], the application of a simplified model [36], or displaying a reference target with known temperature or known size as a guide. We contrarily use the air absorption of LWIR for depth estimation. Although our goal is depth estimation, the temperature of the object can be also estimated at the same time. This can be an advantage over conventional thermography because the correct temperature is measured even when the target depth is unknown.

Spectral information in the LWIR region is useful for some applications. While thermography requires the user to set the emissivity of the object or assumes uniform emissivity, two-color or ratio thermography recovers both emissivity and temperature using two bands [27] in LWIR. Spectral information in the LWIR region is also used for gas detection [48]. Our method uses multi-bands in LWIR for depth sensing because air absorption of LWIR depends on the wavelength.

## 3. LWIR Light Transport

We start from a brief review of thermal radiation theory and temperature measurement. After that, we build an imaging model using the attenuation by the air.

**Thermal radiation theory [26]** Planck's law explains the spectral radiant exitance (emittance) $M_e$ of a black body at the absolute temperature $T$ as

$$M_e(\lambda; T) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}, \quad (1)$$

where $\lambda$ is the wavelength of light, $c$ is the speed of light, $h$ is the Planck constant[1], and $k$ is the Boltzmann constant[2].

As most objects are not black bodies, their radiation $E$ is less than that of a black body:

$$E(\lambda; \epsilon, T) = \epsilon M_e(\lambda; T), \quad (2)$$

where $\epsilon$ is the emissivity, which is object dependent.

**Thermography [58]** A thermographic camera is designed to measure the intensity of thermal radiation. A typical sensor measures the intensity of LWIR (typically, 8–14 µm) as

$$E(\epsilon, T) = \int_{\lambda_1}^{\lambda_2} E(\lambda; \epsilon, T) d\lambda, \quad (3)$$

where $E$ is the intensity of the radiated LWIR light. As almost all radiation energy is within this integration range, the temperature $T$ can be obtained using the Stefan-Boltzmann law, in theory. In practice, the correspondence between the intensity $E$ and temperature $T$ is calibrated using a black body in the factory. Thermal cameras assume that the emissivity is known: they either use a constant value or have a dialog that can be used to select the target's material to adjust the emissivity.

### 3.1. Imaging model

The air is an attenuating medium and some LWIR light is absorbed through the air as it travels from the target to the camera [58]. Using the Lambert-Beer law, the transmission of the medium is expressed as

$$i_{\text{out}}(\lambda) = \exp(-\sigma_{\text{air}}(\lambda)d) i_{\text{in}}(\lambda), \quad (4)$$

where $i_{\text{in}}$ and $i_{\text{out}}$ are the input and transmitted intensity of light, respectively, $\sigma_{\text{air}}$ is the extinction coefficient of the medium, and $d$ is the thickness of the medium. Thermal cameras either cancel this attenuation by allowing the user to manually input the depth of interest or simply ignore it. We contrarily take this into account to estimate the depth of the target.

[1]Planck constant: $h = 6.62607015 \times 10^{-34}$Js

[2]Boltzmann constant: $k = 1.380649 \times 10^{-23}$JK$^{-1}$

Based on Eqs. (1), (2), and (4), the observed intensity $I$ at wavelength $\lambda$ is expressed as

$$I(\lambda) = R_v(\lambda) \exp(-\sigma_{\text{air}}(\lambda)d) \epsilon M_e(\lambda; T), \quad (5)$$

where $R_v(\lambda)$ is the sensitivity that associates the energy density with the observed intensity. The sensitivity $R_v(\lambda)$ depends on the imaging system (e.g., image sensor, lens, and spectral filter) and is known by calibration beforehand. We aim to recover the depth $d$ with additional unknowns, emissivity $\epsilon$ and temperature $T$, from the observation $I(\lambda)$.

## 4. Depth from Multi-spectral LWIR Measurements

Equation (5) has three unknown variables: $T$, $\epsilon$, and $d$. These variables cannot be obtained using only a single observation. This means that we cannot identify whether the target object is far away or the radiation from the target is low.

We address this problem using multi-spectral LWIR observations to estimate the variables, as shown in Fig. 1. A key idea is to use the wavelength dependence of the extinction coefficient of the air. Because spectral radiance follows Planck's law and the attenuation by the air is wavelength dependent, we can separate the effect of the depth and the amount of radiation from the target.

### 4.1. Assumptions

Before explaining our proposed method, we clarify the assumptions of the scene. These assumptions are reasonable as they are also assumed in thermography and the coverage of the material is very wide.

- As the major attenuating medium is water and carbon dioxide [36], the attenuating spectrum of the air $\sigma_{\text{air}}(\lambda)$ is known. It can be measured in advance, as well as being obtained from databases of extinction coefficients of the air [13, 18].
- The target object is a gray body, that is, the emissivity of the target is not wavelength-dependent in LWIR. Only a few materials, such as metal, do not satisfy this assumption.
- The target object does not reflect LWIR light from any other heat source.

### 4.2. Proposed method

Suppose that we obtain two measurements using different wavelengths: $I(\lambda_i)$ and $I(\lambda_j)$. By dividing one measurement by another, we cancel out the emissivity $\epsilon$:

$$\begin{aligned} \hat{I}_{i,j} &= \frac{I(\lambda_i)}{I(\lambda_j)} \\ &= \frac{R_v(\lambda_i) \exp(-\sigma_{\text{air}}(\lambda_i)d) M_e(\lambda_i; T)}{R_v(\lambda_j) \exp(-\sigma_{\text{air}}(\lambda_j)d) M_e(\lambda_j; T)}, \end{aligned} \quad (6)$$

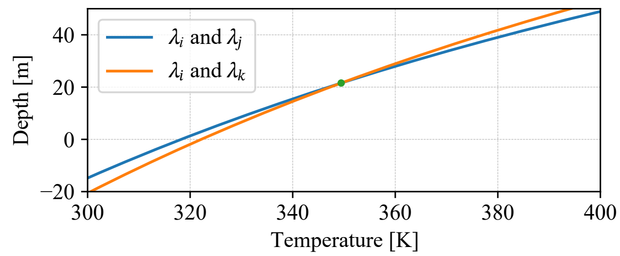



Figure 2. A plot of the solution space. The two lines represent the different combinations of filters. The intersection is the estimate of the temperature and depth.

where $\hat{I}_{i,j}$ is the divided ratio image. By transforming the expression, we can represent the depth $d$ as

$$d_{i,j}(T) = \frac{1}{\sigma_{\text{air}}(\lambda_j) - \sigma_{\text{air}}(\lambda_i)} \ln\left(\frac{R_v(\lambda_j)M_e(\lambda_j;T)}{R_v(\lambda_i)M_e(\lambda_i;T)}\hat{I}_{i,j}\right). \tag{7}$$

This equation describes a curved line as the solution space of $T$ and $d$ as shown in Fig. 2, because $\sigma_{\text{air}}$ and $R_v$ are known and $M_e$ is calculable using Eq. (1). As we have multiple wavelength observations with sufficiently different amounts of air absorption, for example, $\lambda_i$, $\lambda_j$, and $\lambda_k$, Eq. (7) also represents another set of wavelengths. For example, we have two solution spaces of depth $d$ and temperature $T$ that use three wavelength combinations, and the crossing point of these curves is a unique solution. We can find the temperature $T$ and depth $d$ that satisfies the following equation:

$$d_{i,j}(T) = d_{i,k}(T). \tag{8}$$

Optionally, we can estimate the emissivity $\epsilon$ by substituting the estimated depth $d$ and temperature $T$ into Eq. (5).

**Effect of broad bandpass filters** In the above discussion, we implicitly assume that the measurement is implemented using an ideal filter, whose response function is a delta function. If a bandpass filter of broader bandwidth is used, we have to integrate the imaging model with respect to the wavelength; therefore, Eq. (5) becomes

$$I_s(\lambda_i) = \int_0^\infty f_i(\lambda)R_v(\lambda)\exp(-\sigma_{\text{air}}(\lambda)d)\epsilon M_e(\lambda;T)d\lambda, \tag{9}$$

where $f_i$ is the spectral response of the bandpass filter.

Although this model is no longer the Lambert-Beer model, the observation is still monotonically decreasing with respect to the depth, and the magnitude of attenuation varies depending on the filter. Therefore, it is possible to estimate the depth numerically. In our environment, we found that the observation using a filter of a sufficiently narrow bandwidth can be approximated using Eq. (5); hence, we do not use the extended imaging model.

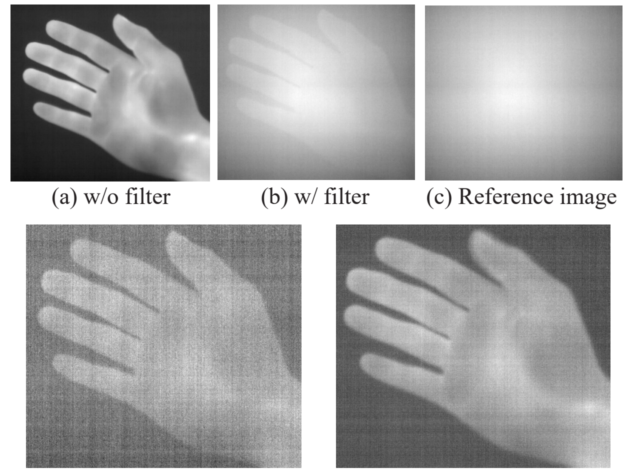



Figure 3. Example of the unique noise in LWIR images. (a) Without filer: a clear image can be obtained. (b) With filter: as the radiation from the camera is reflected by the optics, the image becomes unclear (narcissus effect). (c) Reference image of the shutter plane in front of the optics. (d) Subtraction of (c) from (b): image remains noisy because of $1/f$ noise. (e) Result using the external shutter: unwanted effects are canceled and a clear image is obtained.

**Optimization** As the method above is naïve and sensitive to noise, the estimated values can be physically incorrect values, such as a negative depth. Therefore, we adopt an optimization technique to regulate scene parameters. We formulate the optimization problem as

$$\hat{d}, \hat{\epsilon}, \hat{T} = \operatorname*{argmin}_{d,\epsilon,T} \sum_i^N \|I(\lambda_i) - I'(\lambda_i : d, \epsilon, T)\|_2^2 \tag{10}$$
$$\text{s.t.}\ \ 0 \le \epsilon \le 1, d \ge 0, T \ge 0,$$

where $N$ is the number of observations and $I'$ is a rendered image using Eq. (5). We initialize all parameters using the naïve result obtained using Eq. (8) and estimated using constrained gradient descent.

## 5. Unique Noise Characteristics of the LWIR Camera

As an LWIR camera is very different from an RGB camera, several issues exist that do not appear in traditional imaging. In this section, we introduce two unique effects in LWIR imaging and explain how to deal with them.

### 5.1. Narcissus effect

In LWIR observation, radiation from the camera body is reflected by the surface of the optics, including lens and filters. This is known as the narcissus effect [58]. Figure 3(b) shows an example of the narcissus effect. This effect is generally canceled using the non-uniformity correction (NUC) [58] function in the camera. NUC is calibrated to cancel the camera's radiation in the factory. Because it is calibrated for the built-in optics, it does not work well if another optic is added to the system. Therefore, we have to perform this correction for each filter explicitly.

To cancel the narcissus effect, we require an image with a black body plane with the known temperature immediately in front of the optics. By subtracting two images (*i.e.* scene and reference) captured through a filter, we can obtain only the filter transmitted component. We can obtain pure transmitted component $I(\lambda_i)$ by subtracting the scene and reference images:

$$I(\lambda_i) = I_s(\lambda_i) - I_r(\lambda_i), \tag{11}$$

where $I_s$ and $I_r$ are the observed images with the narcissus effect of the scene and reference, respectively. Fig. 3(d) shows an example of canceling the narcissus effect by simply subtracting Fig. 3(c) from Fig. 3(b).

### 5.2. $1/f$ noise

Another unique noise characteristic of an LWIR camera is $1/f$ noise. In particular, this noise dominates the total noise of the microbolometer sensor [22, 41]. $1/f$ noise is a low-frequency fluctuation; hence, it strongly appears when images taken at long intervals are processed. For example, averaging multiple images does not improve the signal-to-noise ratio (SNR) but degrades the image, which is completely different from RGB imaging. Considering the processing of NUC, this noise is very noticeable, as shown in Fig. 3(d), because there is a very long interval between the two images.

### 5.3. Noise mitigation using an external shutter

To mitigate both the narcissus effect and $1/f$ noise, we propose to use an external shutter in front of the optics. The shutter is operated open and closed, alternately, and we obtain a total of $M$ pairs of a scene and reference images. Then, we suppress the narcissus effect by taking a subtraction of each pair. Note that the 1/f noise is ignorable each of them is captured in a short time. Finally, we suppress the $1/f$ noise by taking the average of them. This corresponds to locking-in high frequency signals. Fig. 3(e) shows the result of noise reduction using $M = 10$ pairs.

## 6. Depth Resolution Analysis

Before conducting the experiment, we confirmed the feasibility of our method in terms of depth resolution. The depth resolution depends on the sensitivity of the sensor, in addition to the temperature and depth of the object. To analyze the depth resolution, we first explain a standard sensor noise measure for LWIR cameras known as the *noise equivalent temperature difference* (NETD). NETD represents the minimum temperature that changes the intensity at the same level of system noise and is typically provided as a camera specification.

**Definition of NETD [58]** We consider the area of the detector $A_d$, the f-number of the imaging lens $F$, and the spectral sensitivity of the sensor $R_v(\lambda)$. The power of LWIR that irradiates to the detector is represented as

$$P(\lambda) = \frac{A_d}{4F^2} M_e(\lambda; T). \tag{12}$$

Assuming that the LWIR camera captures the intensity from $\lambda_1$ to $\lambda_2$, the sensor output voltage is represented as

$$V_s = \int_{\lambda_1}^{\lambda_2} R_v(\lambda) P(\lambda) d\lambda. \tag{13}$$

The change in $V_s$ when the temperature changes by $\Delta T$ is represented as

$$\Delta V_s = \frac{\partial V_s}{\partial T} \Delta T. \tag{14}$$

As the microbolometer sensor has almost flat spectral sensitivity, Eq. (14) is typically simplified as

$$\Delta V_s = \frac{A_d R_v}{4F^2} \frac{\partial \int_{\lambda_1}^{\lambda_2} M_e(\lambda; T) d\lambda}{\partial T} \Delta T. \tag{15}$$

The change in the output voltage when the temperature changes by $1\,\mathrm{K}$ is expressed as

$$V_t = \frac{\Delta V_s}{\Delta T}. \tag{16}$$

The temperature change when the voltages of the output and noise are balanced (*i.e.*, NETD) is defined as

$$\mathrm{NETD} \triangleq \frac{V_n}{V_t} \tag{17}$$

$$= \frac{4F^2 V_n}{A_d R_v} \frac{1}{\frac{\partial \int_{\lambda_1}^{\lambda_2} M_e(\lambda; T) d\lambda}{\partial T}}, \tag{18}$$

where $V_n$ is the root mean square noise voltage. The first factor represents the sensor characteristics and the second factor is numerically obtainable.

**Definition of *noise equivalent depth difference* (NEDD)** We extend the discussion to the depth resolution of our method. Considering the air absorption into the model, the power of LWIR that reaches the camera is rewritten as

$$P(\lambda) = \frac{A_d}{4F^2} e^{-\sigma_{\mathrm{air}} d} M_e(\lambda; T). \tag{19}$$

Equation (13) also holds for this model. We define the NEDD analogously to NETD. The change in $V_s$ when the depth of the target changes by $\Delta d$ is represented as

$$\Delta V_s = \frac{\partial V_s}{\partial d}\Delta d. \tag{20}$$

NEDD is defined as

$$\text{NEDD} \triangleq \frac{V_n}{V_d}, \tag{21}$$

where $V_d$ is the change of the output voltage caused by $1\,\mathrm{m}$ depth change defined as

$$V_d = \frac{\Delta V_s}{\Delta d}. \tag{22}$$

Based on the flat sensitivity of the microbolometer sensor, Eq. (21) is expanded as

$$\text{NEDD} = \frac{4F^2V_n}{A_dR_v}\frac{1}{\frac{\partial \int_{\lambda_1}^{\lambda_2} e^{-\sigma_{\text{air}}d} M_e(\lambda;T)d\lambda}{\partial d}}. \tag{23}$$

The second factor of this equation can be also numerically obtained. Substituting Eq. (18) into Eq. (23), the NEDD can be expressed using the NETD measure as

$$\text{NEDD} = \frac{\frac{\partial \int_{\lambda_1}^{\lambda_2} M_e(\lambda;T)d\lambda}{\partial T}}{\frac{\partial \int_{\lambda_1}^{\lambda_2} e^{-\sigma_{\text{air}}d} M_e(\lambda;T)d\lambda}{\partial d}}\text{NETD}. \tag{24}$$

**NEDD using bandpass filters** The SNR becomes worse when bandpass filters are placed in front of the camera as they block a large amount of energy. The voltage of the sensor output with a bandpass filter is represented as

$$V_f = \int_{\lambda_1}^{\lambda_2} f(\lambda)R_v(\lambda)P(\lambda)d\lambda, \tag{25}$$

where $f(\lambda)$ is the spectral transmittance of the bandpass filter. Following the NEDD definition, the NEDD using the filter is expressed as

$$\text{NEDD}_f = \frac{V_n}{V_d'} \tag{26}$$

$$= \frac{4F^2V_n}{A_dR_v}\frac{1}{\frac{\partial \int_{\lambda_1}^{\lambda_2} e^{-\sigma_{\text{air}}d} M_e(\lambda;T)f(\lambda)d\lambda}{\partial d}}, \tag{27}$$

where

$$V_d' = \frac{\Delta V_f}{\Delta d}. \tag{28}$$

Similar to the above specific example, the NEDD of any bandpass filter observations can be numerically calculated.

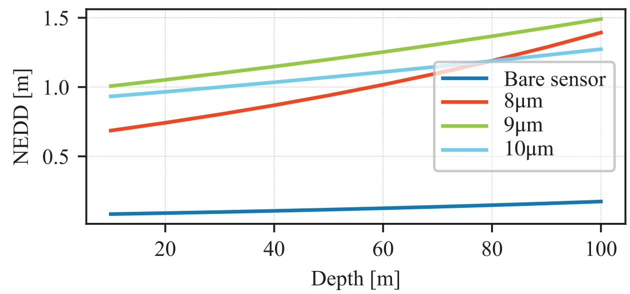



Figure 4. A theoretical plot of the observable depth difference of our measuring system. Each observation has an NEDD of less than approximately $1.5\,\mathrm{m}$, which determines the total depth resolution of the method.

**Specific example** A typical LWIR sensor records the intensity from 8 μm to 14 μm. The value of the numerator of the first factor in Eq. (24) at 300 K in this bandwidth is

$$\frac{\partial \int_{8\mu\text{m}}^{14\mu\text{m}} M_e(\lambda;300\text{K})d\lambda}{\partial T} \approx 2.123. \tag{29}$$

Assuming the target depth is at 20m and $\sigma_{\text{air}} = 0.008\text{m}^{-1}$ the denominator is,

$$\frac{\partial \int_{8\mu\text{m}}^{14\mu\text{m}} e^{-\sigma_{\text{air}}d} M_e(\lambda;300\text{K})d\lambda}{\partial d} \approx 0.939. \tag{30}$$

If the NETD of the camera is $40\,\mathrm{mK}$ ($= 0.04\,\mathrm{K}$), the NEDD of a bare LWIR sensor is calculated by Eq. (24) as,

$$\text{NEDD} = \frac{2.123}{0.939}\text{NETD} = 0.090\,\mathrm{m}. \tag{31}$$

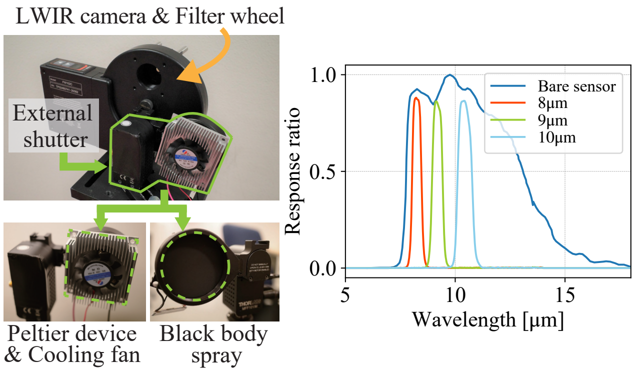
For a bandpass filter of the center wavelength (CWL) $10.400\,\mu\mathrm{m}$ and full width at half maximum (FWHM) $737\,\mathrm{nm}$,

$$\frac{\partial \int_{\lambda_1}^{\lambda_2} e^{-\sigma_{\text{air}}d} M_e(\lambda;300\text{K})f(\lambda)d\lambda}{\partial d} \approx 0.088. \tag{32}$$

The NEDD with the bandpass filter is calculated as $\text{NEDD}_{10\mu\text{m}} = \frac{2.123}{0.088}0.04 = 0.965$ m. This demonstrates that it is feasible to use the NEDD for large-scale scenes. Figure 4 shows the plot of NEDD of our system with the target temperature at $300\,\mathrm{K}$. The feasibility of our method is confirmed by this analysis.

## 7. Real-world Experiments

We build a multi-spectral LWIR imaging system and evaluate the effectiveness of our proposed method in real-world experiments.



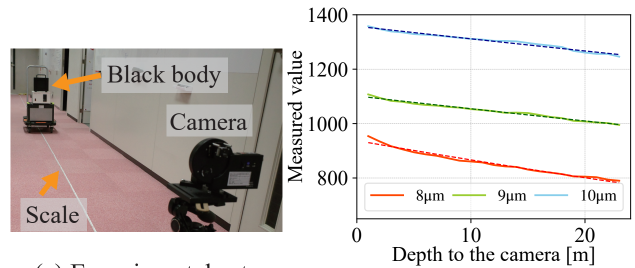

(a) Imaging system (b) Spectral response

Figure 5. Imaging system and profiles of each filter. (a) The multi-spectral imaging system consists of an LWIR camera, filter wheel with three bandpass filters, and external shutter. The external shutter is painted with black body spray and kept at low temperature using a Peltier device. (b) The response ratio of the bare LWIR sensor with the lens and transmittance of each filter.

## 7.1. Imaging system and calibration

**Multi-spectral LWIR imaging system** Our multi-spectral LWIR imaging system is shown in Fig. 5(a). The system consists of a thermal imaging camera (FLIR Boson 640, NETD $= 40\,\mathrm{mK}$) and a filter wheel with three narrow bandpass filters. We select three filters with the best NEDD values in our laboratory ($8\,\mu\mathrm{m}$ (CWL $8.248\,\mu\mathrm{m}$, FWHM $452\,\mathrm{nm}$), $9\,\mu\mathrm{m}$ (CWL $9.127\,\mu\mathrm{m}$, FWHM $545\,\mathrm{nm}$), and $10\,\mu\mathrm{m}$ (CWL $10.400\,\mu\mathrm{m}$, FWHM $737\,\mathrm{nm}$)). The spectral response ratio of each filter is shown in Fig. 5(b). These filters are automatically changed using the filter wheel. An external shutter in front of the camera removes the narcissus effect and reduces $1/f$ noise. The shutter is sprayed with black body paint. To obtain a strong signal, the shutter is cooled using a Peltier device.

**Calibration** In the calibration step, we obtain the sensitivity $R_v$ and extinction coefficients $\sigma_{\mathrm{air}}$ of all wavelengths. We use a black body furnace, whose surface temperature is controllable. Fig. 6(a) shows the experimental setup. By placing the black body at $0\,\mathrm{m}$, $R_v$ can be obtained directly. By placing the black body at several known depths, and fitting the Lambert-Beer law to the measured value, we can obtain the extinction coefficients at each wavelength. We set the temperature of the black body at $80\,°\mathrm{C}$. The fitting result is illustrated in Fig. 6(b) and the calculated extinction coefficients are shown in Table 2. The fitted result shows that the extinction coefficients are different at each wavelength.




(a) Experimental set up (b) Attenuation

Figure 6. Experiment for measuring the extinction coefficients. (a) Experimental setup. (b) Measured values attenuated by the air. Dashed lines represent the fit curve of the Lambert-Beer law for each filter.

Table 2. Fitting results for the extinction coefficients.

| wavelength | $8\,\mu\mathrm{m}$ | $9\,\mu\mathrm{m}$ | $10\,\mu\mathrm{m}$ |
|---|---|---|---|
| $\sigma_{\mathrm{air}}(\lambda)$ | $0.00787\mathrm{m}^{-1}$ | $0.00436\mathrm{m}^{-1}$ | $0.00346\mathrm{m}^{-1}$ |

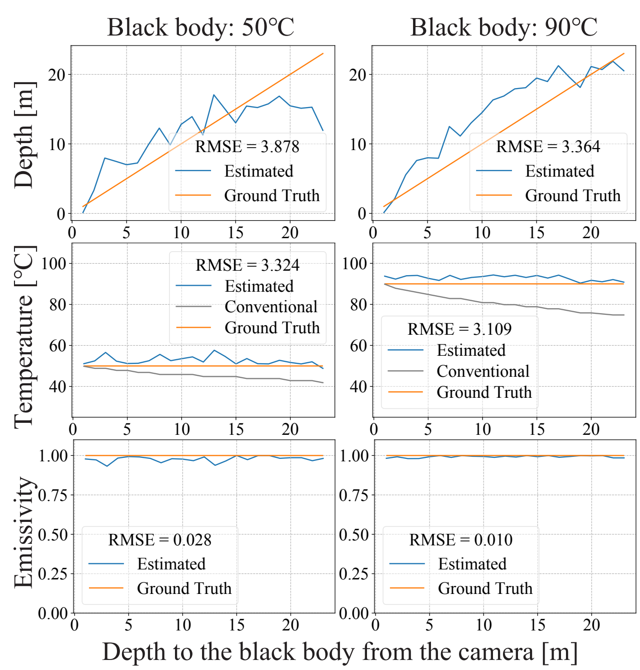



Figure 7. Results for the black body. Temperature is set to $50\,°\mathrm{C}$ (left column) and $90\,°\mathrm{C}$ (right column). Estimated depth, temperature, and emissivity are shown. Root mean squared error (RMSE) is calculated for each result. In the case of $90\,°\mathrm{C}$, in particular, we demonstrate the clear effectiveness of our method.

## 7.2. Results

**Black body target** We first evaluate the method using a black body target, whose emissivity is almost 1. we set the temperature of the black body to $50\,°\mathrm{C}$ and $90\,°\mathrm{C}$, and placed it at several distances from the camera. Figure 7

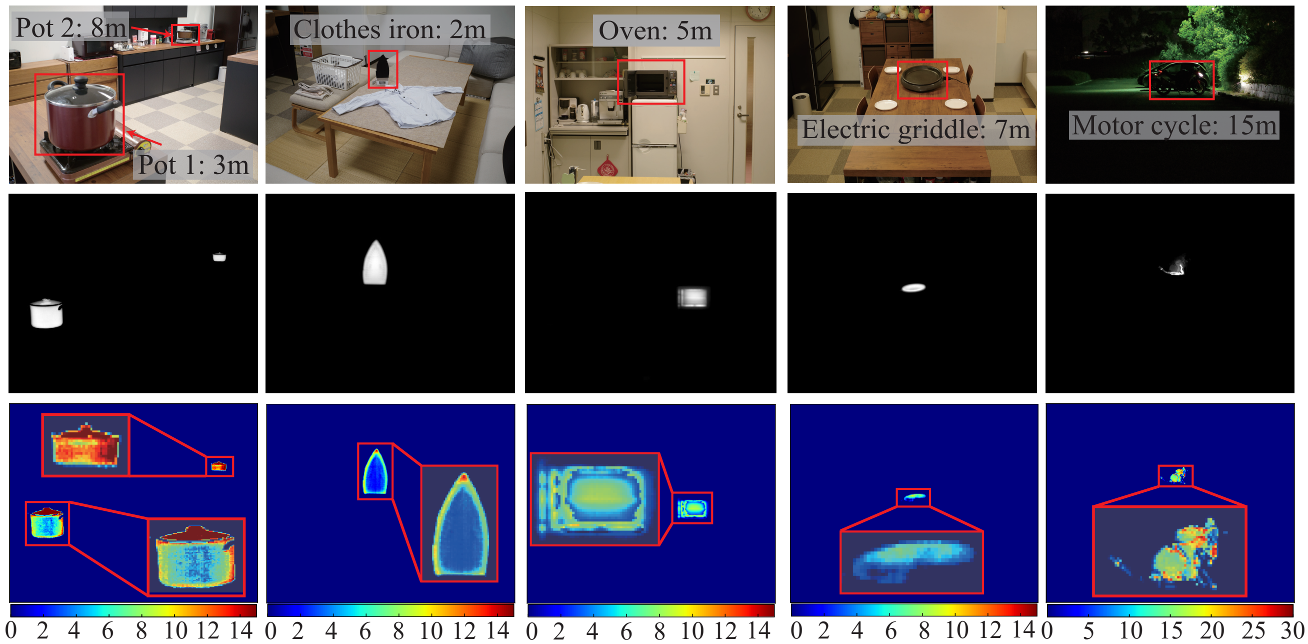



Figure 8. Result for some practical targets. Top: the photo of the target object with its ground truth of the depth. Middle: the thermal image (w/o filter). Bottom: the estimated depth. Red rectangles are the zoom up of the target region.

shows the results for the estimated depth, temperature, and emissivity. Regardless of the target temperature, the target depth is estimated well with respect to the target depth. The temperature and emissivity are estimated at the correct values. The result for $50\,^\circ\mathrm{C}$ is slightly noisier than that for $90\,^\circ\mathrm{C}$ because the SNR depends on the target temperature. This demonstrated that our method works better as the temperature is higher.

Our proposed method also has an advantage in the scenario of long range temperature measurement. When the target is measured at a long distance with conventional thermography, the result becomes lower than its actual temperature due to the attenuation. Compared with conventional thermography, the temperature is flat with respect to the depth as shown in the middle row of Fig. 7.

**Practical scenes** Figure 8 shows the applicability of our method to real-world objects. The target objects are a pot, oven, clothes iron, electric griddle, and motorcycle. The temperatures of the objects are approximately between $80\,^\circ\mathrm{C}$ and $120\,^\circ\mathrm{C}$. In the two-pot scene, the depths of the pots are estimated as $3\,\mathrm{m}$ and $8\,\mathrm{m}$, which shows that our method distinguishes the depth difference. In the motorcycle scene, although it is very difficult to measure the depth using ordinary passive approaches because the scene is very dark, our method estimates the depth of the motorcycle as $15\,\mathrm{m}$. All objects are estimated well considering approximately $0.7\,\mathrm{m}$ NEDD of our system as shown in the analysis.

## 8. Conclusion

We propose a novel and unique approach for fully passive, texture-less, and far-range depth sensing using an LWIR camera. We demonstrate that the attenuation of thermal radiation is a cue for the scene depth. We build a multispectral LWIR measurement system and demonstrated the effectiveness of our proposed method in real-world experiments.

As this is the first attempt at passive LWIR depth sensing, there are many open challenges to applying it at an industrial level. A major problem to be solved is the low SNR of the LWIR measurement, which makes it difficult to measure low-temperature objects. On the other hand, considering the recent advancement of the LWIR sensor, we believe the measurable temperature will reach air temperature in the future.

Another interesting future direction is the combination of our method with learning-based approaches. As multispectral LWIR images fundamentally contain the physics information of the depth, it is possible to build a physics-based learning model to estimate spatially consistent depth images and/or estimate low-temperature objects.

## Acknowledgment

This work is partly supported by JST CREST JPMJCR1764 and SICORP JPMJSC2003.

## References


[1] Jens Ackermann and Michael Goesele. A Survey of Photometric Stereo Techniques. *Foundations and Trends in Computer Graphics and Vision*, 9(3-4):149–254, 2015. 2

[2] Yuta Asano, Yinqiang Zheng, Ko Nishino, and Imari Sato. Shape from Water: Bispectral Light Absorption for Depth Recovery. In *ECCV*, pages 635–649, 2016. 2

[3] Yuta Asano, Yinqiang Zheng, Ko Nishino, and Imari Sato. Depth sensing by near-infrared light absorption in water. *IEEE TPAMI*, 43(8):2611–2622, 2021. 2

[4] Olivier Aubreton, Alban Bajard, Benjamin Verney, and Frederic Truchetet. Infrared system for 3D scanning of metallic surfaces. *Mach. Vision and Appl.*, 24(7):1513–1524, Oct 2013. 2

[5] 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*, volume 12369, pages 554–571, 2020. 2

[6] Seung-Hwan Baek and Felix Heide. Polka Lines: Learning Structured Illumination and Reconstruction for Active Stereo. In *CVPR*, pages 5753–5763, June 2021. 2

[7] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. Single-Shot Hyperspectral-Depth Imaging With Learned Diffractive Optics. In *ICCV*, page 10, 2021. 2

[8] Alban Bajard, Olivier Aubreton, Youssef Bokhabrine, Benjamin Verney, Gonen Eren, Aytul Erçil, and Frederic Truchetet. Three-dimensional scanning of specular and diffuse metallic surfaces using an infrared technique. *Opt. Eng.*, 51(6):063603, Jun 2012. 2

[9] Bruce Baumgart. Geometric modeling for computer vision. aim-249. Technical report, Stanford Artificial Intelligence Laboratory, 1974. 1

[10] Julie Chang and Gordon Wetzstein. Deep Optics for Monocular Depth Estimation and 3D Object Detection. In *ICCV*, pages 10192–10201, Oct. 2019. 2

[11] Wenzheng Chen, Parsa Mirdehghan, Sanja Fidler, and Kiriakos N. Kutulakos. Auto-tuning structured light by optical stochastic gradient descent. In *CVPR*, Jun 2020. 2

[12] Jing Wei Chin, Kwan Long Wong, Tsz Tai Chan, Kristian Suhartono, and Richard H.Y. So. An Infrared Thermography Model Enabling Remote Body Temperature Screening Up to 10 Meters. In *CVPRW*, pages 3870–3876, June 2021. 2

[13] T. Delahaye, R. Armante, N.A. Scott, N. Jacquinet-Husson, A. Chédin, L. Crépeau, C. Crevoisier, V. Douet, A. Perrin, A. Barbe, V. Boudon, A. Campargue, L.H. Coudert, V. Ebert, J.-M. Flaud, R.R. Gamache, D. Jacquemart, A. Jolly, F. Kwabia Tchana, A. Kyuberis, G. Li, O.M. Lyulin, L. Manceron, S. Mikhailenko, N. Moazzen-Ahmadi, H.S.P. Müller, O.V. Naumenko, A. Nikitin, V.I Perevalov, C. Richard, E. Starikova, S.A. Tashkun, Vl.G. Tyuterev, J. Vander Auwera, B. Vispoel, A. Yachmenev, and S. Yurchenko. The 2020 edition of the GEISA spectroscopic database. *Journal of Molecular Spectroscopy*, 380:111510, 2021. 2, 3

[14] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *NeurIPS*, page 2366–2374, 2014. 2

[15] 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, Oct. 2020. 2

[16] 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. *Opt. Express*, 17(14):11457–68, Jul 2009. 2

[17] Yasutaka Furukawa and Carlos Hernández. Multi-View Stereo: A Tutorial. *Foundations and Trends in Computer Graphics and Vision*, 9(1-2):1–148, 2015. 2

[18] I.E. Gordon, L.S. Rothman, R.J. Hargreaves, R. Hashemi, E.V. Karlovets, F.M. Skinner, E.K. Conway, C. Hill, R.V. Kochanov, Y. Tan, P. Wcisło, A.A. Finenko, K. Nelson, P.F. Bernath, M. Birk, V. Boudon, A. Campargue, K.V. Chance, A. Coustenis, B.J. Drouin, J.–M. Flaud, R.R. Gamache, J.T. Hodges, D. Jacquemart, E.J. Mlawer, A.V. Nikitin, V.I. Perevalov, M. Rotger, J. Tennyson, G.C. Toon, H. Tran, V.G. Tyuterev, E.M. Adkins, A. Baker, A. Barbe, E. Canè, A.G. Császár, A. Dudaryonok, O. Egorov, A.J. Fleisher, H. Fleurbaey, A. Foltynowicz, T. Furtenbacher, J.J. Harrison, J.–M. Hartmann, V.–M. Horneman, X. Huang, T. Karman, J. Karns, S. Kassi, I. Kleiner, V. Kofman, F. Kwabia–Tchana, N.N. Lavrentieva, T.J. Lee, D.A. Long, A.A. Lukashevskaya, O.M. Lyulin, V.Yu. Makhnev, W. Matt, S.T. Massie, M. Melosso, S.N. Mikhailenko, D. Mondelain, H.S.P. Müller, O.V. Naumenko, A. Perrin, O.L. Polyansky, E. Raddaoui, P.L. Raston, Z.D. Reed, M. Rey, C. Richard, R. Tóbiás, I. Sadiek, D.W. Schwenke, E. Starikova, K. Sung, F. Tamassia, S.A. Tashkun, J. Vander Auwera, I.A. Vasilenko, A.A. Vigasin, G.L. Villanueva, B. Vispoel, G. Wagner, A. Yachmenev, and S.N. Yurchenko. The HITRAN2020 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, page 107949, 2021. 2, 3

[19] Mohit Gupta and Nikhil Nakhate. A Geometric Perspective on Structured Light Coding. In *ECCV*, volume 11220, pages 90–107, 2018. 2

[20] Harel Haim, Shay Elmalem, Raja Giryes, Alex M. Bronstein, and Emanuel Marom. Depth Estimation From a Single Image Using Deep Learned Phase Coded Mask. *IEEE Trans. Comput. Imaging*, 4(3):298–310, Sept. 2018. 2

[21] Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Horaud. *Time of Flight Cameras: Principles, Methods, and Applications*. SpringerBriefs in Computer Science. Springer, Oct. 2012. 1, 2

[22] Charles M. Hanson. Implications of 1/f noise in uncooled thermal imaging. In *Infrared Technology and Applications XLIV*, volume 10624, pages 142 – 153. International Society for Optics and Photonics, 2018. 5

[23] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2003. 1, 2

[24] Caner Hazirbas, Sebastian Georg Soyer, Maximilian Christian Staab, Laura Leal-Taixé, and Daniel Cremers. Deep depth from focus. In *ACCV*, Dec 2018. 2

[25] Berthold K. P. Horn. Obtaining shape from shading information. *The Psychology of Computer Vision*, pages 115–155, 1975. 1

[26] John R. Howell, M. Pinar Mengüç, and Robert Siegel. *Thermal Radiation Heat Transfer, 6th Edition*. CRC Press, 2016. 3

[27] Gordon B. Hunter, Charly D. Allemand, and Thomas W. Eagar. Multiwavelength Pyrometry: An Improved Method. *Opt. Eng.*, 24(6):241081, Dec. 1985. 2

[28] Masaki Kaga, Takahiro Kushida, Tsuyoshi Takatani, Kenichiro Tanaka, Takuya Funatomi, and Yasuhiro Mukaigawa. Thermal non-line-of-sight imaging from specular and diffuse reflections. *IPSJ Trans. Comput. Vis. Appl*, 11(1):8, Nov. 2019. 2

[29] Namil Kim, Yukyung Choi, Soonmin Hwang, and In So Kweon. Multispectral Transfer Network: Unsupervised Depth Estimation for All-day Vision. In *AAAI*, page 9, 2018. 2

[30] M. Kuo, R. Kawahara, S. Nobuhara, and K. Nishino. Non-rigid shape from water. *IEEE TPAMI*, 43(07):2220–2232, Jul 2021. 2

[31] Byeong-Uk Lee, Kyunghyun Lee, Jean Oh, and In So Kweon. CNN-Based Simultaneous Dehazing and Depth Estimation. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 9722–9728, May 2020. 2

[32] Miao Liao, Liang Wang, Ruigang Yang, and Minglun Gong. Light Fall-off Stereo. In *CVPR*, pages 1–8, Jun 2007. 2

[33] Yawen Lu and Guoyu Lu. An Alternative of LiDAR in Nighttime: Unsupervised Depth Estimation Based on Single Thermal Image. In *WACV*, pages 3832–3842, Jan. 2021. 2

[34] Tomohiro Maeda, Yiqin Wang, Ramesh Raskar, and Achuta Kadambi. Thermal Non-Line-of-Sight Imaging. In *ICCP*, pages 1–11, May 2019. 2

[35] Yue Ming, Xuyang Meng, Chunxiao Fan, and Hui Yu. Deep learning for monocular depth estimation: A review. *Neurocomputing*, 438:14–33, May 2021. 2

[36] W. Minkina and D. Klecha. Atmospheric transmission coefficient modelling in the infrared for thermovision measurements. *Journal of Sensors and Sensor Systems*, 5(1):17–23, Jan. 2016. 2, 3

[37] Parsa Mirdehghan, Wenzheng Chen, and Kiriakos N. Kutulakos. Optimal Structured Light a la Carte. In *CVPR*, pages 6248–6257, June 2018. 2

[38] Seiichi Mita, Xu Yuquan, Kazuhisa Ishimaru, and Sakiko Nishino. Robust 3d perception for any environment and any weather condition using thermal stereo. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 2569–2574, June 2019. 2

[39] 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):687–694, Apr. 2002. 2

[40] Daisuke Miyazaki, Robby T. Tan, Kenji Hara, and Katsushi Ikeuchi. Polarization-based inverse rendering from a single view. In *ICCV*, volume 2, pages 982–987, 2003. 2

[41] Frank Niklaus, Christer Jansson, Adit Decharat, Jan-Erik Källhammer, Håkan Pettersson, and Göran Stemme. Uncooled infrared bolometer arrays operating in a low to medium vacuum atmosphere: performance model and tradeoffs. In *Infrared Technology and Applications XXXIII*, volume 6542, pages 588 – 599. International Society for Optics and Photonics, 2007. 5

[42] Mark Nishimura, David B. Lindell, Christopher Metzler, and Gordon Wetzstein. Disambiguating Monocular Depth Estimation with a Single Transient. In *ECCV*, volume 12366, pages 139–155, 2020. 2

[43] Robert P. O'Shea, Shane G. Blackburn, and Hiroshi Ono. Contrast as a depth cue. *Vision Research*, 34(12):1595–1604, 1994. 2

[44] Alex Paul Pentland. A New Sense for Depth of Field. *IEEE TPAMI*, PAMI-9(4):523–531, July 1987. 1, 2

[45] Matteo Poggi, Fabio Tosi, Konstantinos Batsos, Philippos Mordohai, and Stefano Mattoccia. On the Synergies between Machine Learning and Binocular Stereo for Depth Estimation from Images: a Survey. *IEEE TPAMI*, pages 1–1, 2021. 2

[46] S. Rahmann and N. Canterakis. Reconstruction of specular surfaces using polarization imaging. In *CVPR*, volume 1, pages I–149–I–155, 2001. 1, 2

[47] Ilya Reshetouski, Hideki Oyaizu, Kenichiro Nakamura, Ryuta Satoh, Suguru Ushiki, Ryuichi Tadano, Atsushi Ito, and Jun Murayama. Lensless Imaging with Focusing Sparse URA Masks in Long-Wave Infrared and Its Application for Human Detection. In *ECCV*, volume 12364, pages 237–253, 2020. 2

[48] Thomas A. Russell, Lenore McMackin, Bob Bridge, and Richard Baraniuk. Compressive hyperspectral sensor for LWIR gas detection. In *Compressive Sensing*, volume 8365, pages 55 – 67. International Society for Optics and Photonics, 2012. 2

[49] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666–2680, Aug. 2010. 2

[50] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep Photometric Stereo Network. In *ICCVW*, pages 501–509, Oct. 2017. 2

[51] Philip Saponaro, Scott Sorensen, Abhishek Kolagunda, and Chandra Kambhamettu. Material classification with thermal imagery. In *CVPR*, pages 4649–4656, June 2015. 2

[52] Daniel Scharstein and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *IJCV*, 47:7–42, 2002. 2

[53] Zhanghao Sun, David B. Lindell, Olav Solgaard, and Gordon Wetzstein. Spadnet: deep rgb-spad sensor fusion assisted by monocular depth estimation. *Opt. Express*, 28(10):14948–14962, May 2020. 2

[54] Shiyu Tan, Yicheng Wu, Shoou-I Yu, and Ashok Veeraraghavan. CodedStereo: Learned Phase Masks for Large Depth-of-field Stereo. In *CVPR*, pages 7166–7175, June 2021. 2

[55] 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*, pages 1–1, 2019. 2

[56] Wayne Treible, Philip Saponaro, Scott Sorensen, Abhishek Kolagunda, Michael ONeal, Brian Phelan, Kelly Sherbondy, and Chandra Kambhamettu. CATS: A Color and Thermal Stereo Benchmark. In *CVPR*, pages 134–142, Jul 2017. 2

[57] Shimon Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405–426, Jan. 1979. 2

[58] Michael Vollmer and Klaus-Peter Möllmann. *Infrared Thermal Imaging: Fundamentals, Research and Applications, 2nd Edition*. Wiley-VCH Verlag GmbH & Co. KGaA, Feb. 2017. 2, 3, 5

[59] Xiang Wang, Chen Wang, Bing Liu, Xiaoqing Zhou, Liang Zhang, Jin Zheng, and Xiao Bai. Multi-view stereo in the Deep Learning Era: A comprehensive revfiew. *Displays*, 70:102102, Dec. 2021. 2

[60] Andrew P. Witkin. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17(1):17–45, 1981. 1

[61] Robert J. Woodham. Photometric Method For Determining Surface Orientation From Multiple Images. *Opt. Eng.*, 19(1), Feb. 1980. 2

[62] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. PhaseCam3D — Learning Phase Masks for Passive Single View Depth Estimation. In *ICCP*, pages 1–12, May 2019. 2

[63] Yalin Xiong and Steven Shafer. Depth from focusing and defocusing. In *CVPR*, pages 68 – 73, Jun 1993. 2

[64] Ruo Zhang, Ping-Sing Tsai, J.E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE TPAMI*, 21(8):690–706, Aug. 1999. 2

[65] Yinda Zhang, Sameh Khamis, Christoph Rhemann, Julien Valentin, Adarsh Kowdle, Vladimir Tankovich, Michael Schoenberg, Shahram Izadi, Thomas Funkhouser, and Sean Fanello. ActiveStereoNet: End-to-End Self-supervised Learning for Active Stereo Systems. In *ECCV*, volume 11212, pages 802–819, 2018. 2

[66] Chaoqiang Zhao, Qiyu Sun, Chongzhen Zhang, Yang Tang, and Feng Qian. Monocular depth estimation based on deep learning: An overview. *Science China Technological Sciences*, pages 1–16, 2020. 2

[67] Chao Zuo, Shijie Feng, Lei Huang, Tianyang Tao, Wei Yin, and Qian Chen. Phase shifting algorithms for fringe projection profilometry: A review. *Optics and Lasers in Engineering*, 109:23–59, Oct. 2018. 2