A Theory of Joint Light and Heat Transport for Lambertian Scenes

Mani Ramanagopal, Sriram Narayanan, Aswin C. Sankaranarayanan, and Srinivasa G. Narasimhan
Carnegie Mellon University, Pittsburgh, PA 15213, USA
{manikans, snochurn, saswin, srinivas}@andrew.cmu.edu

Abstract

We present a novel theory that establishes the relationship between light transport in visible and thermal infrared, and heat transport in solids. We show that heat generated due to light absorption can be estimated by modeling heat transport using a thermal camera. For situations where heat conduction is negligible, we analytically solve the heat transport equation to derive a simple expression relating the change in thermal image intensity to the absorbed light intensity and heat capacity of the material. Next, we prove that intrinsic image decomposition for Lambertian scenes becomes a well-posed *problem if one has access to the absorbed light. Our theory generalizes to arbitrary shapes and unstructured illumination. Our theory is based on applying energy conservation principle at each pixel independently. We validate our theory using real-world experiments on diffuse objects made of different materials that exhibit both direct and global components (inter-reflections) of light transport under unknown complex lighting.*

1. Introduction

Printed on paper, this text appears black because the ink does not reflect much light. So what happens to the light striking the ink? It gets absorbed and converted into heat, thereby disappearing from the visible light transport system. Starting from the early works in 1970s [3, 20, 23, 27, 31, 35], decades of research[6, 18, 33] have attempted to separate surface reflectance and shading from images by modeling shapes[30], illuminations [8] and their interactions [13]. However, in the general case, decomposing light transport is fundamentally an ill-posed problem, thus requiring handcrafted [23] or learned priors[1, 2, 4, 9]. *But what if we can somehow observe the light lost to absorption?*

Analogous to light transport, heat transport models the generation and flow of heat through a medium and its exchange with the surrounding [5, 34]. In the heat transport system, the heat generated due to light absorption is no different from any other type of heat generation. While heat itself cannot be seen, all objects radiate infrared light based

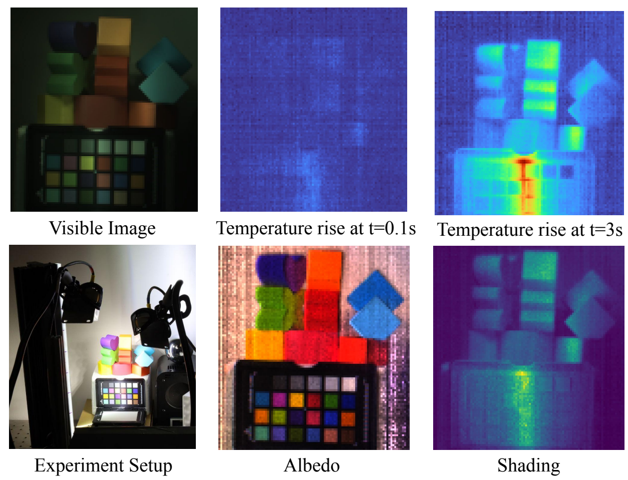


Figure 1. The visible image (brightened for visualization) captures the reflected light, which is the product of albedo and shading at that pixel. The absorbed light gets converted into heat and raises the temperature, which can be observed using a co-located thermal camera. The illumination is turned on at $t = 0$. The temperature rise at $t = 0.1s$ and $t = 3s$ are shown using *turbo* colormap. Our novel theory establishes the relationship between light and heat transport and provides an analytical solution to compute albedo and shading for complex shapes and unknown illumination.

on their surface temperature, and that can be measured using a thermal camera [34]. By modeling heat transport, we make the first attempt to estimate the intensity of light absorbed by an object, thus establishing the connection between light and heat transport.

We develop a novel theory that proves having access to absorbed light turns single view intrinsic image decomposition into a pixel-wise *well-posed* problem, even for arbitrary shape and illumination. Our key insight is that all the complexities of the *reflected* light transport are also present in the *absorbed* light, in the same functional form but simply scaled by the complement of the albedo. Consider the color chart seen in Fig. 1. The amount of irradiance due to the line light is approximately equal for the black and white patches. While the visible image records a low intensity for the black patch, the corresponding increase in intensity in the thermal images is high, and vice versa. Leveraging the principle of energy conservation, the sum of reflected light and absorbed light at each scene point must equal its irradiance, which is also called as shading. Similarly, we can compute the ratio of reflected light to irradiance, which is also called as surface reflectance or albedo.

A key ingredient in our approach is the ability to estimate the intensity of heat generated due to light absorption. In the general case, estimating it requires solving the heat transport equation which does not have an analytical solution for unknown shapes [5, 34]. However, in the absence of heat conduction, we show that the analytical solution to the heat transport equation for a constant source is a transient response that follows a 2-parameter exponential curve. Therefore, the source intensity can be estimated with as little as three frames from a thermal video. In practice, conduction occurs in all real-world objects albeit to a smaller degree in insulators and regions with low temperature gradient. Therefore, we limit the influence of conduction by focusing on the transient response of each pixel immediately after turning on light. A key limitation of our approach is that we require the system to be at thermal equilibrium before the light is turned on and other sources of heat generation, if any, remain constant. This is required to ensure the rise in temperature is only due to the absorbed light.

Prior works in computational thermal imaging have studied the thermal transient response of objects to heating. Dashpute et al. [11] heat planar objects using a laser and capture a 1 min long thermal video to estimate its thermal diffusivity and emissivity. Of most relevance to our work, Tanaka et al. [32] heat objects using infrared lamps and record a 10 mins long thermal video. They decompose these videos using curve fitting into ambient, specular, diffuse and global components, where the latter two are assumed to be exponential curves. But this decomposition is akin to direct-global separation which is different from intrinsic image decomposition. Also, they use the extracted diffuse component as input to a photometric stereo algorithm. Note that their estimated "albedo" corresponds to absorptivity in the infrared spectrum and their photometric stereo is limited to distant point light sources at known directions (separate video for each direction). In contrast, our theory establishes and exploits the causal relationship between light and heat transport. And we apply our theory to albedo-shading separation in the visible spectrum for arbitrary unknown illumination using a single 4 sec thermal video and a single visible image.

Several works in vision [17, 36, 37] and robotics [22, 25] fuse spatial features from the visible and thermal images in order to improve robustness of downstream tasks, such as object detection [24], to lighting and weather conditions. However, these methods do not reason about the relationship between the two spectrums from a physics perspective.

We validate our theory through real world experiments using a co-located setup of a visible and thermal camera. Our target objects, even though diffuse, are made of different materials, contain direct and global light transport (inter-reflections), low and high spatial frequency and unstructured illumination, all of which are unknown.

2. Joint Light and Heat Transport

In this section, we briefly introduce the relationship between light and heat transport. While light energy is carried via photons, heat is thermal energy exchanged via molecular vibrations. Visible light (VIS, $0.4-0.7\mu$m) transport can model the light scattered by the scene from a source towards the camera. The light absorbed by the scene gets converted to heat which is then exchanged via conduction, convection, retention (*i.e.* increase in temperature) and radiation, and is governed by the heat transport equation. Similar to VIS transport, Longwave Infrared light (LWIR, $8-14\mu$m) transport can be used to model the radiation emitted by objects, based on their temperature, towards a thermal camera.

Our first contribution is an algorithm, described in Sec. 3, for estimating the intensity of absorbed light using only a thermal video. This involves two steps: 1) inferring temperatures using LWIR light transport, and 2) inferring source intensity using heat transport equation. As all objects in the scene constantly exchange heat, it is hard to disambiguate heat generated by light absorption from other sources of heat at equilibrium. However, if we disturb the equilibrium by turning on the visible light at a known time, then the resulting rise in temperature allows us to estimate heat generated only due to our illumination.

Our second contribution is a novel theory, described in Sec. 4, that decomposes VIS transport for arbitrary shapes and illumination. We derive simple analytical expressions for albedo and shading using a visible image and the absorbed light intensity estimated from a thermal video captured by a co-located thermal camera.

3. Estimating Absorbed Light Intensity

Consider a scene initially at thermal equilibrium. At a time t_1, the illumination, which is constant with time, is turned on and a thermal video is captured. We assume the illumination is focused primarily at the target scene and therefore the temperature of the surrounding remains constant. Our objective is to estimate the spatially varying absorbed light (heat source) intensity using a single thermal video.

3.1. Thermal Images to Temperature Changes

In LWIR light transport, all surfaces including the camera and the scene emit (and reflect) radiation. The pixel intensity in the n^{th} frame $I_n(\mathbf{p})$ of a thermal video can be written as:

$$I_n(\mathbf{p}) = \alpha U(T_n(\mathbf{x})) + U_s, \tag{1}$$

where $T_n(\mathbf{x})$ is the temperature at time t_n, α is the effective emissivity, U_s denotes the radiation from the surrounding, and $U(T)$ is a non-linear function that approximates the integral of the Planck radiation law.

For a small range around T_*, $U(T)$ can be linearly approximated as:

$$U(T) = k_1(T - T_*) + k_2, \tag{2}$$

where k_1 and k_2 are camera-specific constants that depend on T_*. We refer the reader to Appendix A for more details. Combining Eq. (1) and Eq. (2), we get

$$I_n(\mathbf{p}) - I_m(\mathbf{p}) = k_1\alpha(T_n(\mathbf{x}) - T_m(\mathbf{x})). \tag{3}$$

The above equation shows that change in pixel intensity is linearly related to change in scene temperature. Note that commonly used thermal cameras are uncooled microbolometers that exhibit thermal inertia [28, 29], where the measured intensities have a small delay with respect to changes in the scene. This effect is ignored for the purposes of this paper.

3.2. Heat Transport Equation without Conduction

Consider an infinitesimal volume at a scene point with area δ_A and depth δ_z. The heat transport equation at that point can be written as [5, 34]:

$$\begin{aligned} C_v\delta_A\delta_z\frac{\partial T}{\partial t} = \kappa\delta_A\delta_z\Delta T + \delta_A h_c(T_s - T) + \\ \delta_A\sigma\epsilon(T_s^4 - T^4) + \delta_A S, \end{aligned} \tag{4}$$

where C_v is the volumetric heat capacity, T is the temperature, κ is the thermal conductivity, Δ denotes the laplacian operator at that point, h_c is the convection coefficient, T_s is the surrounding temperature, σ is the Stefan-Boltzmann constant, ϵ is the surface emissivity, and S is the intensity of heat generated via light absorption. Note that all the terms are expressed in units of W. For an opaque Lambertian scene, all the light absorption happens near the surface.

Note that the magnitude of heat conduction is proportional to the local temperature laplacian. Analytically solving Eq. (4) requires knowing the shape since the laplacian operator depends on the local curvature. Ignoring conduction, makes the heat equation pixel-wise independent and lends itself to an analytical solution independent of shape. Moreover, many real-world materials, such as paints, plastics, paper and wood, have low thermal conductivity. As the object is initially at equilibrium, local temperature laplacians start at zero and increase with time if and only if neighboring pixels have different material properties and/or receive different amounts of light. Therefore, we consider a short thermal video immediately after light is turned on when conduction can be ignored.

Dividing by area and ignoring conduction, Eq. (4) is:

$$C_v\delta_z\frac{\partial T}{\partial t} = h_c(T_s - T) + \sigma\epsilon(T_s^4 - T^4) + S. \tag{5}$$

Since temperature rise due to light absorption is typically small ($\leq 15\text{K}$ within 4 sec in our experiments), we linearize the radiation term around a nominal temperature T_* to get

$$\sigma\epsilon(T_s^4 - T^4) \approx 4\sigma\epsilon T_*^3(T_s - T), \tag{6}$$

where the absolute error due to linearization is $\leq 4\%$. This simplifies Eq. (5) to

$$H\frac{\partial T}{\partial t} + PT = S + PT_s, \tag{7}$$

where $H = C_v\delta_z$ and $P = (h_c + 4\sigma\epsilon T_*^3)$.

3.3. Analytical Solution

Solving Eq. (7) at a single pixel (refer Appendix B for derivation), we get

$$T_n - T_1 = \left(\frac{S}{P} + T_s - T_1\right)\left(1 - e^{-\frac{P}{H}(t_n - t_1)}\right). \tag{8}$$

Since we assume the system is initially at thermal equilibrium, we can set $T_s = T_1$. Now, substituting Eq. (3) into the above equation, we get

$$I_n - I_1 = \frac{Sk_1\alpha}{P}\left(1 - e^{-\frac{P}{H}(t_n - t_1)}\right) \tag{9}$$

Therefore, given a thermal video $\{I_1, \ldots, I_n\}$ and corresponding time stamps $\{t_1, \ldots, t_n\}$, we use gradient descent for curve fitting at each pixel independently:

$$I_n - I_1 = c_1(1 - e^{-\frac{t_n - t_1}{c_2}}). \tag{10}$$

3.4. Recovering S from Curve Fitting

The results of curve fitting provide c_1 and c_2 at each pixel. From Eq. (9) and Eq. (10), note that $c_1 = \frac{Sk_1\alpha}{P}$. Recovering S from c_1 would require knowledge of k_1, α and P, where P depends on h_c, ϵ and T_*. In theory, all these quantities could vary per-pixel. However, the spatial variation in S, which depends on albedo in visible spectrum and illumination, is much greater than that of others. In this paper, we assume the quantity $\beta = \frac{k_1\alpha}{P}$ is common for all pixels such that β is the constant of proportionality between S and c_1. Note that Lambertian scenes typically correspond to rough surfaces which have high emissivity. Also, it is known that most paints have similarly high emissivity values of > 0.9 irrespective of their albedo in the visible spectrum [34]. As the object is initially at equilibrium, we can assume T_*, and hence k_1, is common for all pixels. In the absence of wind, we reasonably assume convection, if it exists at all, to be uniform throughout the scene.

Table 1. Various lighting configurations typically modeled in shape-from-intensity problems. It is trivial to verify the equations for spatially varying albedo $\rho(x)$ and shading $\eta(x)$ remains the same irrespective of the complexity of shape or illumination when estimates of both image irradiance $I_v(x)$ and absorbed light intensity $\tilde{S}(x)$ are available. Here, γ is the camera gain, β is the unknown scale factor in the estimation of $\tilde{S}(x)$, $\zeta = \frac{\gamma}{\beta}$ is the relative scale factor, E is the source intensity, $\mathbf{s}$ is the light source direction, $\mathbf{n}$ is the surface normal, ω is light source direction for extended source, η is the shading term and η^* is the scene irradiance. Inter-reflections are modeled as spatially varying source intensities. All the above cases can be extended to model cast and attached shadows using a shadowing function $W(x)$ without changing the expressions for ρ and η.

Illumination	Image Irradiance $I_v(x)$	Estimated Absorbed Light $\tilde{S}(x)$	Albedo $\rho(x)$	Shading $\eta(x)$
Far source	$\gamma\frac{\rho(x)}{\pi}E(\mathbf{s}\cdot\mathbf{n}(x))$	$\beta(1-\rho(x))E(\mathbf{s}\cdot\mathbf{n}(x))$		
Multiple sources	$\gamma\frac{\rho(x)}{\pi}\sum_l E_l(\mathbf{s}_l\cdot\mathbf{n}(x))$	$\beta(1-\rho(x))\sum_l E_l(\mathbf{s}_l\cdot\mathbf{n}(x))$		
Near sources	$\gamma\frac{\rho(x)}{\pi}\sum_l E_l(\mathbf{s}_l(x)\cdot\mathbf{n}(x))$	$\beta(1-\rho(x))\sum_l E_l(\mathbf{s}_l(x)\cdot\mathbf{n}(x))$	$\frac{\pi I_v(x)}{\pi I_v(x)+\zeta\tilde{S}(x)}$	$\pi I_v(x)+\zeta\tilde{S}(x)$
Extended Sources	$\gamma\frac{\rho(x)}{\pi}\int_\omega E(\omega)(\omega\cdot\mathbf{n}(x))$	$\beta(1-\rho(x))\int_\omega E(\omega)(\omega\cdot\mathbf{n}(x))$		
Inter-reflections	$\gamma\frac{\rho(x)}{\pi}\int_\omega E(x,\omega)(\omega\cdot\mathbf{n}(x))$	$\beta(1-\rho(x))\int_\omega E(x,\omega)(\omega\cdot\mathbf{n}(x))$		
General illumination	$\frac{\rho(x)}{\pi}\eta(x)$	$\beta(1-\rho(x))\eta(x)/\gamma$		

4. Albedo-Shading Separation

Consider an opaque Lambertian scene imaged by a camera from a fixed view. We assume that the camera is sensitive to all the wavelengths present in the light sources *i.e*. we primarily consider LEDs or CFL bulbs when using visible cameras. We first consider the case where the albedo and the camera response are independent of wavelength in Sec. 4.1 and then extend our theory to wavelength-dependent albedo functions in Sec. 4.2. The words image and camera correspond to visible spectrum in this section.

4.1. Grayscale Albedo and Camera Response

The image intensity I_v, which is proportional to the power received by the camera per unit area, at a pixel $\mathbf{p}(\mathbf{x})$ focused at a scene point $\mathbf{x}$ is:

$$I_v(\mathbf{p}(\mathbf{x})) = \frac{\rho(\mathbf{x})}{\pi}\eta(\mathbf{x}),\ \text{s.t. } \eta(\mathbf{x}) \equiv \gamma\eta^*(\mathbf{x}) \tag{11}$$

where $\rho(\mathbf{x})$ and $\eta(\mathbf{x})$ are the spatially varying albedo and shading, $\gamma > 0$ is the camera gain representing the optics and sensor electronics in the camera, and $\eta^*(\mathbf{x})$ is the true scene irradiance received by $\mathbf{x}$. Note that we do not restrict the lighting geometry in any way and the shading $\eta(\mathbf{x})$ term is unstructured. In the rest of the paper, we use $\mathbf{p}$ in place of $\mathbf{p}(\mathbf{x})$.

The pixel value in an image describes the energy *reflected* towards the camera by a scene point. Since the surface is opaque, there is no transmission and the remaining energy gets *absorbed* and is converted into heat. Recall from Sec. 3 that $S(\mathbf{x})$ denotes the power absorbed per unit area , *i.e*. intensity, by $\mathbf{x}$. Let $\tilde{S}(\mathbf{x})$ be proportional to it, and is given by:

$$\tilde{S}(\mathbf{x}) = \beta S(\mathbf{x}),\ \text{s.t. } S(\mathbf{x}) = (1-\rho(\mathbf{x}))\eta^*(\mathbf{x}). \tag{12}$$

During operation, light fixtures also generate some thermal energy which increases its temperature and thereby increasing its blackbody radiation. However, the magnitude of this additional heat generated at $\mathbf{x}$ is negligible and hence ignored in this paper. Next, we can express $\tilde{S}$ using shading as

$$\tilde{S}(\mathbf{x}) = \beta(1-\rho(\mathbf{x}))\frac{\eta(\mathbf{x})}{\gamma} = \frac{(1-\rho(\mathbf{x}))\eta(\mathbf{x})}{\zeta}, \tag{13}$$

where $\zeta = \frac{\gamma}{\beta}$ is the relative scale factor.

In the trivial case where $\mathbf{x}$ receives no light (neither direct nor global illumination), the shading term $\eta(\mathbf{x}) = 0$ and the albedo cannot be estimated. Whenever $\eta(\mathbf{x}) > 0$, we can re-write Eqs. (11) and (13) as

$$\pi I_v(\mathbf{p})\frac{1}{\eta(\mathbf{x})} - \rho(\mathbf{x}) = 0 \tag{14}$$

$$\zeta\tilde{S}(\mathbf{x})\frac{1}{\eta(\mathbf{x})} + \rho(\mathbf{x}) = 1 \tag{15}$$

Solving the above system of equations, we get:

$$\eta(\mathbf{x}) = \pi I_v(\mathbf{p}) + \zeta\tilde{S}(\mathbf{x}) \tag{16}$$

$$\rho(\mathbf{x}) = \frac{\pi I_v(\mathbf{p})}{\pi I_v(\mathbf{p}) + \zeta\tilde{S}(\mathbf{x})}. \tag{17}$$

If $I_v(\mathbf{p})$, $\tilde{S}(\mathbf{x})$ and ζ are known, the above equations provide a direct method to compute spatially varying albedo and shading components for complex shapes and arbitrary illumination. To emphasize its applicability further, Table 1 lists several types of lighting conditions typically modeled in shape-from-intensity problems and demonstrates that Eqs. (16), (17) hold in all cases.

4.2. Towards General Albedo Functions

Let the camera have K channels with known spectral responses $\Gamma_k(\lambda)$. Recall that each wavelength present in the light sources must fall within at least one of the channels. The image irradiance at $\mathbf{p}$ in channel k can be written as:

$$I_v^k(\mathbf{p}) = \gamma\int_\lambda\int_\Omega \frac{\rho(\mathbf{x},\lambda)}{\pi}\Gamma_k(\lambda)L(\mathbf{x},\lambda,\omega)d\omega d\lambda, \tag{18}$$

where $\rho(\mathbf{x}, \lambda)$ is the diffuse albedo as a function of wavelength, $L(\mathbf{x}, \lambda, \omega)$ is the spectral radiance at $\mathbf{x}$, and ω denotes the direction along the outer hemisphere. The corresponding estimate of absorbed power per unit area is:

$$\tilde{S}(\mathbf{x}) = \beta \int_\lambda \int_\Omega (1 - \rho(\mathbf{x}, \lambda)) L(\mathbf{x}, \lambda, \omega) d\omega d\lambda, \tag{19}$$

Shading at a point $\mathbf{x}$ is influenced by the emission spectrum of the light sources, the relative geometry between $\mathbf{x}$ and the light sources, and the albedo of other points in the scene due to inter-reflections. While this general case remains an open problem, in the rest of this section we ignore inter-reflections and assume all light sources have a common emission spectrum $l(\lambda)$ *i.e.*

$$\int_\Omega L(\mathbf{x}, \lambda, \omega) d\omega = \eta^*(\mathbf{x}) l(\lambda). \tag{20}$$

Note that, the illumination is still arbitrary in terms of their locations, sizes and angular radiant intensity functions. Substituting Eq. (20) into Eq. (18) and Eq. (19), we can write

$$I_v^k(\mathbf{p}) = \int_\lambda \frac{\rho(\mathbf{x}, \lambda)}{\pi} \Gamma_k(\lambda) \eta(\mathbf{x}) l(\lambda) d\lambda, \tag{21}$$

$$\tilde{S}(\mathbf{x}) = \frac{\int_\lambda \eta(\mathbf{x}) l(\lambda) d\lambda - \int_\lambda \rho(\mathbf{x}, \lambda) \eta(\mathbf{x}) l(\lambda) d\lambda}{\zeta}. \tag{22}$$

As a continuous-valued function of wavelength, the diffuse albedo $\rho(\mathbf{x}, \lambda)$ is infinite-dimensional, which requires further assumptions to enable tractable computations. We rely on a body of work [10, 14–16] that shows that reflectance spectra lie close to a low-dimensional subspace. Denoting the basis for this subspace as $\Phi_\rho(\lambda) = \{\tilde{\rho}_1(\lambda), \ldots, \tilde{\rho}_M(\lambda)\}$, we can express the diffuse albedo as [21]:

$$\rho(\mathbf{x}, \lambda) = \sum_{m=1}^{M} \tilde{\rho}_m(\lambda) a_{\mathbf{x},m} = \Phi_\rho(\lambda) \mathbf{a}_\mathbf{x} \tag{23}$$

where $\mathbf{a}_\mathbf{x} \in \mathbb{R}^M$ are the unknown coefficients of interest. This simplifies Eq. (21) and Eq. (22) into

$$I_v^k(\mathbf{p}) = \eta(\mathbf{x}) \mathbf{E}_k^T \frac{\mathbf{a}_\mathbf{x}}{\pi}, \tag{24}$$

$$\tilde{S}(\mathbf{x}) = \frac{\eta(\mathbf{x})(L - \mathbf{F}^T \mathbf{a}_\mathbf{x})}{\zeta}, \tag{25}$$

where $\mathbf{E}_k$, $\mathbf{F}$ and L can be computed a priori as follows:

$$\mathbf{E_k}[i] = \int_\lambda l(\lambda) \Gamma_k(\lambda) \tilde{\rho}_i(\lambda) d\lambda, \tag{26}$$

$$\mathbf{F}[i] = \int_\lambda l(\lambda) \tilde{\rho}_i(\lambda) d\lambda, \tag{27}$$

$$L = \int_\lambda l(\lambda) d\lambda. \tag{28}$$

Whenever $\eta(\mathbf{x}) > 0$, Eqs. (24) and (25) can be written as

$$\pi I_v^k(\mathbf{p}) \xi(\mathbf{x}) - \mathbf{E}_k^T \mathbf{a}_\mathbf{x} = 0, \ \forall k \tag{29}$$

$$\zeta \tilde{S}(\mathbf{x}) \xi(\mathbf{x}) + \mathbf{F}^T \mathbf{a}_\mathbf{x} = L, \tag{30}$$

where $\xi(\mathbf{x}) = 1/\eta(\mathbf{x})$. Note that we have a system of $K+1$ linear equations with $M+1$ unknowns, namely $\mathbf{a}_\mathbf{x} \in \mathbb{R}^M$ and $\xi(\mathbf{x})$. Therefore, whenever $K \geq M$, the system of equations can be solved to obtain albedo and shading (reciprocal of $\xi(\mathbf{x})$) at each pixel independently for complex shapes and illumination. Specifically, we use non-negative least squares solver for this problem. For most vision applications, which use a 3-channel RGB camera, we choose a corresponding basis set Φ_ρ with $M = 3$. Our theory could be used with multispectral cameras with more channels when higher fidelity in albedo is desired. While the above derivation relies on Eq. (20), it is still practically useful in many real-world scenes where inter-reflections exist as we will show in Sec. 5.

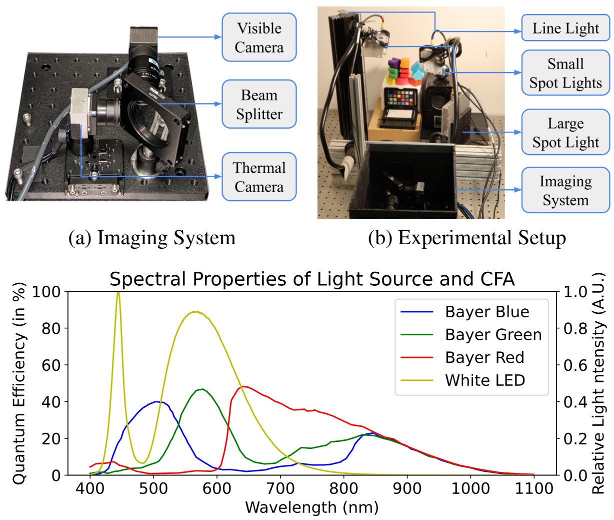
In the special case of a monochrome camera capturing a scene where albedo depends on wavelength, the shading at each pixel can be expressed as a weighted sum of I and S irrespective of the emission spectrum of the light source. We refer the reader to Appendix C for more details.

5. Experimental Results

To validate our theory, we perform experiments on several complex scenes with challenging illumination. Our scenes are mostly diffuse, but contain noticeable non-Lambertian features that test the practical utility of our theory to real-world objects. Our emphasis is on estimating the absorbed light intensity and performing albedo-shading separation.

Hardware Details: Our imaging system consists of an IDS UI-3130CP color camera with 600×800 resolution fitted with an 8mm Tamron lens, a FLIR Boson thermal camera having $\leq 50\text{mK}$ NETD with 512×640 resolution fitted with an 18mm ($24°$ HFOV) integrated lens and a BSP-DI-25-2 gold dichroic beamsplitter from ISP Optics. The cameras are coarsely aligned using an optic stage and a homography is used for fine alignment. We use LED lights from Advanced Illumination, namely a high intensity line light (LL167G96-WHI), a large spot light (SL-S100150M-WHI) and two small spot lights (SL-S050075M-WHI). The relative emission spectrum of the lighting and the spectral response of the color filter array in the visible camera were obtained from their technical datasheets, see Fig. 2.

Data Capture and Preprocessing: The visible camera was radiometrically calibrated [12]. To capture the full dynamic range of the illumination, we acquired a stack of 15



(a) Imaging System (b) Experimental Setup

(c) Spectral properties of light and CFA

Figure 2. Our imaging system consists of a visible camera and a thermal camera colocated using a gold dichroic beamsplitter. The light sources are placed close to the target scene so that the rise in temperature due to light absorption is detectable in the thermal camera. All the light sources have the same emission spectrum. The relative emission spectrum of the white LED and the quantum efficiency curves of the Color Filter Array are obtained from the corresponding datasheets.

images in BayerRG format with a geometric progression of exposures that span 0.05ms to over 180ms. All the images were demosaiced using gradient-corrected linear interpolation [26] and subsequently quantized to 8-bit images. The resulting LDR images were composited into a single linear HDR image using the previously estimated camera response function. Finally, this image is warped into the perspective of the thermal camera using a homography.

The thermal camera is allowed to reach steady state operating temperature after powering on, which can take up to 30mins. The entire experimental setup including the target object is allowed to reach thermal equilibrium before beginning data collection. A flat field correction is performed a few seconds prior to turning on the light and a thermal video is recorded. The thermal camera is operated in the high gain state and the raw 16-bit data is captured at 60Hz. Both the thermal images and the warped HDR image were downsized $4\times$ using local mean computation to suppress noise and alleviate errors in co-location.

Implementation Details: We manually identify the first frame when light was turned on and use the pixel-wise median of the preceding frames as the initial frame I_1. We use 200 frames since light was turned on for fitting the 2-parameter curve. We implement the curve fitting using gradient descent in PyTorch. We consider a 3 dimensional basis set for albedo with $\tilde{\rho}_b(\lambda) = \mathbb{I}[400\text{nm} \leq \lambda < 530\text{nm}]$,

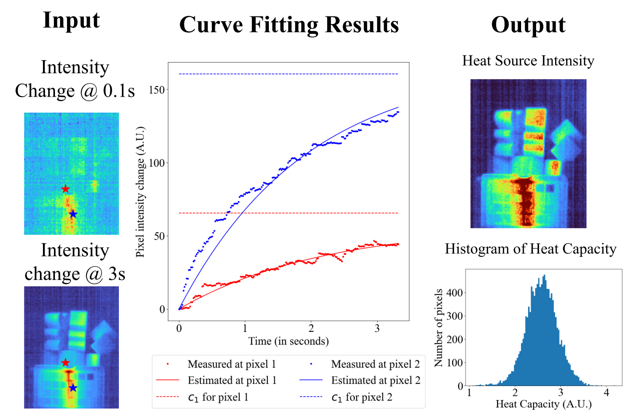


Figure 3. The first column shows change in intensity at $0.1s$ and at $3s$ after turning on the lights. Note that we process each pixel independently. Two points of significance are marked in blue and red respectively. The middle column shows the curve fitting results for the highlighted points. The input intensities are shown as dots and the estimated intensities are shown as a continuous curves. The dashed lines correspond to the steady-state intensity that would be reached and equals c_1 from Eq. (10). The last column shows c_1 for each pixel as a 2D image and a histogram of c_2.

$\tilde{\rho}_g(\lambda) = \mathbb{I}[530\text{nm} \leq \lambda < 620\text{nm}]$, $\tilde{\rho}_r(\lambda) = \mathbb{I}[620\text{nm} \leq \lambda < 1100\text{nm}]$, where $\mathbb{I}[\cdot]$ is the indicator function. Please refer to Appendix D for details of $\mathbf{E}_k$, $\mathbf{F}$ and L. The relative factor ζ can be calibrated for a co-located imaging system with a color chart under controlled lighting. Alternatively, ζ can be treated as a hyper-parameter and tuned using cross-validation.

5.1. Heat Source Estimation results

Figure 3 shows the result of our curve fitting for the wooden blocks scene. The estimated constants c_2, which is proportional to heat capacity, appear like white noise and its corresponding histogram plot resembles a Gaussian distribution. This could be due to high levels of noise in thermal videos as well as similar magnitudes of spatial variation in P and H. On the other hand, the estimated per-pixel constants c_1 have visual similarity to a shading image, although noisy.

5.2. Albedo-Shading separation results

Quantitative Evaluation: For comparison, we chose two methods: (i) the classical, even if dated, Retinex algorithm [23], which is well-suited for the color chart scene. (ii) Ordinal Shading [7], a SOTA learning-based approach which requires a large training dataset. We use the pre-trained model here. We use the scale-invariant Mean Squared Error (si-MSE) from [19] as our metric. It is hard to obtain ground truth albedo and shading for general scenes under unknown lighting. And publicly available datasets do not have co-located thermal videos. Therefore, we first evaluate albedo using the color chart under 4 different illuminations. As shown in Fig. 4, our albedo estimates are

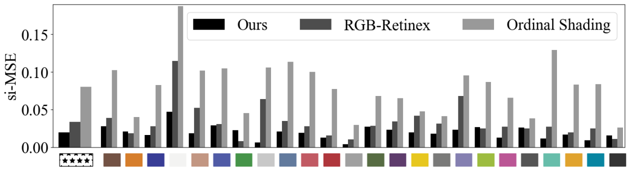


Figure 4. The first column is the mean value across colors (Ours: **0.020**, Retinex: **0.034**, Ordinal Shading: **0.080**).

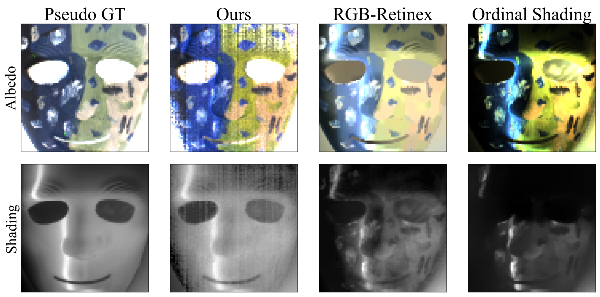


Figure 5. Our method operates per-pixel while other methods use hand-crafted or learnt spatial priors. Note the residual albedo in their estimated shading (images brightened for visualization).

	Ours	RGB-Retinex	Ordinal Shading
Albedo	**0.084**	0.253	0.399
Shading	**0.0005**	0.0030	0.0080

Table 2. si-MSE values for Albedo and Shading using pseudo ground truth data obtained for the painted mask scene.

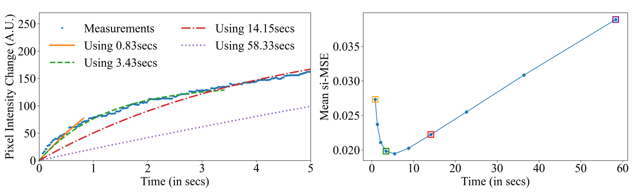


Figure 6. (a) Curve fitting results of the same pixel for different video lengths. (b) Albedo error (against color chart ground truth) vs. length of input video.

significantly better than the other methods.

Next, we obtain pseudo-ground truth similar to [19] i.e. with a scene painted white (ground truth shading) and re-painted with texture. Fig. 5 and Tab. 2 illustrate that our method outperforms SOTA methods both qualitatively and quantitatively.

Ablation on length of video: Longer duration leads to higher spatial thermal gradients that induce more conduction while shorter duration has lower signal to noise ratio. As seen in Fig. 6a, the curve deviates further away from the initial measurements when using longer videos. The accuracy of the fit correlates with the accuracy of albedo estimate for the color chart (see Fig. 6b). Our experiments use 200 frames which corresponds to the green plots.

Qualitative evaluation: Figure 7 summarizes the albedo shading separation results for the four target scenes. As shown in the first two rows, we are given a HDR image from the visible camera and the corresponding absorbed light intensity is estimated from a thermal video using curve fitting as discussed earlier. And the last two rows show results that validate Eqs. (30) which are derived for general functions of albedo and camera reponse with wavelength.

In the first scene, the interior of a mask is painted with white and black acrylic paints and the line light is directed at the portion of the image painted white. As highlighted in the callout, the concave portion corresponding to nose appears flat in the estimated albedo image for both the monochrome and RGB cases. Note that the temperature of the background wall does not raise sufficiently in all of the scenes, which makes it challenging for our approach. The thick wall would also have a high heat capacity which exacerbates the challenge. In the second scene, a cardboard sheet with printing on one side is folded to resemble the shape of W. The inner V groove would have inter-reflections while the outer faces are convex.

In the third scene, a collection of solid colored wooden blocks are stacked into a complex geometry with both cast and attached shadows. This result indirectly shows that ignoring heat conduction for solid objects still allows one to recover the absorbed light intensity precisely. In the final scene, we use a stack of disks made of soft plastic. Different patterns are embossed onto the circumference of the disk. As highlighted in the callout, the shape information corresponding to the embossing is correctly separated into the shading term while the albedo term appears flat. These results demonstrate the broad applicability of our theory to everyday scenes with complex shapes and illumination.

Grayscale approximation: Fig. 8 shows the estimated albedo and shading using grayscale approximation (Eqs. (16) and (17)). Recall that the grayscale approximation does not require knowledge of the emission spectrum of the light sources and the estimated shading is similar to that using Eqs (30). The monochrome image is approximated by taking the mean value across color channels. Corresponding results for all the scenes are provided in Appendix E.

6. Conclusion

This paper studies the theoretical connection between light transport in visible spectrum, heat transport in solids and light transport in the thermal infrared spectrum. We proved that having an estimate of absorbed light turns single image intrinsic image decomposition into a well-posed problem for arbitrary shape and illumination for lambertian scenes. To estimate absorbed light, we derive an analytical expression for surfaces with negligible heat conduction by modeling heat transport immediately after turning on illumination.

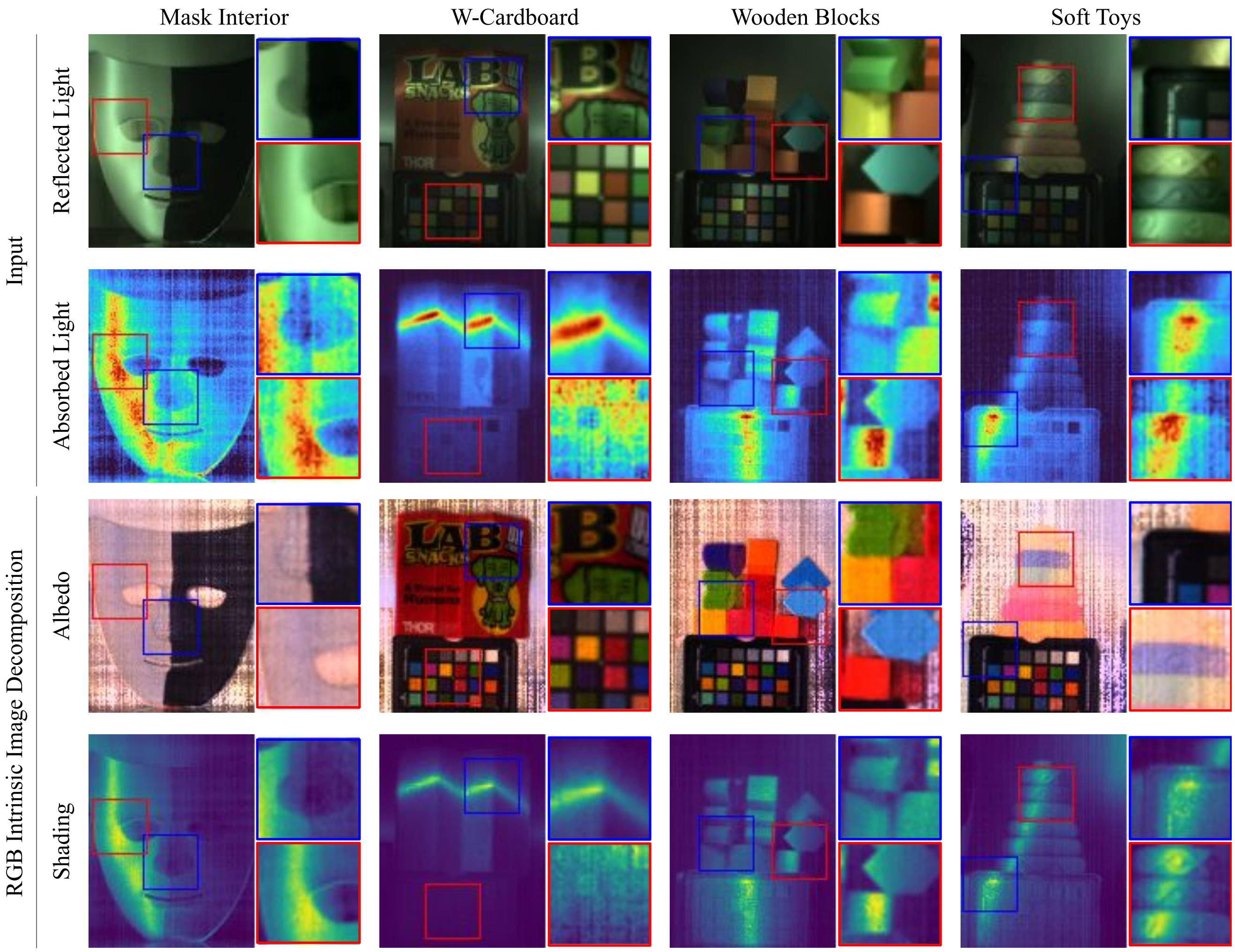


Figure 7. The first row shows the HDR visible image (brightened for visualization). Note that the colorchart is not an input to our method. The second row shows the estimated heat source intensity (turbo colormap) obtained using the method in Sec. 3. The last two rows correspond to solving Eqs. (30) using non-negative least squares method. The estimated albedo is clipped to the range $[0, 1]$. The callouts for the visible image, heat source intensity, and shading are normalized individually to aid visualization.

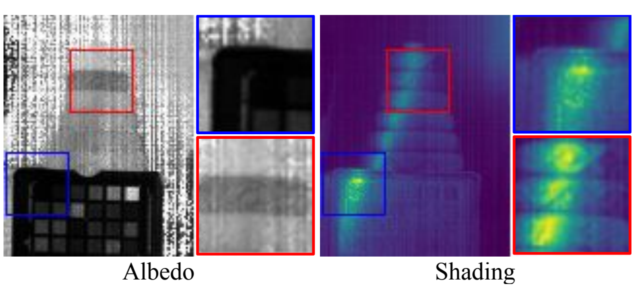


Figure 8. Albedo-Shading result for the soft toys scene using the grayscale approximation.

Experiments showed that albedo and shading can be measured from a single view given a visible image and a short thermal video from a co-located imaging system.

Just like we have shown an example of how modeling heat transport can help solve challenges in visible light transport, we believe research in visible light transport can help Infrared Thermography by improving accuracy of temperature measurement or observing heat transfer within inhomogenous surfaces. Extending our theory to the full light transport, including general BRDFs, translucent materials and subsurface scattering are just a few of the exciting new directions that this research opens up.

Acknowledgements

This work was partly supported by NSF grants IIS-2107236, CCF-1730147, and NSF-NIFA AI Institute for Resilient Agriculture. The authors would like to thank Mark Sheinin for helpful discussions.

References

[1] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014. 1

[2] Anil S Baslamisli, Hoang-An Le, and Theo Gevers. Cnn based learning using reflection and retinex models for intrinsic image decomposition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6674–6683, 2018. 1

[3] Peter N Belhumeur, David J Kriegman, and Alan L Yuille. The bas-relief ambiguity. *International journal of computer vision*, 35(1):33–44, 1999. 1

[4] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)*, 33(4): 1–12, 2014. 1

[5] Theodore L. Bergman. *Introduction to Heat Transfer*. Wiley, 2011. 1, 2, 3

[6] Nicolas Bonneel, Balazs Kovacs, Sylvain Paris, and Kavita Bala. Intrinsic decompositions for image editing. In *Computer Graphics Forum*, pages 593–609. Wiley Online Library, 2017. 1

[7] Chris Careaga and Yağız Aksoy. Intrinsic image decomposition via ordinal shading. *ACM Transactions on Graphics*, 43 (1):1–24, 2023. 6

[8] Robert Carroll, Ravi Ramamoorthi, and Maneesh Agrawala. Illumination decomposition for material recoloring with consistent interreflections. *ACM Trans. Graph.*, 30(4):43, 2011. 1

[9] Jason Chang, Randi Cabezas, and John W Fisher III. Bayesian nonparametric intrinsic image decomposition. In *European conference on computer vision*, pages 704–719. Springer, 2014. 1

[10] Hamilton Y Chong, Steven J Gortler, and Todd Zickler. The von kries hypothesis and a basis for color constancy. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007. 5

[11] Aniket Dashpute, Vishwanath Saragadam, Emma Alexander, Florian Willomitzer, Aggelos Katsaggelos, Ashok Veeraraghavan, and Oliver Cossairt. Thermal spread functions (tsf): Physics-guided material classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1641–1650, 2023. 2

[12] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, page 369–378, USA, 1997. ACM Press/Addison-Wesley Publishing Co. 5

[13] Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Jorge Lopez-Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. Multi-view intrinsic images of outdoors scenes with an application to relighting. *ACM Transactions on Graphics*, page 16, 2015. 1

[14] Graham D Finlayson, Mark S Drew, and Brian V Funt. Color constancy: enhancing von kries adaption via sensor transformations. In *Human Vision, Visual Processing, and Digital Display IV*, pages 473–484, 1993. 5

[15] Graham D Finlayson, Mark S Drew, and Brian V Funt. Diagonal transforms suffice for color constancy. In *IEEE International Conference on Computer Vision*, pages 164–171, 1993.

[16] Graham Fyffe, Xueming Yu, and Paul Debevec. Single-shot photometric stereo by spectral multiplexing. In *IEEE International Conference on Computational Photography (ICCP)*, 2011. 5

[17] Rikke Gade and Thomas B Moeslund. Thermal cameras and applications: a survey. *Machine vision and applications*, 25: 245–262, 2014. 2

[18] Elena Garces, Carlos Rodriguez-Pardo, Dan Casas, and Jorge Lopez-Moreno. A survey on intrinsic images: Delving deep into lambert and beyond. *International Journal of Computer Vision*, 130(3):836–868, 2022. 1

[19] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2335–2342. IEEE, 2009. 6, 7

[20] Berthold KP Horn and Michael J Brooks. *Shape from shading*. MIT press, 1989. 1

[21] Zhuo Hui, Kalyan Sunkavalli, Sunil Hadap, and Aswin C Sankaranarayanan. Illuminant spectra-based source separation using flash photography. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6209–6218, 2018. 5

[22] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1037–1045, 2015. 2

[23] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971. 1, 6

[24] Qiao Liu, Zhenyu He, Xin Li, and Yuan Zheng. Ptb-tir: A thermal infrared pedestrian tracking benchmark. *IEEE Transactions on Multimedia*, 22(3):666–675, 2019. 2

[25] Yawen Lu and Guoyu Lu. Superthermal: Matching thermal as visible through thermal feature exploration. *IEEE Robotics and Automation Letters*, 6(2):2690–2697, 2021. 2

[26] Henrique S Malvar, Li-wei He, and Ross Cutler. High-quality linear interpolation for demosaicing of bayer-patterned color images. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages iii–485. IEEE, 2004. 6

[27] Shree K Nayar, Katsushi Ikeuchi, and Takeo Kanade. Shape from interreflections. *International Journal of Computer Vision*, 6:173–195, 1991. 1

[28] Manikandasriram Srinivasan Ramanagopal, Zixu Zhang, Ram Vasudevan, and Matthew Johnson-Roberson. Pixel-wise motion deblurring of thermal videos. *arXiv preprint arXiv:2006.04973*, 2020. 3

[29] Vishwanath Saragadam, Akshat Dave, Ashok Veeraraghavan, and Richard G. Baraniuk. Thermal image processing via physics-inspired deep networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 4057–4065, 2021. 3

[30] Steven M Seitz, Yasuyuki Matsushita, and Kiriakos N Kutulakos. A theory of inverse light transport. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, pages 1440–1447. IEEE, 2005. 1

[31] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985. 1

[32] 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 Transactions on Pattern Analysis and Machine Intelligence*, 43(6):2075–2085, 2019. 2

[33] Ayush Tewari, Ohad Fried, Justus Thies, Vincent Sitzmann, Stephen Lombardi, Kalyan Sunkavalli, Ricardo Martin-Brualla, Tomas Simon, Jason Saragih, Matthias Nießner, et al. State of the art on neural rendering. In *Computer Graphics Forum*, pages 701–727. Wiley Online Library, 2020. 1

[34] Michael Vollmer and Klaus-Peter Mollmann. *Fundamentals of Infrared Thermal Imaging*, chapter 1, pages 1–106. John Wiley & Sons, Ltd, 2017. 1, 2, 3

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

[36] Pengyu Zhang, Jie Zhao, Dong Wang, Huchuan Lu, and Xiang Ruan. Visible-thermal uav tracking: A large-scale benchmark and new baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8886–8895, 2022. 2

[37] Xingchen Zhang, Ping Ye, and Gang Xiao. Vifb: A visible and infrared image fusion benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 104–105, 2020. 2

A Theory of Joint Light and Heat Transport for Lambertian Scenes

Supplementary Material

A. Image Formation in a Thermal Camera

The field of Infrared Thermography is focused on recovering precise temperature measurements from the intensities recorded by a thermal camera. For an in-depth understanding of this subject, we refer the reader to [34]. In the following, we summarize the key concepts as pertaining to our system.

Let ϵ denote the emissivity of the object and T_n denote the corresponding surface temperature at time t_n. The corresponding intensity I_{thr} returned by an ideal thermal camera is written as

$$\begin{aligned} I_{thr}(t_n) =& r_{bs}\bigg(\tau_{atm}\big(\epsilon U(T_n) + (1-\epsilon)U(T_{refl})\big) + \\ & (1-\tau_{atm})U(T_{atm})\bigg) + \tau_{bs}U(T_{sys}) + \\ & (1-r_{bs}-\tau_{bs})U(T_{bs}), \end{aligned} \tag{31}$$

where r_{bs}, τ_{bs}, and T_{bs} are the reflectivity, transmissivity and temperature of the beam splitter respectively. τ_{atm} and T_{atm} denote the transmissivity and temperature of the atmospheric medium between the camera and the object. T_{sys} is the effective temperature of the imaging system and T_{refl} is the effective temperature of radiation from the surrounding that is incident on the object. Note that the above expression, also called the radiometric chain, models the factors external to the camera, such as the different components of radiation that are incident at the lens of the thermal camera.

Factors internal to the camera, such as its operating temperature, lens characteristics, its type (microbolometer vs. cooled photon detectors), and sensor electronics, also influence the final intensity recorded by the camera. The resulting cumulative effect is modeled by the function U, called the radiometric function of the thermal camera.

In our experiment, we assume all components other than the object surface remain at constant temperature. This assumption is reasonable as our experiments are limited to short duration (<3.5 seconds). Therefore Eq. (31) can be written as

$$I_{thr}(t_n) = \alpha U(T_n) + U_s \tag{32}$$

$$\text{where} \quad \alpha = r_{bs}\tau_{atm}\epsilon \tag{33}$$

$$\begin{aligned} U_s = & r_{bs}\big(\tau_{atm}(1-\epsilon)U(T_{refl}) + \\ & (1-\tau_{atm})U(T_{atm})\big) + \tau_{bs}U(T_{sys}) + \\ & (1-r_{bs}-\tau_{bs})U(T_{bs}). \end{aligned} \tag{34}$$

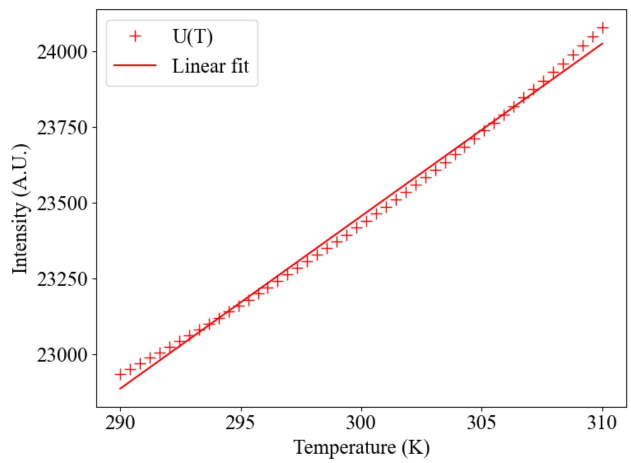


Figure 9. The radiometric function of a typical thermal camera for a small temperature range around room temperature. The linear fit defined with $T_* = 300$K introduces $< 0.2\%$ absolute error in intensity with respect to the mean pixel intensity in that temperature range.

A.1. Radiometric Function of Thermal Camera

The thermal camera's radiometric function U maps the temperature of a blackbody to the corresponding pixel intensity the camera would measure under ideal conditions. It is parameterized by the planckian form of the Sakumo-Hattori equations and is given by:

$$U(T) = \frac{R}{\exp(\frac{B}{T}) - F} + O, \tag{35}$$

where R, B, F and O are camera calibration parameters. Note that other forms of Sakumo-Hattori equations exist but the above expression is the one typically used by the camera manufacturer.

The function is typically defined over a broad range of temperatures, say $[-40°C, 150°C]$. However, the function can be linearized around a nominal temperature T_*. In all our experiments, the rise in pixel intensity due to light absorption was less than ≈ 1000 counts. Fig. 9 shows the plot of U for typical values of R, B, F, O and a small temperature range around room temperature. The linear fit defined around $T_* = 300$K agrees well with the non-linear function. The maximum error introduced due to linearization expressed as percentage of mean pixel intensity is 0.2%.

B. Analytical Solution to Eq. (7)

Eq. (7) from Sec. 3 can be written as

$$\frac{\partial T}{\partial t} = \frac{P(T_s - T)}{H} + \frac{S}{H}. \quad (36)$$

This can be written in standard form as

$$\int_{T_1}^{T_n} \frac{dT}{A - BT} = \int_{t_1}^{t_n} dt, \quad (37)$$

where $A = \frac{PT_s + S}{H}$ and $B = \frac{P}{H}$. The solution to this differential equation is written as

$$\frac{1}{-B} \log\left(\frac{A - BT_n}{A - BT_1}\right) = (t_n - t_1) \quad (38)$$

$$A - BT_n = (A - BT_1)e^{-B(t_n - t_1)} \quad (39)$$

Taking BT_n to the other side and subtracting BT_1 from both sides, we get

$$A - BT_1 = (A - BT_1)e^{-B(t_n - t_1)} + B(T_n - T_1) \quad (40)$$

$$B(T_n - T_1) = (A - BT_1)(1 - e^{-B(t_n - t_1)}) \quad (41)$$

Dividing both sides by B and substituting for A and B in the above equation, we get

$$T_n - T_1 = (\frac{S}{P} + T_s - T_1)(1 - e^{-\frac{P}{H}(t_n - t_1)}) \quad (42)$$

C. General Albedo, Flat Camera response

In applications with focus on shape or illumination, removing the effect of spatially varying albedo from the input image is a useful first step. In such cases, if we have a camera with a flat response across all wavelengths present in the illumination, we can derive a simple expression to directly compute the shading image.

Consider a monochrome camera with a constant spectral response such that $\Gamma(\lambda) = \Gamma_0 \ \forall \lambda$. This simplies Eq. (18) from Sec. 4 to

$$I(\mathbf{p}) = \frac{\gamma}{\pi}\Gamma_0 \int_\lambda \rho(\mathbf{x}, \lambda) L(\mathbf{x}, \lambda) d\lambda. \quad (43)$$

Combining Eq. (22) from Sec. 4 and Eq. (43), we can write

$$L(\mathbf{x}) = \int_\lambda L(\mathbf{x}, \lambda) d\lambda = \frac{\pi I(\mathbf{p})}{\gamma \Gamma_0} + \frac{\tilde{S}(\mathbf{x})}{\beta}, \quad (44)$$

where $L(\mathbf{x})$ is the total scene irradiance across all λ. Note that $L(\mathbf{x})$ contains all the information about shape and illumination. We have shown that it can be computed independent of albedo from a single view without any assumption about shape or illumination.

D. Computing E_k, F and L

Let $\Gamma_b(\lambda), \Gamma_g(\lambda)$, and $\Gamma_r(\lambda)$ be the sensor response functions corresponding to the BGR channels in the camera and let $l(\lambda)$ be the emission spectrum of the white LEDs obtained from the technical datasheets (see Fig. 2c). The wavelengths of interest can be partitioned into $\Lambda = \Lambda_B \cup \Lambda_G \cup \Lambda_R$, where $\Lambda_B = [400\text{nm}, 530\text{nm}), \Lambda_G = [530\text{nm}, 620\text{nm})$, and $\Lambda_R = [620\text{nm}, 1100\text{nm})$. Note that we include wavelengths in near infrared as well in our definitions since the sensor response functions are non-zero at those wavelengths.

Since $l(\lambda)$ is known, we can directly compute L as

$$L = \int_\Lambda l(\lambda) d\lambda. \quad (45)$$

In our experiments, we use $\Phi_\rho(\lambda) = \{\tilde{\rho}_b(\lambda), \tilde{\rho}_g(\lambda), \tilde{\rho}_r(\lambda)\}$ as the basis set for representing albedo as a function of wavelength. These basis functions are defined as

$$\tilde{\rho}_b(\lambda) = \mathbb{I}[\lambda \in \Lambda_B], \quad (46)$$

$$\tilde{\rho}_g(\lambda) = \mathbb{I}[\lambda \in \Lambda_G], \quad (47)$$

$$\tilde{\rho}_b(\lambda) = \mathbb{I}[\lambda \in \Lambda_R], \quad (48)$$

where $\mathbb{I}[\cdot]$ is the indicator function. Let $\mathbf{a_x} = \begin{bmatrix}\rho_b & \rho_g & \rho_r\end{bmatrix}^T$ be the corresponding vector of coefficients.

Since our basis functions are made up of indicator functions, their effective role is to restrict the integration limits in the definition of $\mathbf{E}_k$ and $\mathbf{F}$. We can now define these vectors as

$$\mathbf{E}_b = \left[\int_{\Lambda_B} l(\lambda)\Gamma_b(\lambda)d\lambda \quad \int_{\Lambda_G} l(\lambda)\Gamma_b(\lambda)d\lambda \quad \int_{\Lambda_R} l(\lambda)\Gamma_b(\lambda)d\lambda\right] \quad (49)$$

$$\mathbf{E}_g = \left[\int_{\Lambda_B} l(\lambda)\Gamma_g(\lambda)d\lambda \quad \int_{\Lambda_G} l(\lambda)\Gamma_g(\lambda)d\lambda \quad \int_{\Lambda_R} l(\lambda)\Gamma_g(\lambda)d\lambda\right] \quad (50)$$

$$\mathbf{E}_r = \left[\int_{\Lambda_B} l(\lambda)\Gamma_r(\lambda)d\lambda \quad \int_{\Lambda_G} l(\lambda)\Gamma_r(\lambda)d\lambda \quad \int_{\Lambda_R} l(\lambda)\Gamma_r(\lambda)d\lambda\right] \quad (51)$$

$$\mathbf{F} = \left[\int_{\Lambda_B} l(\lambda)d\lambda \quad \int_{\Lambda_G} l(\lambda)d\lambda \quad \int_{\Lambda_R} l(\lambda)d\lambda\right] \quad (52)$$

E. Grayscale Separation Results

Figure 10 summarizes the albedo shading separation results for the four target scenes using the grayscale approximation. Recall that the analytical expressions for the grayscale approximation are simpler and do not require knowledge of the emission spectrum or the camera response. Yet, as seen in the results, the estimated shading is similar to that obtained using the system of equations for general albedo functions.

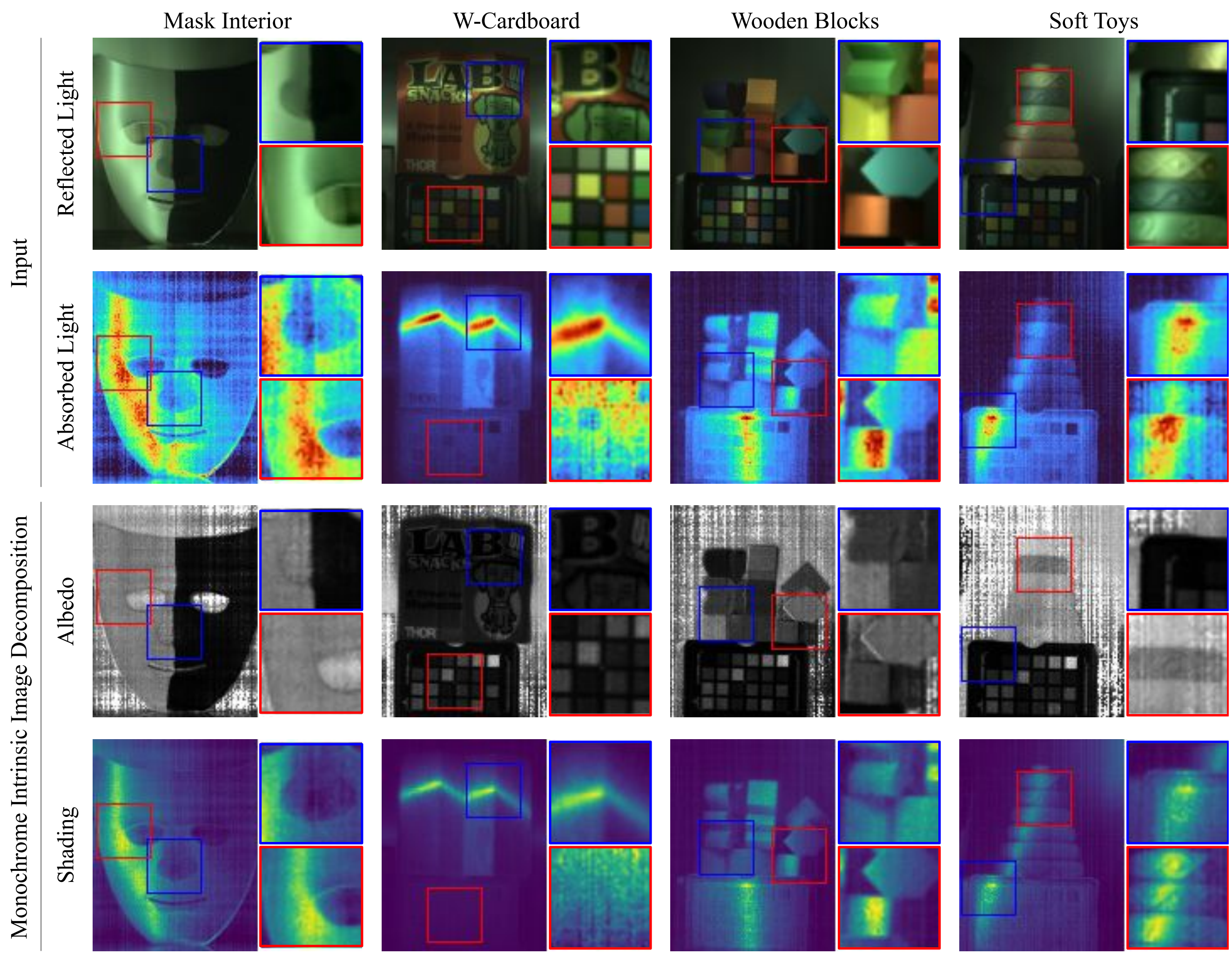


Figure 10. The first row shows the HDR visible image (brightened for visualization). Note that the colorchart is not an input to our method. The second row shows the estimated heat source intensity (turbo colormap) obtained using the method in Sec. 3. The last two rows correspond to using Eqs. (17) and Eqs. (16) respectively. The estimated albedo is clipped to the range $[0, 1]$. The callouts for the visible image, heat source intensity, and shading are normalized individually to aid visualization.