TheRF: Thermal Radiance Fields

Xin-Yi Pan

Abstract—Our work introduces a novel 3D thermal reconstruction method that uses images in the Long-Wave Infrared (LWIR) regime. By leveraging on information from both the visible and Infrared (IR) domains, sufficiently good thermal camera poses can be obtained, demonstrating significant improvements in 3D thermal reconstruction both qualitatively and quantitatively.

1 Introduction

Thermal imaging exposes features of our world that are invisible to the naked eye, and to RGB cameras. By capturing long-wave infrared (IR) light, in the 8-15 μm wavelength range, thermal cameras expose heat sources and material properties, and can see in the dark as well as through many occlusive media such as fog. These properties make thermal imaging a valuable tool in a range of applications including search and rescue, surveillance and defense [1], agricultural monitoring [2], and infrastructure inspection for safety and energy efficiency [3].

In many of these applications, multiple thermal images are collected from different viewpoints during the course of inspection or exploration. These applications stand to benefit from access to 3D thermal field reconstructions that combine these multi-view thermal images into a unified and consistent 3D thermal volume; this task of 3D thermal field recovery is the focus of our work.

Recently, 3D radiance field reconstruction has made great strides based on RGB images taken with visible light cameras. However, 3D reconstruction from thermal (long-wave IR) images remains challenging due to the low resolution of thermal cameras. In addition to limiting image quality, this low thermal resolution limits the number of robust 2D image features that can be used to recover thermal camera poses via structure from motion algorithms like COLMAP [4]. Further, even with known thermal camera poses, directly extending existing radiance field models to include a thermal channel produces limited quality 3D reconstructions because many materials interact differently with thermal versus visible light. For example, a glass of water is transparent to visible light but thermally opaque, while fog is visibly opaque but thermally transparent.

We propose strategies to address both of these challenges, recovering accurate thermal camera poses by calibrating the relative poses of a thermal camera and an RGB camera, and gracefully combining information from the two spectra while respecting and recovering material-specific properties. Our method also improves 3D thermal reconstruction quality by leveraging information from the visible spectrum for super-resolution, as RGB cameras are often of far higher spatial resolution compared to thermal cameras.

2 Related Work

2.1 Thermal Imaging

Thermal cameras detect and measure the heat signature of objects, and convert this emitted infrared (IR) energy into a thermal image reflecting varying levels of IR radiation [5]. Thermal images can provide insights into scenes and objects that are invisible to visible-light cameras [6]. The contactless nature of thermal imaging adds to its attractiveness in a diverse range of applications, including security and surveillance, preventive maintenance, building inspection, monitoring rock masses, and archaeology [6] [7] [8] [9] [10] [11].

Thermal cameras that measure long-wave infrared (LWIR) wavelengths (8 - 15 μm) use very expensive germanium lenses that transmit in the IR spectrum but block visible light [12] [13]. The longer wavelength also requires each element in the detector array to be larger than those required for the visible light spectrum [14]. These factors contribute to the significantly lower resolution [12] [15] and higher production cost of thermal cameras [16] compared to visible-light cameras. Considering the limited resolution available in thermal cameras, they are often paired with other sensors to improve the effective resolution beyond what the thermal camera alone can provide [17] [18].

In particular, 3D scene reconstruction from IR images is a compelling method to achieve such sensor fusion and enhanced resolution, alongside a volumetric scene representation that is of independent value in many applications. Prior work has proposed methods for 3D thermal imaging, such as contact maps from functional grasping [19], rather than non-contact camera measurements, and to employ thermal reflections for visualization rather than direct imaging [20]. Our focus is to produce high-fidelity 3D thermal reconstructions of diverse scenes by combining measurements from a low-resolution LWIR camera and a higher-resolution visible light camera, without any direct scene contact.

2.2 3D Reconstruction and Novel-View Synthesis

3D reconstruction is really useful as a way to combine information from multiple views into a richer model of a cohesive scene for many applications. Two main ways of doing this are a "classical" sparse feature matching approach like Structure from Motion (SfM), Simultaneous Localization and Mapping (SLAM), Multi-view Stereo (MVS) which usually produces a sparse point cloud, or with a dense

volumetric or surface reconstruction based on differentiable rendering.

SfM with MVS is a technique that reconstructs 3D surface models from 2D images. SfM computes projection matrices and 3D points using corresponding points in each view from 2D images [21], while MVS uses calibrated images for dense 3D reconstruction of scenes. Some limitations include the quality of the reconstructed scene being limited to that of the input images and camera parameters computed from SfM algorithms, as well as reconstruction assumptions such as scene rigidity [22] [23]. COLMAP is a specific instantiation of SfM that is used in nearly all radiance field models as a way to recover camera parameters [24] [25]. One limitation of COLMAP is the tendency to fail on images of low-resolution and scenes that lack textures [26].

Our work focuses on reconstructing radiance fields as a dense representation of a scene that leverages differentiable rendering instead of feature mapping. Many methods have been proposed for parameterizing a radiance field, including multilayer perceptrons [27] [28] [29] [30] [31], and great progress has been made in producing faithful and high-resolution radiance fields based on high-resolution visible light camera images.

Neural Radiance Field (NeRF) facilitates the modeling of scenes as implicit representations by projecting output colors and densities obtained from querying 5D coordinates comprising of 3D position coordinates and 2D viewing directions along camera rays [27]. Recent advances in NeRF-based implicit 3D reconstruction have achieved closer photo-realistic results to view-synthesis problems in aspects as such improving NeRF performance in representing finer details [32] [33] [34]. Additionally, a modular PyTorch framework, Nerfstudio, has been developed to facilitate the implementation of NeRF-based methods [35].

Most work that has been done focuses on the visible spectrum, which limits the insights that can be gleaned from reconstructed 3D scenes. Extending radiance field reconstruction to represent other parts of the spectrum which are invisible to the human eye would expose useful and otherwise inaccessible information [36].

2.3 Multispectral NeRF

Capturing data beyond the visible spectrum can be helpful in identifying features that might be transparent in the visible spectrum [37] [38] [39] [40] [41]. Integrated sensors have been proposed to extend visible light (RGB color) cameras to incorporate information beyond the visible spectrum.

For instance, X-NeRF [42] reconstructs 3D from multispectral images by optimizing a transform between RGB and other sensors including Near IR (NIR) cameras [43], while assuming known camera intrinsics. NIR images tend to have higher resolution as compared to long- and mid-wave IR (LWIR and MWIR) images due to its shorter wavelength [44] [45]. Despite this higher resolution training data of NIR compared to our LWIR setting, the novel views rendered from X-NeRF's 3D multispectral reconstruction remain relatively low-resolution, suggesting there is room for improved processing of IR images. Additionally, X-NeRF adopts the same assumption as vanilla NeRF: that each material absorbs different wavelengths of light equally, having a shared density parameter regardless of wavelength. While this assumption is a reasonable one for imaging a small range of wavelengths, it becomes problematic for sensor fusion across a wider spectrum. Additionally, although NIR based imaging allows for resolution similar to visible cameras and have been used in Night vision goggles and LIDAR [46], thermal imaging cameras detect LWIR and MWIR thermal emissions from objects [47] which carries essentially a different set of information hidden from our eyes and visible light cameras.

Consumer-grade thermal cameras are often lower cost and have poorer image resolution [10] [11]. This poses multiple challenges when to comes to both obtaining the camera poses and subsequent 3D reconstruction results [48] [42]. There is a need for a combination of both visible and Infrared (IR) wavelength ranges, with points from RGB augmented with thermal information, for higher accuracy reconstruction with sparse input images [2]. We propose a method for thermal 3D reconstruction of scenes using both RGB and thermal images. While have been approaches that used similar insights, such as dehazing [49], hyperspectral imaging [50] and 3D reconstruction of a person via reflections [51], we demonstrate a method for 3D thermal field reconstruction that separately models material interactions with thermal and visible spectra to improve reconstruction quality.

3 Proposed Method

Our method extends NeRF [27] to the combined RGBT (red, green, blue, and thermal/LWIR) domain. We begin by describing the main idea of our method as compared to standard visible-light RGB radiance fields, and then describing our implementation in more detail.

3.1 Extending Radiance Fields to a Wider Spectrum

Existing radiance field models, including NeRF and its many variations, typically focus on modeling radiance in the visible spectrum as three color channels (red, green, and blue). In using the standard volume rendering formula based on the Beer-Lambert Law (see section 3.2), these models implicitly assume that each point in space is equally absorptive of all three of these colors of light. While this is a good approximation for RGB visible light, to which most materials are either opaque or transparent, there are certain materials, such as stained glass, for which the approximation is no longer valid. For example, a red stained glass window transmits red light but occludes green and blue light, violating the equal-absorption assumption.

When we begin to consider radiance fields across a wider spectrum, including our setting of RGB and LWIR thermal radiance field modeling, we find that more materials exhibit differing absorption behavior across this wider spectrum. We model this behavior by explicitly endowing each spatial location with separate densities (absorption coefficients) for each wavelength, while introducing regularization to encourage these wavelength-specific densities to remain similar for most materials.

3.2 Image Formation Modeling

NeRF [27] in the traditional RGB setting represents a scene as a volumetric radiance field

$$F_\Theta^{\text{rgb}} : (\mathbf{x}, \mathbf{d}) \mapsto (\mathbf{c}_{\text{rgb}}, \sigma_{\text{rgb}})$$

mapping a 3D point $\mathbf{x} \in \mathbb{R}^3$ and viewing direction $\mathbf{d} \in \mathbb{R}^3$ to a volume density σ_{rgb} and view-dependent emitted color $\mathbf{c}_{\text{rgb}} = (r, g, b) \in \mathbb{R}^3$. The scene is rendered along a camera ray $\mathbf{r} = \mathbf{o} + t\mathbf{d}$ with origin $\mathbf{o} \in \mathbb{R}^3$ and direction $\mathbf{d} \in \mathbb{R}^3$ via standard volumetric rendering **[?]**

$$\mathbf{c}_{\text{rgb}}(\mathbf{r}) = \int_0^\infty T(t)\sigma_{\text{rgb}}(\mathbf{r}(t))\mathbf{c}_{\text{rgb}}(\mathbf{r}(t), \mathbf{d})dt \quad (1)$$

$$\text{where} \quad T(t) = \exp\left[\int_0^t -\sigma_{\text{rgb}}(\mathbf{r}(t'))dt'\right] \quad (2)$$

which is approximated numerically via N samples along the ray via

$$\mathbf{c}_{\text{rgb}}(\mathbf{r}) \approx \sum_{i=1}^{N} w_i \mathbf{c}_{\text{rgb}}(\mathbf{r}(t_i), \mathbf{d}) \quad (3)$$

$$\text{where} \quad w_i = T_i(1 - \exp(-\sigma_{\text{rgb}}(\mathbf{r}(t_i)))(t_{i+1} - t_i)) \quad (4)$$

$$\text{and} \quad T_i = \exp\left[-\sum_{j-1}^{i-1} \sigma_{\text{rgb}}(\mathbf{r}(t_i))(t_{i+1} - t_i)\right]. \quad (5)$$

Now, to extend NeRF from the RGB to the RGBT domain, given a ray $\mathbf{r}$, we wish to render the RGB color (r, g, b) plus the color from a thermal image τ. We treat this as 4-D color $\mathbf{c} = (r, g, b, \tau)$. Hence we introduce $c_{\text{therm}} := \tau \in \mathbb{R}$:

$$\mathbf{c}(\mathbf{r}) = (\mathbf{c}_{\text{rgb}}(\mathbf{r}), c_{\text{therm}}(\mathbf{r})) \quad (6)$$

Now, we observe that in the visible-light spectrum, objects of interest tend to absorb wavelengths of light similarly, but the same is not true in the combined infrared-and-visible-light spectrum. For example, many objects are opaque to visible light but transparent to infrared light, or vice versa, as illustrated in figures 2 and 3. Thus we propose to render $c_{\text{therm}}(\mathbf{r})$ with a separate density σ_{therm}, distinct from σ_{rgb}:

$$c_{\text{therm}}(\mathbf{r}) = \int_0^\infty T(t)\sigma_{\text{therm}}(\mathbf{r}(t))c_{\text{therm}}(\mathbf{r}(t), \mathbf{d})dt \quad (7)$$

$$\text{where} \quad T(t) = \exp\left[\int_0^t -\sigma_{\text{therm}}(\mathbf{r}(t'))dt'\right]. \quad (8)$$

We thus represent the RGBT scene as a radiance field

$$F_\Theta : (\mathbf{x}, \mathbf{d}) \mapsto (\mathbf{c}_{\text{rgb}}, c_{\text{therm}}, \sigma_{\text{rgb}}, \sigma_{\text{therm}}). \quad (9)$$

3.3 Optimization and Regularization

To optimize F_Θ, we minimize the following objective:

$$\mathcal{L} = \mathcal{L}_{\text{rgb}} + \mathcal{L}_{\text{therm}} + \lambda_{\text{density}}\mathcal{L}_{\text{density}} + \lambda_{\text{tv}}\mathcal{L}_{\text{tv}}. \quad (10)$$

Here $\mathcal{L}_{\text{rgb}}$ and $\mathcal{L}_{\text{therm}}$ are the standard pixel-wise color losses against calibrated ground-truth images

$$\mathcal{L}_{\text{rgb}} = \frac{1}{|\mathcal{R}|}\sum_{\mathbf{r}\in\mathcal{R}} \left|\mathbf{c}_{\text{rgb}}(\mathbf{r}) - \mathbf{c}^{\text{gt}}_{\text{rgb}}(\mathbf{r})\right| \quad (11)$$

$$\mathcal{L}_{\text{therm}} = \frac{1}{|\mathcal{R}|}\sum_{\mathbf{r}\in\mathcal{R}} \left|\mathbf{c}_{\text{therm}}(\mathbf{r}) - \mathbf{c}^{\text{gt}}_{\text{therm}}(\mathbf{r})\right|. \quad (12)$$

$\mathcal{L}_{\text{density}}$ is a regularizer encouraging the RGB and thermal densities to resemble each other

$$\mathcal{L}_{\text{density}} = \frac{1}{|\mathcal{X}|}\sum_{\mathbf{x}\in\mathcal{X}} \left\|\sigma_{\text{rgb}}(\mathbf{x}) - \sigma_{\text{therm}}(\mathbf{x})\right\|_1. \quad (13)$$

This regularizer is motivated by the observation illustrated in figure 3 that while some materials (like the glass bowl) do absorb visible and infrared light differently, most materials are similarly absorptive of both visible and infrared wavelengths. For those objects, we can leverage the information in the RGB channels to guide the thermal reconstruction and achieve thermal superresolution.

$\mathcal{L}_{\text{tv}}$ is a total variation regularizer on the RGB and thermal densities

$$\mathcal{L}_{\text{tv}} = \sum_{\sigma\in\{\sigma_{\text{rgb}},\sigma_{\text{therm}}\}} \frac{1}{|\mathcal{X}|}\sum_{\mathbf{x}\in\mathcal{X}} \sum_{\substack{i\in\{1,2,3\}\\ s\in\{-1,+1\}}} |\sigma(\mathbf{x}) - \sigma(\mathbf{x} + \delta s e_i)|. \quad (14)$$

where e_i are the standard basis vectors in $\mathbb{R}^3$ and δ is a sampling distance determined by the implicit voxel size of F_Θ – the maximum grid resolution of our Nerfacto-based model [35], see Section 4.2 for details.

4 Implementation

4.1 Camera Calibration

We calibrate both RGB and thermal cameras separately. For each camera, we find the focal lengths f_x, f_y, principal points c_x, c_y, and distortion parameters k_1, k_2, k_3, p_1, p_2. From these, we find the relative rolation and translation between both cameras.

$$s\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}f_x & 0 & c_x\\0 & f_y & c_y\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}R & T\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \quad (15)$$

$$x' = x(1 + k_1r^2 + k_2r^4 + k_3r^6) + 2p_1xy + p^2(r^2 + 2x^2) \quad (16)$$

$$y' = y(1 + k_1r^2 + k_2r^4 + k_3r^6) + p_1(r^2 + 2y^2) + 2p_2xy \quad (17)$$

4.2 Training

We use Nerfstudio, a modular PyTorch framework which provides a plug and play platform, to implement Neural Radiance Fields (NeRF) - based methods. In particular, we build our method on Nerfacto which promotes a good balance between training efficiency and performance [52].

5 Experimental Results

5.1 Our Method

We use a mix of qualitative images and quantitative metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) Index, and Learned Perceptual Image Patch Similarity (LPIPS) to gauge the performance of both RGB and thermal 3D reconstruction (see table 1. Here, we included rendered views from a mix of indoors and outdoor.

The indoor scenes are particularly interesting in showing information hidden in one spectrum but revealed in another. For instance, figure 2 shows a sheet that is opaque in the visible domain and transparent in the thermal domain, while figure 3 shows a bowl that is transparent in the visible domain but opaque in the thermal domain.

We use a 90:10 train-test-split for our datasets, and trained for 30k epochs with cross channel loss of $1e^{-6}$, density loss of $2.5e^{-5}$, thermal loss of 100. This combination of parameters have been found to give the best PSNR values for the Pyrex dataset which we use as the gauge.

5.2 Baseline Comparisons

In Table 1, we quantitatively compare our method to baseline methods for the case where images are contained within a RGB or thermal images. In figures 4, 5, we show qualitative comparisons to RGB baseline methods. In figures 6, 7, we show qualitative comparisons to thermal baseline methods. While our method has similar performance with the RGB baseline, our method performs significantly better as compared to the thermal baseline. This is consistent with the understanding that thermal images have poor resolution which makes finding their poses and doing 3D construction very challenging. This further supports the motivation of our project which is to come up with a model that produces better thermal reconstruction.

5.3 Comparison with other methods

In Table 1, we quantitatively compare our method to X-NeRF's cross-spectra method [42] and the values are reported in table 1. We note that our method generally performs better in terms of the PSNR and LPIPS, however, it is uncertain which method performs better in terms of the SSIM metric.

6 Discussion

Although our implementation is built on Nerfacto, our modifications to handle thermal radiance are applicable to any radiance field model. While we focus on LWIR thermal imaging, our modifications are likely applicable to other multispectral imaging settings in which the same material interacts differently with different wavelengths. For example, our method may be relevant to multi-energy X-ray computed tomography, in which different tissues absorb different X-ray wavelengths to varying degrees, or even to RGB radiance field modeling of certain materials that absorb red, green, and blue visible wavelengths differently, such as stained glass.

While our approach uses RGB information obtained in the presence of a source of light, our work demonstrates the possibility of 3D thermal scene reconstruction from long-wave IR (LWIR) images. This in itself holds much information that might be invisible to the naked eye or our usual visible light cameras. This represents a significant step forward in 3D thermal scene reconstruction, with much potential for wide-ranging applications and possibilities. One future direction could be to extend our method to more challenging scenes and applications, such as larger scale inspection facilitated by higher resolution thermal cameras mounted on drones. These advances hold the potential to enable and empower us to understand more about the world around us.

7 Conclusion

We present a novel method that addresses critical challenges of 3D thermal reconstruction. By recovering accurate thermal camera poses through inter-camera calibration, and by integrating information from both spectra while taking into account the wavelength-dependent material properties, we have achieved significant improvement in 3D thermal reconstruction both quantitatively and qualitatively.

Acknowledgments

This project was done with people outside of this class, and we gratefully appreciate the fruitful and inspiring discussions with Professor Gordon Wetzstein and the opportunity to work on this project that greatly interests us.

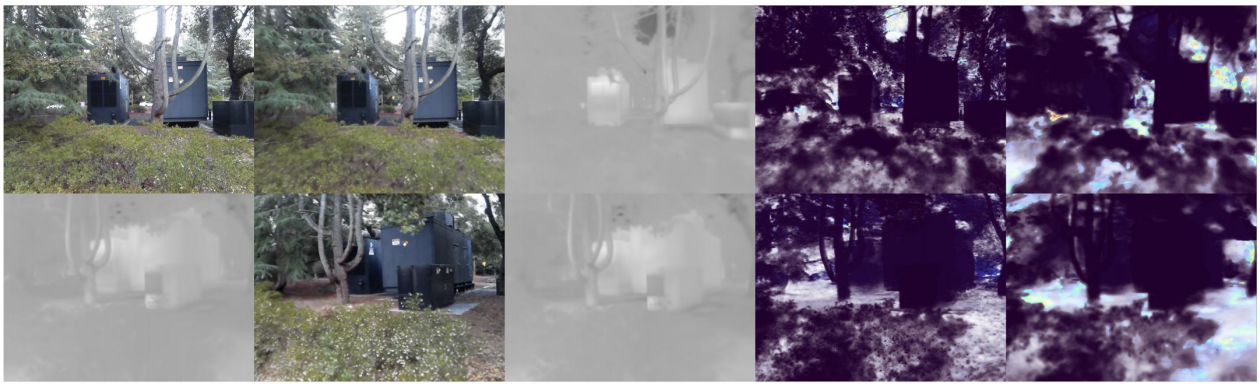

Fig. 1. Outdoor scene of a few generators: columns from left: ground truth (RGB (top row), thermal (bottom row), reconstructed RGB, reconstructed thermal, RGB depth, thermal depth. Images obtained using Weights & Biases (wandb) which displays only either of the ground truths in the first column.

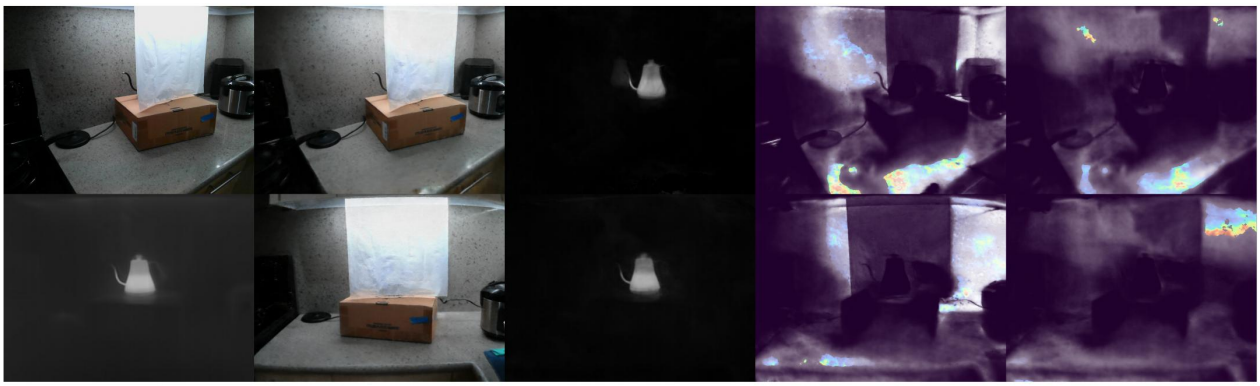

Fig. 2. Indoor scene of a hot kettle placed behind a sheet that is opaque in visible domain but transparent in thermal domain: columns from left: ground truth (RGB (top row), thermal (bottom row), reconstructed RGB, reconstructed thermal, RGB depth, thermal depth. Images obtained using Weights & Biases (wandb) which displays only either of the ground truths in the first column.

Fig. 3. Indoor scene of an ice pack in a bowl that is transparent in visible domain but opaque in thermal domain: columns from left: ground truth (RGB (top row), thermal (bottom row), reconstructed RGB, reconstructed thermal, RGB depth, thermal depth. Images obtained using Weights & Biases (wandb) which displays only either of the ground truths in the first column.

REFERENCES

[1] "Search and rescue— teledyne flir." [Online]. Available: https://www.flir.com/surveillance/search-and-rescue/

[2] J. M. Jurado, A. López, L. Pádua, and J. J. Sousa, "Remote sensing image fusion on 3d scenarios: A review of applications for agriculture and forestry," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, p. 102856, 2022.

[3] "Thermal imaging inspection checklist." [Online]. Available: https://www.fluke.com/en-us/learn/blog/thermal-imaging/thermal-imaging-inspection

[4] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4104–4113.

[5] Fluke, "What is thermal imaging? how a thermal image is

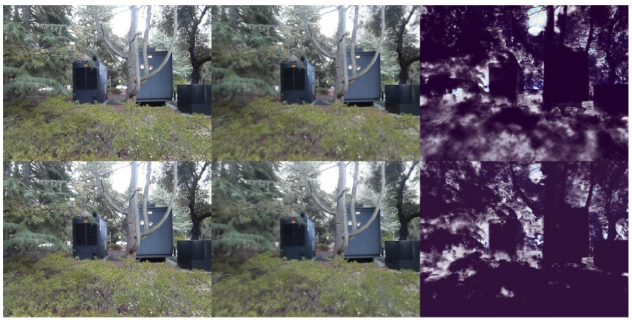

Fig. 4. Outdoor scene of a few generators with our model (top) and RGB baseline model (bottom) - columns from left: groundtruth, reconstructed RGB, RGB depth.

Fig. 5. Indoor scene of a hot kettle placed behind a sheet that is opaque in visible domain but transparent in thermal domain with our model (top) and RGB baseline model (bottom) - columns from left: ground truth, reconstructed RGB, RGB depth.

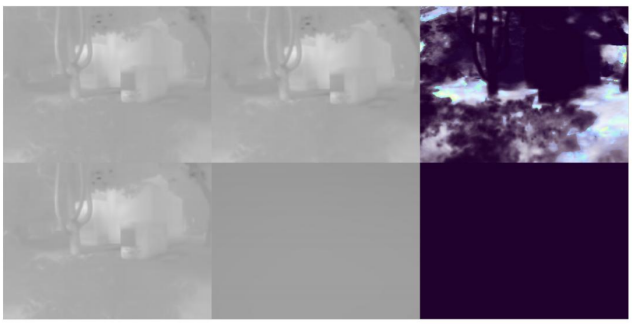

Fig. 6. Outdoor scene of a few generators with our model (top) and thermal baseline model (bottom) - columns from left: ground truth, reconstructed thermal, thermal depth.

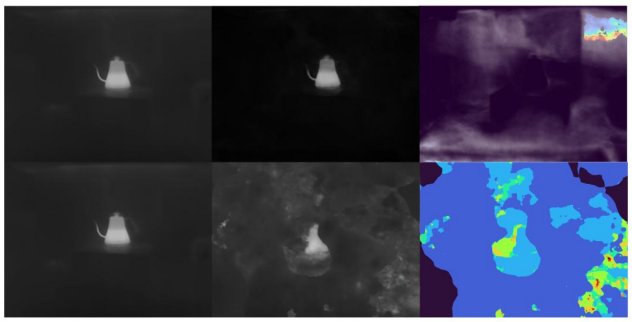

Fig. 7. Indoor scene of a hot kettle placed behind a sheet that is opaque in visible domain but transparent in thermal domain with our model (top) and thermal baseline model (bottom) - columns from left: ground truth, reconstructed thermal, thermal depth.

TABLE 1
Comparisons

Method			pyrex	sheet	heater	generators
Ours	PSNR(↑)	RGB	22.30	23.79	22.70	18.58
		Thermal	37.77	21.11	35.35	32.83
	SSIM(↑)	RGB	0.715	0.765	0.775	0.419
		Thermal	0.994	0.542	0.986	0.971
	LPIPS(↓)	RGB	0.437	0.382	0.407	0.574
		Thermal	0.015	0.121	0.012	0.039
RGB only	PSNR(↑)	RGB	21.30	24.33	24.31	18.15
		Thermal	-	-	-	-
	SSIM(↑)	RGB	0.689	0.780	0.805	0.389
		Thermal	-	-	-	-
	LPIPS(↓)	RGB	0.449	0.365	0.370	0.597
		Thermal	-	-	-	-
Thermal only	PSNR(↑)	RGB	-	-	-	-
		Thermal	22.04	19.46	19.60	21.69
	SSIM(↑)	RGB	-	-	-	-
		Thermal	0.970	0.781	0.958	0.883
	LPIPS(↓)	RGB	-	-	-	-
		Thermal	0.096	0.185	0.179	0.153
X-NeRF	PSNR(↑)	RGB	19.11	15.97	18.89	15.56
		Thermal	30.10	21.93	27.53	27.08
	SSIM(↑)	RGB	0.661	0.622	0.679	0.472
		Thermal	0.997	0.927	0.962	0.998
	LPIPS(↓)	RGB	0.535	0.490	0.378	0.700
		Thermal	0.033	0.064	0.053	0.061

captured." [Online]. Available: https://www.fluke.com/en-us/learn/blog/thermal-imaging/how-infrared-cameras-work

[6] ——, "What is thermal imaging? thermal cameras and how they work," Jan 2024. [Online]. Available: https://www.fluke.com/en-us/learn/blog/thermal-imaging/how-infrared-cameras-work

[7] J. Casana, A. Wiewel, A. Cool, A. C. Hill, K. D. Fisher, and E. J. Laugier, "Archaeological aerial thermography in theory and practice," *Advances in Archaeological Practice*, vol. 5, no. 4, pp. 310–327, 2017.

[8] C. Brooke, "Thermal imaging for the archaeological investigation of historic buildings," *Remote Sensing*, vol. 10, no. 9, p. 1401, 2018.

[9] A. Adán, B. Quintana, J. Garcia Aguilar, V. Pérez, and F. J. Castilla, "Towards the use of 3d thermal models in constructions," *Sustainability*, vol. 12, no. 20, p. 8521, 2020.

[10] G. Grechi, M. Fiorucci, G. M. Marmoni, and S. Martino, "3d thermal monitoring of jointed rock masses through infrared thermography and photogrammetry," *Remote Sensing*, vol. 13, no. 5, p. 957, 2021.

[11] O. González, M. I. Lizarraga, S. Karaman, and J. Salas, "Thermal radiation dynamics of soil surfaces with unmanned aerial systems," in *Pattern Recognition: 11th Mexican Conference, MCPR 2019, Querétaro, Mexico, June 26–29, 2019, Proceedings 11*. Springer, 2019, pp. 183–192.

[12] F. Nilsson, *Intelligent network video: Understanding modern video surveillance systems*. crc Press, 2008.

[13] B. Mesnik, "Thermal versus optical ip cameras." [Online]. Available: https://kintronics.com/thermal-versus-optical-ip-cameras/

[14] "Comparing sensitivity of thermal imaging camera modules." [Online]. Available: https://www.flir.com/discover/cores-components/Comparing-Sensitivity-of-Thermal-Imaging-Cameras-Modules/

[15] F. L. Liu, *Single-Shot 3D Microscopy: Optics and Algorithms Co-Design*. University of California, Berkeley, 2022.

[16] R. Schmidt, "How patent-pending technology blends thermal and visible light." [Online]. Available: https://www.fluke.com/en-us/learn/blog/thermal-imaging/how-patent-pending-technology-blends-thermal-and-visible-light

[17] "Flir one pro thermal imaging camera for smartphones — teledyne flir." [Online]. Available: https://www.flir.com/products/flir-one-pro/?vertical=condition%2Bmonitoring&segment=solutions

[18] FLIR, "Flir one® series thermal imaging cameras for ios® or android™ smartphones." [Online]. Available: https://www.flir.com/flir-one/

[19] S. Brahmbhatt, C. Ham, C. C. Kemp, and J. Hays, "Contactdb: Analyzing and predicting grasp contact via thermal imaging," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8709–8719.

[20] T. Maeda, Y. Wang, R. Raskar, and A. Kadambi, "Thermal non-line-of-sight imaging," in *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2019, pp. 1–11.

[21] D. Robertson and R. Cipolla, "Practical image processing and computer vision," in *chapter Structure from Motion*. John Wiley & Sons Australia, 2009.

[22] Y. Furukawa, C. Hernández *et al.*, "Multi-view stereo: A tutorial," *Foundations and Trends® in Computer Graphics and Vision*, vol. 9, no. 1-2, pp. 1–148, 2015.

[23] S. Wang, H. Jiang, and L. Xiang, "Ct-mvsnet: Efficient multi-view stereo with cross-scale transformer," in *International Conference on Multimedia Modeling*. Springer, 2024, pp. 394–408.

[24] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm, "Pixelwise view selection for unstructured multi-view stereo," in *European Conference on Computer Vision (ECCV)*, 2016.

[25] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[26] Z. Cheng, C. Esteves, V. Jampani, A. Kar, S. Maji, and A. Makadia, "Lu-nerf: Scene and pose estimation by synchronizing local unposed nerfs," *arXiv preprint arXiv:2306.05410*, 2023.

[27] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[28] Sara Fridovich-Keil and Alex Yu, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa, "Plenoxels: Radiance fields without neural networks," in *CVPR*, 2022.

[29] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, "Tensorf: Tensorial radiance fields," 2022.

[30] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Trans. Graph.*, vol. 41, no. 4, pp. 102:1–102:15, Jul. 2022. [Online]. Available: https://doi.org/10.1145/3528223.3530127

[31] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, July 2023. [Online]. Available: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

[32] M. Suhail, C. Esteves, L. Sigal, and A. Makadia, "Light field neural rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8269–8279.

[33] K. Gao, Y. Gao, H. He, D. Lu, L. Xu, and J. Li, "Nerf: Neural radiance field in 3d vision, a comprehensive review," *arXiv preprint arXiv:2210.00379*, 2022.

[34] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5855–5864.

[35] M. Tancik, E. Weber, E. Ng, R. Li, B. Yi, J. Kerr, T. Wang, A. Kristoffersen, J. Austin, K. Salahi, A. Ahuja, D. McAllister, and A. Kanazawa, "Nerfstudio: A modular framework for neural radiance field development," in *ACM SIGGRAPH 2023 Conference Proceedings*, ser. SIGGRAPH '23, 2023.

[36] H. Zhu, Y. Sun, C. Liu, L. Xia, J. Luo, N. Qiao, R. Nevatia, and C.-H. Kuo, "Multimodal neural radiance field," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 9393–9399.

[37] Y. Zhang, S. Müller, B. Stephan, H.-M. Gross, and G. Notni, "Point cloud hand–object segmentation using multimodal imaging with thermal and color data for safe robotic object handover," *Sensors*, vol. 21, no. 16, p. 5676, 2021.

[38] J. Á. S. Carmona, E. Quirós, V. Mayoral, and C. Charro, "Assessing the potential of multispectral and thermal uav imagery from archaeological sites. a case study from the iron age hillfort of villasviejas del tamuja (cáceres, spain)," *Journal of Archaeological Science: Reports*, vol. 31, p. 102312, 2020.

[39] M. McLeester, J. Casana, M. R. Schurr, A. C. Hill, and J. H. Wheeler III, "Detecting prehistoric landscape features using thermal, multispectral, and historical imagery analysis at midewin national tallgrass prairie, illinois," *Journal of Archaeological Science: Reports*, vol. 21, pp. 450–459, 2018.

[40] G. Patrucco, A. Gómez, A. Adineh, M. Rahrig, and J. L. Lerma, "3d data fusion for historical analyses of heritage buildings using thermal images: The palacio de colomina as a case study," *Remote Sensing*, vol. 14, no. 22, p. 5699, 2022.

[41] N. Sutherland, S. Marsh, G. Priestnall, P. Bryan, and J. Mills, "Infrared thermography and 3d-data fusion for architectural heritage: A scoping review," *Remote Sensing*, vol. 15, no. 9, p. 2422, 2023.

[42] M. Poggi, P. Zama Ramirez, F. Tosi, S. Salti, L. Di Stefano, and S. Mattoccia, "Cross-spectral neural radiance fields," in *International Conference on 3D Vision*, 2022, 3DV.

[43] Microsoft, "Azure kinect dk depth camera." [Online]. Available: https://learn.microsoft.com/en-us/azure/kinect-dk/depth-camera

[44] J. Oncea, "Swir, mwir, and lwir: One use case for each." [Online]. Available: https://www.photonicsonline.com/doc/swir-mwir-and-lwir-one-use-case-for-each-0001

[45] I. Electro-Optics, "Nir (near-infrared imaging (fog/haze filter)." [Online]. Available: https://www.infinitioptics.com/technology/nir-near-infrared

[46] O. Optical, "Optical material: Infrared optics." [Online]. Available: https://www.emf-corp.com/optical-materials/optical-material-infrared-optics/

[47] A. Optics, "The differences between swir, mwir, and lwir cameras." [Online]. Available: https://www.axiomoptics.com/blog/the-differences-between-swir-mwir-and-lwir-cameras/

[48] A. López, J. M. Jurado, C. J. Ogayar, and F. R. Feito, "An optimized approach for generating dense thermal point clouds from uav-imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 182, pp. 78–95, 2021.

[49] F. Dümbgen, M. El Helou, N. Gucevska, and S. Süsstrunk, "Near-infrared fusion for photorealistic image dehazing," Tech. Rep., 2018.

[50] S. Hu, R. Hou, L. Ming, S. Meifang, and P. Chen, "A hyperspectral image reconstruction algorithm based on rgb image using multi-scale atrous residual convolution network," *Frontiers in Marine Science*, vol. 9, p. 1006452, 2023.

[51] R. Liu and C. Vondrick, "Humans as light bulbs: 3d human reconstruction from thermal reflection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 12 531–12 542.

[52] M. Tancik, E. Weber, E. Ng, R. Li, B. Yi, T. Wang, A. Kristoffersen, J. Austin, K. Salahi, A. Ahuja *et al.*, "Nerfstudio: A modular framework for neural radiance field development," in *ACM SIGGRAPH 2023 Conference Proceedings*, 2023, pp. 1–12.