TRNeRF: Restoring Blurry, Rolling Shutter, and Noisy Thermal Images with Neural Radiance Fields
Supplementary Material

Spencer Carmichael[1][†]   Manohar Bhat[1]   Mani Ramanagopal[2]   Austin Buchan[1]
Ram Vasudevan[1]   Katherine A. Skinner[1]
[1]University of Michigan, Ann Arbor, MI USA   [2]Carnegie Mellon University, Pittsburgh, PA USA

In this supplementary material, we provide additional details on our data collection (Sec. A), evaluation methods (Sec. B), and baseline implementations (Sec. C). We also provide additional qualitative comparisons (Sec. D), which show the results of Instant-NGP [11] and a thermal deblurring method [14].

A. Dataset Collection Details

A.1. Hardware and Configuration

In our data collection rig, the backpack mounted hardware includes a portable power supply, Intel NUC computer, and an Arty Z7 FPGA development board connected to the cameras through a custom adapter PCB. The sensor platform includes two microbolometer thermal cameras (FLIR ADK), two monochrome cameras (FLIR Blackfly S GigE), and an IMU (VectorNav VN-100). We provide the IMU data to support future work, but it is not utilized in this paper.

The FLIR ADK thermal cameras (part number 40640U050-6PAAX) have a 640×512 resolution [7], $50° \times 40°$ field-of-view [7], a 1 meter near-focus distance [1], and an 8 ms thermal time constant [7]. The cameras read out the first pixel 0.5 ms after a trigger signal is received and take 27.8 μs to read out a single row[1] [5]. The two thermal cameras are placed side-by-side to support the comparison of different camera settings with minimal parallax, yielding a baseline of 3.5 cm between them. We operate the thermal cameras in high gain mode. In their default configuration, the thermal cameras employ a number of onboard post-processing stages to reduce noise. This includes factory-calibrated corrections, automatically per-

formed non-uniformity corrections (NUCs), and image filters.

The factory-calibrated corrections compensate for pixel-wise offset variation, pixel-wise responsivity variation stemming from the sensor and lens assembly, heat radiating from surfaces inside the camera assembly, and defective pixels [7]. We keep all of these factory-calibrated corrections enabled. Some of the factory-calibrated corrections are dependent on the camera's operating temperature, and the correction tables are automatically switched as the operating temperature changes [7]. Moreover, some of the factory-calibrated corrections are optimized for steady-state conditions and thermal shock (rapidly changing operating temperature) is to be avoided [4]. For these reasons, we allow time for the cameras' operating temperatures to equilibrate as we bring them into new environments.

The NUCs involve a shutter that is closed to present a uniform thermal signal to the camera, allowing non-uniformity to be identified [7]. The NUCs aid in correcting residual fixed pattern noise (FPN), but freeze the camera output for ∼0.5 seconds and cause abrupt changes to applied corrections. Therefore, we disable automatic NUCs, and perform a single manual NUC before recording the three sequences (slow, medium, and fast) in each scene.

The image filters target column and row correlated noise, random spatial noise, and temporal noise [7]. The impact of these filters to training neural radiance fields (NeRFs) has not been explored in prior literature. We enable these filters in the left thermal camera and disable them in the right to allow a comparison to be made.

The FLIR Blackfly S GigE monochrome cameras (part number BFS-PGE-16S2M) are paired with Computar lenses (part number A4Z2812CS-MPIR). The cameras have a 1440×1080 resolution and their lenses are adjusted to provide an approximate $70° \times 55°$ field-of-view. The monochrome cameras are spaced 28 cm apart to support stereo visual structure from motion as described in the main text of the paper. Auto-exposure and auto-gain are enabled,

---

[1]FLIR provides contradictory information regarding this timing. Another source [6] states that the first pixel is read out 0.3 ms after the trigger signal and that it takes 32 μs to read out a single row. Both sets of numbers yield similar results.

such that exposure is increased first, up to a maximum of 1 ms, and then gain is increased, up to a maximum of 18 dB. The small maximum exposure time ensures sharp images, even during fast camera motion.

A.2. Time Synchronization

We use the FPGA to send a 60 Hz square wave signal to each sensor. The cameras capture an image on each falling edge. The IMU is configured to output data at a 400 Hz rate. Included in the IMU data is a count of the falling edges received on the square wave signal and the time since the last falling edge.

To synchronize all of the data in post-processing, we follow a strategy similar to that outlined in [3]. Specifically, we apply a blanking period to the square wave signal at the start of each recorded sequence. The blanking period can be identified in the recorded data, providing an association across all sensors.

This association allows the first images captured after the blanking period to be aligned. Next, we identify all other simultaneously triggered images by counting forward while accounting for dropped frames. The timestamps of the left monochrome images can then be assigned to the simultaneously triggered images across the other cameras. In practice, however, we found that this results in a 1 frame offset between the thermal and monochrome cameras. We were able to verify this offset through multiple means (including TRNeRF's performance, and reprojection errors of the Aprilgrid corners), but could not identify the cause. The problem is easily resolved by associating the *second* image captured after the blanking period in each thermal camera with the first image captured after the blanking period in each monochrome camera. We have applied this correction in the published data.

Similarly, the blanking period can be identified in the IMU data and each IMU measurement can be associated with the most recently captured image. The timestamp of an IMU measurement is then determined by adding the time since the last falling edge to the timestamp of the last captured left monochrome image. Occasionally, the IMU began to respond inconsistently to falling edges midway through a recorded sequence. In such cases, we linearly extrapolate the later IMU measurement timestamps from the timestamps assigned prior to the inconsistency. Additionally, through the calibration process described in the next section we identified a 4 ms offset between the IMU and camera timestamps. This offset has been corrected in the published data.

A.3. Calibration

We follow the approach described in [3] to calibrate the cameras. Specifically, we use an Aprilgrid board [8, 9, 12] that is constructed of aluminum and vinyl in order to ap-

pear in both the visible and thermal spectra. After applying a heat gun to the board, we record a sequence of the board being moved slowly in front of the cameras. We use Kalibr [8] to process the data and solve for the intrinsics and extrinsics of all four cameras. To adapt the thermal data for Kalibr's AprilTag detection, we convert the thermal images to 8-bit, as described in the main text, and invert them. We also apply Contrast Limited Adaptive Histogram Equalization (CLAHE) to the monochrome images to mitigate the impact of reflections on AprilTag detection.

Additionally, to calibrate the IMU intrinsics and extrinsics, we begin by recording 18 hours of IMU data with the IMU resting on a damped surface. We process this data with [2] to estimate the IMU noise parameters. We then recorded a sequence of the static board with six degree-of-freedom camera motion. We finally use Kalibr [8], with the previously obtained calibration results, to process the IMU and monochrome data in this sequence to determine the IMU extrinsics.

A.4. Two-Point NUCs

As noted in the main text, we use two-point NUCs to aid in pseudo-ground truth generation. A two-point NUC [13] is a procedure that utilizes hot and cold uniform temperature sources to estimate FPN modeled as constant pixel-wise gains $g_{u',v'}$ and offsets $o_{u',v'}$:

$$n_{u',v',i} = g_{u',v'} m_{u',v',i} + o_{u',v'} \tag{1}$$

When imaging the hot and cold uniform temperature sources, the noise-free pixel value $m_{u',v',i}$ is the same across all pixels, and therefore Eq. (1) can be written as:

$$n_{u',v',h} = g_{u',v'} m_h + o_{u',v'} \tag{2}$$
$$n_{u',v',c} = g_{u',v'} m_c + o_{u',v'} \tag{3}$$

where h and c denote images of the hot and cold sources, respectively. We assume that the mean pixel value of these images is equal to the noise-free value, that is:

$$\bar{n}_h = m_h \tag{4}$$
$$\bar{n}_c = m_c \tag{5}$$

The pixel-wise gains and offsets can then be solved for as follows:

$$n_{u',v',h} - n_{u',v',c} = g_{u',v'} (\bar{n}_h - \bar{n}_c) \tag{6}$$
$$g_{u',v'} = \frac{n_{u',v',h} - n_{u',v',c}}{\bar{n}_h - \bar{n}_c} \tag{7}$$
$$o_{u',v'} = n_{u',v',c} - g_{u',v'} \bar{n}_c \tag{8}$$

and the FPN can then be removed from images by inverting Eq. (1):

$$m_{u',v',i} = \frac{n_{u',v',i} - o_{u',v'}}{g_{u',v'}} \tag{9}$$

Figure 1. Approximate hot (top) and cold (bottom) uniform temperature sources used to compute the two-point NUC.

(a) Per-pixel gain

(b) Per-pixel offset

(c) Raw image
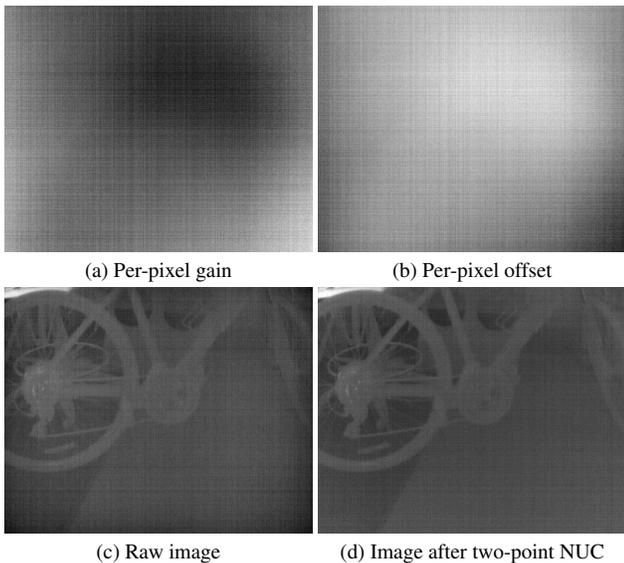
(d) Image after two-point NUC

Figure 2. The results of performing a two-point NUC in the slow indoor sequence with the right thermal camera. Specifically, this shows the computed gains and offsets and an example of applying them to a raw image.

In practice, it is best to record multiple images of each source to mitigate the impact of temporal noise. Then $\bar{n}_h$ and $\bar{n}_c$ are computed as the mean pixel value across all the images and $n_{u',v',h}$ and $n_{u',v',c}$ are computed as the pixel-wise means.

In our dataset, we create approximate uniform temperature sources by adhering squares of vinyl to aluminum sheets. The aluminum's high thermal conductivity distributes heat evenly and the vinyl's high emissivity ensures the surface is not reflective in the thermal spectrum. In the indoor scene, we place one aluminum sheet over a heated blanket and the other over a bag of ice, as pictured in Fig. 1.

	Outdoor		Indoor	
	Linear Speed (m/s)	Angular Speed (deg/s)	Linear Speed (m/s)	Angular Speed (deg/s)
Slow	0.18	8	0.16	9
Medium	0.91	33	0.85	37
Fast	2.81	120	2.84	148

Table 1. Average linear and angular speeds in the subsets of each sequence used in our experiments

In the outdoor scene, we place one sheet in sunlight and the other in shade. The surfaces were not perfectly uniform in temperature due to defects in the vinyl and condensation that formed on the cold source indoors. We translate and rotate the cameras when recording the uniform sources to mitigate this non-uniformity by averaging it out.

We record the uniform sources before and after recording the scene in each sequence. To aid pseudo-ground truth generation, we separately compute the two-point NUC with both sets of data and choose the results that qualitatively produce the best noise removal, as shown in Fig. 2.

A.5. Data Collection Procedure

To record the three sequences (slow, medium, and fast) in each scene, we use the following procedure:

1. Power on the thermal cameras, and allow time for them to reach thermal equilibrium and run automatic NUCs.

2. Disable automatic NUCs and manually execute a final NUC

3. Then, for each sequence:

   (a) (Indoor) Apply a heat gun to the calibration board.

   (b) Begin recording data.

   (c) Execute the blanking period as described in Sec. A.2.

   (d) Record the uniform hot and cold temperature sources for approximately 20 seconds each, as described in Sec. A.4.

   (e) Record the scene (4 minutes for slow, 3 minutes for medium, and 2 minutes for fast). Each recording begins with moderate speed six degree-of-freedom movements intended to aid IMU initialization.

   (f) Record the uniform hot and cold temperature sources again.

   (g) Stop recording data.

Note that the board is heated to increase contrast and enable AprilTag detection. In the outdoor scene the board was instead placed in direct sunlight for this purpose (although
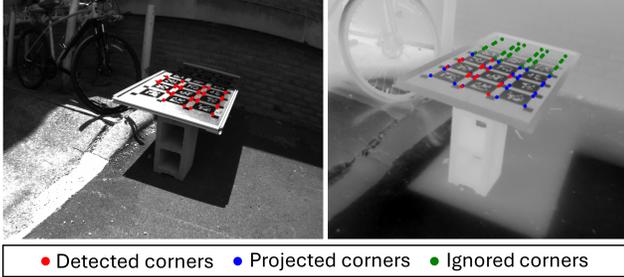
Figure 3. An example of computing the detection metric in the outdoor scene where the low contrast rows of the board are ignored.

it was later partially shaded, as explained in Sec. B). We also maintained a ≥ 1 meter distance between the cameras and the majority of the scene content to ensure all-in-focus images. Table 1 lists the average linear and angular speeds (of the left monochrome camera) in the subsets of each sequence used in our experiments.

B. Evaluation Details

There are some elements of the collected data that complicate the evaluation methods. For example, the conditions of the outdoor scene changed slowly between the recorded sequences. Most significantly, a shadow cast by an adjacent building traveled across the Aprilgrid board. This change in the scene is not reflected in the pseudo ground truth, which likely has a subtle impact on the LPIPS metric [16]. The shadow also cooled sections of the board, rendering the affected rows of the board undetectable. We therefore ignored these rows of the board in the evaluation, as shown in Fig. 3. Similarly, the heating of the board in between the recorded indoor sequences produces subtle changes in temperature. These subtle changes are seen between sequences, but also within individual sequences as the board cools.

As mentioned in the main text, to avoid using inaccurate pseudo ground truth images, we only evaluate against ones rendered at viewpoints sufficiently similar to the slow sequence training images. Specifically, for a given image from the medium or fast sequences, we evaluate it against the pseudo ground truth if its pose is close to at least one pose from the slow sequence trajectory. We define two poses as close if the relative translation and rotation between them are both beneath certain thresholds. In the fast sequences, we use a translation threshold of 19 cm and a rotation threshold of 9 degrees. In the medium sequences, we use a translation threshold of 10 cm and a rotation threshold of 10 degrees.

C. Baseline Implementation Details

USB-NeRF [10], provides the option to enable or disable trajectory optimization. We therefore tested it with

both and present the best results for each sequence. We found that trajectory optimization improved performance in the medium and fast sequences, and harmed performance in the slow sequences. We suspect that this is due to the large number of training images in the slow sequences (13.4k) leading to relatively few pose updates per image.

GSotM [15] requires an input point cloud for good performance. We use the point cloud obtained by applying COLMAP to the left monochrome images as described in the main text. We estimate the thermal values of the points by projecting them into the slow sequence images. GSotM is designed with several components that can be individually enabled [15]. We experimented with many combinations and obtained the best performance by enabling rolling shutter correction, blur correction, and velocity optimization, disabling pose optimization, and initializing the velocity estimates at zero. As GSotM assumes the photoelectric image formation model with a finite exposure time [15], we also tested it with the exposure time set to 0, 10, and 20 ms and found that 10 ms yielded the best results. To account for the large number of training images, we ran GSotM for 100k iterations (five times the default).

D. Additional Qualitative Comparisons

To our knowledge, the method in [14], which we will refer to as Pixel-Wise Deblurring, is the only prior thermal deblurring method with available code. In the main text, we mention that this method requires a high frame rate and was originally demonstrated with 200 Hz images. It is therefore not expected that it would perform well with 60 Hz data, but we include the results from running it on our dataset here for completeness. We run Pixel-Wise Deblurring with the default parameters in the provided codebase. We find that it takes around 2 hours to deblur a window of 33 images. We therefore limit our evaluation to a small number of qualitative results as shown in the rightmost column of Fig. 4. The results contain significant artifacts and are in some cases significantly shifted from the pseudo ground truth.

Fig. 4 also shows the qualitative results obtained by Instant-NGP without our modifications. Unsurprisingly, without accounting for any degradations, Instant-NGP is unable to learn a sharp scene.

References

[1] Boson+ Optical Specifications. https://www.flir.com/support-center/oem/boson-plus-optical-specs/. Accessed: 2024-06-01. 1

[2] Russell Buchanan. Allan variance ros, Nov. 2021. 2

[3] Spencer Carmichael, Austin Buchan, Mani Ramanagopal, Radhika Ravi, Ram Vasudevan, and Katherine A Skinner. Dataset and benchmark: Novel sensors for autonomous vehicle perception. *The International Journal of Robotics Research*, 2024. 2
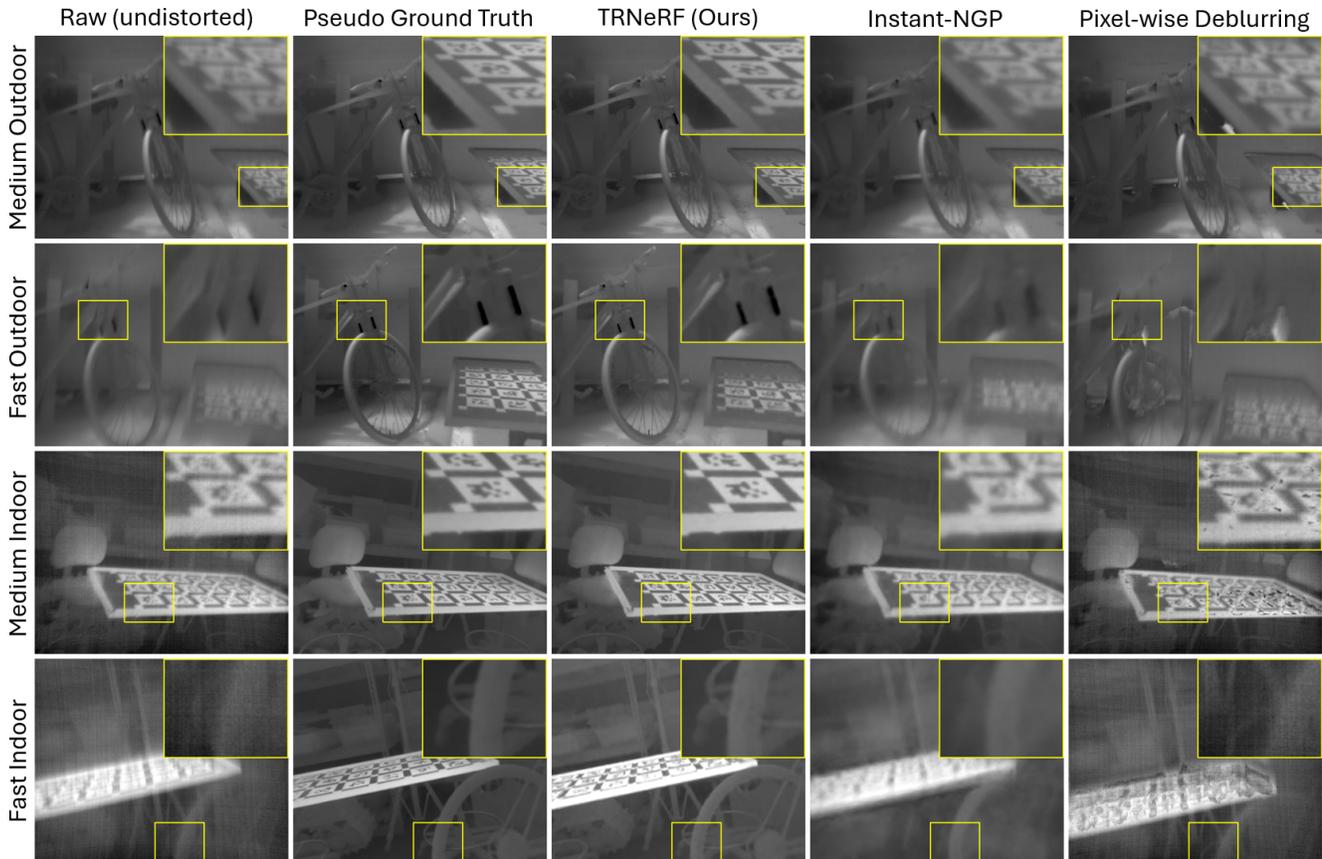
Figure 4. Additional qualitative results (Instant-NGP [11] and Pixel-Wise Deblurring [14]) on the medium and fast sequences from the indoor and outdoor scenes.

[4] FLIR. *Boson Image Non-Uniformities (INUs) Application Note*. 1

[5] FLIR. *FLIR Boson External Sync Application Note*. 1

[6] FLIR. *Getting Started guide for the FLIR ADK with USB, GMSL, and Ethernet Connector Options*. 1

[7] FLIR. *Teledyne FLIR Boson and Boson+ Thermal Imaging Core Product Data Sheet*. 1

[8] Paul Furgale, Joern Rehder, and Roland Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286, Nov. 2013. 2

[9] Jianzhu Huai, Yuxin Shao, Grzegorz Jozkow, Binliang Wang, Dezhong Chen, Yijia He, and Alper Yilmaz. Geometric wide-angle camera calibration: A review and comparative study. *Sensors*, 24(20), 2024. 2

[10] Moyang Li, Peng Wang, Lingzhe Zhao, Bangyan Liao, and Peidong Liu. USB-NeRF: Unrolling Shutter Bundle Adjusted Neural Radiance Fields. In *International Conference on Learning Representations*, 2024. 4

[11] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Transactions on Graphics*, 41(4):1–15, July 2022. 1, 5

[12] Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3400–3407, May 2011. 2

[13] David L. Perry. Linear theory of nonuniformity correction in infrared staring sensors. *Optical Engineering*, 32(8):1854, 1993. 2

[14] Manikandasriram Srinivasan Ramanagopal, Zixu Zhang, Ram Vasudevan, and Matthew Johnson Roberson. Pixel-Wise Motion Deblurring of Thermal Videos. In *Proceedings of Robotics: Science and Systems*, Corvalis, Oregon, USA, July 2020. 1, 4, 5

[15] Otto Seiskari, Jerry Ylilammi, Valtteri Kaatrasalo, Pekka Rantalankila, Matias Turkulainen, Juho Kannala, Esa Rahtu, and Arno Solin. Gaussian splatting on the move: Blur and rolling shutter compensation for natural camera motion. In *European Conference on Computer Vision*, pages 160–177, 2024. 4

[16] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 4