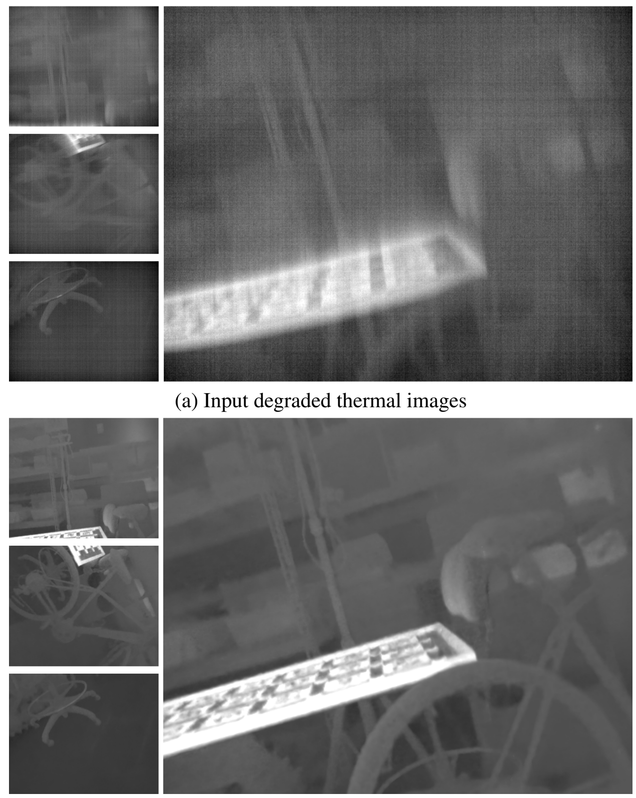
TRNeRF: Restoring Blurry, Rolling Shutter, and Noisy Thermal Images with Neural Radiance Fields

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

Abstract

Thermal cameras offer unique detection capabilities in building inspections, search and rescue operations, and autonomous vehicle perception. Of the different types of thermal cameras, uncooled microbolometers are often chosen due to their relative affordability, small size, and low power consumption. However, microbolometers suffer from motion blur, rolling shutter distortions, and fixed pattern noise, which limit the conditions of their use. Nearly all prior methods for microbolometer image restoration account for only one of these degradations, and current techniques addressing microbolometer blur and rolling shutter are limited. This paper presents TRNeRF, a thermal image restoration method that jointly addresses all three degradations by incorporating the microbolometer image formation model with Neural Radiance Fields (NeRFs). To evaluate TRNeRF, this paper introduces a new real-world dataset that is uniquely designed to support two novel quantitative evaluation strategies for thermal image restoration. Experiments demonstrate that TRNeRF is able to recover sharp, global shutter, and clear thermal images, even under extremely aggressive camera motion that causes existing methods to fail. The code and dataset are available at: `https://umautobots.github.io/trnerf`.



(a) Input degraded thermal images

(b) Output TRNeRF restored images

Figure 1. Given a sequence of degraded thermal images (a) and a known camera trajectory, TRNeRF trains an accurate thermal scene representation and restores the input images (b).

1. Introduction

Thermal cameras can detect heat loss and moisture in building inspections [6,46], highlight hidden fires, humans, and animals in search and rescue operations [17,41], and see pedestrians through fog in autonomous vehicle applications [12]. Cooled thermal cameras offer high image quality, but uncooled microbolometers are more affordable, smaller, and consume less power, making them appealing in many of these applications [6,52]. However, microbolometers suffer from motion blur [37], rolling shutter distortions [33], and noise [2], which greatly impact image quality under fast, or even moderate, camera motion and in environments with low thermal contrast.

In the existing literature on thermal image restoration, noise removal [2, 7, 8, 10, 16, 38, 50], motion deblurring [1, 15, 22, 32, 37, 45, 56], and rolling shutter correction [32, 33] are considered separately, with only one exception that accounts for both motion blur and rolling shutter [32]. Moreover, the current methods for deblurring and rolling shutter correction either ignore the unique blur formation model of

† Corresponding author: `specarmi@umich.edu`
This work was supported by a grant from Ford Motor Company via the Ford-UM Alliance under award N028603.

the microbolometer [1, 45, 56], or are limited to particular camera specifications and scenarios [22, 32, 33, 37]. Quantitative analysis with real microbolometer images is also lacking, with nearly all prior work on deblurring restricting quantitative results to synthetic images that are artificially degraded [1, 22, 45, 56].

Meanwhile, the recent popularity of Neural Radiance Fields (NeRFs) [28] and 3D Gaussian Splatting (3DGS) [13] has led to interest in methods that use these frameworks to restore visible spectrum images [4, 18–21, 23, 26, 27, 30, 34, 35, 42, 43, 48, 49, 55]. These methods train sharp NeRF and 3DGS scene representations directly from degraded images, allowing the degraded images to be restored by rendering them at the inference stage. While recent papers have begun to explore training NeRF models with thermal images [9, 24, 25, 51, 53, 57, 58], none have considered the unique image formation characteristics of microbolometers or been demonstrated with degraded images.

In this paper, we introduce Thermal Restoration NeRF (TRNeRF), a method that incorporates the microbolometer image formation model within a NeRF framework. Given a sequence of degraded thermal images and a known camera trajectory, our method trains a NeRF network that is able to render sharp, global shutter, and clear thermal images, as shown in Fig. 1. Specifically, inspired by existing work on NeRF-based restoration [20, 23, 48], we model the microbolometer's blur, rolling shutter, and fixed pattern noise in the rendering pipeline during training, and remove these modifications at the inference stage. To test our method's restoration performance on real degraded microbolometer images, we introduce a new dataset and two novel quantitative evaluation strategies that do not require ground truth restored images. We demonstrate that TRNeRF can successfully restore real degraded microbolometer images, even under extremely aggressive camera motion that causes existing methods to fail.

2. Related Work

2.1. Visible-Spectrum Image Restoration with NeRF and 3DGS

In recent years, Neural Radiance Fields (NeRF) [28] and 3D Gaussian Splatting (3DGS) [13] have emerged as highly effective methods for scene reconstruction and novel view synthesis and inspired a wave of research into improving these methods and applying them in various domains.

Most relevant to our paper, one line of research seeks to train NeRF and 3DGS from degraded images [4, 18–21, 23, 26, 27, 30, 34, 35, 42, 43, 48, 49, 55]. In general, NeRF and 3DGS train by rendering individual pixels or full images from the same perspectives as the input images to compute a loss. A common strategy to handle degraded images is to model the degradation in the rendering pipeline during training, such that the loss can be computed against the degraded inputs, and to remove it during inference, such that high quality images can be generated.

An early paper on this topic was Deblur-NeRF [27]. Deblur-NeRF handles defocus blur and motion blur by training a multi-layer perceptron (MLP) to output sparse, view-dependent, and pixelwise blur kernels [27]. During training, a single blurry pixel value is estimated by querying the MLP to obtain multiple weighted rays, which are individually rendered and then blended together [27]. This strategy has been further explored in several subsequent papers [18, 19, 21, 26, 34, 35]. Concurrent papers BAD-NeRF [48] and ExBluRF [20] demonstrated that when addressing motion blur alone, directly modeling the blur as a discrete integral of rendered pixel values over the exposure interval outperforms the blur kernel strategy. This technique has been improved [49], extended to dynamic scenes [43], and adapted for 3DGS [4, 30, 55]. In our method, we assume all-in-focus images and similarly approximate motion blur as a discrete integral over multiple rendered pixel values, but we adapt this to the microbolometer's unique blur formation model.

USB-NeRF accounts for rolling shutter readout by jointly optimizing a single continuous camera trajectory across all images [23]. This enables a pixel to be rendered using the pose at the precise time it was read out [23]. Gaussian Splatting on the Move [42] addresses both motion blur and rolling shutter together by jointly optimizing per-image camera poses and velocities and introducing a screen-space approximation of pixel-wise motion blur. We use a strategy similar to USB-NeRF to account for rolling shutter, but combine it with the motion blur compensation described above.

We select NeRF [28] over 3DGS [13] as the foundation of our method as NeRF's pixel-wise rendering lends itself directly to rolling shutter correction and, unlike 3DGS, NeRF does not require an input point cloud to obtain good performance. In addition to accounting for motion blur and rolling shutter, we also jointly optimize the thermal camera's fixed pattern noise.

2.2. Thermal NeRF & Thermal Image Restoration

Several methods have been recently proposed for training a NeRF with thermal images either exclusively [53], or with one or more additional modalities [9, 24, 25, 51, 57, 58]. However, prior work on thermal NeRFs has not considered the unique image formation characteristics of microbolometers or been demonstrated with degraded images.

The problem of noise removal in uncooled thermal images has received significant attention [2, 7, 8, 10, 16, 38, 50], but it has not been considered together with motion blur and rolling shutter distortions. Most similar to our method with respect to noise removal, DeepIR jointly optimizes fixed pattern noise and a neural representation of noise-free thermal images using multiple thermal images with small motions [38]. The method assumes all images are related to the first frame using a simple geometric transform, which is not suitable for large motions. While our method also uses a neural representation, we model motion-induced degradations in addition to noise.

Prior work has also addressed thermal deblurring [1, 15, 22, 32, 37, 45, 56]. The methods most similar to ours are those that also consider the unique blur formation model of the microbolometer, however, they are currently limited to particular camera specifications and scenarios [22, 32, 37]. For example, the method in [37] requires a high frame rate (200 Hz in their experiments) not typically supported by low-cost microbolometers, and [22] assumes purely rotational camera motion. Notably, nearly all existing papers on thermal deblurring also restrict their quantitative evaluation to artificially degraded images originally captured with a cooled thermal camera or slow moving uncooled thermal camera [1, 15, 22, 45, 56]. Additionally, to our knowledge, only two papers have investigated rolling shutter correction with microbolometers [32, 33]. The algorithm in [33] is designed only for a stationary camera observing a static object as it cools. A technique in [32] considers rolling shutter and motion blur, but it is limited to objects against a constant background moving in a straight line with known pixel velocity.

We propose a thermal restoration method that jointly accounts for motion blur, rolling shutter, and noise. Moreover, we introduce a new dataset to test our method exclusively on real degraded microbolometer images and introduce two novel evaluation schemes that do not require ground truth restored images.

3. Preliminary

3.1. Microbolometer Image Formation Model

Let the vector $\underline{\mathbf{x}}_w = [x_w, y_w, z_w, 1]^T$ be the *homogeneous* (denoted by the underline) world frame coordinates of an object in a scene. Given an ideal pinhole camera model, the infrared (IR) radiation emerging from this object would be projected into the camera according to:

$$\underline{\mathbf{x}}_p = \mathbf{K}\mathbf{x}_n = \mathbf{K}\mathbf{x}_c\frac{1}{z_c}, \quad \underline{\mathbf{x}}_c = \mathbf{T}_w^c(t)\underline{\mathbf{x}}_w \tag{1}$$

where $\mathbf{T}_w^c(t)$ is the transformation matrix from the world frame to the camera frame at time t, $\mathbf{x}_c$ are the camera frame coordinates, $\mathbf{x}_n$ are the normalized coordinates, $\mathbf{K}$ is the camera intrinsic matrix, and $\underline{\mathbf{x}}_p = [u, v, 1]^T$ are the pixel coordinates.

Any real lens will suffer from distortion, and this can typically be modeled as:

$$\underline{\mathbf{x}}'_p = \mathbf{K}\mathbf{x}'_n = \mathbf{K}g(\mathbf{x}_n) \tag{2}$$

where $\underline{\mathbf{x}}'_p = [u', v', 1]^T$ are the distorted pixel coordinates and $g(\mathbf{x}_n)$ represents a distortion model, e.g., the radial-tangential distortion model. The inverse projection is then given by:

$$\mathbf{x}_n = g^{-1}(\mathbf{x}'_n) = g^{-1}(\mathbf{K}^{-1}\underline{\mathbf{x}}'_p) \tag{3}$$

which typically does not have a closed form solution, but can be solved iteratively.

In microbolometer thermal cameras, the pixels are made of a material that has a strongly temperature-dependent electrical resistance. Incoming IR radiation heats the pixels and their electrical resistances are periodically measured to obtain an image [37]. The temperatures of the pixels do not reach a steady state instantaneously and this gives rise to motion blur in the image. Specifically, the noise-free value of a pixel (u', v') at time t can be given by [37]:

$$m_{u',v',t} = \frac{1}{\tau}\int_{-\infty}^{t} \exp\left(\frac{s-t}{\tau}\right) p_{u',v',s}\, ds \tag{4}$$

where τ is the thermal time constant, a property of the camera, and $p_{u',v',t}$ is a value directly proportional to the power incident on pixel (u', v') at time t. Note that the pixels of a microbolometer are always exposed and that the power incident on a pixel at any previous time has an exponentially decaying impact on the current pixel value. This is in contrast to the motion blur model of visible spectrum cameras, which involves a finite exposure period over which all incident power has equal contribution [37].

Further complicating the image formation model, microbolometer cameras employ a rolling shutter readout [33, 47]. That is, in a given image i, if $t_{0,0,i}$ is the time the top left pixel $(0, 0)$ is measured, the pixel (u', v') will be measured at time:

$$t_{u',v',i} = t_{0,0,i} + u'\Delta t_{\text{pix}} + v'w\Delta t_{\text{pix}} \tag{5}$$

where Δt_{pix} is the readout delay between each pixel and w is the width of the image.

Finally, the value measured at pixel (u', v') is impacted by various sources of noise, which can be modeled as an offset [2]:

$$n_{u',v',t} = m_{u',v',t} + o_{u',v',t} \tag{6}$$

where the noise $o_{u',v',t}$ can be broken down into slowly varying stripe noise (i.e., offsets that are correlated across rows and columns), quickly varying stripe noise, high-frequency random noise, and a slowly varying and spatially smooth component [2]. The slowly varying portions of this noise are often referred to as Fixed Pattern Noise (FPN). Note that in practical use, and in our experiments, much of the FPN is factory calibrated or estimated via shutter-based non-uniformity corrections (NUCs). Still, remaining uncorrected FPN or errors in the corrections appears as *residual* FPN that must be accounted for.

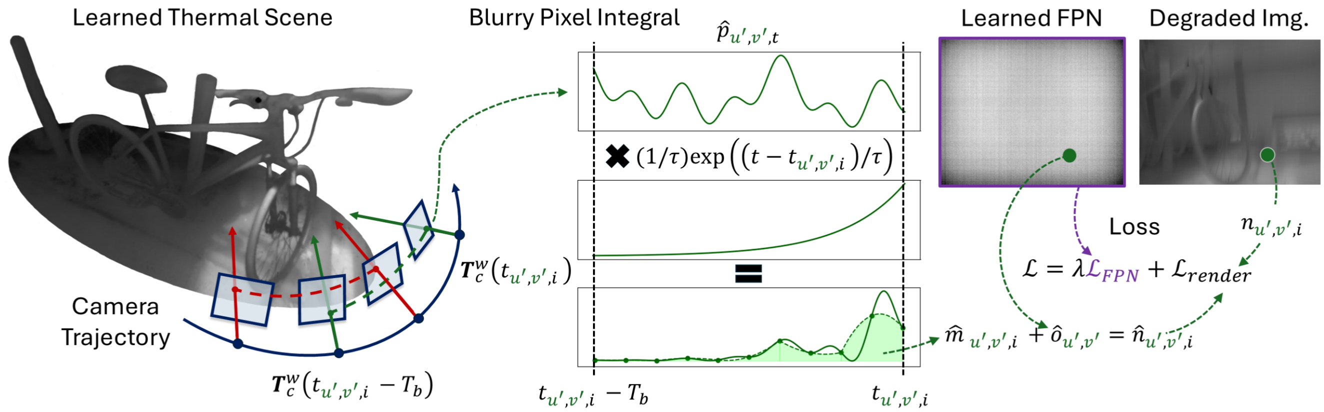


Figure 2. The rendering pipeline of TRNeRF. The left side of the figure depicts the learned scene and the views traversed by two pixels read out at different times. The remaining sections show how a single degraded pixel value is estimated and contributes to the loss.

3.2. Neural Radiance Fields

Our method is built upon Neural Radiance Fields (NeRFs) [28], which we introduce here briefly. Given multiple global shutter images $\{\mathbf{I}_i\}_{i=1}^{N_{\text{img}}}$ of a scene with corresponding camera poses $\{\mathbf{T}_{c_i}^w\}_{i=1}^{N_{\text{img}}}$, NeRF offers a way to represent the scene implicitly using an MLP. The NeRF network f_θ, parameterized by weights θ, is trained to predict the volume density σ and view-dependent color $\mathbf{q}$ of a point $\mathbf{x}_w$ viewed from the direction $\mathbf{d}_w$:

$$(\mathbf{q}, \sigma) = f_\theta(\mathbf{x}_w, \mathbf{d}_w) \tag{7}$$

The NeRF network is trained by using these predictions to render pixel values and compare them against the measured pixel values in the images $\{\mathbf{I}_i\}_{i=1}^{N_{\text{img}}}$. A pixel value is rendered by projecting a ray through the pixel into the scene, applying Eq. (7) at multiple points sampled along the ray, and accumulating the results. Specifically, the ray corresponding to the pixel (u', v') in image i is given by:

$$\mathbf{r}(s) = \mathbf{t}_{w\to c_i}^w + s\mathbf{d}, \quad \mathbf{d} = \frac{\mathbf{R}_{c_i}^w \mathbf{x}_n}{\left\|\mathbf{R}_{c_i}^w \mathbf{x}_n\right\|_2} \tag{8}$$

where $\mathbf{R}_{c_i}^w$ and $\mathbf{t}_{w\to c_i}^w$ are the rotation and translation embedded in the pose $\mathbf{T}_{c_i}^w$ and $\mathbf{x}_n$ is obtained by Eq. (3). After applying Eq. (7) at samples $\{\mathbf{r}(s_j)\}_{j=1}^{N_s}$ to compute colors $\{\mathbf{q}_j\}_{j=1}^{N_s}$ and volume densities $\{\sigma_j\}_{j=1}^{N_s}$ the pixel value is finally estimated as:

$$\hat{\mathbf{c}}_{u',v',i} = \sum_{j=1}^{N_s} \exp\left(-\sum_{k=1}^{j-1} \sigma_k \delta_k\right)(1 - \exp(-\sigma_j \delta_j))\mathbf{q}_j \tag{9}$$

where $\delta_j = s_{j+1} - s_j$ is the distance between samples.

Many papers have improved upon the original NeRF technique. Our method is conceptually agnostic to the underlying NeRF architecture, but we select Instant-NGP [29] to benefit from its relatively fast training and rendering.

4. Method

In this section, we present the details of TRNeRF, our method to restore thermal images via NeRF. The essential goal of our method is to train a NeRF network f_θ from degraded microbolometer images such that we can use it to render new images that remove distortions and approximate what would have been captured from an ideal thermal camera. To achieve this, we treat the rendered pixel value given by Eq. (9) as an estimate of $p_{u',v',t}$ from Eq. (4) and we extend the rendering pipeline to output noisy, blurry, rolling shutter pixel values that can be supervised by the real captured images subject to degradation. An overview of the technique is depicted in Fig. 2. At the inference stage, the extensions are removed and the network can be used to directly render sharp, global shutter, and clear thermal images.

4.1. Motion Blur and Rolling Shutter Correction

To incorporate the microbolometer motion blur and rolling shutter readout into the NeRF rendering pipeline, we begin by assuming that our training images $\{\mathbf{I}_i\}_{i=1}^{N_{\text{img}}}$, with timestamps $\{t_{0,0,i}\}_{i=1}^{N_{\text{img}}}$, and known poses $\{\mathbf{T}_{c_i}^w\}_{i=1}^{N_{\text{img}}}$, were captured at a high enough rate to support accurate interpolation. With this assumption, we apply cubic interpolation to the translations $\{\mathbf{t}_{w\to c_i}^w\}_i^{N_{\text{img}}}$ and spherical linear interpolation (slerp) to the rotations $\{\mathbf{R}_{c_i}^w\}_i^{N_{\text{img}}}$ to obtain the camera pose $\mathbf{T}_c^w(t)$ at any time t during the recording period.

The ability to compute $\mathbf{T}_c^w(t)$ enables Equations 8 and 9 to be evaluated at any time to obtain $\hat{p}_{u',v',t}$. With this, we can then estimate the blurry, but noise-free, value of the pixel (u', v') at time t with a discrete approximation of the integral in Eq. (4):

$$\hat{m}_{u',v',t} = \frac{1}{\tau}\sum_{l=0}^{N_b-1} w_l \exp\left(\frac{-l\Delta t_b}{\tau}\right)\hat{p}_{u',v',t-l\Delta t_b}\Delta t_b \tag{10}$$

where Δt_b is the step between blur samples, $\{w_l\}_{l=0}^{N_b-1} = \{\frac{1}{3}, \frac{4}{3}\,\frac{2}{3}, \frac{4}{3}\,\frac{2}{3}, ..., \frac{4}{3}, \frac{1}{3}\}$ are weights according to Simpson's rule, and N_b is the number of blur samples, which must be odd to support Simpson's rule. We use Simpson's rule as it is generally a more accurate numerical integration method than a Riemann sum ($w_l = 1, \forall l$). The number of blur samples N_b and the interval of integration $T_b = \Delta t_b(N_b - 1)$ are hyperparameters. Note that T_b must be large enough to well approximate the unbounded integral in Eq. (4). $T_b = 5\tau$ captures $> 99\%$ of the value in the case of constant $\hat{p}$.

Ultimately, we need to compute a loss between rendered and captured pixels in specific images. To render the blurry value of the pixel (u', v') in the image i while accounting for the rolling shutter readout of the microbolometer, we substitute $t_{u',v',i}$ from Eq. (5) into Eq. (10). We denote the result of this substitution as $\hat{m}_{u',v',i}$.

4.2. Fixed Pattern Noise Correction

To account for the slowly varying FPN offset represented in Eq. (6), we introduce a learnable per-pixel offset $\hat{o}_{u',v'}$ that we apply to get the final rendered value:

$$\hat{n}_{u',v',i} = \hat{m}_{u',v',i} + \hat{o}_{u',v'} \tag{11}$$

Finally, the loss is computed as:

$$\mathcal{L} = \mathcal{L}_{\text{render}} + \lambda \mathcal{L}_{\text{FPN}} \tag{12}$$

$$\mathcal{L}_{\text{render}} = \frac{1}{|P|} \sum_{(u',v',i) \in P} (\hat{n}_{u',v',i} - n_{u',v',i})^2 \tag{13}$$

$$\mathcal{L}_{\text{FPN}} = \frac{1}{wh} \left| \sum_{u'=0}^{w-1} \sum_{v'=0}^{h-1} \hat{o}_{u',v'} \right| \tag{14}$$

where $\mathcal{L}_{\text{render}}$ is the rendering loss, $\mathcal{L}_{\text{FPN}}$ is a loss computed from the learned FPN offsets, P is a set of pixels randomly sampled across images and pixel coordinates, w and h are the height and width of the images, and λ is a loss coefficient. $\mathcal{L}_{\text{render}}$ supervises the rendered degraded pixel values $\hat{n}_{u',v',i}$ with the measured values $n_{u',v',i}$ across the set P, while $\mathcal{L}_{\text{FPN}}$ enforces that the mean of the learned offsets is zero to maintain the average intensity of the scene. If freely optimized, offsets with a positive or negative mean would drive the learned scene intensity in the opposite direction.

5. Experiments

5.1. Dataset

To collect our dataset, we designed a rig with a backpack mounted computer and hand-held sensor platform pictured in Fig. 3. The sensor platform includes two 640×512 microbolometer thermal cameras (FLIR ADK), two 1440×1080 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.

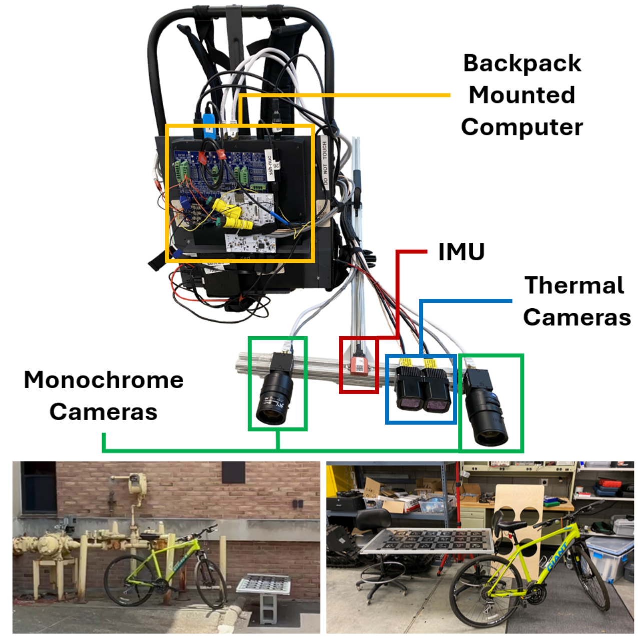


Figure 3. Image of the data collection rig (top) and reference RGB images of the outdoor (bottom left) and indoor (bottom right) scenes.

The thermal cameras are placed side-by-side to support the comparison of different camera settings with minimal parallax. In particular, we disable the default noise filters in the right thermal camera and keep them enabled in the left. The monochrome cameras are global shutter, and we limit their exposure times to 1 ms to ensure sharp images. We trigger image capture simultaneously across all cameras at 60 Hz and follow the approach described in [3] to synchronize the recorded data and calibrate all sensors. This calibration involves an Aprilgrid board [5, 11, 31] that is constructed of aluminum and vinyl in order to appear in both the visible and thermal spectra [3].

We record data in a sunny outdoor scene with high thermal contrast, and an air-conditioned indoor scene with low thermal contrast. Reference RGB images of these scenes are shown in Fig. 3. We place the multi-spectral calibration board in each scene to support quantitative evaluation as described in Sec. 5.3. We also place uniform hot and cold sources near each scene. We record these sources before and after each sequence to support pseudo ground truth generation, as described in Sec. 5.3. In each scene we record three sequences, denoted slow, medium, and fast, with increasingly aggressive six degree-of-freedom camera motion. For our experiments, we select a 3.7 minute (13.4k image) subset of each slow sequence and a 1 minute (3.6k image) subset of each medium and fast sequence. Additional details on the dataset collection are provided in our supplementary material.

To obtain the pose estimates required by NeRF, it is common to apply the structure-from-motion method COLMAP [40] to the training images. However, COLMAP fails with thermal images collected under fast motion due to motion blur and rolling shutter distortions. Instead, we obtain scaled poses for the left monochrome images, and transform these poses to each thermal camera using the calibrated extrinsics. Specifically, for each scene, we run COLMAP with 1k images from the left monochrome camera sampled across all sequences. Then, we localize all remaining monochrome images, left and right, against the COLMAP reconstruction using hloc [39]. Finally, we scale the poses by comparing the median stereo baseline between the estimated left and right poses against the calibrated baseline. After scaling, we assess the accuracy of the poses by comparing the estimated relative poses between the left and right cameras against the calibrated extrinsics. We found the accuracy in position to be 2 mm and the accuracy in rotation to be 0.06° on average. Note that by running COLMAP with images from each sequence together, all of the estimated poses for a given scene are in a common world frame. This fact is exploited in Sec. 5.3 to generate pseudo ground truth images.

For the FLIR ADK cameras, $\tau = 8$ ms and $\Delta t_{\text{pix}} = \Delta t_{\text{row}}/w$, where $w = 640$ is the image width and $\Delta t_{\text{row}} = 27.8\ \mu$s is the time required to read out a single row. The first pixel is read out 0.5 ms after the trigger signal while the monochrome images begin exposure immediately. We account for this delay when interpolating the poses.

5.2. Implementation Details

We build TRNeRF off of the nerfstudio [44] implementation of Instant-NGP [29]. Typically, Instant-NGP is trained with 8-bit, 3-channel RGB images, which are converted to 32-bit floating point and mapped from the range $[0, 255]$ to the range $[0, 1]$. Most thermal cameras capture 16-bit, single-channel images. For compatibility with Instant-NGP, we duplicate the thermal images across the 3 channels and perform the operations described in Equations 10, 11, and 13 separately in each channel. Additionally, the captured values in thermal images typically span a small subset of the full 16-bit range. Therefore, after conversion to 32-bit floating point, we map the captured values to the range $[0, 1]$ as follows:

$$n = \frac{n_{\text{orig}} - n_{\text{low}}}{n_{\text{high}} - n_{\text{low}}} \tag{15}$$

where n_{orig} is the captured value, n is the value used in training, and n_{low} and n_{high} are set to the minimum and maximum values captured in the scene across all three sequences. Notably, we avoid converting the training images to 8-bit.

When rendering a restored image, we take the pixel-wise mean across the 3 channels, invert Eq. (15), and convert to 16-bit. For evaluation and visualization, we apply Eq. (15), multiply by 255, and convert to 8-bit. To improve contrast in the evaluation and visualization images, we set n_{low} and n_{high} to the 0.1 and 99.9 percentiles.

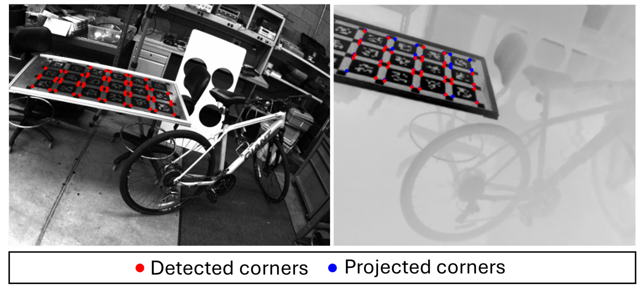


Figure 4. An example of computing the detection metric with a monochrome image (left) and inverted thermal image (right).

For training, we set the number of blur samples to $N_b = 19$, the interval of integration to $T_b = 5\tau = 40$ ms, and the FPN loss coefficient to $\lambda = 1 \times 10^{-3}$. We render 500 pixels per iteration for a total of 60k iterations. Note that each rendered pixel requires $N_b = 19$ rays when blur compensation is enabled, and only 1 otherwise. We use the Adam optimizer [14] to update our FPN offset parameters and set the learning rate to start at 1×10^{-4} and decay exponentially to 5×10^{-5}. We use the default values in nerfstudio for all remaining hyperparameters.

5.3. Evaluation Methods

To evaluate TRNeRF, and the baselines described in Sec. 5.4, we render a restored image for each training image. Specifically, each restored image is distortion-free, global shutter, and rendered from the pose of the camera when the trigger signal was received. We introduce two quantitative evaluation methods summarized below with further details given in our supplementary material.

The first method utilizes the multi-spectral Aprilgrid board that we placed in each scene. The board can be detected in both the sharp visible spectrum images and the restored thermal images. For every left monochrome image in which we can detect the board and estimate its pose, we project all of the board's corners into the restored thermal image corresponding to the same trigger. We retain the projected corners that correspond to AprilTags with all four corners lying in the thermal camera's field-of-view. We then attempt to detect the board in the thermal image, and we determine the number of the projected corners that were successfully detected. We aggregate the number of projected and detected corners over all images in a sequence to compute the detection percentage. This percentage reflects how well areas of the board were restored. The process for a single pair of images is visualized in Fig. 4. Note that the thermal image is inverted for AprilTag detection.

The second method involves generating pseudo ground truth images. This method relies on the assumption that

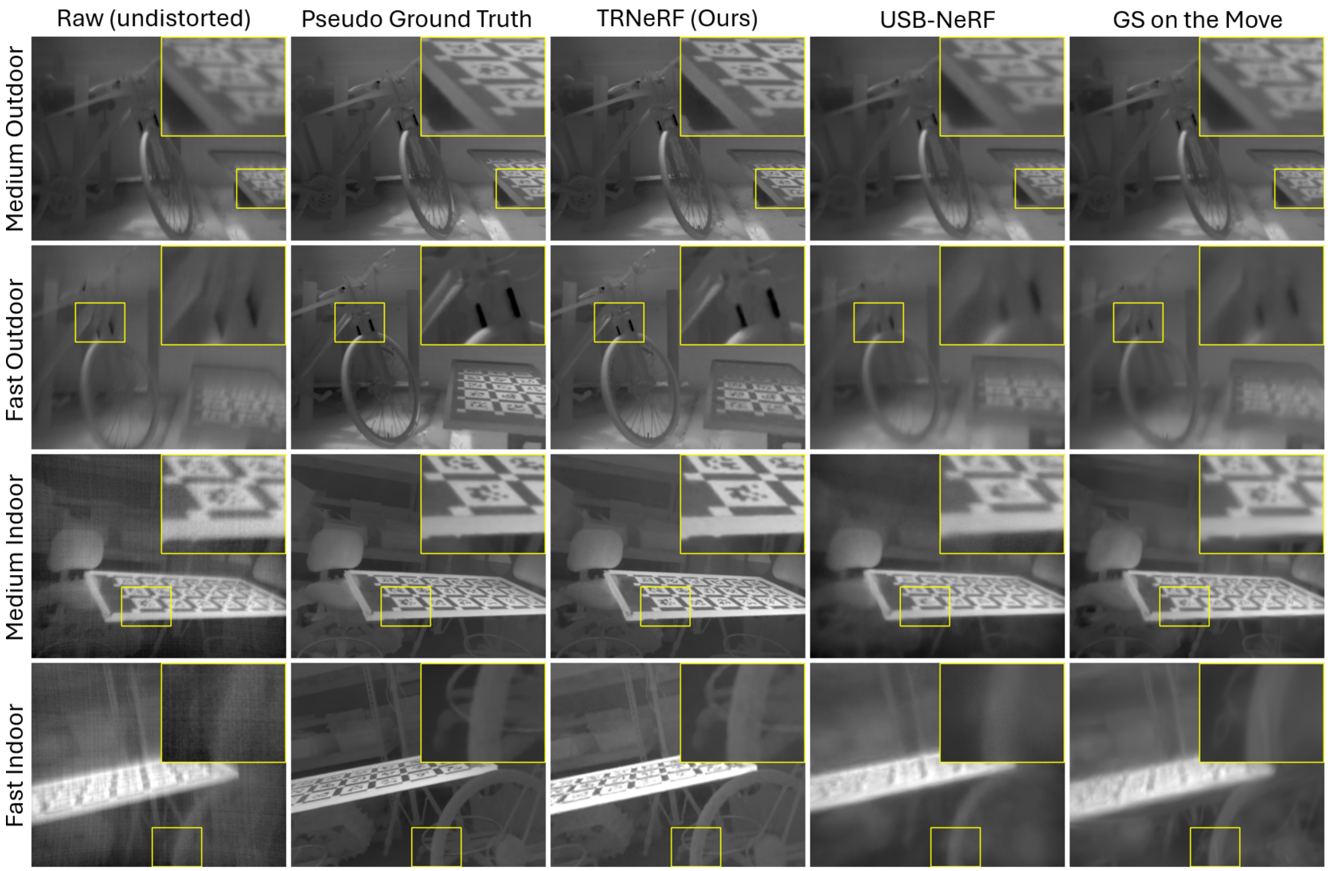


Figure 5. Qualitative results on the medium and fast sequences from the indoor and outdoor scenes.

an accurate thermal NeRF model can be trained on each slow sequence. Under this assumption, we can use these NeRF models to render pseudo ground truth images for the medium and fast sequences. To avoid self-comparison, we use Instant-NGP [29], without any of our modifications, to generate the pseudo ground truth images. To account for FPN, we use the recordings of the hot and cold sources to perform a two-point NUC [36] and apply the corrections to the slow sequence training images at the input to Instant-NGP. Note that we use the two-point NUC only in the generation of pseudo ground truth, and not in our proposed method, as we assume hot and cold sources will not always be available. We report the average LPIPS value [54] between the pseudo ground truth and restored images. To avoid using inaccurate pseudo ground truth images, we restrict this evaluation to viewpoints sufficiently similar to the slow sequence training images.

Method	Detection % ↑			LPIPS ↓	
	Slow	Med.	Fast	Med.	Fast
Outdoor					
Raw	74.3	16.0	**0.2**	0.155	0.336
Instant-NGP	69.6	0.0	0.0	0.129	0.281
USB-NeRF	65.5	5.9	0.0	0.155	0.355
GS on the Move	74.7	3.5	0.0	0.093	0.283
TRNeRF (Ours)	**74.8**	**58.6**	0.0	**0.042**	**0.078**
Indoor					
Raw	49.2	3.5	0.0	0.672	0.736
Instant-NGP	51.8	0.0	0.0	0.253	0.393
USB-NeRF	39.8	0.0	0.0	0.501	0.536
GS on the Move	59.2	0.0	0.0	0.206	0.328
TRNeRF (Ours)	**64.9**	**41.7**	0.0	**0.068**	**0.117**

Table 1. Detection percentage (higher is better) and LPIPS (lower is better) for each method on each sequence.

5.4. Comparisons

Existing methods for thermal deblurring and rolling shutter correction either lack available code or are limited to camera specifications and scenarios that are unmet by our dataset. Furthermore, the only available thermal NeRF implementations were designed to fuse with RGB images [9, 25]. Therefore, we turn to state-of-the-art NeRF and 3DGS based restoration methods for our baselines. Specifically, we compare TRNeRF against USB-NeRF [23], and Gaussian Splatting on the Move [42], which we abbreviate here as GSotM. We also compare against Instant-NGP [29] without our modifications. The training images were input to all methods in the same way, as described in Sec. 5.2. All methods were run with the right thermal camera, as this re-

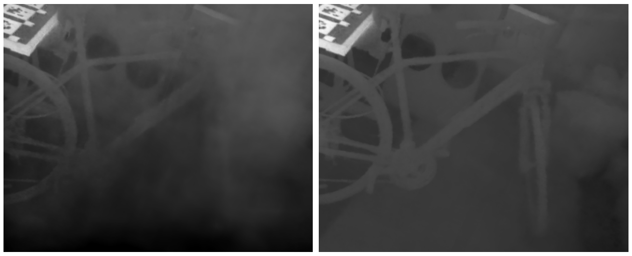


Figure 6. A restored image from the medium indoor sequence without FPN correction (left) and with FPN correction (right).

sulted in better performance (see Sec. 5.5). Further details on the implementations and additional qualitative comparisons are provided in our supplementary material.

The quantitative results of our experiments are given in Tab. 1, alongside the results of applying our evaluation methods to the raw images (after undistortion). The detection percentage results with the raw images confirm that the slow sequence images are not substantially degraded, especially in the high thermal contrast outdoor scene where the FPN is less significant. Correspondingly, the baselines all perform well on the slow sequences by the detection percentage metric.

In the medium sequences, TRNeRF attains substantial detection percentages while the baseline results are low or zero. We note that the detections in the raw images are limited to brief periods of slower camera motion. USB-NeRF performs poorly, suggesting that rolling shutter correction alone is insufficient to restore microbolometer images. GSotM accounts for rolling shutter and motion blur, but also achieves limited success. We believe this is because GSotM does not account for FPN, uses a screen-space approximation that is invalid under aggressive motion, and models blur with the photoelectric image formation model, which is inapplicable to microbolometers [42].

None of the methods, including ours, can produce detections in the fast sequences. The board may be an especially challenging object to restore, as the aluminum is reflective in the thermal spectrum. However, it is clear from the LPIPS and qualitative results, shown in Fig. 5, that TRNeRF still achieves significant improvement in image restoration in these sequences, whereas portions of the raw images and baseline results are so degraded as to be nearly indiscernible.

5.5. Ablation Study

To assess the components of our method, we tested with rolling shutter correction only, motion blur correction only (with Simpson's rule), and the two corrections combined using a Riemann sum and using Simpson's rule. Additionally, we tested our full method with the training images converted to 8-bit (as done for evaluation and visualization) and with the images from the left thermal camera, which had the

Ablation Method	Detection % ↑			LPIPS ↓	
	Slow	Med.	Fast	Med.	Fast
Outdoor					
Rolling shutter (RS)	56.4	11.3	0.0	0.119	0.274
Blur, Simpsons (BS)	71.1	0.0	0.0	0.104	0.350
RS + Blur, Riemann	75.8	59.4	**4.0**	0.166	0.147
RS + BS	74.2	**62.3**	0.0	0.054	0.091
TRNeRF, 8-bit	**76.2**	60.2	0.0	0.044	**0.073**
TRNeRF, w/ filters	75.0	50.1	0.0	0.047	0.083
TRNeRF	74.8	58.6	0.0	**0.042**	0.078
Indoor					
Rolling shutter (RS)	54.1	0.7	0.0	0.232	0.341
Blur, Simpsons (BS)	47.7	0.0	0.0	0.338	0.468
RS + Blur, Riemann	64.2	40.2	**1.5**	0.217	0.261
RS + BS	60.0	38.3	0.0	0.182	0.250
TRNeRF, 8-bit	64.6	37.2	0.0	0.141	0.308
TRNeRF, w/ filters	45.2	14.0	0.0	0.207	0.223
TRNeRF	**64.9**	**41.7**	0.0	**0.068**	**0.117**

Table 2. Detection percentage (higher is better) and LPIPS (lower is better) for each method of the ablation study on each sequence.

FLIR ADK's default noise filters enabled. The quantitative results of this ablation study are given in Tab. 2.

The results show that correcting rolling shutter or motion blur alone is insufficient to restore microbolometer images. When accounting for both, using a Riemann sum achieves a nonzero detection percentage on the fast sequences. The Riemann sum may better approximate the step function witnessed by a pixel traversing the Aprilgrid pattern. Nonetheless, the LPIPS results obtained with Simpson's rule suggest that it is the better option in general. The remaining methods are only significantly differentiated in the indoor scene, where the importance of correcting FPN is evident in the LPIPS results. The difference is also clearly apparent in the qualitative results, as shown in Fig. 6. Additionally, the indoor results suggest that converting the training images to 8-bit and keeping noise filters enabled in the camera significantly harms performance. While the filters appear to improve the captured images, they may also introduce biases that NeRF is less robust to than random noise.

6. Conclusions

We present TRNeRF, a method to restore microbolometer thermal images degraded by motion blur, rolling shutter distortions, and fixed pattern noise. We propose extensions to the NeRF rendering pipeline that account for these degradations, allowing an implicit representation of the original scene to be trained directly from the degraded images. To validate the restoration performance of our method, we introduce a new dataset and two novel quantitative evaluation schemes. Our experiments demonstrate that TRNeRF can restore sharp, global shutter, and clear thermal images, even under extremely aggressive camera motion.

Acknowledgment: We thank Katy Frank for her help in collecting the dataset.

References

[1] Ganbayar Batchuluun, Young Won Lee, Dat Tien Nguyen, Tuyen Danh Pham, and Kang Ryoung Park. Thermal Image Reconstruction Using Deep Learning. *IEEE Access*, 8:126839–126858, 2020. 1, 2, 3

[2] Lijing Cai, Xiangyu Dong, Kailai Zhou, and Xun Cao. Exploring video denoising in thermal infrared imaging: Physics-inspired noise generator, dataset, and model. *IEEE Transactions on Image Processing*, 33:3839–3854, 2024. 1, 2, 3

[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. 5

[4] Wenbo Chen and Ligang Liu. Deblur-GS: 3D Gaussian Splatting from Camera Motion Blurred Images. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1):1–15, May 2024. 2

[5] 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. 5

[6] Rikke Gade and Thomas B. Moeslund. Thermal cameras and applications: a survey. *Machine Vision and Applications*, 25(1):245–262, Jan. 2014. 1

[7] Juntao Guan, Rui Lai, and Ai Xiong. Wavelet Deep Neural Network for Stripe Noise Removal. *IEEE Access*, 7:44544–44554, 2019. 1, 2

[8] Juntao Guan, Rui Lai, Ai Xiong, Zesheng Liu, and Lin Gu. Fixed pattern noise reduction for infrared images based on cascade residual attention CNN. *Neurocomputing*, 377:301–313, Feb. 2020. 1, 2

[9] Mariam Hassan, Florent Forest, Olga Fink, and Malcolm Mielle. Thermonerf: Multimodal neural radiance fields for thermal novel view synthesis. *arXiv preprint arXiv:2403.12154*, 2024. 2, 7

[10] Zewei He, Yanpeng Cao, Yafei Dong, Jiangxin Yang, Yanlong Cao, and Christel-Löic Tisse. Single-image-based nonuniformity correction of uncooled long-wave infrared detectors: a deep-learning approach. *Applied Optics*, 57(18):D155–D164, Jun 2018. 1, 2

[11] 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. 5

[12] Kelsey M. Judd, Michael Thornton, and Austin Richards. Automotive sensing: assessing the impact of fog on LWIR, MWIR, SWIR, visible, and lidar performance. In *Infrared Technology and Applications XLV*, page 50, Baltimore, United States, May 2019. SPIE. 1

[13] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4):1–14, Aug. 2023. 2

[14] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, Jan. 2017. arXiv:1412.6980 [cs]. 6

[15] Kangwook Ko, Kyujin Shim, Kangil Lee, and Changick Kim. Large-Scale Benchmark for Uncooled Infrared Image Deblurring. *IEEE Sensors Journal*, 23(24):30119–30128, Dec. 2023. 1, 3

[16] Xiaodong Kuang, Xiubao Sui, Yuan Liu, Qian Chen, and Guohua GU. Single Infrared Image Optical Noise Removal Using a Deep Convolutional Neural Network. *IEEE Photonics Journal*, 10(2):1–15, Apr. 2018. 1, 2

[17] Helge A. Lauterbach, C. Bertram Koch, Robin Hess, Daniel Eck, Klaus Schilling, and Andreas Nüchter. The Eins3D project — Instantaneous UAV-Based 3D Mapping for Search and Rescue Applications. In *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 1–6, Sept. 2019. 1

[18] Byeonghyeon Lee, Howoong Lee, Usman Ali, and Eunbyung Park. Sharp-nerf: Grid-based fast deblurring neural radiance fields using sharpness prior. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3709–3718, 2024. 2

[19] Dogyoon Lee, Minhyeok Lee, Chajin Shin, and Sangyoun Lee. DP-NeRF: Deblurred Neural Radiance Field with Physical Scene Priors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12386–12396, Vancouver, BC, Canada, June 2023. 2

[20] Dongwoo Lee, Jeongtaek Oh, Jaesung Rim, Sunghyun Cho, and Kyoung Mu Lee. Exblurf: Efficient radiance fields for extreme motion blurred images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17639–17648, 2023. 2

[21] Jungho Lee, Dogyoon Lee, Minhyeok Lee, Donghyung Kim, and Sangyoun Lee. SMURF: Continuous Dynamics for Motion-Deblurring Radiance Fields, Mar. 2024. arXiv:2403.07547 [cs]. 2

[22] Kangil Lee, Yuseok Ban, and Changick Kim. Motion Blur Kernel Rendering Using an Inertial Sensor: Interpreting the Mechanism of a Thermal Detector. *Sensors*, 22(5):1893, Feb. 2022. 1, 2, 3

[23] 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. 2, 7

[24] Xiangjie Li, Shuxiang Xie, Ken Sakurada, Ryusuke Sagawa, and Takeshi Oishi. Implicit neural fusion of rgb and far-infrared 3d imagery for invisible scenes. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024. 2

[25] Yvette Y Lin, Xin-Yi Pan, Sara Fridovich-Keil, and Gordon Wetzstein. Thermalnerf: Thermal radiance fields. In *2024 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2024. 2, 7

[26] Achleshwar Luthra, Shiva Souhith Gantha, Xiyun Song, Heather Yu, Zongfang Lin, and Liang Peng. Deblur-NSFF: Neural Scene Flow Fields for Blurry Dynamic Scenes. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3646–3655, Waikoloa, HI, USA, Jan. 2024. 2

[27] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V Sander. Deblur-nerf: Neural radiance

fields from blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12861–12870, 2022. 2

[28] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2, 4

[29] 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. 4, 6, 7

[30] Jeongtaek Oh, Jaeyoung Chung, Dongwoo Lee, and Kyoung Mu Lee. DeblurGS: Gaussian Splatting for Camera Motion Blur, Apr. 2024. arXiv:2404.11358 [cs]. 2

[31] Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3400–3407, May 2011. 5

[32] Beate Oswald-Tranta. Temperature reconstruction of infrared images with motion deblurring. *Journal of Sensors and Sensor Systems*, 7(1):13–20, Jan. 2018. 1, 2, 3

[33] J. Peeters, E. Louarroudi, D. De Greef, S. Vanlanduit, J. J. J. Dirckx, and G. Steenackers. Time calibration of thermal rolling shutter infrared cameras. *Infrared Physics & Technology*, 80:145–152, Jan. 2017. 1, 2, 3

[34] Cheng Peng and Rama Chellappa. PDRF: Progressively Deblurring Radiance Field for Fast Scene Reconstruction from Blurry Images. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(2):2029–2037, June 2023. 2

[35] Cheng Peng, Yutao Tang, Yifan Zhou, Nengyu Wang, Xijun Liu, Deming Li, and Rama Chellappa. Bags: Blur agnostic gaussian splatting through multi-scale kernel modeling. In *European Conference on Computer Vision*, page 293–310, 2024. 2

[36] David L. Perry. Linear theory of nonuniformity correction in infrared staring sensors. *Optical Engineering*, 32(8):1854, 1993. 7

[37] 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, 2, 3

[38] Vishwanath Saragadam, Akshat Dave, Ashok Veeraraghavan, and Richard G. Baraniuk. Thermal Image Processing via Physics-Inspired Deep Networks. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 4040–4048, Montreal, BC, Canada, Oct. 2021. 1, 2, 3

[39] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From Coarse to Fine: Robust Hierarchical Localization at Large Scale. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12708–12717, Long Beach, CA, USA, June 2019. 6

[40] Johannes L. Schonberger and Jan-Michael Frahm. Structure-from-Motion Revisited. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, Las Vegas, NV, USA, June 2016. 6

[41] Christian Schönauer, Emanuel Vonach, Georg Gerstweiler, and Hannes Kaufmann. 3D building reconstruction and thermal mapping in fire brigade operations. In *Proceedings of the 4th Augmented Human International Conference*, pages 202–205, Stuttgart Germany, Mar. 2013. ACM. 1

[42] 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. 2, 7, 8

[43] Huiqiang Sun, Xingyi Li, Liao Shen, Xinyi Ye, Ke Xian, and Zhiguo Cao. Dyblurf: Dynamic neural radiance fields from blurry monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7517–7527, 2024. 2

[44] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–12, 2023. 6

[45] Nisha Varghese, Mahesh Mohan M. R., and A. N. Rajagopalan. Fast Motion-Deblurring of IR Images. *IEEE Signal Processing Letters*, 29:459–463, 2022. 1, 2, 3

[46] Stephen Vidas and Peyman Moghadam. HeatWave: A handheld 3D thermography system for energy auditing. *Energy and Buildings*, 66:445–460, Nov. 2013. 1

[47] Michael Vollmer and Klaus-Peter Möllmann. *Infrared Thermal Imaging: Fundamentals, Research and Applications*. Wiley, 1 edition, Dec. 2017. 3

[48] Peng Wang, Lingzhe Zhao, Ruijie Ma, and Peidong Liu. BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4170–4179, Vancouver, BC, Canada, June 2023. 2

[49] Yining Wang, Jinyi Zhang, and Yuxi Jiang. KT-NeRF: multiview anti-motion blur neural radiance fields. *Journal of Electronic Imaging*, 33(03), May 2024. 2

[50] Pengfei Xiao, Yecai Guo, and Peixian Zhuang. Removing stripe noise from infrared cloud images via deep convolutional networks. *IEEE Photonics Journal*, 10(4):1–14, 2018. 1, 2

[51] Jiacong Xu, Mingqian Liao, Ram Prabhakar Kathirvel, and Vishal M. Patel. Leveraging Thermal Modality to Enhance Reconstruction in Low-Light Conditions. In *European Conference on Computer Vision*, pages 321–339, 2024. 2

[52] P.V. Karthik Yadav, Isha Yadav, B. Ajitha, Abraham Rajasekar, Sudha Gupta, and Y. Ashok Kumar Reddy. Advancements of uncooled infrared microbolometer materials: A review. *Sensors and Actuators A: Physical*, 342:113611, Aug. 2022. 1

[53] Tianxiang Ye, Qi Wu, Junyuan Deng, Guoqing Liu, Liu Liu, Songpengcheng Xia, Liang Pang, Wenxian Yu, and Ling Pei. Thermal-NeRF: Neural Radiance Fields from an Infrared Camera, Mar. 2024. arXiv:2403.10340 [cs]. 2

[54] 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. 7

[55] Lingzhe Zhao, Peng Wang, and Peidong Liu. Bad-gaussians: Bundle adjusted deblur gaussian splatting. In *European Conference on Computer Vision*, pages 233–250, 2024. 2

[56] Yuqing Zhao, Guangyuan Fu, Hongqiao Wang, Shaolei Zhang, and Min Yue. Infrared Image Deblurring Based on Generative Adversarial Networks. *International Journal of Optics*, 2021(1):9946809, 2021. 1, 2, 3

[57] Haidong Zhu, Yuyin Sun, Chi Liu, Lu Xia, Jiajia Luo, Nan Qiao, Ram Nevatia, and Cheng-Hao Kuo. Multimodal neural radiance field. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9393–9399, 2023. 2

[58] Mert Özer, Maximilian Weiherer, Martin Hundhausen, and Bernhard Egger. Exploring Multi-modal Neural Scene Representations With Applications on Thermal Imaging, Mar. 2024. arXiv:2403.11865 [cs]. 2