

## Photometric Correction for Infrared Sensors

Jincheng Zhang<sup>1</sup>, Andrew R Willis<sup>1</sup>, Kevin Brink<sup>2</sup>

<sup>1</sup>University of North Carolina at Charlotte, <sup>2</sup>Air Force Research Laboratory

<sup>1</sup>9201 University City Blvd, Charlotte, NC 28223

<sup>2</sup>Eglin AFB, FL, 32542

<sup>1</sup>{jzhang72, arwillis}@uncc.edu, <sup>2</sup>kevin.brink@us.af.mil

### Abstract

*Infrared thermography has been widely used in several domains to capture and measure temperature distributions across surfaces and objects. This methodology can be further expanded to 3D applications if the spatial distribution of the temperature distribution is available. Structure from Motion (SfM) is a photometric range imaging technique that makes it possible to obtain 3D renderings from a cloud of 2D images. To explore the possibility of 3D reconstruction via SfM from infrared images, this article proposes a photometric correction model for infrared sensors based on temperature constancy. Photometric correction is accomplished by estimating the scene irradiance as the values from the solution to a differential equation for microbolometer pixel excitation with unknown coefficients and initial conditions. The model was integrated into an SfM framework and experimental evaluations demonstrate the contribution of the photometric correction for improving the estimates of both the camera motion and the scene structure. Further, experiments show that the reconstruction quality from the corrected infrared imagery achieves performance on par with state-of-the-art reconstruction using RGB sensors.*

### 1. Introduction

Infrared (IR) sensors have essential applications in a wide variety of sensing contexts and have recently seen much interest as a sensor to facilitate autonomous ground and aerial vehicle navigation. Heightened interest is largely due to the IR sensor’s capability to sense accurate scene image data under low light conditions. This is particularly useful in contexts where illumination is unavailable, e.g., navigating at night and when the visible light sources, surface appearance textures, and surface reflections introduce difficulties to visible light algorithms, e.g., camouflaged targets [17, 21, 25, 26].

Infrared image sensors consist of a grid of radiation-

sensitive optoelectronic components that are sensitive to radiated energy having wavelengths from the IR frequency band which spans wavelengths  $\lambda \in [0.7\mu\text{m}, 1000\mu\text{m}]$ . The IR frequency band is commonly divided into the NIR (Near IR,  $\lambda \in [0.7\mu\text{m}, 2.5\mu\text{m}]$ ), MWIR (Mid-Wavelength IR,  $\lambda \in [2.5\mu\text{m}, 5\mu\text{m}]$ ), LWIR (Long-Wavelength IR,  $\lambda \in [8\mu\text{m}, 15\mu\text{m}]$ ), and the FIR (Far IR,  $\lambda \in [15\mu\text{m}, 1000\mu\text{m}]$ ).

This article proposes a new photometric correction for thermographic infrared cameras. Thermographic imaging devices use microbolometer sensors to detect radiation in the MWIR and LWIR frequency ranges. For simplicity, we will refer to microbolometer sensors simply as IR sensors. *It is important to note that other sensing devices are required to capture images in the NIR and FIR wavelengths and the results of this article do not apply to these IR sensor types [18].*

Similar to related work for RGB cameras [14], photometric correction for these IR image sensors promises to improve the accuracy of measured pixels by estimating the underlying scene irradiance responsible for generating a pixel value. Accurate estimates of the scene irradiance promise to improve the output quality of many existing computer vision algorithms [4, 28]. This is especially true for the large class of algorithms that assume scene points are “view-point invariant,” i.e., the Lambertian assumption, for modest changes in the camera pose.

The proposed photometric correction is tailored to microbolometer pixel sensors whose characteristic response is known in the literature [6, 22] and distinct from the response of RGB image sensors. Photometric correction for these image sensors promises to allow vision algorithms developed for popular RGB image sensors to translate to infrared sensors better. In fact, our results indicate that some aspects of the estimates provided by infrared sensors may be more accurate than corresponding estimates from RGB sensors (see Section 4).

The contributions of this article include novel theoretical and experimental results. These results include:

- A novel photometric correction model for infrared (mi-

microbolometer) sensors is proposed.

- A novel application of this model to SfM problems and a comparison of SfM accuracy with IR vs. RGB sensors.
- Experimental work showing the proposed photometric correction model can improve algorithm performance for SfM problems.
- Experimental work showing the proposed photometric correction model can provide performance benefits over traditional RGB sensors for SfM problems.

To our knowledge, this work constitutes the first quantitative analysis of SfM for IR sensors in the literature. The impact of this photometric correction is important to several sensor types, sensing conditions, and sensing contexts which include: (1) uncooled IR sensors, (2) IR sensors that move or observe moving scenes, (3) IR sensors that operate in high-temperature environments and (4) IR sensors that operate at high frame rates. Each of these circumstances can lead to circumstances where the value of sensed pixels from prior image frames can significantly offset the value of subsequent measurements leading to ghosting effects of the moving object in the video [1]. The proposed IR photometric correction approach seeks to compensate for these effects.

## 2. Related work

The contribution of this work is to propose and apply a photometric correction algorithm specific to IR image sensors. Other than [8] where a linear sensor response model is used to calibrate IR sensors, we have been unable to find other references within the computer vision literature detailing similar approaches. For this reason, we use the related work section to discuss research on models to characterize the response microbolometer image sensor pixel values. Experimental work integrates the proposed photometric correction to a state-of-the-art SfM algorithm and a component of the literature review discusses SfM algorithms and how advances in photometric correction for RGB sensors have served to improve SfM estimates.

### 2.1. Infrared Imaging with Microbolometers

An infrared camera is a device that converts infrared radiation into a visual image that depicts temperature variations across an object or scene. Infrared radiation is a characteristic of all objects that have a temperature higher than absolute zero (zero Kelvin or -273 Celsius) [29]. Thermal energy radiated by scene objects is focused onto the sensor image plane where a grid of microbolometer sensors is placed to convert the optical energy focused onto each grid

element into a pixel voltage indicative of the object temperature.

The physical response of each pixel element is governed by a heating and cooling mechanism as the camera shutter is opened and closed. Similar to RGB optical sensors, microbolometer sensors integrate incident radiation into stored charge when the camera shutter is open and dissipate the stored charge when the camera shutter is closed. This process generates a response analogous to an RC circuit driven by a square wave excitation where the square wave period is determined by the exposure time and its amplitude is determined by the radiated energy of the scene onto the pixel sensor. The rate of integration and dissipation is driven by a sensor-specific time constant,  $\tau$ .

One shortcoming of microbolometer sensors is their response time. RGB pixel sensors based on CMOS technology have been shown to have time constants  $\tau \in [1\mu s, 500\mu s]$  with the median sensor performance across RGB sensors  $\tau \approx 10\mu s$  at room temperature [7]. Typical response times for microbolometer sensors are  $\tau \in [8ms, 15ms]$  which is more than two orders of magnitude larger. Long response times lead to “ghosting” of sensed values where the value of a pixel in a past frame persists in the current frame creating a spatio-temporal blur of moving objects in sensed IR images.

### 2.2. Structure from Motion

Structure from Motion is the process of reconstructing a 3D structure from its projections into a series of images taken from different viewpoints. It commonly starts with feature extraction and matching, followed by geometric verification [27], where the feature matching searches for feature correspondences by finding the most similar feature between image pairs with scene overlapping and comparing the similarities between features. This requires the detected feature to be accurate and reliable throughout the image sequence. We note work in the literature on the related problem of visual odometry from IR image data which estimates the “M” or camera motion estimation component of the SfM problem [19, 20, 24]. Work in [10] demonstrates that RGB SfM algorithms may be used on IR cameras for use in fire-fighting scenarios with some success.

The photometric correction has proven to be important for improving the result of SfM estimates [11, 13, 15]. Direct Sparse Odometry (DSO) [11] achieves state-of-the-art performance by taking full advantage of photometric camera calibration, including lens attenuation, gamma correction [14], and brightness correction. The lens attenuation and gamma correction both require prior knowledge about the sensor and lens being used for solving SfM problems, the brightness correction, instead, is an empirical model that takes into account automatic exposure changes and compensates the pixel values in order to make them more con-

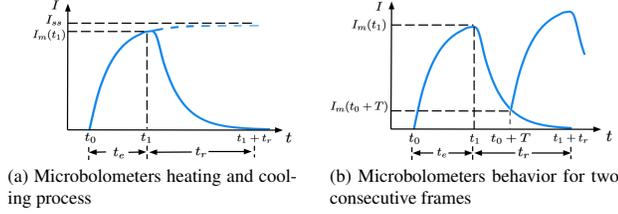


Figure 1. An illustration of microbolometer pixels' behavior during the heating and cooling process. (a) Pixels do not have enough time to reach the temperature of the scene being measured. (b) In two (or more) consecutive frames the later frame(s) has a memory of the energy residual in the previous frame(s).

sistent and stable across a sequence of images. The brightness correction function in DSO is given by  $e^{-a}(I - b)$  where  $a, b$  are the correction parameters, and  $I$  is the pixel value. In their work, the authors show that a naive brightness constancy assumption used in other approaches like LSD-SLAM [12] or SVO [16] significantly decreases the SfM accuracy. This article builds on these concepts by translating this intuition from CMOS-based RGB image sensors to microbolometer-based IR image sensors. A photometric correction model for microbolometer pixel values is proposed in order to improve the performance for SfM problems using IR data.

### 3. Methodology

Photometric correction is necessary for uncooled infrared image sensors because the microbolometers used by these sensors have a response that is entirely different from photon detector imagers such as RGB cameras. Specifically, uncooled microbolometer devices require a certain portion of each frame time to integrate signals. Hence for high-speed temperature measurement with short frame time, pixels in a microbolometer usually do not have enough time to reach the temperature of the scene to be measured (a steady temperature state) before the pixels receive new radiance from objects in the next frame. Moreover, microbolometers do not have a mechanism to reset the integrated signal from the previous frame, and therefore signals captured in previous frames will have a residual impact on the microbolometer pixel reading in the current frame. Both of these factors result in an “inaccuracy” of sorts in pixel values in the images generated by infrared sensors as they are not determined solely by the “current scene” and this results in unreliable reconstruction results via structure from motion.

The proposed model for photometric correction of microbolometer measurements consists of two parts: (1) a heating model which characterizes the IR pixel response during a frame exposure and (2) a cooling model which characterizes the IR pixel response when the sensor is not

exposed. This section describes these models using continuous differential equations and combines these models into a comprehensive photometric correction model for IR pixels measured at arbitrary framerates. We discuss camera models that allow these corrections to be position invariant and, under these circumstances, algorithms can quickly apply photometric correction across all image pixels using lookup tables to improve their performance. We then integrate this model into the DSO SfM solution and detail how the SfM problem is modified by the integration of this new photometric correction.

#### 3.1. Model for Pixel Heating

Heating the IR pixel is modeled as a differential equation with a unit step forcing function,  $\mu(t)$ , where the amplitude of the step is proportional to the scene irradiance. We then seek to calculate the steady pixel value that reflects the unknown irradiance of the scene point which corresponds to the asymptotic of the step response. The rate of the convergence of the pixel value to the steady state response is determined by the *heating time constant*,  $\tau_h$ . The process of pixel heating up can be depicted in Fig. 1(a).

The model denotes the initial measurement time as  $t = t_0$  and uses the “black body” assumption to set the initial value of the IR pixel, i.e.,  $I_m(t_0) = 0$ .

A first-order differential equation model is provided in Eq. 1 to characterize heating an IR pixel.

$$\begin{aligned} \tau_h \frac{dI_m(t)}{dt} + I_m(t) &= I_{ss} \\ I_m(t_0) &= 0 \end{aligned} \quad (1)$$

where  $I_m(t)$  is the pixel intensity value measured by the camera sensor at time  $t$ , and  $I_{ss}$  is the steady state pixel intensity value if the microbolometers were given sufficient heating time.

Let  $t_e = t_1 - t_0$  denote the exposure time, meaning that the shutter is open from the beginning time  $t_0$  to time  $t_1$ . Solving the first-order differential equation at  $t = t_1$  gives

$$I_{ss} = \frac{I_m(t_0 + t_e)}{1 - e^{-\frac{t_e}{\tau_h}}} \quad (2)$$

By modeling the heating process of a microbolometer, we find the value of the forced response at a steady state due to the excitation of the pixel from the associated portion of the 3D scene.

#### 3.2. Model for Pixel Cooling

Cooling the IR pixel is modeled as the natural response of the same differential equation after the heating forcing function,  $\mu(t)$ , has been set to zero. We then seek to calculate the measured pixel value at the time  $t = t_0 + T$  where  $T$  is the time of each frame. This pixel value will

then contribute as a non-zero initial condition for the subsequent image at time  $t_0 + T$ . The decay rate of the pixel value is determined by the *cooling time constant*,  $\tau_c$ . The process of pixel cooling is depicted in Fig. 1(a).

The model denotes the excitation of the pixel at the time that the forcing function is removed as  $t = t_1$  when the shutter is closed, and uses the value of the measured pixel at  $t = t_1$  as the initial value of the IR pixel, i.e.,  $I_m(t_1) = I_0$ .

Similarly, another first-order approximation is used to describe the cooling process.

$$\begin{aligned} \tau_c \frac{dI_m(t)}{dt} + I_m(t) &= 0 \\ I_m(t_1) &= I_0 \end{aligned} \quad (3)$$

Let  $t_r = t_0 + T - t_1$  denote the sensor readout time when the shutter is closed. Solving this first-order differential equation at time  $t = t_0 + T$  gives

$$I_m(t_0 + T) = I_0 e^{-\frac{t_r}{\tau_c}} \quad (4)$$

By modeling the cooling process of a microbolometer, we find the pixel value at the end of each frame period, which is also the initial measurement for the next frame.

### 3.3. Complete Video Sensing Model

We consider sensors that record images at a framerate of  $f_s$  or equivalently having a temporal sample period  $T = \frac{1}{f_s}$ . The time interval between each frame,  $T$ , is further subdivided into a measurement or exposure time during which the shutter is open,  $t_e$ , and a readout time during which the shutter is closed,  $t_r$ . During the exposure time period, the microbolometer is heated. During the period that the sensor reads out the pixel values, the microbolometer is cooling. Fig. 1(b) shows the pixel value convergence when the microbolometer is heating and the heat residual from the previous frames left on the new frame when the microbolometer is cooling. We seek to calculate the steady state pixel value excited by a scene point with the prior heat residual removed. Assuming that the effect from previous frames is dominated by the most recent prior frame when there are no drastic temperature gradients in the scenes, the earlier frames are ignored in the model. The complete video sensing model in Eq. 5 merges these two models into a comprehensive model for the pixel response,  $I'_i(t_e, t_r)$  at frame  $i$  while recording a time-sequence of images.

$$I'_i(t_e, t_r) = \frac{I_i - I_{i-1} \left( e^{-\frac{t_{r,i-1}}{\tau_c}} \right)}{\left( 1 - e^{-\frac{t_{e,i}}{\tau_h}} \right)} \quad (5)$$

where  $I_{i-1}$  and  $I_i$  are two continuous frame,  $t_{r,i-1}$  is the readout time in the previous frame  $i - 1$ , and  $t_{e,i}$  is the exposure time of the current frame  $i$ .

By combining the heating model and cooling model, the irradiance, or the temperature of the measured scene point,

can be more accurately reflected by the stable pixel value  $I'_i$ .

In the remainder of this paper,  $I_i$  will always refer to the photometrically corrected image  $I'_i$ , except where otherwise stated.

### 3.4. IR Sensor-Based Structure From Motion (SfM)

In DSO the authors develop advanced photometric correction models for both camera lens and RGB pixel sensing compensation. This is coupled with camera calibration data to perform highly-accurate SfM at real-time rates with impressive 3D structure reconstruction results. To leverage such a system and apply it to infrared sensors, the brightness transfer model used for RGB sensors in DSO is replaced by our infrared sensor photometric correction model with the following modifications in the SfM algorithm.

- Added a time history (previous sensed values) to tracked pixels.
- Modified the optimization approach to use the derivatives and Hessian of our photometric correction model.

To reconstruct a 3D scene from infrared images using structure from motion, a map is computed to associate two pixels in different frames that both correspond to the same 3D scene point. The photometric error between them is defined in a similar way as [11]. For a point,  $\mathbf{p}$  in reference frame  $I_i$ , observed in target frame  $I_j$ , the photometric error, given by Eq. 6, is formulated as the weighted Sum of Squared Differences (SSD) over a small neighborhood of pixels.

$$\begin{aligned} E_{\mathbf{p}j} &:= \sum_{\mathbf{p} \in \mathcal{N}_{\mathbf{p}}} w_{\mathbf{p}} \| I_j[\mathbf{p}'] - I_{j,o}[\mathbf{p}'] - \beta (I_i[\mathbf{p}] - I_{i,o}[\mathbf{p}']) \|_{\gamma} \\ I_{j,o}[\mathbf{p}'] &= e^{-\frac{t_{r,j-1}}{\tau_c}} \cdot I_{j-1}[\mathbf{p}'] \\ I_{i,o}[\mathbf{p}'] &= e^{-\frac{t_{r,i-1}}{\tau_c}} \cdot I_{i-1}[\mathbf{p}'] \\ \beta &= \frac{1 - e^{-\frac{t_{e,j}}{\tau_h}}}{1 - e^{-\frac{t_{e,i}}{\tau_h}}} \end{aligned} \quad (6)$$

where  $\mathcal{N}_{\mathbf{p}}$  is the set of pixels in the SSD, and  $\| \cdot \|_{\gamma}$  is the Huber norm. In addition to using robust Huber penalties, a gradient-dependent weighting  $w_{\mathbf{p}}$  [11] is applied.

To minimize the photometric error between the corresponding points in two frames, the optimizer then optimizes the heating and cooling time constants  $\tau_c$  and  $\tau_h$ , instead of the brightness transfer variables in DSO. The optimization is accomplished using the Gauss-Newton algorithm in a sliding window [23].



Figure 2. An image pair sample from the FLIR ADAS dataset: IR image (left) and RGB image (right).

## 4. Experiments and Results

In this section, the proposed photometric model for infrared sensors is evaluated on two datasets, the FLIR ADAS Dataset v1.3 [2] and the BU-TIV dataset [30]. Both datasets contain RGB and thermal images for the same scene. The FLIR dataset contains a video sequence of cameras mounted on a vehicle moving in an area during nighttime while the BU-TIV dataset contains a video sequence of a daytime street scene recorded by stationary cameras. The experimental results are obtained on a 32-core Intel Xeon Silver 4110 CPU.

### 4.1. Evaluations on FLIR Dataset

The FLIR ADAS Dataset consists of a video sequence of images taken from an IR camera and an RGB camera mounted to the front of a vehicle. The dataset was acquired with an RGB and IR camera mounted on a vehicle (car) where the IR sensor was a Teledyne FLIR Tau 2 thermal camera and the RGB a Teledyne FLIR Blackfly camera. Both RGB and IR videos were recorded at 30 frames per second (fps) under generally clear conditions during the night. Experiments use a subset of the complete ADAS dataset that corresponds to a video sequence of 600 images where the vehicle drives straight down a road at night. Example images from this video sequence are shown in Fig. 2. The results are summarized in two experiments:

- Evaluations on the reconstruction quality of the road show that our photometric correction enables DSO to track more points on IR data and improve the accuracy of reconstruction.
- Evaluations on the camera motion show that with the proposed correction model, the trajectory is more stable and less deviated in terms of being at a certain distance away from the RGB camera trajectory.

#### 4.1.1 Evaluation Metrics

Our photometric correction algorithm for IR pixel value correction was integrated into the code for the DSO algo-

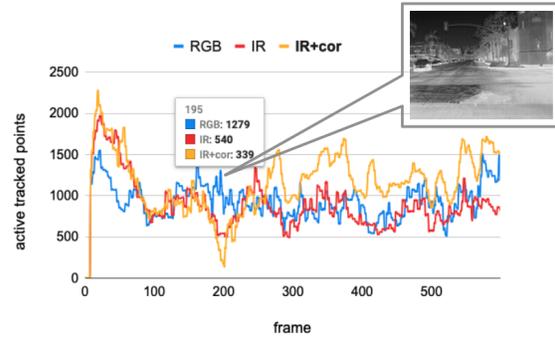


Figure 3. The number of tracked features is plotted for each video frame for “RGB” (blue), “IR” (red), and “IR+cor” (yellow) SfM estimates. The plot shows that the proposed IR photometric correction allows more stable feature detection which results in more tracked features leading to denser SfM estimates.

riethm [11] as a representation of a state-of-the-art SfM algorithm for RGB image sensors. Experiments are performed using RGB and IR image data as input to the DSO algorithm. We consider 3 outputs: (1) the SfM estimates using the input RGB images, referred to as “RGB”, (2) the SfM estimates using the input IR images without the proposed correction, referred to as “IR”, and (3) the SfM estimates using the IR input images with the proposed correction, referred to as “IR+cor”.

As the only public infrared dataset containing certain data (a series of images taken from different viewpoints) that can be used for solving SfM problems, the FLIR ADAS dataset, however, has some limitations that pose challenges to our experiments: (1) the exposure time information of each frame is not available; (2) the 3D scan of the scene is not provided, and (3) the ground truth of the camera pose is not provided. Missing the accurate exposure time can undermine the advantage of the proposed photometric correction model for IR sensors. The lack of 3D scan and camera pose ground truth makes it difficult to analyze the exact improvements brought by the correction model.

To overcome these drawbacks, based on the appearance of the objects shown in the dataset, for example, the street appears to be flat and the vehicle was driving toward one direction on the same lane, three hypotheses are made for our experiments:

- **Exposure Time Hypothesis:** The exposure time of each frame is around 10 milliseconds.
- **Planar Road Hypothesis:** The road where the video was recorded can be approximated as a planar surface.
- **Straight-line Trajectory Hypothesis:** The trajectory of the camera, when the vehicle is driving straight on a road, can be approximated as a line.

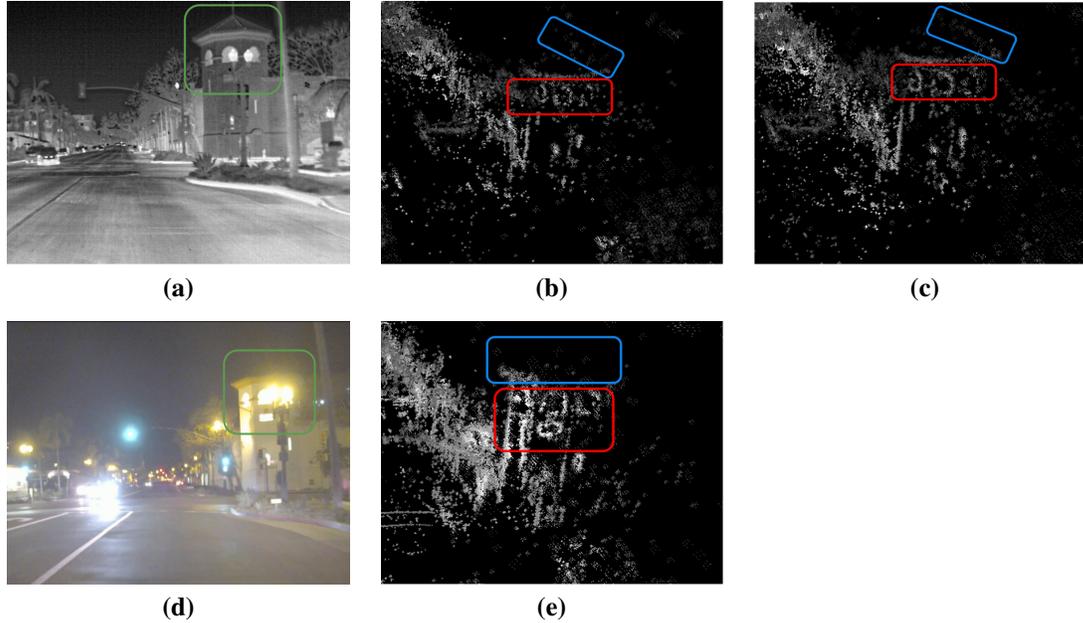


Figure 4. **Qualitative examples.** (a) Infrared image from the FLIR ADAS Dataset. (b) “*IR*” point cloud reconstruction. (c) “*IR+cor*” point cloud reconstruction. (d) RGB image of the same scene from the dataset. (e) “*RGB*” point cloud reconstruction.

According to the FLIR Tau2 camera document [3], the exposure time of the Tau2 camera is measured to be about 10 milliseconds. The value is then used as an approximation of the exposure time for each frame in the dataset. The Planar Road Hypothesis allows us to define a plane that fits the structure of the road to serve as a ground truth for the reconstruction quality evaluation. The Straight-line Trajectory Hypothesis provides an opportunity to evaluate the accuracy of the camera pose by looking into the deviation of the trajectory from the line.

#### 4.1.2 Structure Estimate Evaluation

The Planar Road Hypothesis in Section 4.1.1 is used to evaluate the structure reconstruction accuracy. The “ground truth” of the road is defined by: (1) segmenting the road area within the point cloud, (2) sampling from each point cloud the **same** amount of points located within the region around the road surface, and (3) merging the points from all three point clouds and fitting a plane to the combined road points. This way a common reference of the road surface is available for evaluation.

The “kfs” and “total pts” rows of Table 1 show that the proposed IR photometric correction model allows for more scene points to be tracked for the SfM estimation algorithm. The “kfs” row denotes the total number of keyframes for the image sequence and the “total” pts row indicates the total number of tracked features for the image sequence. As shown in the table, the SfM algorithm using the IR correc-

	RGB	IR	<b>IR+cor</b>
kfs	219	181	<b>246</b>
total pts	55562	42411	<b>65229</b>
road pts	1395	1304	<b>1890</b>
RMSE	0.0153	0.0137	<b>0.0119</b>
std.	0.0088	0.0073	<b>0.0065</b>

Table 1. Statistics of the road reconstruction performance.

tion tracks 53.8% more points (total pts) and 35.9% more keyframes (kfs) than IR input images without photometric correction. Similarly, the SfM algorithm using IR correction tracks 17.4% more points and 12.3% more keyframes than RGB input images using the RGB image photometric correction. We also note that similar numbers are found for the point clouds identified as inside the segmented region around the road surface (road points). These results indicate that the IR photometric correction allows more points to be tracked and the resulting SfM solution, therefore, yields a denser 3D point cloud for both the camera motion trajectory (number of keyframes) and the scene structure.

The number of actively tracked features over the 600-frame video sequence is plotted in Fig. 3. Actively tracked points are used for both camera motion and scene structure estimation as each new frame in the video is measured. Overall running the “*IR+cor*” enables more points for tracking while both the “*IR*” and the “*RGB*” track fewer points but are comparable to each other. An interesting undulation

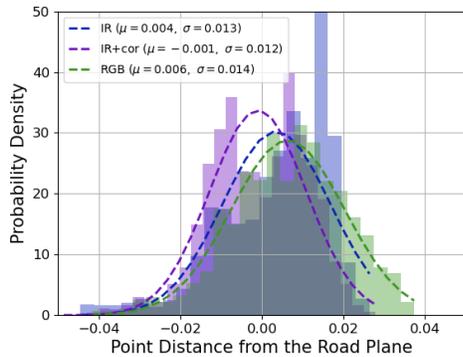


Figure 5. Histogram and approximated Gaussian distributions of the fitting error of SfM-estimated 3D points to a common planar surface (the road plane). RGB (green) and IR (blue) show more error variance than the proposed IR photometric correction approach (purple).

of the curves shown in the figure occurs at frame index 200. This corresponds to a sequence of images within a large two-way street intersection. The RGB image sensor can track better in this particular context due to rich structural appearance data provided by the white lines and cross-walk textures on the ground which have little thermal contrast. This explains the decrease in tracked points for frames 150-220 from the IR image sensor.

To evaluate the accuracy of the structure reconstruction identical sections of the estimated reconstruction data in the vicinity of the road surface were segmented from the complete structure estimate. In each case, the segmented surface points were compared against a pre-defined road plane. Evaluation of performance was done by computing the Root-Mean-Square deviation (RMSE) and the standard deviation (std) of the perpendicular distance between reconstructed 3D points and the road surface. As presented in Table 1, “*IR+cor*” improves the reconstruction accuracy of the road by 15.1% over the “*IR*” approach and 28.5% over the “*RGB*” approach. The road points detected in “*IR+cor*” exhibit less noise relative to the road plane model having 10.9% and 26.1% less deviation than the results of the “*IR*” and “*RGB*” methods respectively. These results indicate that the proposed IR photometric correction model reduces the noise in the estimated scene structure.

This conclusion can be further supported by the histogram of the fitting error and error distribution in Fig. 5, from which we can also see that DSO on the IR data with infrared photometric correction achieves the best performance in terms of accuracy (closest to zero mean value) and stability (least standard deviation). Although RGB data can sometimes provide more features for detection and tracking (more tracked points), the reconstruction quality is not as well as the results from infrared data. This can be because

the RGB intensity values are very prone to bad illumination conditions such as at night, which was the case when the dataset was recorded.

### 4.1.3 Motion Estimate Evaluation

Performance analysis for the motion estimates uses a Straight-line Trajectory Hypothesis in Section 4.1.1 for the vehicle motion and examines the observed errors of the estimated camera motion for each approach. In the video, the vehicle where the two cameras were mounted was driving on a straight street, as a result of which the trajectory of the cameras can be approximated to be straight as well. Towards this end, a 3D line segment is fit to the positions of the estimated camera trajectories for three approaches: (1) *IR*, (2) *IR+cor*, and (3) *RGB*. Table 2 shows the RMSE (root-mean-square error) and standard deviation (std) of the perpendicular distances between camera positions and the estimated 3D line model.

	RGB	IR	<b>IR+cor</b>
RMSE	0.0128	0.0109	<b>0.0106</b>
std.	0.0061	0.0049	<b>0.0045</b>

Table 2. Statistics of the trajectory estimation performance.

The RMSE error row of Table 2 indicates that the camera motion estimates for the infrared data using the proposed photometric correction model exhibit less error relative to the line model by a factor of 2.8% for the IR method and 17.2% for the RGB method. The standard deviation (std.) row of Table 2 indicates that the variability in the camera motion for the IR photometric correction is also less than that for the other two approaches. The reduction in noise for both IR estimates relative to the RGB data suggests that IR image data in this low-light context may provide advantages over RGB data for SfM estimation and that the proposed photometric correction further improves the SfM estimation performance in accuracy and stability.

### 4.1.4 Qualitative Analysis

The point cloud reconstructions from the IR and RGB data are illustrated in Fig. 4 respectively. The green box in both Fig. 4(a) and Fig 4(d) includes the building of our interest. The red and blue boxes in the other three figures highlight the area where the reconstruction results differ. Compared to Fig. 4(b), Fig. 4(c) shows sharper edges on the windows in the building (red box) and contains more points to reflect the edge of the top of the building (blue box). This indicates that the photometric correction model we propose can enable SfM algorithms to track more features (points) for reconstruction and the features are less noisy and more stable. As shown in Fig. 4(e), the top of the building (blue box) is

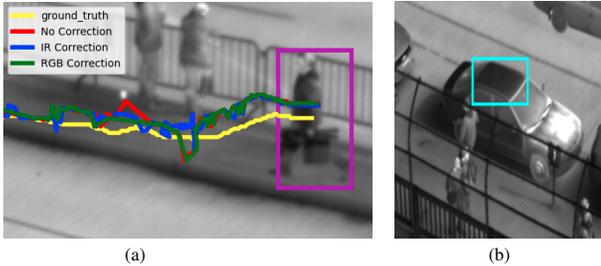


Figure 6. Excerpts from the IR video in the BU-TIV dataset. (a) shows pedestrian (magenta box) tracking results relative to the ground truth (yellow), for the proposed correction (blue), RGB correction (green), and no correction (red). Table 3 provides quantitative performance metrics. (b) shows the roof of the parked car (cyan box) where the temperature is tracked.

completely missing and unrecognizable, and the point cloud of the windows in the building is very noisy (red box). This is because the top area of the building is interfered with by the illumination from the street light and RGB sensors can easily suffer from such illumination conditions and will fail to detect reliable features. The infrared sensors, however, are more robust in this scenario as they are sensitive to thermal contrast instead of photons. Note that the example image here comes from the same intersection area in the street as the image illustrated in Fig. 3, the point clouds here further prove that in this area of the street, more active points are tracked in the RGB data than the IR data due to the fact that rich structure but little thermal contrast are available in this area, as explained previously in Section 4.1.2

## 4.2. Evaluations on BU-TIV Dataset

To show that the proposed photometric correction model can also be applied to other IR image-processing contexts, the proposed photometric correction model is applied to the BU-TIV dataset for solving a human being tracking problem. The BU-TIV dataset is designated for the object-tracking problem using infrared data and consists of video sequences in different scenes that were recorded with FLIR SC8000 cameras. A subset of the dataset of a daytime crowded street view during a marathon competition was used. The results are summarized in two experiments: (1) Experiment 1 tracks pedestrians in IR image sequences. (2) Experiment 2 tracks the observed temperature for a stationary target in the IR video.

Experiment 1 evaluates the proposed photometric correction for pedestrian tracking from an IR image sequence. Results are shown in Fig. 6(a) for three distinct photometric correction approaches: (1) RGB correction (*RGB*), (2) no correction (*IR*), and (3) the proposed correction (*IR+cor*). Table 3 shows the quality of each estimate as measured by three distinct metrics: (1) the Discrete Fréchet (DF) distance [9], (2) the Dynamic Time Warping (DTW) [5], and

	RGB	IR	IR+cor
DF [9]	23.77	39.82	<b>22.84</b>
DTW [5]	4444.43	4339.27	<b>4257.07</b>
Mean Distance	8.81	8.82	<b>8.51</b>
Tracked frames count	496	537	<b>555</b>

Table 3. Tracking differences measured by different algorithms between the estimated trajectories and ground truth.

(3) a custom metric referred to as the Mean Distance. Mean distance calculates the average distance between the person’s estimated and actual positions for all corresponding frames. Table 3 indicates that the proposed IR correction improves the performance across all three metrics, where lower scores are better. The last row shows the number of frames where the person is tracked and again the proposed IR correction algorithm outperforms the other cases.

Experiment 2 tracks the temperature of a parked car roof as shown in Fig. 6(b) and seeks to analyze the stability of the temperature before and after applying the proposed IR photometric correction. Table 4 shows that the proposed photometric correction improves the stability of pixel intensity and, by extension, the estimate of the unknown constant temperature of the observed object.

	RGB	IR	IR+cor
std.	0.884	0.918	<b>0.877</b>

Table 4. Standard deviation of observed intensities for the roof of a parked car denoted as a cyan box in Fig. 6(b) over time.

## 5. Conclusion

This article proposes a photometric correction model appropriate for microbolometer sensors typically integrated into infrared cameras. A new theoretical model for the pixel response is proposed and the parameters of this model are characterized. The photometric correction model was integrated into a state-of-the-art SfM algorithm where it was shown to improve upon the structure and camera motion estimates. Prior literature has made clear that photometric correction is an important component in improving the performance of SfM for RGB sensors and the results of this article indicate that appropriate models for infrared photometric correction also improve estimates in the infrared frequency regime. We hypothesize that the impact of the proposed infrared photometric correction further generalizes to potentially improve other computer vision algorithms when applied to IR image sensors, particularly those with the “view invariance” or Lambertian assumption for the radiance of scene points over short motion distances.

## References

- [1] Are Boson, Tau 2, and Quark 2 and Lepton rolling shutter cameras or framing cameras? [https://flir.custhelp.com/app/answers/detail/a\\_id/3090/related/1](https://flir.custhelp.com/app/answers/detail/a_id/3090/related/1). Accessed: 2022-05-31. **2**
- [2] TELEDYNE FLIR Dataset. <https://flir.box.com/s/suwst0b3k9rko35homhr3rnyytf3102d>. Accessed: 2022-05-31. **5**
- [3] Time Constant Design of Tau2 and Quark2. [https://flir.custhelp.com/app/answers/detail/a\\_id/3171/related/1](https://flir.custhelp.com/app/answers/detail/a_id/3171/related/1). Accessed: 2022-11-06. **6**
- [4] Steven S. Beauchemin and John L. Barron. The computation of optical flow. *ACM computing surveys (CSUR)*, 27(3):433–466, 1995. **1**
- [5] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, USA., 1994. **8**
- [6] N Boudou, A Durand, S Tinnes, S Cortial, A Gorecki, M Cueff, and A Viro. Ulis bolometer improvements for fast imaging applications. In *Infrared Technology and Applications XLV*, volume 11002, pages 366–375. SPIE, 2019. **1**
- [7] Calvin Yi-Ping Chao, Honyih Tu, Thomas Wu, Kuo-Yu Chou, Shang-Fu Yeh, and Fu-Lung Hsueh. Cmos image sensor random telegraph noise time constant extraction from correlated to uncorrelated double sampling. *IEEE Journal of the Electron Devices Society*, 5(1):79–89, 2017. **2**
- [8] Manash Pratim Das, Larry Matthies, and Shreyansh Daftry. Online photometric calibration of automatic gain thermal infrared cameras. *IEEE Robotics and Automation Letters*, 6(2):2453–2460, 2021. **2**
- [9] Thomas Eiter and Heikki Mannila. Computing discrete fréchet distance. 1994. **8**
- [10] Erika Emilsson and Joakim Rydell. Chameleon on fire—thermal infrared indoor positioning. In *2014 IEEE/ION Position, Location and Navigation Symposium-PLANS 2014*, pages 637–644. IEEE, 2014. **2**
- [11] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):611–625, 2017. **2, 4, 5**
- [12] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *European conference on computer vision*, pages 834–849. Springer, 2014. **3**
- [13] Jakob Engel, Jörg Stückler, and Daniel Cremers. Large-scale direct slam with stereo cameras. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1935–1942, 2015. **2**
- [14] Jakob Engel, Vladyslav Usenko, and Daniel Cremers. A photometrically calibrated benchmark for monocular visual odometry. *arXiv preprint arXiv:1607.02555*, 2016. **1, 2**
- [15] P. Favaro, Hailin Jin, and S. Soatto. A semi-direct approach to structure from motion. In *Proceedings 11th International Conference on Image Analysis and Processing*, pages 250–255, 2001. **2**
- [16] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. Svo: Fast semi-direct monocular visual odometry. In *2014 IEEE international conference on robotics and automation (ICRA)*, pages 15–22. IEEE, 2014. **3**
- [17] Robert Grimmig, Bruce McIntosh, Abhijit Mahalanobis, and Ronald G Driggers. Lwir sensor parameters for deep learning object detectors. *OSA Continuum*, 4(2):529–541, 2021. **1**
- [18] Caroline Hyll. *Infrared Emittance of Paper: Method Development, Measurements and Application*. PhD thesis, KTH Royal Institute of Technology, 2012. **1**
- [19] Shehryar Khattak, Christos Papachristos, and Kostas Alexis. Keyframe-based direct thermal-inertial odometry. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3563–3569. IEEE, 2019. **2**
- [20] Shehryar Khattak, Christos Papachristos, and Kostas Alexis. Keyframe-based thermal-inertial odometry. *Journal of Field Robotics*, 37(4):552–579, 2020. **2**
- [21] Sungho Kim, Woo-Jin Song, and So-Hyun Kim. Infrared variation optimized deep convolutional neural network for robust automatic ground target recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8, 2017. **1**
- [22] Margaret Kohin and Neal R Butler. Performance limits of uncooled vox microbolometer focal plane arrays. In *Infrared Technology and Applications XXX*, volume 5406, pages 447–453. SPIE, 2004. **1**
- [23] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. Keyframe-based visual-inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3):314–334, 2015. **4**
- [24] Tarek Mouats, Nabil Aouf, Lounis Chermak, and Mark A Richardson. Thermal stereo odometry for uavs. *IEEE Sensors Journal*, 15(11):6335–6347, 2015. **2**
- [25] Nicolas Pinchon, Olivier Cassignol, Adrien Nicolas, Frédéric Bernardin, Patrick Leduc, Jean-Philippe Tarel, Roland Brémond, Emmanuel Bercier, and Johann Brunet. All-weather vision for automotive safety: which spectral band? In *International Forum on Advanced Microsystems for Automotive Applications*, pages 3–15. Springer, 2018. **1**
- [26] Arturo Rankin, Andres Huertas, Larry Matthies, Max Bajracharya, Christopher Assad, Shane Brennan, Paolo Bellutta, and Gary W Sherwin. Unmanned ground vehicle perception using thermal infrared cameras. In *Unmanned Systems Technology XIII*, volume 8045, pages 19–44. Spie, 2011. **1**
- [27] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. **2**
- [28] Jae Kyu Suhr. Kanade-lucas-tomasi (klt) feature tracker. *Computer Vision (EEE6503)*, pages 9–18, 2009. **1**
- [29] Ming Wilson. Temperature measurement. *Anaesthesia & Intensive Care Medicine*, 22(3):202–207, 2021. **2**
- [30] Zheng Wu, Nathan Fuller, Diane Theriault, and Margrit Betke. A thermal infrared video benchmark for visual analysis. In *CVPR PBVS Workshop*, pages 201–208, 2014. **5**