

Pixel-Wise Motion Deblurring of Thermal Videos

Manikandasriram S.R.¹, Zixu Zhang¹, Ram Vasudevan², and Matthew Johnson-Roberson³

Robotics Institute¹, Mechanical Engineering², Naval Architecture and Marine Engineering³

University of Michigan, Ann Arbor, Michigan, USA 48109.

{srmani, zixu, ramv, mattjr}@umich.edu

<https://fcav.engin.umich.edu/projects/thermal-vision>

Abstract—Uncooled microbolometers can enable robots to see in the absence of visible illumination by imaging the “heat” radiated from the scene. Despite this ability to see in the dark, these sensors suffer from significant motion blur. This has limited their application on robotic systems. As described in this paper, this motion blur arises due to the *thermal inertia* of each pixel. This has meant that traditional motion deblurring techniques, which rely on identifying an appropriate spatial blur kernel to perform spatial deconvolution, are unable to reliably perform motion deblurring on thermal camera images. To address this problem, this paper formulates reversing the effect of thermal inertia at a single pixel as a Least Absolute Shrinkage and Selection Operator (LASSO) problem which we can solve rapidly using a quadratic programming solver. By leveraging sparsity and a high frame rate, this pixel-wise LASSO formulation is able to recover motion deblurred frames of thermal videos without using any spatial information. To compare its quality against state-of-the-art visible camera based deblurring methods, this paper evaluated the performance of a family of pre-trained object detectors on a set of images restored by different deblurring algorithms. All evaluated object detectors performed systematically better on images restored by the proposed algorithm rather than any other tested, state-of-the-art methods.

I. INTRODUCTION

Robots need to operate in a variety of lighting and weather conditions. Since individual sensing modalities can suffer from specific shortcomings, robotic systems typically rely on multiple modalities [29]. Visible cameras, for instance, often fail to detect objects hidden in shadows, in poor lighting conditions such as those that arise at night, or those behind sun glare [25]. LIDARs, on the other hand, suffer from spurious returns from visual obscurants such as snow, rain, and fog [6].

To address these sensor limitations, one can apply thermal infrared cameras that operate in the Long Wave Infrared (LWIR, $7.5\mu\text{m} - 14\mu\text{m}$) region [12]. These cameras capture the thermal blackbody radiation emitted by all object surfaces at “*earthly*” temperatures and can therefore operate even in the complete absence of visible light. Note that these thermal cameras are different from night vision cameras that either operate in Near-Infrared (NIR, $0.75\mu\text{m} - 1.4\mu\text{m}$) or use image intensifiers in the visible spectrum [5]. Unlike visible and night-vision cameras that rely on reflected light from a few light sources, thermal cameras depend primarily on emitted radiation, which eliminates shadows and reduces lighting artifacts. The longer wavelengths used by these cameras also allow them to see better through visual obscurants [23].

Despite their potential applicability, these sensors suffer from significant motion blur, as illustrated in Fig. 1. As a



Fig. 1: An illustration of the proposed motion deblurring algorithm for microbolometers. The top image shows a visible image captured at 30fps with auto exposure. The middle image shows a thermal image captured at 200fps which suffers from significant non-uniform motion blur. As described in this paper, the lack of exposure control and the difference in the physics of image formation aggravates the image degradation due to motion blur in thermal images. The bottom image shows the result from our algorithm that processes each pixel independently. Our model-based approach is able to eliminate the blur without any visible artefacts.

result, the research community has focused on scenarios where the relative magnitudes of camera motion, scene geometry and camera resolution can be restricted to limit motion blur [19, 9, 15, 33]. Unfortunately, such restrictions can be difficult to accommodate during mobile robot applications. In addition, the inability to control exposure time and the lack of a global shutter make microbolometers even more challenging to use.

To address these motion blur problems, one can rely on another class of thermal cameras called cooled photon detectors [28]. These sensors use the photoelectric effect, i.e. the same physics underlying CCD and CMOS sensors, which enables controllable exposure time and global shutter and, as a result, limits motion blur. However, these cooled photon detectors are prohibitively expensive since they require cooling the sensor to 77K to eliminate self-emissions [4]. Therefore they are used only in high-end applications.

Uncooled microbolometers, on the other hand, due to their affordability, low weight and low power consumption, have brought thermal imaging to consumer applications. As a result, they are the predominant type of thermal cameras used by the robotics and vision research community today [12]. These sensors use the bolometer principle, where the incoming

radiation heats up a detector material that has a temperature dependent electrical resistance [34]. Unlike the photoelectric effect, temperature changes take a non-negligible amount of time, which leads to a lag in these cameras with respect to changes in the scene. This paper illustrates that this lag, which we refer to as *thermal inertia*, gives rise to motion blur in thermal images. For applications with continuous motion, such as autonomous vehicles, the sensors never reach steady state. The measured temperatures might then be incorrect, making these images an inaccurate representation of the scene. Unfortunately, designing microbolometers with lower thermal inertia without increasing sensor noise remains a significant technological challenge [7]. Therefore, motion deblurring algorithms are required to accurately recover the instantaneous scene representation from the data measured by these microbolometers.

There has been limited success in developing general motion deblurring algorithms for microbolometers [34]. Motion blur is typically modelled as a latent sharp image convolved with an unknown blur kernel, which can vary across the image. As a result, motion deblurring requires estimating the blur kernel and using deconvolution to recover the latent sharp image. For instance, Oswald-Tranta [22] construct an exponential blur kernel using the thermal time constant for uniform motions of a target object against a constant background. Nihei et al. [21] use an Inertial Measurement Unit (IMU) to estimate the camera motion and provide an approximate correction equation to recover the steady state value of a target pixel. In addition, one could consider applying algorithms developed for visible cameras to thermal images [24]. For example, Yan et al. [38] formulate deblurring as a maximum *a posteriori* problem that alternatively optimizes the latent image and the blur kernel with an additional image prior, which promotes sparsity of image gradients. Whyte et al. [36] consider non-uniform blurring caused due to 3D rotation of the camera against a static scene. Bahat et al. [3] estimate per-pixel blur kernels to accommodate dynamic objects in the scene. More recently, there have been a number of learning based methods proposed for single-image deblurring [20, 30, 14] and multi-image/video deblurring [39, 35].

This paper diverges from these conventional deblurring approaches that perform kernel estimation and subsequent latent image recovery, and instead focuses on devising a method that can correct for the microbolometer-specific cause for motion blur. In particular, we show that the lag introduced by the bolometer principle at each individual pixel directly gives rise to the motion blur. By reversing this effect of thermal inertia pixel-wise without any spatial constraints, from a sequence of high frame rate measurements, we recover motion deblurred images without explicitly modeling the motion of the camera, scene depth, or scene dynamics. To this end, we construct an under-determined system of linear equations, which has infinitely many solutions, by discretizing at a high rate the differential equation that models the thermal image formation process. We then leverage temporal sparsity to select a unique solution while being robust to measurement noise by using the

Least Absolute Shrinkage and Selection Operator (LASSO) and solve it using quadratic programming solvers. Our formulation works across a variety of scenes with arbitrary 6D relative motions between the camera and the dynamic objects within the scene. In addition, our method is uniquely suited for robotic vision tasks where an image is often represented by a sparse set of keypoints. While existing approaches at-least require deconvolution over small patches covering the sparse keypoints, our proposed approach just requires the data from the sparse pixels of interest.

The key contributions of our paper are summarized below:

- In Section III, we show that reversing the effect of thermal inertia at an independent pixel can be formulated as solving an under-determined system of linear equations.
- In Section IV, we show that assuming temporal sparsity of the signal pixel-wise is sufficient to recover motion deblurred frames from thermal videos using LASSO without requiring any training data.
- In Section V, we illustrate that our model based approach outperforms state-of-the-art learning based single-image and multi-image deblurring algorithms. In particular, we show that a variety of object detectors performed systematically better when using the images deblurred by our proposed algorithm rather than any of the five tested, state-of-the-art methods.

The remainder of our paper is organized as follows. Section II reviews the origins of motion blur in visible cameras and microbolometers based on their physics of image formation. Our key contributions are presented in Sections III, IV and V. Finally, Section VI concludes the paper with a brief discussion on the insights offered by our proposed framework.

II. IMAGE FORMATION MODELS

This section reviews the underlying physics of image formation to identify the cause of motion blur as summarized in Fig. 2. In particular, we focus on two distinct Image Formation Models (IFM): the Photoelectric IFM for visible cameras and the Microbolometer IFM for thermal cameras. We present both models to emphasize the distinct reasons that give rise to motion blur in each IFM.

A. Photoelectric Image Formation Model

In Photoelectric IFM (P-IFM), the sensor is exposed for a finite duration to incoming Electro-Magnetic (EM) radiation from the scene and the number of photons “collected” is digitized into an image. The registers are reset to zero before the next frame is captured.

Let $\Phi(p)$ denote the total power from a point p in the scene and T_e denote the exposure time. In the case of global shutter and a linear camera response function, the value of pixel (i, j) can be written as

$$I(i, j) = \frac{1}{T_e} \int_0^{T_e} \Phi(p_{i,j}(s)) ds, \quad (1)$$

where $p_{i,j}(s)$ is the scene point being imaged by pixel (i, j) at time s . In the absence of relative motion during exposure, $p_{i,j}$

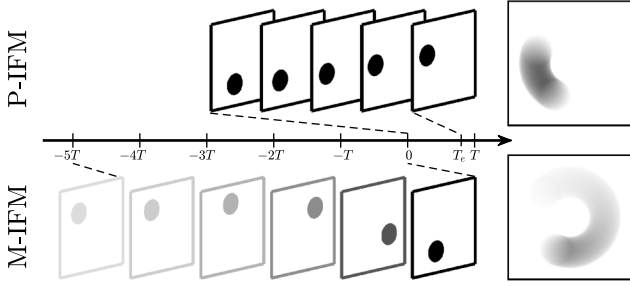


Fig. 2: An illustration of motion blur in photoelectric sensors (top) and microbolometers (bottom). Each frame on the left represents a time slice of instantaneous power from the scene. P-IFM averages over time slices during exposure while M-IFM gives more weight to recent frames as depicted by the transparency in this figure. The blurred images on the right were generated by pixelwise average for P-IFM and by simulating a microbolometer with $\tau = T$ for M-IFM. Note that the blur looks symmetric for P-IFM since each slice is equally weighted while the motion blur in M-IFM appears as a blur-tail.

corresponds to a single scene point and the integrand above simplifies to a constant.

When there is relative motion during exposure, $p_{i,j}$ instead corresponds to a sequence of scene points, resulting in motion blur. To describe this more explicitly, let $\delta s \in \mathbb{R}$ with $\frac{T_e}{\delta s} \in \mathbb{N}$, and suppose we divide the total exposure time into $\frac{T_e}{\delta s}$ intervals each of length δs , such that for each $k \in \{0, \dots, \frac{T_e}{\delta s} - 1\}$, $p_{i,j}(s) = p_{i,j}^k$ is constant for all $s \in [k\delta s, (k+1)\delta s)$. Then

$$I(i, j) = \frac{1}{T_e} \sum_{k=0}^{\frac{T_e}{\delta s} - 1} \Phi(p_{i,j}^k) \delta s. \quad (2)$$

The objective of motion deblurring algorithms is to recover an image of the scene at the start of exposure or equivalently $\Phi(p_{i,j}^0)$. Over the duration of exposure, the relative motion causes different pixels to image a single scene point. Therefore, the scene point imaged by (i, j) at the k -th interval would have been imaged by pixel (x, y) at the start of exposure. In other words, every k -th slice can be obtained by warping $\Phi(p_{i,j}^0)$. In the computer vision literature [24], $\Phi(p_{i,j}^0)$ is referred as the latent sharp image, L , and the blurred image is written as 2D convolution

$$I(i, j) = \iint H(i - x, j - y) L(x, y) dx dy, \quad (3)$$

where H is referred to as a Point Spread Function (PSF) or blur kernel. The blur kernel models how much time pixel (i, j) was exposed to the scene point that would have been imaged at (x, y) at the start of the exposure. As a result, this explicitly models the relative motion and scene dynamics.

B. Microbolometer Image Formation Model

In Microbolometer IFM (M-IFM), the sensor is constantly exposed to incoming EM radiation from the scene. The resulting change in the temperature of a pixel changes its electrical resistance. A Read-Out Integrated Circuit (ROIC) measures the current electrical resistance at regular intervals and computes corresponding temperature values to be returned by the camera. When the incoming radiation is continuously changing,

the microbolometer is unable to reach an equilibrium before the ROIC takes the next measurement. Moreover, there is no “reset” mechanism at the end of a frame. The microbolometer can be modeled by [7, 32, 34]

$$\frac{C_{th}}{G_{th}} \frac{d\Delta T}{dt}(t) + \Delta T(t) = \frac{\alpha}{G_{th}} \Phi(p(t)), \quad (4)$$

where C_{th} , G_{th} and α are constants representing the thermal capacitance, thermal conductance, and absorptivity of the microbolometer respectively, with Φ and p as in the previous subsection. $\Delta T(t) = T_m(t) - T_s$ is the difference in temperature at time t between the pixel membrane T_m and a constant substrate temperature T_s .

We can define the steady state temperature of the pixel membrane $T_m^{ss}(p(t))$ corresponding to the power $\Phi(p(t))$ as

$$T_m^{ss}(p(t)) = T_s + \frac{\alpha}{G_{th}} \Phi(p(t)). \quad (5)$$

Inserting in (4), we get

$$\tau \frac{dT_m}{dt}(t) + T_m(t) = T_m^{ss}(p(t)), \quad (6)$$

where $\tau = \frac{C_{th}}{G_{th}}$ is the thermal time constant of the microbolometer. The general solution to (6) can be obtained by integration-by-parts as:

$$T_m(t) = \frac{1}{\tau} \int_{-\infty}^t e^{\frac{s-t}{\tau}} T_m^{ss}(p(s)) ds. \quad (7)$$

Note that (7) is the solution of a first order linear differential equation, and T_m^{ss} is proportional to Φ from (5). As a result, a step change in Φ would require a delay of 5τ before T_m reaches within 1% of the steady state value. This lag in the temperature of the microbolometer T_m with respect to changes in the incoming power Φ is what we call the *thermal inertia* of each pixel.

Every pixel follows the same model and hence the value of pixel (i, j) returned by the camera at time t can be written as

$$I(i, j) = \frac{1}{\tau} \int_{-\infty}^t e^{\frac{s-t}{\tau}} T_m^{ss}(p_{i,j}(s)) ds. \quad (8)$$

Since T_m^{ss} is proportional to Φ , as a result of (5), the integral in (8) can be interpreted as a weighted sum of the incoming power over the history of the pixel with the weight exponentially decaying into the past.

In the absence of relative motion over the last 5τ seconds, the pixel reaches steady state such that $I(i, j) \approx T_m^{ss}(p_{i,j}(t))$. However, when there is relative motion, the value of a pixel depends on $p_{i,j}$, resulting in apparent motion blur. The objective of motion deblurring for thermal images is to recover $T_m^{ss}(p_{i,j}(t))$ for each i and j .

C. Spatial vs Temporal Spread

In P-IFM, the exposure time is divided into small intervals within which the incoming power is described by a single scene point. During motion, a single pixel’s intensity is described by the power of a variety of points in the world. To identify a latent sharp image, one can identify a blur kernel

and apply spatial deconvolution. One could follow a similar approach for (8) wherein the blur kernel would additionally need to model the time dependent exponential weight factor and the time constant. The effective exposure time that needs to be considered would also be larger than the typical values of T_e . In addition, the challenges associated with estimating H to describe the 6D relative motion in dynamic scenes would only further aggravate the problem.

This work takes a different tact. The reset mechanism at the end of a frame in P-IFM renders each frame as an independent measurement of the scene. However, two consecutive frames taken at nT and $(n+1)T$ in M-IFM have a common history $[(n+1)T - 5\tau, nT]$ where T is the sampling period of the camera and is chosen such that $T < 5\tau$. For instance, if $T = \tau$, then 5 consecutive measurements have different portions of $T_m^{ss}(p(nT))$ in it. Suppose we have an algorithm that can undo this temporal spread to recover $T_m^{ss}(p(nT))$ given a set of measurements $T_m(nT)$, then such an algorithm achieves motion deblurring for thermal images without modeling the relative motion of the camera with the scene. The next section poses this inverse problem as selecting a unique signal from the solutions of an under-determined system of linear equations.

III. REVERSING THE EFFECT OF THERMAL INERTIA

To understand how to reverse the thermal inertia in microbolometers, let's consider the following differential equation

$$\tau \frac{dy}{dt}(t) + y(t) = x(t), \quad (9)$$

where τ is a known constant, x is the unknown signal of interest and y is the signal from which we have discrete samples $\{y(t_0), y(t_1), \dots, y(t_N)\}$. Our objective is to recover the signal x within the right open interval $[t_0, t_N)$ that respects (9) and the given samples of y . Note that (9) is equivalent to (6) from previous section.

By integrating (9), we get for $t_0 \leq t \leq t_N$

$$y(t) = y(t_0)e^{-\frac{t-t_0}{\tau}} + \frac{e^{-\frac{t}{\tau}}}{\tau} \int_{t_0}^t e^{\frac{\alpha}{\tau}} x(\alpha) d\alpha. \quad (10)$$

This gives rise to N constraints, one for each $t_i \in \{t_1, \dots, t_N\}$

$$y(t_i) = y(t_0)e^{-\frac{t_i-t_0}{\tau}} + \frac{e^{-\frac{t_i}{\tau}}}{\tau} \int_{t_0}^{t_i} e^{\frac{\alpha}{\tau}} x(\alpha) d\alpha. \quad (11)$$

This is an under-constrained problem as there are many signals x that satisfy (11). Therefore, we proceed by constructing a feasible set of such signals.

Let the indicator function be denoted as

$$\mathbb{I}(x_{\min} \leq x < x_{\max}) = \begin{cases} 1 & \text{if } x_{\min} \leq x < x_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

Consider the class of piece-wise constant signals $f_n : [t_0, t_N) \rightarrow \mathbb{R}$ for $n \in \mathbb{Z}$ defined by

$$f_n(\alpha) = \sum_{k=0}^{K_n-1} a_k \mathbb{I}\left(t_0 + \frac{k\Delta t}{K_n} \leq \alpha < t_0 + \frac{(k+1)\Delta t}{K_n}\right), \quad (13)$$

where $K_n = 2^n$ and $\Delta t = t_N - t_0$. This function can be equivalently represented by the vector $A = [a_0, a_1, \dots, a_{K_n-1}]^T \in \mathbb{R}^{K_n}$. Note that the set of functions that can be represented by vectors drawn from \mathbb{R}^{K_n} is a subset of the set of functions that can be represented by vectors drawn from \mathbb{R}^{K_m} when $m \geq n$.

Next, consider the following important property of the class of functions we have just defined that follows from [11, Theorem 2.26]:

Lemma 1. *Let $g : [t_0, t_N) \rightarrow \mathbb{R}$ be a Lebesgue integrable function. Then for every $\epsilon > 0$, there exists $n \in \mathbb{Z}$ and function f_n as in (13) with $\|g - f_n\|_1 < \epsilon$.*

The above lemma allows us to associate any signal x within $[t_0, t_N)$ with a function f_n within machine precision for a sufficiently large n . This allows us to construct our feasible set of signals as a subset of \mathbb{R}^{K_n} .

Next, we describe how to formulate the set of N constraints in (11) as a set of linear equations:

Theorem 1. *For each $m \in \mathbb{Z}$, consider f_m as in (13) with a vector of coefficients $A_{f_m} \in \mathbb{R}^{K_m}$. If f_m satisfies (11), then A_{f_m} satisfies the following equation:*

$$Y = VA_{f_m}, \quad (14)$$

where $Y \in \mathbb{R}^N$ and $V \in \mathbb{R}^{N \times K_m}$ are defined as

$$[Y]_j = y(t_{j+1}) - y(t_0)e^{-\frac{t_{j+1}-t_0}{\tau}} \quad (15a)$$

$$[V]_{j,k} = e^{-\frac{t_{j+1}-t_0}{\tau}} (e^{\gamma_{\max}} - e^{\gamma_{\min}}) \mathbb{I}(\gamma_{\min} < \gamma_{\max}), \quad (15b)$$

where $0 \leq j < N$, $0 \leq k < K_m$, the subscripts outside the square-brackets represent the corresponding component in the matrix, and

$$\gamma_{\min} = \max\left\{\frac{k}{K_m} \frac{\Delta t}{\tau}, 0\right\} \quad (16a)$$

$$\gamma_{\max} = \min\left\{\frac{k+1}{K_m} \frac{\Delta t}{\tau}, \frac{t_{j+1}-t_0}{\tau}\right\}. \quad (16b)$$

Proof: Following Lemma 1, there exists n such that we can substitute for x in (10) and using change of variables $\beta = \frac{\alpha-t_0}{\Delta t}$ and using $s = t - t_0$, we get

$$y(t) = y(t_0)e^{-\frac{s}{\tau}} + \frac{\Delta t}{\tau} e^{-\frac{s}{\tau}} \int_0^{\frac{s}{\Delta t}} e^{\frac{\beta\Delta t}{\tau}} f_n(t_0 + \beta\Delta t) d\beta. \quad (17)$$

Substituting for f_n , the second term in the above equation becomes

$$\sum_{k=0}^{K_n-1} a_k e^{-\frac{s}{\tau}} \int_0^{\frac{s}{\Delta t}} e^{\frac{\beta\Delta t}{\tau}} \mathbb{I}\left(\frac{k}{K_n} \leq \beta < \frac{k+1}{K_n}\right) \frac{\Delta t}{\tau} d\beta. \quad (18)$$

Using another change of variables $\gamma = \frac{\beta\Delta t}{\tau}$,

$$\sum_{k=0}^{K_n-1} a_k e^{-\frac{s}{\tau}} \left(\int_{\gamma_{\min}^{n,k,s}}^{\gamma_{\max}^{n,k,s}} e^{\gamma} d\gamma \right) \mathbb{I}(\gamma_{\min}^{n,k,s} < \gamma_{\max}^{n,k,s}), \quad (19)$$

where

$$\gamma_{\min}^{n,k,s} = \max\left\{\frac{k}{K_n} \frac{\Delta t}{\tau}, 0\right\} \quad (20a)$$

$$\gamma_{\max}^{n,k,s} = \min\left\{\frac{k+1}{K_n} \frac{\Delta t}{\tau}, \frac{s}{\tau}\right\}. \quad (20b)$$

The solution to the integral can now be written in closed form yielding

$$y(t) - y(t_0)e^{-\frac{s}{\tau}} = \sum_{k=0}^{K_n-1} a_k e^{-\frac{s}{\tau}} (e^{\gamma_{\max}^{n,k,s}} - e^{\gamma_{\min}^{n,k,s}}) \mathbb{I}(\gamma_{\min}^{n,k,s} < \gamma_{\max}^{n,k,s}). \quad (21)$$

Substituting for t with values t_1, t_2, \dots, t_N , we retrieve (14) with $m = n$. Therefore, for every f_n satisfying (11), the corresponding A_{f_n} satisfies (14) by construction. ■

The above theorem allows us to represent the constraints in (11) as a system of linear equations (14). The choice of n divides the interval $[t_0, t_N)$ into uniform intervals of size $\frac{\Delta t}{2^n}$ with discontinuities at integral multiples of $\frac{\Delta t}{2^n}$. Therefore, we align the observation time instances $\{t_0, t_1, \dots, t_N\}$, which are typically uniformly spaced, with the discontinuities of f_n by choosing N to be a power of 2. Recall that increasing n strictly increases the set of signals that are represented. Therefore, the set of all possible feasible solutions to (11) is obtained by letting $n \rightarrow \infty$. The next sections shows that temporal sparsity can be leveraged to select a unique signal from this feasible set.

IV. MOTION DEBLURRING USING TEMPORAL SPARSITY

Recall from Section II-B that motion blur arises from the thermal inertia of each pixel. In the previous section, we saw that there are infinitely many signals that agree with the data returned by the camera as well as the camera parameter τ . Yet, following Section II-C, we need to choose a unique signal from this feasible set using only that pixel's temporal data. We address this by pruning the feasible set based on two observations presented here.

To understand our first observation, consider the time between two consecutive observations $[t_i, t_{i+1}]$. A large value of n would divide this interval into many small intervals, each of length $\frac{\Delta t}{2^n}$. Suppose we consider a function $f_n : [t_0, t_N) \rightarrow \mathbb{R}$ as defined in (13) and its corresponding vector representation $A = [a_0, \dots, a_{K_n-1}]^T \in \mathbb{R}^{K_n}$. Note when n is large, any pair a_p and a_q corresponding to intervals between $[t_i, t_{i+1}]$ can be increased in equal and opposite directions while effectively cancelling each other out at t_{i+1} . Therefore, given a signal with representation A satisfying (11), there are many pairs of coefficients that can be changed when n is large to obtain new signals that satisfy (11).

Next, to describe our second observation consider a typical scene to be imaged. The surfaces in the scene are locally at similar temperatures with distinct jumps at object or material boundaries that appear as edges in the thermal image. For most applications, these edges capture sufficient information

Algorithm 1 Pixelwise Deblurring using QP

Input: $\{t_0, \dots, t_N\}, \{y(t_0), \dots, y(t_N)\}, T, \log_2(N) \in \mathbb{N}, \lambda$

Output: $A^* \in \mathbb{R}^{K_n} \iff x : [t_0, t_N) \rightarrow \mathbb{R}$

Choose n such that $\frac{\Delta t}{2^n} \ll T$

Construct V, Y from (14)

Construct Haar Matrix H of size $K_n \times K_n$

Setup objective to minimizing $\|Y - VH^T D\|_2 + \lambda \|D\|_1$

Solve quadratic program to obtain D^*

return $A^* = H^T D^*$

in the image. The trajectory of scene points imaged by a single pixel would have temperatures that change drastically only at the few instances when the imaged scene point passes through a boundary. Based on the above two observations, we propose the following assumption about the signal of interest:

Assumption 1. *The signal of interest, T_m^{ss} , can be approximated by a piecewise constant signal that satisfies (11) with the minimum number of transitions between consecutive components of its representation vector.*

To understand this assumption, recall that as a result of Lemma 1 any integrable function can be approximated arbitrarily well using piecewise constant functions as the number of discontinuities is allowed to go to infinity. Reducing the number of such discontinuities would discourage redundant transitions from coefficients a_p to a_q as in the first observation. By minimizing the number of step changes, we also leverage the second observation that the temperature of a pixel in a thermal image does not typically have many large changes as a function of time.

Computing the signal of interest that satisfies this assumption can be formulated as the solution to the following optimization problem:

$$A^* = \underset{A \in \mathbb{R}^{K_n}}{\operatorname{argmin}} \rho(A) \quad (22)$$

s.t. $VA = Y,$

where A^* denotes the vector representing the approximation to the signal of interest that satisfies Assumption 1, V, Y and K_n are as in Theorem 1 for a large value of n , and $\rho : \mathbb{R}^{K_n} \rightarrow \mathbb{W}$ counts the number of transitions between consecutive components of the representation vector A (i.e. $\sum_{k=1}^{K_n-1} \mathbb{I}\{a_{k-1} \neq a_k\}$). This is a non-convex optimization problem. In this paper, we instead solve a closely related convex optimization problem described below.

Let D denote the coefficients of A in the Haar system such that $D = HA$ where H is the Haar matrix of size $K_n \times K_n$ [10]. Note that a signal with few transitions would also be sparse in the Haar system. Therefore, we can relax (22) to

$$D^* = \underset{D \in \mathbb{R}^{K_n}}{\operatorname{argmin}} \|VH^T D - Y\|_2 + \lambda \|D\|_1, \quad (23)$$

where λ is a hyper-parameter that balances between respecting (14) and temporal sparsity from Assumption 1. In particular, we have relaxed the equality constraint in (22) to account

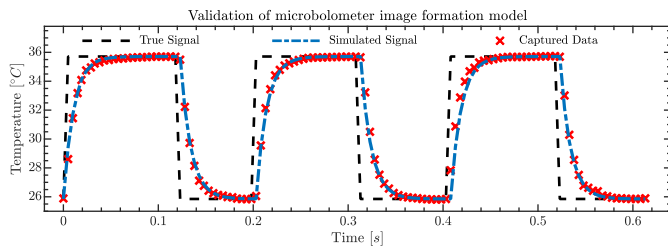


Fig. 3: Temporal plot of a single pixel from a real-world microbolometer when subjected to a series of rectangular pulses. The simulated signal for $\tau = 11\text{ms}$ agrees well with the captured real-world data.

for measurement noise. This optimization problem is called a LASSO problem [31] and can be solved efficiently using a Quadratic Programming (QP) solver [1]. Algorithm 1 summarizes the proposed approach to deblur a single pixel.

V. EXPERIMENTAL RESULTS

The experimental results are divided into five subsections. First, we validate our Microbolometer Image Formation Model. We then illustrate the proposed approach using synthetic data for a single pixel. Next, we introduce the datasets and baselines presented in the results. Next, we present deblurring results for typical outdoor scenes. Finally, we present quantitative evaluation of state-of-the-art object detectors on images restored using different deblurring algorithms.

All the experiments were carried out using a radiometrically calibrated FLIR A655sc camera capturing frames every 5ms at 640×120 resolution. The object parameters were set to unity for both emissivity and atmospheric transmissivity. The following parameters were used throughout the experiments: $\tau = 11\text{ms}$, $N = 17$, $n = 7$ and $\lambda = 0.5$. Recall that the proposed approach can deblur a frame using any consecutive set of $N + 1$ measurements of which it is a part. In our experiments, the past N frames were used, unless noted otherwise, so that future information is not required. The quadratic programs were solved using IBM CPLEX v12.10 through the Python API [1]. On average over 20 clips of 17 frames each, processing a single pixel took 212ms including problem setup on an Intel Xeon 8170M server processor running at 2.1GHz. The running time can be further reduced by using optimised code with the C++ API.

A. Validating M-IFM

Microbolometers utilise a number of internal signal processing algorithms to compute the temperature values that we receive as an image. Despite this, we show here that the overall response of a microbolometer to changes in scene still follows (6). To this end, we setup the camera to view a constant warm surface, such as a powered on monitor and move a board at a cooler temperature in front of it repeatedly such that an individual pixel experiences a series of rectangular pulses. We manually approximated the rectangular pulse and fit the M-IFM to the captured data to get $\tau = 11\text{ms}$. Fig. 3 illustrates good agreement between the captured data and the

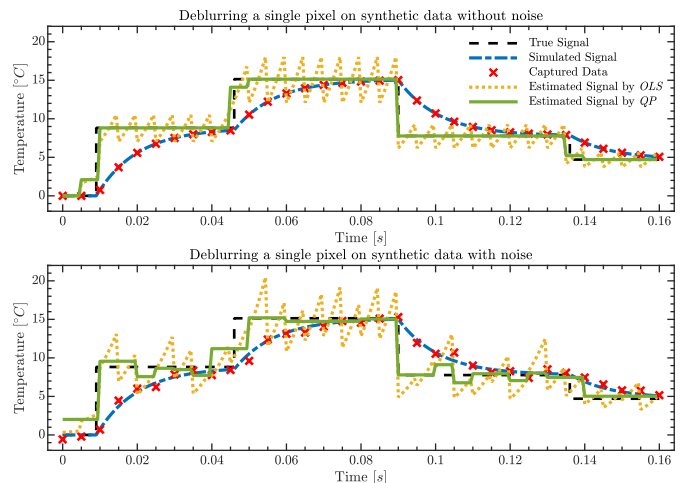


Fig. 4: An illustration that the ordinary least squares solution is unable to capture the true underlying signal even without noise. Whereas the QP solution aligns well with the ground truth signal even after the introduction of noise.

signal simulated using $\tau = 11\text{ms}$ thus validating our M-IFM. We use this value of τ in the rest of the experiments.

B. Deblurring at a single pixel

In this experiment, we analyse the proposed method for deblurring a single pixel. We generated a piece-wise constant signal and simulated the microbolometer response with $\tau = 11\text{ms}$ with additive Gaussian noise of standard deviation 0.5K. For comparison, we process the same signal with and without noise. The Ordinary Least Squares (OLS) solution to (14) does not align well with the true signal as seen in Fig. 4 even without noise. While the true signal has 4 transitions, the number of transitions in OLS is much higher. However, the QP solution agrees well with the true signal in both the noisy and noiseless scenarios. The λ values used here were 0.001 and 2 for the noiseless and noisy signals, respectively. This demonstrates the utility of sparsity as presented in Section IV.

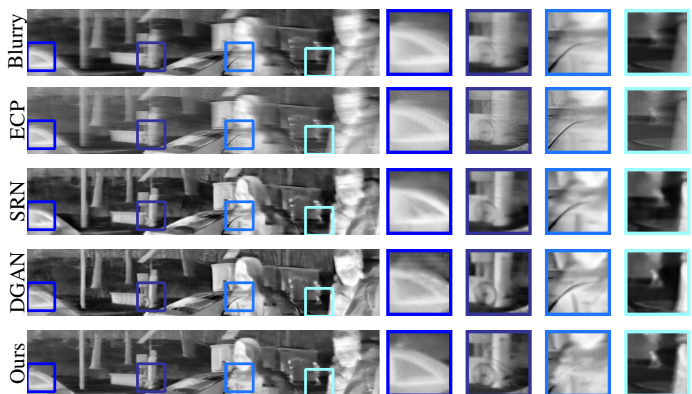


Fig. 5: An illustration of the output of various deblurring algorithms on a thermal image (top row) of an outdoor scene with objects at different depths. In particular, call out 2 (third column from left) illustrates a bicycle frame that is recovered by our approach while only the front wheel is recovered by some of the other approaches. Call out 3 (fourth column from left) depicts the deblurring result when the motion blur in the input image sequence blended the car and the person in front of it.

C. Dataset and Baselines

We collected a test dataset, which we refer to as the *Blurry Thermal Data set*, with generic camera motions in outdoor scenes with moving pedestrians and vehicles. Note that obtaining ground truth thermal images without motion blur using microbolometers is infeasible as there is no control over exposure time. One could use a cooled photon detector, but that would require pixel-wise correspondence between the microbolometer and the expensive cooled photon detector, which is a challenging task. Therefore, we are unable to quantitatively evaluate the deblurring algorithms directly.

As a proxy, we evaluate the utility of such deblurring algorithms by benchmarking the performance of state-of-the-art object detectors on blurred and deblurred thermal images using our proposed algorithm and a variety of existing, state-of-the-art approaches. To this end, we annotated a subset of 406 images using the COCO Annotator [8] for a total of 1168 pedestrians and 831 vehicles. We compared our approach against several classes of deblurring algorithms. Single-image deblurring algorithms: 1) non-learning based method (ECP [38]), 2) learning based approaches (SRN [30], DGAN [14]). Video deblurring algorithms: 3) learning based methods (STFAN [39], EDVR [35]). Since thermal images are in units of temperature, we map the temperature range of the images to $[0, 1]$ and used the floating point representation directly, whenever possible, in the baseline algorithms to avoid loss of precision due to intermediate 8-bit image formats.

D. Motion Deblurring of Natural Scenes

This section qualitatively compares the performance of the various deblurring approaches. For brevity, we compare performance with the non-learning algorithm ECP and two strong learning-based algorithms, SRN and DGAN. Comparison with all baseline algorithms are presented as supplementary material for reference. Figs. 5 and 6 illustrate images captured at an intersection with pedestrians and cars moving in different directions at different depths causing highly non-uniform blur. The non-learning based, ECP algorithm merely sharpens the image without removing any blur as it assumes uniform blur kernel. In all the call outs, our proposed model-based approach achieves equal or superior deblurring without any artifacts suffered by learning based algorithms.

E. Object Detection under Severe Motion Blur

We consider three state-of-the-art object detectors, namely Faster R-CNN [27], RetinaNet [18] and Yolov3 [26]. For Faster R-CNN and RetinaNet, we use their state-of-the-art implementation in Detectron2 [37] with Feature Pyramid Networks [17] and a ResNet-101 [13] backbone. The object detectors were trained on the FLIR Thermal Data set [2] and their mean Average Precision(mAP) scores on the corresponding FLIR Thermal validation set were 46.57%, 42.54% and 38.38% for object detection, respectively. Note that we only evaluate for *Person* and *Car* categories with our Blurry Thermal Data set. All evaluations were done using the stringent COCO [16] metric which takes average over

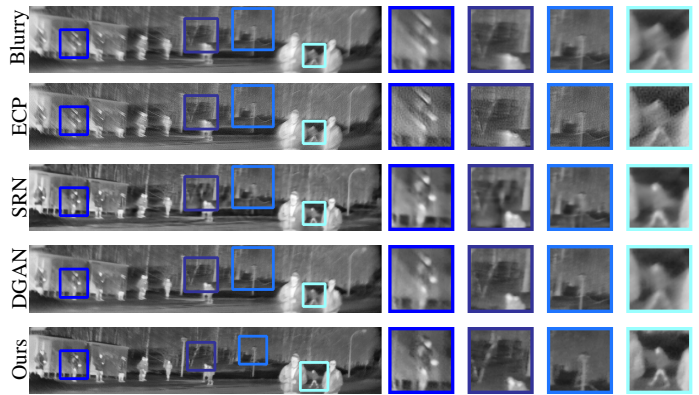


Fig. 6: An illustration of the output of various deblurring algorithms on a thermal image (top row) when an arbitrary camera motion. ECP sharpens the image without removing blur. SRN introduces artifacts seen in call outs 2 and 3 (columns 3 and 4 from the left, respectively). DGAN has residual blur in call out 4 (column 5 from the left). In comparison our proposed approach shows no artifacts or blur.

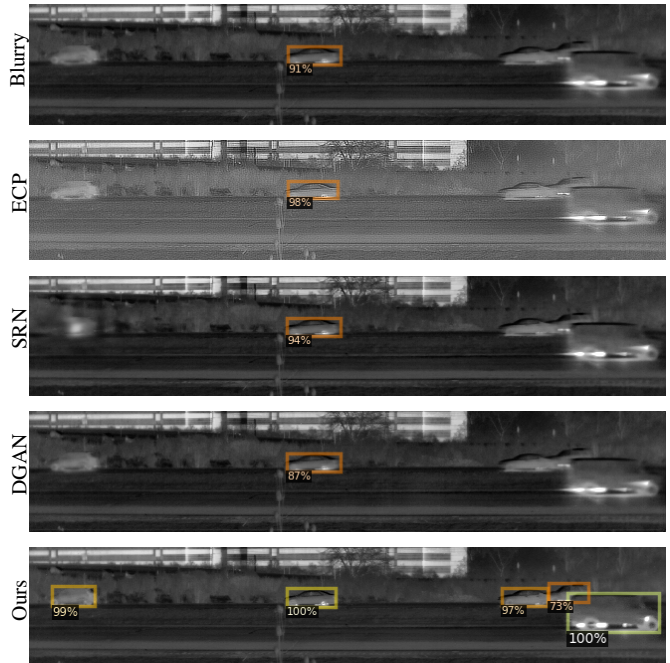


Fig. 7: An illustration of object detection results from RetinaNet using the original image, the output of various state-of-the-art deblurring techniques, and our proposed approach. The stationary camera is viewing an expressway with fast moving cars (top image). The images deblurred using our proposed approach enables the detector to find all the cars in the scene.

Intersection-over-Union (IoU) thresholds every 0.05 from 0.5 to 0.95.

The trained object detectors were tested on the Blurry Thermal Data set. Fig. 7 depicts object detection results for RetinaNet on a highway scene. All the vehicles are detected when using images restored by our approach. Fig. 8 illustrates object detection results for Faster R-CNN for a test image containing multiple pedestrians approaching an intersection. Our proposed approach aids the detector to find all three pedestrians while the other approaches were unable to deblur

Detector	Image Type	Person	Car	Total	ΔAP
Faster R-CNN	Blurred	13.62	26.88	20.25	-
	ECP [38]	11.27	24.11	17.69	-2.56
	STFAN [39]	15.18	28.23	21.71	1.46
	EDVR [35]	14.80	29.25	22.02	1.77
	SRN [30]	17.46	30.99	24.23	3.98
	DGAN [14]	16.94	31.40	24.16	3.91
	Ours	21.81	40.53	31.17	10.92
RetinaNet	Blurred	12.99	29.92	21.46	-
	ECP [38]	10.60	28.08	19.34	-2.12
	STFAN [39]	14.69	30.20	22.44	0.98
	EDVR [35]	14.32	31.61	22.96	1.5
	SRN [30]	14.99	31.96	23.48	2.02
	DGAN [14]	16.18	32.84	24.51	3.05
	Ours	20.61	41.15	30.88	9.42
YoloV3	Blurred	13.14	24.80	18.97	-
	ECP [38]	12.35	21.68	17.02	-1.95
	STFAN [39]	14.10	25.58	19.84	0.87
	EDVR [35]	14.22	26.03	20.13	1.16
	SRN [30]	15.93	26.73	21.33	2.36
	DGAN [14]	15.20	27.64	21.42	2.45
	Ours	18.63	34.55	26.59	7.62

TABLE I: The Average Precision scores for three state-of-the-art object detectors when using blurred images or images deblurred by our proposed approach and various state-of-the-art techniques. ΔAP shows the change in performance between using deblurred images and the original blurred images. Deblurring using our proposed algorithm increases performance by more than twice that achieved by using any tested, state-of-the-art deblurring methods.

the pedestrians. The performance of the detectors with the original blurred images and the images after deblurring using different algorithms are summarized in Table I.

To summarize, the performance of all three detectors on the blurry images is significantly lower than their performance on the FLIR validation set. This emphasizes the need for motion deblurring algorithms. The non-learning based ECP generates deblurred images that actually degrade the performance of all three detectors when compared to just applying the blurred images directly to each of the detectors. The learning-based algorithms improve object detection to various degrees with DGAN and SRN performing better than other baselines. However, our proposed approach more than doubles this improvement in Average Precision scores for all three detectors. While using the strong baseline DGAN to restore blurred images increases the performance by 3.91%, 3.05% and 2.45% for Faster R-CNN, RetinaNet and YoloV3, respectively. Using our proposed approach to restore blurred images increases the performance by 10.92%, 9.42% and 7.62% for Faster R-CNN, RetinaNet and YoloV3, respectively.

VI. DISCUSSION AND CONCLUSION

Motion is inherent to mobile robots where the world continuously evolves in time. In this context, microbolometers provide a unique sensing modality as they are constantly exposed to the scene, accumulating the evolution in time. Until now, the thermal inertia of microbolometers has largely been considered as an unwanted phenomenon that needs to be minimized if not eliminated. Our work shows that spreading the information in time can in-fact be leveraged using intelligent algorithms to side-step the prominent challenges in motion deblurring - 6D relative motion and dynamic objects

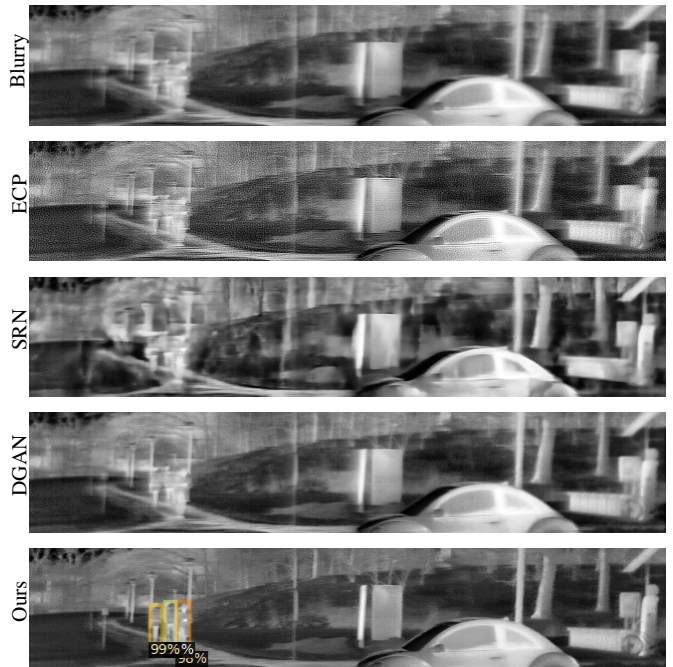


Fig. 8: An illustration of the object detection results from Faster R-CNN using the original image, the output of various state-of-the-art deblurring techniques, and our proposed approach. SRN introduces large artefacts while residual blur is still visible in the DGAN and ECP results. Our proposed approach eliminates blur and aids the detector to identify all three pedestrians in the image which were otherwise missed.

in the scene. The pixel-wise formulation of our approach allows parallel processing to be utilised to speed up motion deblurring. Furthermore, a sparse set of keypoints could be deblurred on-demand providing an attractive utility for robotic vision.

Motion blur in thermal cameras is a primary impediment for their wider adoption in robotics despite their ability to see in the dark. The lack of exposure control and global shutter features further aggravates the problem. In this work, we showed that the motion blur in thermal cameras is caused due to the thermal inertia of each pixel that is unique to the physics of image formation in these cameras. We showed that reversing this effect amounts to solving an underdetermined system of linear equations which have infinitely many solutions. By leveraging high frame rate and temporal sparsity, we formulated LASSO optimization problem and solved it using Quadratic Programming(QP) to estimate the incident power at each pixel independently and yet recovered motion deblurred images that are spatially coherent. Our work shows that just temporal sparsity of signals is sufficient to achieve motion deblurring without modeling the relative motion or scene dynamics.

ACKNOWLEDGMENTS

This work was supported by a grant from Ford Motor Company via the Ford-UM Alliance under award N022884. The authors would also like to thank the reviewers for their useful comments and suggestions.

REFERENCES

- [1] Ibm ilog cplex. www.cplex.com, 2019.
- [2] FLIR ADAS. Flir thermal dataset for algorithm training. <https://www.flir.com/oem/adas/adas-dataset-form/>, 2018.
- [3] Yuval Bahat, Netalee Efrat, and Michal Irani. Non-uniform blind deblurring by reblurring. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 3306–3314. IEEE Computer Society, 2017. doi: 10.1109/ICCV.2017.356. URL <https://doi.org/10.1109/ICCV.2017.356>.
- [4] Jarosław Bareła, Mariusz Kastek, Krzysztof Firmanty, and Michał Krupiński. Comparison of parameters of modern cooled and uncooled thermal cameras. In *Electro-Optical and Infrared Systems: Technology and Applications XIV*, volume 10433, page 1043307. International Society for Optics and Photonics, 2017.
- [5] Erwan Bernard, Nicolas Rivière, Mathieu Renaudat, Michel Péalat, and Emmanuel Zenou. Active and thermal imaging performance under bad weather conditions. In *6th International Symposium on Optronics in Defence and Security (OPTRO 2014)*, Paris, FR, 2014. URL <https://oatao.univ-toulouse.fr/11729/>. Thanks to 3AF.
- [6] Mario Bijelic, Tobias Gruber, and Werner Ritter. A benchmark for lidar sensors in fog: Is detection breaking down? In *2018 IEEE Intelligent Vehicles Symposium, IV 2018, Changshu, Suzhou, China, June 26-30, 2018*, pages 760–767. IEEE, 2018. doi: 10.1109/IVS.2018.8500543. URL <https://doi.org/10.1109/IVS.2018.8500543>.
- [7] N. Boudou, A. Durand, S. Tinnes, S. Cortial, A. Gorecki, M. Cueff, and A. Virot. ULIS bolometer improvements for fast imaging applications. In Bjørn F. Andresen, Gabor F. Fulop, and Charles M. Hanson, editors, *Infrared Technology and Applications XLV*, volume 11002, pages 366 – 375. International Society for Optics and Photonics, SPIE, 2019. doi: 10.1117/12.2520360. URL <https://doi.org/10.1117/12.2520360>.
- [8] Justin Brooks. COCO Annotator. <https://github.com/jsbroks/coco-annotator/>, 2019.
- [9] Yukyung Choi, Namil Kim, Soonmin Hwang, Kibaek Park, Jae Shin Yoon, Kyoungwan An, and In So Kweon. KAIST multi-spectral day/night data set for autonomous and assisted driving. *IEEE Trans. Intell. Transp. Syst.*, 19(3):934–948, 2018. doi: 10.1109/TITS.2018.2791533. URL <https://doi.org/10.1109/TITS.2018.2791533>.
- [10] Charles K Chui. *An introduction to wavelets*. Elsevier, 2016.
- [11] Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.
- [12] Rikke Gade and Thomas B. Moeslund. Thermal cameras and applications: a survey. *Mach. Vis. Appl.*, 25(1):245–262, 2014. doi: 10.1007/s00138-013-0570-5. URL <https://doi.org/10.1007/s00138-013-0570-5>.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016. doi: 10.1109/CVPR.2016.90. URL <https://doi.org/10.1109/CVPR.2016.90>.
- [14] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 8877–8886. IEEE, 2019. doi: 10.1109/ICCV.2019.00897. URL <https://doi.org/10.1109/ICCV.2019.00897>.
- [15] Chenglong Li, Xinyan Liang, Yijuan Lu, Nan Zhao, and Jin Tang. RGB-T object tracking: Benchmark and baseline. *Pattern Recognit.*, 96, 2019. doi: 10.1016/j.patcog.2019.106977. URL <https://doi.org/10.1016/j.patcog.2019.106977>.
- [16] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. In David J. Fleet, Tomás Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*, volume 8693 of *Lecture Notes in Computer Science*, pages 740–755. Springer, 2014. doi: 10.1007/978-3-319-10602-1_48. URL https://doi.org/10.1007/978-3-319-10602-1_48.
- [17] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 936–944. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.106. URL <https://doi.org/10.1109/CVPR.2017.106>.
- [18] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(2):318–327, 2020. doi: 10.1109/TPAMI.2018.2858826. URL <https://doi.org/10.1109/TPAMI.2018.2858826>.
- [19] Qiao Liu, Zhenyu He, Xin Li, and Yuan Zheng. PTB-TIR: A thermal infrared pedestrian tracking benchmark. *IEEE Trans. Multimedia*, 22(3):666–675, 2020. doi: 10.1109/TMM.2019.2932615. URL <https://doi.org/10.1109/TMM.2019.2932615>.
- [20] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 257–265. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.35. URL <https://doi.org/10.1109/CVPR.2017.35>.
- [21] Ryota Nihei, Yutaka Tanaka, Hiroyuki Iizuka, and Takeshi Matsumiya. Simple correction model for blurred images of uncooled bolometer type infrared cameras. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXX*, volume 11001, page 1100118. Interna-

- tional Society for Optics and Photonics, 2019.
- [22] Beate Oswald-Tranta. Temperature reconstruction of infrared images with motion deblurring. *Journal of Sensors and Sensor Systems*, 7(1):13–20, 2018. Copyright - Copyright Copernicus GmbH 2018; Last updated - 2018-01-11.
- [23] Dragana B. Peric, Branko Livada, Miroslav V. Peric, and Sasa Vujic. Thermal imager range: Predictions, expectations, and reality. *Sensors*, 19(15):3313, 2019. doi: 10.3390/s19153313. URL <https://doi.org/10.3390/s19153313>.
- [24] A. N. Rajagopalan and Rama Chellappa, editors. *Motion Deblurring: Algorithms and Systems*. Cambridge University Press, 2014. ISBN 9781107360181. doi: 10.1017/CBO9781107360181. URL <https://doi.org/10.1017/CBO9781107360181>.
- [25] Manikandasriram Srinivasan Ramanagopal, Cyrus Anderson, Ram Vasudevan, and Matthew Johnson-Roberson. Failing to learn: Autonomously identifying perception failures for self-driving cars. *IEEE Robotics Autom. Lett.*, 3(4):3860–3867, 2018. doi: 10.1109/LRA.2018.2857402. URL <https://doi.org/10.1109/LRA.2018.2857402>.
- [26] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *CoRR*, abs/1804.02767, 2018. URL <http://arxiv.org/abs/1804.02767>.
- [27] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(6):1137–1149, 2017. doi: 10.1109/TPAMI.2016.2577031. URL <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [28] Antoni Rogalski. History of infrared detectors. *Opto-Electronics Review*, 20(3):279–308, 2012.
- [29] Francisca Rosique, Pedro J. Navarro, Carlos Fernández, and Antonio Padilla. A systematic review of perception system and simulators for autonomous vehicles research. *Sensors*, 19(3):648, 2019. doi: 10.3390/s19030648. URL <https://doi.org/10.3390/s19030648>.
- [30] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 8174–8182. IEEE Computer Society, 2018. doi: 10.1109/CVPR.2018.00853. URL http://openaccess.thecvf.com/content_cvpr_2018/html/Tao_Scale-Recurrent_Network_for_CVPR_2018_paper.html.
- [31] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- [32] Nezh Topaloglu. *Analysis and Modeling of Uncooled Microbolometers with Tunable Thermal Conductance*. PhD thesis, University of Waterloo, 2009.
- [33] Wayne Treible, Philip Saponaro, Scott Sorensen, Abhishek Kolagunda, Michael O’Neal, Brian Phelan, Kelly Sherbondy, and Chandra Kambhamettu. CATS: A color and thermal stereo benchmark. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 134–142. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.22. URL <https://doi.org/10.1109/CVPR.2017.22>.
- [34] Michael Vollmer and Klaus-Peter Möllmann. *Infrared thermal imaging: fundamentals, research and applications*. John Wiley & Sons, 2017.
- [35] Xintao Wang, Kelvin C. K. Chan, Ke Yu, Chao Dong, and Chen Change Loy. EDVR: video restoration with enhanced deformable convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 1954–1963. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPRW.2019.00247. URL http://openaccess.thecvf.com/content_CVPRW_2019/html/NTIRE/Wang_EDVR_Video_Restoration_With_Enhanced_Deformable_Convolutional_Networks_CVPRW_2019_paper.html.
- [36] Oliver Whyte, Josef Sivic, Andrew Zisserman, and Jean Ponce. Non-uniform deblurring for shaken images. *Int. J. Comput. Vis.*, 98(2):168–186, 2012. doi: 10.1007/s11263-011-0502-7. URL <https://doi.org/10.1007/s11263-011-0502-7>.
- [37] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. <https://github.com/facebookresearch/detectron2>, 2019.
- [38] Yanyang Yan, Wenqi Ren, Yuanfang Guo, Rui Wang, and Xiaochun Cao. Image deblurring via extreme channels prior. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 6978–6986. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.738. URL <https://doi.org/10.1109/CVPR.2017.738>.
- [39] Shangchen Zhou, Jiawei Zhang, Jinshan Pan, Wangmeng Zuo, Haozhe Xie, and Jimmy S. J. Ren. Spatio-temporal filter adaptive network for video deblurring. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 2482–2491. IEEE, 2019. doi: 10.1109/ICCV.2019.00257. URL <https://doi.org/10.1109/ICCV.2019.00257>.