Far-Infrared Image Deblurring Based on a Thermal Temporal Transient Model by Using Deep Unfolding

Tenshin Kuroiwa*, Kosuke Kurihara*, Yoshihiro Maeda†, Daisuke Sugimura‡, Takayuki Hamamoto*
*Tokyo University of Science, Tokyo, Japan
†Shibaura Institute of Technology, Tokyo, Japan
‡Tokyo Metropolitan University, Tokyo, Japan
4324519@ed.tus.ac.jp, kuriharak@rs.tus.ac.jp, ymaeda@shibaura-it.ac.jp, d-sugimura@tmu.ac.jp, hamamoto@rs.tus.ac.jp

***Abstract*—Uncooled-type far-infrared cameras offer advantages in compactness and cost efficiency; however, their slow temporal response to incident infrared radiation causes image blur. In this paper, we propose an image deblurring method for uncooled FIR imaging systems based on deep unfolding, which can introduce learnable structures into the iterative algorithm based on physics models. By adopting a deep unfolding algorithm to an iterative algorithm constructed based on an uncooled far-infrared sensor model, we can exploit the strengths of both the deep learning framework and the thermal transient model for uncooled-type far-infrared sensors. Experimental results demonstrate the effectiveness of our method.**

***Index Terms*—Deep unfolding, Far-infrared cameras, Image deblurring.**

I. Introduction

Far-infrared cameras capture thermal radiation emitted by heat sources without the need for external illumination. This unique characteristic allows the camera to capture thermal images clearly even in low-light and visually challenging environments. Due to this advantage, far-infrared cameras are utilized in a wide range of sensing applications [1]–[4].

Far-infrared cameras are categorized into two types: cooled and uncooled. The uncooled far-infrared cameras, which are based on a microbolometer sensor, are compact and cost-effective because of their operation without an active cooling system. However, these cameras exhibit a slow temporal response to incident infrared radiation. Consequently, when objects move with a speed that exceeds the temporal response of the camera, it leads to motion blur, as discussed in various studies [5]–[8].

Several studies have proposed a deblurring method for uncooled far-infrared cameras [5], [6]. The method [5] applied a deblurring algorithm originally developed for visible-light images to far-infrared data. The method [6] adopted a deep neural network (DNN)-based approach to deblur far-infrared images. However, both approaches have limitations. The former was originally designed for visible wavelengths and thus did not consider the characteristics of an uncooled far-infrared sensor. The latter approach required a large amount of training data. Moreover, they did not incorporate the physics model of an uncooled far-infrared sensor, leading to a lack of both reliability and interpretability.

Deblurring methods that incorporate the temporal response characteristics of far-infrared cameras have been proposed [7], [8]. The method [8] utilizes a thermal transient model to represent pixel-level thermal dynamics and formulates the deblurring process as an optimization problem with a temporal smoothness regularization. Although these methods introduce thermal dynamics specific to far-infrared cameras, the performance of such approaches is sensitive to the values of regularization parameters and optimization settings.

In this paper, we propose a deblurring method for uncooled far-infrared imaging based on deep unfolding, which can exploit the strengths of a physics model and data-driven learning approaches. Deep unfolding reformulates iterative optimization algorithms into deep network by unfolding each iteration into a corresponding neural network layer, allowing parameters such as step sizes and regularization weights to be learned during training [9], [10]. We employ a deep unfolding framework to an optimization algorithm derived from the physics model of uncooled far-infrared sensors. This enables performance improvements through data-driven learning while retaining interpretability based on physics models.

II. Proposed Method

We aim to estimate the energy of incident far-infrared radiation by removing motion blur from the observed temporal signals in a pixel-wise manner. We show an overview of the proposed method in Fig. 1.

This research was supported by the JSPS Grants-in-Aid for Scientific Research (JP24K02964, JP24K23903, JP25K21231).

This article has been accepted for publication in IEEE SENSORS 2025. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/SENSORS59705.2025.11330828

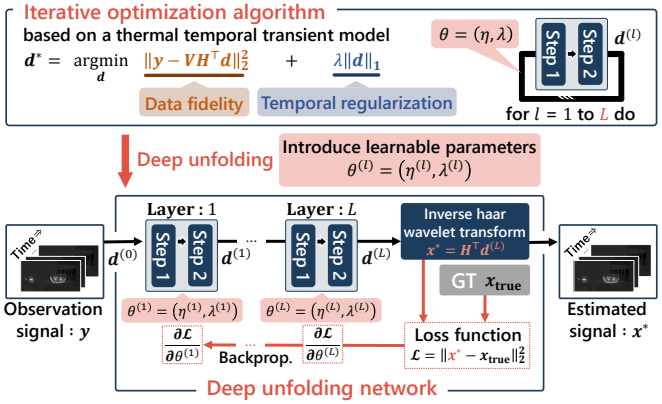


Fig. 1: Overview of proposed method.

A. Problem Formulation Based on Far-infrared Sensor Model

The sensor observation signal $\{y(t_n)\}_{n=0}^{N}$ [K], where $t_n \in \{t_0, \ldots, t_N\}$ denotes the time index, exhibits time delay and attenuation with respect to continuous incident radiant energy $x(t)$ [W], where t represents continuous time. This is due to the thermal transient response of the uncooled far-infrared sensor, which can be explained on the basis of a physics model of a far-infrared sensor.

We model the thermal transient response of each pixel using the first-order delay differential equation [8], [11], [12]. Recovering the continuous incident signal $x(t)$ from the finite sequence of sensor observation $\{y(t_n)\}_{n=0}^{N}$ is inherently ill-posed. To facilitate estimation, we divide the interval $[t_0, t_N)$ into K equally spaced subintervals, and approximate $x(t)$ by a piecewise constant function $f_K(t)$ over these subintervals, i.e., $x(t) \approx f_K(t)$. The function $f_K(t)$ is represented as

$$f_K(t) = \sum_{k=0}^{K-1} x_k \, \mathbb{I}\left(t_0 + \frac{kT}{K} \leq t < t_0 + \frac{(k+1)T}{K}\right), \quad (1)$$

where $T = t_N - t_0$ and x_k denote the length of the observation and the piecewise constant approximation of $x(t)$ over the k-th subinterval, respectively. In addition, $\mathbb{I}(t_{\min} \leq t < t_{\max})$ represents the indicator function defined as

$$\mathbb{I}(t_{\min} \leq t < t_{\max}) = \begin{cases} 1 & \text{if } (t_{\min} \leq t < t_{\max}) \\ 0 & \text{otherwise} \end{cases}. \quad (2)$$

The observation vector $\boldsymbol{y} \in \mathbb{R}^N$ is obtained by removing the influence of the initial conditions from $\{y(t_n)\}_{n=1}^{N}$. The $n = \{0, \ldots, N-1\}$-th component of $\boldsymbol{y}$, denoted by $[\boldsymbol{y}]_n$, is represented as

$$[\boldsymbol{y}]_n = y(t_{n+1}) - y(t_0)\, e^{-\frac{t_{n+1}-t_0}{\tau}}, \quad (3)$$

where τ [s] denotes the thermal time constant, defined as the ratio of the heat capacitance C [J/K] to the heat conductance G [W/K]: $\tau = \frac{C}{G}$. This parameter characterizes the temporal response to incident radiant energy $x(t)$ [13].

Using the coefficient vector $\boldsymbol{x} = [x_0, x_1, \ldots, x_{K-1}]^\top \in \mathbb{R}^K$, a linear relationship between the observation vector $\boldsymbol{y}$ and $\boldsymbol{x}$ is represented as follows:

$$\boldsymbol{y} = \boldsymbol{V}\boldsymbol{x}\ , \quad (4)$$

where $\boldsymbol{V} \in \mathbb{R}^{N \times K}$ represents the temporal convolution based on the thermal transient response of the uncooled far-infrared sensor. The (n,k)-th element of $\boldsymbol{V}$, denoted by $[\boldsymbol{V}]_{n,k}$, is represented as

$$[\boldsymbol{V}]_{n,k} = e^{-\frac{t_{n+1}-t_0}{\tau}} \left(e^{\gamma_{\max}} - e^{\gamma_{\min}}\right) \mathbb{I}(\gamma_{\min} < \gamma_{\max}), \quad (5)$$

where $\gamma_{\min}$ and $\gamma_{\max}$ are defined as

$$\gamma_{\min} = \max\left(\frac{kT}{K\tau},\ 0\right), \quad (6)$$

$$\gamma_{\max} = \min\left(\frac{(k+1)T}{K\tau},\ \frac{t_{n+1}-t_0}{\tau}\right). \quad (7)$$

Equation (4) becomes underdetermined when $N < K$. To address this, we assume that $\boldsymbol{x}$ has temporal smoothness [8]. We incorporate this assumption by applying the Haar wavelet transform to $\boldsymbol{x}$, obtaining $\boldsymbol{d} = \boldsymbol{H}\boldsymbol{x}$, where $\boldsymbol{H} \in \mathbb{R}^{K \times K}$ denotes the Haar transform matrix. To promote temporal smoothness in $\boldsymbol{d}$, an ℓ_1 norm regularization term is introduced, leading to the following optimization problem:

$$\boldsymbol{d}^* = \underset{\boldsymbol{d} \in \mathbb{R}^K}{\operatorname{argmin}} \left\|\boldsymbol{y} - \boldsymbol{V}\boldsymbol{H}^\top \boldsymbol{d}\right\|_2^2 + \lambda \left\|\boldsymbol{d}\right\|_1, \quad (8)$$

where λ is the regularization parameter. The estimated signal $\boldsymbol{x}^*$ is reconstructed by inverse transformation: $\boldsymbol{x}^* = \boldsymbol{H}^\top \boldsymbol{d}^*$.

B. Iterative Optimization Algorithm

We construct an iterative optimization algorithm for the optimization problem described in Section II-A. Specifically, we employ the iterative shrinkage thresholding algorithm (ISTA) [14]. ISTA is a proximal gradient method suitable for an objective function involving a non-differentiable ℓ_1-norm.

The solution $\boldsymbol{d}^{(l)}$ in the $l = \{1, \ldots, L\}$-th iteration is obtained as follows. We first compute the gradient of the differentiable term in the objective function in Eq. (8), i.e., $\left\|\boldsymbol{y} - \boldsymbol{V}\boldsymbol{H}^\top \boldsymbol{d}\right\|_2^2$. The auxiliary variable $\boldsymbol{u}^{(l)}$ in the l-th iteration is obtained as

$$\boldsymbol{u}^{(l)} = \boldsymbol{H}\boldsymbol{V}^\top(\boldsymbol{V}\boldsymbol{H}^\top \boldsymbol{d}^{(l-1)} - \boldsymbol{y})\ . \quad (9)$$

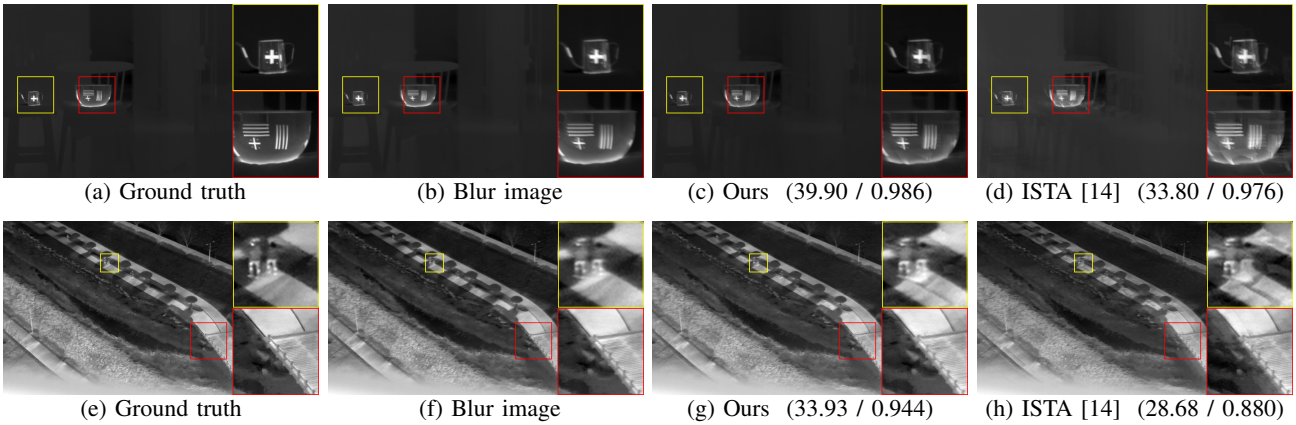
Using $\boldsymbol{u}^{(l)}$, $\boldsymbol{d}^{(l-1)}$ is updated by applying the proximal operator for non-differentiable ℓ_1-norm as

$$\boldsymbol{d}^{(l)} = \operatorname{prox}_{\lambda, \|\cdot\|_1}(\boldsymbol{d}^{(l-1)} - \eta \boldsymbol{u}^{(l)}), \quad (10)$$

where η denotes the step size. In addition, $\operatorname{prox}_{\lambda, \|\cdot\|_1}(\boldsymbol{a})$ represents the proximal operator defined as

$$\operatorname{prox}_{\lambda, \|\cdot\|_1}(a_k) = \begin{cases} a_k + \lambda & \text{if } (a_k < -\lambda) \\ 0 & \text{if } (-\lambda \leq a_k \leq \lambda), \\ a_k - \lambda & \text{if } (\lambda < a_k) \end{cases} \quad (11)$$

where a_k denotes the k-th element of $\boldsymbol{a}$.



(a) Ground truth (b) Blur image (c) Ours (39.90 / 0.986) (d) ISTA [14] (33.80 / 0.976)

(e) Ground truth (f) Blur image (g) Ours (33.93 / 0.944) (h) ISTA [14] (28.68 / 0.880)

Fig. 2: Example deblurring results with PSNR [dB] / SSIM. Top: Eval-1, Bottom: Eval-3.

C. Network Construction Based on Deep Unfolding

We construct a deep network based on the deep unfolding framework. Specifically, we unfold the iterative optimization algorithm described in Section II-B into a deep network, introducing learnable parameters.

A single layer of our network is designed based on the single iteration composed of Eqs. (9) and (10). We construct the l-th layer as

$$\boldsymbol{d}^{(l)} \leftarrow \mathcal{F}_{\theta^{(l)}}\left(\boldsymbol{d}^{(l-1)}\right), \tag{12}$$

where $\mathcal{F}_{\theta^{(l)}}(\cdot)$ denotes the operation performed in the l-th layer, implementing Eqs. (9) and (10) with learnable parameters $\theta^{(l)} = \{\lambda^{(l)}, \eta^{(l)}\}$. We construct our deep network by stacking L instances of the aforementioned single layer.

The entire network is trained by minimizing the mean squared error (MSE) between the estimated and ground truth signals using the backpropagation algorithm. The MSE loss is defined as

$$\mathcal{L}_{\mathrm{MSE}} = \|\boldsymbol{x}^* - \boldsymbol{x}_{\mathrm{true}}\|_2^2, \tag{13}$$

where $\boldsymbol{x}^*$ denotes the deblurred signal estimated by the network, and $\boldsymbol{x}_{\mathrm{true}}$ represents the corresponding ground truth.

III. Evaluation Experiment

We evaluated the effectiveness of our method using the public dataset. Experiments were conducted on an Intel Core i7-6700KF CPU and an NVIDIA GeForce RTX 3070 GPU.

A. Experiment Settings

Dataset: We utilized the SBTI dataset [7], which contains a pair of sharp (i.e., ground-truth) and blurred image sequences of four scenes. The ground-truth images were captured by an uncooled far-infrared camera (FLIR A655sc) with a thermal time constant τ of 0.008. This camera operates at 50 fps, with a resolution of 640×480 pixels and video durations ranging from 28 to 40 seconds. The blurred image sequences were synthesized from ground-truth images based on the thermal transient model as described in Section II-A.

TABLE I: Quantitative comparison results

	PSNR [dB]		**SSIM**	
Data	**Ours**	**ISTA [14]**	**Ours**	**ISTA [14]**
Eval-1	35.85	32.96	0.982	0.972
Eval-2	27.29	26.65	0.874	0.801
Eval-3	30.41	25.49	0.910	0.829
Eval-4	27.83	27.15	0.921	0.885
Average	**30.35**	28.06	**0.922**	0.872

Network training: We initialized the step size η and the regularization parameter λ as $\eta^{(l)} = 2.0$ and $\lambda^{(l)} = 0.01$ ($\forall l \in \{1, \ldots, L\}$). The number of input frames was set to $N = 4$, and the number of layers was set to $L = 100$. Training was performed using the AdamW optimizer with a learning rate of 0.001, a batch size of 8192, and 30 epochs.

Evaluation setup: As a comparison method, we employed a physics model-based method with fixed parameters as described in Section II-B. To be a fair assessment, we set the step size and regularization parameters to $\eta = 2.0$ and $\lambda = 0.01$, which are the same as the initial values in the proposed method.

The performance of our method was evaluated using 4-fold cross-validation, where each image sequence was used as a test set once, while the remaining three sequences were used for training. For evaluation metrics, we used PSNR and SSIM, which are widely used metrics for image quality assessment.

B. Experimental Results

Table I presents a comparison result using PSNR and SSIM metrics. Higher values of PSNR and SSIM indicate better image quality. It can be seen that our method exhibits higher PSNR and SSIM values for all evaluation videos than the comparison method, indicating superior deblurring performance.

Fig. 2 presents deblurring results for the Eval-1 and Eval-3 sets. In both scenes, our method effectively suppresses motion blur. As shown in the red rectangular region of Fig. 2 (h), the comparison method exhibits a noticeable residual blur. In con-

trast, our method achieves accurate deblurring, demonstrating the effectiveness of the deep unfolding-based approach.

IV. Conclusion

We proposed a deblurring method that employs deep unfolding to an iterative algorithm, which is based on the thermal transient model of an uncooled far-infrared sensor. Our method enables parameter learning within a physically grounded framework. We demonstrated the effectiveness of our method in both qualitative and quantitative evaluations.

References

[1] A. S. Bhadoriya, V. Vegamoor, and S. Rathinam, "Vehicle detection and tracking using thermal cameras in adverse visibility conditions," *Sensors*, vol. 22, no. 12, 2022.

[2] M. J. Sousa, A. Moutinho, and M. Almeida, "Thermal infrared sensing for near real-time data-driven fire detection and monitoring systems," *Sensors*, vol. 20, no. 23, 2020.

[3] K. S. Ahmed, F. F. Sherif, M. S. Abdallah, Y.-I. Cho, and S. M. ElMetwally, "An innovative thermal imaging prototype for precise breast cancer detection: Integrating compression techniques and classification methods," *Bioeng.*, vol. 11, no. 8, 2024.

[4] W. Choi, B. Jang, I. Jung, H. Sung, and Y. Jang, "Evaluation of preferences for a thermal-camera-based abnormal situation detection service via the integrated fuzzy ahp/topsis model," *Appl. Sci.*, vol. 13, no. 20, 2023.

[5] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8174–8182.

[6] S. Zhou, J. Zhang, J. Pan, W. Zuo, H. Xie, and J. Ren, "Spatio-temporal filter adaptive network for video deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2482–2491.

[7] K. Lee, Y. Ban, and C. Kim, "Motion blur kernel rendering using an inertial sensor: Interpreting the mechanism of a thermal detector," *Sensors*, vol. 22, no. 5, 2022.

[8] M. S. Ramanagopal, Z. Zhang, R. Vasudevan, and M. J. Roberson, "Pixel-Wise Motion Deblurring of Thermal Videos," in *Proc. Robot.: Sci. Syst.*, July 2020.

[9] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, 2021.

[10] J. Zhang and B. Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1828–1837.

[11] M. Vollmer and K.-P. Möllmann, "Advanced methods in ir imaging," in *Infrared Thermal Imaging: Fundamentals, Research and Applications*. Chichester, UK: John Wiley & Sons, Ltd, 2010, ch. 3, pp. 157–237.

[12] N. Topaloglu, P. M. Nieva, M. Yavuz, and J. P. Huissoon, "Modeling of thermal conductance in an uncooled microbolometer pixel," *Sens. Actuators A Phys.*, vol. 157, no. 2, pp. 235–245, 2010.

[13] D. L. Waldron and D. J. Lohrmann, "Per-pixel time constant measurement of bolometer cameras," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7725–7739, 2020.

[14] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.