Estimating Fog Parameters from a Sequence of Stereo Images

Yining Ding, João F. C. Mota, Andrew M. Wallace, and Sen Wang*

***Abstract*—We propose a method which, given a sequence of stereo foggy images, estimates the parameters of a fog model and updates them dynamically. In contrast with previous approaches, which estimate the parameters sequentially and thus are prone to error propagation, our algorithm estimates all the parameters simultaneously by solving a novel optimisation problem. By assuming that fog is only locally homogeneous, our method effectively handles real-world fog, which is often globally inhomogeneous. The proposed algorithm can be easily used as an add-on module in existing visual Simultaneous Localisation and Mapping (SLAM) or odometry systems in the presence of fog. In order to assess our method, we also created a new dataset, the Stereo Driving In Real Fog (SDIRF), consisting of high-quality, consecutive stereo frames of real, foggy road scenes under a variety of visibility conditions, totalling over 40 minutes and 34k frames. As a first-of-its-kind, SDIRF contains the camera's photometric parameters calibrated in a lab environment, which is a prerequisite for correctly applying the atmospheric scattering model to foggy images. The dataset also includes the counterpart clear data of the same routes recorded in overcast weather, which is useful for companion work in image defogging and depth reconstruction. We conducted extensive experiments using both synthetic foggy data and real foggy sequences from SDIRF to demonstrate the superiority of the proposed algorithm over prior methods. Our method not only produces the most accurate estimates on synthetic data, but also adapts better to real fog. We make our code and SDIRF publicly available[1] to the community with the aim of advancing the research on visual perception in fog.**

***Index Terms*—Fog parameter estimation, atmospheric scattering model, foggy dataset, photometric calibration, vehicular perception, image defogging, depth reconstruction.**

I. Introduction

FOG is formed when small water droplets are suspended in the air. They interact with light, for example via scattering, causing severe visual degradation, which in turn poses significant challenges to visual perception. Foggy scenarios, despite their low probability of occurrence, are thus important edge cases that cannot be ignored for extremely safety-oriented systems such as autonomous vehicles. The amount of visual degradation depends on the depth of the corresponding scene point, as explained by the atmospheric scattering model.

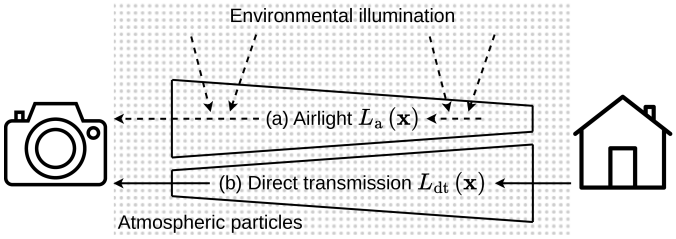

Fig. 1. The atmospheric scattering model. (a) Airlight: The atmospheric particles act as a light source by reflecting environmental illumination towards the camera, causing the radiance of the airlight $L_{\mathrm{a}}(\mathbf{x})$ to increase with distance. (b) Direct transmission: The atmospheric particles also scatter away the incident light that traverses from a scene point to the camera, causing the radiance of direct transmission $L_{\mathrm{dt}}(\mathbf{x})$ to attenuate with distance.

Atmospheric scattering model. Fig. 1 illustrates the atmospheric scattering model [1], which decomposes the total radiance originating on a scene point and reaching the camera under foggy conditions into direct transmission and airlight. The quantity of light transmitted via direct transmission (resp. airlight) decreases (resp. increases) with the distance from the scene point to the camera. Formally, let $\mathbf{x} \in \mathbb{Z}_+^2$ denote the pixel coordinates (in the image plane of the camera) associated with the scene point. Then, the total radiance $L(\mathbf{x}) \in \mathbb{R}_+$ reaching that point can be decomposed as

$$L(\mathbf{x}) = L_{\mathrm{dt}}(\mathbf{x}) + L_{\mathrm{a}}(\mathbf{x}) = L_{\mathrm{c}}(\mathbf{x})\, t(\mathbf{x}) + L_{\infty}\left(1 - t(\mathbf{x})\right), \tag{1}$$

where $L_{\mathrm{dt}}(\mathbf{x}) \in \mathbb{R}_+$ is the direct transmission, $L_{\mathrm{a}}(\mathbf{x}) \in \mathbb{R}_+$ is the airlight, $L_{\mathrm{c}}(\mathbf{x}) \in \mathbb{R}_+$ is the fog-free radiance of the scene point, $L_{\infty} \in \mathbb{R}_+$ is the radiance of the atmospheric light, i.e., the airlight at infinite distance, and $t(\mathbf{x}) \in (0,1)$ is the transmission coefficient, which controls the combination between $L_{\mathrm{c}}(\mathbf{x})$ and L_{∞} as a function of the distance $d(\mathbf{x}) \in \mathbb{R}_{++}$ between the scene point and the camera:

$$t(\mathbf{x}) = \exp\left(-\beta d(\mathbf{x})\right). \tag{2}$$

The parameter $\beta \in \mathbb{R}_{++}$ is the scattering coefficient and measures the density of fog, being related to visibility $V_{\mathrm{MOR}} \in \mathbb{R}_{++}$ (also known as the meteorological optical range [2]) as

$$V_{\mathrm{MOR}} = -\ln(0.05)/\beta, \tag{3}$$

where we assume that V_{MOR} is measured in meters.

For simplicity, in the rest of this paper we will omit $\mathbf{x}$ from any variables, e.g., in (1) and (2), whenever their dependence on the pixel coordinates is clear from context.

Y. Ding is with the Edinburgh Centre for Robotics, the School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. (e-mail: yd2007@hw.ac.uk).

J. F. C. Mota and A. M. Wallace are with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. (e-mail: {j.mota, a.m.wallace}@hw.ac.uk).

S. Wang is with the Sense Robotics Lab, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: sen.wang@imperial.ac.uk).

Manuscript received xx xx, 2024; revised xx xx, 2024.

* Corresponding author

[1]https://github.com/SenseRoboticsLab/estimating-fog-parameters

Note that we limit the scope of this work to daytime fog, mist or haze. In other inclement weather conditions, such as rain and snow, where the particles present in the atmosphere typically demonstrate significant spatial and temporal inhomogeneity, the above atmospheric model no longer applies [1].

Fog parameters. The fog parameters are L_∞ and β, and their knowledge is key both to defog the input images and to construct an accurate depth map of the scene. Estimating the fog parameters accurately is thus crucial for improving the safety of autonomous vehicles and mobile robots operating in challenging weather. Specifically, estimating L_∞ is an important step in most non-deep learning-based image defogging methods such as [3], [4]. β is also an essential parameter because, as (2) suggests, it determines the relation between the transmission coefficient t and the distance d (from which the scene depth can be calculated given the pixel coordinates $\mathbf{x}$ and the camera's intrinsic parameters). Consequently, an accurate estimate of β is a prerequisite for simultaneous defogging and stereo reconstruction methods [5]–[7].

Prior work on single image defogging [1], [3], [4], [8], [9] typically assumes that β is constant (i.e., a homogeneous medium) over horizontal paths. In our work, leveraging a sequence of images, we adapt this assumption to local homogeneity. That is, we assume that the fog is homogeneous only within a local space, corresponding to the local map constructed in the first step of our method (Section III-A). Experimental results show that our method effectively handles real-world fog, even when the fog is inhomogeneous over larger areas, thus validating our assumption.

Intensity vs radiance in fog parameter estimation. Although (1) was derived to study the scattering phenomenon in the field of photometry/radiometry, almost all existing literature on fog parameter estimation or defogging applies it directly to pixel intensity values. Hence, (1) is rewritten as

$$I = Jt + A\,(1-t)\,, \tag{4}$$

where $I \in [0, 255]$ is the observed intensity of the scene point, $J \in [0, 255]$ is the fog-free intensity of the scene point, and $A \in [0, 255]$ is the intensity of the atmospheric light. These quantities are counterparts of L, L_c and L_∞ in (1). (1) and (4) can be applied to either a grayscale image or each colour channel in an RGB image independently [8]. However, using (4) rather than (1) implicitly neglects any non-linearities incurred in the mapping from scene radiance to pixel intensity saved in a compressed image format such as JPEG or PNG. There can exist many sources of non-linearities in an image sensing pipeline, the most prominent one being the gamma correction [10]. Nevertheless, to the best of our knowledge, no existing fog parameter estimation method takes gamma correction into account, and no existing foggy dataset for autonomous driving provides the photometric parameters of the camera. We will demonstrate that estimating β based on (4) rather than (1) results in a different optimisation problem (Section III-C1) and introduces a bias (Section V-F2).

Problem statement. Given a sequence of stereo images taken by a vehicle driving in fog, our goal is to estimate the parameters β, L_∞ and relevant L_cs of the atmospheric scattering model in (1) and (2), and to update them dynamically.

Our approach and contributions. We seek to find a set of distance-radiance curves, defined by (1) and (2) and characterised by the above parameters, that most closely fits the observed data, which is generated from a local 3D feature map built by a visual Simultaneous Localisation and Mapping (SLAM) or odometry system (e.g., [11]). As the ego-vehicle moves in an environment, the local map [Fig. 2(a)] is updated, and so are the observations [Fig. 2(b)], our regression problem [Fig. 2(c)], and our parameter estimation results. As no existing real foggy dataset meets our needs, we collected our own dataset and are releasing it to the public.

We summarise our contributions as follows.

- We propose an optimisation-based method which estimates all the fog parameters *simultaneously*. Compared to prior approaches, all of which adopt a *sequential* estimation strategy, our method is less sensitive to error propagation. The proposed method is purely model-based, with its estimated fog parameters constrained via physical principles. Our only assumption is local homogeneity of the fog, a constraint which real fog in general satisfies.
- We demonstrate through comprehensive experimental results that our method a) outperforms competitive methods both quantitatively and qualitatively on simulated data; b) achieves the best performance (qualitatively) on real data. Specifically, it distinguishes thin from thick fog better than prior methods and is able to respond adaptively to spatially variant fog. Also, its estimate of atmospheric light is closer to the colour of the horizon.
- We publish the Stereo Driving In Real Fog (SDIRF) dataset, the first foggy dataset comprising consecutive stereo images of real road scenes under various visibility conditions. Our dataset also includes the counterpart clear images of the same routes recorded in overcast weather. Additionally, we calibrate the camera's photometric parameters to make SDIRF photometrically ready for the deployment of the atmospheric scattering model.

This paper is an extension of our previous work [12], in which only synthetic data was used to evaluate the proposed fog parameter estimation method. Here, we refine our method by making its initialisation fully automatic, release the new, real SDIRF dataset, and add extensive evaluations on it.

Organisation. In the next section, we review literature on fog parameter estimation and discuss existing datasets that are publicly available in the field of autonomous driving. In Section III, we explain in detail the proposed fog parameter estimation methodology. In Section IV, we introduce our self-collected SDIRF dataset and describe how we carried out the camera's photometric calibration. In Section V, we conduct extensive experiments to evaluate our fog parameter estimation method and compare it with its competitors using both synthetic and real data. We conclude in Section VI.

II. Related Work

A. Fog Parameter Estimation

Almost all existing methods operate at pixel intensity level, i.e., (4), without knowledge of the photometric parameters, that is, they estimate A rather than L_∞ or β. Early approaches

estimate A from multiple images of the same scene acquired under different conditions, such as visibility [13] or manually changed polarisation [14]. Such methods are thus inapplicable in autonomous vehicle or mobile robot scenarios. In addition, some methods rely on very strong assumptions, such as the presence of a sky region in the image. In the rest of this section, we focus on existing work that processes a single image, a stereo pair of images or a sequence of images acquired by an onboard camera.

Estimation of A. This is a critical step in non-deep learning-based single image defogging methods. To this end, [9] obtains A from the pixels that have the highest intensity in the input image, whereas [3] relies on the dark channel prior to locate the most haze-opaque region in the image and then computes A from these pixel intensities. In turn, [15] estimates A as the brightest pixel value among all local minima, and [4] locates A in RGB space by leveraging the observation that fog transforms the distribution of pixel intensities from tight clusters to stretched lines (dubbed "haze-lines"). Given the limited amount of information embedded in a single image, some of these approaches have demonstrated their general effectiveness in estimating A and therefore are adopted by later conventional methods such as [16], and some pioneering deep learning-based methods such as [17]. Even some video defogging methods such as [6] directly follow [3]'s approach in estimating A, due to its robustness and simplicity. Similarly, [18] applies firstly [15]'s method to compute an A value from the current frame. To impose temporal consistency, they then refine their estimate of A by calculating a weighted average of this A value and the A estimate from the previous frame.

Estimation of β. Whenever the value of t can be inferred directly from I [cf. (4)], most defogging methods (e.g., [3], [4]) bypass the estimation of the parameter β in (2). This topic can be categorised into *perceptual* estimation and *quantitative* estimation. Methods including [19]–[21] achieve referenceless prediction of perceptual fog density from a single image. Although their predicted perceptual fog density indices may correlate well with human judgements, the authors make no attempt to show how these perceptual indices can be mapped to a *numerical* value of β. To the best of our knowledge, the *quantitative* estimation of β is hardly addressed in the existing literature. As (2) implies, β is the key linkage between the problem of defogging and the problem of scene depth estimation, and consequently an accurate estimate of its value plays a crucial part in various existing simultaneous defogging and stereo reconstruction methods [5]–[7]. In general, estimating β entails observing the same object (more precisely, the same J) at a range of known distances, which makes this task extremely challenging at best and not always possible when only a single image or even only a stereo pair of images is available. As a special case, [22] estimates β from just a single image but requires the image to contain both the sky and the road, the latter assumed homogeneous and flat (so that a known depth can be associated with each image row from the road after calibration). These are indeed very strong and application-specific constraints, making the method inapplicable to general scenes. In contrast, [6] uses a sequence of images and performs structure-from-motion to facilitate observations of the same object at a range of known distances. After A is estimated following [3], they use each pair of observations, whose inverse depth difference is large enough, to estimate β by inverting the atmospheric scattering model. Then all the estimates are gathered, from which they build a histogram of β and choose the value from the highest bin.

To summarise, estimating the fog parameters from a sequence of images [6], [18] is more robust compared to using a single image or a stereo pair of images [3], [4], [9], [15], [22], because more information is available and there are fewer assumptions or constraints to be made. Nevertheless, existing methods still have a few shortcomings. [18] estimates A only and introduces a weighted average scheme to enforce its temporal consistency. However, as a key factor in controlling such consistency, the weight itself becomes a learnable parameter and requires fine-tuning for overall optimal performance in different scenarios. As will be shown in Section V, the strategy for estimating A and β proposed in [6] has severe drawbacks. Firstly, A is still estimated from a single image (i.e., the current frame), and thus possibly temporally inconsistent. Secondly, estimating β requires a previous estimate of A; any error in the latter estimate thus propagates to β. See our supplementary material for a theoretical qualitative analysis of the error propagation.

Distinct from the existing methods that estimate the fog parameters sequentially, we propose an optimisation-based method that simultaneously estimates them. It assumes only local homogeneity of the fog, which is very realistic.

B. Foggy Datasets for Autonomous Driving

Publicly available datasets have vastly aided the research and development of perception algorithms for autonomous vehicles. The overwhelming majority of them [23]–[29], however, do not contain any foggy scene, the presence of which can pose significant challenges to a driver-assistance system.

Real fog. Real fog happens rarely. Only a few existing datasets include real foggy scenes and they are listed below. DrivingStereo [30] contains four stereo sequences that are labelled "foggy". Nevertheless, in all of them the visibility is still relatively good. BDD100K [31] uses a monocular camera, from which the depth of the scene can be recovered only up to a scale using the pinhole camera model. SeeingThroughFog [32] features a number of stereo foggy sequences. However, the consecutive frames of only its left camera are published. RADIATE [33] pays particular attention to radar imaging in adverse weather. It has four stereo foggy sequences, only one of which was recorded while the ego-vehicle was on the move. Unfortunately, in that sequence there is consistently a considerable amount of water residual on the camera casing, which significantly blocks the view.

Synthesised fog. Synthesised fog has been widely used with the aim of enriching foggy data, typically in the following three ways. a) Real scenes with artificial fog: [34], [35] deploy a fog machine to generate artificial fog in a controlled environment. Although images are recorded in a wide range of visibility conditions, they only include very few preset scenes. More importantly, the scenes remain static, which dramatically

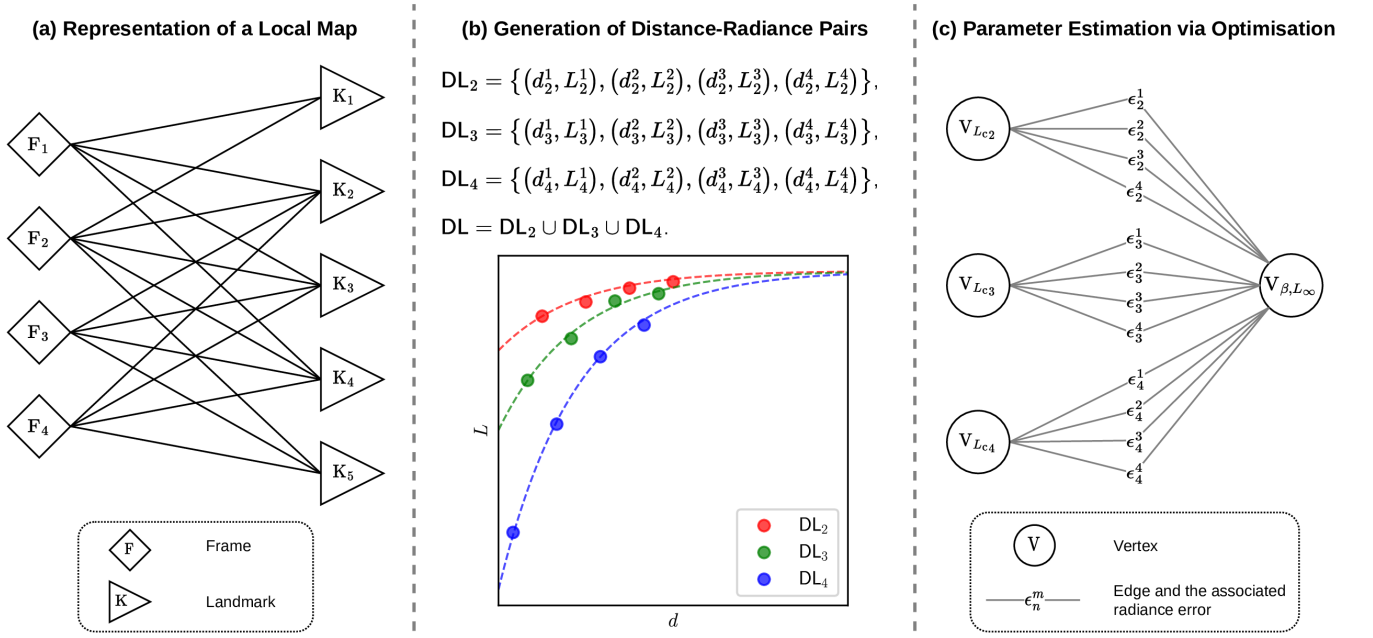

Fig. 2. Overview of our three-step method. (a) An example of a local map represented as a bipartite graph consisting of four frames, five landmarks and 17 edges which describe their observation relations (Section III-A). (b) The corresponding distance-radiance pairs (Section III-B) and their scatter plot. The distance-radiance pairs of the same landmark share the same colour. The dashed curves are generated by (12) using ground truth values. There is no distance-radiance pair associated with K_1 or K_5 because they appear in a very small number of frames (< 4). (c) The corresponding optimisation problem (Section III-C) depicted by a hypergraph. Note these figures are *illustrative*. In reality, the local map typically contains many more landmarks, with each landmark being observed in many more frames. The graph is thus much larger in practice.

limits their use. b) Real scenes with simulated fog: To this end, dense ground truth depth must be obtained first before adding fog to clear images according to (4). Most work adopting this approach considers indoor scenes rather than outdoor ones because dense depth measurements are less difficult to acquire [36], [37]. Indoor scenes, however, usually have a very limited depth range and therefore are, in general, not representative of outdoor ones. In contrast, [38], [39] add simulated fog to outdoor scenes. The authors rely on either monocular depth estimation [40] or stereoscopic inpainting [41] to generate dense pseudo-ground truth depth maps. However, such a process can introduce undesirable artefacts in the synthesised foggy images due to erroneous depth data. c) Simulated scenes with simulated fog: In order to prepare training data (or at least part of it to be used for pre-training) for data-hungry deep neural networks, some researchers [42], [43] add simulated fog to simulated scenes for which dense ground truth depth data is available. However, in no way can such completely simulated data replace real recordings for real-world, extremely safety-oriented applications.

To summarise, no existing dataset focuses specifically on real road scenes recorded by a vehicle driving in fog comprising high-resolution, consecutive left and right images. To address the above concerns, we present SDIRF. Our dataset has the following two additional features.

- We calibrated the camera's photometric parameters to make SDIRF photometrically ready for the deployment of the atmospheric scattering model.
- We also collected the counterpart clear images in overcast weather of the same routes. Such data can be useful for quantitative evaluation of downstream depth estimation and image defogging tasks.

III. Methodology

In a nutshell, given a sequence of stereo foggy images, our method simultaneously estimates the parameters β, L_∞ and relevant L_cs of the atmospheric scattering model in (1) and (2), and dynamically updates them as new images become available. Fig. 2 depicts from left to right the three steps of our method: the representation of a local map (Section III-A), the generation of distance-radiance pairs (Section III-B), and the parameter estimation via optimisation (Section III-C).

A. Representation of a Local Map

We first use the sequence of stereo images to build a local 3D feature map of the scene using a visual SLAM/odometry system such as [11]. We build a local map, rather than a global one, for two reasons: a) a local map entails a locally homogeneous fog model, as opposed to a globally homogeneous one; b) the dimensions of the resulting optimisation problem are smaller and thus can be solved more efficiently.

A local map is a collection of observations describing which local frames observe which local landmarks. It is represented as a bipartite graph G [Fig. 2(a) shows an example]:

$$\mathsf{G} = (\mathsf{F}, \mathsf{K}, \mathsf{E}), \tag{5}$$

where F denotes a set of left image frames, K denotes a set of 3D landmarks, and E denotes a set of edges each connecting a frame in F to a landmark in K. More specifically, an edge $(m, n) \in \mathsf{E}$ exists between the mth frame $\mathrm{F}_m \in \mathsf{F}$ and the nth landmark $\mathrm{K}_n \in \mathsf{K}$ only if F_m observes K_n.

Let $\mathsf{E}_n \subseteq \mathsf{E}$ denote the set of edges incident to K_n. $|\mathsf{E}_n|$ is therefore the number of frames that observe K_n, and E can be partitioned as

$$\mathsf{E} = \bigcup_{n=1}^{|\mathsf{K}|} \mathsf{E}_n. \tag{6}$$

B. Generation of Distance-Radiance Pairs

Using the local map built in the previous step, we now generate observations that will serve as data in the subsequent optimisation step. More specifically, these observations are distance-radiance pairs [Fig. 2(b)]. We use DL_n to denote the set of distance-radiance pairs associated with landmark K_n:

$$\mathsf{DL}_n = \{(d_n^m, L_n^m) \,:\, (m,n) \in \mathsf{E}_n\}, \tag{7}$$

where d_n^m denotes the Euclidean distance between K_n and F_m, and L_n^m denotes the radiance of K_n observed in F_m. The distance d_n^m can be calculated from the output of a sparse feature-based visual SLAM/odometry system, which typically consists of camera poses and landmark positions. I_n^m, which denotes the pixel intensity of landmark K_n's corresponding 2D feature point in frame F_m, is also typically available. Next, we will explain how we derive L_n^m from I_n^m.

Gamma expansion/compression. We use $g : [0, 255] \to \mathbb{R}$ to denote the mapping from pixel intensity I to radiance L. g is essentially a gamma expansion [10], the inverse operation of a digital camera's image signal processor (ISP):

$$g(I) :- \alpha I^{\gamma} + \zeta, \tag{8}$$

where $\alpha > 0$, $\gamma > 1$ (hence the name "gamma expansion") and $\zeta \in \mathbb{R}$ are the photometric parameters that characterise g. Inversely, we use $g^{-1} : \mathbb{R} \to [0, 255]$ to map radiance to intensity:

$$g^{-1}(L) :- \left(\frac{L-\zeta}{\alpha}\right)^{\frac{1}{\gamma}}. \tag{9}$$

Calibrating the camera entails discovering the photometric parameters α, γ and ζ. See Section IV-B for the details of our calibration procedure.

We use DL to denote the overall set of distance-radiance pairs, which can be expressed as the union of some (disjoint) sets in (7):

$$\mathsf{DL} = \bigcup_{n\,:\,|\mathsf{E}_n| \geq \xi_{\mathrm{F}}} \mathsf{DL}_n, \tag{10}$$

where $\xi_{\mathrm{F}} \in \mathbb{Z}_{++}$ is a threshold that ensures that a landmark is considered only if it is observed in at least ξ_{F} frames. As a result, the number of disjoint DL_ns that form DL is typically smaller than $|\mathsf{K}|$.

Finally, in order to make the estimation of the fog parameters reliable, we require DL to consist of at least $\xi_{\mathrm{K}} \in \mathbb{Z}_{++}$ disjoint DL_ns:

$$|\{n \,:\, |\mathsf{E}_n| \geq \xi_{\mathrm{F}}\}| \geq \xi_{\mathrm{K}}. \tag{11}$$

This condition is a prerequisite for the next step, estimating the fog parameters via optimisation.

C. Parameter Estimation via Optimisation

This step estimates the fog parameters β and L_∞, together with the clear radiance L_{c} of the relevant landmarks, by minimising a cost function that uses the observations DL generated in the previous step. The problem of interest can be represented as a hypergraph, in which vertices represent variables to optimise, and edges represent observation errors [44]. An edge connects two vertices that contribute to the underlying observation error. In such a hypergraph [Fig. 2(c) shows an example], there are two types of vertices.

- $\mathrm{V}_{\beta, L_\infty}$ encodes the fog parameters β and L_∞. When the fog is locally homogeneous, β and L_∞ are invariant within a local space and, in that case, there is only one such vertex.
- $\mathrm{V}_{L_{\mathrm{c}n}}$ encodes the clear radiance $L_{\mathrm{c}n}$ of the nth landmark K_n. The number of occurrences of such vertex is the same as the number of disjoint subsets in DL (10).

According to the atmospheric scattering model in (1) and (2), we can compute the predicted radiance value of the nth landmark K_n observed in the mth frame F_m from the distance d_n^m between K_n and F_m, the scattering coefficient β, the atmospheric light radiance L_∞, and K_n's clear radiance $L_{\mathrm{c}n}$:

$$\begin{aligned} {}_{\text{pred}}L_n^m &= L_{\mathrm{c}n} \exp\left(-\beta d_n^m\right) + L_\infty \left(1 - \exp\left(-\beta d_n^m\right)\right) \\ &= \left(L_{\mathrm{c}n} - L_\infty\right) \exp\left(-\beta d_n^m\right) + L_\infty. \end{aligned} \tag{12}$$

We define an error term $\epsilon_n^m \in \mathbb{R}$ to be the difference between the observed radiance L_n^m and the corresponding estimated radiance ${}_{\text{pred}}L_n^m$:

$$\begin{aligned} \epsilon_n^m &= L_n^m - {}_{\text{pred}}L_n^m \\ &= L_n^m - \left[\left(L_{\mathrm{c}n} - L_\infty\right) \exp\left(-\beta d_n^m\right) + L_\infty\right]. \end{aligned} \tag{13}$$

We can see that each ϵ_n^m depends on β, L_∞ and $L_{\mathrm{c}n}$, and is therefore associated with an edge between $\mathrm{V}_{\beta, L_\infty}$ and $\mathrm{V}_{L_{\mathrm{c}n}}$ in the hypergraph. We define

$$\mathcal{E}_n = \{\epsilon_n^m \,:\, (m,n) \in \mathsf{E}_n\} \tag{14}$$

as the set of radiance errors associated with K_n. Also, $\mathcal{E}$ will represent the overall set of radiance errors, which can be expressed as the union of some (disjoint) sets in (14):

$$\mathcal{E} = \bigcup_{n\,:\,|\mathcal{E}_n| \geq \xi_{\mathrm{F}}} \mathcal{E}_n. \tag{15}$$

We define each residual term to be a loss function, e.g., Huber loss or square loss, $\ell : \mathbb{R} \to \mathbb{R}_+$ of ϵ_n^m. The total cost function is a weighted sum of all residual terms. Our goal is to solve:

$$\begin{aligned} \underset{\beta, L_\infty, \{L_{\mathrm{c}n} \,:\, \mathcal{E}_n \subset \mathcal{E}\}}{\text{minimise}} \quad & \textstyle\sum_{n\,:\,\mathcal{E}_n \subset \mathcal{E}} \sum_{m\,:\,\epsilon_n^m \in \mathcal{E}_n} w_n^m \ell\left(\epsilon_n^m\right) \\ \text{subject to} \quad & l_\beta \leq \beta \leq u_\beta \\ & l_{L_\infty} \leq L_\infty \leq u_{L_\infty} \\ & l_{L_{\mathrm{c}n}} \leq L_{\mathrm{c}n} \leq u_{L_{\mathrm{c}n}}, \end{aligned} \tag{16}$$

where $\{L_{\mathrm{c}n} \,:\, \mathcal{E}_n \subset \mathcal{E}\}$ are the radiances of the relevant landmarks, $w_n^m \in \mathbb{R}_+$ is the weight associated with ϵ_n^m, and ls and us are the lower and upper bounds of the parameters, respectively. In the following, we first analyse (16) and then describe our initialisation scheme. We also explain how we set each l and u, our two-stage strategy for solving (16), and finally how we set each weight w_n^m.

1) Analysis of (16)*:* We show that a) (16) is non-convex; b) when the gamma correction in (8) is non-linear, i.e., $\gamma \neq 1$, solving (16) using ϵ_n^m computed in the radiance domain or in the intensity domain yields problems that are not equivalent.

Non-convexity. Consider an arbitrary error term ϵ_n^m and, without loss of generality, the square loss. Then, each unweighted term in the objective of (16) can be written as

$$f\left(\beta, L_\infty, L_{\mathrm{c}n}\right) = \left(L_n^m - \left[\left(L_{\mathrm{c}n} - L_\infty\right)\exp\left(-\beta d_n^m\right) + L_\infty\right]\right)^2. \tag{17}$$

A function is convex if and only if it is convex when restricted to any line intersecting its domain [45, §3.1.1]. Consider then (17) restricted to the line $\beta = L_\infty = \nu, L_{\mathrm{c}n} = 0$, for $\nu \geq 0$:

$$\phi\left(\nu\right) := f\left(\nu, \nu, 0\right) = \left(b + \nu\exp\left(-a\nu\right) - \nu\right)^2,$$

where $a, b \geq 0$ are constants. The 2nd-order derivative of ϕ with respect to ν is

$$\phi''\left(\nu\right) = 2\left[\exp\left(-a\nu\right)\left(1 - a\nu\right) - 1\right]^2 + 2\big[b + \nu\exp\left(-a\nu\right) - \nu\big]a\exp\left(-a\nu\right)\left(\nu - 2\right).$$

Setting, for example, $a = b = 1$, one obtains $\phi''\left(0.2\right) \approx -2.6 < 0$, where the function is concave, and $\phi''\left(1\right) \approx 1.73 > 0$, where the function is convex. Thus, f is neither concave nor convex. In other words, it is non-convex. See our supplementary material for a visual example of non-convexity.

Impact of gamma correction. As above, consider an arbitrary error term ϵ_n^m and a square loss. When we use radiance, the corresponding unweighted term in (16) is given by (17). Suppose now that we apply no gamma correction. In this case, we directly use intensity and apply (4), making the corresponding term in (16)

$$f_{\text{int}}(\beta_{\text{int}}, A, J_n) = \left(I_n^m - \left[\left(J_n - A\right)\exp\left(-\beta_{\text{int}} d_n^m\right) + A\right]\right)^2, \tag{18}$$

where β_{int} denotes the scattering coefficient when we use intensity. Substituting I_n^m with $g^{-1}\left(L_n^m\right)$ defined in (9), (18) becomes

$$\left[\left(\frac{L_n^m - \zeta}{\alpha}\right)^{\frac{1}{\gamma}} - \left[\left(J_n - A\right)\exp\left(-\beta_{\text{int}} d_n^m\right) + A\right]\right]^2. \tag{19}$$

When $\gamma = 1$ (i.e., g is affine), (19) can be written as

$$\alpha^{-2}\Bigg[L_n^m - \Big[\left(\left(\alpha J_n + \zeta\right) - \left(\alpha A + \zeta\right)\right)\exp\left(-\beta_{\text{int}} d_n^m\right) + \left(\alpha A + \zeta\right)\Big]\Bigg]^2. \tag{20}$$

Comparing (17) and (20), we observe that the two minimisation problems are related by a positive scaling and the following one-to-one mappings for the variables: $\beta \leftarrow \beta_{\text{int}}$, $L_{\mathrm{c}n} \leftarrow \alpha J_n + \zeta$ and $L_\infty \leftarrow \alpha A + \zeta$. Thus, when $\gamma = 1$, (17) and (18) yield equivalent problems [45, §4.1.3].

In the typical case where $\gamma \neq 1$, the gamma correction alters the structure of the model such that the above equivalence no longer holds. To see this, let $h\left(\gamma\right) := \left[\left(L_n^m - \zeta\right)/\alpha\right]^{1/\gamma}$ and $c := \left(L_n^m - \zeta\right)/\alpha$. A Taylor expansion of $h\left(\gamma\right)$ around $\gamma = 1$ yields

$$\begin{aligned} &h\left(1\right) + h'\left(1\right)\left(\gamma - 1\right) + \frac{1}{2}h''\left(1\right)\left(\gamma - 1\right)^2 + \cdots \\ &= c - \left(\ln c\right)c\left(\gamma - 1\right) + \frac{1}{2}\left(\ln c\right)\left(2 + \ln c\right)c\left(\gamma - 1\right)^2 + \cdots. \end{aligned} \tag{21}$$

We can see that if any of the 1st-order and the subsequent higher-order terms in (21) is non-zero, then (17) and (18) no longer yield equivalent problems. In Section V-F2, we will show this experimentally and investigate how the estimate of β is affected by this non-linearity.

2) Initialisation: We will solve (16) with an iterative algorithm, e.g., Levenberg–Marquardt. However, since (16) is non-convex, it is critical to initialise β, L_∞ and $\{L_{\mathrm{c}n} : \mathcal{E}_n \subset \mathcal{E}\}$ properly. Our initialisation strategy is as follows. Assuming the fog to be locally homogeneous, if we have access to previous estimates $\hat{\beta}$, $\hat{L}_\infty$ and $\{\hat{L}_{\mathrm{c}n} : \mathcal{E}_n \subset \mathcal{E}\}$, obtained from the last run of our fog estimation process, we use these values as initialisation. Otherwise (i.e., if the fog estimation process has never run before, or if $L_{\mathrm{c}n}$ has never been estimated before), we set $\beta = 0.014$, which is the geometric mean of its lower and upper bounds (Section III-C3). For L_∞, we use the radiance of all landmarks observed from the maximal distance. And for $L_{\mathrm{c}n}$, we use the radiance of the corresponding landmark observed from the minimal distance.

3) Parameter Bounds: We set $l_\beta = 0.001$ and $u_\beta = 0.2$, which, according to (3), corresponds to a visibility range of $[15, 3000]$ meters, values that are conservative.

Next, we set the bounds for each $L_{\mathrm{c}n}$, and build a candidate set for l_{L_∞} at the same time, as explained below. For each $L_{\mathrm{c}n}$, we first determine if it is lower or higher than L_∞ by computing the slope k_n of the line going through the intensities observed at the maximal and the minimal distances:

$$k_n = \left(g^{-1}\left(L_n^{d_{\max}}\right) - g^{-1}\left(L_n^{d_{\min}}\right)\right) / \left(d_{\max} - d_{\min}\right), \tag{22}$$

which will then be compared to a threshold $\eta \in \mathbb{R}_{++}$. If $k_n > \eta$ (i.e., strongly positive), we set $l_{L_{\mathrm{c}n}} = g\left(0\right)$ and $u_{L_{\mathrm{c}n}} = L_{\mathrm{c}n}^{d_{\min}}$, and we add $L_{\mathrm{c}n}^{d_{\max}}$ to the candidate set for l_{L_∞}. If $k_n < -\eta$ (i.e., strongly negative), we set $l_{L_{\mathrm{c}n}} = L_{\mathrm{c}n}^{d_{\min}}$ and $u_{L_{\mathrm{c}n}} = g\left(255\right)$. If neither, we set $l_{L_{\mathrm{c}n}} = g\left(0\right)$ and $u_{L_{\mathrm{c}n}} = g\left(255\right)$. See our supplementary material for more details.

Finally, we let l_{L_∞} be the median value of its candidate set, and $u_{L_\infty} = g\left(255\right)$. We noticed if we set u_{L_∞} in a similar way to l_{L_∞} (i.e., let u_{L_∞} be the median value of its candidate which consists of $L_{\mathrm{c}n}^{d_{\max}}$ when $k_n < -\eta$), its value is often underestimated. We think this is caused by the fact that objects that are brighter than L_∞ are rare in a foggy scene.

4) Two-stage Optimisation: We adopt a two-stage optimisation strategy following [11]. In the first stage, we let ℓ be the Huber loss (with parameter δ) in order to mitigate the effect of outlier observations. In the second stage, we let ℓ be the square loss and perform optimisation using inlier observations only. After the first stage, our system keeps track of the number of times each relevant observation, i.e., each relevant edge in the bipartite graph of Fig. 2(a), is classified as an inlier, which is done by evaluating each residual term and comparing it with

δ. That number, $c_n^m \in \mathbb{Z}_+$, for landmark K_n observed in frame F_m, will be used to set the weight w_n^m in (16), as explained below.

5) Residual Weights: The partial derivatives of (13) are

$$\frac{\partial \epsilon_n^m}{\partial \beta} = d_n^m \left(L_{\mathrm{c}n} - L_\infty\right) \exp\left(-\beta d_n^m\right), \tag{23a}$$

$$\frac{\partial \epsilon_n^m}{\partial L_\infty} = \exp\left(-\beta d_n^m\right) - 1, \tag{23b}$$

$$\frac{\partial \epsilon_n^m}{\partial L_{\mathrm{c}n}} = -\exp\left(-\beta d_n^m\right). \tag{23c}$$

We argue that the larger the radiance difference between a landmark's L_{c} and L_∞, the more suitable that landmark is for estimating β. As can be seen from (23a), the partial derivative of ϵ_n^m with respect to β is proportional to $(L_{\mathrm{c}n} - L_\infty)$. This suggests that when $L_{\mathrm{c}n}$ is close to L_∞ this term diminishes, causing difficulties in finding the optimal β. Intuitively, when $L_{\mathrm{c}n}$ is close to L_∞, the range of the predicted radiance ${}_{\text{pred}}L_n^m$ flattens out according to (12) and therefore ϵ_n^m contains very little information on the inference of β.

In light of this, we heuristically set the weight in our first optimisation stage to be the product of the following two terms: the absolute difference between the previous estimates $\hat{L}_{\mathrm{c}n}$ and $\hat{L}_\infty$, and the current inlier count of the corresponding observation plus one:

$$w_n^m = |\hat{L}_{\mathrm{c}n} - \hat{L}_\infty| \cdot (c_n^m + 1)\,. \tag{24}$$

It can be seen that the first term is landmark-dependent, while the second term is observation-dependent.

Our weighting scheme in the first stage is apparently related to iteratively reweighted minimisation strategies proposed in the sparse regression literature, for example [46], [47], which comes with theoretical guarantees for convergence even for some non-convex problems. Such a strategy, however, involves solving a sequence of optimisation problems, making it computationally expensive. We therefore opt to solve just one instance of (16) at a given frame. Specifically, each weight w_n^m, which is computed from estimates from a previous frame, remains constant while solving (16) for the current frame.

In our second optimisation stage where only inlier observations are used, we use uniform weighting.

Results of our ablation study (Section V-D4) show that our weighting scheme performs better than naively uniformly weighting all residual terms in both optimisation stages.

IV. SDIRF Dataset

In this section, we introduce the data collection and the photometric calibration of SDIRF. More details can be found in our supplementary material.

A. On-road Data Collection

We collected the on-road data during September 2023 in Rosyth, Queensferry and Penicuik near Edinburgh, Scotland. Foggy data was collected first, and the counterpart clear data in overcast weather was collected two weeks later by traversing the same routes.

We used an off-the-shelf stereo camera ZED 2i which features an electronic synchronised rolling shutter and a built-in inertial measurement unit (IMU). It was placed behind the windshield of a car and connected to a laptop installed with the ZED software development kit (SDK) and the ZED Robot Operating System (ROS) wrapper. Considering the way the windshield inclines and that the mounting holes are located at the bottom of the camera but not at its top, it was mounted upside down for convenience and safety reasons (see our supplementary material for a picture of the setup). We accordingly set the "camera_flip" parameter to "true" to allow the SDK to account for this upside-down setup to generate the correct stereo images. The left frames of these images, whose poses are estimated by a SLAM algorithm in our method and to which the 3D positions of all landmarks are referenced, were actually acquired by the right camera because of this upside-down setup. Therefore, our photometric calibration was later performed on the right camera (Section IV-B).

We disabled the camera's auto-white balance and auto-exposure functionalities. Instead, we used a fixed white balance, and manually adjusted the exposure time, in conjunction with the gain, according to the lighting condition of the scene. We took note of the combination of exposure time and gain used to collect each data sequence. Later we performed a photometric calibration for each combination of these two parameters (Section IV-B).

During the collection, the data, together with the timestamps, was logged to ROS bags, which were later parsed to generate the following files.

- Rectified[2] left and right images at a frame rate of 15 Hz saved as PNG files, each at a resolution of 1920×580[3].
- IMU data at a rate of 400 Hz saved as CSV files.
- Magnetometer data at a rate of 50 Hz saved as CSV files.

In total, 52 data episodes were collected. Apart from just one episode that contains only a foggy sequence, the remaining 51 episodes comprise a foggy sequence and a counterpart clear sequence of the same route. The total duration of the foggy and the clear videos are 2578 seconds and 2443 seconds, respectively. Further, by visually examining the images, we subjectively classify the 52 foggy sequences into thin fog (20 sequences totalling 1101 seconds) and thick fog (32 sequences totalling 1477 seconds). See Fig. 3 for sample images.

B. Photometric Calibration

As (1) shows, the atmospheric scattering model operates in the radiance domain. However, the ZED 2i camera features an onboard ISP and can only save the digitally post-processed intensity data, but not the raw radiance data. Therefore, for the on-road data that we recorded, we have to infer the

[2]The stereo rectification was performed by the camera's SDK using its factory calibrated stereo parameters.

[3]The original image size was 1920×1080. The images were later cropped by 290 pixels at the top (to remove the mostly-sky region) and by 210 pixels at the bottom (to remove the car's bonnet and interior reflections caused by the windshield). The final image size after cropping has an aspect ratio of $1920/580 \approx 3.31$, which is very close to the aspect ratio ($1241/376 \approx 3.30$) of the images in the KITTI odometry benchmark [23]. See our supplementary material for an example image that illustrates the effect of our cropping.

Fig. 3. Sample images of SDIRF. We show eight foggy-clear image pairs, which are grouped into four columns. Each column contains a thin fog situation (first row) and a thick fog situation (third row). The second and fourth rows show the corresponding clear images taken in overcast weather.

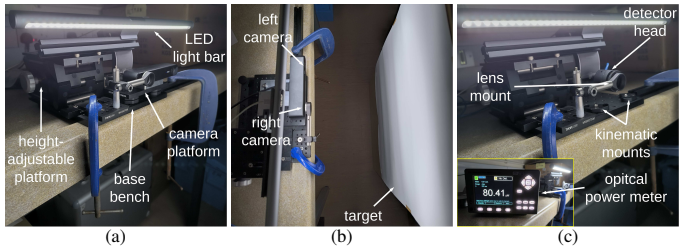

Fig. 4. Our photometric calibration setup.

radiance values from the intensity values. This process requires the photometric parameters of the camera to be known, and we achieve this by performing a photometric calibration in a controlled laboratory where the radiance values can be measured by an optical power meter. Our calibration process focuses on recovering the parameters of gamma correction, which introduces the largest amount of non-linearity in the mapping between radiance and intensity [10]. We will explain our calibration setup, the experiments we conducted, and how we infer the photometric parameters from the experimental data.

1) Setup: Our calibration setup is shown in Fig. 4. The only light source was an LED light bar with 20 different, adjustable levels of brightness. It emitted diffused light against a big white sheet (in matte finish), which was used as the target. The light bar was affixed to a height-adjustable platform mounted onto the base bench using screws. The camera was mounted the right way up onto a platform which, in turn, was attached to the base bench via three kinematic mounts. Using kinematic mounts ensured that the camera's position relevant to the base bench remained the same every time it was removed then reattached. When attached, its image plane was roughly aligned to the sheet and the principal axis of the right camera intersected the centre of the sheet. To measure the optical power perceived by the camera, we used an optical power meter paired with an optical power detector. A lens mount (with no lens mounted) was affixed to the base bench and placed just in front of the right camera's position for attaching the detector head.

2) Experiments: We calibrated the photometric parameters for each combination of exposure time and gain that had been used when collecting the on-road data. The rest of the camera settings were configured identically to the ones used in our data collection. To avoid image saturation, the target was placed further away from the test bench when we used longer exposure time. During calibration, we alternated between the following two modes.

- Photo mode [Figs. 4(a) and 4(b)]. In this mode, the detector head is removed from the lens mount, the camera's platform is attached to the base bench, and we let the camera take a stereo pair of photos of the target. In particular, the right camera images the target through the hole of the lens mount.
- Power measurement mode [Fig. 4(c)]. In this mode, the camera's platform is removed from the base bench, the detector head is attached to the lens mount, and we record the measured optical power reading from the power meter screen.

Fixing the exposure time and gain to a given combination, Fig. 5 shows the images of the right frame taken at the 20 different levels of brightness.

3) Photometric Parameter Characterisation: For each colour channel of each image, we need to associate an intensity value with each optical power measured. We compute the intensity by averaging the pixel values of a circular area with a radius of 500 pixels (see the magenta outline in the very last image of Fig. 5) within the target region in an image. Given the intensity values and the corresponding optical power readings, we use least squares fitting to estimate α, γ and ζ in (8).

Fig. 6 plots the data points obtained from the image series in Fig. 5 and the fitted curves. We can see that all curves bend upwards (i.e., $\gamma > 1$), which is in line with the expectation of a gamma expansion.

V. Experiments

We now describe our experiments. After introducing the data used for evaluation and competitive methods, we describe our implementation details. Then, we present thorough results

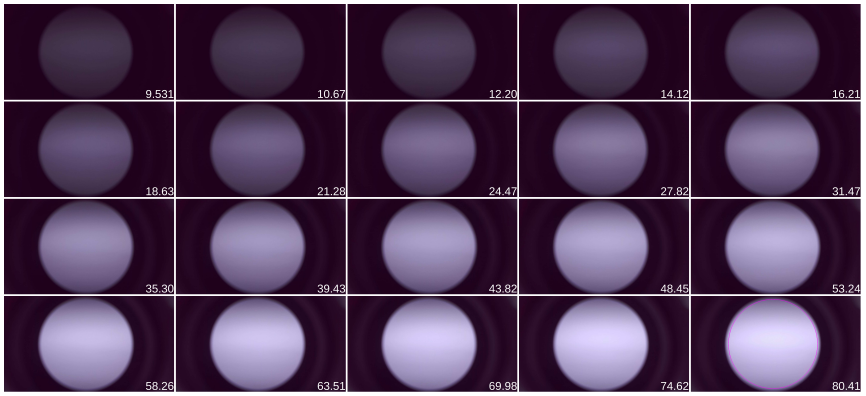

Fig. 5. The right images taken at the 20 different levels of brightness, overlaid with the corresponding measured optical power (in μW).

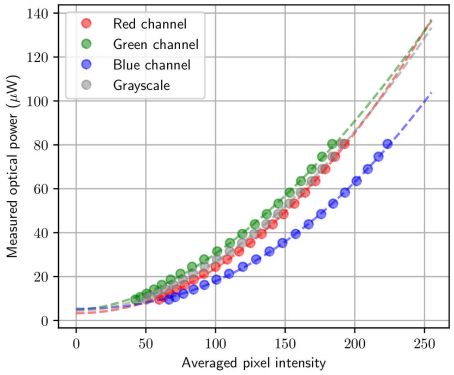

Fig. 6. The data points obtained from the image series in Fig. 5 and the fitted curves.

(a) (b) (c)

Fig. 7. Sample synthetic foggy images for evaluation. (a) VKITTI2 ($V_{\text{MOR}} = 40$ m). (b) KITTI-CARLA ($V_{\text{MOR}} = 60$ m). (c) DRIVING ($V_{\text{MOR}} = 80$ m).

on both synthetic and real data. Finally, we report two additional experiments that further showcase the superiority of our method over others.

A. Data for Evaluation

We use both synthetic and real data for evaluation.

To generate *synthetic* foggy images we use the following three datasets: the Virtual KITTI 2 dataset [48] (VKITTI2), the KITTI-CARLA dataset [49] and the Driving dataset (DRIVING) [50]. They all contain sequences of left and right clear intensity images as well as the corresponding left and right ground truth depth maps. For each clear image, we first compute a distance map from its ground truth depth map, and then synthesise a corresponding foggy image by applying (4) rather than (1) (because the camera's photometric parameters are not available) to each channel. For all three colour channels, we fix A at $255 \times 0.7 = 178.5$, $255 \times 0.8 = 204.0$ and $255 \times 0.9 = 229.5$ for VKITTI2, KITTI-CARLA and DRIVING, respectively. These values of A fall within the typical range $[0.7, 1]$ which previous work, for example [51], [52], extensively adopted to synthesise foggy images. For each dataset, six different visibility levels at $V_{\text{MOR}} = \{30, 40, 50, 60, 70, 80\}$ meters are tested. The corresponding ground truth β values are calculated according to (3). See Fig. 7 for sample synthetic foggy images at various visibility levels, and see our supplementary material for a summary of the synthetic datasets we use for evaluation.

For evaluation on *real* foggy data we use our self-collected SDIRF dataset introduced in the previous section.

B. Competitive Methods

To the best of our knowledge, there is very limited existing work on estimating both A and β. Firstly, we report the results of Berman *et al.* [4] which estimates A only. We further compare our method with the fog estimation strategy proposed by Li *et al.* [6] (estimating both A and β) as well as *our modified version* of that strategy, which resulted in a major improvement over the original one. We made two main modifications: Firstly, as proposed by [53], to estimate A, we use the median, instead of the maximum [3], of the 0.1% pixels with the largest dark channel values. Secondly, when building the histogram of values of β, we discard values of β outside the range $[0.001, 0.2]$. This is motivated by the observation, typically at lower visibility, that a proportion of β values are negative and that there is a large cluster of β centred at the value of zero. In many situations, the zero bin has the highest counts, leading to a wrong estimate of β. Figs. 8(a) and 8(b) show examples of unbounded and bounded β histograms, respectively, at 30 m visibility. As we will show later, this modified version greatly improves the original one's performance and therefore we deem it to be a much stronger baseline method. Note that the above range of β that we use in the modified version of Li's method is consistent with the bounds of β that we set in our method (Section III-C3) when solving (16). This ensures a fair comparison between them.

C. Implementation Details

We empirically set $\xi_{\text{F}} = 4$ in (10), $\xi_{\text{K}} = 15$ in (11), $\eta = 2$ when determining the bounds for each relevant $L_{\text{c}n}$ (Section III-C3), and $\delta = 5$ (in the intensity domain) when defining the Huber loss in the first stage of our optimisation (Section III-C4). We fix these parameters throughout all our experiments.

To make a fair comparison, for all methods we use the stereo ORB-SLAM2 [11] to facilitate multiple observations of the

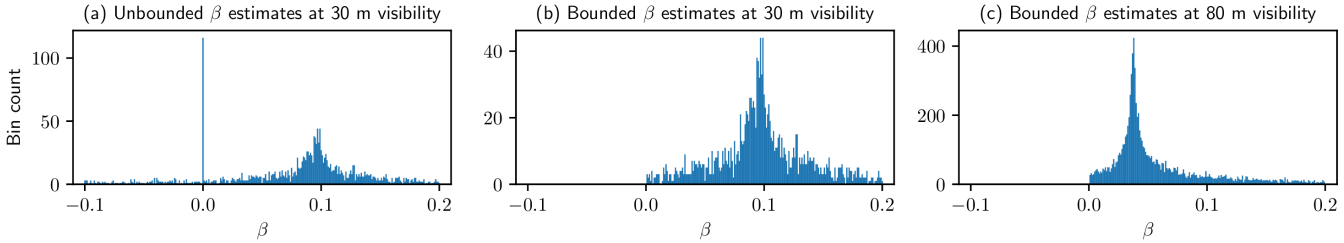

Fig. 8. β histogram examples generated by: (a) Li's method [6]; (b) and (c) our modified version of it. Note that the vertical axes have different scales. (a) Unbounded β estimates at 30 m visibility. The highest bin, which occurs at zero, leads to a wrong estimate of β. (b) Bounded (within the range $[0.001, 0.2]$) β estimates at 30 m visibility. The highest bin occurs at 0.097, which is much closer to the ground truth β value of 0.1. (c) Bounded (within the range $[0.001, 0.2]$) β estimates at 80 m visibility. Comparing (c) to (b), we observe that the total number of β estimates that are used to build the histogram is typically much larger at a higher visibility level.

same landmark from a range of known distances. These observations are established from ORB-SLAM2's local key frames and local map points after its local bundle adjustment. The fog parameters are updated after ORB-SLAM2's local mapping thread only after the ego-vehicle has moved at least five meters from the origin (for the very first estimation) or from the position of the last update (for subsequent estimations). We use the Ceres Solver [54] and choose the Levenberg–Marquardt algorithm to solve (16). See our supplementary material for more implementation details including pseudo-code.

Due to the way we generate synthetic foggy images, we have assumed that both g and g^{-1} are identity mappings for all colour channels. For real foggy data from SDIRF, in contrast, g and g^{-1} are characterised by the parameters α, γ and ζ found during our photometric calibration and therefore are channel-specific. Unless otherwise specified, the fog parameter results we report in the rest of this section are from the *grayscale* foggy images.

D. Evaluation on Synthetic Data

We conduct extensive experiments and present both quantitative and qualitative results.

1) Quantitative Results: We compute the root-mean-square error (RMSE), the mean-absolute error (MAE) and the standard deviation (SD), in both absolute and relative[4] scales, of the β and A[5] estimates. Table I shows the quantitative results of the average β and A error metrics on synthetic datasets[6].

We observe that in most cases our method performs the best, in terms of both estimation accuracy (i.e., smallest errors) and precision (i.e., the lowest standard deviation). The *only* exception is VKITTI2's A error metrics, but our β metrics are still the best in this case. A closer look at VKITTI2's results shows that our bad estimates of A stem from countryside scenes with sparse features (Scene02), or when the ego-vehicle is surrounded by other vehicles moving at similar speeds (Scene18). In either case the ORB-SLAM2's performance has been significantly degraded and therefore produces unreliable distance and/or intensity information. We will investigate how to address this limitation in our future work.

2) Qualitative Results: Fig. 9 illustrates how the estimates of β and A vary with frame at various visibility levels on scene "backwards" in DRIVING.

We observe that the values of β and A estimated by our method are closer to the ground truth values and more stable compared to other methods.

3) Error Metrics given Partial Ground Truth: We test the estimation performance of β on KITTI-CARLA given the ground truth value of A. The quantitative results are shown in the middle block of rows in the middle subtable of Table I.

We observe: a) As expected, all methods perform better when ground truth A is given; b) For Li's and Li's modified methods, there is a significant improvement in β's error metrics when the ground truth A is given. This is not surprising due to their sequential estimation strategy, since an error-free A will indeed benefit the subsequent estimation of β. This observation adds to the evidence that, in their method, any error in the estimate of A can propagate to the estimate of β; c) For our method, such improvement is much less significant. This may suggest that our method, when simultaneously optimising β and A with minimal prior knowledge, is able to find a β value that is not far from the optimal solution.

4) Ablation Study: We conduct an ablation study on KITTI-CARLA to better understand how our optimisation setup affects the performance of the fog parameter estimation. The following additional settings are experimented with: a) One-stage: Only the first stage of our optimisation is preserved; b) Uniform weight: We set $w_n^m = 1$ for all observations in both optimisation stages. The quantitative results are shown in the bottom block of rows in the middle subtable of Table I.

We observe: a) If one-stage optimisation is performed or a uniform weight is used, the estimation results are inferior to those produced by our full method; b) These two settings still outperform all competitive methods, despite trailing behind our full method.

5) Error Metrics vs Visibility: We investigate how the fog parameter estimation performance varies with visibility. Fig. 10 plots the relative RMSE of β and A against visibility.

We observe: a) Our method consistently excels by a large margin in both estimates of β and A for all visibility levels tested; b) Both Li's and Li's modified demonstrate a downward trend in the relative RMSE of β as visibility increases. By

[4]A relative metric is calculated as the ratio of the corresponding absolute metric to the ground truth value, and is shown as percentage in Table I and Fig. 10.

[5]A and L_∞ are essentially the same for synthetic data since g and g^{-1} have been defined as identity mappings.

[6]Result of Town04 in KITTI-CARLA at 30 m visibility is excluded as the ORB-SLAM2 loses tracking and provides no valid observation.

TABLE I
AVERAGE β AND A ERROR METRICS ON SYNTHETIC DATASETS. RELATIVE METRICS ARE SHOWN AS PERCENTAGE. FOR ALL THE METRICS, THE LOWER THE BETTER (↓). THE TABLE ALSO CONTAINS THE RESULTS ON OUR ABLATION STUDY USING KITTI-CARLA.

Dataset	Method	β						A					
		RMSE (↓)		MAE (↓)		SD (↓)		RMSE (↓)		MAE (↓)		SD (↓)	
		Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.
VKITTI2 [48]	Berman's [4]	N/A		N/A		N/A		4.7547	2.66	3.7004	2.07	2.9130	1.63
	Li's [6]	0.0430	66.39	0.0347	52.17	0.0225	37.17	3.8760	2.17	1.8920	1.06	3.3408	1.87
	Li's modified	0.0113	16.74	0.0080	11.47	0.0091	14.07	**1.5921**	**0.89**	**0.8374**	**0.47**	**1.2550**	**0.70**
	Ours	**0.0078**	**10.91**	**0.0061**	**8.57**	**0.0061**	**8.60**	2.5495	1.43	1.9276	1.08	1.8575	1.04
KITTI-CARLA [49]	Berman's [4]	N/A		N/A		N/A		12.6372	6.19	12.3382	6.05	2.6533	1.30
	Li's [6]	0.0499	84.01	0.0452	75.84	0.0199	34.66	15.9104	7.80	15.1337	7.42	4.3683	2.14
	Li's modified	0.0166	28.41	0.0143	24.35	0.0105	18.29	10.3676	5.08	9.6282	4.72	3.3599	1.65
	Ours	**0.0116**	**20.66**	**0.0101**	**18.11**	**0.0058**	**10.33**	**2.5711**	**1.26**	**1.8632**	**0.91**	**2.1219**	**1.04**
	Li's [6] (GT A)	0.0357	59.57	0.0250	41.48	0.0297	50.50	-		-		-	
	Li's modified (GT A)	0.0109	19.24	0.0081	14.47	0.0087	15.37	-		-		-	
	Ours (GT A)	0.0099	17.76	0.0087	15.57	0.0055	9.74	-		-		-	
	Ours (One-stage)	0.0122	21.27	0.0108	18.85	0.0063	10.81	3.4662	1.70	2.5064	1.23	2.8686	1.41
	Ours (Uniform weight)	0.0122	21.67	0.0107	19.05	0.0061	10.66	2.9282	1.44	2.2646	1.11	2.2818	1.12
DRIVING [50]	Berman's [4]	N/A		N/A		N/A		17.6183	7.68	11.6009	5.05	15.5607	6.78
	Li's [6]	0.0465	75.56	0.0387	62.39	0.0260	43.19	14.0963	6.14	11.2103	4.88	8.5101	3.71
	Li's modified	0.0168	26.85	0.0117	18.99	0.0129	20.56	12.9722	5.65	9.5988	4.18	8.7345	3.81
	Ours	**0.0051**	**8.98**	**0.0037**	**6.44**	**0.0033**	**6.66**	**1.9021**	**0.83**	**1.3379**	**0.58**	**1.6629**	**0.72**

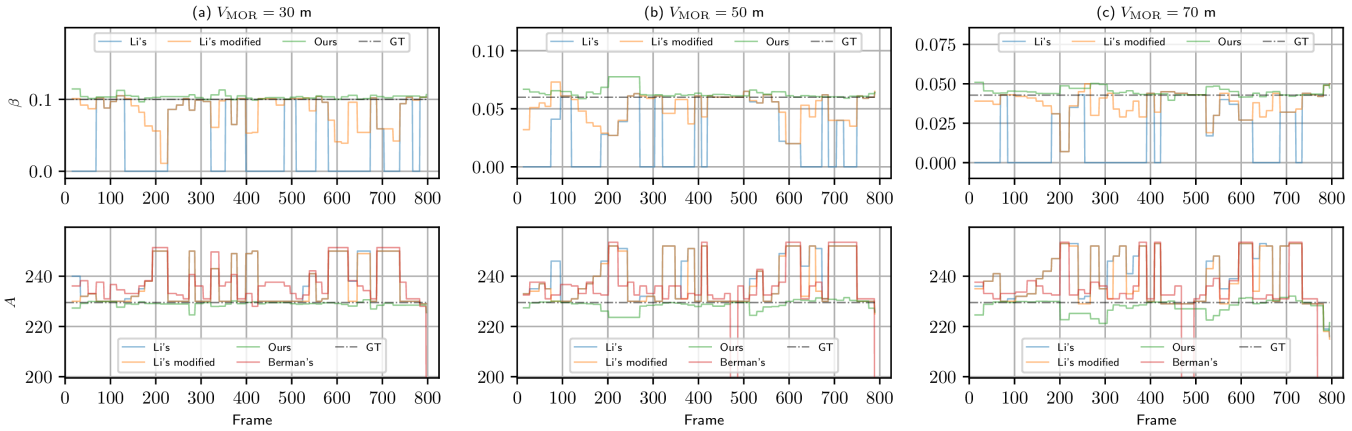

Fig. 9. Evaluating β and A estimates vs frame on scene "backwards" in DRIVING at various visibility levels. (a) 30 m. (b) 50 m. (c) 70 m. Ground truth values are indicated by black dotted lines. We observe that our modified version of Li's method improves the original one's performance slightly in the estimation of A and significantly in the estimation of β. Nevertheless, both of them, as well as Berman's method, result in many large errors. The performance of our method surpasses the rest by a large margin. We also highlight that how an error in the estimate of A propagates to the estimate of β in the results of Li's method and Li's modified method [an underestimate of β, for example around frame 200 in (a), occurs with an overestimate of A] is in line with the theoretical analysis of error propagation that we provide in our supplementary material.

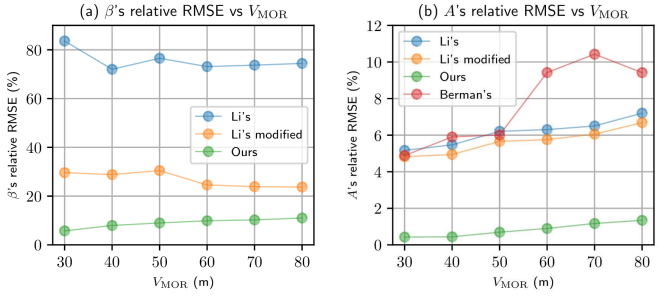

Fig. 10. Evaluating the relative RMSE vs V_{MOR} on DRIVING. (a) β. (b) A.

comparing the histogram in Fig. 8(c) with that in Fig. 8(b), we infer that as visibility increases, the number of estimates of β to build a histogram becomes larger, which in turn improves the performance of the statistics-based estimation method used by the two baseline methods; c) All methods witness an upward trend in the relative RMSE of A as visibility increases, which is expected because images will appear to be less fog-obscured as visibility increases.

E. Evaluation on Real Data

For real foggy data from SDIRF, we focus on *qualitative* evaluation because it is not possible to obtain the ground truth values of the fog parameters for real foggy images taken in an open, uncontrolled environment.

1) Scattering Coefficient Estimation: We examine both *inter-sequence* consistency and *intra-sequence* consistency between the perceptual density of the fog and the estimated β.

Firstly, in Fig. 11 we show the normalised histograms of β estimated by various methods of the eight foggy sequences

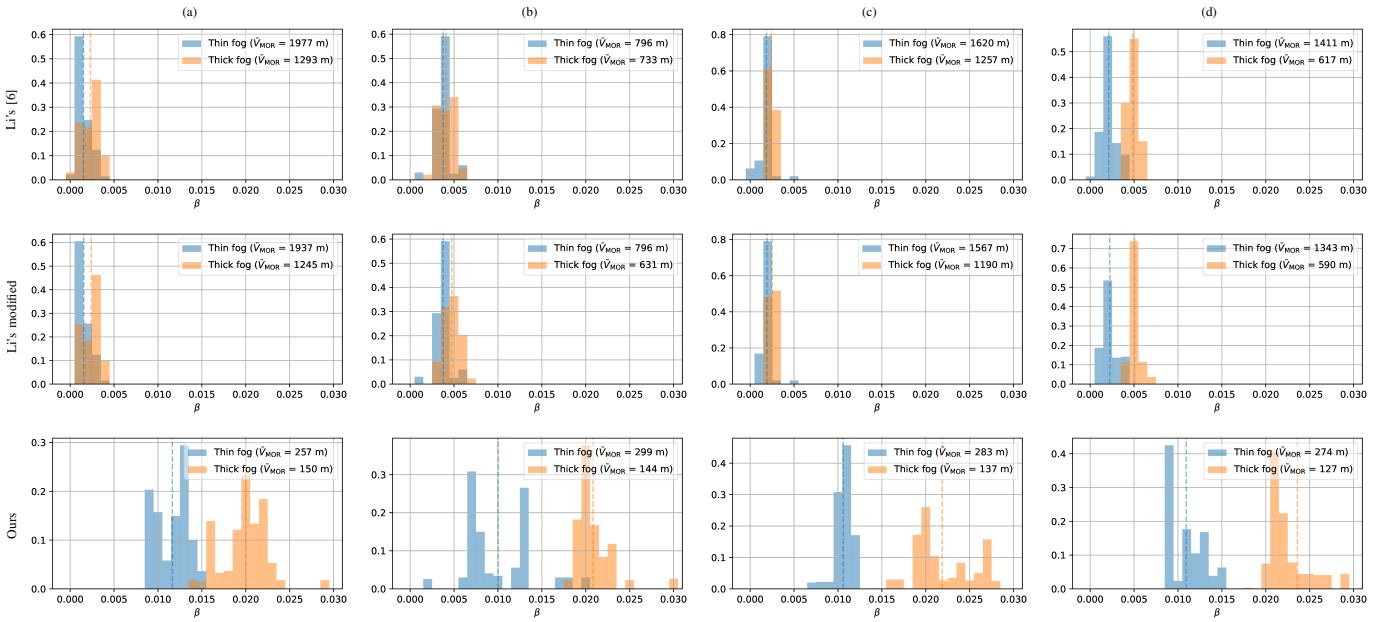

Fig. 11. Evaluating the inter-sequence consistency of β on SDIRF. We plot the normalised histograms of the estimated β of the whole thin/thick foggy sequences whose first left frames are shown in the corresponding columns of Fig. 3. Each row shows the results of a method. Note that all horizontal axes have the same scale. Our method is always the best at distinguishing between thin and thick fog by having the least overlap between the two distributions. Furthermore, in legend we show the mean visibility, $\bar{V}_{\text{MOR}}$, calculated from the mean β (indicated by the dashed vertical line) according to (3). We observe that the mean visibility values from our method are much more reasonable than the rest when compared with the foggy images in Fig. 3.

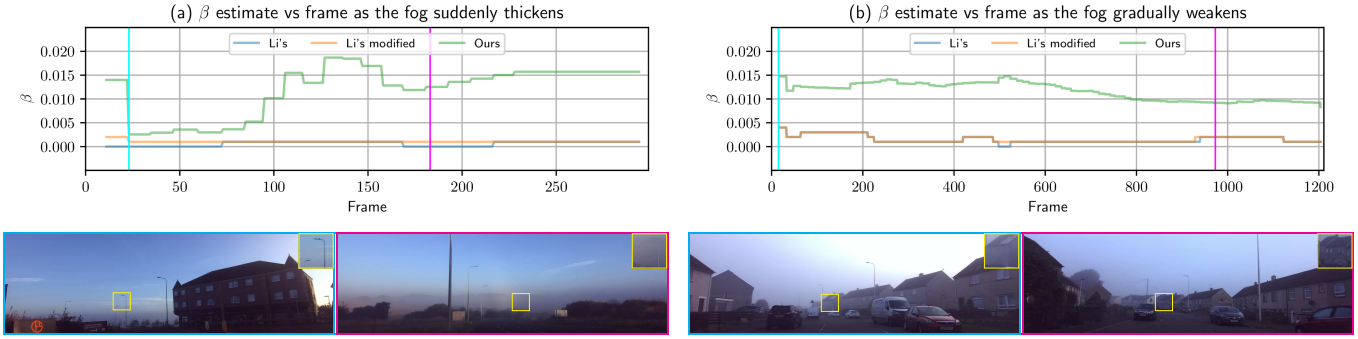

Fig. 12. Evaluating the intra-sequence consistency of β on SDIRF. We show the estimate of β vs frame of two foggy sequences in which the fog demonstrates a noticeable spatial variation in its density. (a) The vehicle traverses an area where the fog suddenly thickens. (b) The vehicle traverses an area where the fog gradually weakens. Shown at the bottom are the foggy images of the frames which are indicated by the vertical cursors in the corresponding plot above. Each image is bordered by the colour that matches the corresponding cursor's. See the close-up of yellow squares to better visually compare the fog density. We observe that our method is the only one that is able to respond adaptively to changes in fog density through updating the estimated values of β. These results also add to the evidence that our assumption of local homogeneity of fog is still valid when the fog is spatially variant.

whose first left frames are shown in Fig. 3. The key observation is that our method demonstrates the best inter-sequence consistency between the visual appearance of the foggy images and the estimated β. In addition, the mean visibility value computed according to (3) using the mean β value estimated by our method is perceptually more sensible than the rest. See our supplementary material for more results.

Secondly, in Fig. 12 we plot the estimated β vs frame of two foggy sequences in which the fog demonstrates a noticeable spatial variation in its density. The key observation is that our method, being the only one that is capable of responding adaptively to spatial variation in fog density, demonstrates the best intra-sequence consistency between the visual appearance of the foggy images and the estimated β.

2) Atmospheric Light Estimation: For real foggy images, atmospheric light is no longer monochrome. Therefore, we apply our method to each colour channel.

Firstly, we visually compare the atmospheric light estimated by various methods. After obtaining the estimate of L_∞ of each colour channel by each method, for visualisation purposes we map it back to pixel intensity A by applying g^{-1} so that its colour can be illustrated. Sample results are shown in Fig. 13. The key observation is that the atmospheric light estimated by our method is the closest to the colour of the horizon, i.e., the most fog-opaque region in a foggy image. We also observe that our method is more robust to changes in the atmospheric light than competing ones.

Next, we take a step further by visually comparing the defogging results using L_∞ estimated by various methods. To facilitate a fair comparison between all methods such that the only difference is L_∞, we follow [3] to estimate t and perform defogging. Again, for visualisation purposes, we apply g^{-1} to

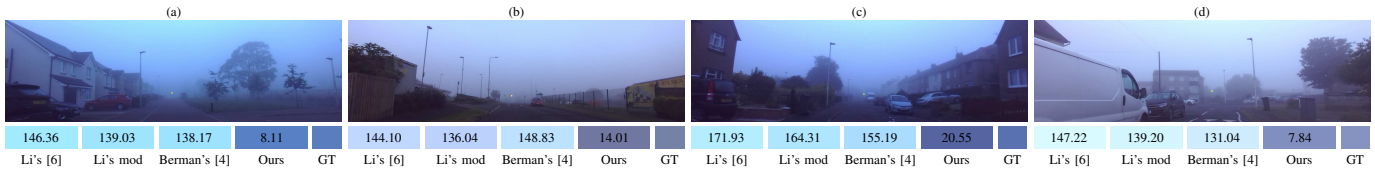

Fig. 13. Evaluating the accuracy of the estimated A on SDIRF. The small rectangles at the bottom are painted the colours of A estimated by various methods. We also show A's pseudo-ground truth colour, which is extracted by visually examining each foggy image then manually selecting a pixel just above the horizon in the central area (see the little yellow square in each foggy image). Each rectangle is overlaid with the Euclidean distance from the corresponding estimate to the pseudo-ground truth. Note that the real foggy images are typically very different from the simulated ones shown in Fig. 7 in a way that the sky region is not of a uniform colour, which is particularly the case of a foggy image taken at dawn. We infer that this phenomenon causes competitive methods to fail as they all estimate the atmospheric light from a single image. The results demonstrate that only our method is able to accurately unveil the atmospheric light. In addition, our method is more robust to changes in the atmospheric light.

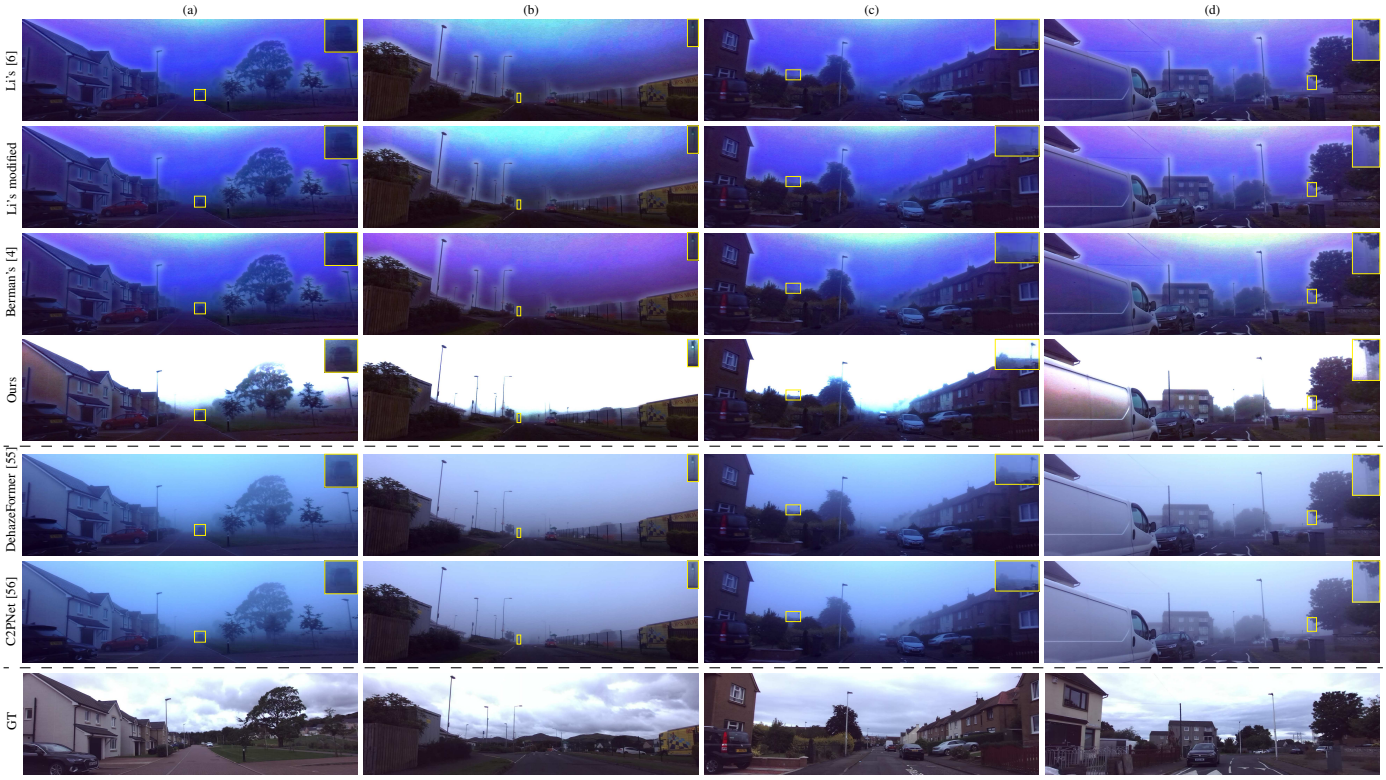

Fig. 14. Evaluating the atmospheric light estimated by various methods (top four rows) on SDIRF by using it to perform defogging following [3]. The input foggy images are shown in the corresponding columns of Fig. 13. The corresponding clear images recorded in overcast weather are shown in the last row to serve as pseudo-ground truth. We observe: a) The defogged images which use the atmospheric light estimated by our method appear to be more accurate in colour compared to others with minimal visual artefacts; b) Using the atmospheric light estimated by our method, fog on distant objects seems to be better removed (see the close-up of yellow squares). In addition, we investigate how two state-of-the-art, end-to-end deep learning-based defogging methods, DehazeFormer [55] and C2PNet [56], perform on the same foggy images (in the intensity domain, as their networks were trained on intensity images). The results are shown in the middle two rows between the dashed lines. The key observation is that for both methods their defogging effect is barely visible.

the defogging results, and the final defogged images are shown in the top four rows in Fig. 14. The clear images recorded in overcast weather are shown in the last row as pseudo-ground truth. The key observation is that using the atmospheric light estimated by our method yields defogged images that are perceptually superior to those produced by other methods.

In addition, we evaluate two state-of-the-art, end-to-end deep learning-based defogging methods, DehazeFormer [55] and C2PNet [56], on the same foggy images (in the intensity domain, as their networks were trained on intensity images). The results are shown in the middle two rows between the dashed lines in Fig. 14. Compared with the foggy images in Fig. 13, their defogging effect is barely visible.

See our supplementary material for more results, including comparisons of the estimated atmospheric light and the defogged images obtained in the radiance and intensity domains.

F. Additional Experiments

We report two additional experiments to demonstrate a) our results on β's wavelength dependence align with what was reported from previous physics experiments; b) the non-linearity introduced by gamma correction cannot be ignored when estimating β.

1) Scattering Coefficient's Wavelength Dependence: In this experiment, we apply our method to each colour channel (RGB) as well as to the grayscale image independently, and investigate how β varies with wavelength.

To this end, we examine all β estimates from a total of 34457 frames evaluated on all foggy sequences of SDIRF.

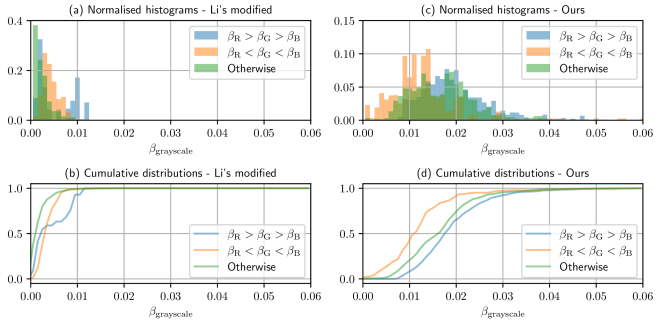

Fig. 15. Investigating β's wavelength dependence on all foggy sequences in SDIRF. The results of Li's modified method are shown in (a) and (b), and the results of our method are shown in (c) and (d). From (c) and (d) we observe that the case $\beta_R > \beta_G > \beta_B$ tends to happen at a larger β value, whereas the case $\beta_R < \beta_G < \beta_B$ tends to happen at a smaller β value. The remaining cases tend to happen at intermediate β values. The above observations are in line with [57, Figure 6.12]. However, they cannot be made from (a) or (b).

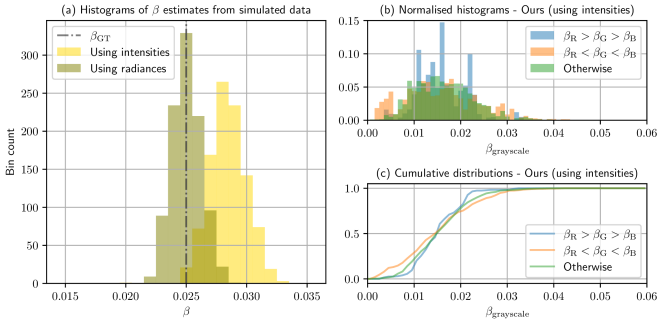

Fig. 16. (a) Investigating how the gamma correction affects the estimate of β using simulated data. We observe that using intensities rather than radiances overestimates β. Comparing (b) and (c) with their counterparts [Figs. 15(c) and 15(d)], we can see that using intensities rather than radiances makes both the histograms and the cumulative distribution curves significantly overlap.

We categorise the β values obtained from each frame into the following three cases: a) $\beta_R > \beta_G > \beta_B$ (i.e., β strictly increases with wavelength); b) $\beta_R < \beta_G < \beta_B$ (i.e., β strictly decreases with wavelength); c) Otherwise. We investigate how these three cases are distributed as the grayscale scattering coefficient $\beta_{grayscale}$, which measures the mean visibility, changes for both Li's modified method and our method. The results are are illustrated in Fig. 15 as normalised histograms [(a) and (c)], and as cumulative distribution curves [(b) and (d)].

The key observation is that the results of our method align with [57, Figure 6.12], which shows that at lower visibility (i.e., a larger β) the relative attenuation of different colour channels tends to increase with wavelength (i.e., $\beta_R > \beta_G > \beta_B$), whereas at higher visibility (i.e., a smaller β) it tends to decrease with wavelength (i.e., $\beta_R < \beta_G < \beta_B$). In contrast, the results of Li's modified method do not show such trends.

2) Gamma Correction: We investigate the effect of the non-linearity caused by gamma correction on the estimate of β.

Firstly, we conduct the following experiment using simulated data. We generate clean data consisting of the radiances of a number of landmarks observed from a range of distances according to (1) [i.e., drawing samples from the dotted lines shown in Fig. 2(b)] with ground truth $\beta_{GT} = 0.025$, which is then corrupted with random Gaussian noise. We then apply g^{-1} (with $\gamma > 1$ in accordance with our photometric calibration results) to convert the radiance data to intensity data. Next, we use our proposed method to estimate two β values, one from the radiance data and one from the intensity data. The experiment is repeated 1000 times. Due to random noise, each instance leads to slightly different values for β. We plot their histogram in Fig. 16(a). We observe that using intensities we tend to overestimate β, whereas using radiances the estimates seem to be unbiased. In fact, estimating β using intensity *always* yields estimates larger than using radiance. Our experiment also reveals that a) the direction of the bias depends on whether $\gamma > 1$ or $\gamma < 1$; b) the amount of the bias increases as γ deviates from 1. See our supplementary material for more details of the experiment and the results.

Finally, in Figs. 16(b) and 16(c) we show the counterparts of Figs. 15(c) and 15(d) but using intensities instead of radiances. We observe that this time the histograms and the cumulative distribution curves significantly overlap, which adds to the evidence that the atmospheric scattering model should be applied to radiances rather than to intensities.

VI. Conclusion

We presented an optimisation-based method for estimating the parameters of fog. While prior methods adopt a sequential estimation strategy that is prone to error propagation, our method simultaneously estimates the parameters by solving a minimisation problem. Extensive experiments show that our method outperforms prior ones on synthetic data both qualitatively and quantitatively, and on real data qualitatively from various aspects. Our method has the potential to be plugged into an existing feature-based visual SLAM/odometry system as an add-on module for its deployment in fog. In addition, we have introduced SDIRF, a dataset consisting of high-quality, consecutive stereo foggy images of real road scenes under a variety of visibility conditions. SDIRF also provides calibrated photometric parameters, which makes it photometrically ready to apply the atmospheric scattering model, as well as counterpart clear images taken in overcast weather of the same routes, which will be useful for companion work in image defogging and depth reconstruction. All of the above features together make SDIRF a first-of-its-kind dataset for the study of visual perception for autonomous driving in fog.

In the future, we will investigate how to improve the resilience of our method when the underlying visual SLAM system struggles to generate accurate distance and/or intensity (hence radiance) information, which is a limitation of our current method. Our experimental results in Section V-D1 suggest that these situations arise in countryside scenes with very sparse features or when the ego-vehicle is surrounded by other vehicles moving at similar speed. To this end, we will consider the following two approaches: a) to more tightly couple our method with a visual SLAM system by jointly optimising the fog parameters, the camera's poses, and the landmark's 3D positions; b) to integrate our method into a visual-inertial SLAM system (e.g., ORB-SLAM3 [58]) that is inherently more robust in the presence of fog.

ACKNOWLEDGEMENTS

The authors thank Aongus McCarthy for offering help and equipment for the photometric calibration. This work is supported in part by U.K.'s Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/S023208/1.

REFERENCES

[1] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, pp. 233–254, 2002.

[2] W. M. Organization, "Measurement of meteorological variables," 2014, last accessed 15 March 2024. [Online]. Available: https://library.wmo.int/index.php?lvl=notice_display&id=19673#.Yzw6YTfMJhE

[3] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[4] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1674–1682.

[5] L. Caraffa and J.-P. Tarel, "Stereo reconstruction and contrast restoration in daytime fog," in *Asian Conference on Computer Vision*. Springer, 2013, pp. 13–25.

[6] Z. Li, P. Tan, R. T. Tan, D. Zou, S. Z. Zhou, and L.-F. Cheong, "Simultaneous video defogging and stereo reconstruction," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4988–4997.

[7] Y. Ding, A. M. Wallace, and S. Wang, "Variational simultaneous stereo matching and defogging in low visibility," in *British Machine Vision Conference*. BMVA Press, 2022.

[8] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003.

[9] R. T. Tan, "Visibility in bad weather from a single image," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[10] R. Szeliski, *Computer vision: algorithms and applications*. Springer Nature, 2022.

[11] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[12] Y. Ding, A. M. Wallace, and S. Wang, "Estimating fog parameters from an image sequence using non-linear optimisation," in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 1578–1586.

[13] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 598–605.

[14] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. I–I.

[15] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2012.

[16] C. Chen, M. N. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *European Conference on Computer Vision*. Springer, 2016, pp. 576–591.

[17] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[18] B. Cai, X. Xu, and D. Tao, "Real-time video dehazing based on spatio-temporal mrf," in *Advances in Multimedia Information Processing-PCM 2016*. Springer, 2016, pp. 315–325.

[19] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888–3901, 2015.

[20] Z. Ling, J. Gong, G. Fan, and X. Lu, "Optimal transmission estimation via fog density perception for efficient single image defogging," *IEEE Transactions on Multimedia*, vol. 20, no. 7, pp. 1699–1711, 2018.

[21] H. Guo, X. Wang, and H. Li, "Density estimation of fog in image based on dark channel prior," *Atmosphere*, vol. 13, no. 5, 2022.

[22] N. Hautiere, J.-P. Tarel, J. Lavenant, and D. Aubert, "Automatic fog detection and estimation of visibility distance through use of an onboard camera," *Machine Vision and Applications*, vol. 17, no. 1, pp. 8–20, 2006.

[23] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[24] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.

[25] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.

[26] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 6433–6438.

[27] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in perception for autonomous driving: Waymo open dataset," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2443–2451.

[28] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 618–11 628.

[29] J. Mao, M. Niu, C. Jiang, H. Liang, J. Chen, X. Liang, Y. Li, C. Ye, W. Zhang, Z. Li *et al.*, "One million scenes for autonomous driving: Once dataset," *arXiv preprint arXiv:2106.11037*, 2021.

[30] G. Yang, X. Song, C. Huang, Z. Deng, J. Shi, and B. Zhou, "Drivingstereo: A large-scale dataset for stereo matching in autonomous driving scenarios," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 899–908.

[31] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2633–2642.

[32] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 679–11 689.

[33] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "Radiate: A radar dataset for automotive perception in bad weather," in *IEEE International Conference on Robotics and Automation*, 2021, pp. 1–7.

[34] L. Daniel, D. Phippen, E. Hoare, A. Stove, M. Cherniakov, and M. Gashinova, "Low-thz radar, lidar and optical imaging through artificially generated fog," in *International Conference on Radar Systems*, 2017, pp. 1–4.

[35] T. Gruber, M. Bijelic, F. Heide, W. Ritter, and K. Dietmayer, "Pixel-accurate depth evaluation in realistic driving scenarios," in *International Conference on 3D Vision*, 2019, pp. 95–105.

[36] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 154–169.

[37] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *IEEE International Conference on Computer Vision*, 2017, pp. 4780–4788.

[38] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.

[39] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, pp. 973–992, 2018.

[40] F. Liu, C. Shen, G. Lin, and I. Reid, "Learning depth from single monocular images using deep convolutional neural fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 10, pp. 2024–2039, 2016.

[41] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: Joint color and depth completion from stereo images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[42] T. Song, Y. Kim, C. Oh, H. Jang, N. Ha, and K. Sohn, "Simultaneous deep stereo matching and dehazing with feature attention," *International Journal of Computer Vision*, vol. 128, pp. 799–817, 2020.

[43] C. Yao and L. Yu, "Foggystereo: Stereo matching with fog volume representation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 13 033–13 042.

[44] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," in *IEEE International Conference on Robotics and Automation*, 2011, pp. 3607–3613.
[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[46] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier analysis and applications*, vol. 14, pp. 877–905, 2008.
[47] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 63, no. 1, pp. 1–38, 2010.
[48] Y. Cabon, N. Murray, and M. Humenberger, "Virtual kitti 2," *arXiv preprint arXiv:2001.10773*, 2020.
[49] J.-E. Deschaud, "KITTI-CARLA: a KITTI-like dataset generated by CARLA Simulator," *arXiv preprint arXiv:2109.00892*, 2021.
[50] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4040–4048.
[51] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.
[52] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *International Journal of Computer Vision*, vol. 128, pp. 240–259, 2020.
[53] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2995–3002.
[54] S. Agarwal, K. Mierle, and T. C. S. Team, "Ceres Solver," 10 2023. [Online]. Available: https://github.com/ceres-solver/ceres-solver
[55] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Transactions on Image Processing*, vol. 32, pp. 1927–1941, 2023.
[56] Y. Zheng, J. Zhan, S. He, J. Dong, and Y. Du, "Curricular contrastive regularization for physics-aware single image dehazing," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5785–5794.
[57] E. J. McCartney, *Optics of the atmosphere: scattering by molecules and particles*. John Wiley & Sons, 1976.
[58] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

Yining Ding received the B.Eng. (Hons.) degree in electronics and electrical engineering from the University of Edinburgh, Edinburgh, U.K., in 2013, the M.Sc. degree in communications and signal processing from Imperial College London, London, U.K., in 2014, and the Ph.D. degree in robotics and autonomous systems from the Edinburgh Centre for Robotics, Edinburgh, U.K., in 2025. He worked as a design engineer on ultrasound signal processing for industrial metrology applications at Renishaw plc, U.K., between 2014 and 2020. His research focuses on robust visual perception in fog, including fog parameter estimation, depth reconstruction, and image defogging.

João F. C. Mota received the M.Sc. and Ph.D. degrees in electrical and computer engineering from the Technical University of Lisbon, Lisbon, Portugal, in 2008 and 2013, respectively, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2013. He is currently an Assistant Professor of Signal and Image Processing with Heriot-Watt University, Edinburgh, U.K. His research interests include theoretical and practical aspects of high-dimensional data processing, inverse problems, optimization theory, machine learning, data science, and distributed information processing and control. Dr. Mota was a recipient of the 2015 IEEE Signal Processing Society Young Author Best Paper Award and is currently Associate Editor for IEEE Transactions on Signal Processing.

Andrew Michael Wallace received the B.Sc. and Ph.D. degrees from the University of Edinburgh, Edinburgh, U.K., in 1972 and 1975, respectively. He was an Emeritus Professor of Signal and Image Processing with Heriot-Watt University, Edinburgh, U.K. His research interests included LiDAR and 3D vision, image and signal processing, and accelerated computing. He had authored or coauthored extensively and had secured funding from EPSRC, the EU and other sponsors. He was a Chartered Engineer and a fellow of the Institute of Engineering Technology.

Sen Wang received the Ph.D. degree in robotics from the University of Essex, Colchester, U.K., in 2015. He is currently an Associate Professor with the Sense Robotics Lab, Imperial College London, London, U.K. His research interests include robot perception and autonomy using probabilistic and learning approaches, especially autonomous navigation, robotic vision, SLAM, and robot learning. He has served as an Associate Editor for IEEE Transactions on Robotics, IEEE Transactions on Automation Science and Engineering and IEEE Robotics and Automation Letters.