Estimating Fog Parameters from an Image Sequence using Non-linear Optimisation - Supplementary Material

Yining Ding Andrew M. Wallace
Edinburgh Centre for Robotics
Heriot-Watt University, Edinburgh, UK
{yd2007, a.m.wallace}@hw.ac.uk

Sen Wang
Sense Robotics Lab
Imperial College London, London, UK
sen.wang@imperial.ac.uk

1. Datasets used for Evaluations

Tab. 1 summarises the datasets we use for evaluations. Note that we downsize the images from KITTI-CARLA by a factor of two, and crop pixels at the bottom to exclude the ego-vehicle's bonnet from the view. The final image size is 492×696. For the other two datasets the original image sizes are kept.

2. Implementation Details

As was mentioned in our paper, our method can potentially be plugged into any visual SLAM/odometry system as an add-on module. Below we use the well-known stereo ORB-SLAM2 [13] as an example to detail our method.

2.1. Extracting a Local Map

This is the first step of our method. A local map is extracted around the ORB-SLAM2's current key frame after its Local BA, at which stage the key frame (KF) poses and the map point (MP) positions are relatively accurate. Given the current key frame (denoted as $\mathrm{KF_c}$.), we first identify its local key frames (denoted as a set $\mathsf{LKF_c}$), then find its local map points (denoted as a set $\mathsf{LMP^c}$). $\mathsf{LKF_c}$ basically consists of all key frames that are covisible[1] to $\mathrm{KF_c}$. $\mathsf{LMP^c}$ basically contains all map points that can be observed by at least one key frame in $\mathsf{LKF_c}$.

The pseudo code of this step is shown in Algorithm 1.

2.2. Generating Distance-Intensity Pairs

This is the second step of our method. We use d_n^m to denote the Euclidean distance between KF_m and MP_n, which can be calculated given the pose of KF_m and the position of MP_n. We use I_n^m to denote the pixel intensity at MP_n's corresponding 2D feature point in KF_m. A bilinear interpolation is performed in case of a non-integer pixel location.

The pseudo code of this step is shown in Algorithm 2.

[1]Two frames are called covisible if there exists at least one map point that is visible to both of them.

2.3. Setting the Intensity Bounds

Algorithm 3 details how we set lower and upper bounds on A and relevant J's for our optimisation.

3. Experiments on the Intensity

As was mentioned in the previous section, when generating the distance-intensity pairs, we use the intensity at a MP's corresponding 2D feature point in the image. From now on we call this intensity the *raw* intensity. We have also experimented with the use of the following two intensities.

1. The *projected* intensity - This is the intensity value at the pixel location where the MP projects onto the image plane of the camera. A bilinear interpolation is performed in case of a non-integer pixel location.

2. The *refined* intensity - This is the intensity value obtained after aligning a local image patch around the feature point, which will be explained in detail below.

3.1. The Refined Intensity

First, out of all KFs that can observe an MP, we choose the KF that has the median distance to it as the reference KF. Then, we consider an image patch around the MP's corresponding feature point in the reference KF to be the reference patch. Next, we reproject the reference patch and then project it onto every query KF, which is basically every KF other than the reference KF that can observe the same MP. Finally, for every query KF we seek a pair of image coordinates which best aligns (*i.e.* minimises the sum of squared differences) its surrounding patch to the projected reference patch by applying the inverse compositional algorithm [1, Fig. 4], which is iterative and we use the corresponding 2D feature point location in the query KF to initialise. Additionally, we normalise each image patch (*i.e.* subtract its mean then divide by its standard deviation) before performing the alignment in order to mitigate the effect

Dataset	Source clear images	Image size (H × W)	A (in the range $[0, 255]$)	V_{met} in meters	Total # of sequences
VKITTI2 [4]	'overcast', all five scenes	375×1242	0.7 (178.5)	$\{30, 40, 50, 60, 70, 80\}$	$5 \times 6 = 30$
KITTI-CARLA [6]	all seven towns	492×696	0.8 (204)		$7 \times 6 = 42$
DRIVING [12]	'finalpass', 15mm focal length, 'slow', 'backwards' and 'forwards' scenes	540×960	0.9 (229.5)		$2 \times 6 = 12$

Table 1. Datasets used for evaluation

Algorithm 1: Extracting a local map

Input: KF_c
Output: LMP^c
$\text{LMP}^c \leftarrow \emptyset$; // initialise LMP^c as an empty set
find LKF_c ; // find the set of all KFs that are covisible to KF_c
foreach $KF_m \in \text{LKF}_c$ **do** // loop through all KFs in LKF_c
  foreach $\langle m, n \rangle \in \text{M}^m$ **do** // loop through all MPs that can be observed by KF_m
    if $MP_n \notin \text{LMP}^c$ **then** // if MP_n has not been added to LMP^c
      $\text{LMP}^c \leftarrow \text{LMP}^c \cup \{\text{MP}_n\}$; // add MP_n to LMP^c

of reduced local contrast caused by fog. We compute the intensity at the optimised image coordinates as the refined intensity. Again, a bilinear interpolation is performed in case of a non-integer pixel location.

Our image patch alignment operation is inspired by the feature alignment step in SVO [8, Sec. IV-B], in which they consider it as a relaxation step that violates the epipolar constraints in order to achieve a higher correlation between the feature patches. In contrast, our intention here is to refine the intensity data in order to robustify the subsequent fog parameter estimation at the cost of a slight deviation from the best KF poses and MP positions found by the local BA. Note that this step is for the sole purpose of refining the intensity data, and does not have to modify the KF poses or the MP positions in the underpinning visual SLAM/odometry system.

3.2. Results

Tab. 2 shows the quantitative results of our intensity experiment evaluated on DRIVING.

We observe: a) Using the projected intensity, the results are not as good as using the raw intensity. This is likely to be caused by errors in the KF poses and MP positions. Bear in mind that the input images are foggy and therefore ORB-SLAM2's performance can be severely compromised; b) Using the refined intensity gives a comparable performance to using the raw intensity in estimating β, but gives the worst performance in estimating A. A plausible explanation is that the image patch alignment does not work well when trying to align two patches that are of quite different local contrasts (because the image local contrast gets degraded *asynchronously* by fog depending on the distance between the scene point and the camera).

We also notice that because the inverse compositional algorithm is iterative, when the local map is of a large scale, generating the refined intensity of each observation would add a considerable amount of computational time.

In conclusion, we choose the raw intensity as our default setting and report its results in the paper.

4. More Qualitative Results

4.1. A and β Estimates vs. Frames

In Fig. 1 we present how β and A estimates vary with frames at various visibilities on the scene 'backwards' in DRIVING.

4.2. Using A and β Estimated by Our Method

To further demonstrate how the fog parameters estimated by our method can be used for downstream image defogging and depth estimation tasks, in Fig. 2 we present more qualitative results after feeding the fog parameters estimated by the proposed method to our previous work [7].

5. The Need of a New Dataset

In Tab. 3 we list a number of popular, *real* datasets that are publicly available for researchers in the autonomous driving and computer vision communities to use.

We make the following observations.

- Most datasets do not contain images recorded in foggy weather.
- Out of the few datasets which contain foggy images, only RADIATE and DrivingStereo contain consecutive, left and right frames. Consecutive frames are

Algorithm 2: Generating distance-intensity pairs

Input: LMP^c
Output: DI

$\mathsf{DI} \leftarrow \emptyset$; // initialise DI as an empty set
foreach $MP_n \in \mathsf{LMP}^c$ **do** // loop through all MPs in LMP^c
  if $|\mathsf{M}_n| < 4$ **then** // if MP_n is observed by too few KFs
    continue;
  $\mathsf{DI}_n \leftarrow \emptyset$; // initialise DI_n as an empty set
  foreach $\langle m, n \rangle \in \mathsf{M}_n$ **do** // loop through all KFs that can observe MP_n
    $\mathsf{DI}_n \leftarrow \mathsf{DI}_n \cup (d_n^m, I_n^m)$; // add the distance-intensity pair to DI_n
  $\mathsf{DI} \leftarrow \mathsf{DI} \cup \mathsf{DI}_n$; // add DI_n to DI

Algorithm 3: Setting the bounds on A and relevant J's

Input: DI
Output: $l_A, u_A, \{l_{J_n}, u_{J_n} \mid \mathsf{DI}_n \in \mathsf{DI}\}$

$\mathsf{LB}_A \leftarrow \emptyset$; // initialise the candidate sets of A's lower bounds as an empty set
foreach $\mathsf{DI}_n \in \mathsf{DI}$ **do** // loop through all distance-intensity pairs
  $k_n \leftarrow (I_n^{d_{\max}} - I_n^{d_{\min}}) / (d_{\max} - d_{\min})$; // compute the slope of the line
  if $k_n > 2$ **then** // if J_n is smaller than A
    $l_{J_n} \leftarrow 0$;
    $u_{J_n} \leftarrow I_n^{d_{\min}}$;
    $\mathsf{LB}_A \leftarrow \mathsf{LB}_A \cup I_n^{d_{\max}}$;
  else if $k_n < -2$ **then** // else if J_n is greater than A
    $l_{J_n} \leftarrow I_n^{d_{\min}}$;
    $u_{J_n} \leftarrow 255$;
  else // undetermined whether J_n is smaller or greater than A, set loose bounds
    $l_{J_n} \leftarrow 0$;
    $u_{J_n} \leftarrow 255$;
if $\mathsf{LB}_A \neq \emptyset$ **then** // if LB_A is not empty
  $l_A \leftarrow \text{median}\,(\mathsf{LB}_A)$;
else // if LB_A is empty
  $l_A \leftarrow 0$;
$u_A \leftarrow 255$;

Intensity mode	β			A		
	RMSE (%)	MAE (%)	SD (%)	RMSE (%)	MAE (%)	SD (%)
Raw (*i.e.* **Ours** in the paper)	**0.0060** (10.66)	**0.0044** (7.69)	**0.0043 (7.74)**	**2.3435 (1.02)**	**1.6025 (0.70)**	**2.0229 (0.88)**
Projected	0.0066 (11.65)	0.0052 (9.07)	0.0044 (7.84)	2.3697 (1.03)	1.6453 (0.72)	2.1615 (0.94)
Refined	0.0062 (**10.00**)	0.0044 (**7.11**)	0.0059 (9.60)	3.3864 (1.48)	2.5242 (1.10)	3.1021 (1.35)

Table 2. Averaged β and A error metrics on DRIVING comparing different intensity modes

required to maximise the performance of a visual SLAM/odometry system. Meanwhile, stereo images with a known baseline are required in estimating β because the distance d needs to be recovered to an absolute scale (*cf*. Eq. (2) in our paper).

- Only one out of the four foggy sequences in RADIATE 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 (see the left of Fig. 3).
- All four sequences that are labelled 'foggy' in DrivingStereo are, however, not really foggy - the visibility is still very good (see the right of Fig. 3).
- Neither RADIATE nor DrivingStereo contains any counterpart sequence of a foggy one but taken in clear weather of the same route. Such data is essential if one wants to quantitatively evaluate the downstream depth estimation or image defogging performance.

We make a final remark that none of the listed datasets comes with information about its camera's photometric calibration. The atmospheric scattering model is actually de-

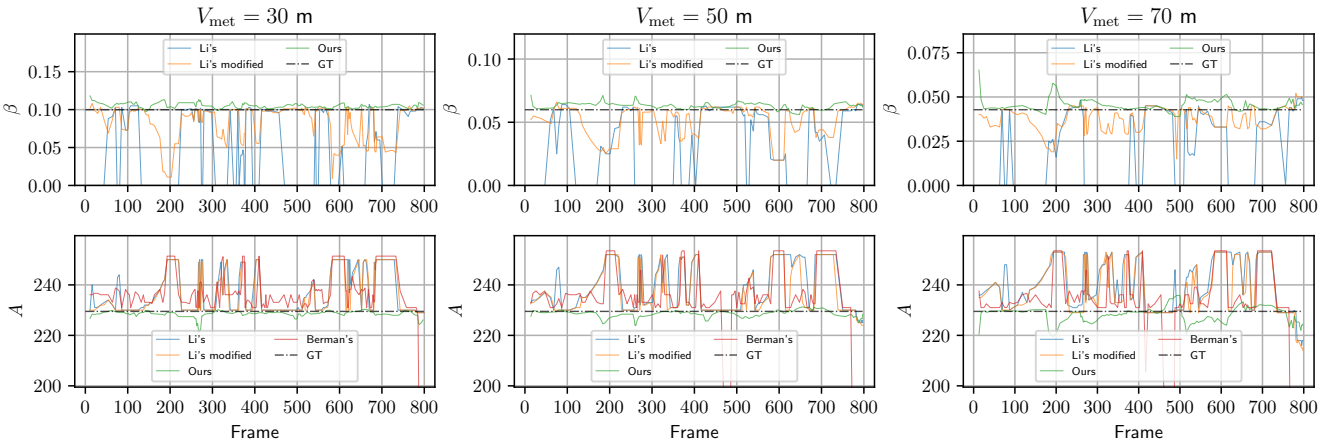

Figure 1. β and A estimates *vs*. frame evaluated on the scene 'backwards' in DRIVING at various visibilities (left to right: 30, 50 and 70 m).

Dataset	Stereo images	Consecutive frames	Foggy images	Counterpart clear images
KITTI Odometry [9]	✓	✓	✗	N/A
Oxford (Radar) RobotCar [2, 10]	✓	✓	✗	N/A
nuScenes [5]	✓	✓	✗	N/A
Waymo [16]	✓	✓	✗	N/A
ONCE [11]	✓	✓	✗	N/A
Seeing Through Fog [3]	✓	✗ (only the left consecutive frames are published)	✓	✗
BDD100K [18]	✗ (monocular)	✓	✓	✗
RADIATE [15]	✓	✓	✓	✗
DrivingStereo [17]	✓	✓	✓	✗

Table 3. Comparison of some popular datasets for autonomous driving applications

rived in the context of radiometry [14]. When a camera captures an image of a scene and saves an object's radiance as a pixel intensity value, there exist a few stages in its sensing pipeline that inevitably introduce non-linear mappings. The most prominent stage is called gamma correction. As a result, ideally the saved images need to be gamma-undone before the atmospheric scattering model can be applied. This inverse mapping process would require the camera to be photometrically calibrated.

To conclude, to the best of our knowledge, none of the existing real datasets is suitable for us to use for evaluation. We therefore resort to synthetic foggy images. Collecting a dataset which addresses all aforementioned issues is a priority.

References

[1] Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 2004. 1

[2] Dan Barnes, Matthew Gadd, Paul Murcutt, Paul Newman, and Ingmar Posner. The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset. In *ICRA*, 2020. 4

[3] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *CVPR*, 2020. 4

[4] Yohann Cabon, Naila Murray, and Martin Humenberger. Virtual kitti 2. *arXiv preprint arXiv:2001.10773*, 2020. 2

[5] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020. 4

[6] Jean-Emmanuel Deschaud. KITTI-CARLA: a KITTI-like dataset generated by CARLA Simulator. *arXiv preprint arXiv:2109.00892*, 2021. 2

[7] Yining Ding, Andrew M. Wallace, and Sen Wang. Variational simultaneous stereo matching and defogging in low visibility. In *BMVC*, 2022. 2, 5

[8] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. Svo: Fast semi-direct monocular visual odometry. In *ICRA*, 2014. 2

[9] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012. 4

[10] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The oxford robotcar dataset. *IJRR*, 2017. 4

[11] Jiageng Mao, Minzhe Niu, Chenhan Jiang, Hanxue Liang, Jingheng Chen, Xiaodan Liang, Yamin Li, Chaoqiang Ye, Wei Zhang, Zhenguo Li, et al. One million scenes

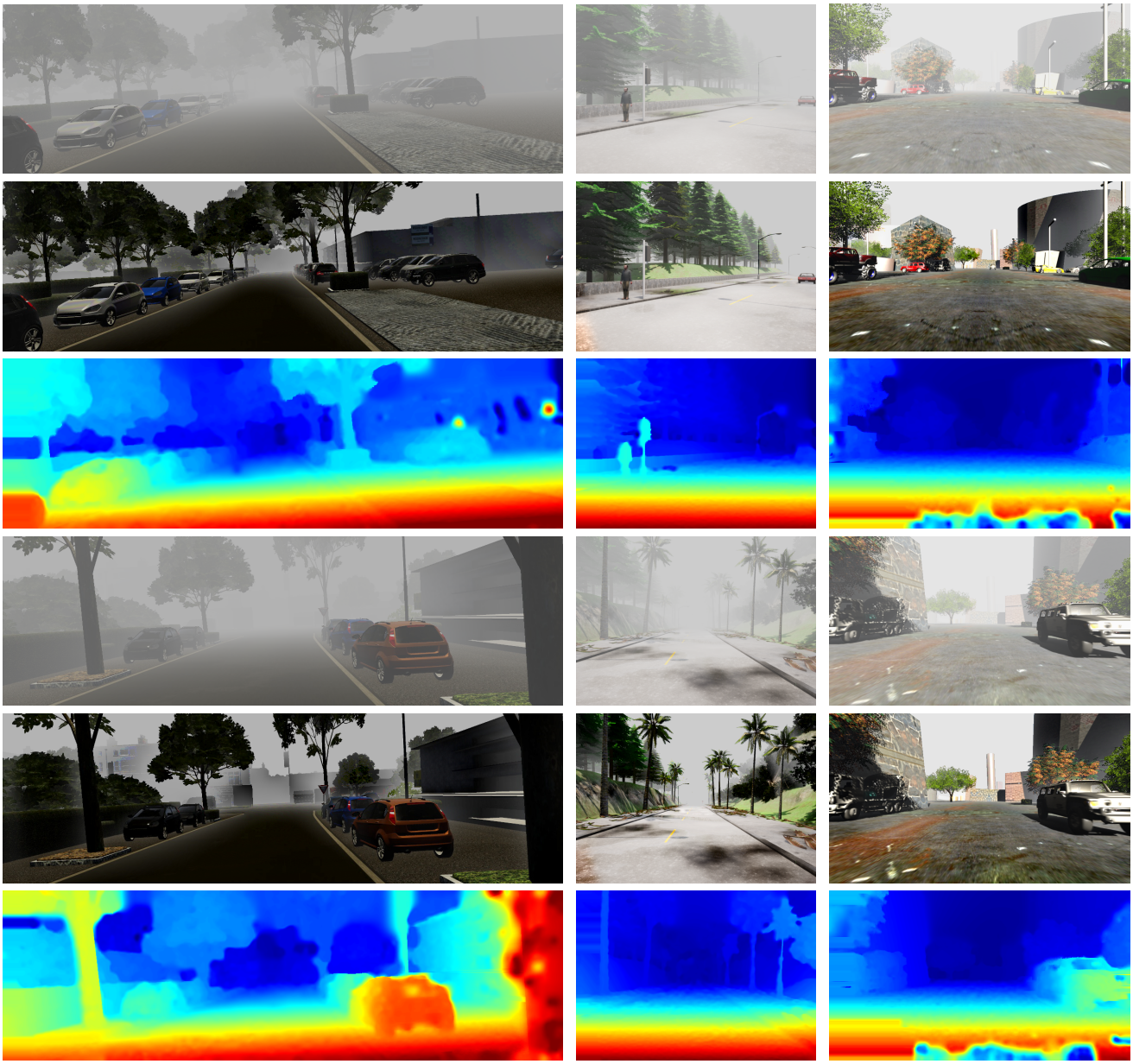

Figure 2. More qualitative results produced by our previous work [7] after feeding the A and β values estimated by the proposed method to it. Columns from left to right: VKITTI2 ($V_{\text{met}} = 40$ m), KITTI-CARLA ($V_{\text{met}} = 60$ m) and DRIVING ($V_{\text{met}} = 80$ m). Rows from top to bottom: the input foggy images, the defogged images and the disparity maps.

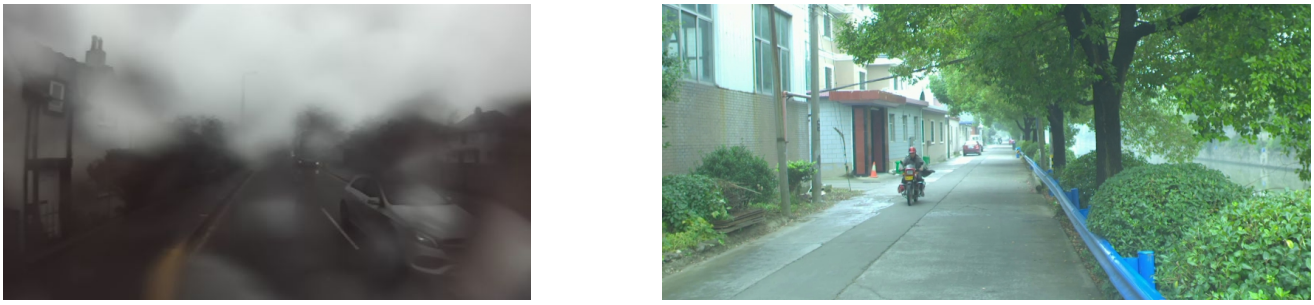

Figure 3. Sample foggy images from RADIATE (left) and DrivingStereo (right)

for autonomous driving: Once dataset. *arXiv preprint arXiv:2106.11037*, 2021. 4

[12] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *CVPR*, 2016. 2

[13] Raul Mur-Artal and Juan D Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE T-RO*, 2017. 1

[14] Srinivasa G. Narasimhan and Shree K. Nayar. Contrast restoration of weather degraded images. *IEEE TPAMI*, 2003. 4

[15] Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, and Andrew Wallace. Radiate: A radar dataset for automotive perception in bad weather. In *ICRA*, 2021. 4

[16] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, 2020. 4

[17] Guorun Yang, Xiao Song, Chaoqin Huang, Zhidong Deng, Jianping Shi, and Bolei Zhou. Drivingstereo: A large-scale dataset for stereo matching in autonomous driving scenarios. In *CVPR*, 2019. 4

[18] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *CVPR*, 2020. 4