Supplementary Material: Variational Simultaneous Stereo Matching and Defogging in Low Visibility

Yining Ding[1]
yd2007@hw.ac.uk
Andrew M. Wallace[1]
a.m.wallace@hw.ac.uk
Sen Wang[2]
sen.wang@imperial.ac.uk

[1] Edinburgh Centre for Robotics
Heriot-Watt University
Edinburgh, UK

[2] Department of Electrical and Electronic Engineering & Imperial-X
Imperial College London
London, UK

In this supplemental material, we present algorithms, parameter setting, implementation, the generation/selection procedure of data on which we perform evaluation, additional experimental results and limitations of our method.

1 The Generalised Condat-Vu Algorithm

The generalised Condat-Vu Algorithm [4, 10] is applied to solve the u minimisation problem (*i.e.* Line 3 in Algorithm 1 in our paper). This is detailed in Algorithm 2, where * denotes the adjoint operator[1], $\Pi_{\{P,Q\}}$ denote point-wise projections onto ℓ_2 balls: $\Pi_P(p) = \frac{p}{\max\{1, \|p\|_2/\lambda_s\}}$ and $\Pi_Q(q) = \frac{q}{\max\{1, \|q\|_2/\lambda_a\}}$, Π_R denotes a point-wise projection onto the set $[-\lambda_t, \lambda_t]^{H\times W\times 2}$, and Π_Z denotes a point-wise projection onto the set $[-\tilde{u}, 1-\tilde{u}]$, which ensures $u \in [0,1]^{H\times W}$. Note that we apply a change of variable such that $z = u - \tilde{u}$ before starting the iterative algorithm and finally recover u after all iterations complete.

2 Disparity Post-processing

In Algorithm 3 we describe the disparity post-processing procedure which is applied to the output of Algorithm 1 in our paper. Note that this post-processing step only affects disparities in the leftmost region.

© 2022. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

[1] The adjoint of a discrete gradient operator with Neumann boundary conditions is a negative divergence operator with Dirichlet boundary conditions: given $\nabla : \mathbb{R}^{H\times W} \to \mathbb{R}^{H\times W\times 2}$ and $\boldsymbol{\nabla} : \mathbb{R}^{H\times W\times 2} \to \mathbb{R}^{H\times W\times 4}$, we have $\nabla^* : \mathbb{R}^{H\times W\times 2} \to \mathbb{R}^{H\times W}$ and $\boldsymbol{\nabla}^* : \mathbb{R}^{H\times W\times 4} \to \mathbb{R}^{H\times W\times 2}$.

Algorithm 2: The generalised Condat-Vu algorithm to solve the u minimisation problem (*i.e.* Line 3 in Algorithm 1 in our paper)

Input: $u^{(k)}$, $a^{(k)}$, w, $s^{(k)}$, $\theta^{(k)}$, $\tilde{u}$
Parameters : ρ, N
Output: $u^{(k+1)}$

1 Initialisation: $u^{(0)} = u^{(k)}$, $v^{(0)} = p^{(0)} = r^{(0)} = 0^{H\times W\times 2}$, $q^{(0)} = 0^{H\times W\times 4}$;
2 Lipschitz constant $L = \max\{w^{(k)}\}/\theta^{(k)}$, dual and primal step sizes $\sigma = \rho/(12+8)$, $\tau = 0.99/(L/2+\rho)$;
3 $z^{(0)} = u^{(0)} - \tilde{u}$;
4 **for** $n = 0, 1, \ldots, N$ **do**
5 $\quad z^{(n+1)} = \Pi_Z\left(z^{(n)} - \tau\left(s^{(k)} + \frac{1}{\theta^{(k)}} w \circ (z^{(n)} + \tilde{u} - a^{(k)}) + \nabla^* \mathsf{G}\left(p^{(n)}\right) + \nabla^* r^{(n)}\right)\right)$;
6 $\quad v^{(n+1)} = v^{(n)} - \tau\left(-\mathsf{G}(p^{(n)}) + \boldsymbol{\nabla}^* q^{(n)}\right)$;
7 $\quad p^{(n+1)} = \Pi_P\left(p^{(n)} + \sigma\left(\nabla \mathsf{G}\left(2z^{(n+1)} - z^{(n)}\right) + \nabla\tilde{u} - \left(2v^{(n+1)} - v^{(n)}\right)\right)\right)$;
8 $\quad q^{(n+1)} = \Pi_Q\left(q^{(n)} + \sigma\boldsymbol{\nabla}\left(2v^{(n+1)} - v^{(n)}\right)\right)$;
9 $\quad r^{(n+1)} = \Pi_R\left(r^{(n)} + \sigma\nabla\left(2z^{(n+1)} - z^{(n)}\right)\right)$;
10 $u^{(k+1)} = z^{(N+1)} + \tilde{u}$;
11 **return** $u^{(k+1)}$

Algorithm 3: Disparity post-processing

Input: Disparities $u_{\text{in}} \in \mathbb{R}^{H\times W}$
Parameters : Maximum disparity $\gamma_{\max}$
Output: Disparities $u_{\text{out}} \in \mathbb{R}^{H\times W}$

1 $u_{\text{out}} \leftarrow u_{\text{in}}$;
2 **for** $h = 1, 2, \ldots, H$ **do**
3 $\quad [u_{\max}, i_{\max}] \leftarrow \text{findMaxValueAndIndex}(u_{\text{out}}[h, 1:\gamma_{\max}+1])$;
4 $\quad$ **if** $i_{max} > 1$ **then**
5 $\quad\quad u_{\text{out}}[h, 1:i_{\max}-1] \leftarrow u_{\max}$;
6 **return** u_{out}

3 Parameter Setting and Implementation

The Census Transform [12] (CT) is applied to each colour channel for the Virtual KITTI 2 dataset [2] (VKITTI2) and to grayscale images for the Pixel-Accurate Depth dataset [5] (PAD) because the images in the latter are much noisier. Γ is chosen to be $\{0, 1, \cdots, 127\}$ (*i.e.* $\gamma_{\max} = 127$) for VKITTI2 and $\{0, 1, \cdots, 63\}$ (*i.e.* $\gamma_{\max} = 63$) for PAD.

The following parameters are chosen empirically and *fixed* in all of our experiments. We use a window size of 7 in CT. In the anisotropic diffusion operator calculation: tensor magnitude parameter is 9 and sharpness parameter is 0.85. In the adaptive support-weight [11] calculation: the support window size is 15, the similarity parameter is 5 and the proximity parameter is 7.5. In Algorithm 1 in our paper: $\lambda_d = 0.4$, $\lambda_s = 1$, $\lambda_a = 8$, $\lambda_t = 0.01$, $\alpha = 0.001$ and $K = 80$. In Algorithm 2: $\rho = 10$ and $N = 150$. In the photo-inconsistency check: $\varepsilon = 5$. In the transmission refinement: $\mu = 0.003$.

The algorithm is implemented (unoptimised) in MATLAB R2022a on a laptop with Intel® Core™ i7-10750H CPU @2.60GHz and NVIDIA® GeForce® RTX 2070 GPU with Max-Q Design. MATLAB's GPU computing is used to parallelise array operations. The computational time is around 92 seconds for foggy stereo matching and around 1.5 seconds for defogging, given an input image size of 494×1032.

4 Data for Evaluation

We describe how we generate/select data from VKITTI2 and PAD on which we perform evaluation.

4.1 Virtual KITTI 2

VKITTI2 contains five sequences of *synthetic* road scenes. We first select clear images (only under the *overcast* weather condition to better resemble real foggy situations, as [7] does) every ten frames in all five sequences. We then synthesise foggy images according to the atmospheric scattering model using the dense ground truth depth maps. *vis* and A are randomly drawn from the uniform distribution in the intervals $(20,40)$ and $(0.7,1)$ respectively. The above procedure generates 214 foggy image pairs in total, on which we perform evaluation.

4.2 Pixel-Accurate Depth Dataset

PAD contains four *real* static scenes recorded in a weather chamber under various fog visibility conditions. Dense ground truth depth of each scene is available because a Leica scan station was used and multiple measurements were accumulated. Note that this is a very challenging dataset because the fog appears to be highly inhomogeneous and the foggy images are very noisy (possibly due to high ISO values). We choose daytime foggy images of the first three scenes (the fourth scene is not used because the right frames all seem severely blurred) from the following five visibilities in meters $\{20,25,30,35,40\}$, hence 150 foggy image pairs in total (ten samples per scene per visibility) for evaluation. We crop each original image such that only the central scene is reserved and its surrounding pixels are discarded.

5 Additional Experiments

We report results of some additional experiments.

5.1 FRIDA3

In addition to the results provided in our paper, we would also like to compare our method with methods such as [8, 9]. However, to our best knowledge, the source code of neither is published. Moreover, their evaluation was performed on different datasets from ours. In order to facilitate a direct comparison with [9], we evaluate our method (using the same experimental setup as VKITTI2, see Sec. 3) on FRIDA3 [3]. Results are shown in Tab. 3.

Task	Metric	[10]	**Ours**
Stereo matching	3PE-all	11.810	**9.330**
	EPE	4.170	**1.308**
Defogging	PSNR	**16.590**	15.438
	SSIM	0.839	**0.848**

Table 3: Results on FRIDA3

5.2 Sensitivity to $\tilde{t}$

All experiments in our paper use [6] to estimate the initial transmission map $\tilde{t}$. In Tab. 4 we show results of our method on VKITTI2 if [11] is used to obtain $\tilde{t}$. We observe that the new results are close to the original ones shown in Tab. 1 in our paper, demonstrating our method's robustness to $\tilde{t}$ from different methods.

Method	SSIM	PSNR
Ours ($\tilde{t}$ from [1])	0.635	10.564

(a) Defogging

Method	D1-all	RMSE	MAE	SILog	Sq Rel	Abs Rel	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Ours ($\tilde{t}$ from [1])	8.849	26.150	7.309	25.599	4.212	8.790	86.916	92.392	95.501

(b) Stereo matching

Table 4: Results on VKITTI2 if [1] is used to estimate $\tilde{t}$

6 Limitations

Although our method demonstrates strong performances, it also has limitations. In terms of defogging, like most existing methods, we assume the scattering coefficient β to be a global constant, thereby failing to deal with inhomogeneous fog very well and leaving visible residual of fog in the defogged image in such case (see the second row last column of Fig. 6 in our paper). As for stereo matching, because of the introduction of the regularisation term $\lambda_t \|\nabla(u-\tilde{u})\|_{1,1}$, false discontinuities in the estimated transmission map $\tilde{t}$ can cause false disparity (hence depth) discontinuities in our method (see the zebra crossing area in the first row last column of Fig. 6 in our paper).

References

[1] Dana Berman, Tali treibitz, and Shai Avidan. Non-local image dehazing. In *CVPR*, 2016.

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

[3] Laurent Caraffa and Jean-Philippe Tarel. Stereo reconstruction and contrast restoration in daytime fog. In *ACCV*, 2012.

[4] Laurent Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 2013.

[5] Tobias Gruber, Mario Bijelic, Felix Heide, Werner Ritter, and Klaus Dietmayer. Pixel-accurate depth evaluation in realistic driving scenarios. In *3DV*, 2019.

[6] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *PAMI*, 2010.

[7] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *IJCV*, 2018.

[8] Taeyong Song, Youngjung Kim, Changjae Oh, and Kwanghoon Sohn. Deep network for simultaneous stereo matching and dehazing. In *BMVC*, 2018.

[9] Taeyong Song, Youngjung Kim, Changjae Oh, Hyunsung Jang, Namkoo Ha, and Kwanghoon Sohn. Simultaneous deep stereo matching and dehazing with feature attention. *IJCV*, 2020.

[10] Bằng Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 2013.

[11] Kuk-Jin Yoon and In So Kweon. Adaptive support-weight approach for correspondence search. *PAMI*, 2006.

[12] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *ECCV*, 1994.