FoggyStereo: Stereo Matching with Fog Volume Representation

Chengtang Yao[1,2], Lidong Yu[2]

[1]Beijing Laboratory of Intelligent Information Technology, Beijing Institute of Technology,

[2]Autonomous Driving Algorithm, NIO

Paper & Code

CVPR 2022, JUNE 19-24, NEW ORLEANS • LOUISIANA

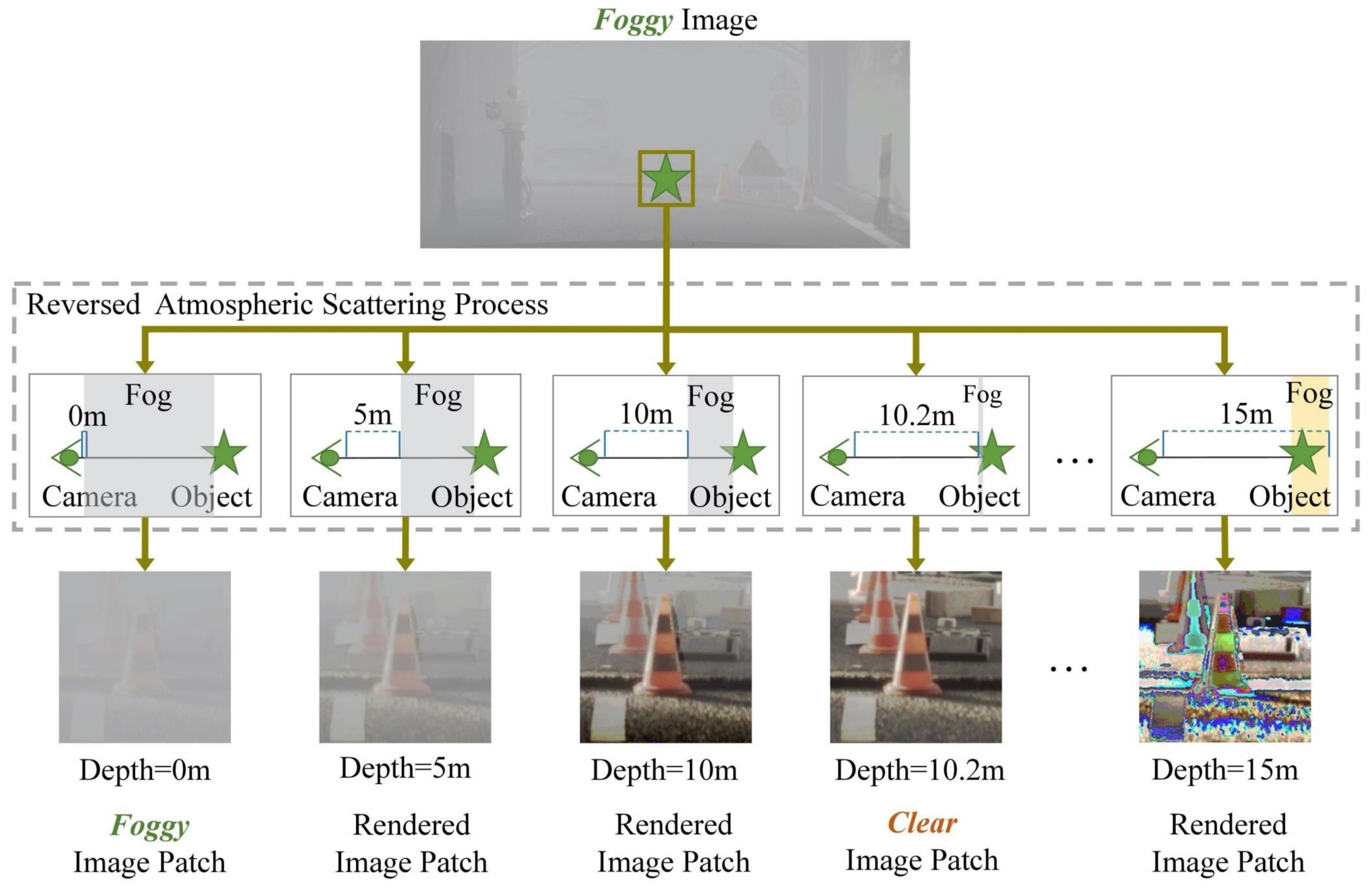
Introduction

Problem Stereo matching in foggy scenes is challenging as the scattering effect of fog blurs the image and makes the matching ambiguous

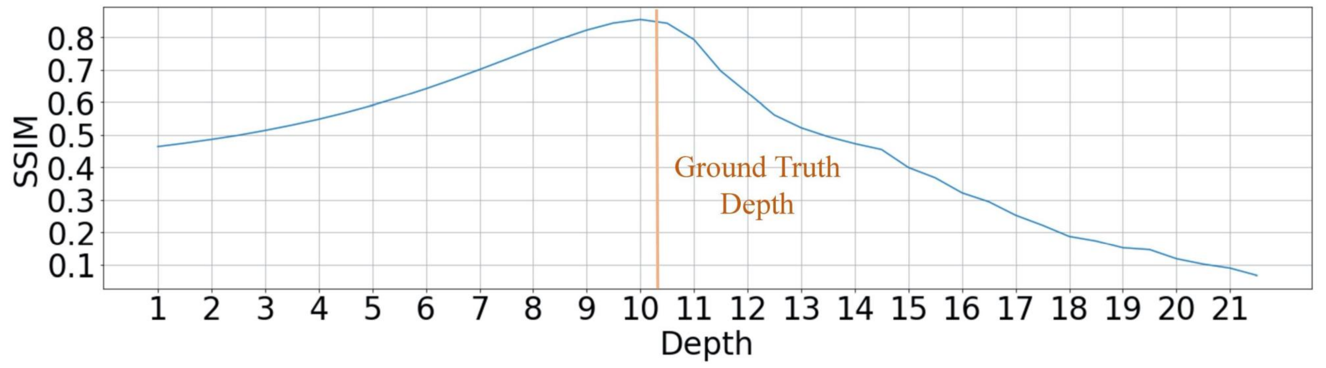
Idea Prior methods mainly deem fog as a noise and discard it to improve matching results. Different from them, we propose to take advantage of fog and explore depth hints for stereo matching.

Motivation During rendering, fog is accumulated along the light path between objects and camera following the *physical atmospheric scattering process*.

When we render the image by reversing this process, fog is removed within a selected depth range. Only the depth close to the real depth will lead to a clear image. In other words, *the quality of the rendered image indicates the correctness of depth used in the rendering*.



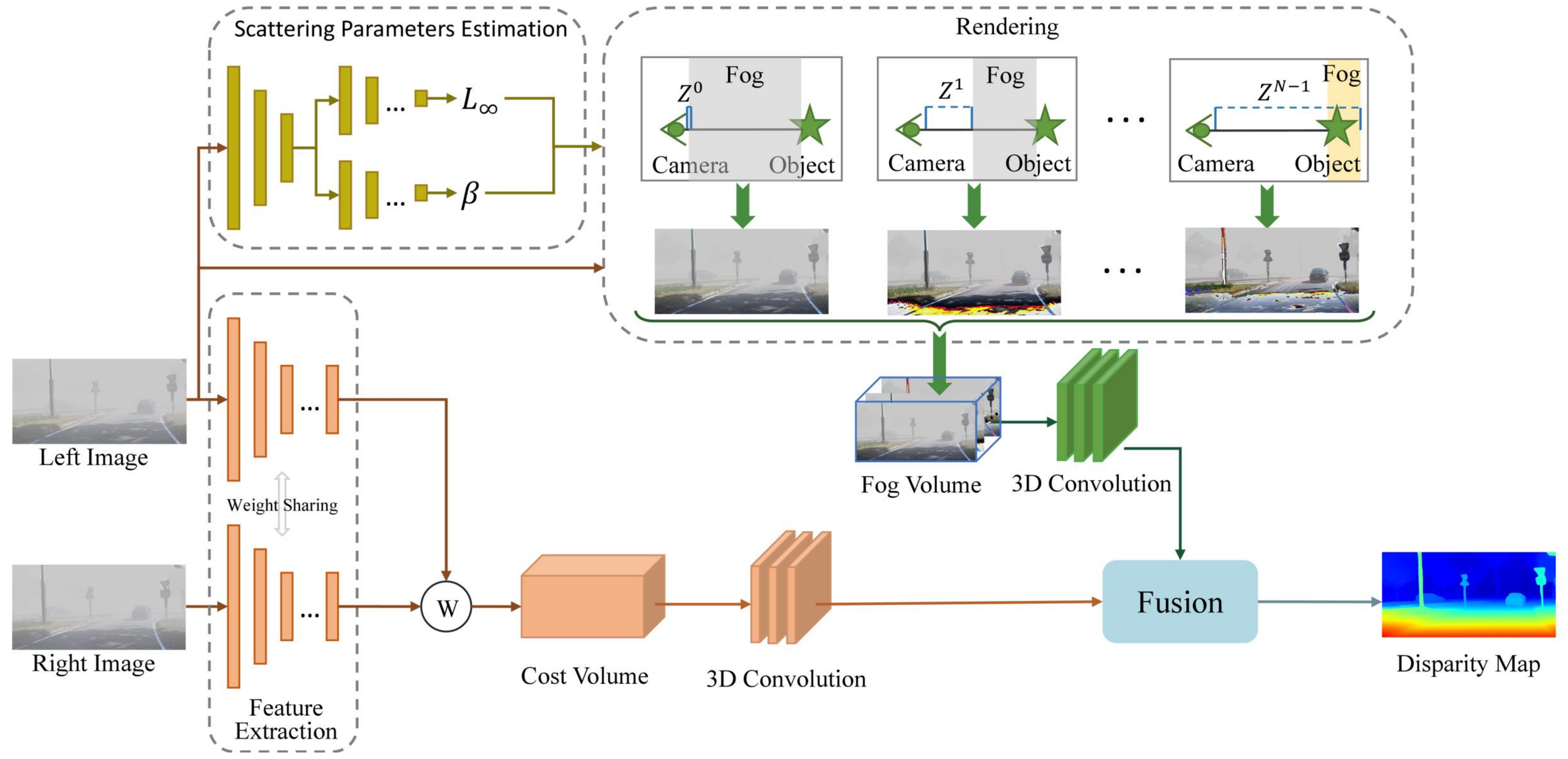
(a) The process and results of rendering.



(b) The distribution of SSIM ~ Depth.

Method

We extract features from left and right images to build a cost volume through warping ⓦ. We estimate atmospheric light L_∞ and attenuation coefficient β from the left image to render a series of images with different depth Z_i. The rendered images are concatenated to build fog volume, which is fused with cost volume for disparity estimation.



Fog Volume Representation

(1) Rendering The atmospheric scattering effect causes the attenuation of light reflected from objects L_t and the accumulation of environmental light L_c:

$$\left.\begin{aligned} T(Z_x) &= e^{-\int_0^{Z_x}\beta(z)dz} \\ L_t(x) &= L_\infty \rho(x) T(Z_x) \\ L_c(x) &= L_\infty(1 - T(Z_x)) \end{aligned}\right\} \Rightarrow \begin{aligned} I(x) &= L_t(x) + L_c(x) \\ &= J(x)T(Z_x) + L_\infty(1 - T(Z_x)). \end{aligned}$$

The rendered image R is computed by reversing the atmospheric scattering:

$$R(x, Z_x^i) = \left(I(x) - L_\infty\left(1 - T(Z_x^i)\right)\right) / T(Z_x^i),$$

$$R(x, Z_x^i) = \ln(|I(x) - L_\infty|) + \int_0^{Z_x^i} \beta(z)dz.$$

(2) Scattering Parameters Estimation We set L_∞ and β as global parameters under the condition of one single light source and a homogeneous transporting medium.

(3) Disparity Candidates Sampling We sample disparity candidates $\{D_x^i\}_{i=0}^{i=N-1}$ to construct cost volume, and convert them into depth $\{Z_x^i\}_{i=0}^{i=N-1}$ to build the fog volume.

(4) Rendered Images Gathering We build the fog volume representation $\mathcal{V}_f$ by stacking rendered images: $\mathcal{V}_f(x, Z) = [R(x, Z_x^0), R(x, Z_x^1), \cdots, R(x, Z_x^{N-1})]$.

Fusion

We fuse the cost volume and the fog volume with uncertainty σ which is the variance along disparity dimension: $\tilde{\mathcal{V}}(x, D_{\{i\}}) = [\sigma_c(x, D_i)\mathcal{V}_c(x, D_i), \sigma_f(x, D_i)\mathcal{V}_f(x, D_i)]$.

Results

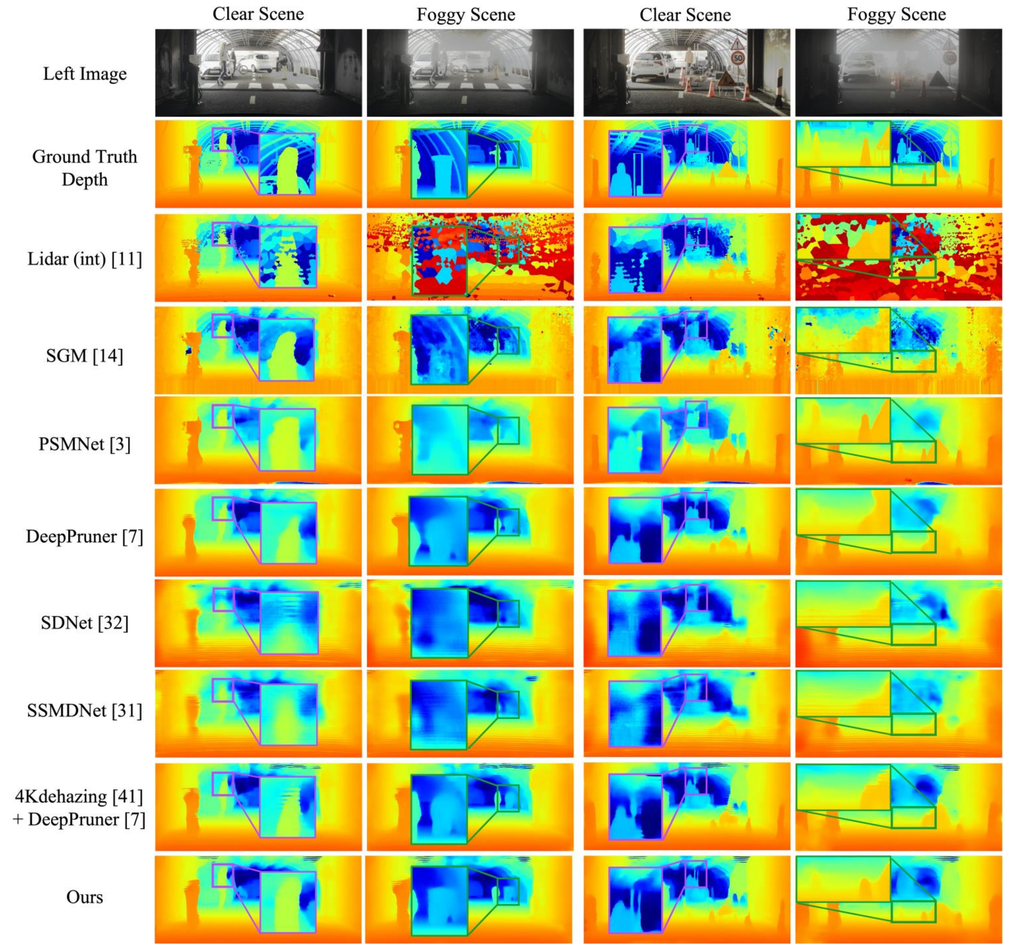
The comparison of algorithms on SceneFlow dataset. We compare the results on clear data and foggy data. * represents our re-implementation results.

Testing	Metrics	Stereo: PSMNet* [3]	Stereo: DeepPruner* [7]	Joint: SDNet [32]	Joint: SSMDNet [31]	Sequential: 4Kdehazing [41] + DeepPruner [7]	Ours
Clear	EPE	0.99	0.98	-	-	1.19	**0.81**
	3px (%)	**4.1**	5.30	-	-	6.2	**4.5**
Foggy	EPE	1.27	3.77	2.68	2.23	1.49	**1.04**
	3px (%)	8.1	14.10	26.43	9.71	10.30	**7.2**

The comparison of algorithms on KITTI 2015 and 2012 datasets.

Methods		KITTI 2015 Foggy 3px (%)	KITTI 2015 Foggy EPE	KITTI 2015 Clear 3px (%)	KITTI 2015 Clear EPE	KITTI 2012 Foggy 3px (%)	KITTI 2012 Foggy EPE	KITTI 2012 Clear 3px (%)	KITTI 2012 Clear EPE
Stereo	PSMNet* [3]	1.3	0.54	**1.0**	0.49	3.3	0.84	3.3	0.86
	DeepPruner* [7]	3.7	0.88	8.8	1.66	4.3	0.94	5.0	1.09
Joint	SDNet [32]	13.4	1.73	-	-	11.0*	1.63*	10.7*	1.60*
	SSMDNet [31]	10.8	1.23	-	-	9.7*	1.55*	9.5*	1.53*
Sequential	4Kdehazing [41] + DeepPruner [7]	7.3	0.951	**1.1**	0.49	3.2	0.91	3.2	0.89
ours		**1.2**	**0.51**	**1.1**	**0.47**	**2.7**	**0.77**	**2.7**	**0.78**

The visualization of depth map on PixelAccurateDeth dataset with real foggy scenes.



The visualization of error rate (EPE) distribution over the depth and the fog thickness.