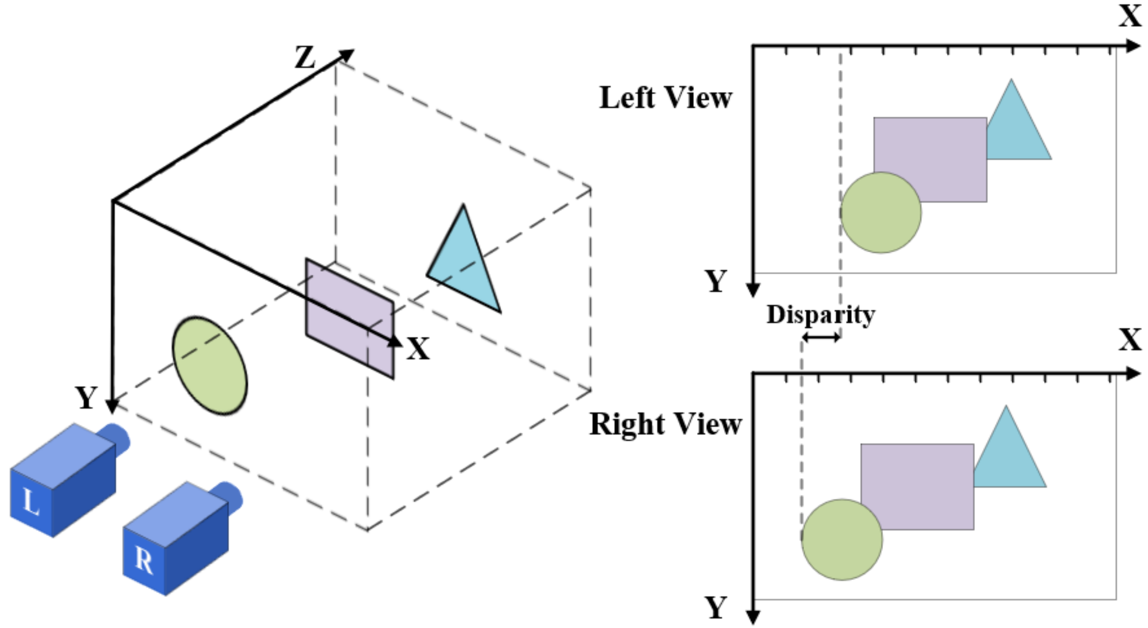
FoggyStereo: Stereo Matching with Fog Volume Representation

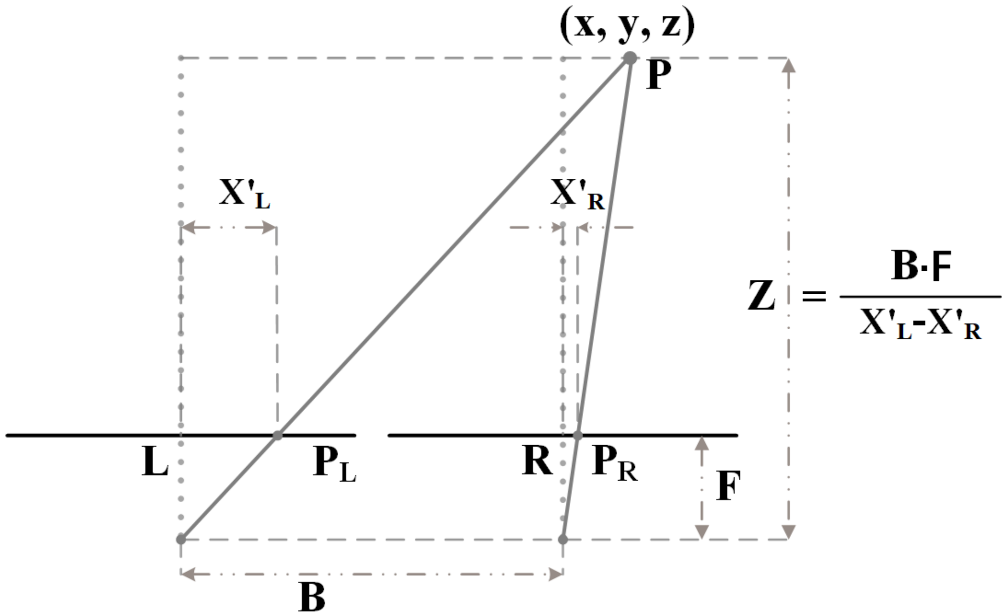
Chengtang Yao[1,2], Lidong Yu[2]

[1]Beijing Laboratory of Intelligent Information Technology,
School of Computer Science, Beijing Institute of Technology
[2]Autonomous Driving Algorithm, NIO

Stereo Matching



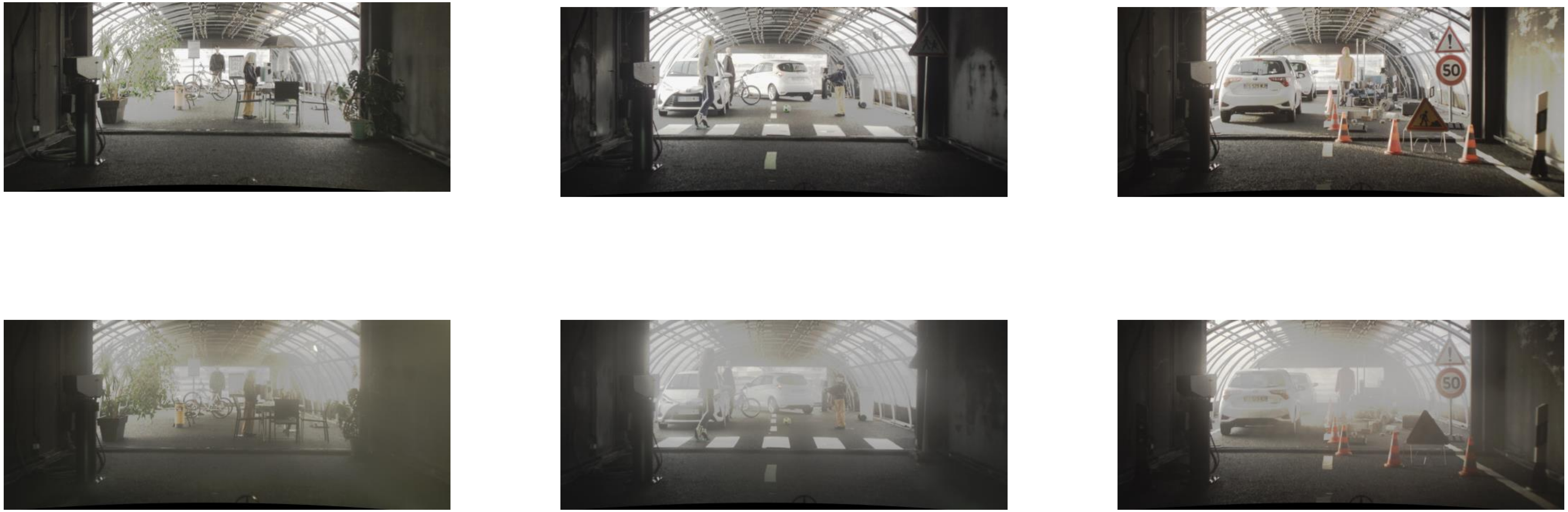
Disparity Estimation



Depth Computation

Problem

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

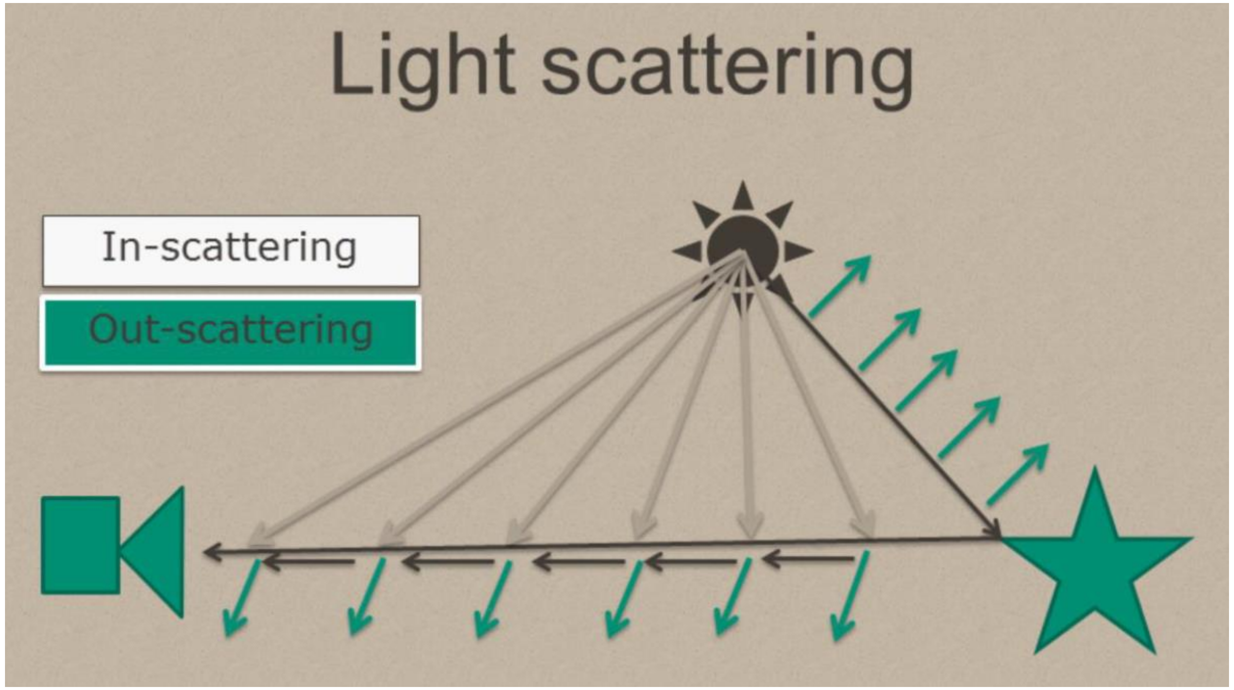


Pictures come from [Gruber et. al. 3DV 2019]

Motivation

Fog is accumulated along the light path between objects and camera following the *physical atmospheric scattering process*.

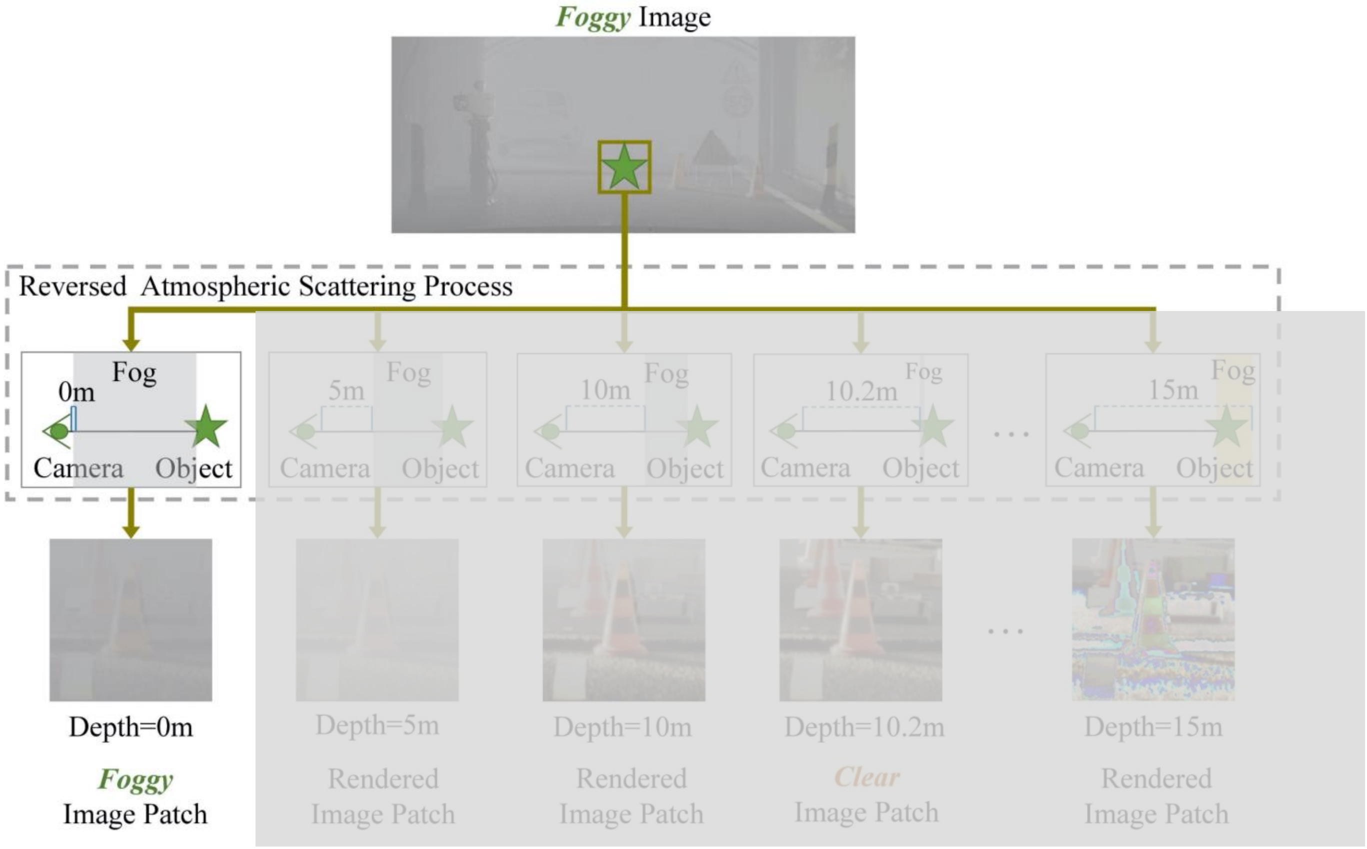
Different depths will lead to different brightness and blur the image at different levels.



Picture comes from [Bartlomiej SIGGRAPH 2014]

Motivation

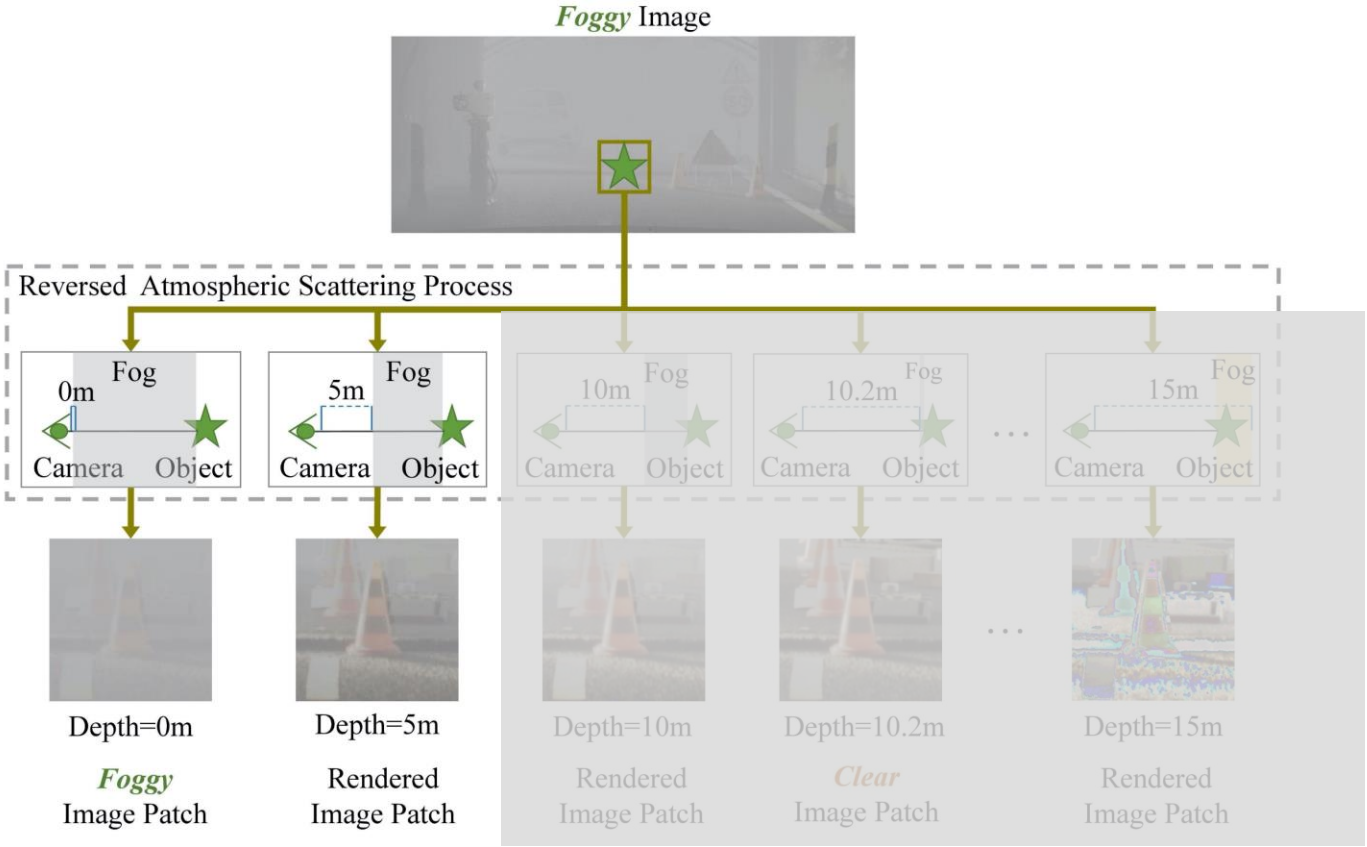
When we render the image by reversing the process, fog is removed within a selected depth range. Only the depth close to the real depth will lead to a clear image.



(a) The process and results of rendering.

Motivation

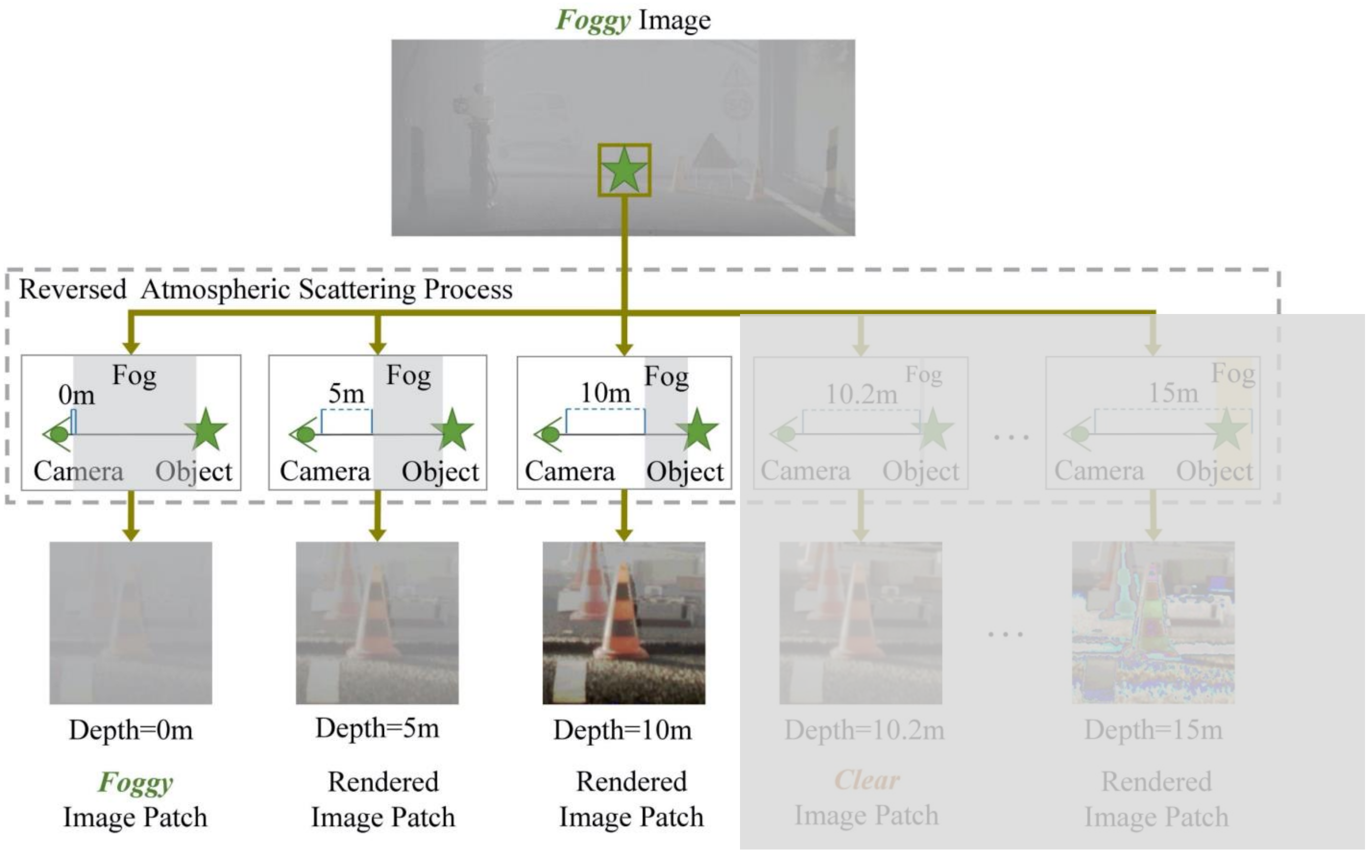
When we render the image by reversing the process, fog is removed within a selected depth range. Only the depth close to the real depth will lead to a clear image.



(a) The process and results of rendering.

Motivation

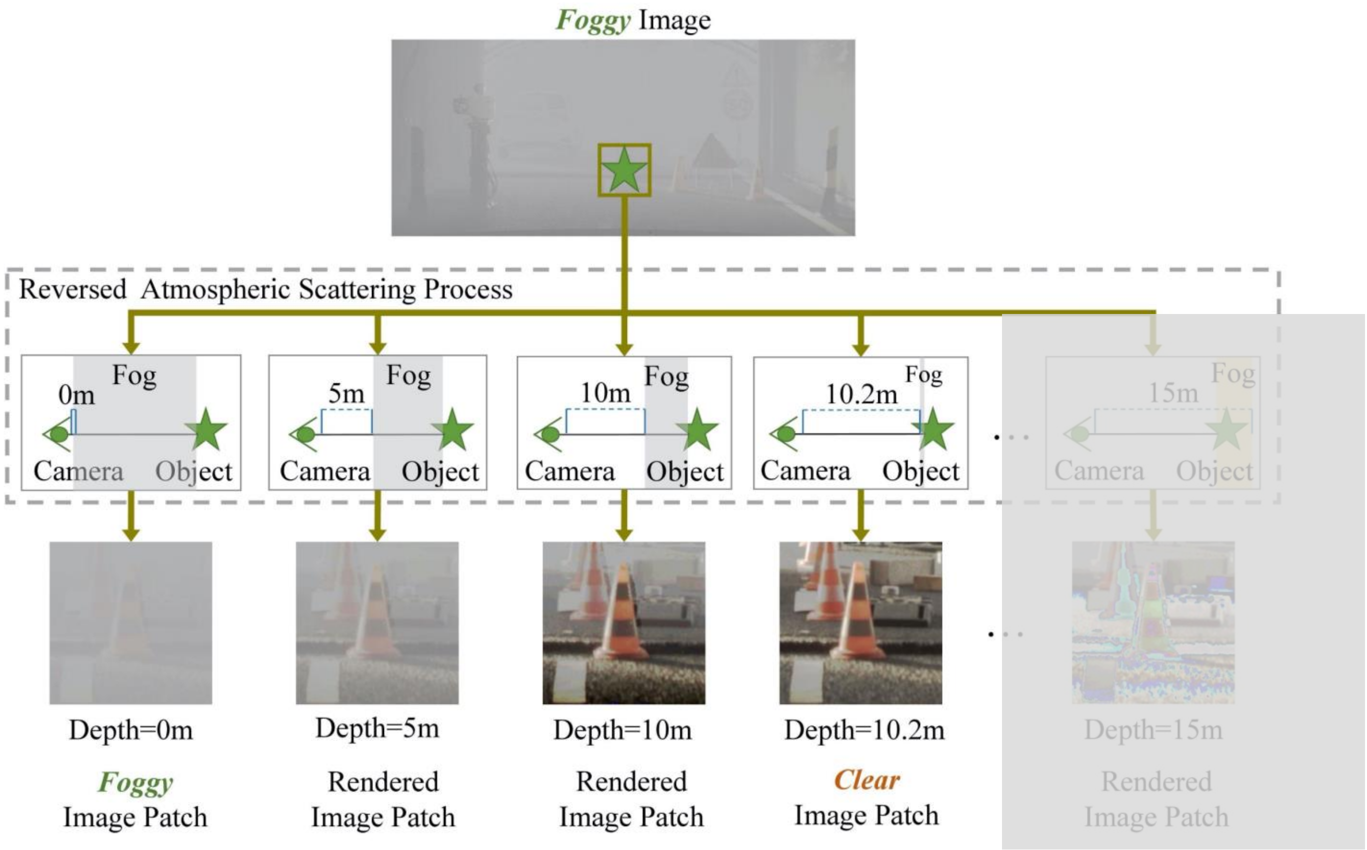
When we render the image by reversing the process, fog is removed within a selected depth range. Only the depth close to the real depth will lead to a clear image.



(a) The process and results of rendering.

Motivation

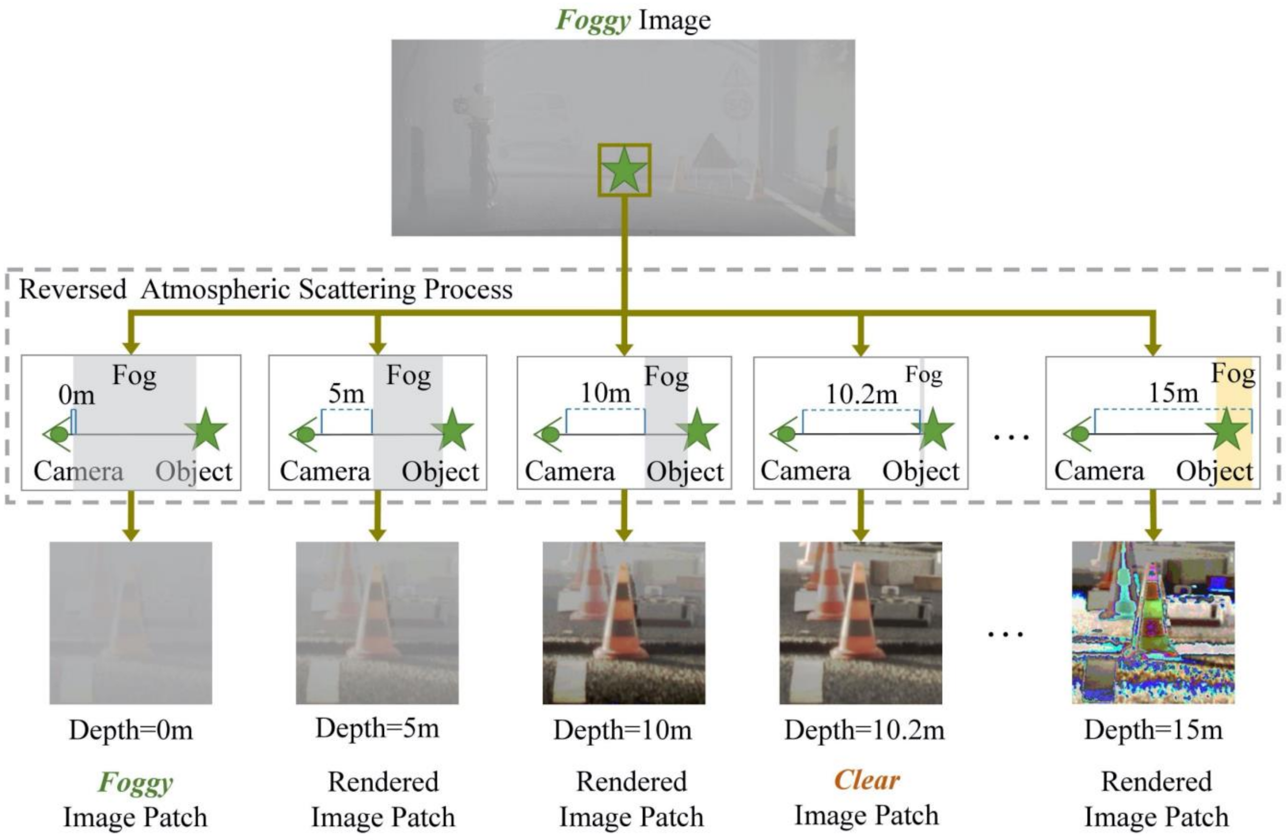
When we render the image by reversing the process, fog is removed within a selected depth range. Only the depth close to the real depth will lead to a clear image.



(a) The process and results of rendering.

Motivation

When we render the image by reversing the process, fog is removed within a selected depth range. Only the depth close to the real depth will lead to a clear image.

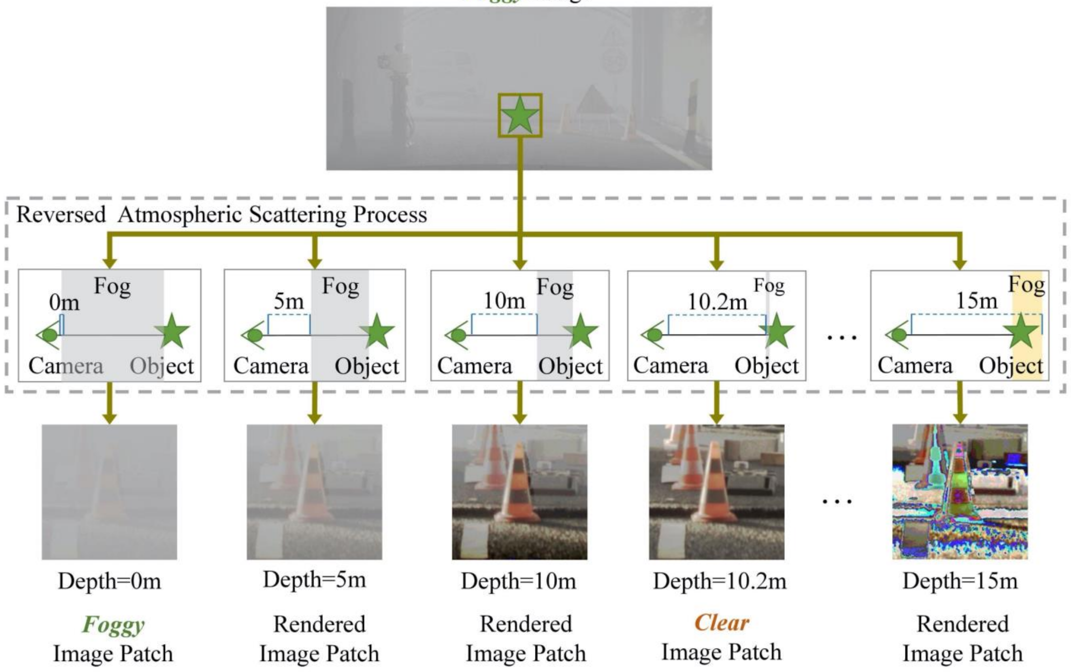


(a) The process and results of rendering.

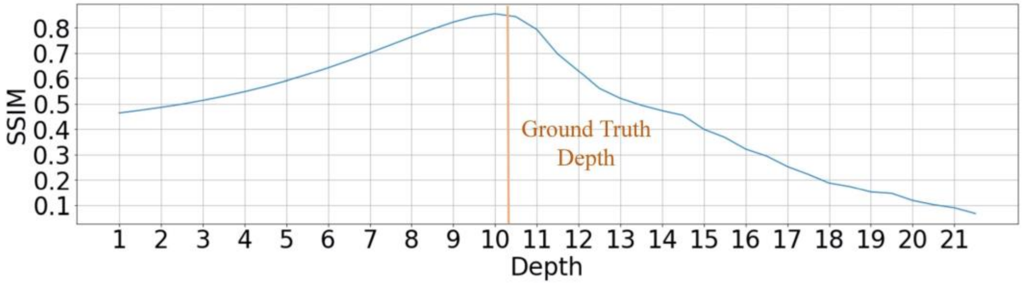
Motivation

When we render the image by reversing the 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.



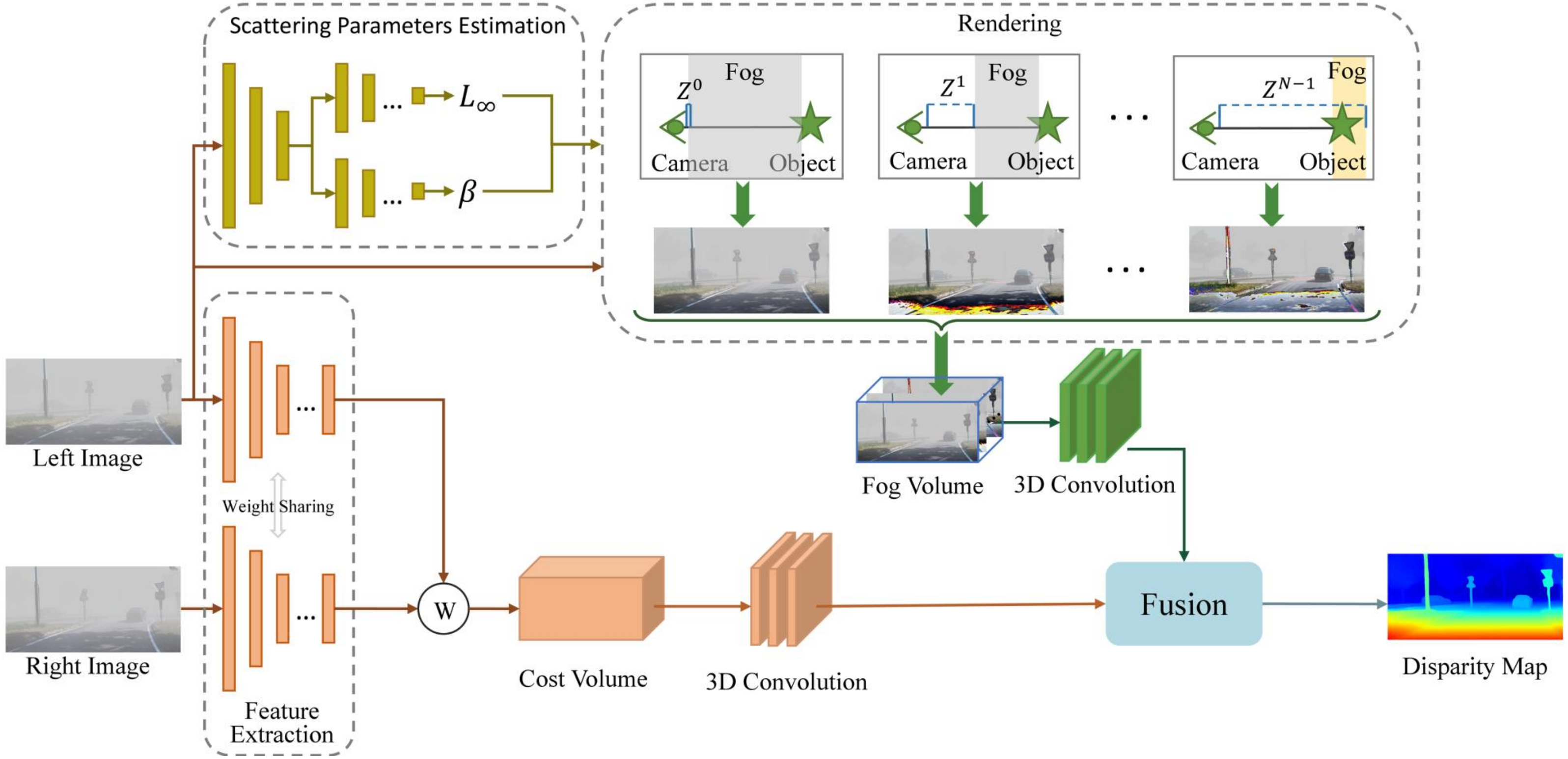
(a) The process and results of rendering.



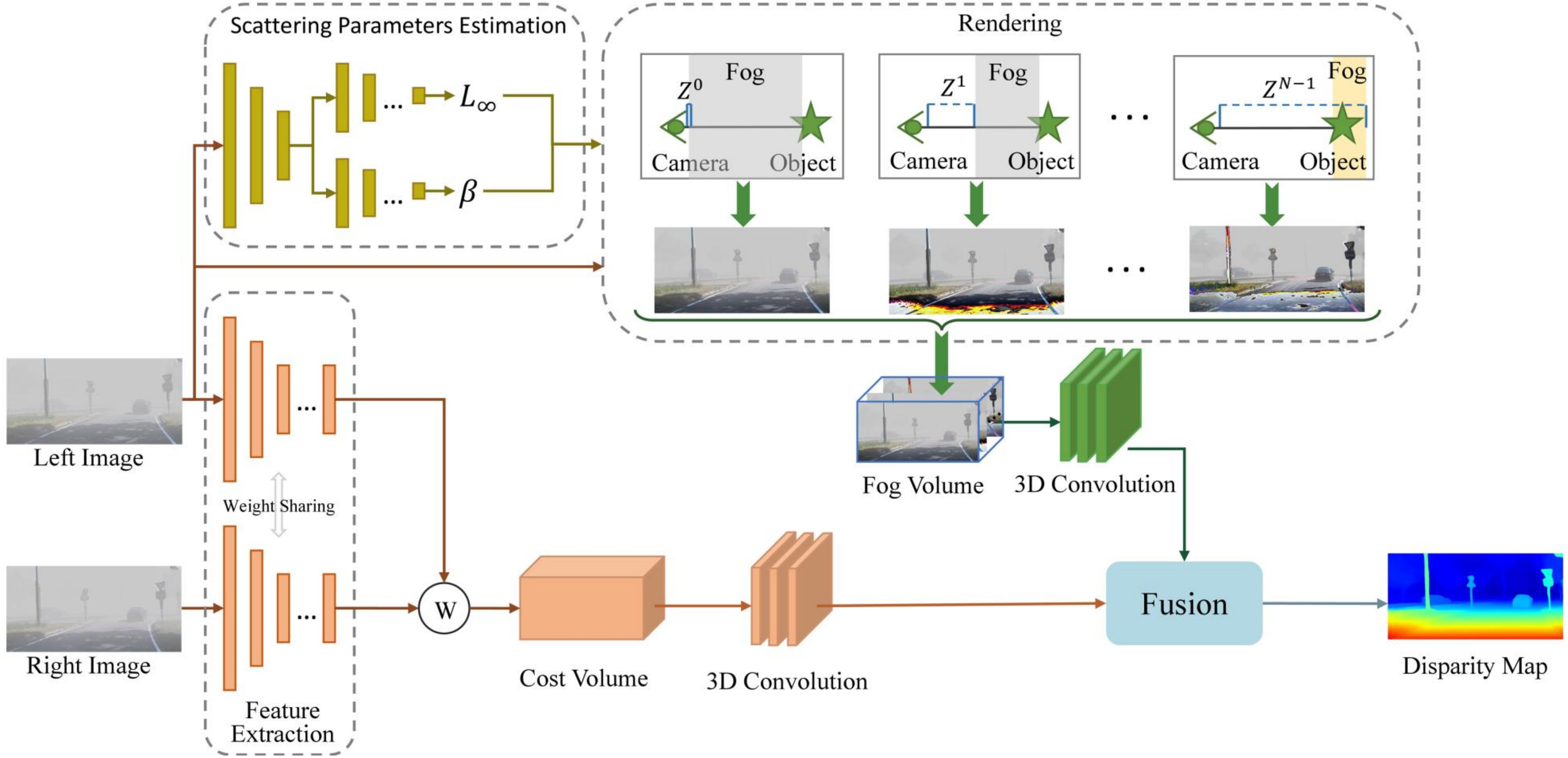
(b) The distribution of SSIM ~ Depth.

Contributions

- We introduce a fog volume representation to collect depth hints from the fog.
- We propose to fuse the cost volume and the fog volume to adapt to both foggy areas and good visible areas.



Model


Scattering Parameters Estimation
L_∞
β
Rendering
Z^0
Fog
Camera
Object
Z^1
Fog
Camera
Object
Z^{N-1}
Fog
Camera
Object
Left Image
Weight Sharing
Right Image
Feature Extraction
W
Cost Volume
3D Convolution
Fog Volume
3D Convolution
Fusion
Disparity Map

Model

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:

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

L_∞ is the atmospheric light, ρ is the reflectance of pixel x on the object surface. The attenuation T is commonly measured by the Beer-Lambert-Bouguer law, where β is the attenuation coefficient.

Model

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:

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

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

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

Model

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:

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

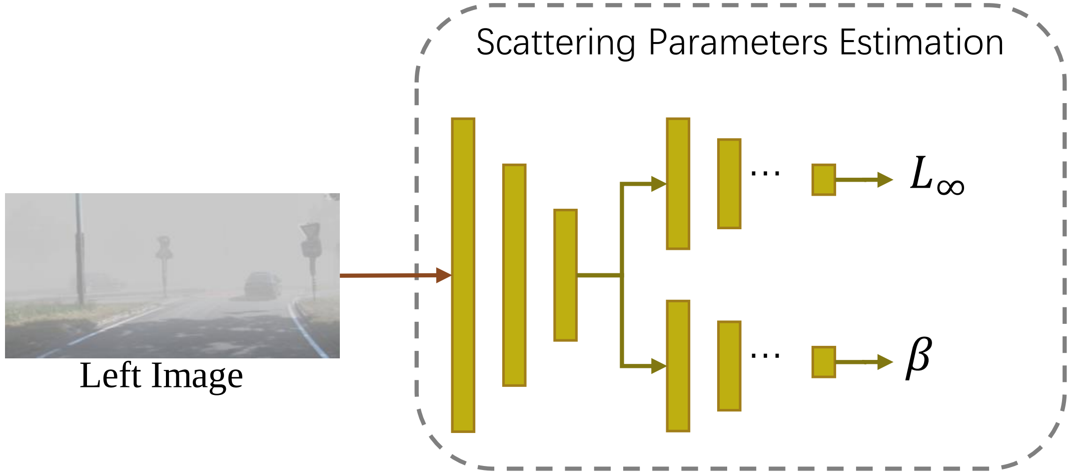
The rendered image R is computed by reversing the atmospheric scattering:

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

Model

Fog Volume Representation

(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. We learn the global parameters from the left image by a fully convolutional network.



Model

Fog Volume Representation

(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 fog volume.

(4) Rendered Images Gathering We build the fog volume representation $\mathcal{V}_f$ by stacking rendered images:

$$\mathcal{V}_f(x, Z) = [R(x, Z_x^0), R(x, Z_x^1), \cdots, R(x, Z_x^{N-1})].$$

Model

Fusion

Cost volume works well in clear areas.

Fog volume works well in foggy areas.

Fusion for adaptation to both clear areas and foggy areas.

We concatenate the cost volume and the fog volume with uncertainty σ which is the variance along disparity dimension:

$$\tilde{\mathcal{V}}(x, D_{\{i\}}) = \left[\sigma_c(x, D_i)\mathcal{V}_c(x, D_i),\ \sigma_f(x, D_i)\mathcal{V}_f(x, D_i)\right].$$

The concatenated volume is subsequently input into a 3D convolution network to jointly leverage the beneficial information from both volumes.

Model

Loss Function

- The predicted disparity map $\widetilde{D}$ is supervised by the ground truth disparity map D using L_1 loss:

$$\mathcal{L}_1 = L_1\left(D, \widetilde{D}\right).$$

- The estimated $\tilde{L}_\infty$ and $\tilde{\beta}$ are supervised by the reconstruction loss of clear image J in RGB space and gray space:

$$\mathcal{L}_2 = L_1\left(\tilde{R}, J'\right) + L_1\left(\tilde{R}_{gray}, {J'}_{gray}\right),$$

$$J' = \ln(|J - L_\infty|).$$

- We supervise the learning of $\tilde{L}_\infty$with the average intensity $\bar{L}_\infty$ of pixels whose disparity is smaller than 1.5 as $L_\infty \approx I_x$ when Z_x is large:

$$\mathcal{L}_3 = L_1\left(\bar{L}_\infty, \tilde{L}_\infty\right).$$

Results

- SceneFlow dataset. * represents our re-implementation results.

Testing	Metrics	Stereo		Joint		Sequential	Ours
		PSMNet* [3]	DeepPruner* [7]	SDNet [32]	SSMDNet [31]	4Kdehazing [41] + DeepPruner [7]	
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**

- KITTI 2015 and 2012 datasets.

Methods		KITTI 2015				KITTI 2012			
		Foggy		Clear		Foggy		Clear	
		3px (%)	EPE	3px (%)	EPE	3px (%)	EPE	3px (%)	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**

Results

- The clear data of PixelAccurateDepth dataset.

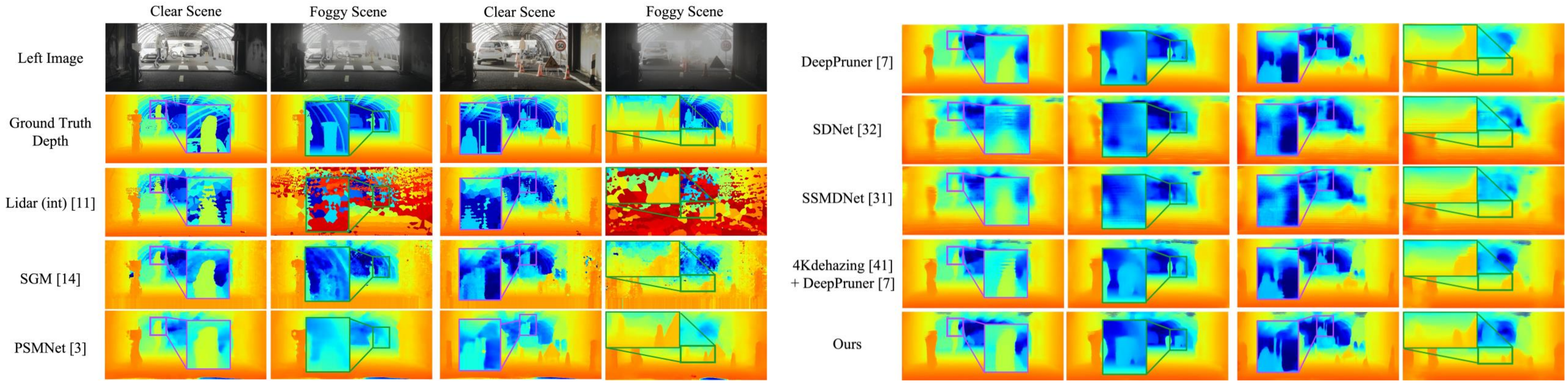
Method		RMSE ↓	tRMSE ↓	MAE ↓	tMAE ↓	logRMSE ↓	SRD ↓	ARD ↓	SIlog ↓	δ_1 (%) ↑	δ_2 (%) ↑	δ_3(%) ↑
Stereo	SGM [14]	1.90	1.40	0.96	0.86	0.14	0.27	8.12	13.32	90.74	98.44	99.50
	PSMNet [3]	2.75	1.96	1.44	1.22	0.18	0.56	9.91	16.07	89.14	97.21	98.80
	DeepPruner* [7]	1.81	1.37	0.80	0.70	0.12	0.21	5.52	11.78	93.57	98.08	99.50
Joint	SDNet* [32]	1.89	1.53	1.03	0.94	0.13	0.26	7.94	12.87	92.52	98.22	**99.57**
	SSMDNet* [31]	1.95	1.53	1.00	0.90	0.12	0.22	7.05	12.17	92.75	**98.53**	**99.68**
Sequential	4Kdehazing [41] + DeepPruner [7]	**1.79**	1.32	0.77	0.67	**0.11**	**0.19**	5.12	**10.95**	**94.41**	**98.45**	**99.66**
Lidar (int.) [11]		1.89	1.36	**0.70**	**0.59**	0.13	0.23	**4.78**	12.58	93.62	98.13	99.36
RGB+Lidar [11]		3.05	2.04	1.61	1.29	0.26	0.53	10.85	24.01	84.69	94.77	97.05
Ours		**1.82**	**1.31**	**0.75**	**0.64**	**0.11**	**0.20**	**5.01**	**11.11**	**94.07**	**98.45**	**99.56**

- The foggy data of PixelAccurateDepth dataset.

Method		RMSE ↓	tRMSE ↓	MAE ↓	tMAE ↓	logRMSE ↓	SRD ↓	ARD ↓	SIlog ↓	δ_1 (%) ↑	δ_2 (%) ↑	δ_3(%) ↑
Stereo	SGM [14]	3.00	1.81	1.56	1.20	0.21	1.00	14.02	20.75	84.34	94.91	97.22
	PSMNet [3]	3.01	2.10	1.65	1.35	0.19	0.61	11.10	16.94	84.95	96.34	98.65
	DeepPruner* [7]	2.61	1.75	1.30	1.00	0.16	0.40	8.10	15.16	87.24	95.61	**98.92**
Joint	SDNet* [32]	2.63	1.88	1.48	1.22	0.18	0.47	10.67	16.86	85.83	95.70	98.50
	SSMDNet* [31]	2.69	1.83	1.42	1.13	0.17	0.42	9.23	16.12	87.42	96.13	98.54
Sequential	4Kdehazing [41] + DeepPruner [7]	3.32	1.81	1.69	1.06	0.23	0.76	9.91	20.71	85.08	92.13	95.01
Lidar (int.) [11]		3.67	2.01	1.68	1.13	0.39	0.91	12.21	35.19	80.57	87.27	91.66
RGB+Lidar [11]		3.81	2.52	2.34	1.83	0.35	0.91	16.88	28.67	69.77	85.16	92.74
Ours		**2.55**	**1.64**	**1.19**	**0.91**	**0.15**	**0.38**	**7.38**	**14.77**	**89.28**	**96.33**	**98.66**
Ours (PixelAccurateDepth Clear)		**1.74**	**1.20**	**0.80**	**0.61**	**0.10**	**0.22**	**4.50**	**9.04**	**93.14**	**97.42**	**99.72**

Results

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

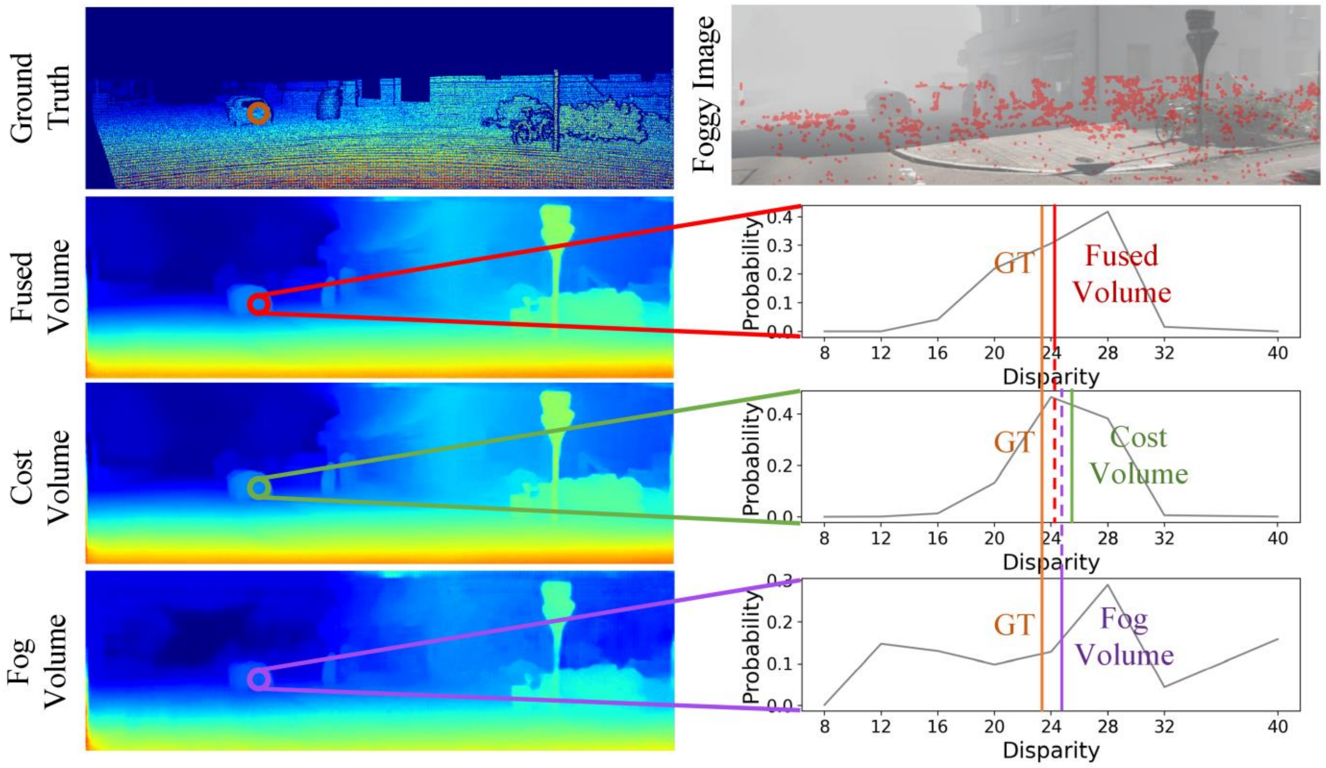


Results

- Influence of fusion.

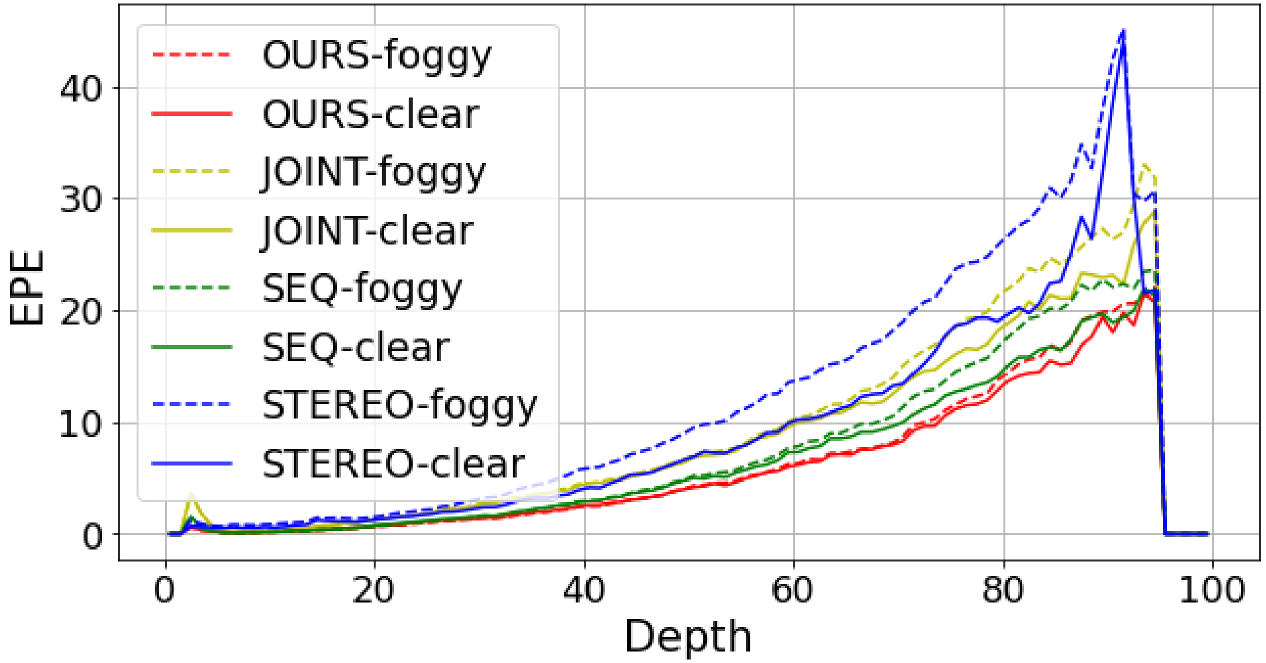
In the foggy image, red points illustrate areas where the results of fused volume are the best while the results of fog volume are better than that of cost volume.

In probability distribution, the ground truth and final predictions are illustrated through the vertical line in a different color.



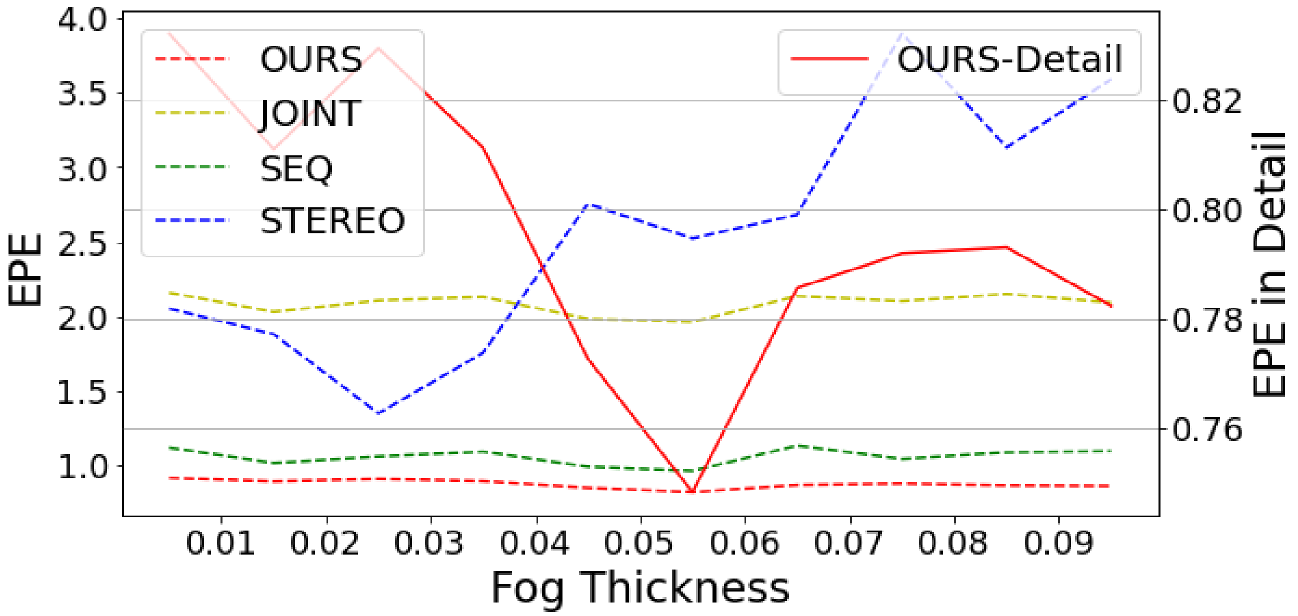
Results

- Influence of depth range.
  - 'OURS' represents the distribution result of our method.
  - 'JOINT' represents the result of SSMDNet.
  - 'SEQ' represents the result of 4Kdehazing + DeepPruner.
  - 'STEREO' represents the result of DeepPruner.



Results

- Influence of fog thickness.
  - 'OURS' represents the distribution result of our method.
  - 'JOINT' represents the result of SSMDNet.
  - 'SEQ' represents the result of 4Kdehazing + DeepPruner.
  - 'STEREO' represents the result of DeepPruner.



Limitations and Discussion

- Model
  - the assumption over atmospheric parameters
    - inhomogeneous scattering median
    - multi-light sources

- Scenes
  - other scattering media
    - Haze
    - Rain
    - Water

Thanks for Your Attention

Paper & Code