Deep Photo Relighting by Integrating Both 2D and 3D Lighting Information

Takashi Machida and Satoru Nakanishi
Toyota Central R&D Labs. Inc.,

Motivation

- The needs for virtual assessment (VA) has been increasing in the field of autonomous driving systems.
- VA can design and evaluate a system/algorithm with virtually generated data.
- Using virtually generated data reduces the workload to collect data from actual driving.

- To generate images, there are two major problems.

➢ Geometrical Problem → F-VIR[1]

Caused by changing camera position.

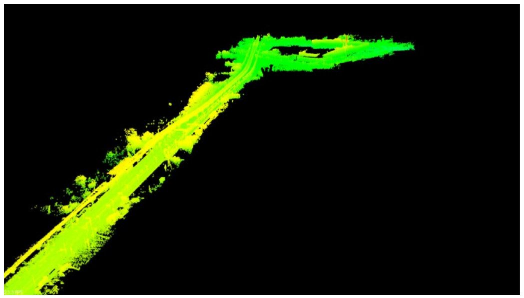
➢ Optical Problem → DPR*

Caused by changing the environment.

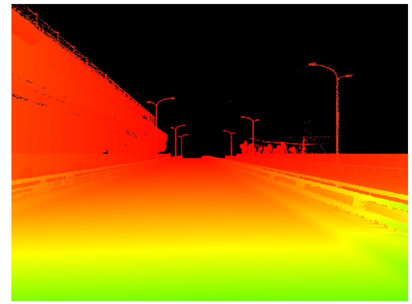
* : Deep Photo Relighting

F-VIR (Free Viewpoint Image Rendering)[1]

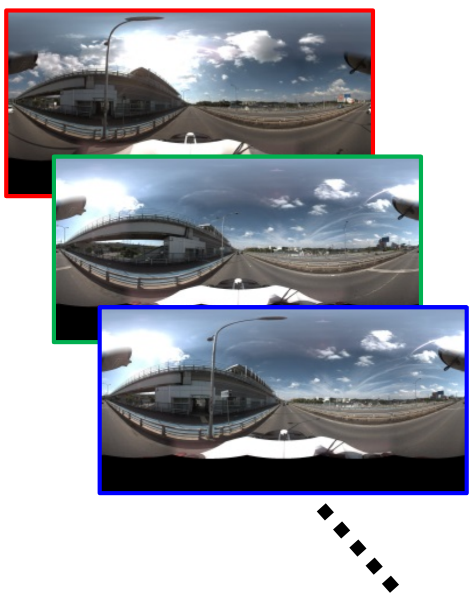
Point cloud (3D information)



1 Generate arbitrary viewpoint depth-image



Omnidirectional images

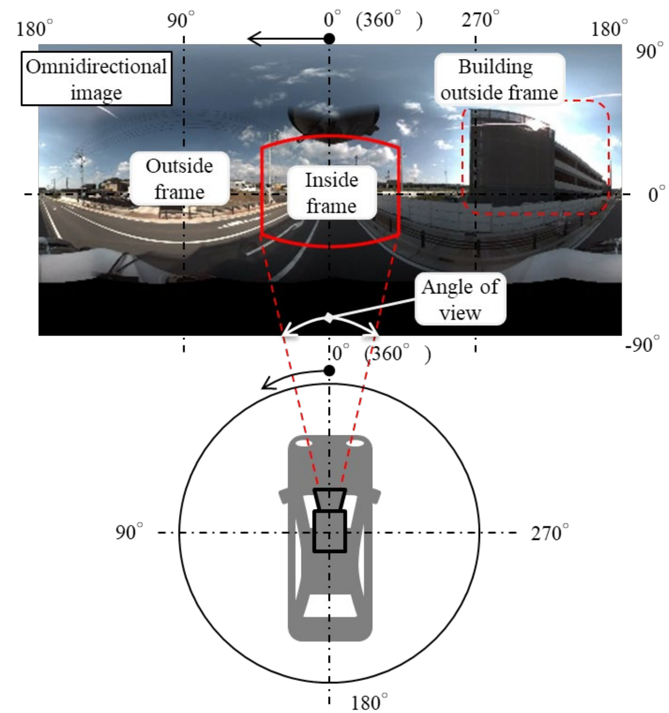


2 Select suitable pixel from multiple omnidirectional images

Problems and Contributions

☐Problems

- The generated image can be reproduced incorrectly because not all the factors are supported.
- Shadow from outside the frame cannot be considered.



☐Contributions

Proposing a practical framework for

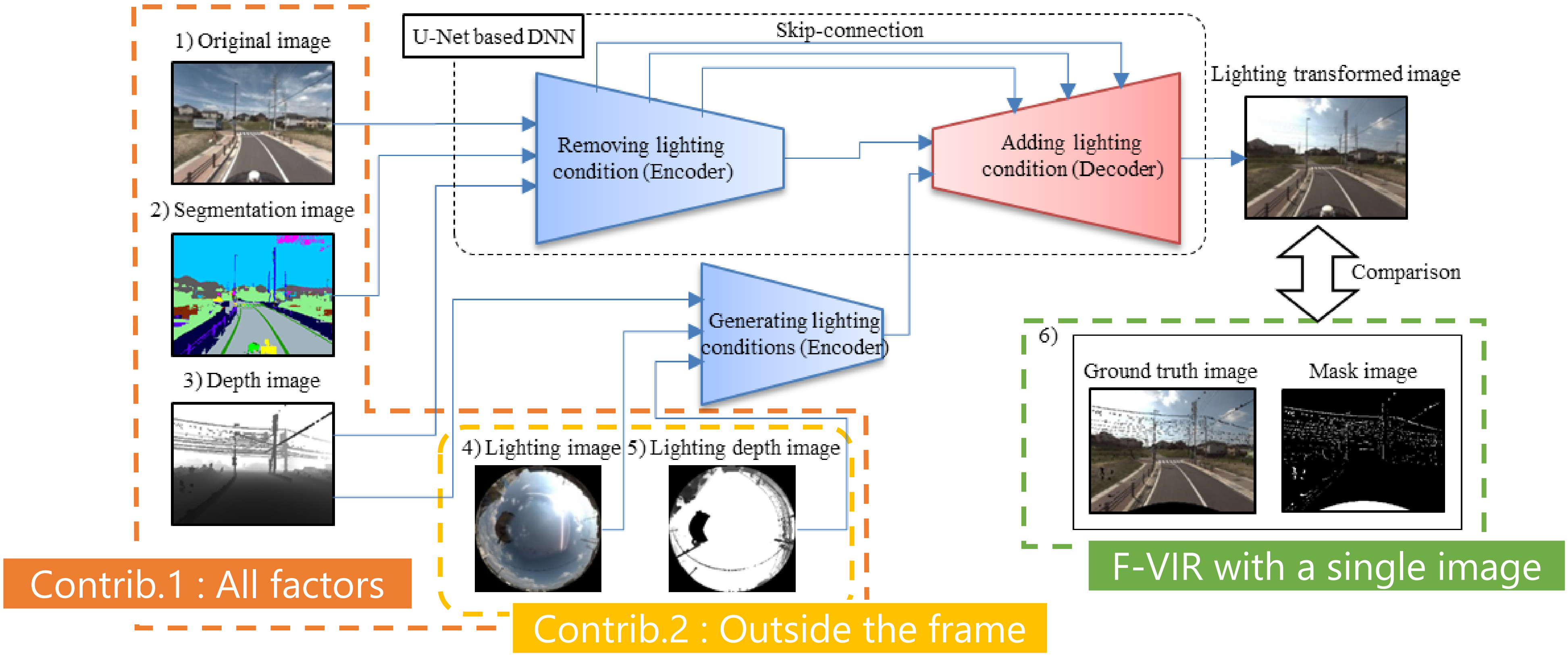
1. Considering all lighting condition factors to correctly transform the lighting conditions of images.
2. Considering the influence of lighting condition not only inside the frame but also outside the frame.

TOYOTA CENTRAL R&D LABS

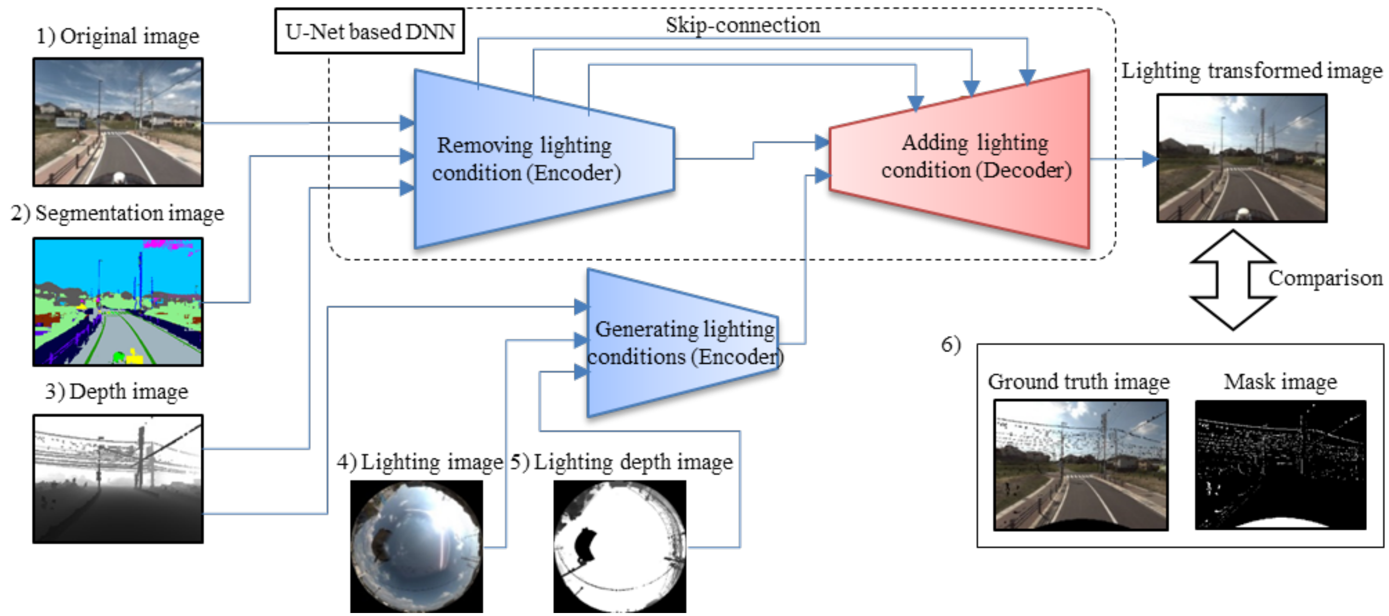
Framework

◻U-Net based DNN

- Two-step process : removing and adding lighting conditions.
- Integrating 2D and 3D lighting information.



2D and 3D input images

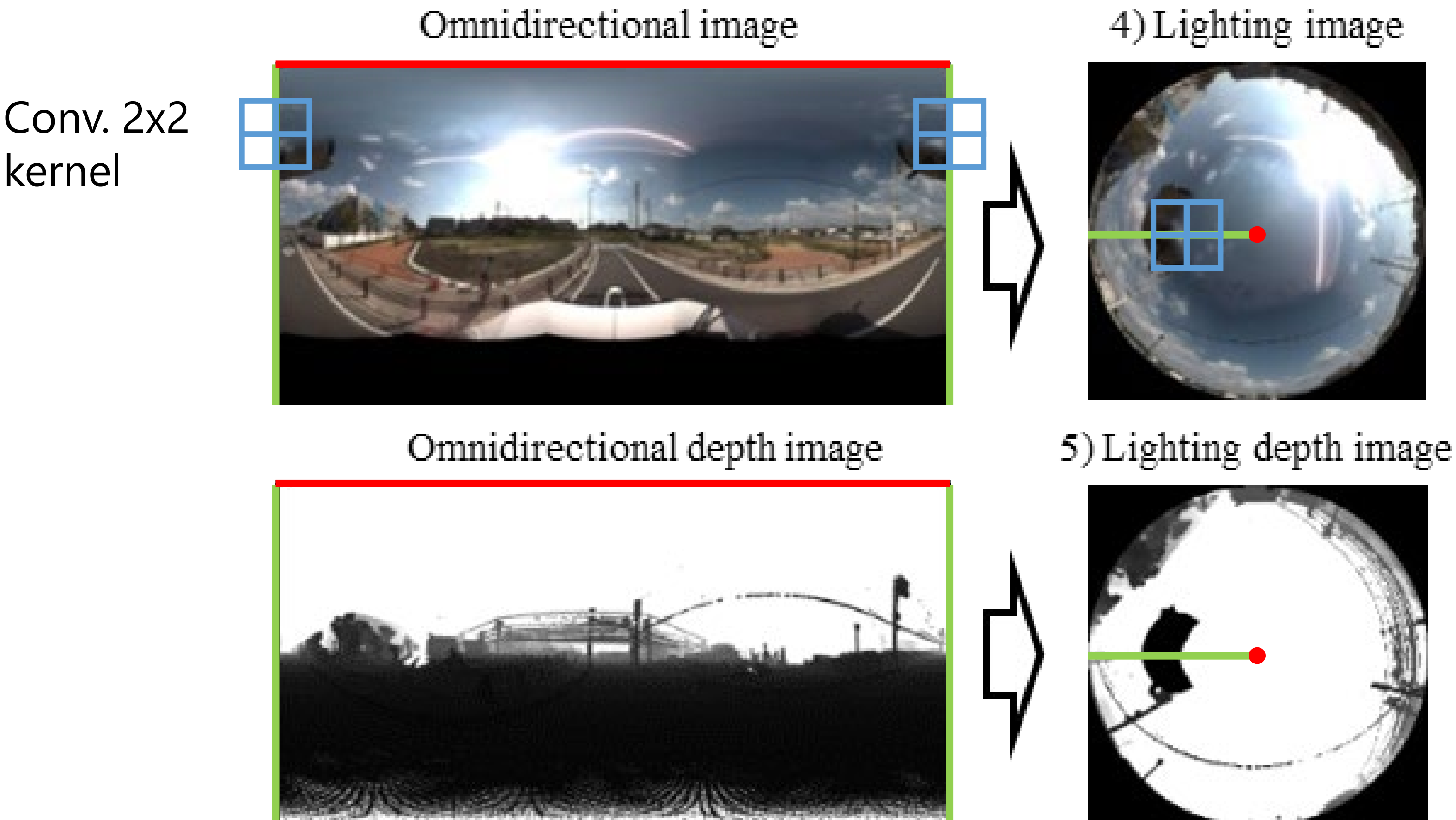


Factors of constructing image		Input image
Viewpoint		Original image
Object material		Original image Segmentation image
Object geometry		Depth image
Lighting condition	Lighting color/intensity	Lighting image Lighting depth image
	Obstructing object shape	Lighting depth image
	Projection surface shape	Depth image

2D and 3D lighting information

- Using the omnidirectional image, the convolution process cannot extract the image feature suitably at the left and right sides.

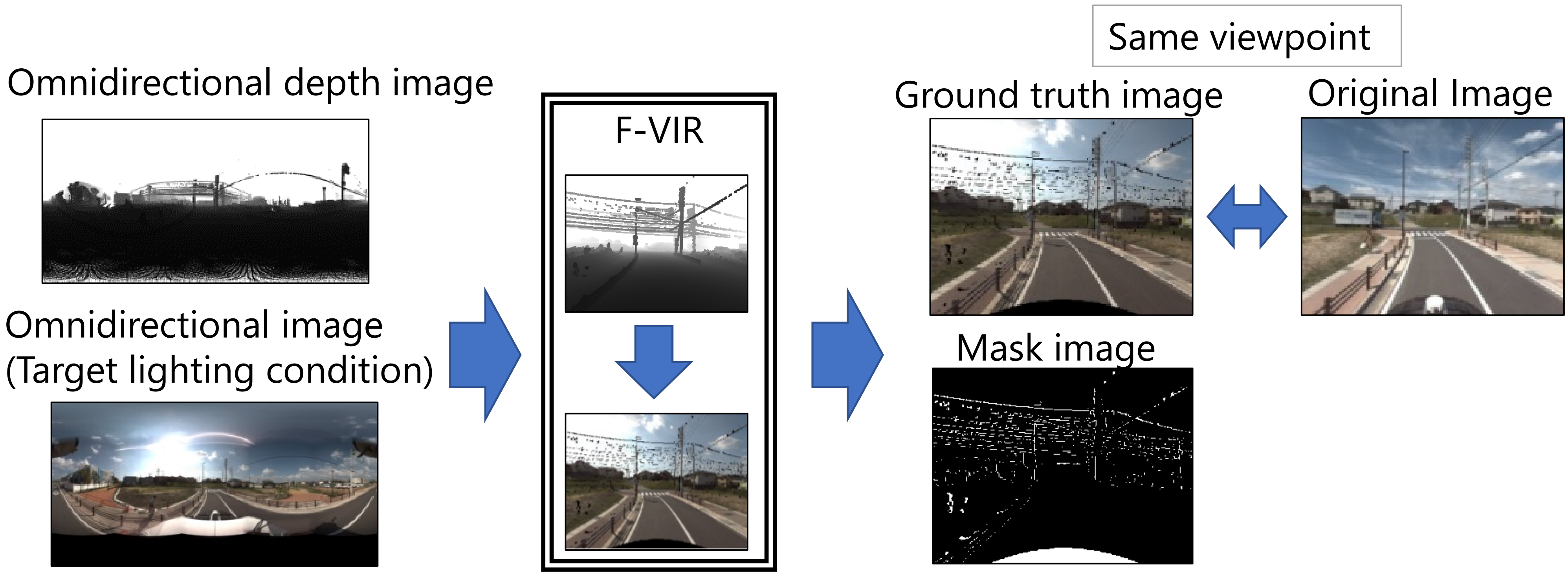
  ➔ Converting zenith format



F-VIR with a single image

- For the learning process, we need images from the same viewpoint with different lighting conditions. However, it's difficult to collect these images during actual driving.

  ➔ Using F-VIR with a single image



Result 1

- DPR can reproduce the shadow due to the obstructing object outside the frame.

(a) Original image (b) Lighting transformed image (c) Ground truth image (d) Omnidirectional image of ground truth at 2 pm

Result 2 : Comparison with CycleGAN[2]

- DPR has few artifacts than CycleGAN.

- DPR is superior to CycleGAN on five indicators.

	L1	PSNR	SSIM	LPIPS	FID	mIoU
DPR	**3.68**	**25.19**	**0.88**	**0.088**	25.97	**0.67**
CycleGAN	7.30	20.60	0.80	0.11	**19.98**	0.60

The FID of DPR is greater than that of CycleGAN because the lighting transformed image is smoothed by the loss function which is based on the average of error in the image.

References

- [1] Oko et al. “Evaluation of image processing algorithms on vehicle safety system based on free-viewpoint image rendering,” *IEEE Intelligent Vehicle Symposium*, 2014
- [2] Zhu et al. “Unpaired image-to-image translation using cycle-consistent adversarial networks,” *ICCV*, 2017