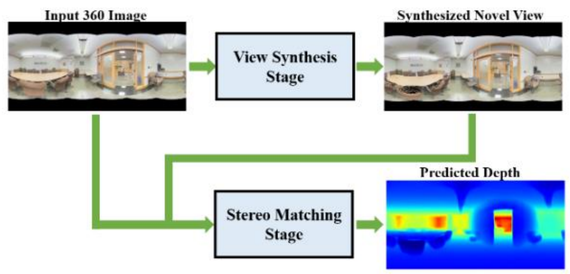
PanoDepth: A Two Stage Approach for Monocular Omnidirectional Depth Estimation

Yuyan Li[1], Zhixin Yan[2], Ye Duan[1], Liu Ren[2]
[1]University of Missouri [2]Bosch Research

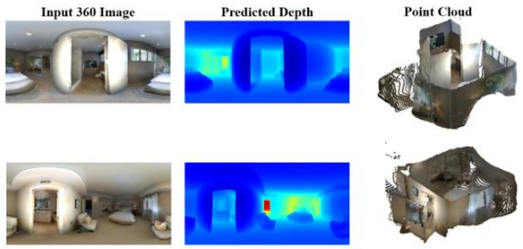
3DV 2021

INTRODUCTION

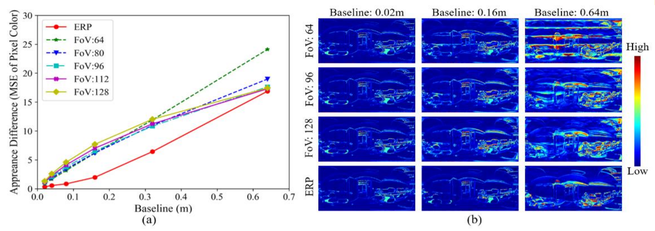
We present a *two-stage* framework, **PanoDepth**, for 360 monocular depth estimation.



Below we give two examples of using **PanoDepth** to process 360 RGB inputs (column 1) and produce 360 Depth outputs (column 2), along with reconstructed point cloud (column 3).



MOTIVATION



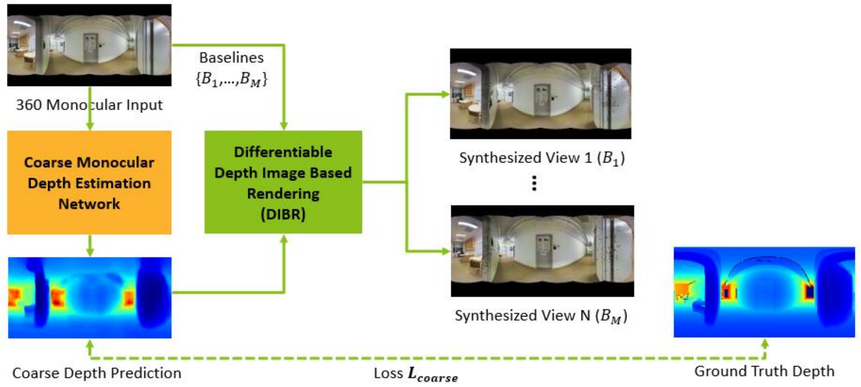
Why two-stage?

- Single-stage monocular methods require large amount of data to map RGB to depth accurately.
- Additional stereo geometry cue is valuable for monocular depth estimation, especially when training data is limited.

Can we successfully synthesize novel view 360 images?

- Large FoV is less sensitive to large baselines, this indicates 360 image which has large FoV, is favorable for view synthesis.
- Under large baseline, 360 image synthesis has less error and artifacts.
- 360 image is well-suited for high-quality novel view synthesis.

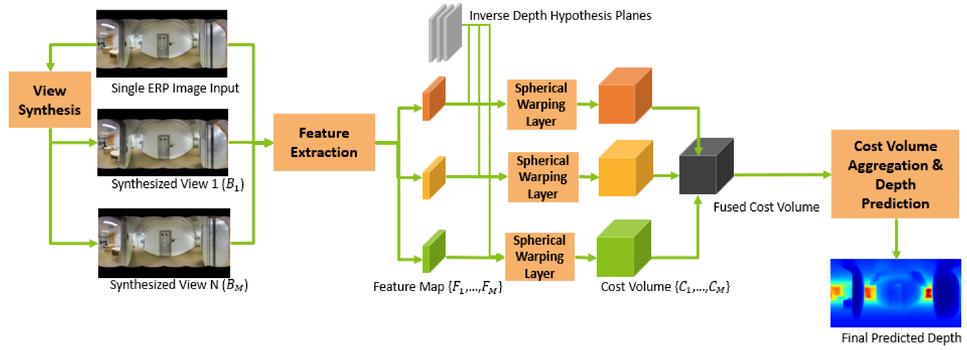
VIEW SYNTHESIS STAGE



View synthesis stage:

- A light-weight monocular coarse depth estimation predicts a coarse depth.
- The coarse depth and the original RGB input is used for novel view synthesis via DIBR.

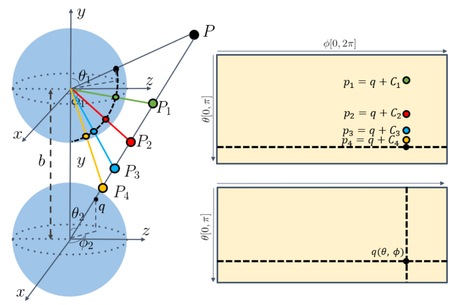
STEREO MATCHING STAGE



Stereo Matching stage:

- All synthesized views and the original view are passed through a multi-view stereo matching network which produces the final high-quality depth.
- A novel customized module, spherical warping layer (SWL), is used to adapt to 360 geometry.
- Cascade mechanism is incorporated to further improve depth quality.

SPHERICAL WARPING LAYER



SWL is a closed-form solution to warp a reference 360 image to a target 360 image.

The Purpose of SWL:

- Enabling multi-view stereo of 360 images.
- Enabling both vertical and horizontal 360 stereo.
- Enabling cascade mechanism.

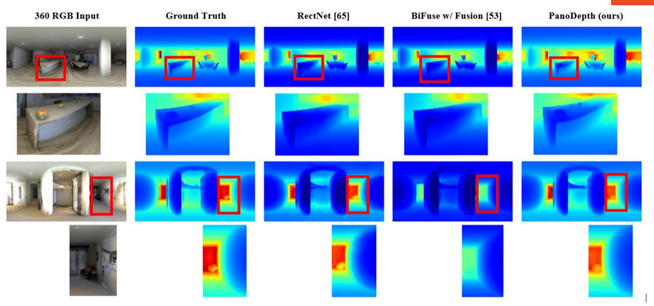
The steps of SWL:

- Direct sampling on the inverse depth to obtain depth hypothesis planes.
- SWL transforms the depth hypothesis to reference view and target view disparity based on spherical geometry.
- Using the spherical disparity, a reference view is mapped to the targe view domain.

CONCLUSION

- We propose a novel, model-agnostic, two-stage network, **PanoDepth**, for 360 monocular depth estimation.
- We introduce **SWL** to adapt to 360 geometry. SWL enables us to investigate multi-view stereo and horizontal stereo of 360 images.
- **PanoDepth** achieves state-of-the-art performances and proves to have strong generalization ability.

EXPERIMENTAL RESULTS



Datasets	Methods	Abs Rel↓	RMSE↓	δ_1 ↑	δ_2 ↑	δ_3 ↑
Stanford2D3D [3]	FCRN [36]	0.1837	0.5774	0.7230	0.9207	0.9731
	RectNet [65]	0.1409	0.4568	0.8326	0.9518	0.9822
	BiFuse with fusion [53]	0.1209	0.4142	0.8660	0.9580	0.9860
	PanoDepth(Ours)	**0.0972**	**0.3747**	**0.9001**	**0.9701**	**0.9900**
360D [65]	FCRN [36]	0.0699	0.2833	0.9532	0.9905	0.9966
	RectNet [65]	0.0702	0.2911	0.9574	0.9933	0.9979
	BiFuse with fusion [53]	0.0615	0.2440	0.9699	0.9927	0.9969
	ODE-CNN [11]	0.0467	**0.1728**	0.9814	**0.9967**	**0.9989**
	PanoDepth(Ours)	**0.0456**	0.1955	**0.9830**	0.9957	0.9984