Spherical Epipolar Rectification for Deep Two-View Absolute Depth Estimation.

Pierre-André Brousseau and Sébastien Roy
International Conference on Computer Vision 2025

Stereo from Motion

Small moving platforms often rely on a single camera to navigate and they would benefit from real-time depth estimation.

Goal: Obtain a dense depth map from two consecutive views of a monocular image sequence obtained from a camera in motion in a static scene.

- Given known camera pose, we aim to rectify the images to leverage traditional and deep stereo matching.

Contribution

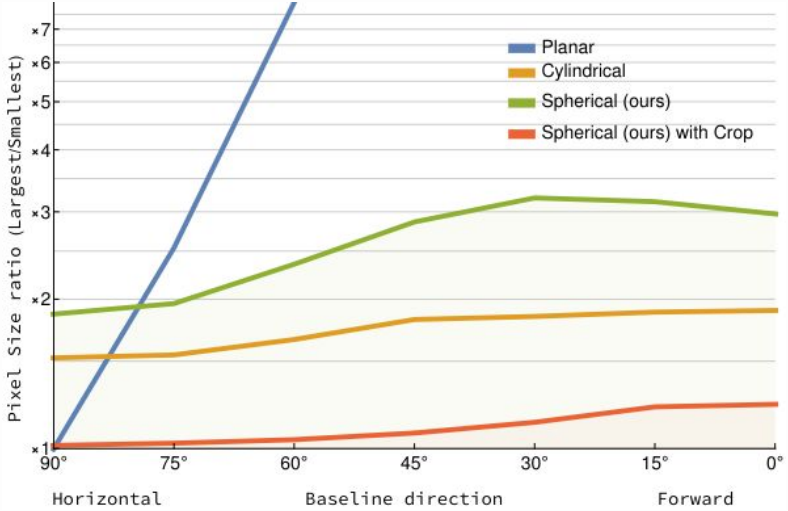
A differentiable spherical epipolar rectification model that makes deep stereo matching possible (training and inference) on single camera image pairs.

- A spherical crop operator that reduces image distortion.
- A spherical rectified stereo matching process for two-view absolute depth estimation.

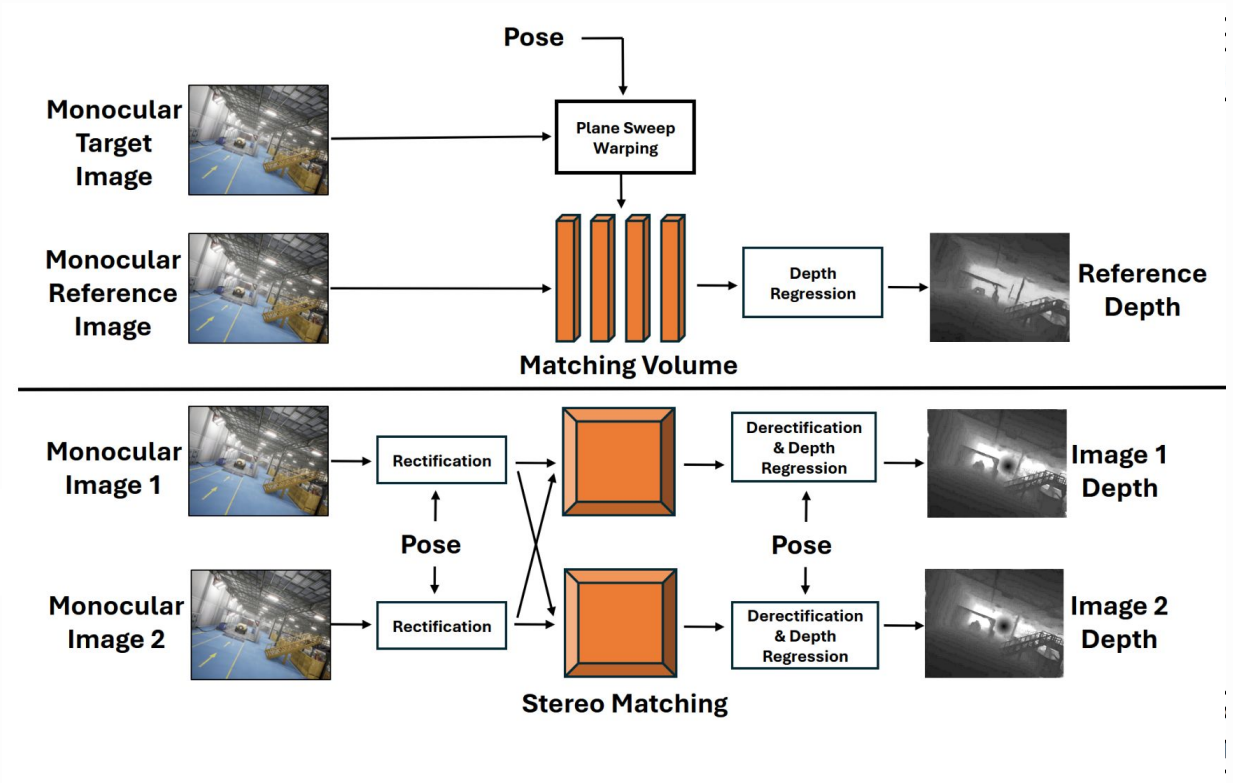
Minimizing Distortion

Planar epipolar rectification breaks down if the epipole enters the image, with severe distortion triggered by even minor forward motion.

Our proposed model minimizes distortion.



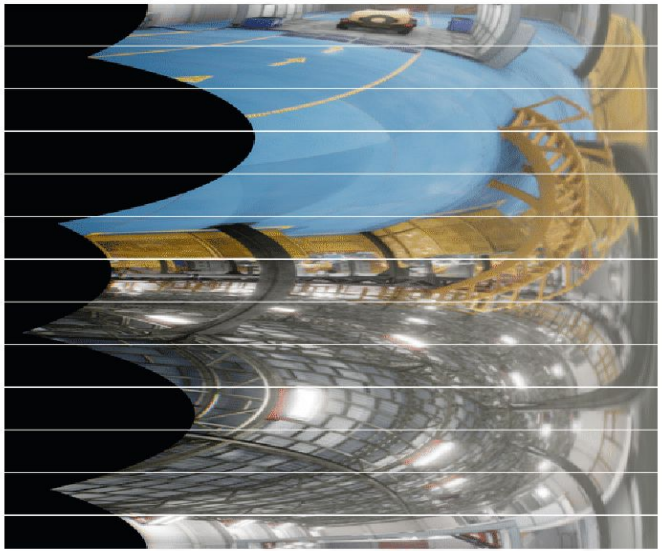
Monocular Two-View Depth Estimation



Usual Approach vs Suggested Approach

Spherical Epipolar Rectification

Image Representation

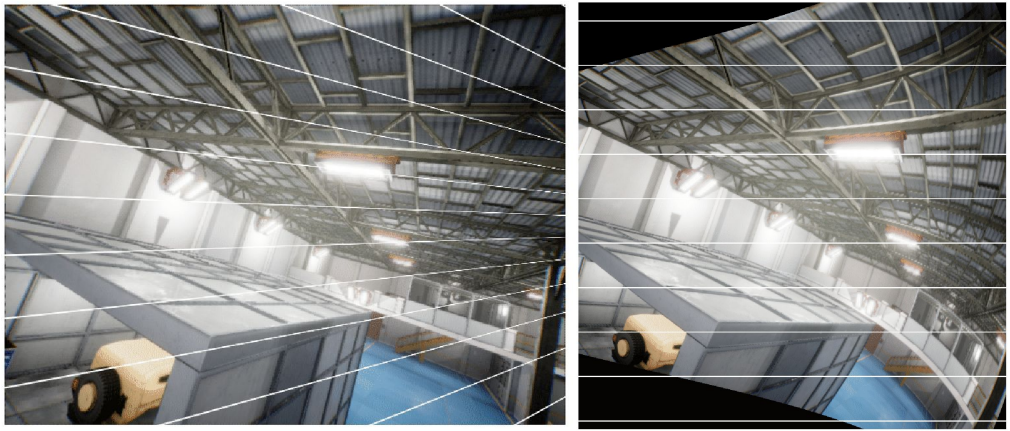


Spherical Representation

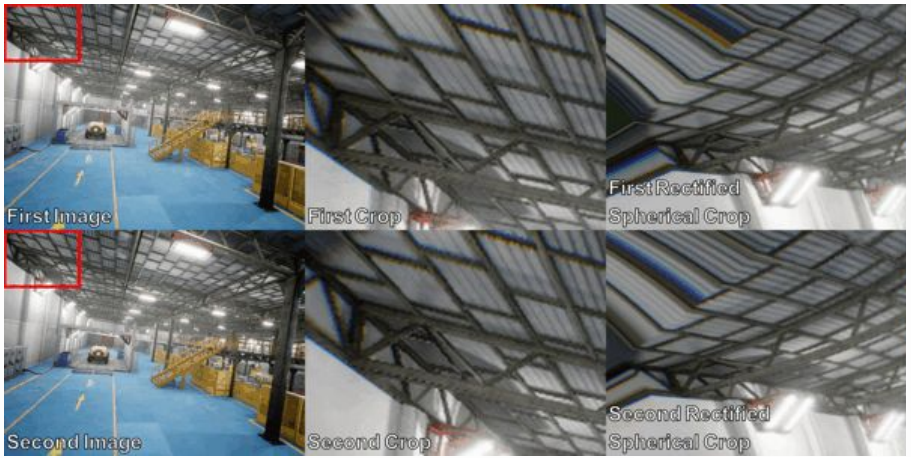
Spherical Epipolar Model

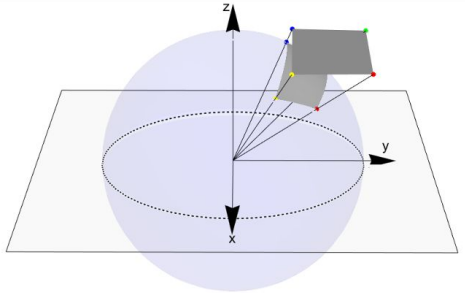
Our rectification model:

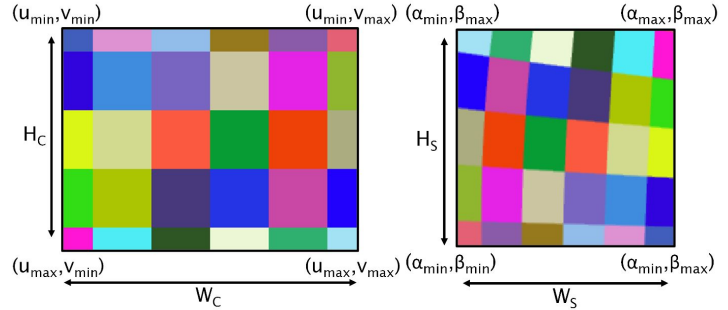
- is valid for all camera geometries.
- is differentiable and can be integrated with deep networks.
- allows for **absolute depth computation from disparity.**
- allows for control over rectified image size.
- disentangles geometry from matching in the depth estimation problem.



Spherical Crop

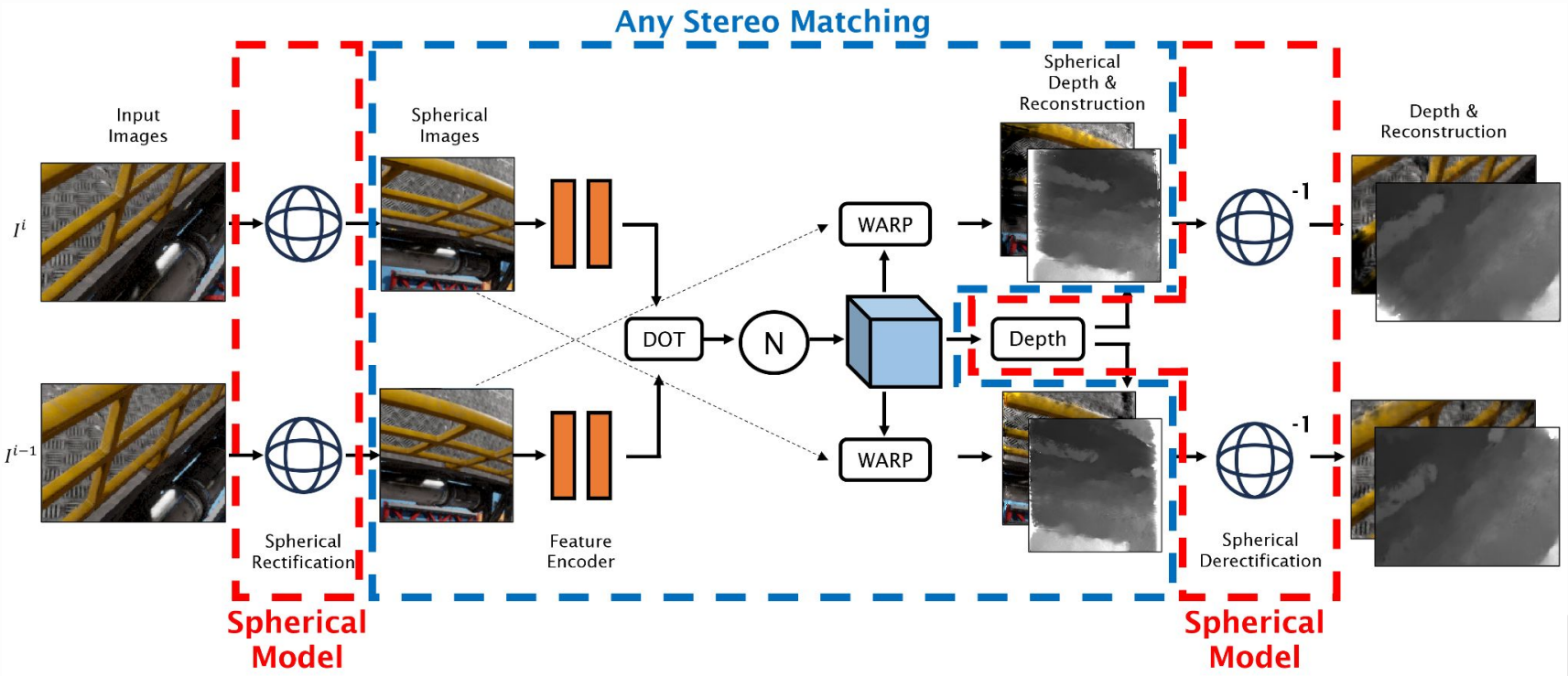




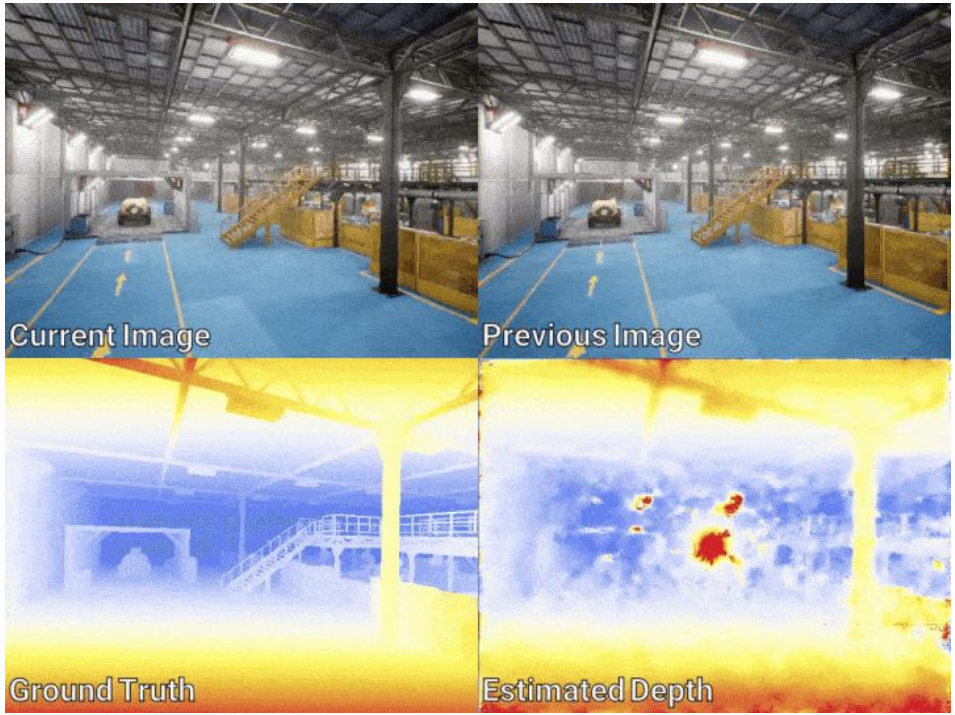


Spherical Rectification Model

Spherical Stereo Architecture



Qualitative Results



*Every depth map is generated using the two most recent image frames.

Results Against Literature

Method		Error - Lower is better ↓				Accuracy - Higher is better ↑			Nbr. of Params.
		Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$	
Depth Anything V2 [196]	Supervised & Scaled[1]	0.33	1.04	2.63	0.30	0.38	0.80	0.98	97.47M
GMDepth [192]	Supervised	0.24	1.20	2.55	0.24	0.60	0.85	0.95	7.32M
Lite-Mono 8M [214]	Supervised	0.23	0.61	1.41	0.34	0.70	0.78	0.81	8.78M
DeepSfm Rev [175]	Supervised	0.15	**0.41**	1.42	0.14	0.80	0.95	**0.99**	16.64M
MonST3R [212]	Supervised & Scaled[1]	0.15	0.49	1.30	0.14	0.79	0.95	0.98	577.81M
Spherical Stereo (ours)	Supervised	**0.10**	**0.41**	**0.98**	**0.12**	**0.92**	**0.96**	0.98	**0.73M**
Spherical Stereo (ours)	Self-Supervised	0.23	3.32	2.39	0.24	0.75	0.84	0.90	**0.73M**

Absolute Depth Performance on TartanAir

Multi-image Coherence

The model is trained with the **current frame (i)** and the **previous frame (i-1)**.
→ We can use more/other frames at inference.

		Error - Lower is better ↓				Accuracy - Higher is better ↑		
Nb. of Frames	Frame Timesteps	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
2	$\{i, i-1\}$	0.23	3.32	2.39	0.24	0.75	0.84	0.90
2	$\{i, i-3\}$	0.40	7.00	2.72	0.29	0.70	0.80	0.86
2	$\{i, i-2\}$	0.30	4.87	2.43	0.25	0.74	0.83	0.89
2	$\{i, i+3\}$	0.32	5.03	2.75	0.28	0.70	0.81	0.87
2	$\{i, i+2\}$	0.24	3.39	2.46	0.25	0.74	0.83	0.89
2	$\{i, i+1\}$	0.22	3.02	2.53	0.25	0.74	0.83	0.89
3	$\{i, i-1, i-2\}$	0.23	3.16	2.26	0.23	0.75	0.85	0.90
3	$\{i, i-2, i+2\}$	0.20	2.05	2.03	0.21	0.77	**0.87**	**0.92**
3	$\{i, i-1, i+1\}$	0.19	2.08	2.21	0.21	0.77	0.86	0.91
5	$\{i, i-1, i+1, i-2, i+2\}$	**0.17**	**1.58**	**2.02**	**0.20**	**0.78**	**0.87**	**0.92**

Absolute Depth Performance on TartanAir

Contact

Pierre-André Brousseau
pierre-andre.brousseau@umontreal.ca

Website: sphericalstereo.github.io

Takeaways

1. A novel spherical epipolar rectification model for solving absolute depth from two views of a monocular image sequence.

2. Differentiable model that can be **effectively integrated with deep stereo matching**.

3. **Reduces image distortion** compared to previous rectification methods.