CVMP2023

LFSphereNet: Real Time Spherical Light Field Reconstruction from a Single Omnidirectional Image

Manu Gond, *Emin Zerman, Sebastian Knorr, Mårten Sjöström*

The 20th ACM SIGGRAPH European Conference on Visual Media Production, London, 2023

Internet

Image source: Mid Sweden University - https://www.miun.se/kontakt/press/nyhetsarkiv/2022-10/battre-arbetsmiljo-och-lagre-kostnader-med-fjarrstyrda-virkeslastare/

Motivation

- Omnidirectional content improves telepresence applications
  - Full 360 Field of View (FoV)
  - Provides 3 Degrees of Freedom (DoF)
  - Easy to use 360 cameras

Fig 1: An omnidirectional Camera

Limitations

No physical Movement

(No 3DoF+ or 6DoF)

Causes discomfort or nausea

Multicamera setup requires higher bandwidth

Fig 2: VR based crane operation from inside the truck cabin[1]

[1]BRUNNSTRÖM, Kjell, et al. Quality of experience of hand controller latency in a virtual reality simulator. In: *Human Vision and Electronic Imaging 2019.*

LFSphereNet

- Generation of a **Spherical Light Field**

Limitations

No physical Movement

(No 3DoF+ or 6DoF)

Causes discomfort or nausea

Multicamera setup requires higher bandwidth

Allows 3DoF+

More natural viewing

One camera setup requires less bandwidth

LFSphereNet

[1]BRUNNSTRÖM, Kjell, et al. Quality of experience of hand controller latency in a virtual reality simulator. In: *Human Vision and Electronic Imaging 2019.*

Recent Studies

- Planar Light Field (LF) Reconstruction
  - Single/Sparse Input to NxN LF
  - Near Real Time
  - Tradeoff between spatio-angular resolution
  - Limited FoV

- Spherical LF Reconstruction
  - Has a FoV of 360°
  - No method for single ODI to dense LF

- Spherical View Synthesis Methods Exist

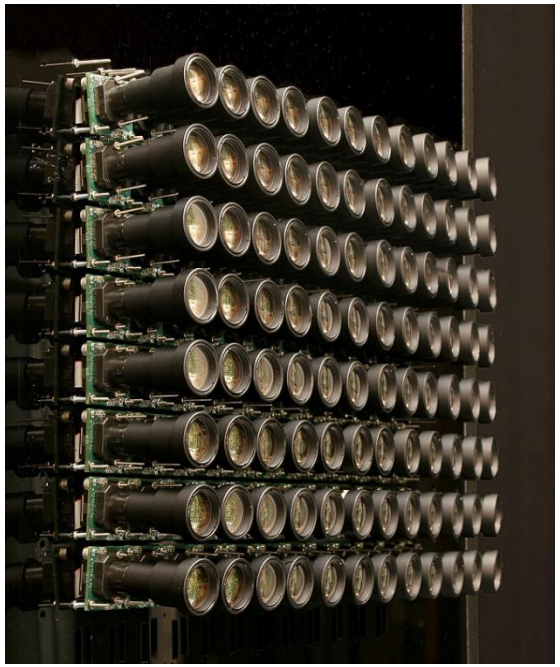


Fig 3: Stanford Multi-Camera Array[2]

[2]http://graphics.stanford.edu/projects/array/

Related Work: PanoSynthVR

- Uses Multi Plane Representation
  - Warp Padding Near Edges
  - Multi Cylinder Image (MCI)

Fig 4: Single ODI

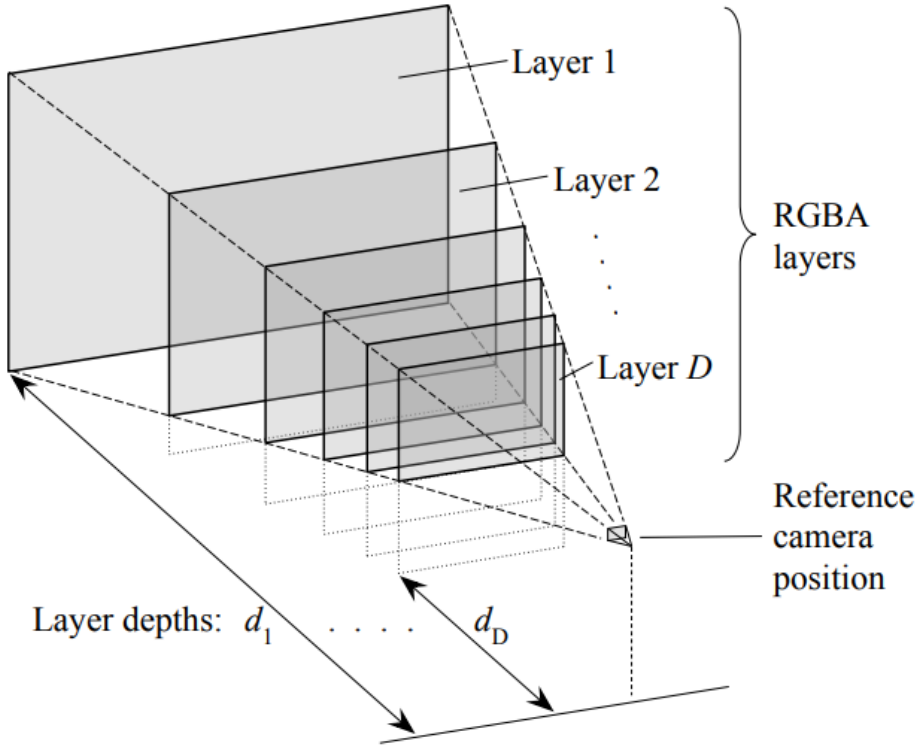


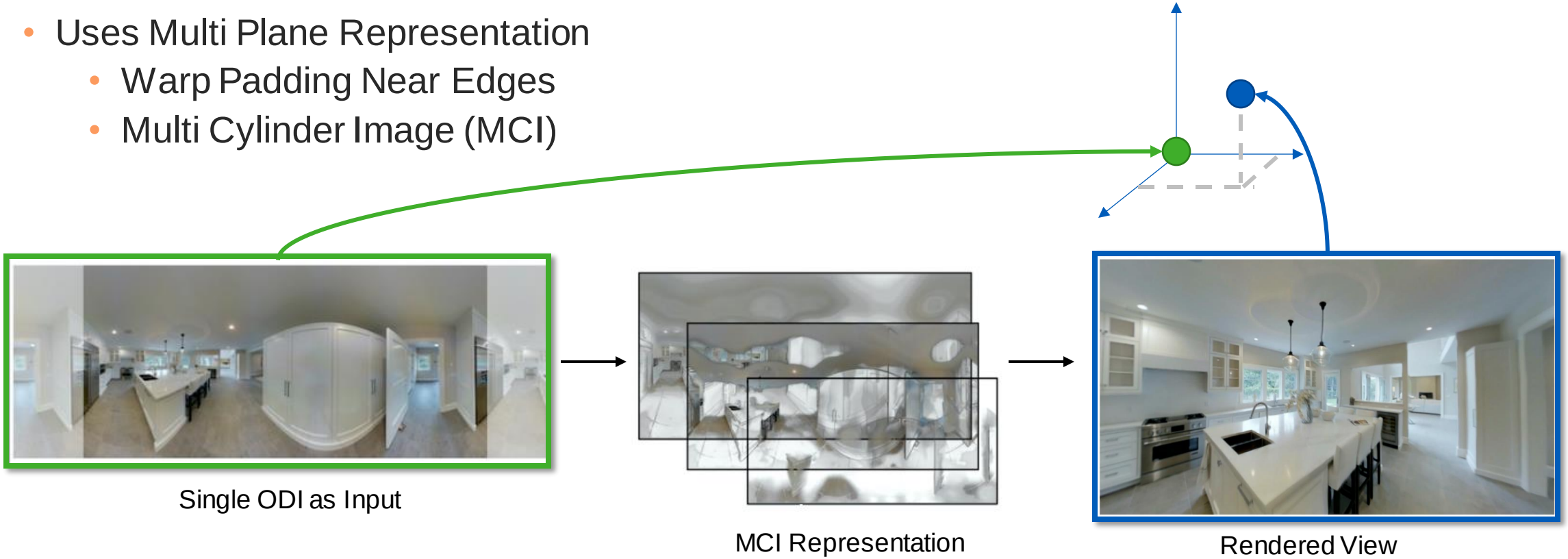
Fig 5: Multi Plane Image Representation[3]

PanoSynthVR: Waidhofer, J., Gadgil, R., Dickson, A., Zollmann, S. and Ventura, J., 2022 PanoSynthVR: Toward Light-weight 360-Degree View Synthesis from a Single Panoramic Input. In *2022 IEEE (ISMAR)* (pp. 584-592)

[3]Tucker, R. and Snavely, N., 2020. Single-view view synthesis with multiplane images. In *Proceedings of the IEEE/CVF* (pp. 551-560).

Related Work: PanoSynthVR

- Uses Multi Plane Representation
  - Warp Padding Near Edges
  - Multi Cylinder Image (MCI)



PanoSynthVR: Waidhofer, J., Gadgil, R., Dickson, A., Zollmann, S. and Ventura, J., 2022 PanoSynthVR: Toward Light-weight 360-Degree View Synthesis from a Single Panoramic Input. In *2022 IEEE (ISMAR)* (pp. 584-592)

Related Work: PanoSynthVR

- Limitations
  - Building the layers takes time
  - Uses Google's MPI Network Backbone
  - Generated ODIs have wrong Geometry

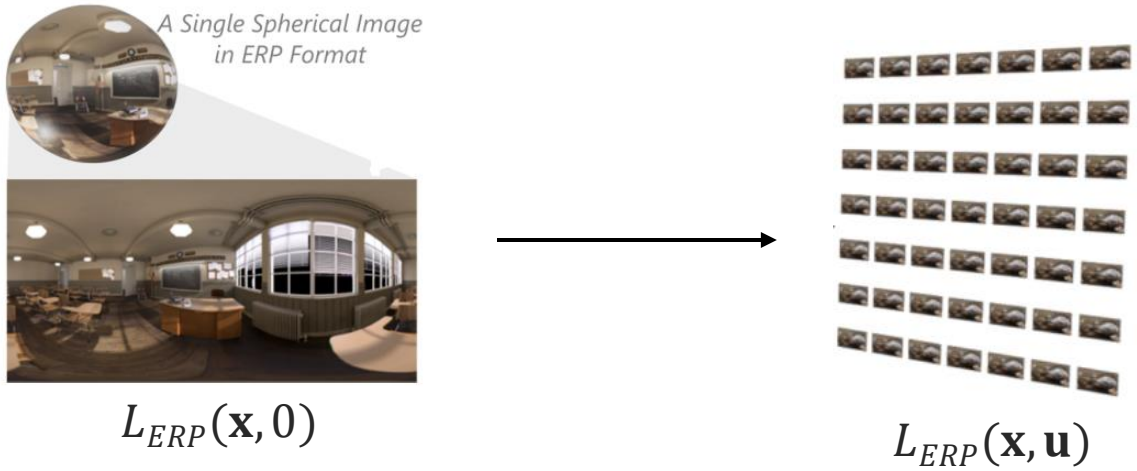
Ground Truth

PanoSynthVR

PanoSynthVR: Waidhofer, J., Gadgil, R., Dickson, A., Zollmann, S. and Ventura, J., 2022 PanoSynthVR: Toward Light-weight 360-Degree View Synthesis from a Single Panoramic Input. In *2022 IEEE (ISMAR)* (pp. 584-592)

Generating a Spherical Light Field

- Given $L_{ERP}(\mathbf{x}, 0)$
  - central image of NxN spherical LF
- Reconstruct $L_{ERP}(\mathbf{x}, \mathbf{u})$ where
  - $\mathbf{x}$: spatial coordinates (x, y)
  - $\mathbf{u}$: angular coordinates (u, v)
- Equirectangular (ERP) images have distortion
  - Limits usage of convolution kernels



$L_{ERP}(\mathbf{x}, 0)$ $L_{ERP}(\mathbf{x}, \mathbf{u})$

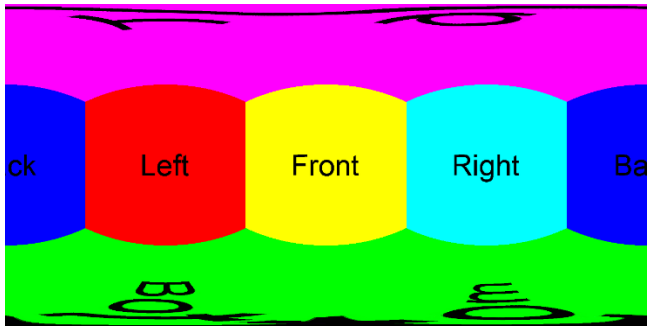
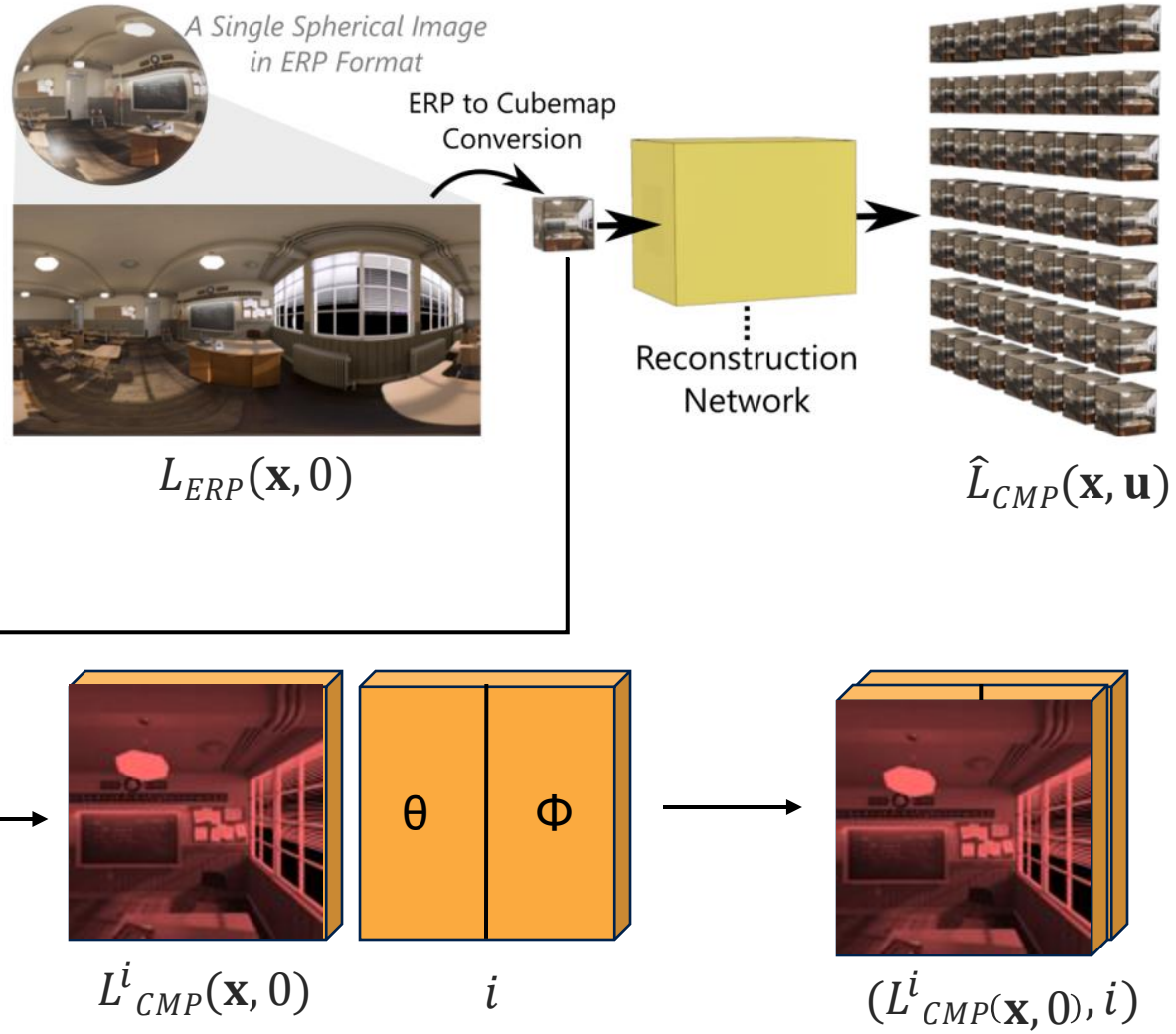


Fig 6: Visualization of Distortions in ERP Image

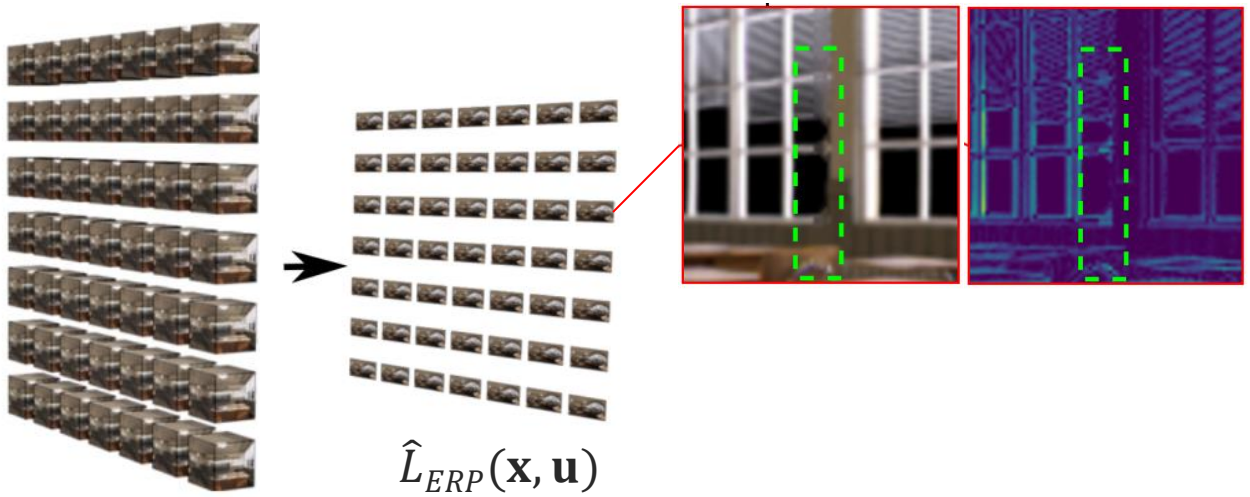
LFSphereNet – Pre-Processing and Reconstruction

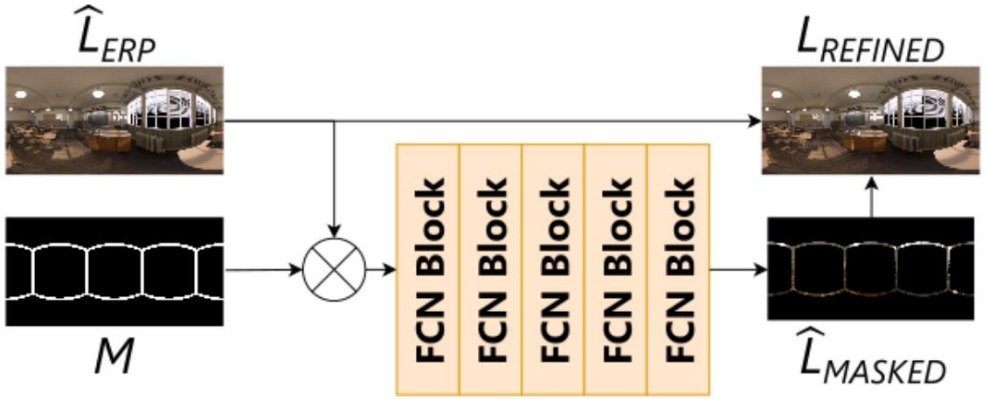
- Cubemap (CMP) Conversion
  - Minimize distortions in the image
- Reconstruction in CMP
  - Single Cube Face input at a time
- Positional Encoding
  - Using single channel map holding positional information (θ,Φ)



LFSphereNet – Post Processing and Refinement

- Reconstructed LF in CMP can be converted back to ERP $\hat{L}_{ERP}(\mathbf{x}, \mathbf{u})$
- The $\hat{L}_{ERP}(\mathbf{x}, \mathbf{u})$ will have minor discontinuity near borders
- The Mask M defines width of region which needs refinement





Data Collection

Created two datasets:

- Synthetic Data
  - Using Blender
  - 5 Different open-source scenes
  - 1024x2048x**7x7** resolution

Fig: Some examples from Synthetic Data

- Real Data
  - Captured data
    - Using a moving station
    - Captured with Insta 360X3
  - 6 different real-world scenes
  - 1024x2048x**1x7** resolution

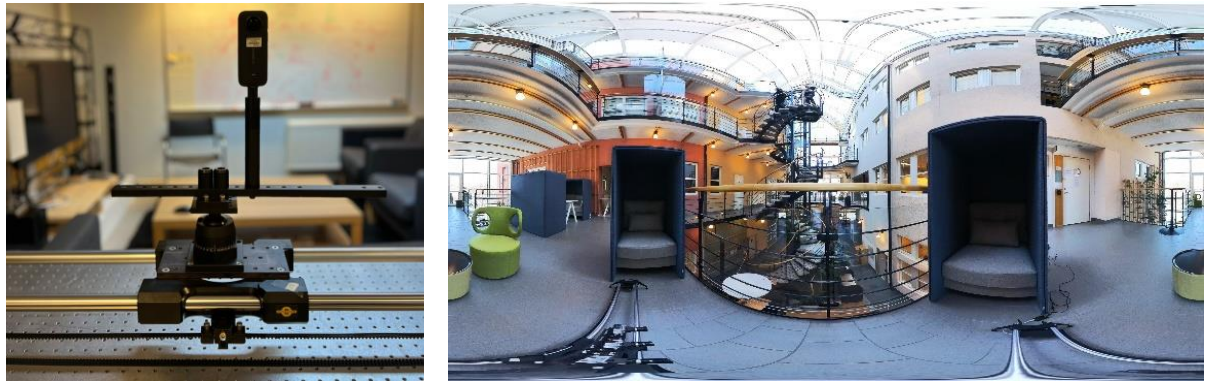


Fig: Some examples of Real Data captured using moving station as shown in left.

Quantitative Results

- Outperforms SOTA in
  - SSIM, MS-SSIM, PSNR
- Lowest Inference Time
  - For 1024x2048x7x7 LF
  - On 1080Ti

Metric	360ViewSynth	PanoSynthVR	LFSphereNet
MAE	0.0922	0.0265	**0.0125**
DISTS	0.1215	**0.0531**	0.0880
LPIPS	0.1935	**0.0670**	0.0825
PSNR	32.89	34.76	**37.45**
SSIM	0.6495	0.7878	**0.9121**
MS-SSIM	0.7690	0.8687	**0.9691**
FSIMC	0.8443	0.9148	**0.9573**
VIFP	0.3710	0.4978	**0.8035**
Runtime	12.4845	2.7077	**0.0606**

~7% improvement

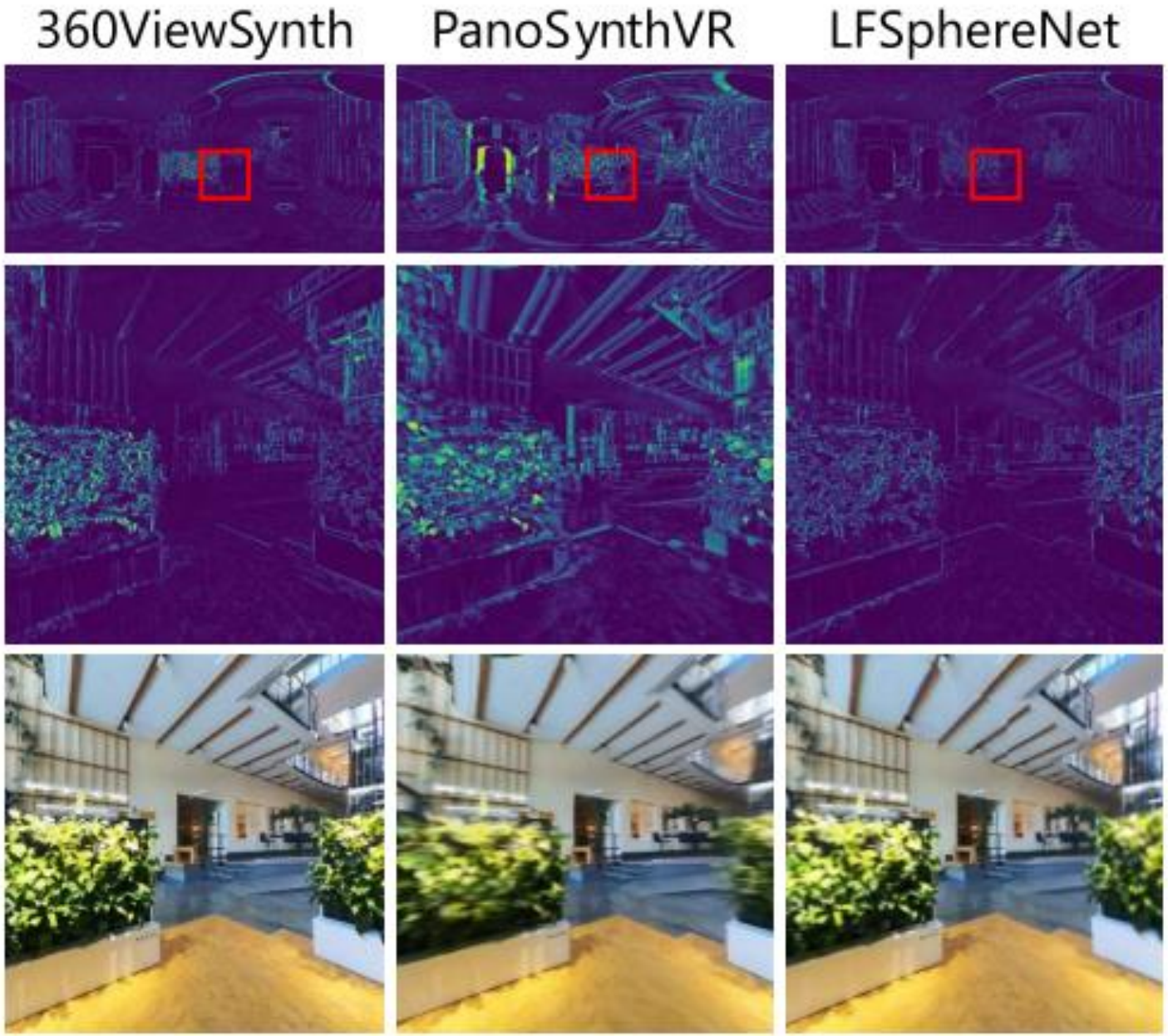
~16% improvement

44x speed-up

Best Values in Bold. Runtime is in seconds

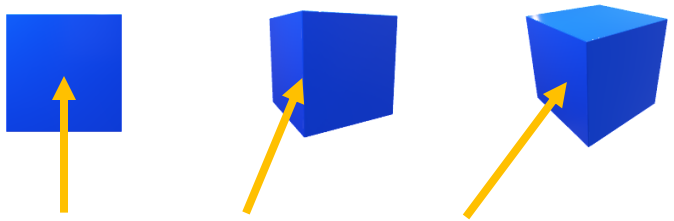
Qualitative Results

- Tested on real-world scene
- Error map around the leaves shows clear difference
- LFSphereNet has the lowest error



Rendering Demo

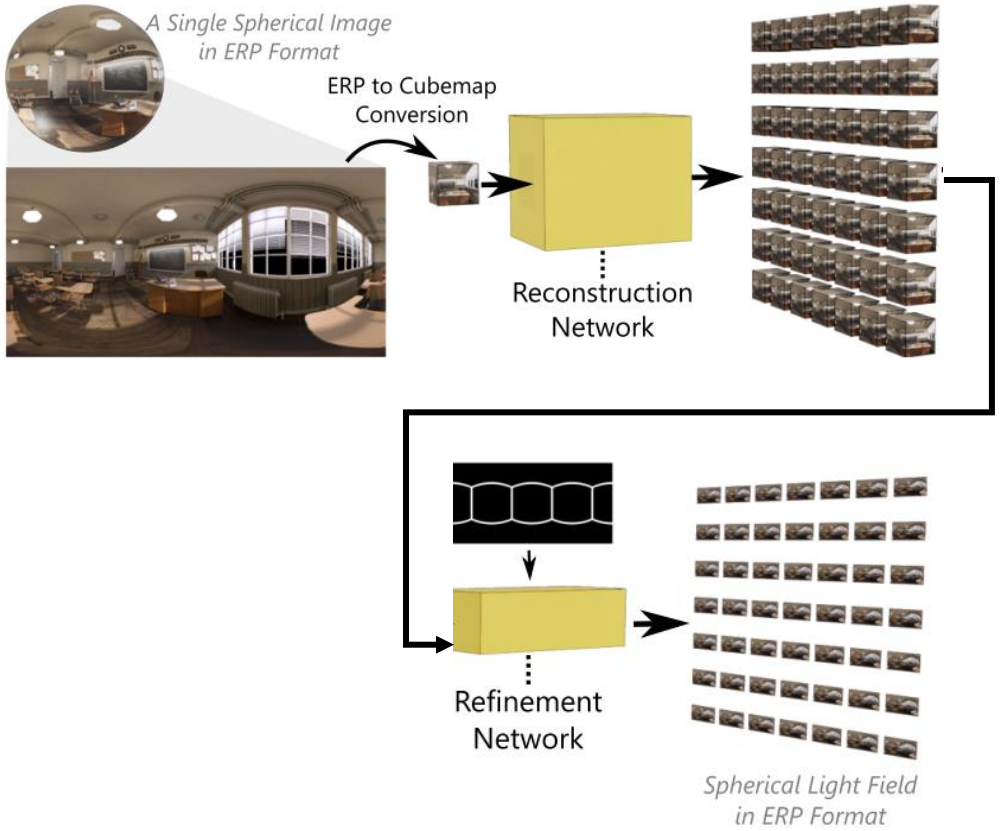
- Inference time is 0.06 secs
  - For 1024x2048x7x7 LF



- We only need max of 3 cube faces at any given time
  - Therefore faster inference
  - Demo on RTX 2070 Super at 60 FPS

Conclusion

- LFSphereNet
  - Achieves 3DoF+ and 6DoF
  - Enables Real-Time Applications to Use LF
- Spherical LF Dataset
  - Can Be Used to Train Different Reconstruction Methods
  - Perform Subjective Tests on LF Rendering



Dataset, supplementary materials can be found in our paper.

CVMP2023

plenoptima

Mittuniversitetet
MID SWEDEN UNIVERSITY

Technische Universität Berlin

Manu Gond

Mid Sweden University

Technical University of Berlin

manu.gond@miun.se