View-dependent MPIs in minutes, not hours

Jochem Gebraad

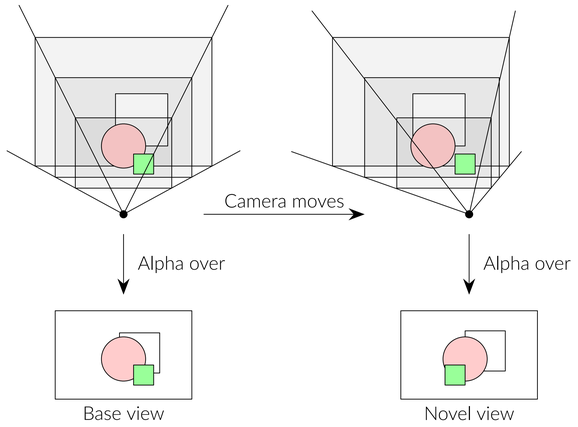
Delft University of Technology

1. The problem

Goal: render novel views of a scene in real time from a few photos.

Multi-plane image (MPI): a stack of semi-transparent RGBA planes at increasing depths, warped and composited front-to-back. Renders in real time, but assumes diffuse surfaces: no reflections or refraction.

View-dependent MPIs exist, but are slow to build. NeX learns geometry *and* an appearance basis jointly, per scene: hours to days.



A diffuse MPI shows every surface the same colour from any viewpoint: RGBA planes warped and composited front-to-back.

2. Research question

Can a decoupled pipeline build view-dependent MPIs quickly while still beating a diffuse MPI?

Idea: stop optimizing geometry and the basis together.

- **Geometry is fixed** from an depth estimator.
- **Appearance is a fixed** basis of the view direction.
- Fit only a few **per-pixel coefficients** ⇒ lightweight regression instead of joint optimization.

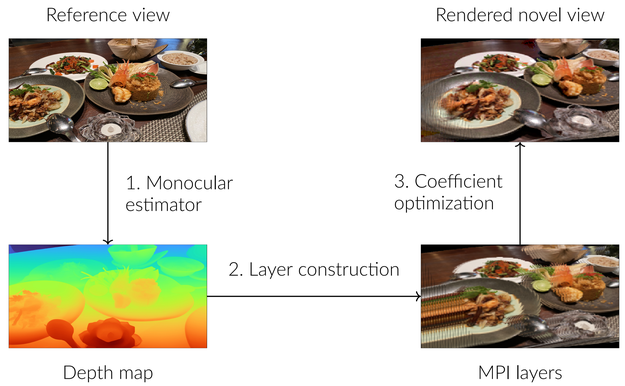
3. Method

Inputs: 1 reference view + its depth map + N nearby views.

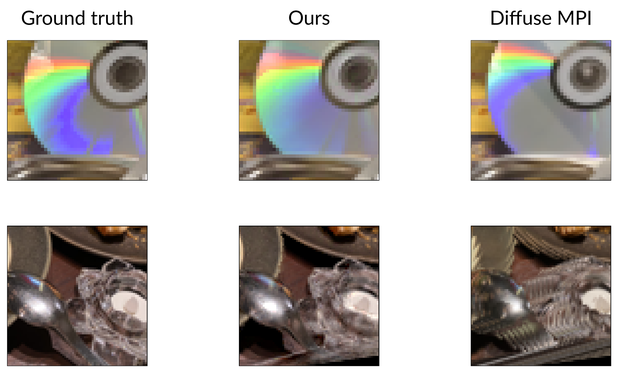
Geometry (fixed): 32 planes uniform in disparity; each starts as the reference RGB, alpha from the depth map. No optimization.

Appearance (fitted): each pixel colour is a view-independent base plus a view-dependent residual. Only the per-pixel α_b are trained, by MSE on held-out views through a differentiable renderer.

Two fixed bases compared: Fourier features vs. spherical harmonics.



Pipeline: estimated depth → fixed MPI geometry → fit per-pixel coefficients to add view dependence.



Held-out novel views: ours recovers reflections the diffuse MPI cannot.

4. Results

Beats a diffuse MPI, fast? Yes: MPI built in seconds to ∼2 min/scene (compared to SOTA hours), rendering at ∼30 fps.

Method	LPIPS ↓	SSIM ↑	PSNR ↑	Build
Diffuse MPI	0.100	0.730	21.0	—
Spherical harmonics	0.060	0.889	**26.0**	∼1.7 min
Fourier (ours)	**0.040**	**0.893**	25.9	∼1 min

Novel-view quality vs. a diffuse MPI from the same geometry. Build = per-scene wall-clock construction time.

Q1: which basis? Fourier has lower LPIPS and faster build at equal PSNR/SSIM: forward-facing views span a narrow cone, without the wasted directions spherical harmonics carry.

Q2: how little is enough? Quality saturates at ∼25 gradient steps and 24 coefficients; more capacity only overfits.

5. Discussion

The method trades quality for a fast build, so it fits many scenes built once (a catalogue, a multi-room tour) rather than a single hero asset. Quality is capped by the borrowed depth and the fixed basis, and the gap to a learned basis is not measured.

6. Conclusion

Fast view-dependent MPIs *are* possible: borrowed depth + a frozen basis recover much of the gain over a diffuse MPI in minutes, not hours by not matching a fully learned basis like NeX.

Supervisor: Petr Kellnhofer

EEMCS, Delft University of Technology, The Netherlands