



Dense and Occlusion-Robust Multi-View Stereo for Unstructured Videos

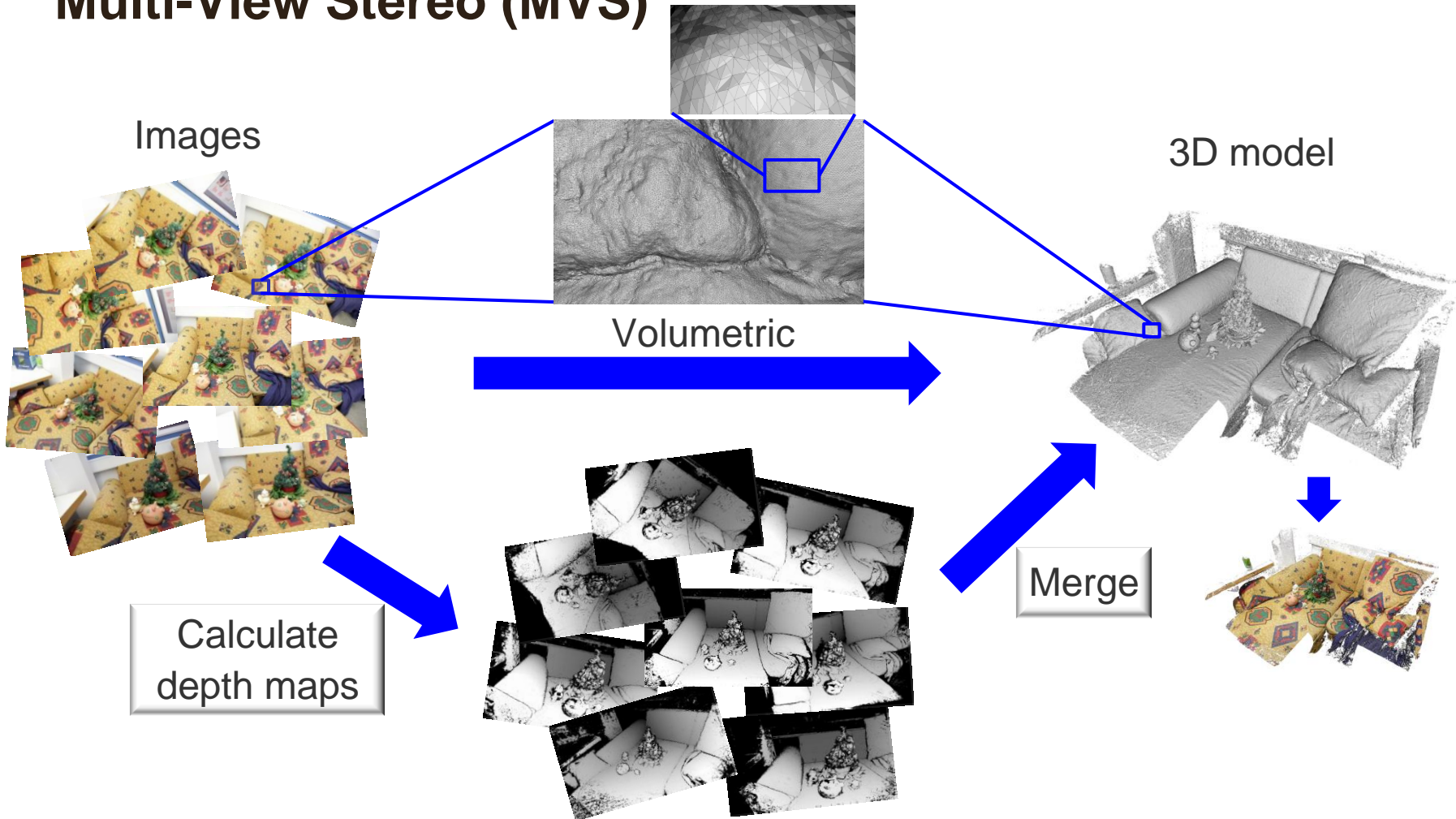
Jian Wei, Benjamin Resch, Hendrik P. A. Lensch

Computer Graphics Group
Tübingen University

CRV 2016



Multi-View Stereo (MVS)





MVS for High-Frame-Rate Videos

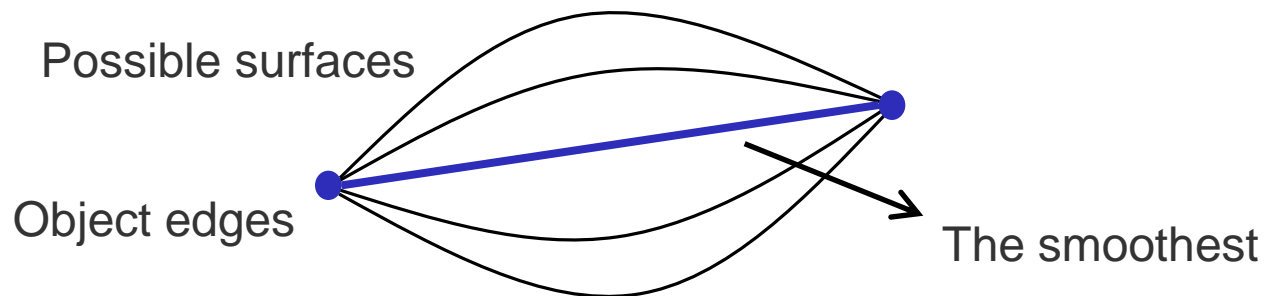
- Our input data
 - Videos:
 - High image resolution
 - High frame rate
 - Unstructured camera trajectories
- Compared with photographs
 - Pros:
 - Easier capturing
 - More information
 - Cons:
 - Smaller baselines
 - Larger amount of data
 - Common problems:
 - Homogeneous surfaces
 - Occlusions





Our Main Idea

- Recovery of homogeneous areas
 - Flat surface assumption:

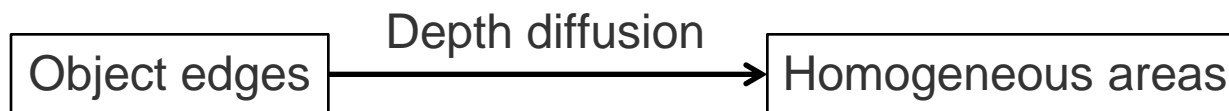




Our Main Idea

- Recovery of homogeneous areas

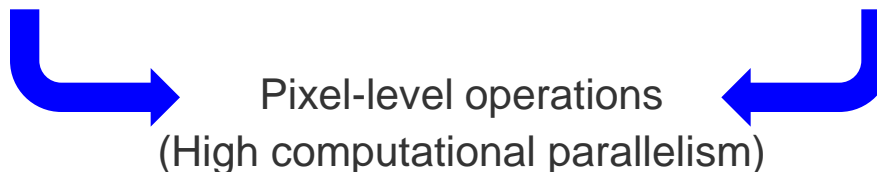
- Solution:



- Requirements:

Sharp discontinuities
Robustness to camera paths

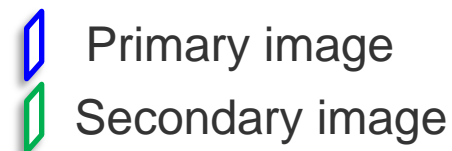
No perspective distortion
Visibility consistency



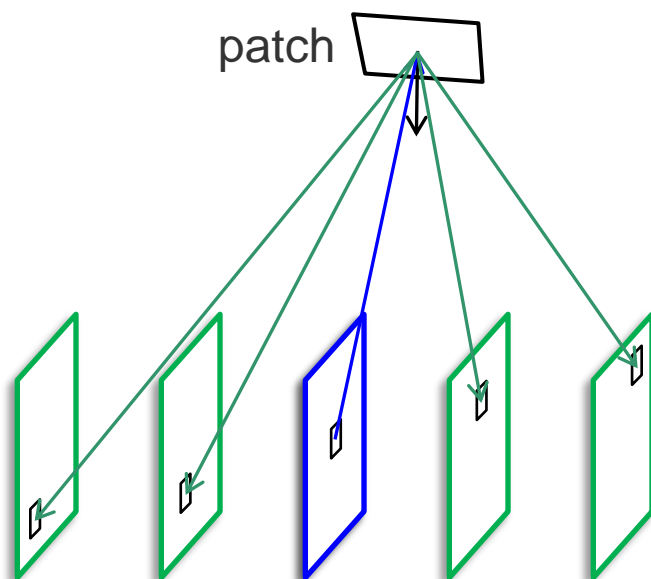
All solved in this paper!



Related Work



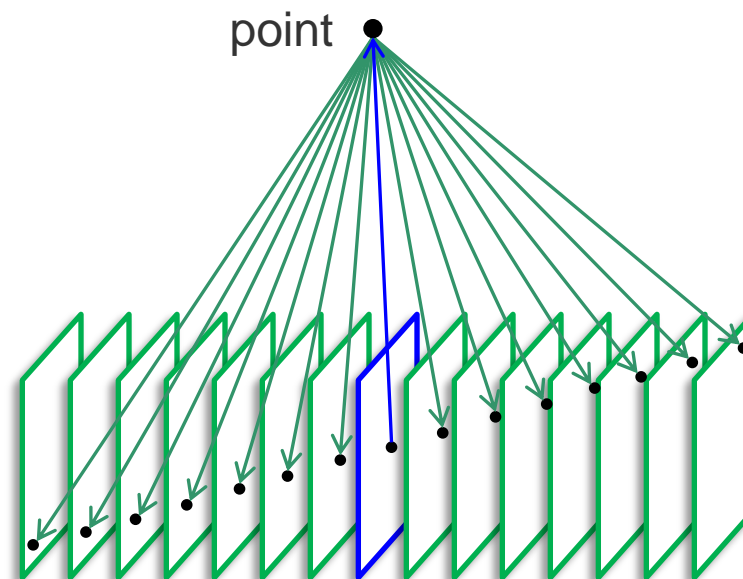
- Textured region reconstruction



- Patch matching

[Bailer *et al.* 2012; Wei *et al.* 2014]

- Surface orientation calculated
- Blur edges



- Ray-based correlation

[Kim *et al.* 2013]

- Sharp edges

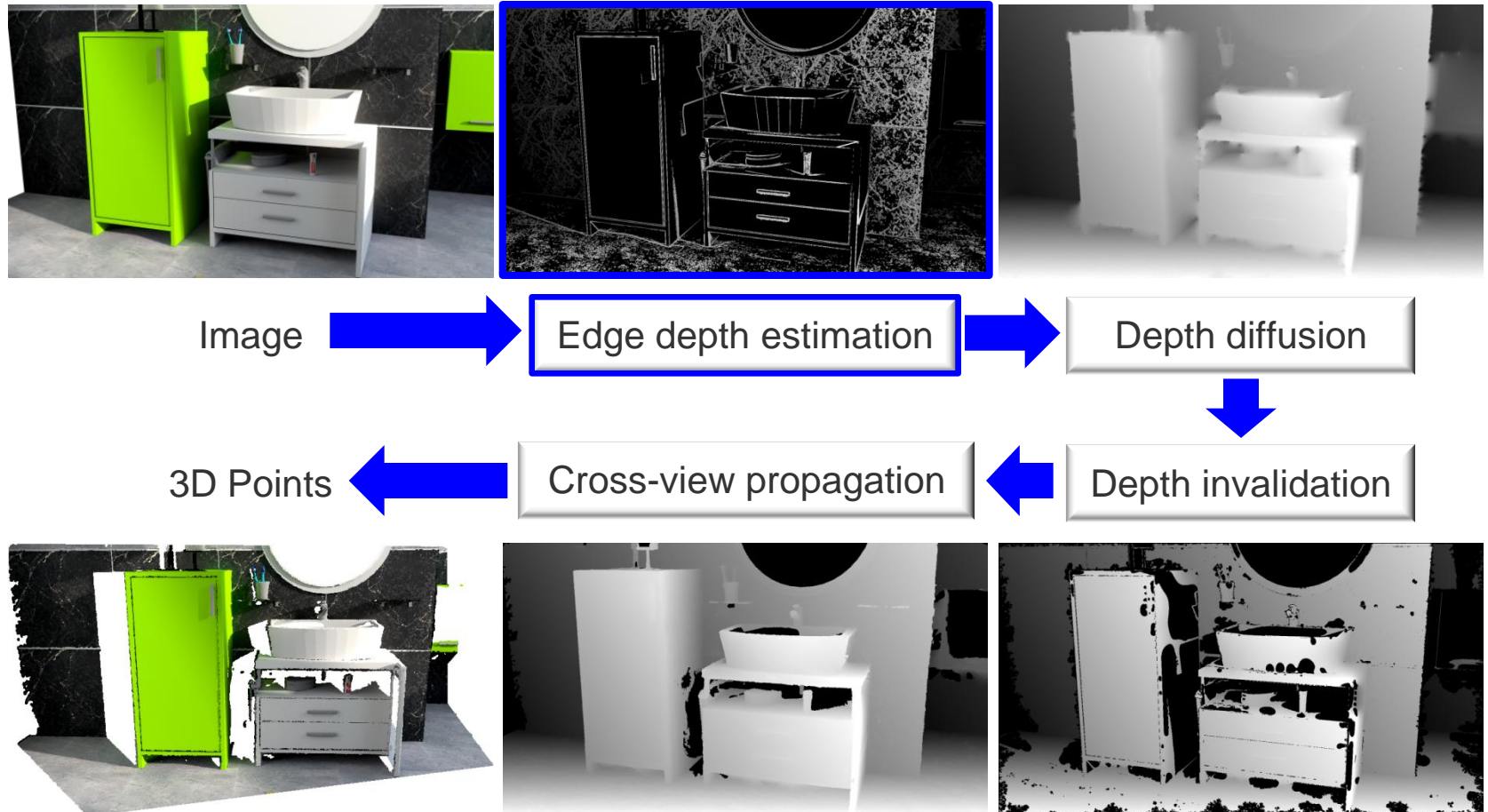


Related Work

- Homogeneous surface recovery
 - Plane segmentation
[Gee *et al.* 2007; Zhang *et al.* 2009; Kundu *et al.* 2014]
 - Multi-resolution framework
[Kim *et al.* 2013; Wei *et al.* 2014; Kang and Medioni 2014]
 - Domain transform filter or diffusion
[Stefanoski *et al.* 2014; Rzeszutek and Androustos 2013, 2015]
- Visibility consistency
 - Careful cross-view depth propagation or filtering
[Kim *et al.* 2013; Wei *et al.* 2014; Rzeszutek and Androustos, 2015]
 - Bundle optimization [Zhang *et al.* 2009]
 - Tetrahedra carving [Hoppe *et al.* 2013]



Pipeline







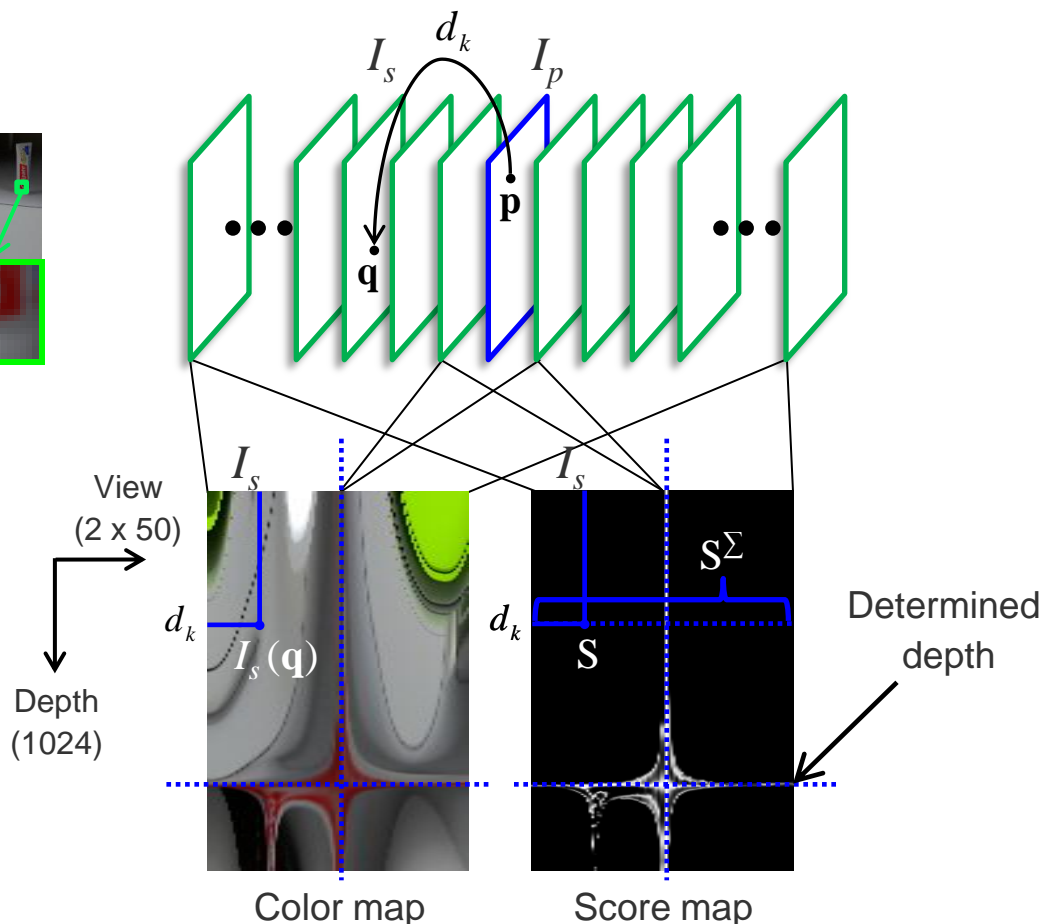
Edge Depth Estimation

- Approach overview
 - Ray-wise depth sweeping [Kim *et al.* 2013]
 - Novelties:
 - Two-scale image selection
 - Dense scale → similar image content → high robustness
 - Sparse scale → large baselines → high depth resolution
 - Robust score aggregation
 - Better handling of large-baseline occlusions / out-of-image projections
 - Pre-processing:
 - Structure from Motion (SfM) [Resch *et al.* 2015]
 - Detect high-gradient pixels: Sobel



Pixel-Wise Color and Score Maps

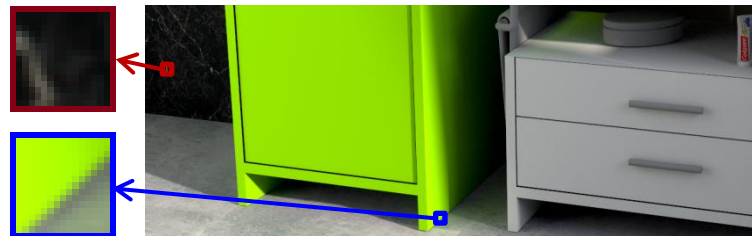
 Primary image
 Secondary image



- Depth score:

$$S = G(I_p(\mathbf{p}) - I_s(\mathbf{q}))$$

Two-Scale Image Selection

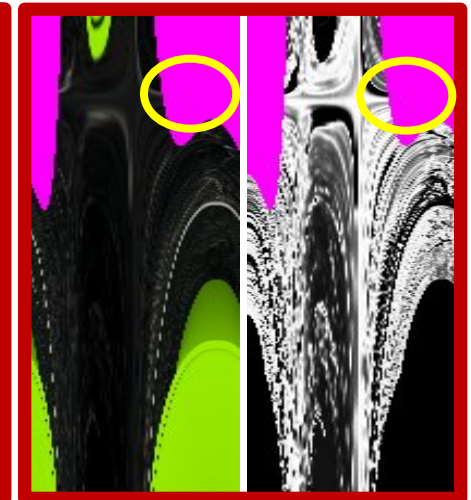
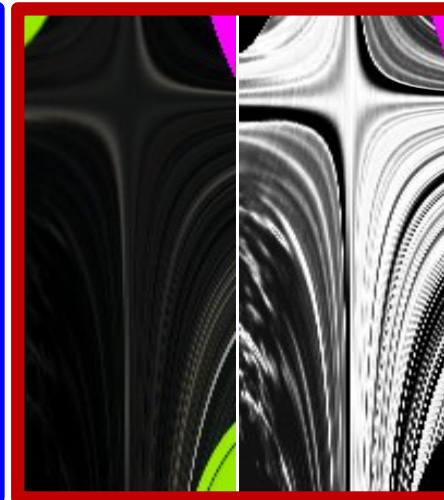
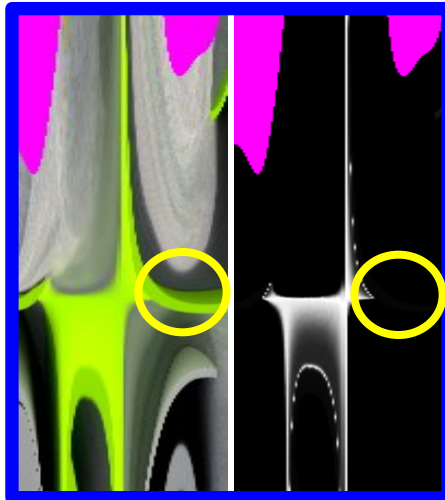
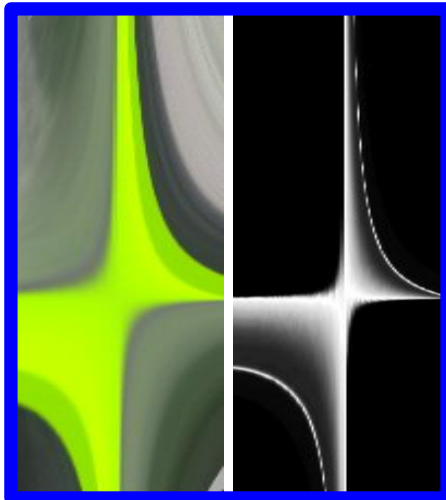


Dense scale

Sparse scale

Dense scale

Sparse scale



Color map

Score map

Color map

Score map

Color map



Score map

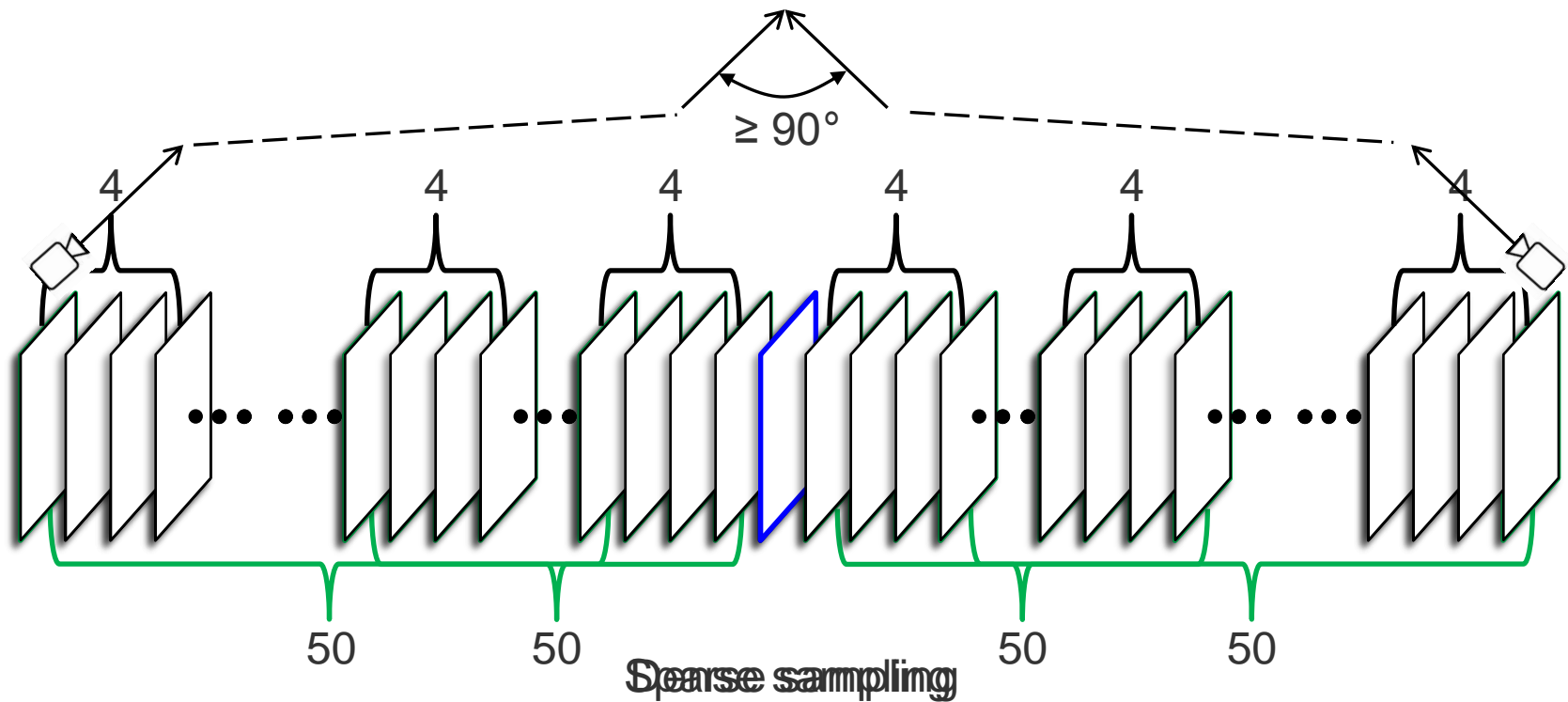
Color map

Score map



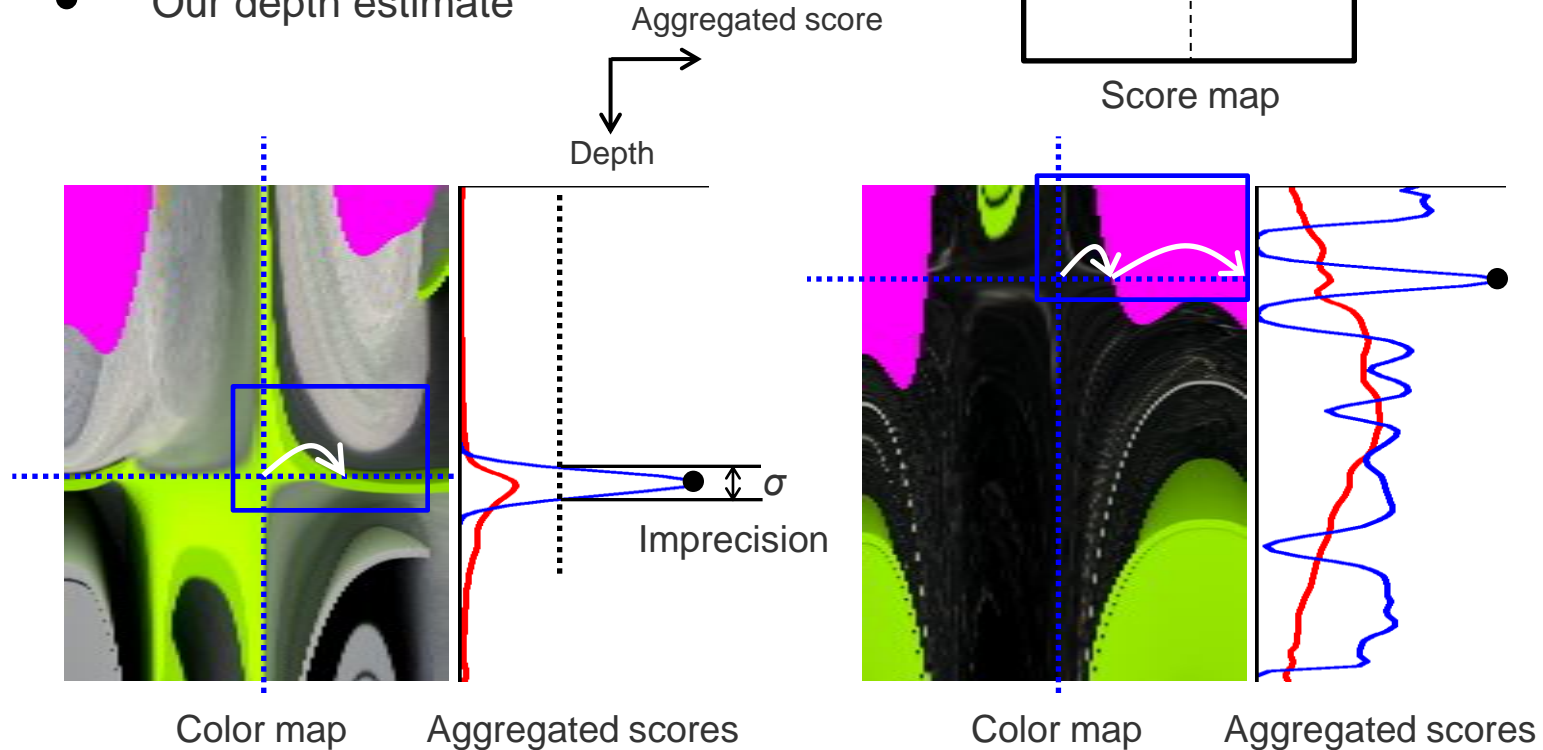
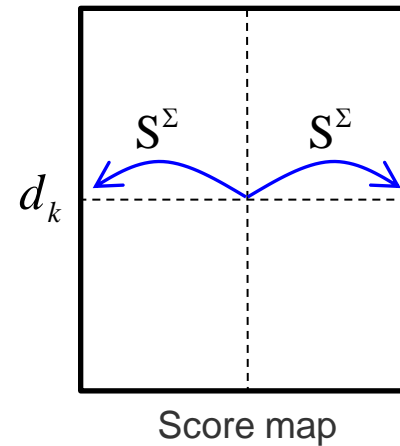
Two-Scale Image Selection

-  Primary image
-  Secondary image



Robust Score Aggregation

- Averaging [Kim *et al.* 2013]
- Our robust aggregation
- Our depth estimate





Pipeline



Image

Edge depth estimation

Depth diffusion

3D Points

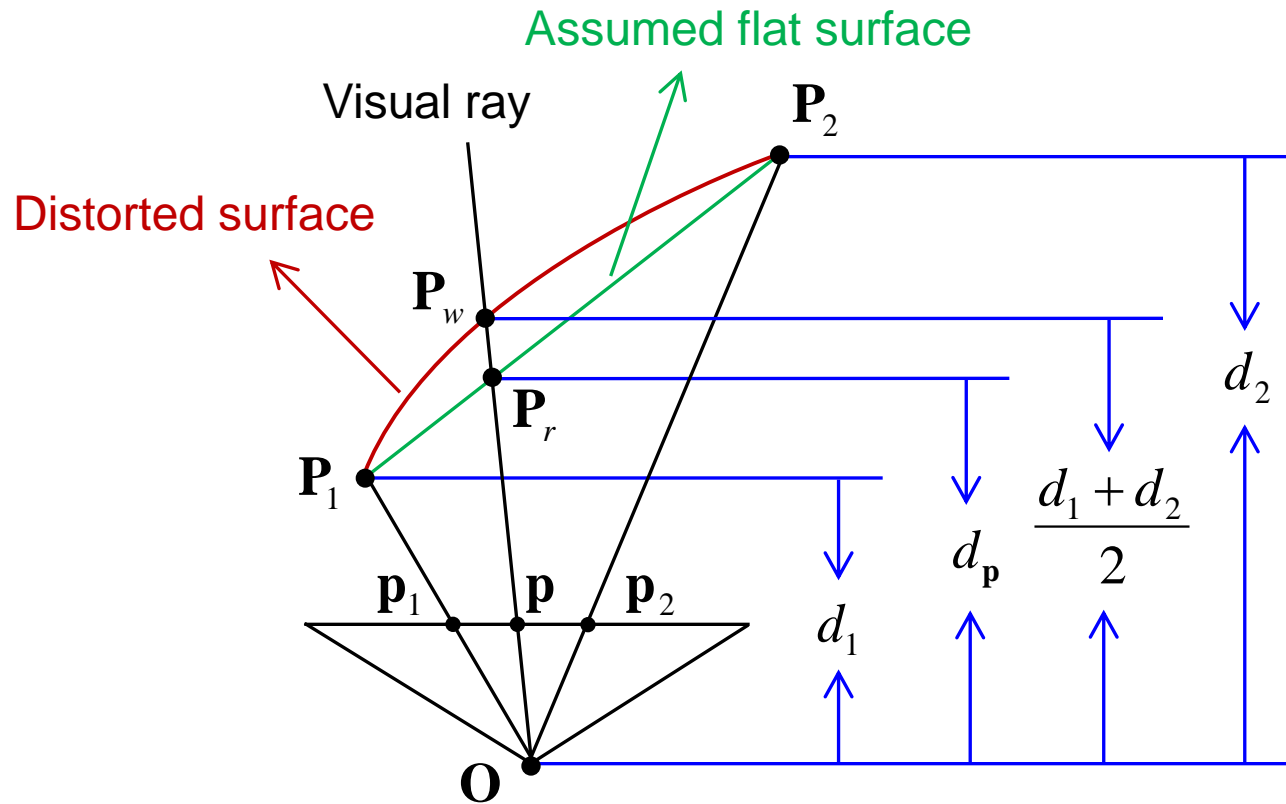
Cross-view propagation

Depth invalidation





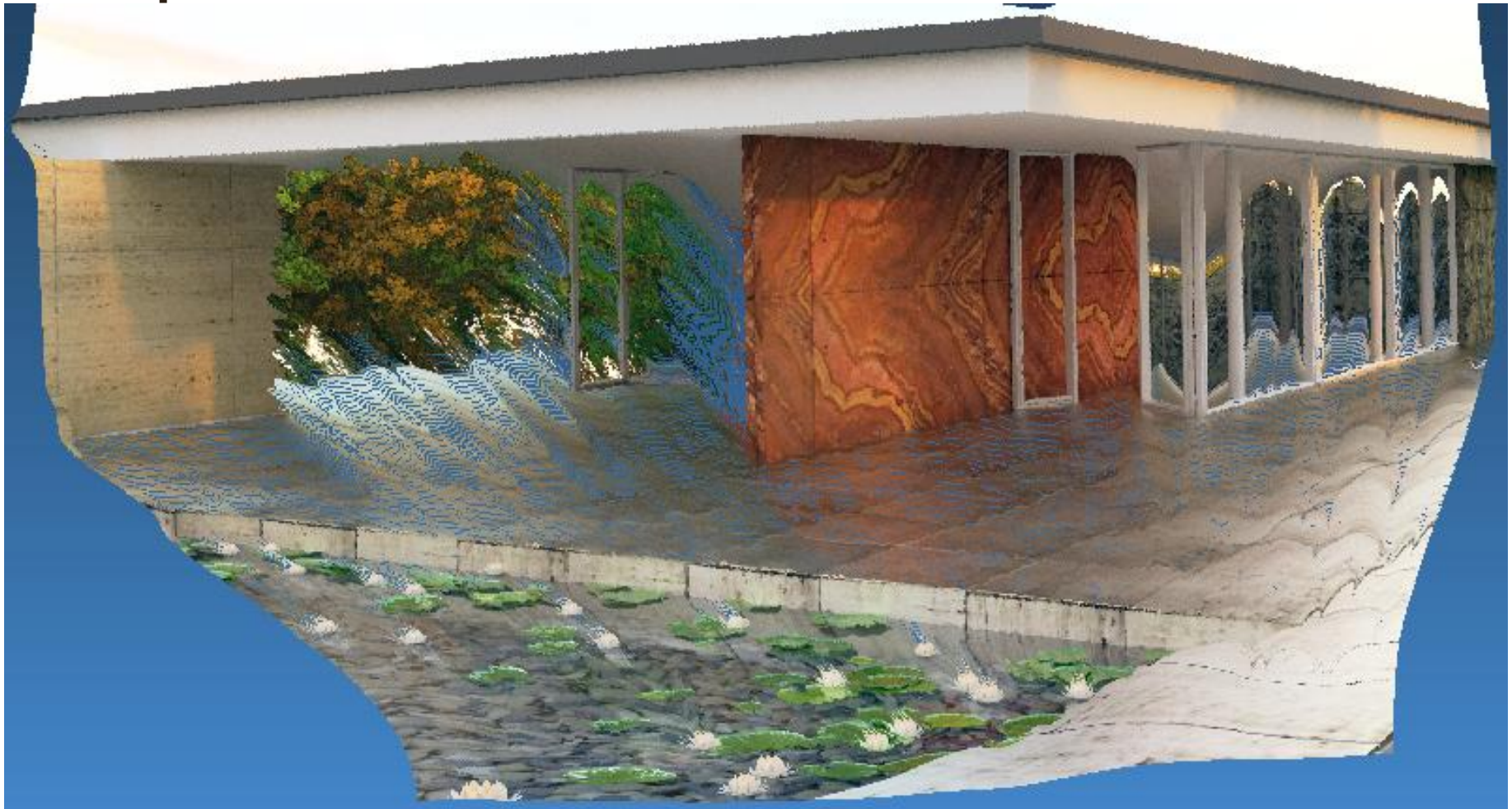
Depth Diffusion





Depth Diffusion

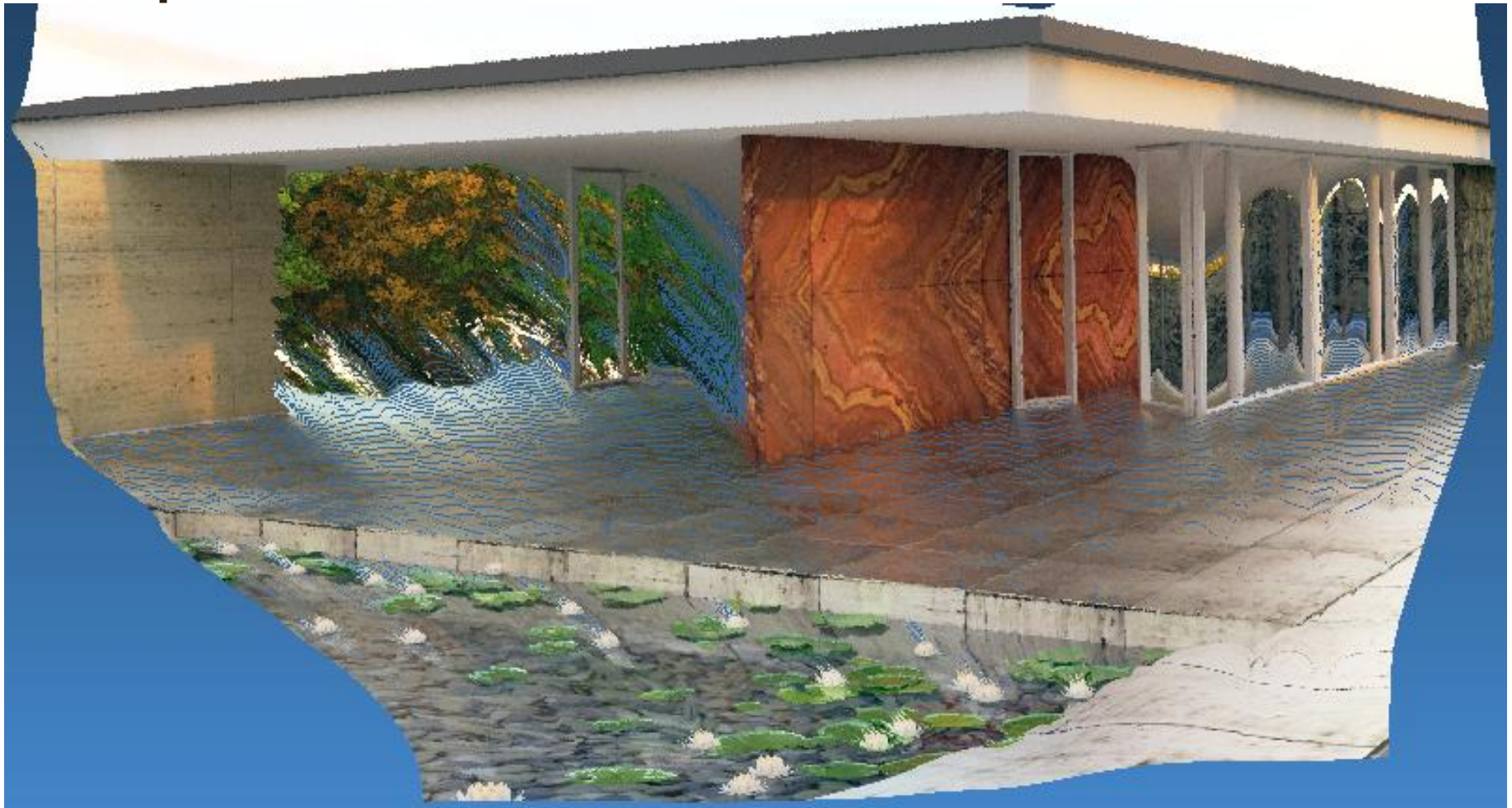
Classic diffusion





Depth Diffusion

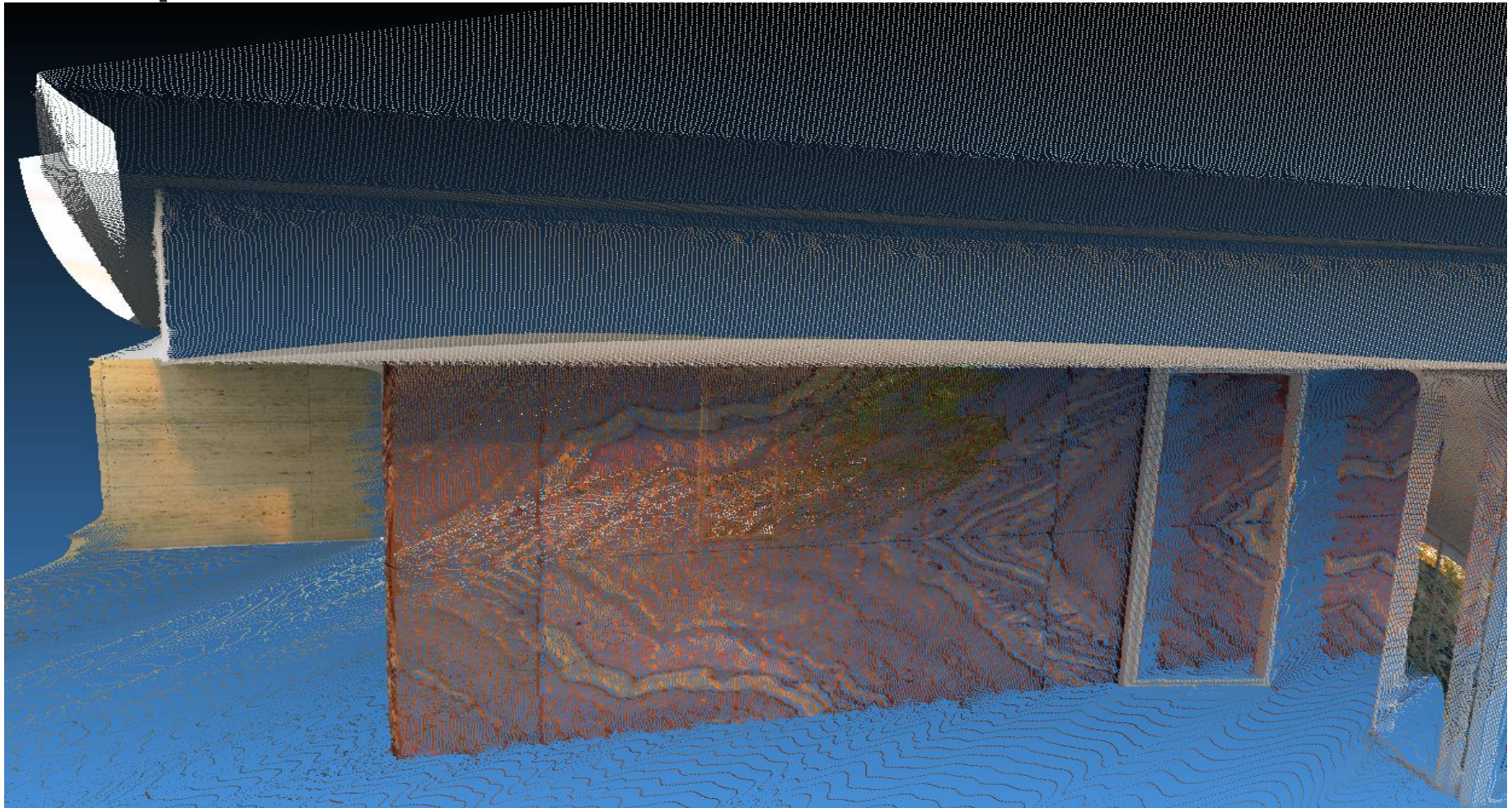
Perspective diffusion





Depth Diffusion

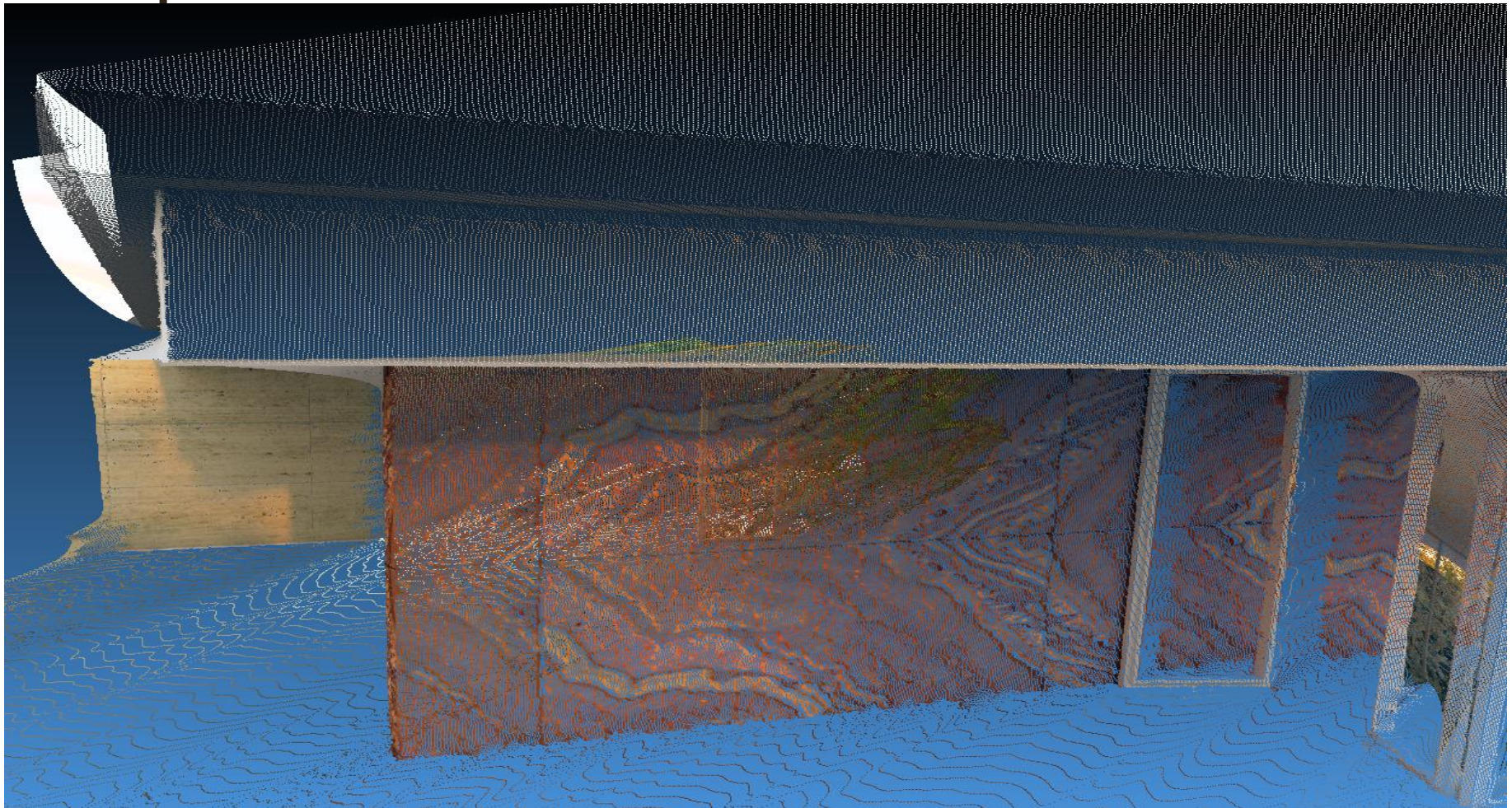
Classic diffusion





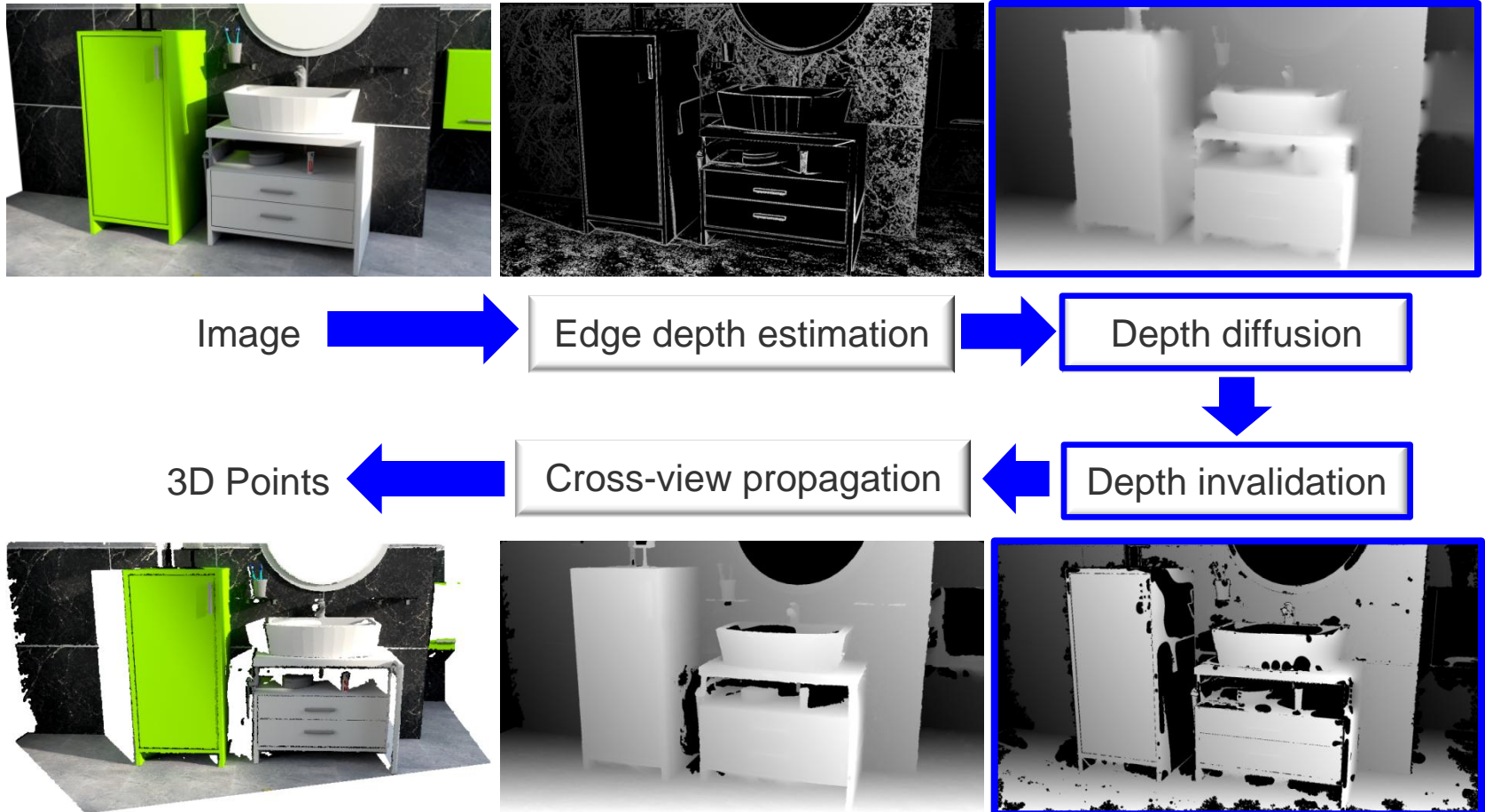
Depth Diffusion

Perspective diffusion





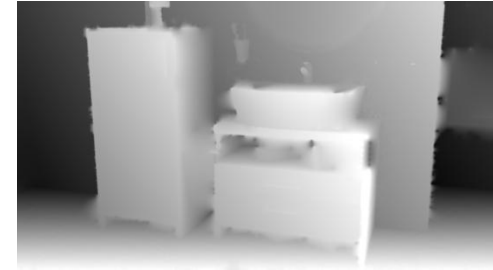
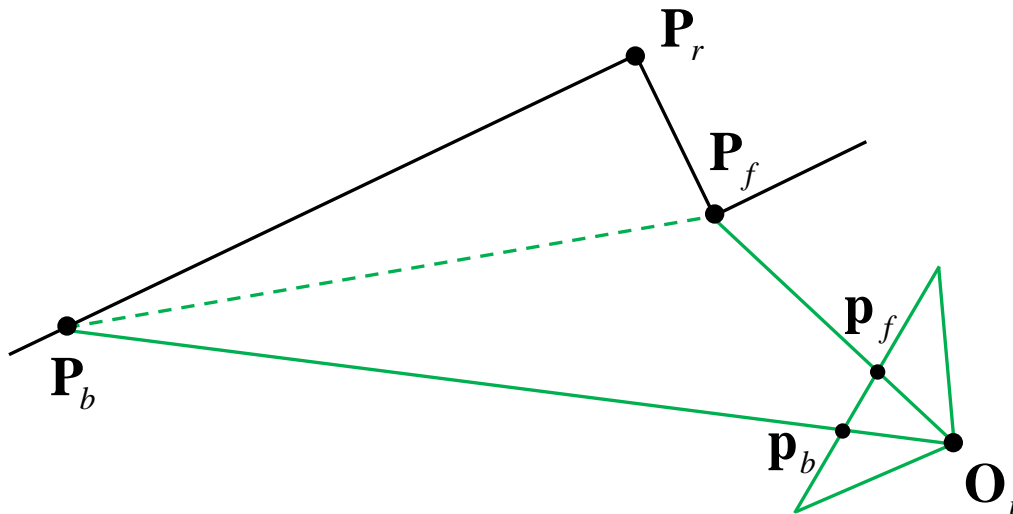
Pipeline





Depth Invalidation

- Occlusion problem



Depth map



Front view

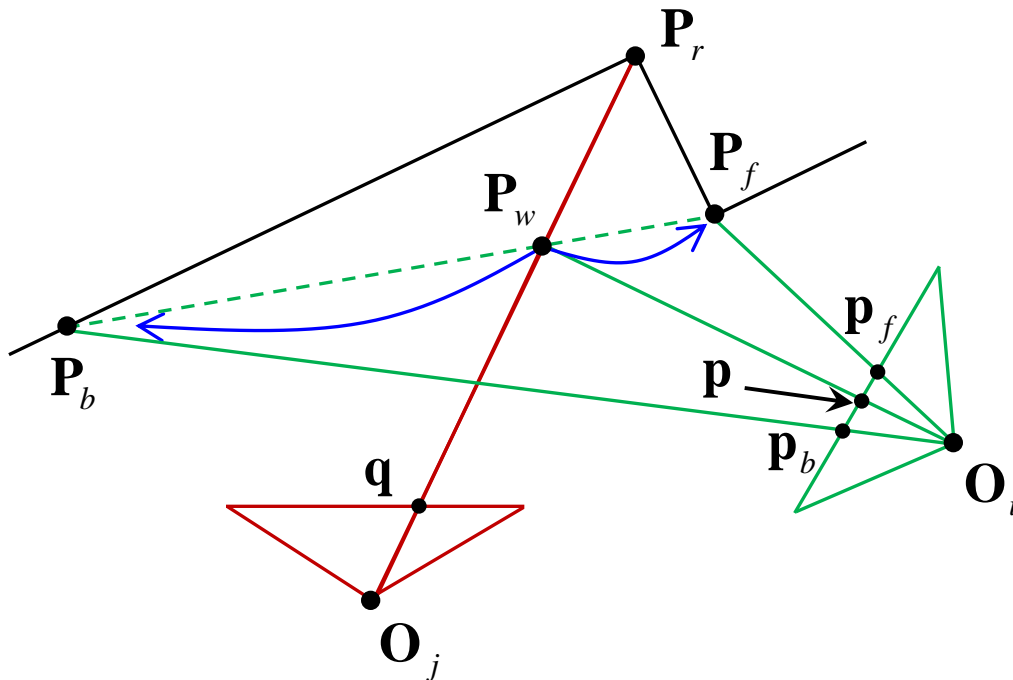


Top view



Depth Invalidation

- Our solution
 - Region growing



Sparse errors



Grow



Dense errors

Remove all large errors!



Depth Invalidation

- Invalid area growing

Local error map



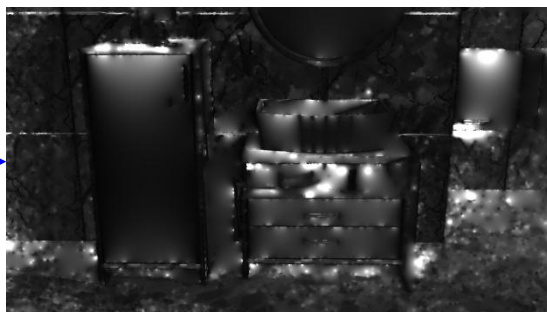
Combination



Diffused combination
(Approximated SD map)



Edge SD map



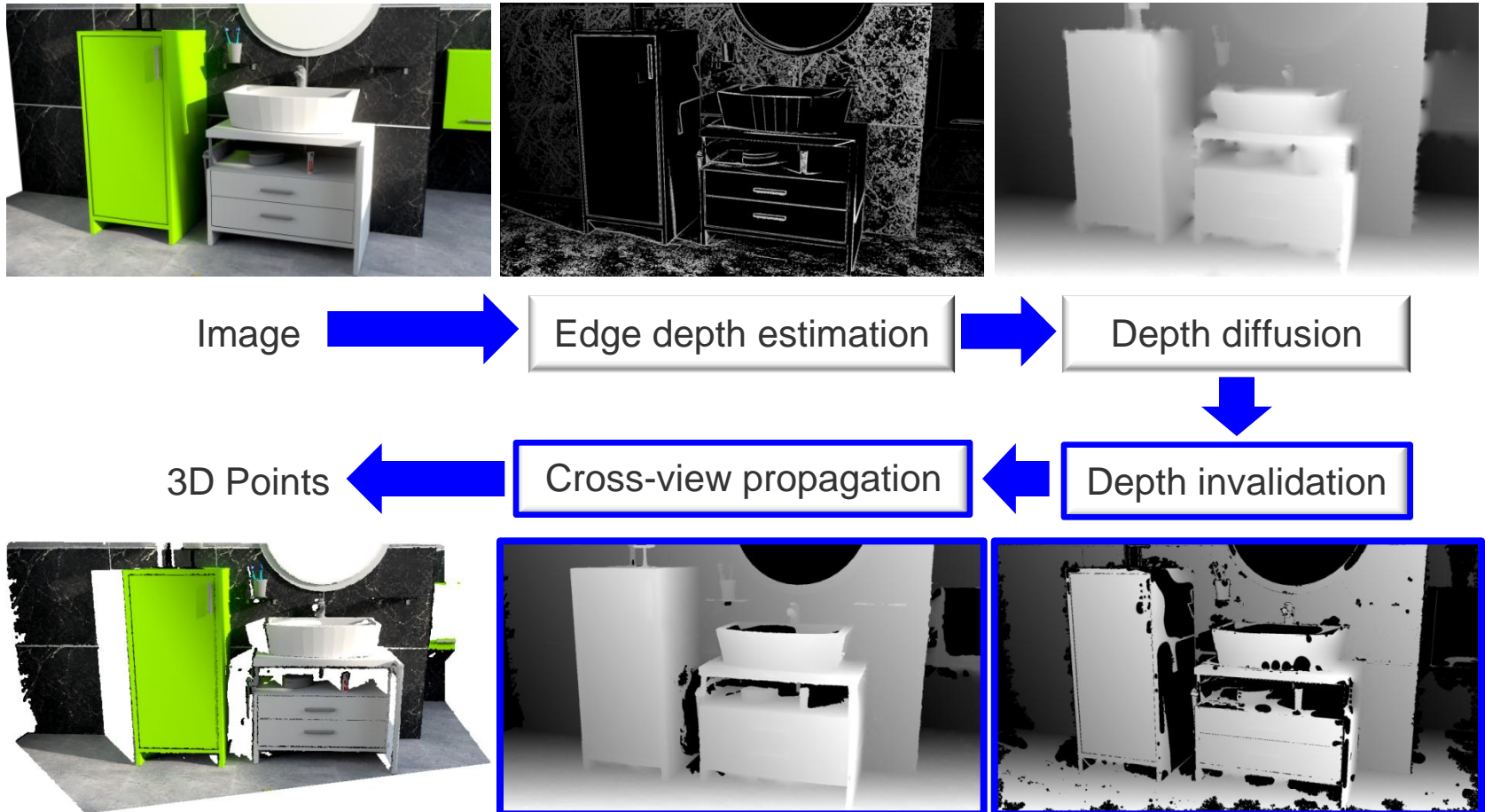
Diffused SD map (expected)



Invalidity mask

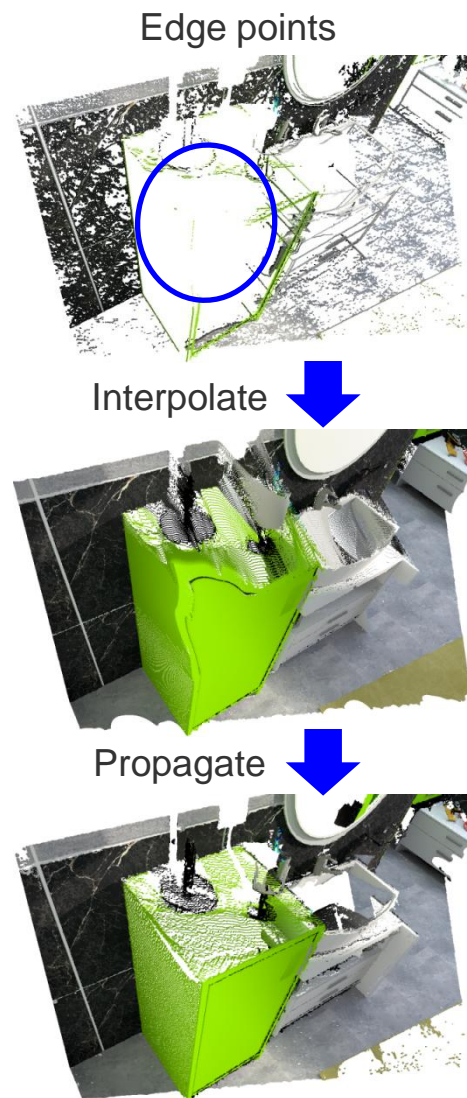
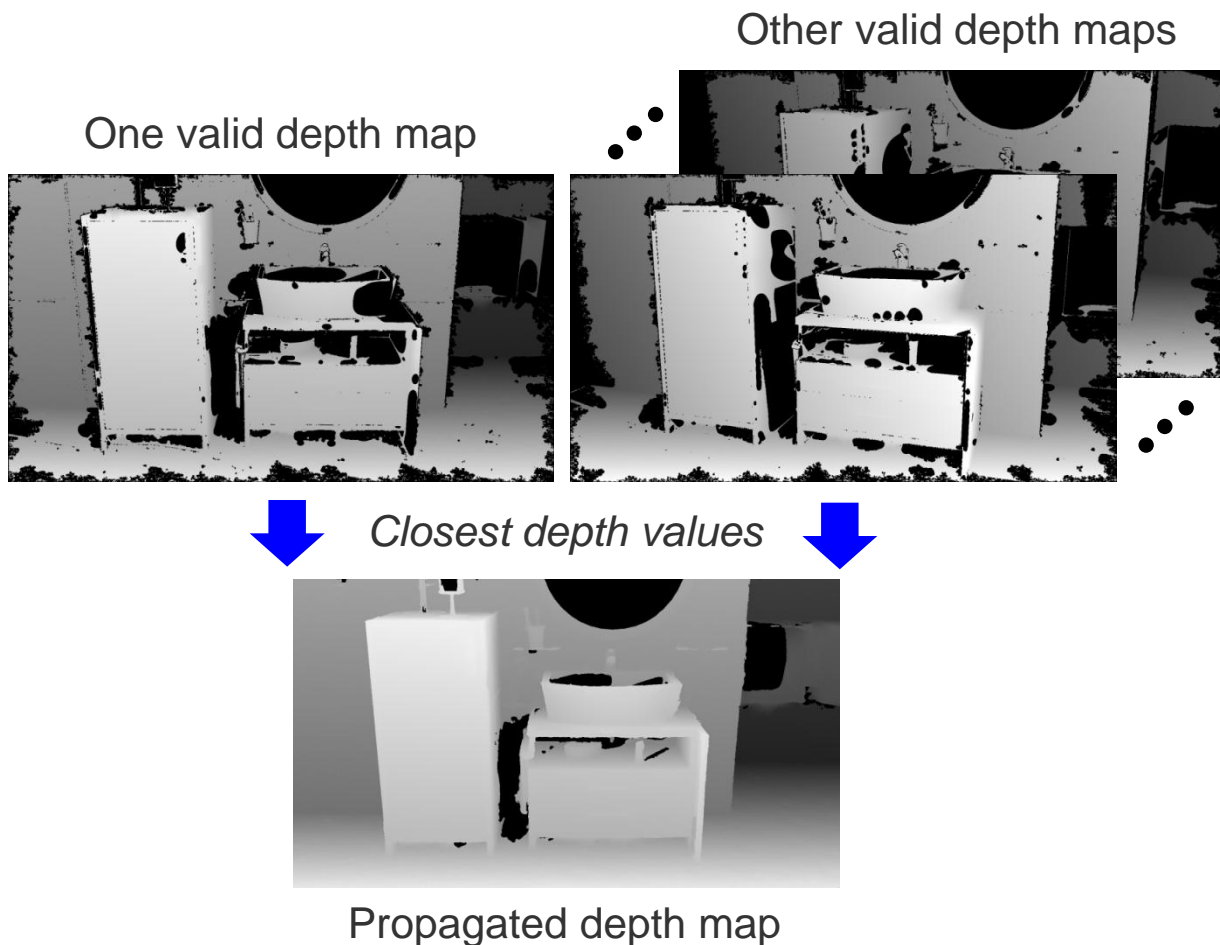


Pipeline





Cross-View Propagation





Evaluation

- Datasets
 - Arbitrary camera trajectories
 - Large homogeneous surfaces + occlusions
 - Resolution: 1920 x 1080



Bathroom

Pabellon

Boxes

Building

Synthetic
(with ground truth)

Real-world
(no ground truth)



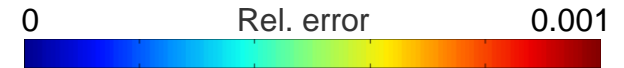
Evaluation

- Compared algorithms
 - [Bailer *et al.* 2012]
 - [Wei *et al.* 2014] (Our previous work)
 - [Kim *et al.* 2013]
 - Our work without depth propagation (w/o DP)
- Output
 - Create 100 depth maps

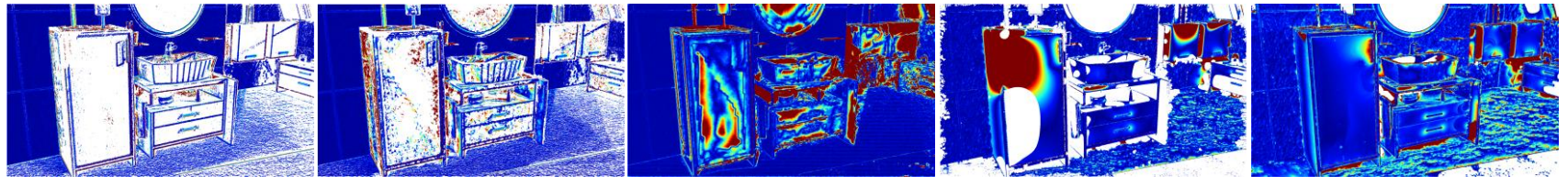
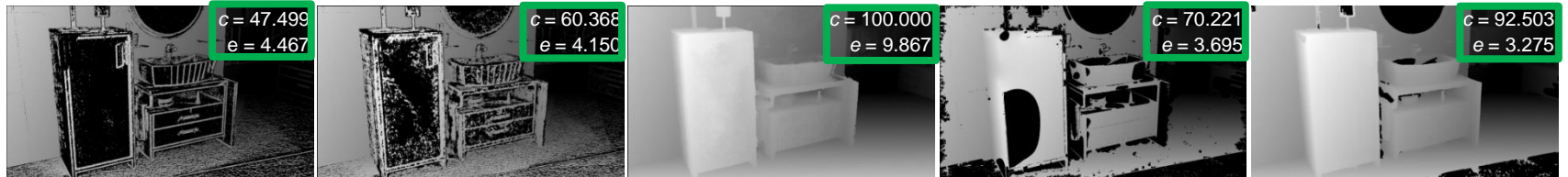
Synthetic Scenes

- Depth and relative error maps

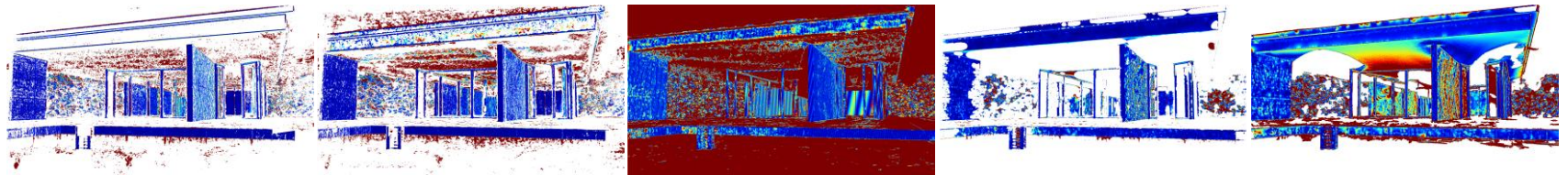
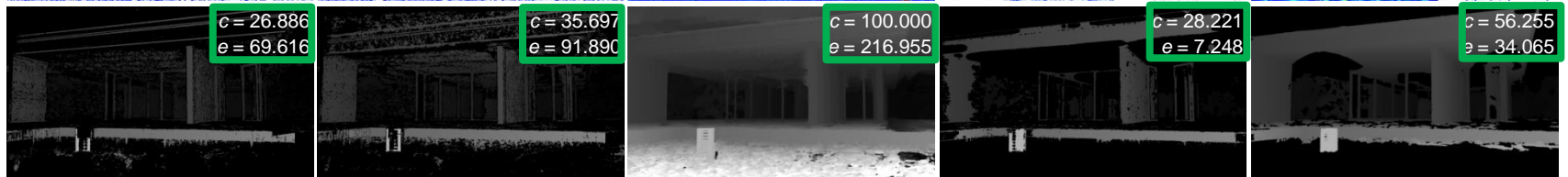
Completeness (%): $c \uparrow$
Mean rel. error ($\times 10^{-3}$): $e \downarrow$



Bathroom



Pabellon



Bailer et al.

Wei et al.

Kim et al.

Ours w/o DP

Ours



Realworld Scenes

Boxes



Building



Wei et al.

Kim et al.

Ours



Runtime

- Implementation platform
 - Kim *et al.*: NVIDIA GeForce GTX 680 GPU
 - Others: NVIDIA GeForce GTX Titan GPU

Runtime (sec.) for calculating one depth map

Method	Bathroom	Pabellon	Boxes	Building
Bailer <i>et al.</i>	68.12	63.28	71.29	75.92
Wei <i>et al.</i>	34.02	35.14	32.84	35.30
Kim <i>et al.</i> (256)	33.72	33.27	31.74	35.77
Kim <i>et al.</i> (1024)	121.55	117.12	119.76	125.39
Ours	10.21	12.53	11.67	11.07



Conclusion

- Efficient and dense recovery of textureless surfaces from videos
 - Two-scale image selection
 - Robustness to camera trajectories
 - Region-growing-based invalidation
 - Pixel-level computations
- Future work
 - Merging of redundant per-view results
 - Extension to large scale



Thank you.