Microsoft

Microsoft Research Faculty Summit 2013

Microsoft

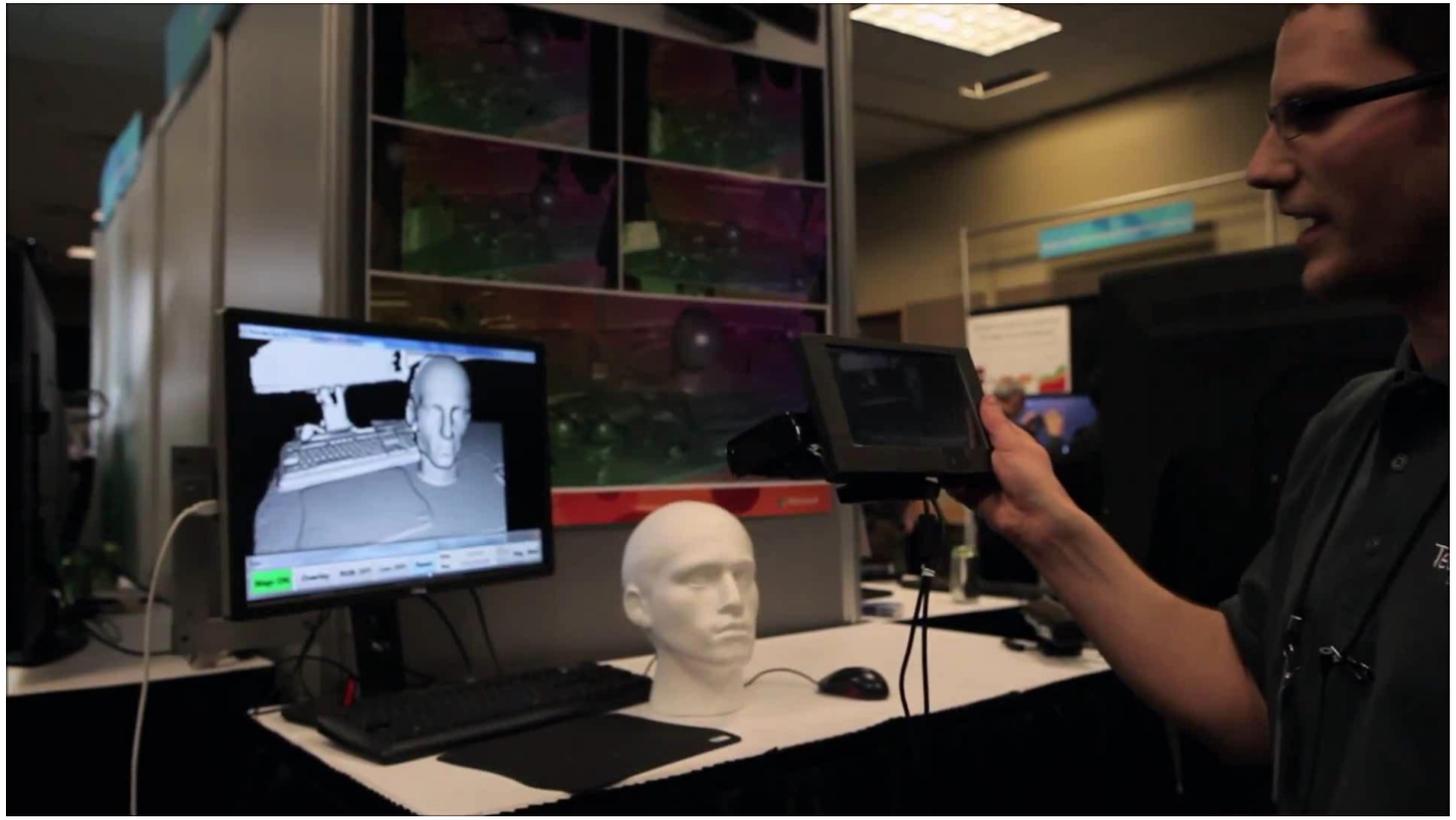
City-scale Kinect Fusion

Andrew Blake
Laboratory Director,
Microsoft Research, Cambridge, UK

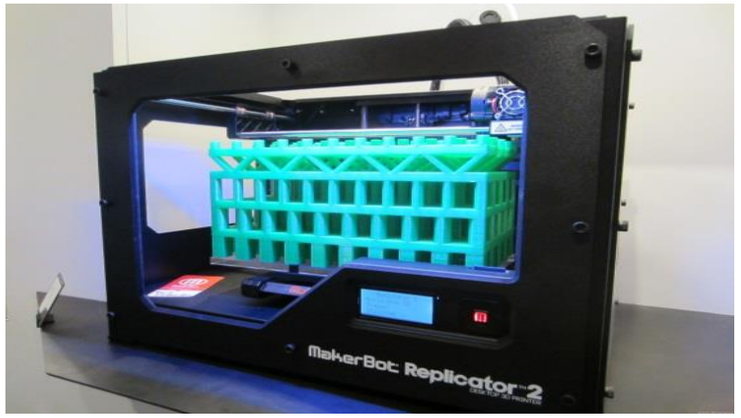
on behalf of:
I3D group, MSR
– led by Shahram Izadi.

Microsoft Research
Faculty Summit

Capturing Surfaces



Augmented Reality (e.g. medical)



3D copier

Noninvasive measurement
(c.f. FuSci)

3D Reconstruction Technology

Hernández, C. and Cipolla, R. “Digital Pygmalion” , 2005

Levoy, M. et al. “Digital Michelangelo”, SIGGRAPH’00

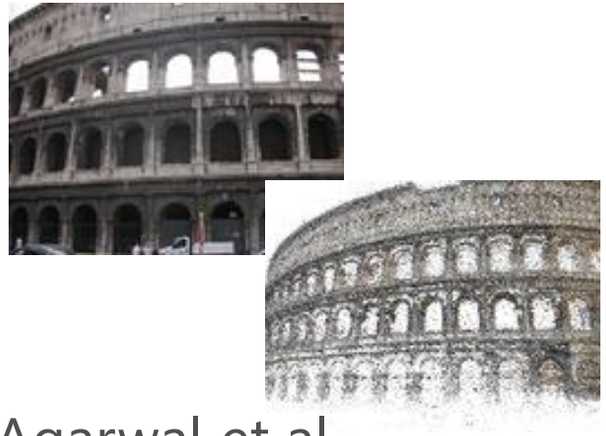
Offline 3D reconstruction

- Multi-view stereo
  - online photos
- Light stages

- Dense reconstruction
- High quality
- Non real-time

Debevec et al., ICT Lab

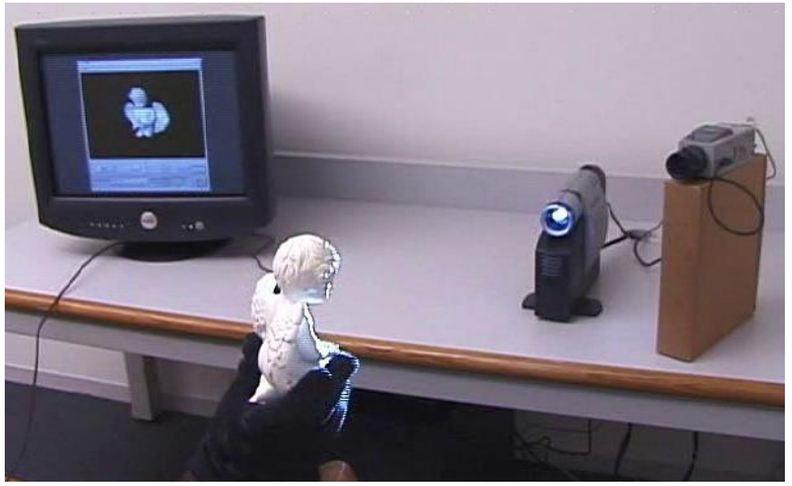
Frahm et al.
Building Rome on a cloudless day
ECCV'10



Agarwal et al.
Building Rome in a Day
ICCV'09

Small object scanners

- Custom active sensors
- Interactive rates



Rusinkiewicz et al. SIGGRAPH'02

Z Corporation ZScanner

- Point-based
- Small scenes
- Non-commodity hardware

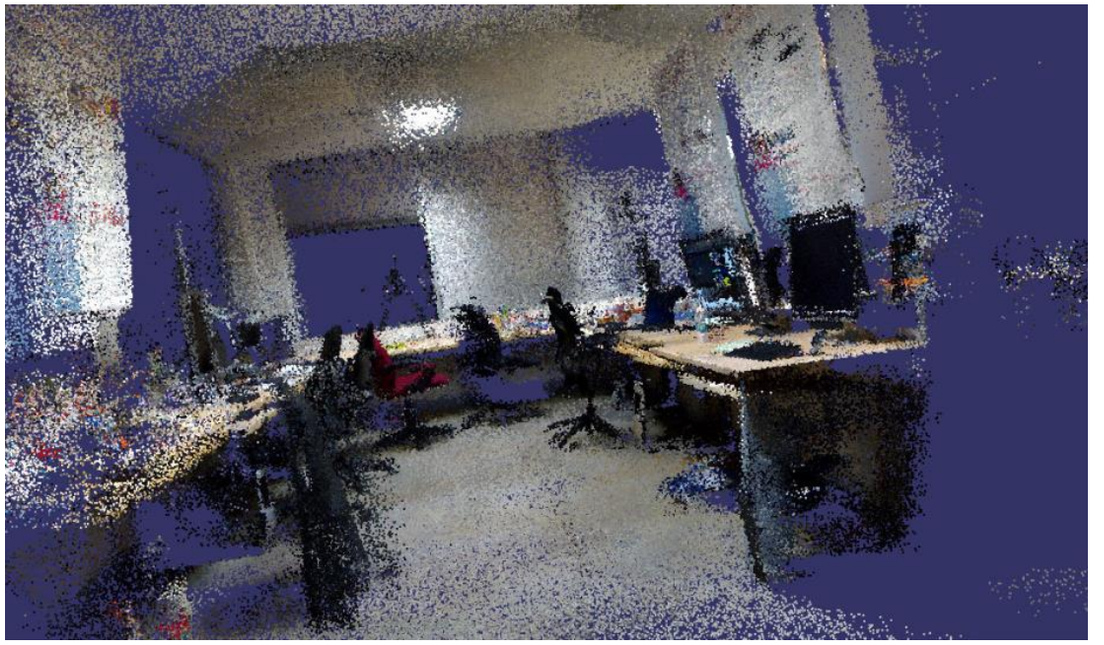
Weise et al., PhD Thesis, ETHZ 2009

Input **Result**

Cui et al. CVPR 2010

Monocular SLAM

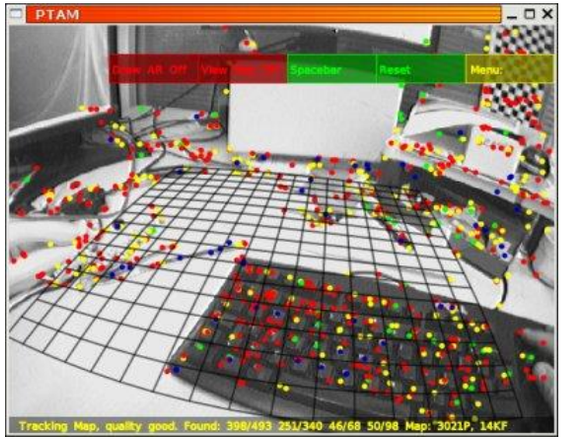
- Single handheld camera
- Markerless
- Simultaneous mapping & reconstruction
- Office sized scenes



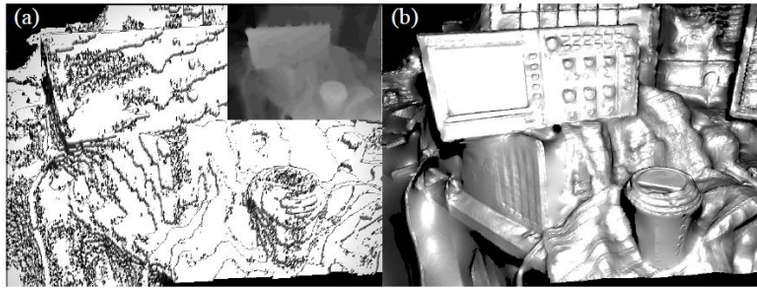
Henry et al. ISER 2010

Want:

- Dense reconstructed surfaces
- Real-time



PTAM: Klein & Murray, ISMAR 2007

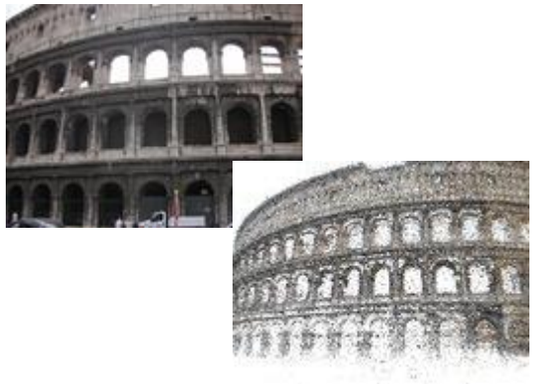


DTAM: Newcombe et al. ICCV 2011

Scale vs. Quality vs. Speed

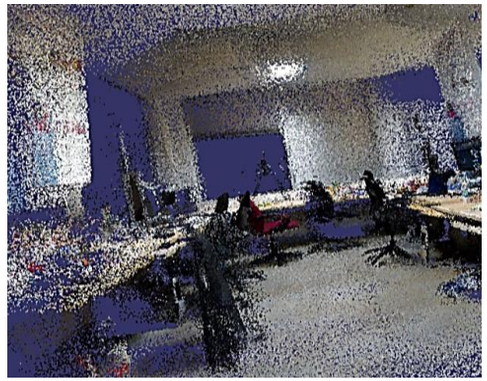
Small Scale
Real-time
High Quality

E.g. Weise et al., PhD Thesis, ETHZ 2009



Large Scale
Offline
High quality

E.g. Agarwal et al. Building Rome in a Day ICCV'09



Larger Scale
Interactive(~10Hz)
Lower quality

E.g. Henry et al. ISER 2010

Large Scale
Real-time(~30Hz)
Lower quality

E.g. Nister, Pollefeys et al. UrbanScape, 2007

?

Large Scale
Real-time
High Quality

Kinect Fusion Pipeline

Input

Raw Depth Image

Compute 3D Point Cloud and Normals

Camera Pose Estimation

Volumetric Fusion

Surface Extraction

Kinect Fusion Pipeline

Input

Raw Depth Image

Compute 3D Point Cloud and Normals

Camera Pose Estimation

Volumetric Fusion

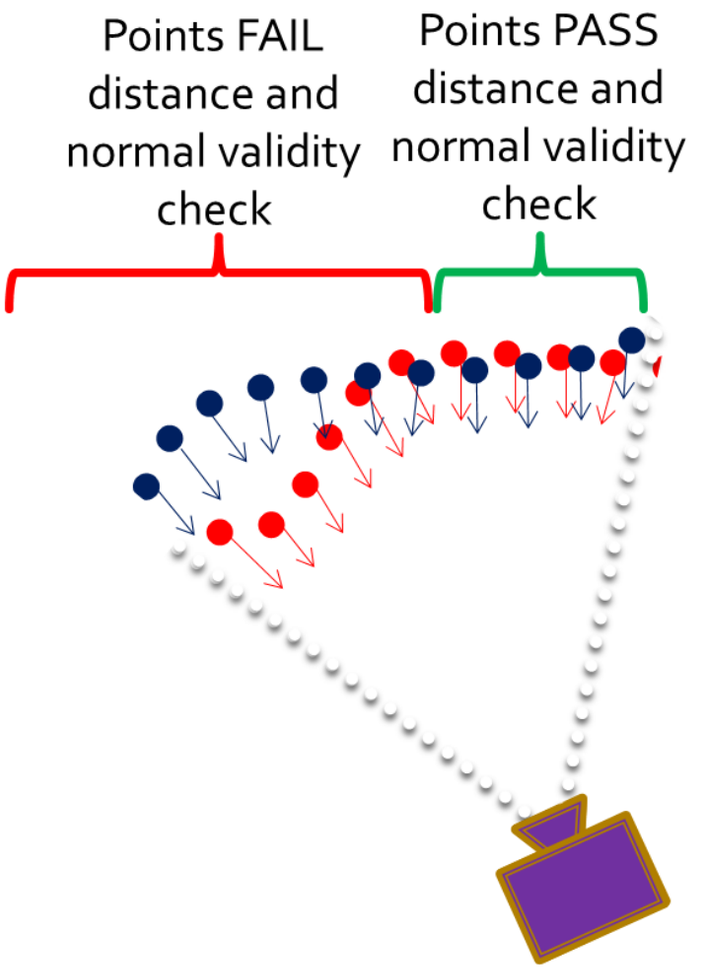
Surface Extraction

Iterated Closest Point

- Transformation T aligns points V_t , normals N_t with (V_{t-1}, N_{t-1})
- Solve for data association
- Point-plane error metric:

$$\arg\min_{T} \sum_{x,y} [\, T\, v_t(x,y) - v_{t-1}(x,y))\ \cdot n_{t-1}(x,y)]^2$$

[Chen and Medioni '91]



Kinect Fusion Pipeline

Input

Raw Depth Image

Compute 3D Point Cloud and Normals

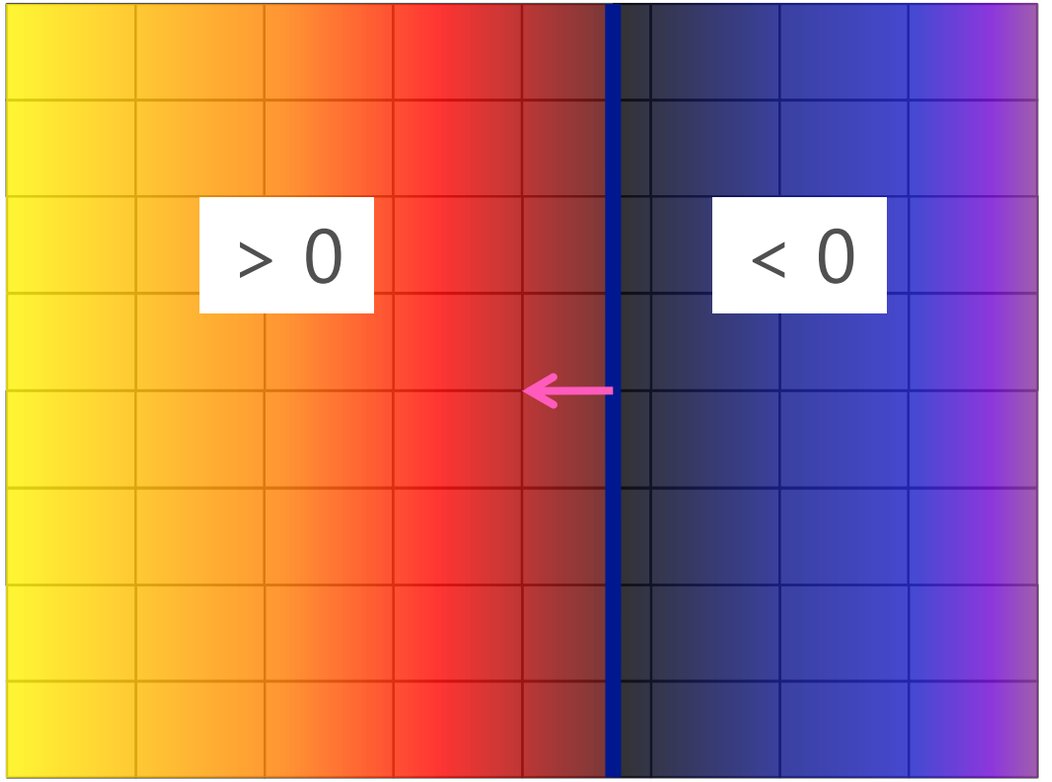
Camera Pose Estimation

Volumetric Fusion

Surface Extraction

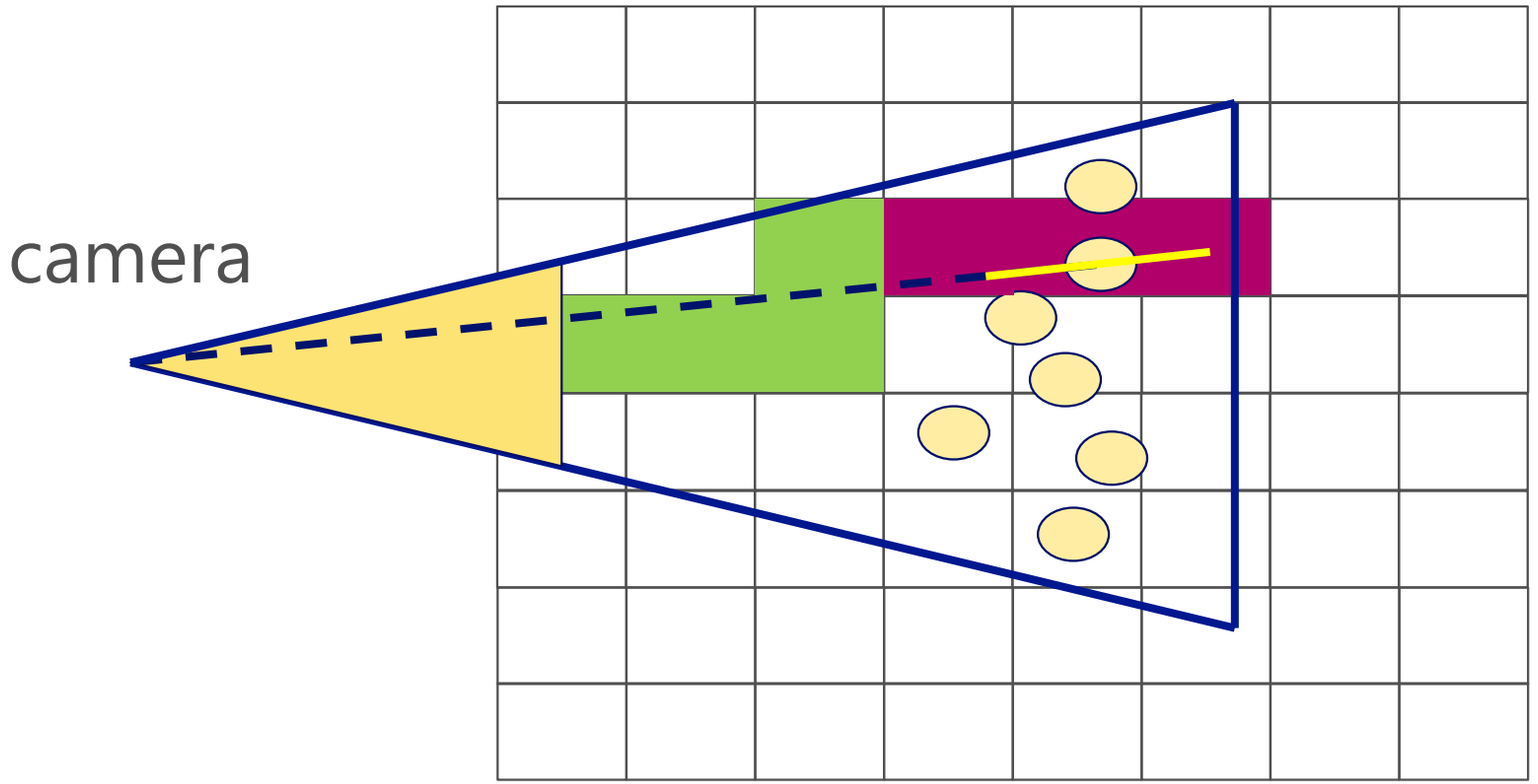
Signed distance map

[Curless & Levoy, SIGGRAPH'96]

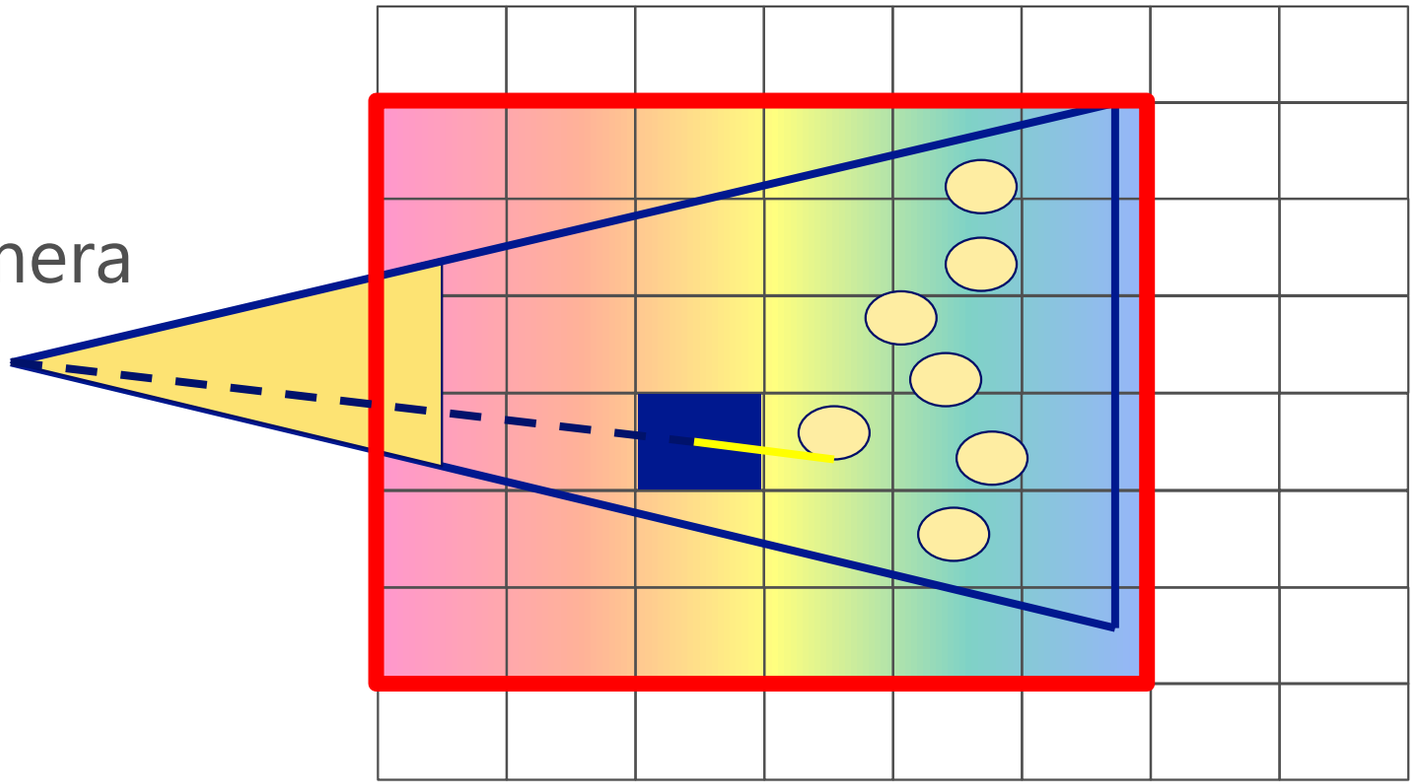


Each voxel stores **signed** distance D to surface
surface is at the **zero crossing**
surface normal is the **gradient** $\nabla D / \|\nabla D\|$

SDF banding and free-space carving



SDF depth map fusion



within frustum bounding box ….

sweep voxels *v* in parallel - GPU

project each voxel center onto screen

compute signed distance *d* to nearby feature

Weighted update of signed distance fn

Kinect Fusion Pipeline

Input

Raw Depth Image

Compute 3D Point Cloud and Normals

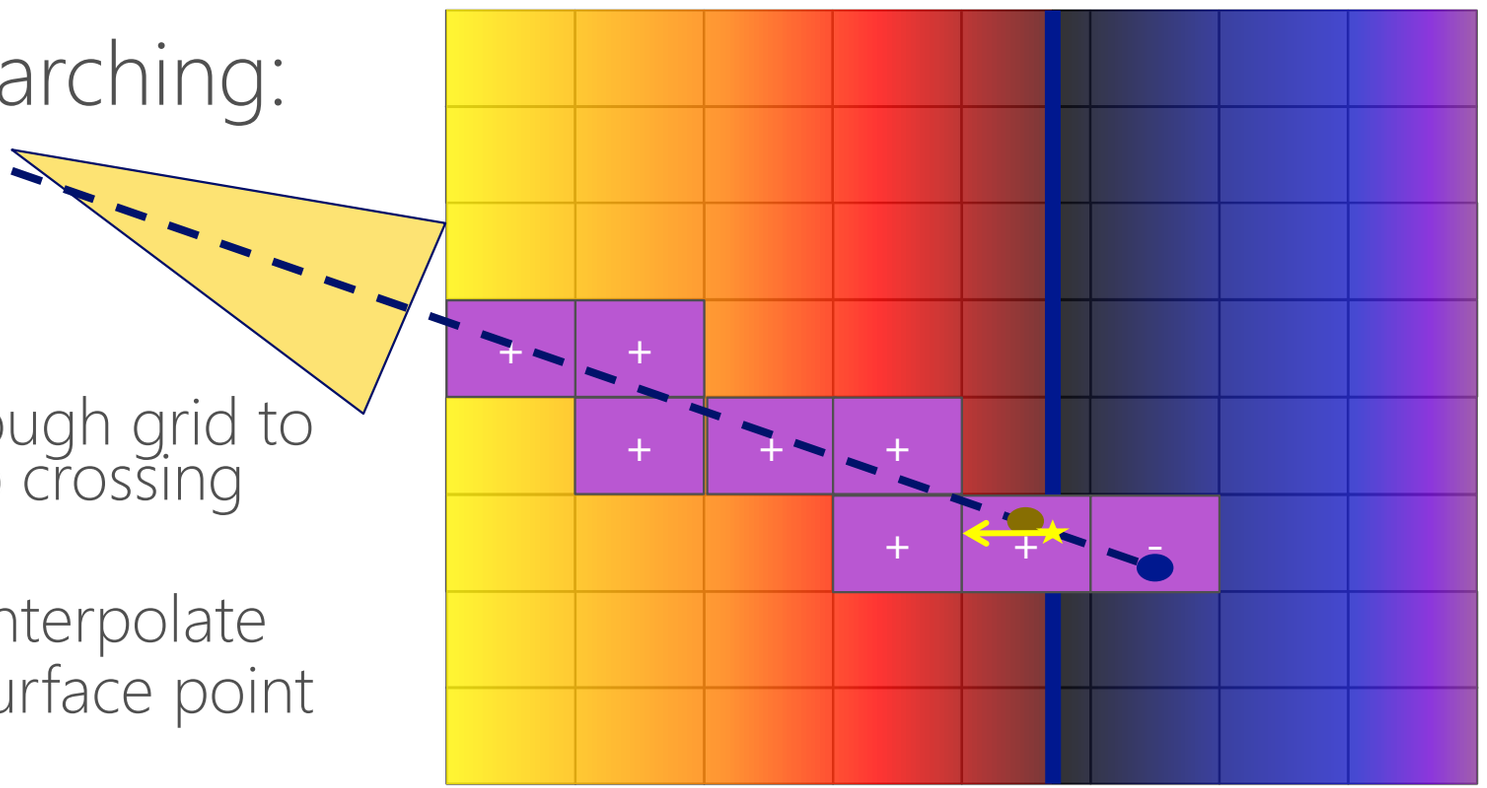
Camera Pose Estimation

Volumetric Fusion

Surface Extraction

Surface extraction

Ray marching:



step through grid to find zero crossing

linearly interpolate to find surface point

filter gradient to compute surface normal

Kinect Fusion Pipeline

Input

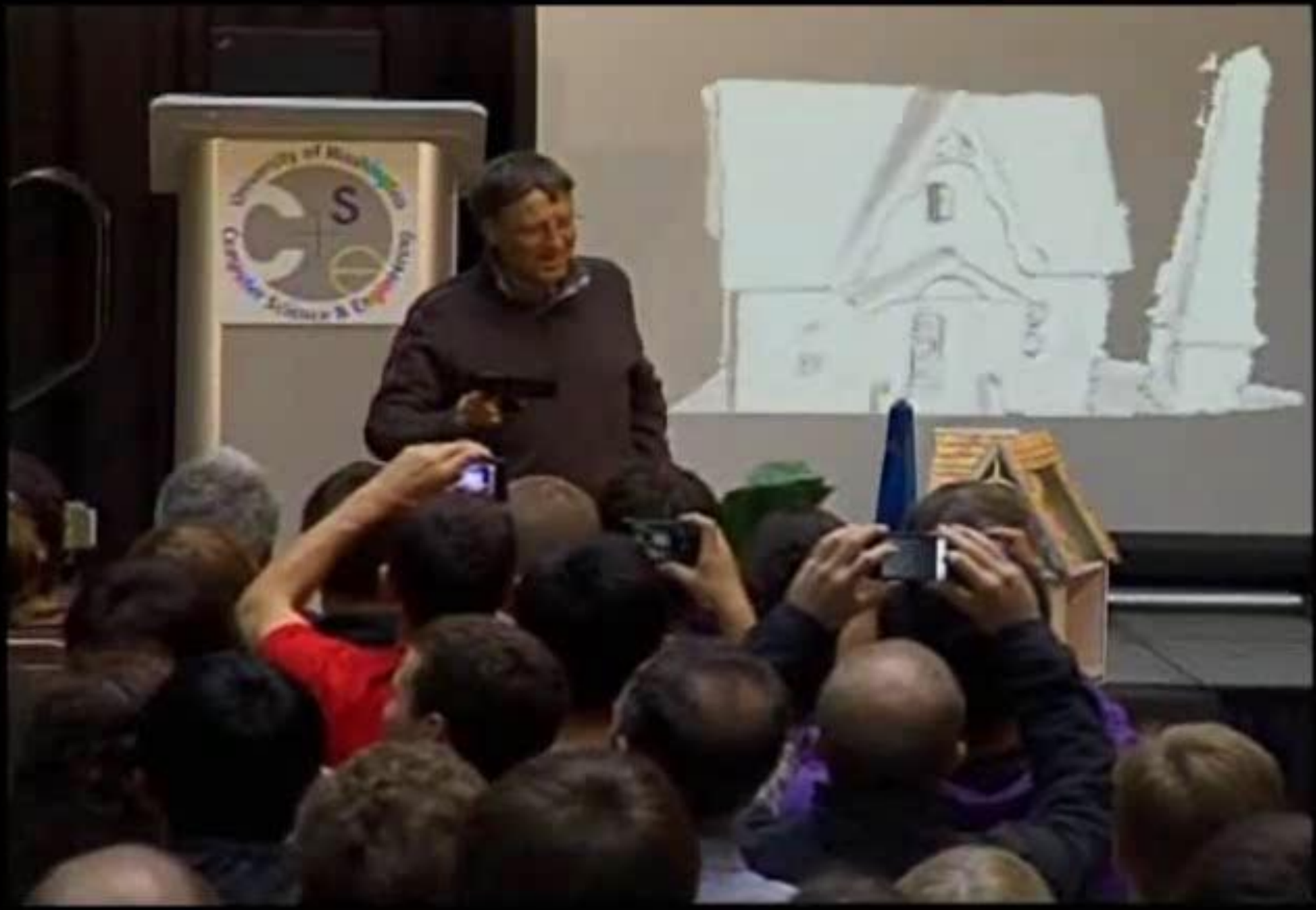
Raw Depth Image

Compute 3D Point Cloud and Normals

Camera Pose Estimation

Volumetric Fusion

Surface Extraction

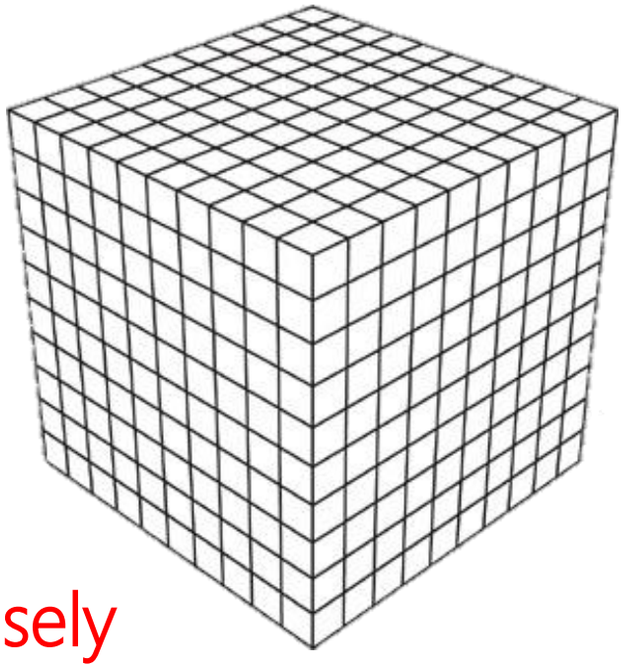


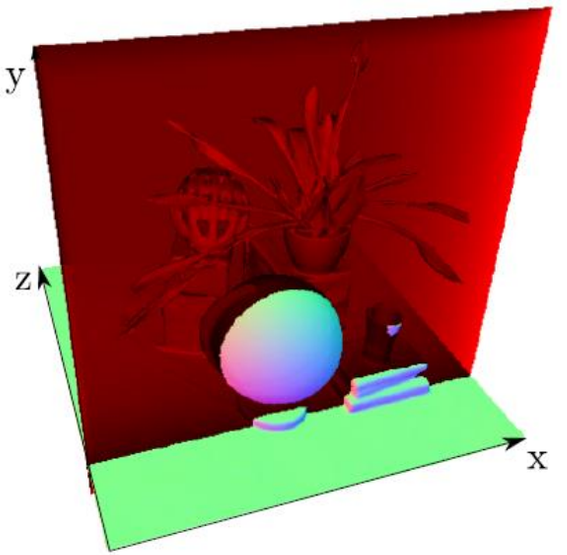
The Challenge of Scalability

-- limitations of the regular grid

512^3 8-byte voxels: 512 MB
(4mm per voxel, 2m volume)

Kinect range: 40cm to 8m





Uniform grid represents space densely

But: Signed Distance Function is sparse

Streaming Hierarchical Fusion

[Chen, Bautembach, Izadi, SIGGRAPH 2013]
[Zhou, Gong, Huang, Guo, 2011]

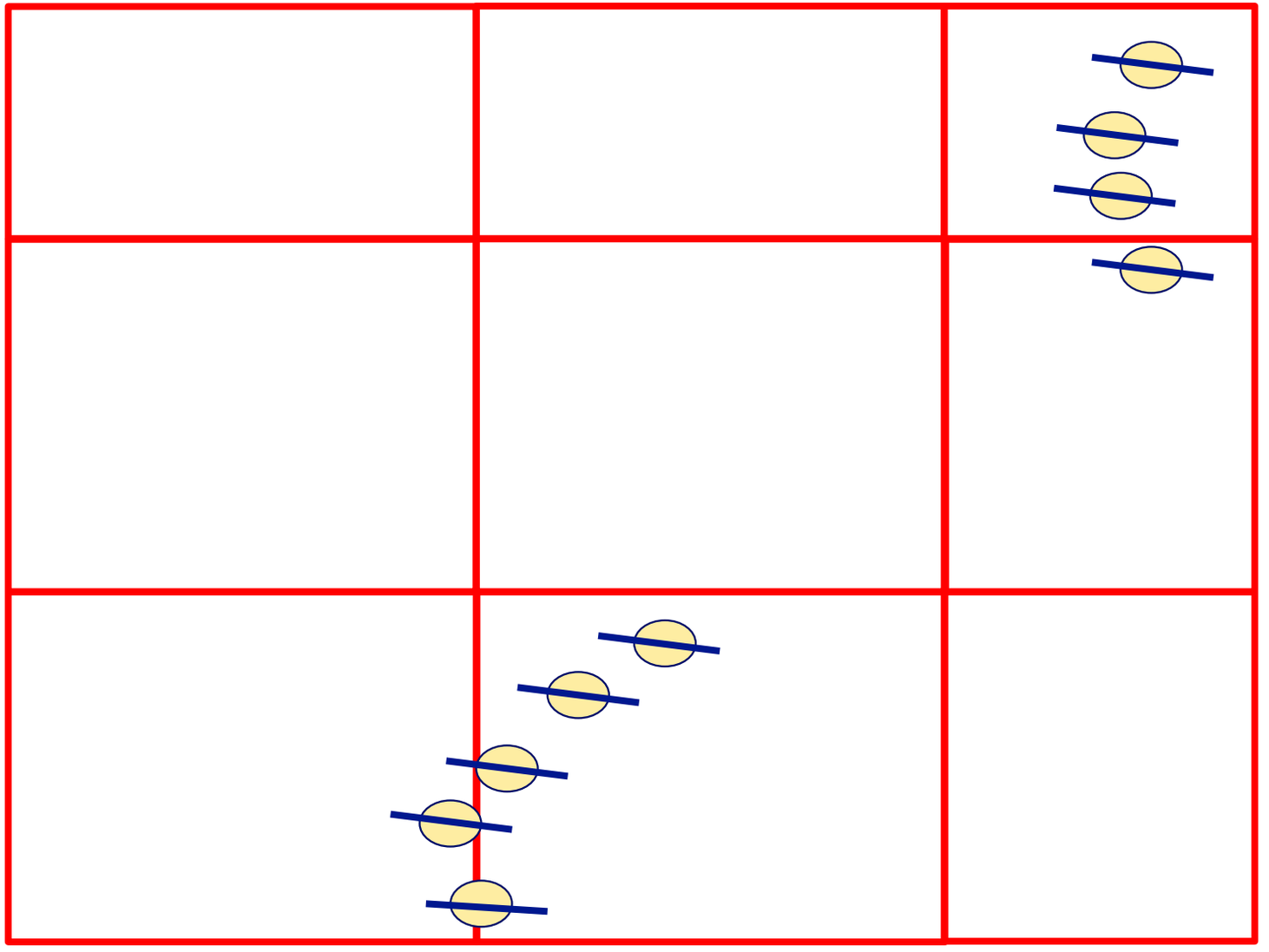
Hierarchical

- compress free space
- high branching factor
- GPU implementation

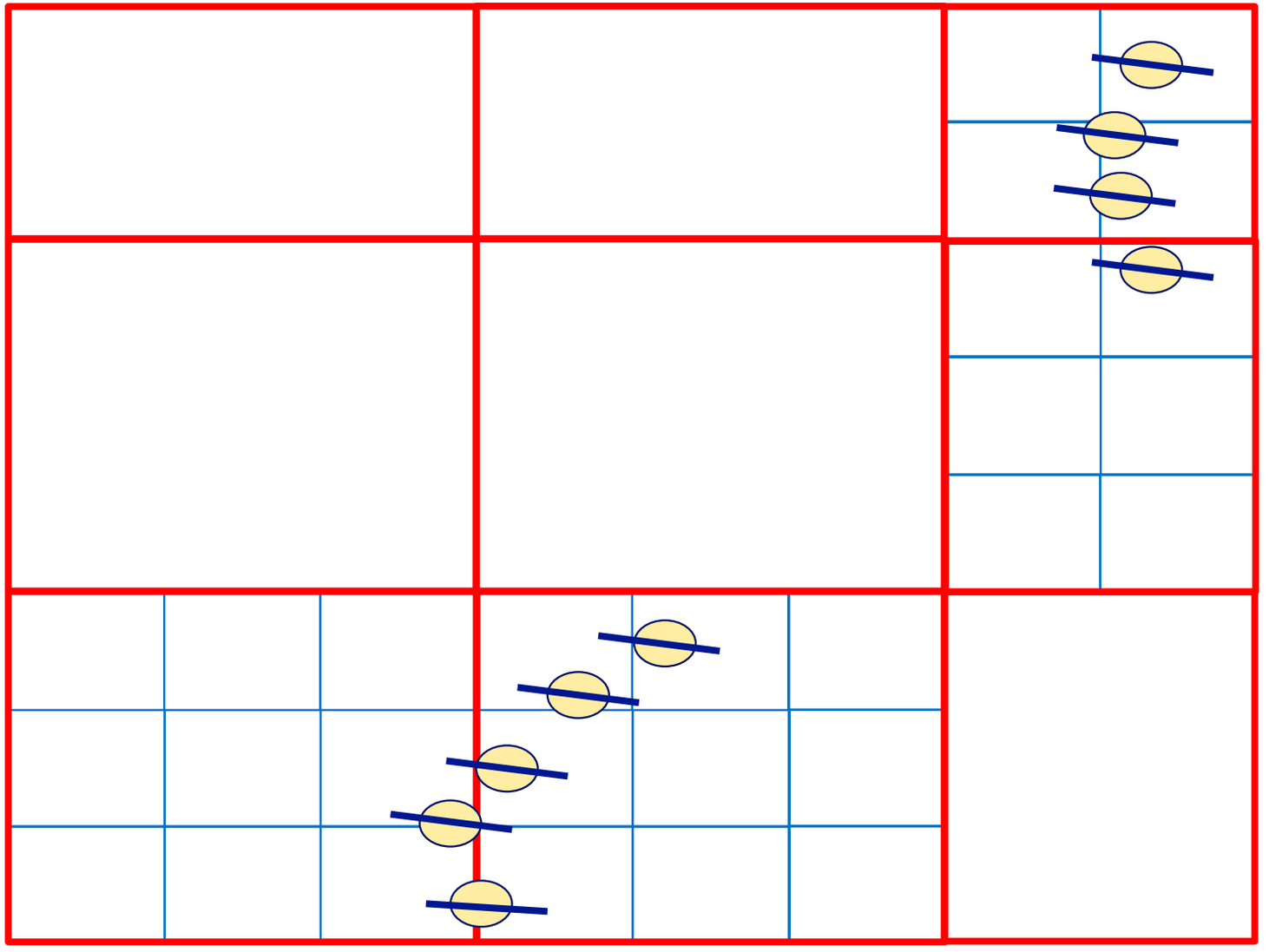
Streaming

- sub-volume on GPU moves with camera

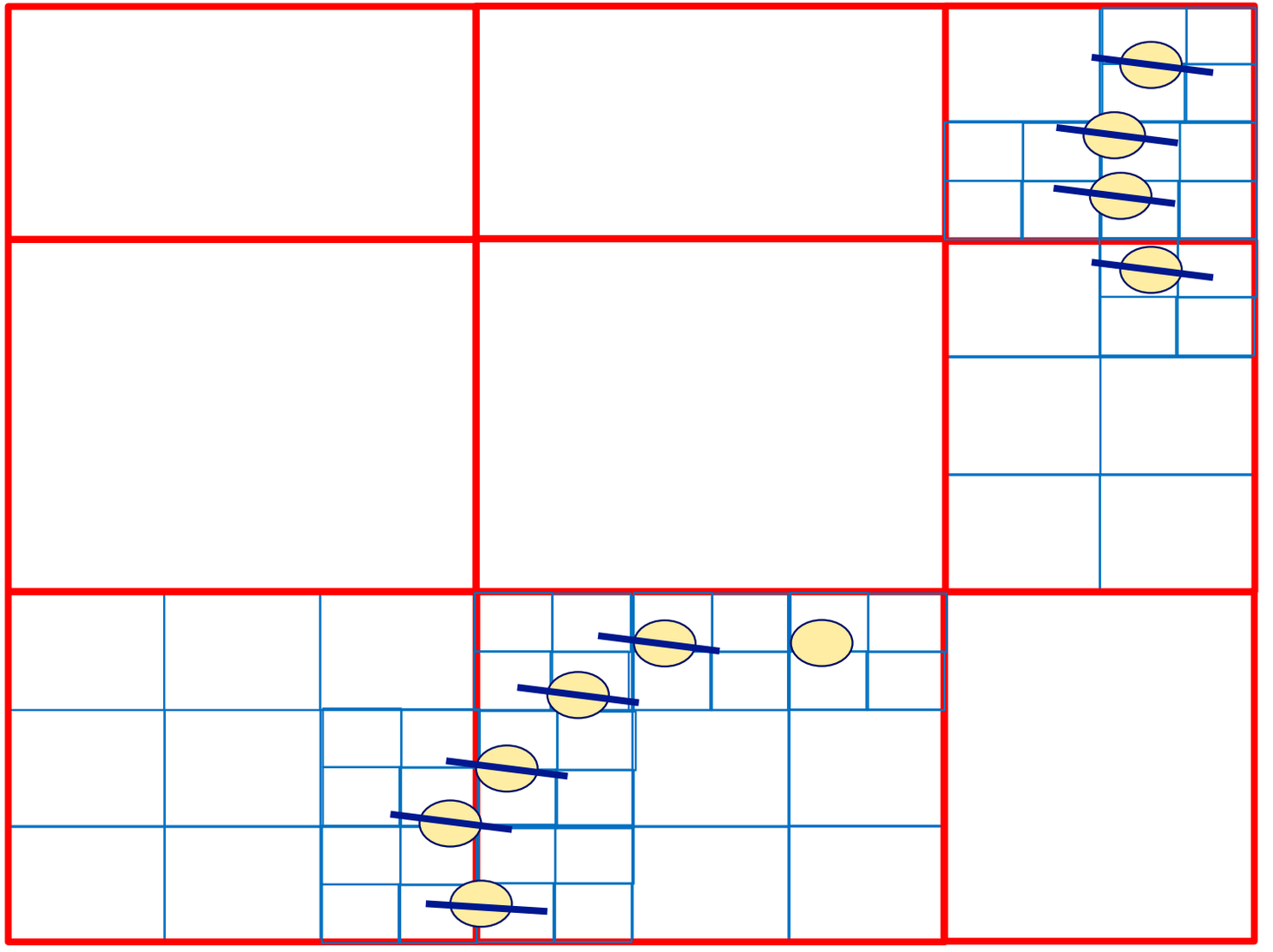
Hierarchical Volumetric Fusion



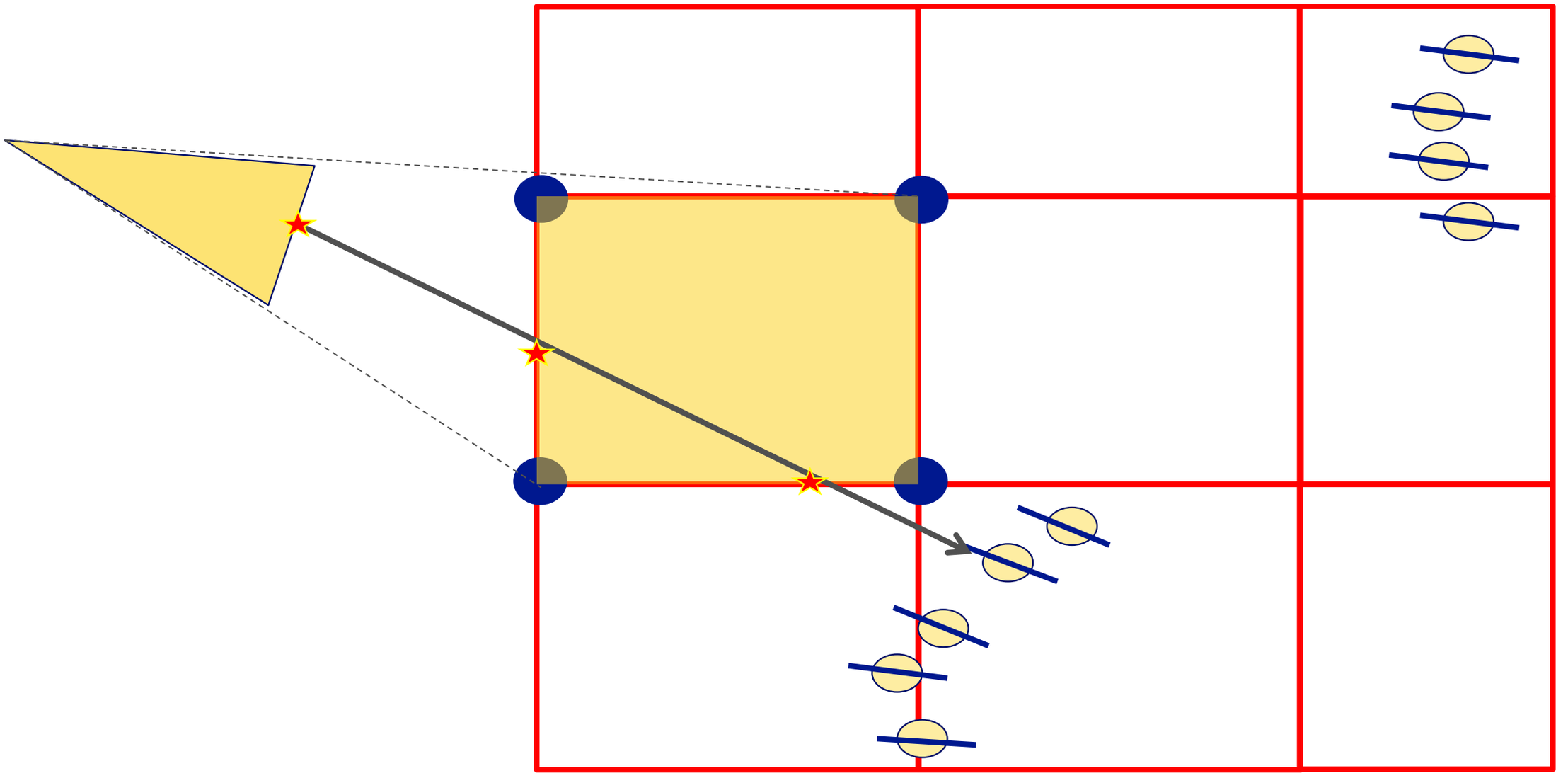
Hierarchical Volumetric Fusion



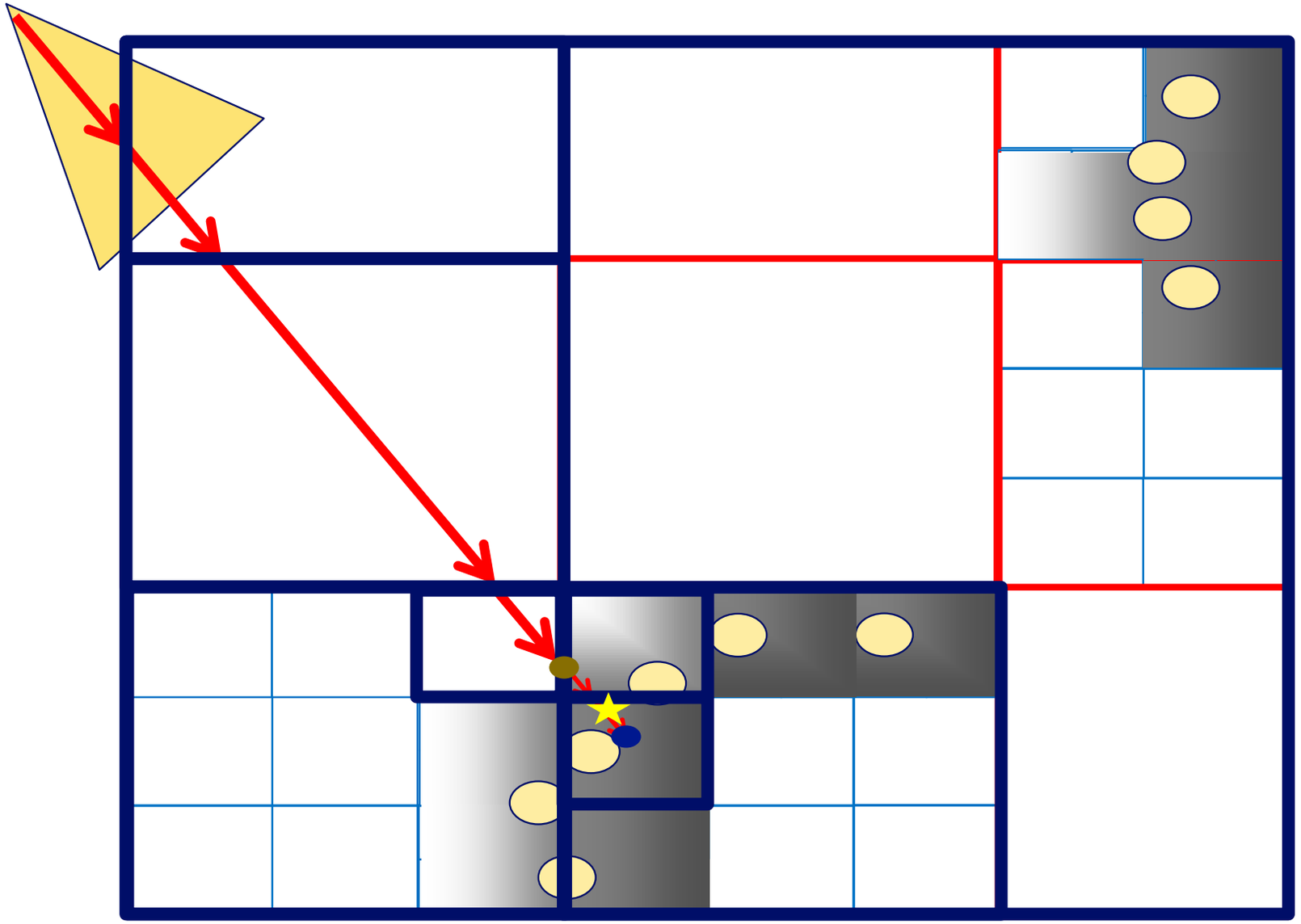
Hierarchical Volumetric Fusion



Subdivision rule

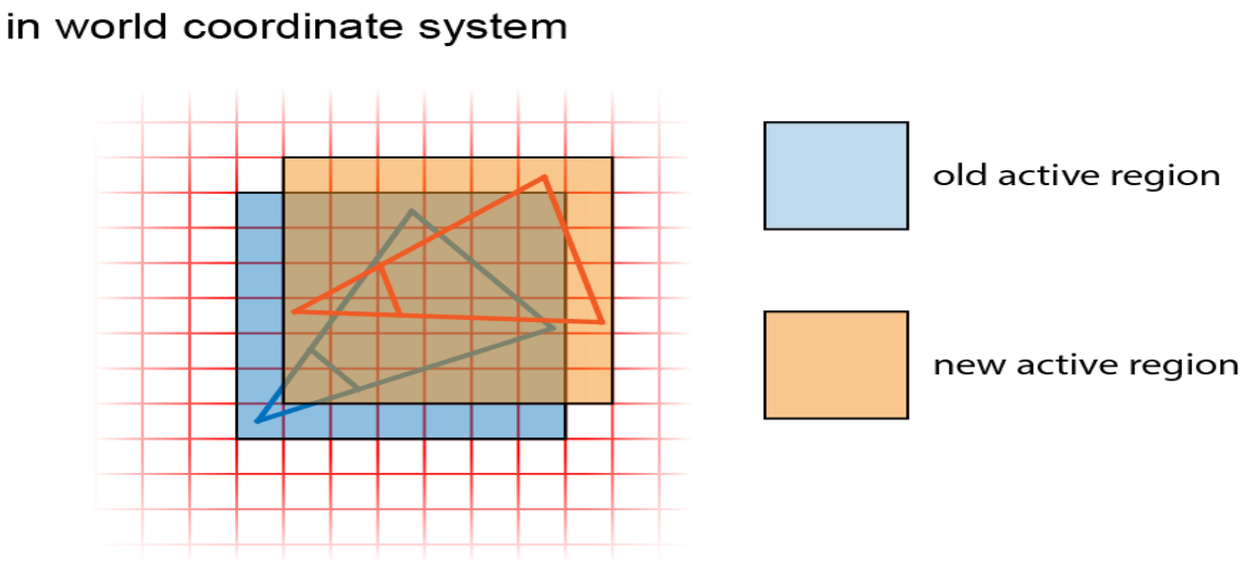


Hierarchical surface extraction



Skip over large regions of free space

Streaming

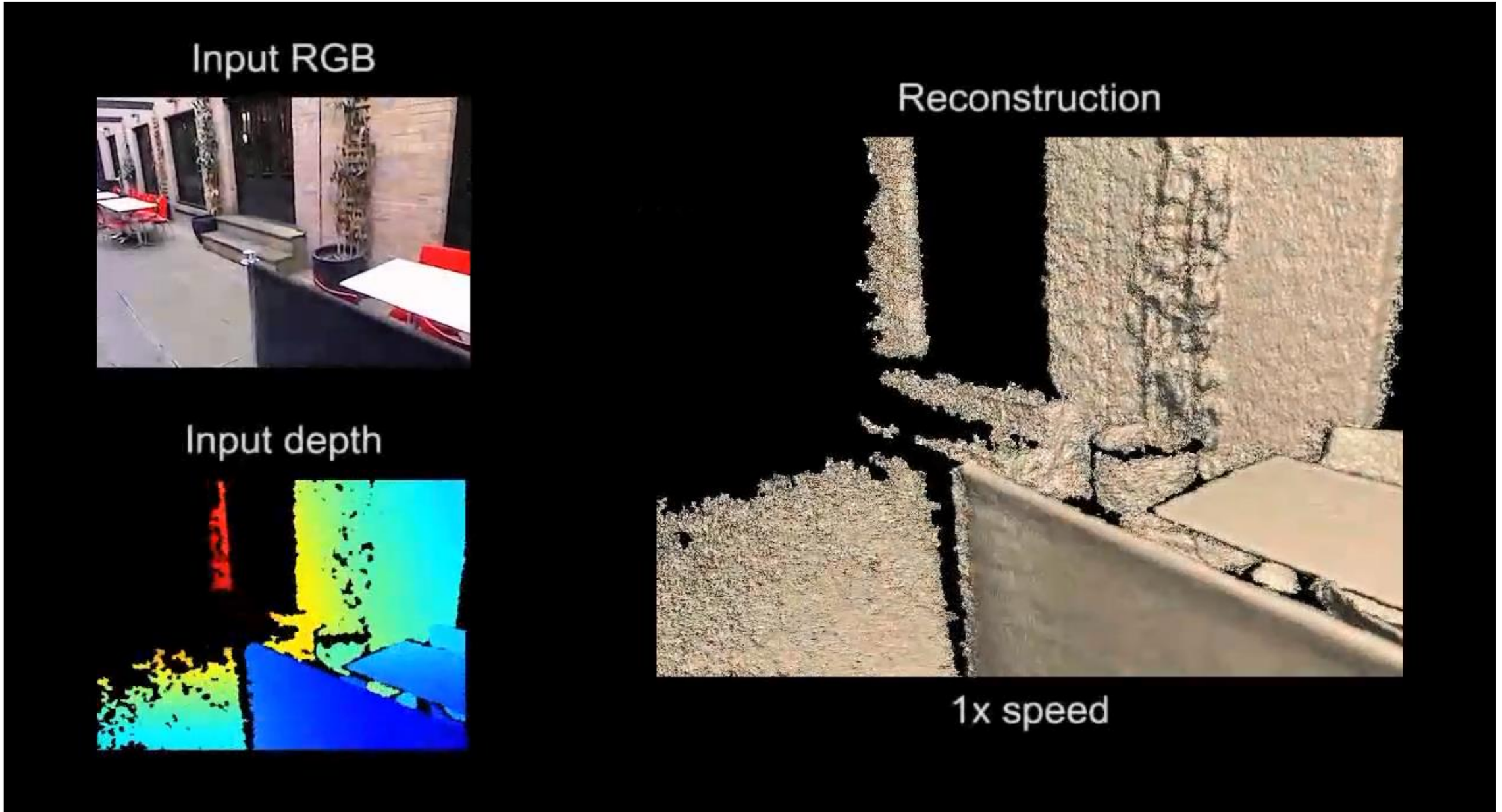


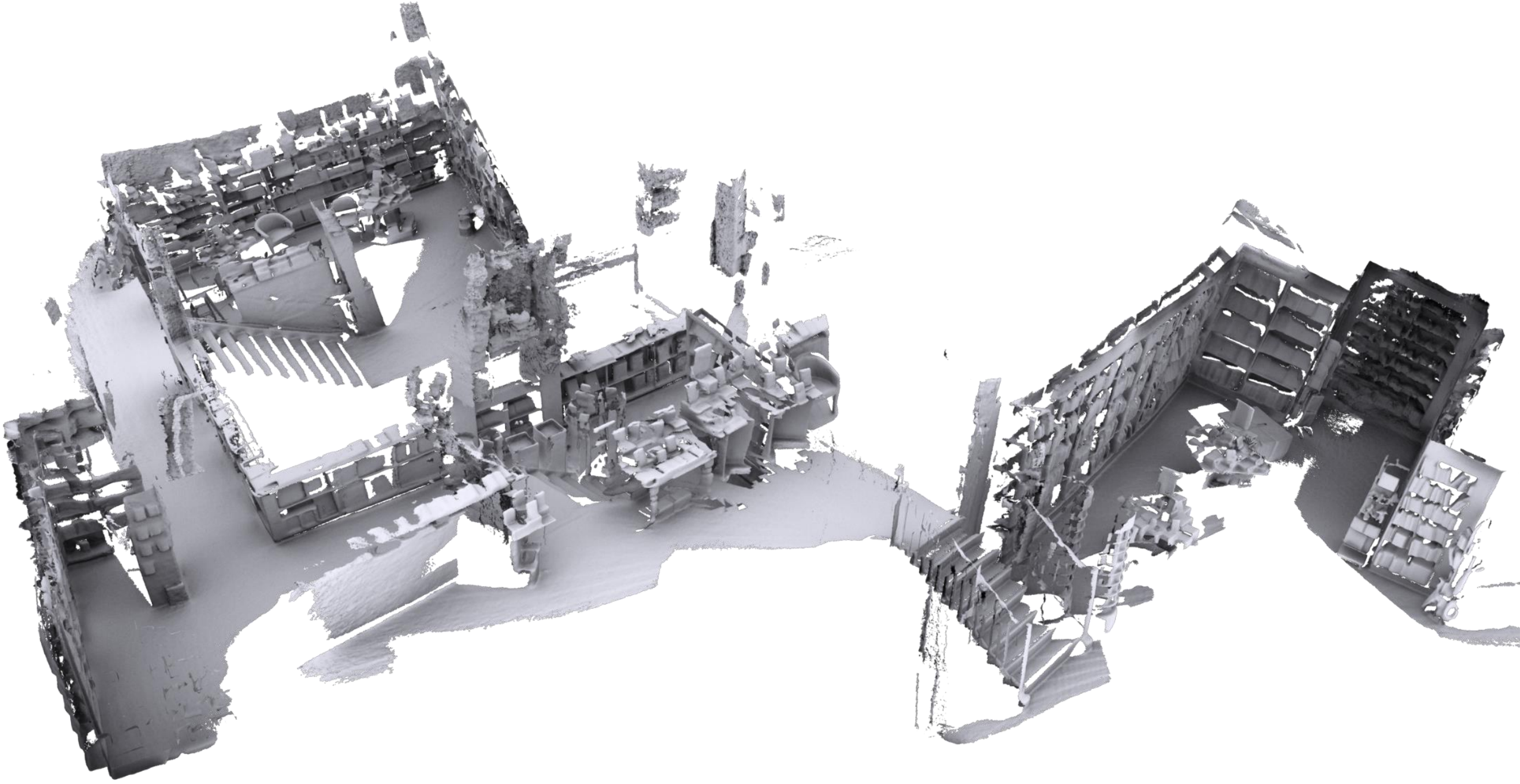
Serialize subtree:
breadth first search

Subtrees are sparse
(< 1 MB each)

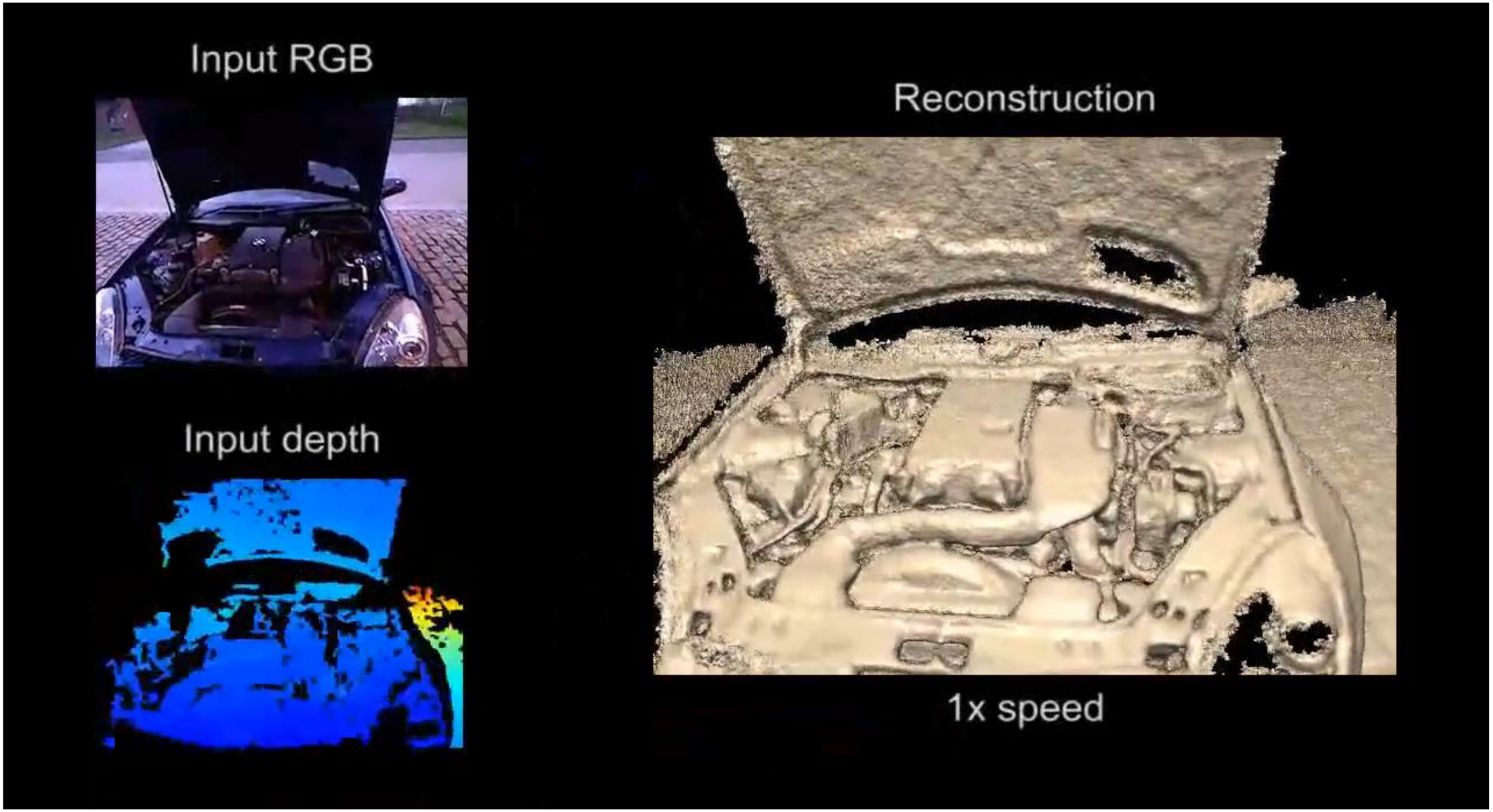
~2 ms to *losslessly* transfer
between GPU and host

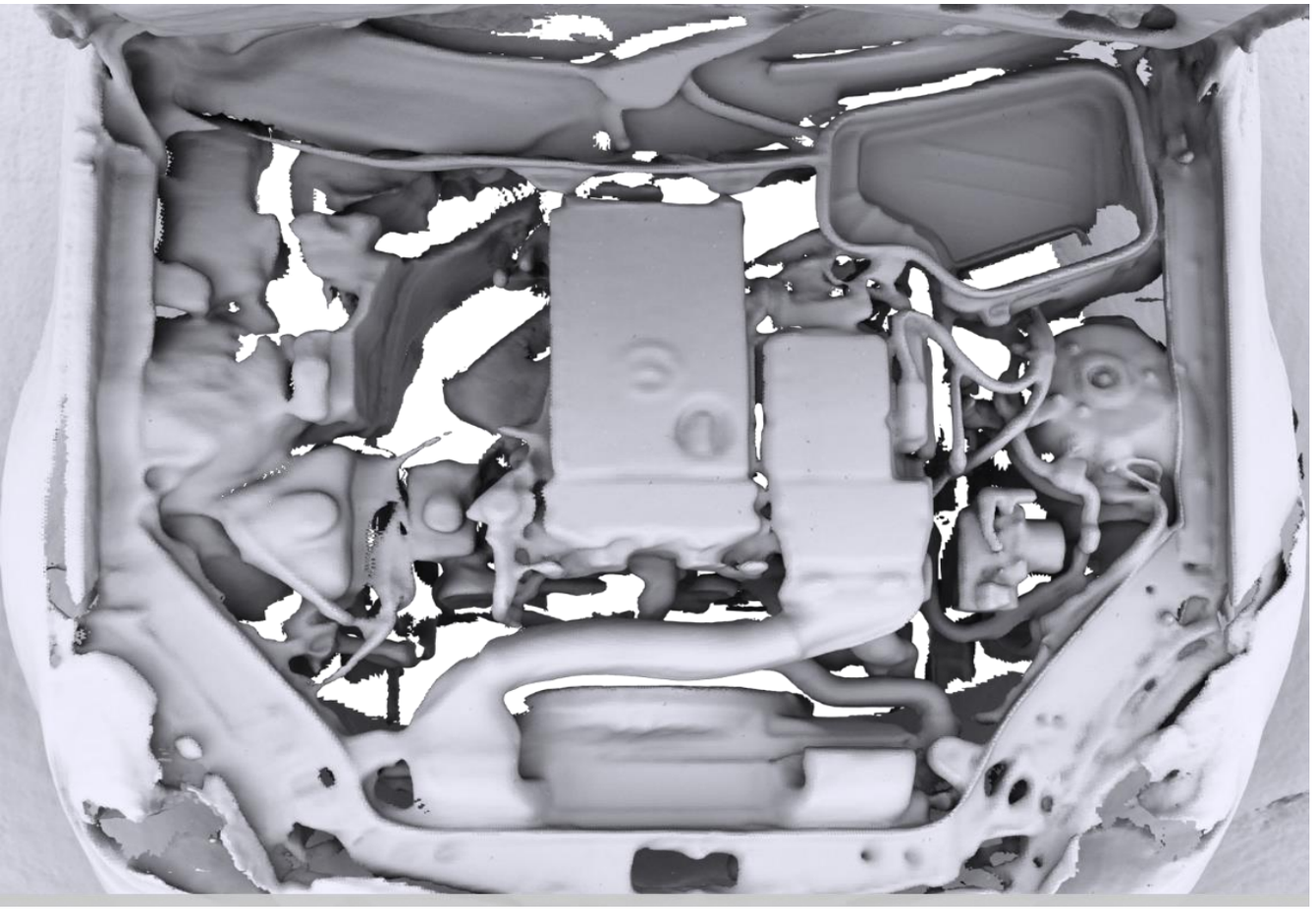
Scanning a bookshop

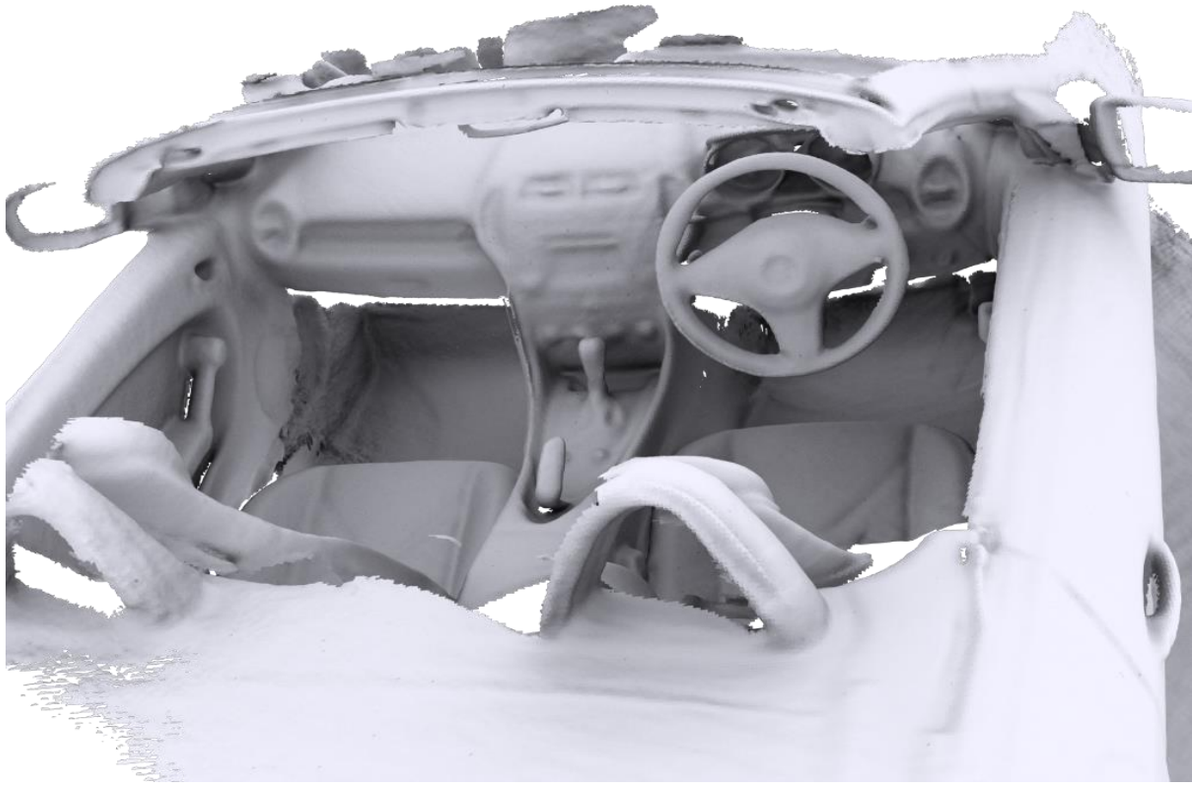




Fusion style





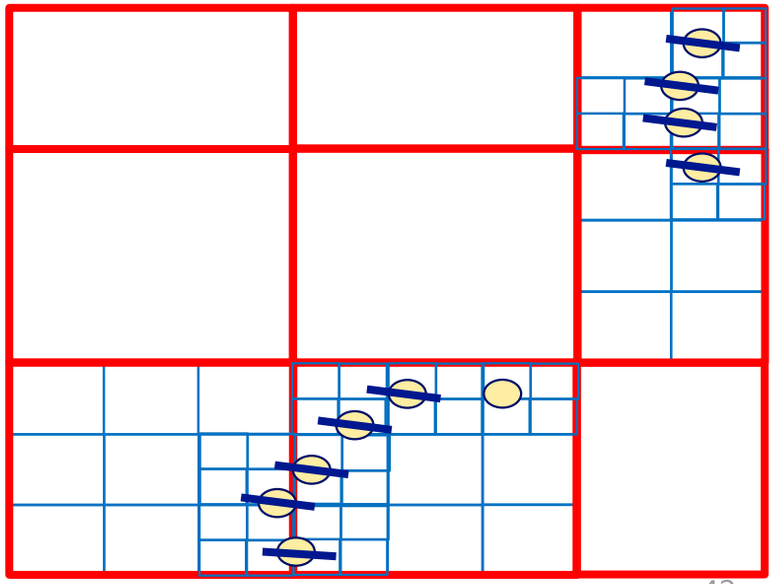


Summary

Real-time large-scale 3D reconstruction

Exploit the sparsity of surface data

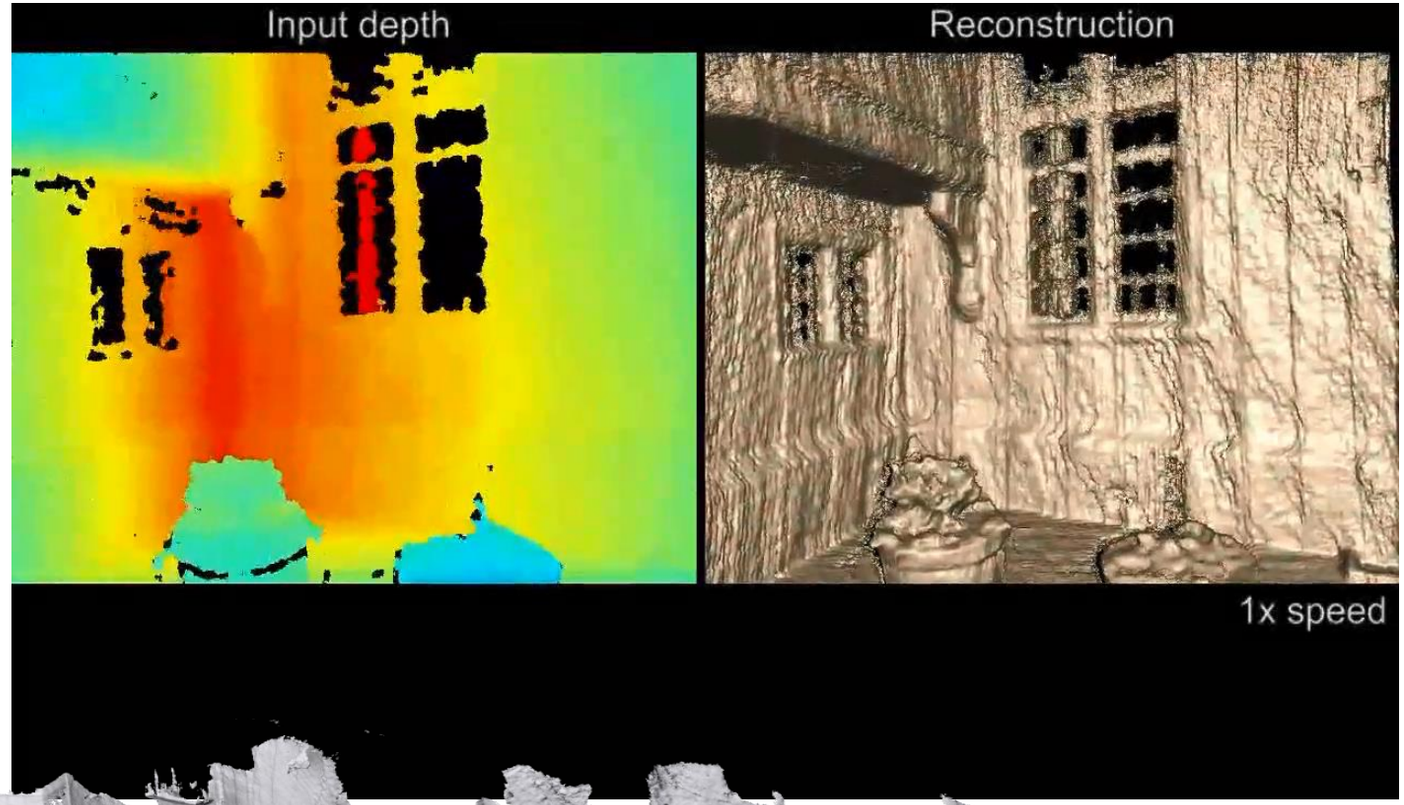
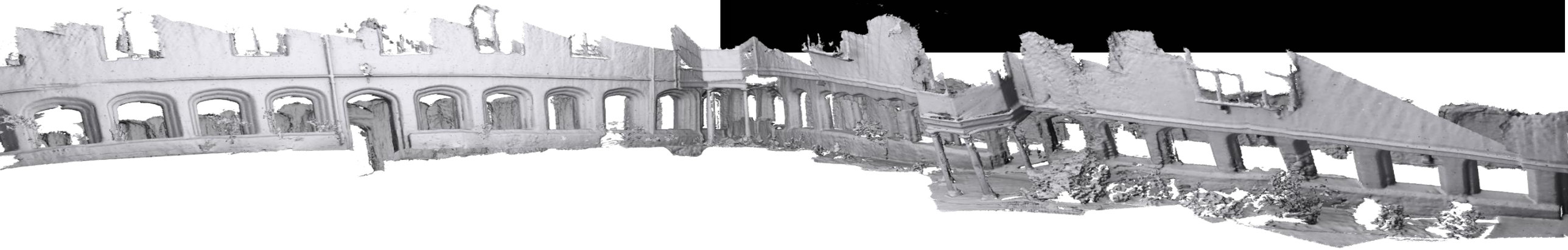
Hierarchical streaming data structure



Issues for the future

Drift and loop closure

Relocalization

Portability

Labelling the world