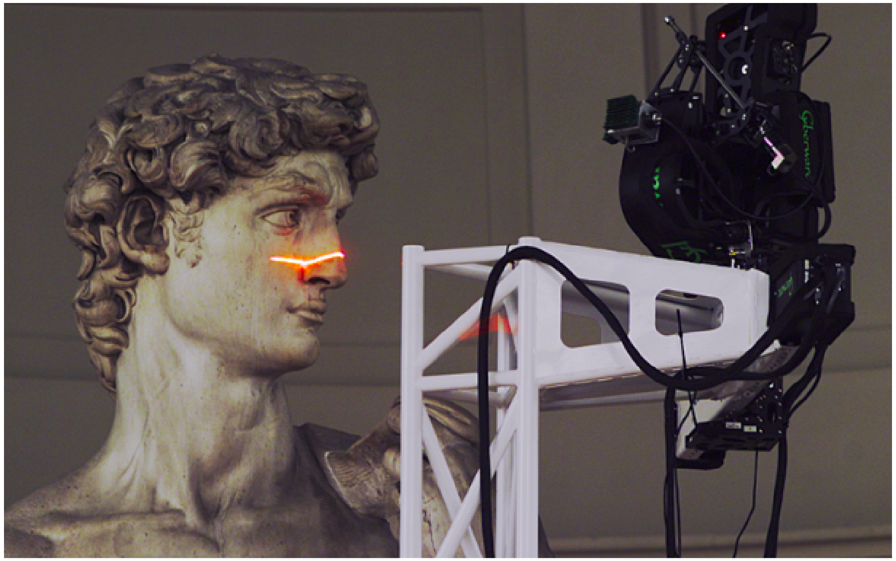
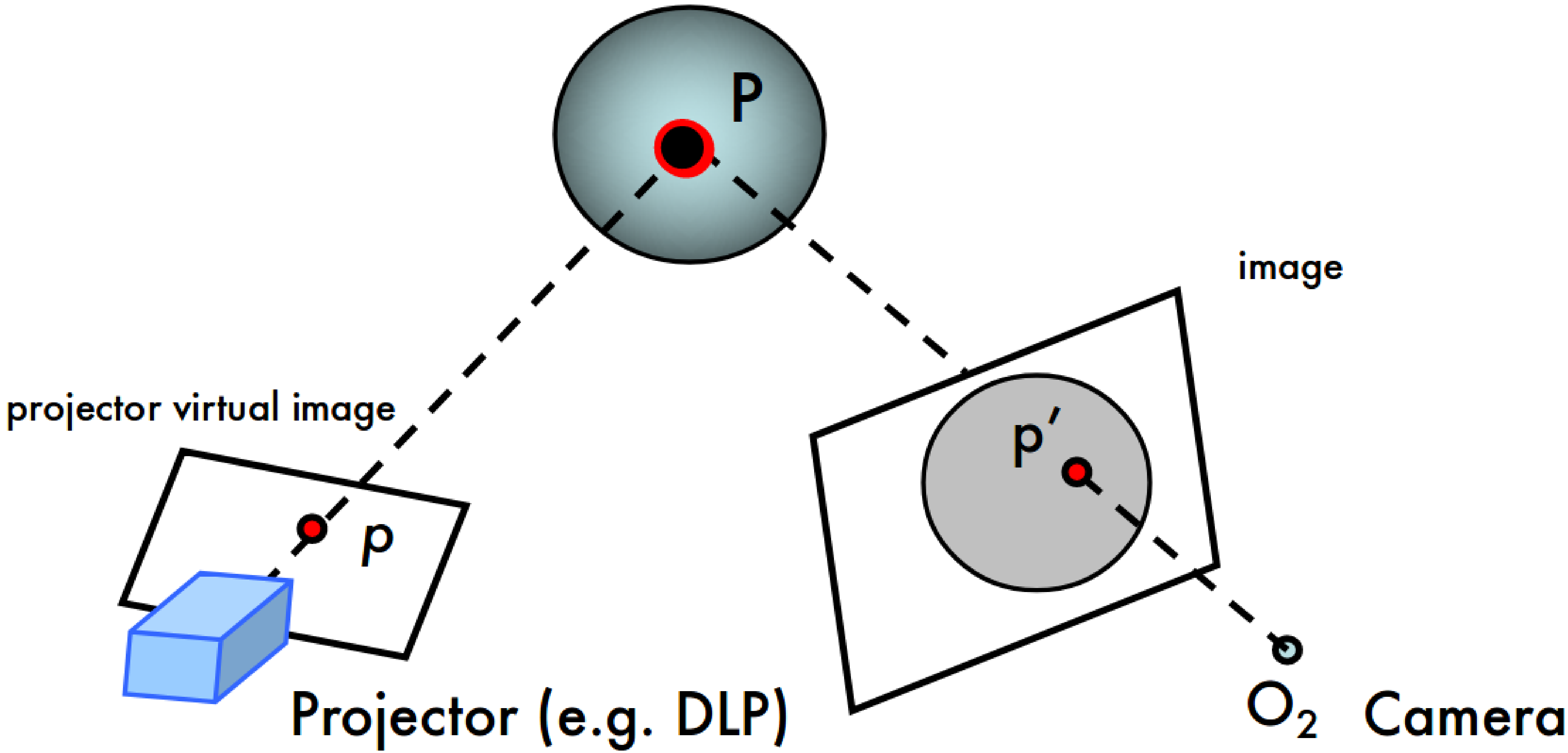
Active Stereo

- Structured Lighting
  - Point
  - Stripe
  - Shadows
  - Color-coded Stripes
- Depth Sensing
  - Kinect

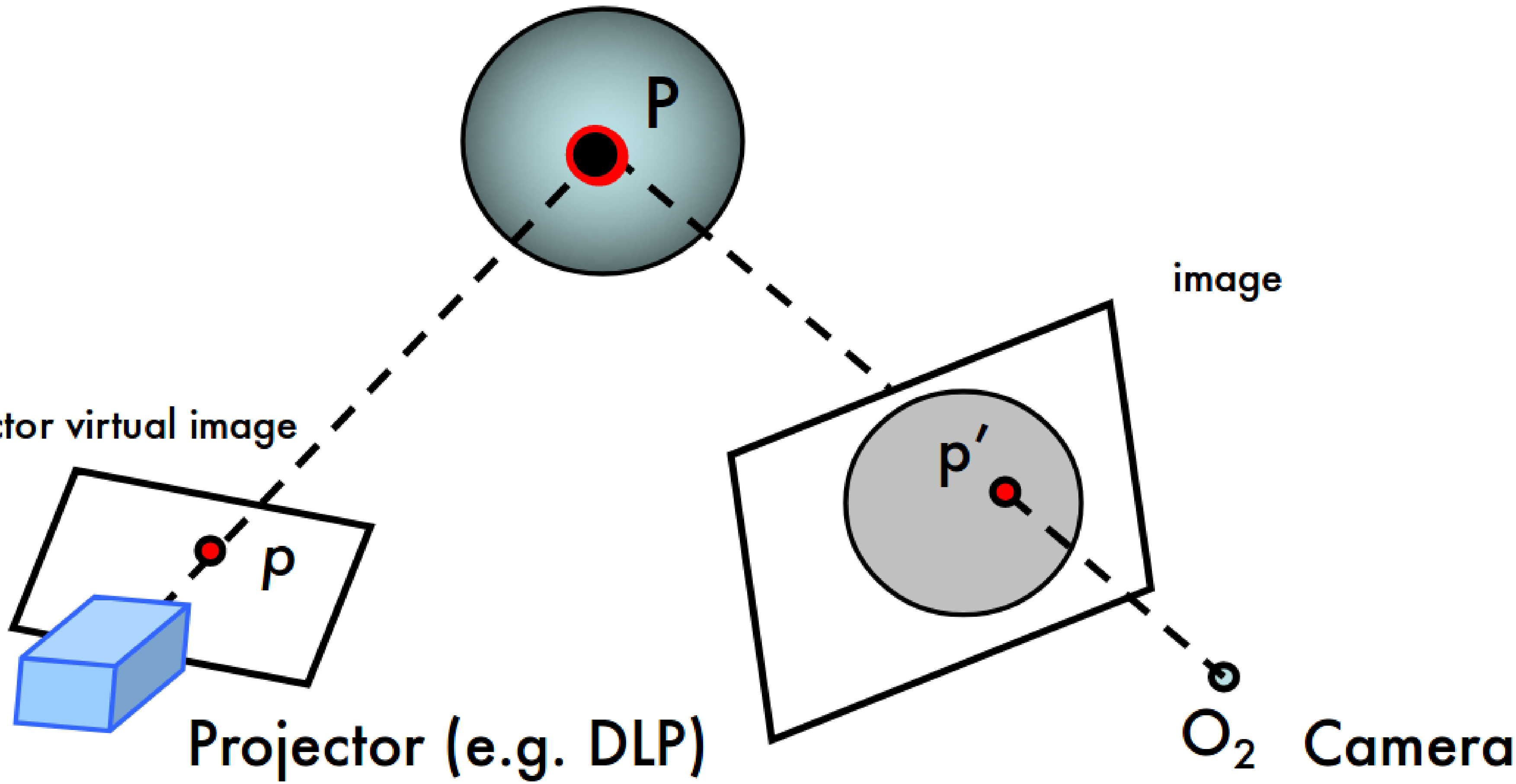


Digital Michelangelo Project (1990)
http://graphics.stanford.edu/projects/mich/

Active stereo (point)



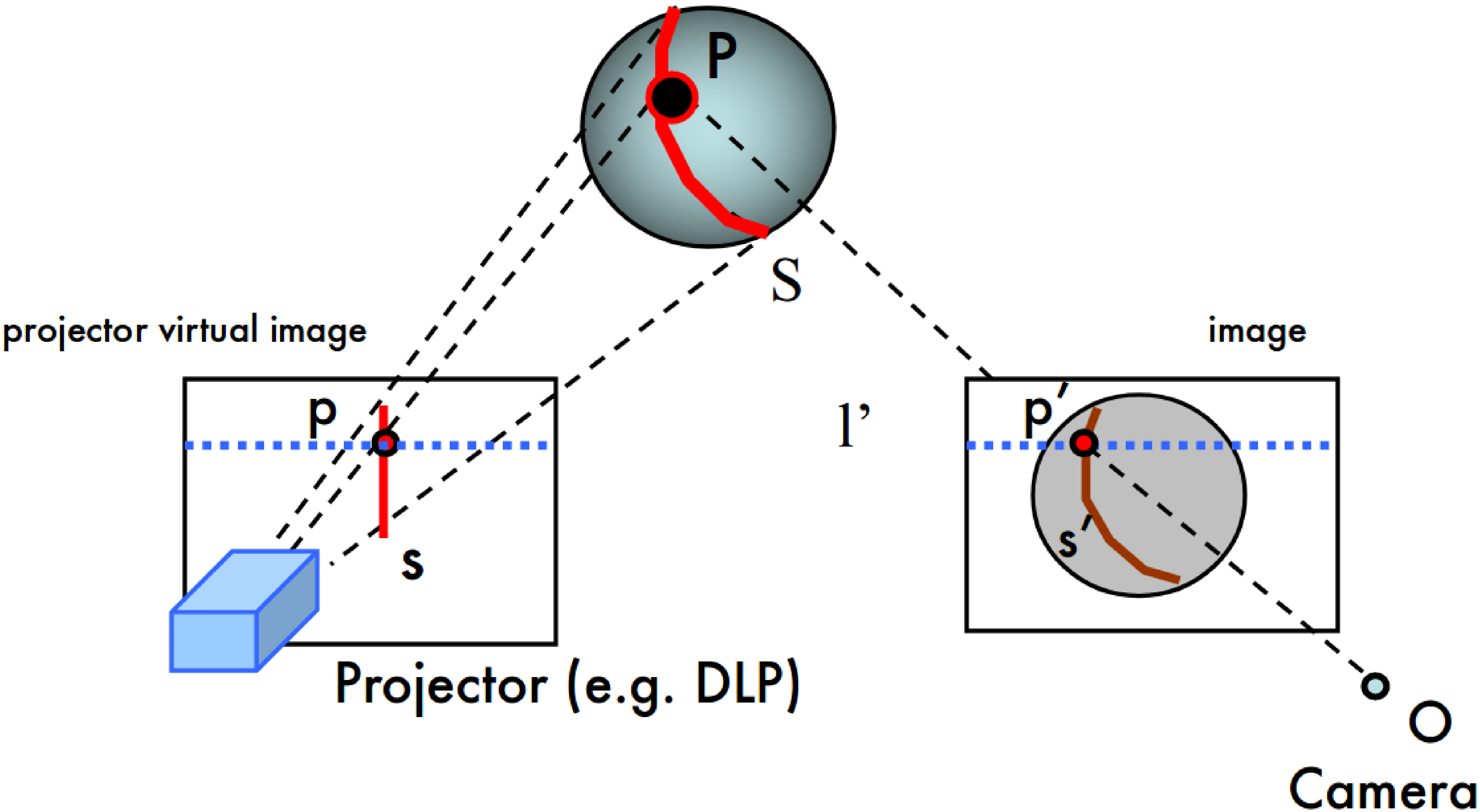
Active stereo (point)



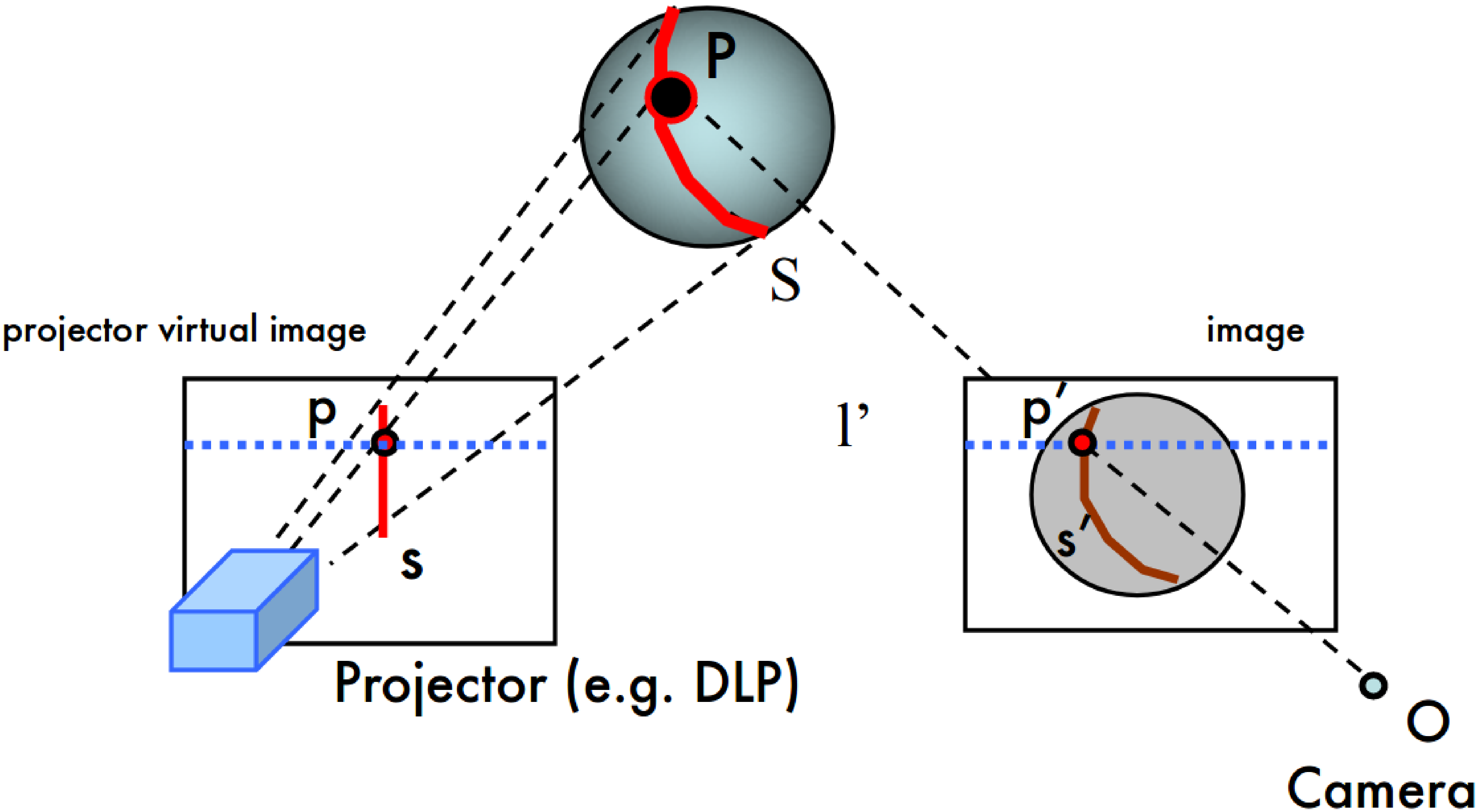
Replace one of the two cameras by a projector

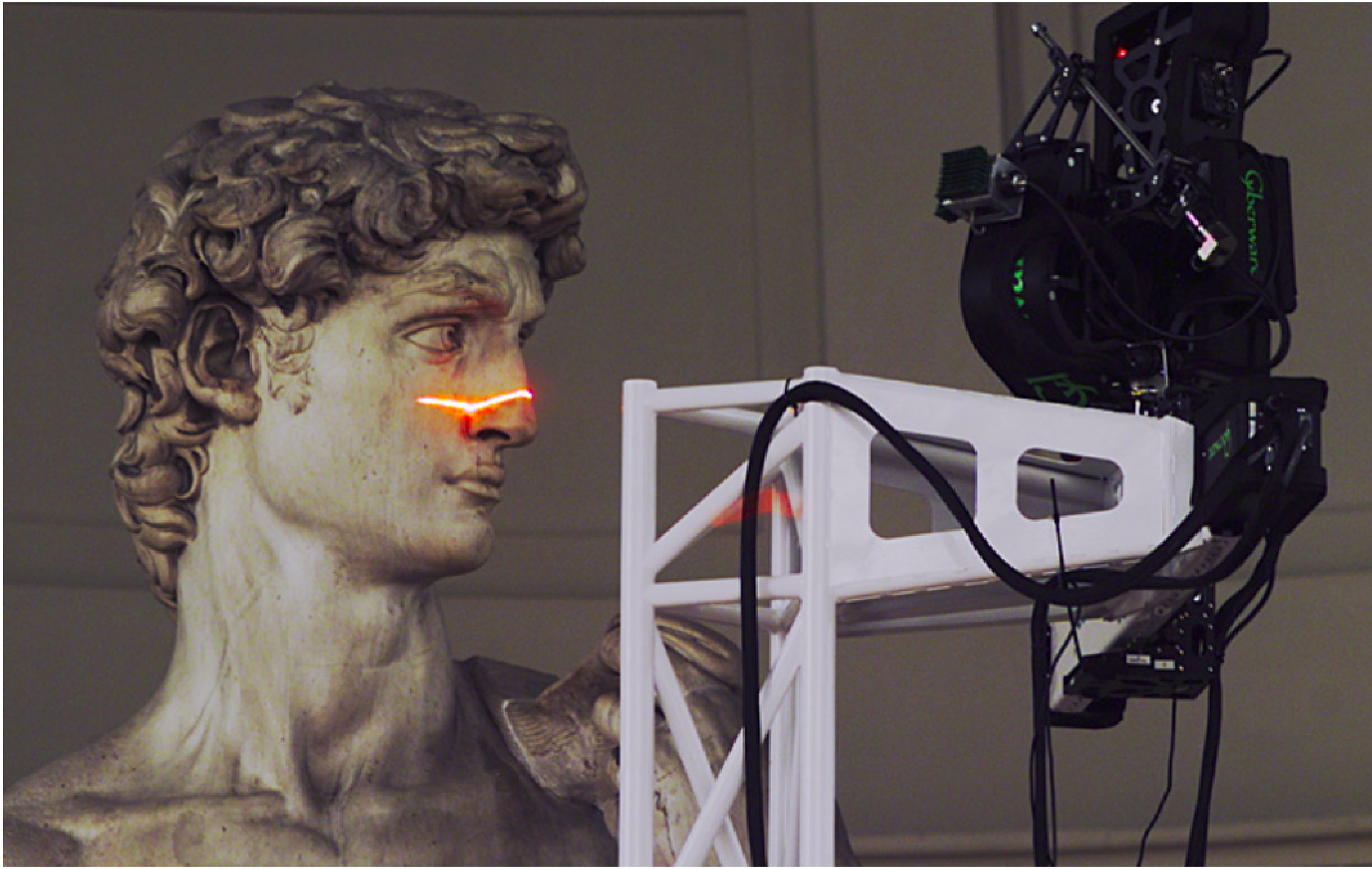
- Projector geometry calibrated
- What's the advantage of having the projector? Correspondence problem solved!

Active stereo (stripe)



Active stereo (stripe)





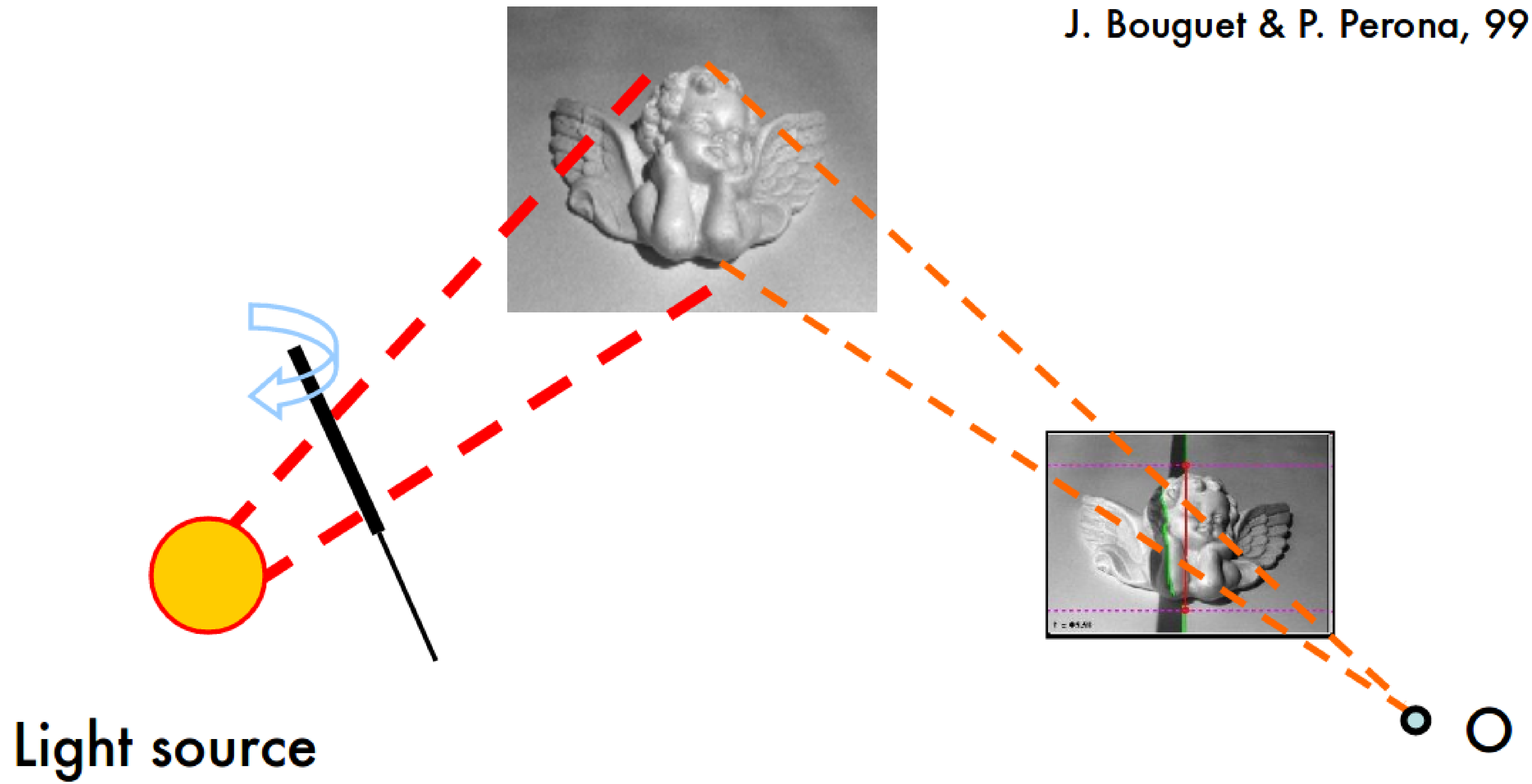
Digital Michelangelo Project (1990)
http://graphics.stanford.edu/projects/mich/

Laser scanning

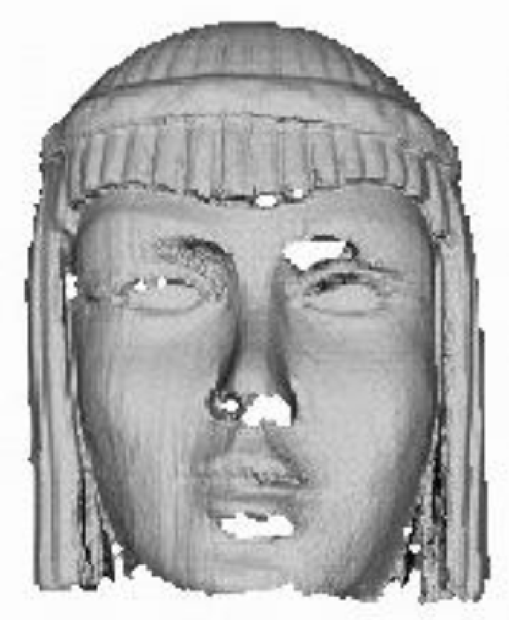
The Digital Michelangelo Project, Levoy et al.

Active stereo (shadows)

J. Bouguet & P. Perona, 99

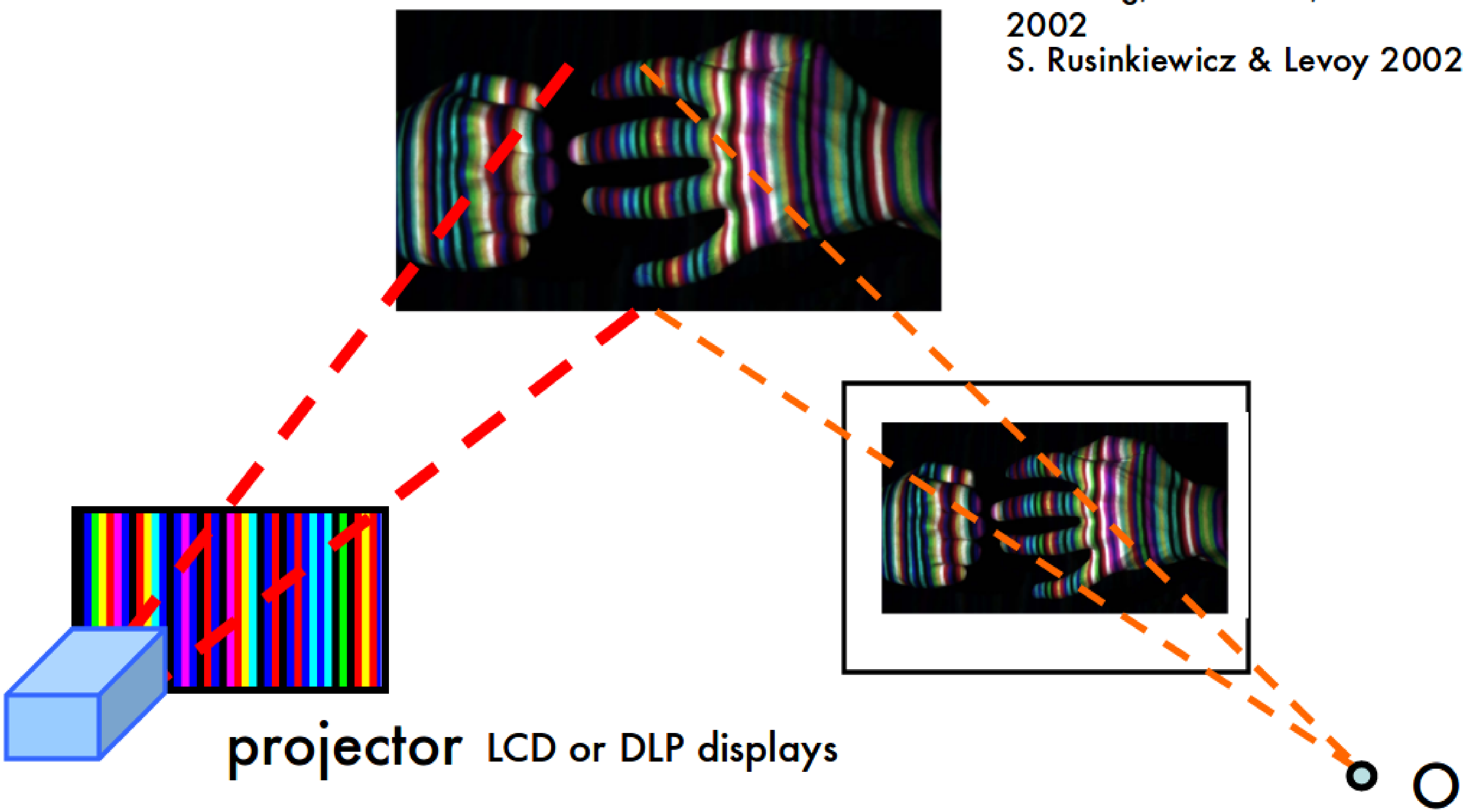


Active stereo (shadows)



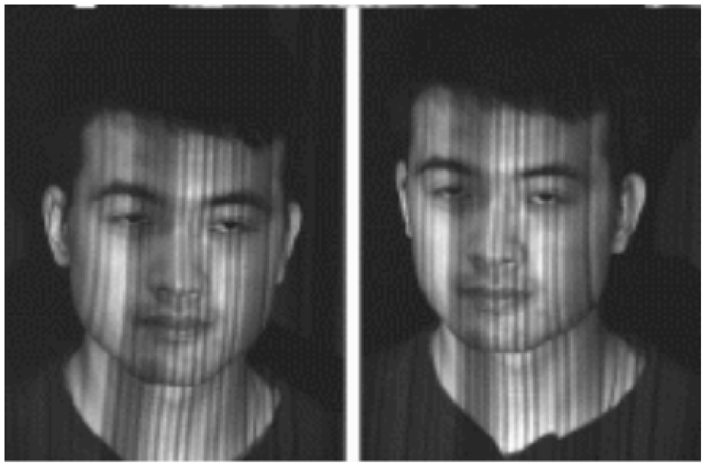
Active stereo (color-coded stripes)

L. Zhang, B. Curless, and S. M. Seitz 2002
S. Rusinkiewicz & Levoy 2002

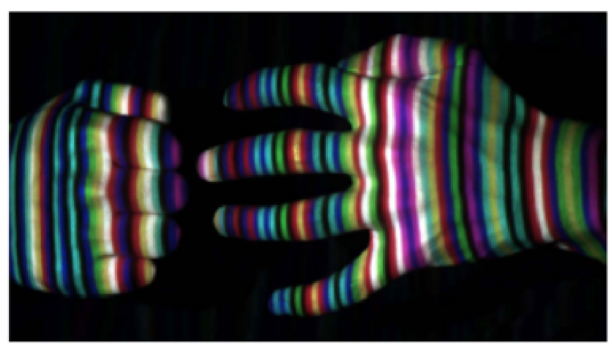


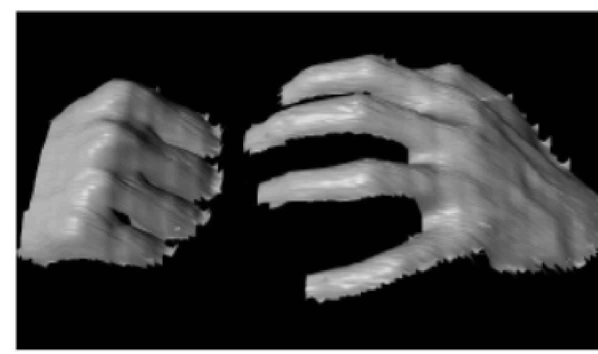
- Dense reconstruction
- Correspondence problem again
- Get around it by using color codes

Active stereo (color-coded stripes)



Rapid shape acquisition: Projector + stereo cameras



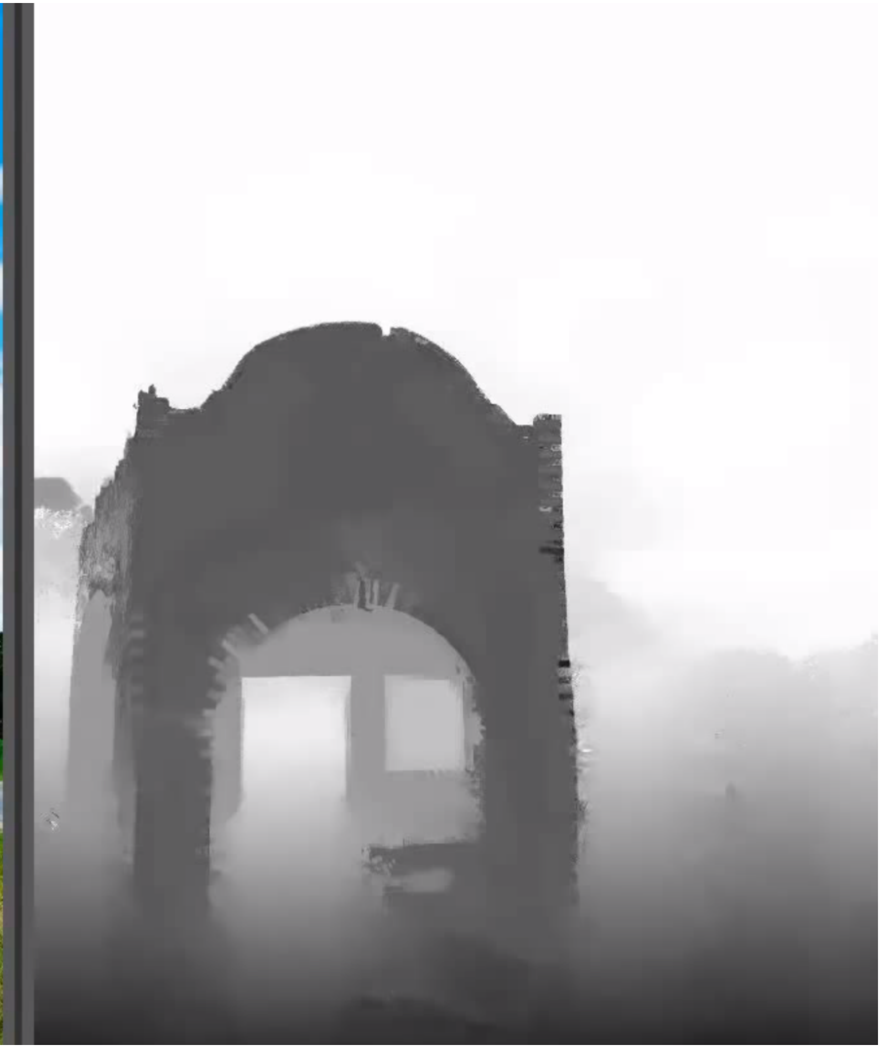


L. Zhang, B. Curless, and S. M. Seitz. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. *3DPVT* 2002

Depth sensing

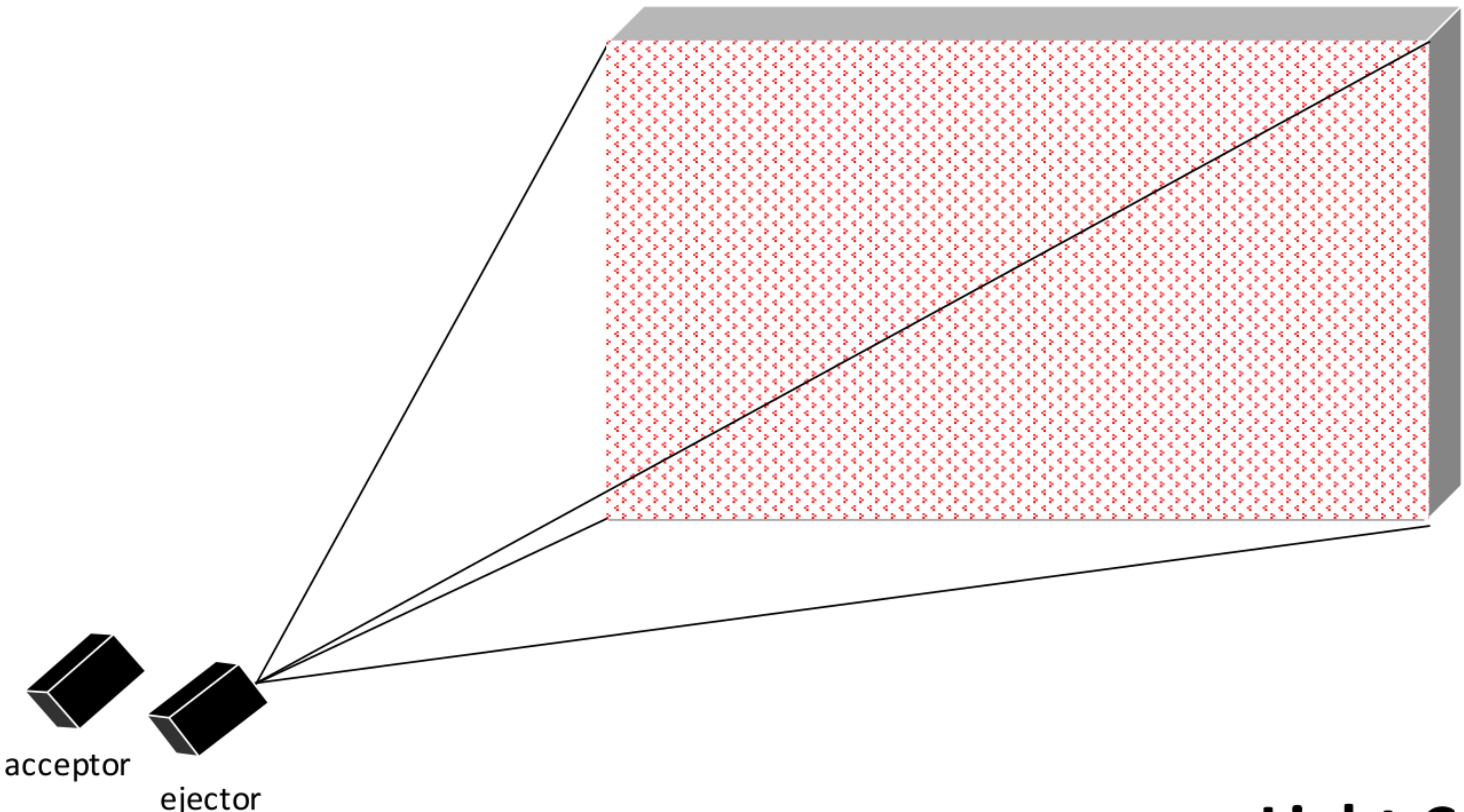
Depth map

RGB



Depth map

KINECT V1



Light Coding

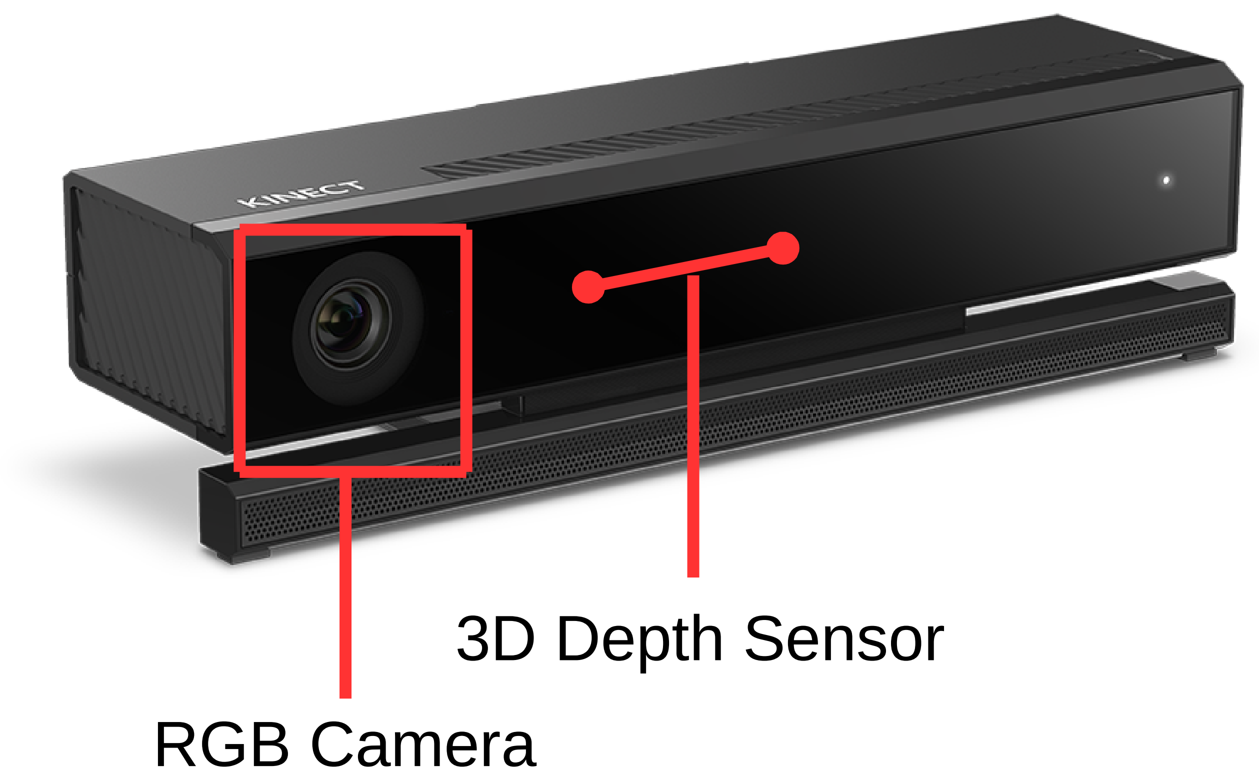
Limitations

- Short distance

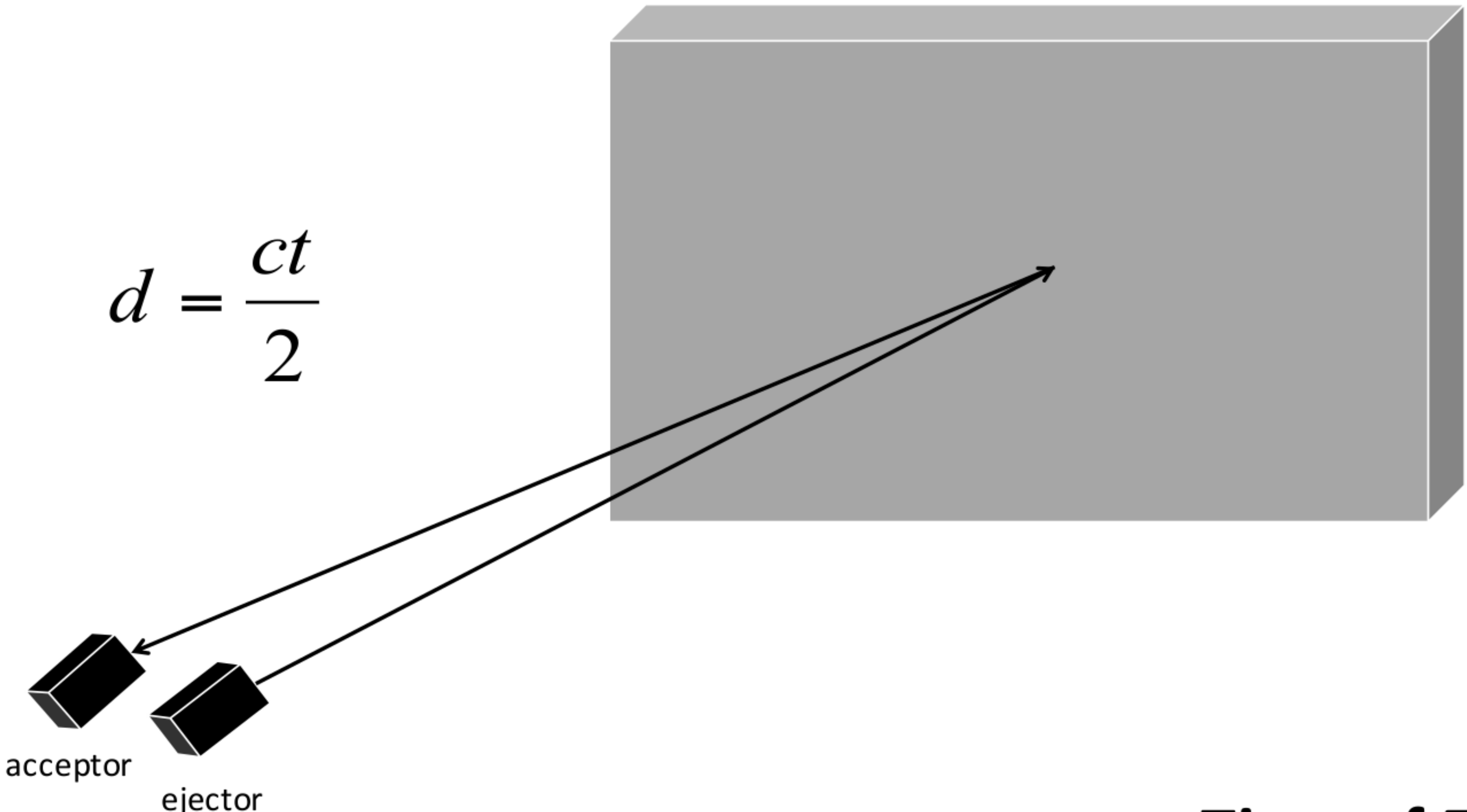
Limitations

- Occlusion results in lossing depth values
- Surface material causes loss of depth value

KINECT V2

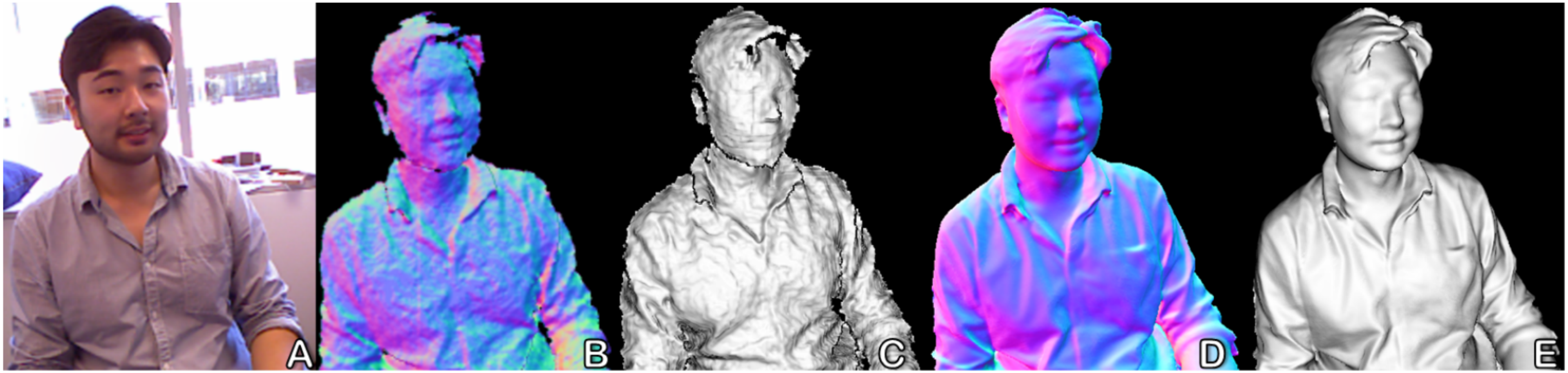


$$d = \frac{ct}{2}$$

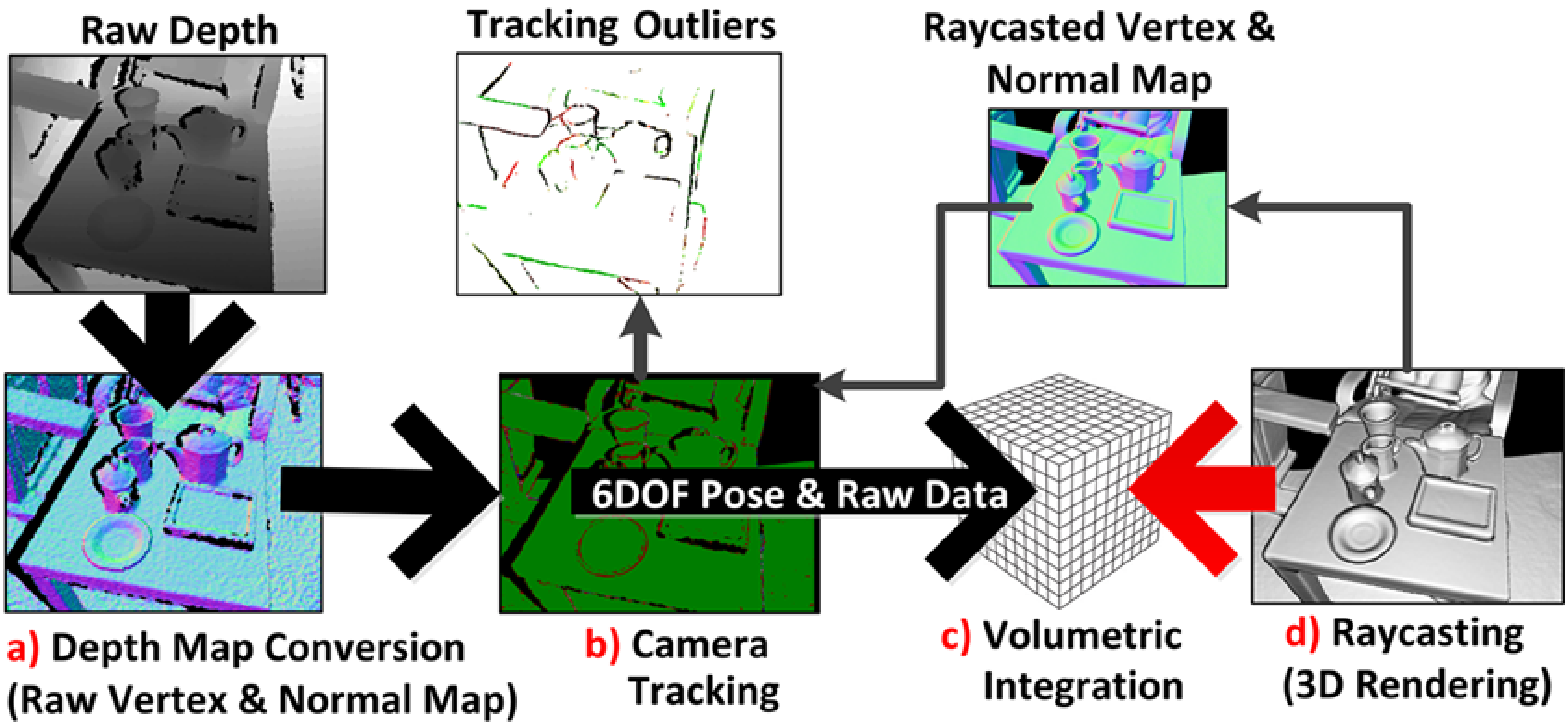


Time of Flight

KincetFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera



Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera (2011)


Raw Depth
Tracking Outliers
Raycasted Vertex &
Normal Map
6DOF Pose & Raw Data
a) Depth Map Conversion
(Raw Vertex & Normal Map)
b) Camera
Tracking
c) Volumetric
Integration
d) Raycasting
(3D Rendering)

Camera Tracking

Listing 1 Projective point-plane data association.

for each image pixel $\mathbf{u} \in$ depth map $\mathbf{D_i}$ **in parallel do**
    if $\mathbf{D}_i(\mathbf{u}) > \mathbf{0}$ **then**
        $\mathbf{v}_{i-1} \leftarrow \mathbf{T}_{i-1}^{-1} \mathbf{v}_{i-1}^{\mathrm{g}}$
        $\mathbf{p} \leftarrow$ perspective project vertex $\mathbf{v}_{i-1}$
        if $\mathbf{p} \in$ vertex map $\mathbf{V}_i$ **then**
            $\mathbf{v} \leftarrow \mathbf{T}_{i-1} \mathbf{V}_i(\mathbf{p})$
            $\mathbf{n} \leftarrow \mathbf{R}_{i-1} \mathbf{N}_i(\mathbf{p})$
            if $||\mathbf{v} - \mathbf{v}_{i-1}^{\mathrm{g}}|| <$ distance threshold **and** $\mathbf{n} \,.\, \mathbf{n}_{i-1}^{\mathrm{g}} <$ normal threshold **then**
                point correspondence found

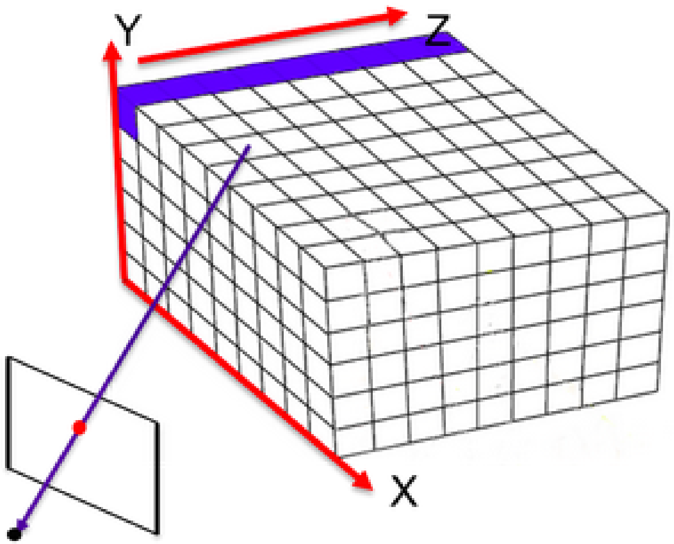
Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera (2011)

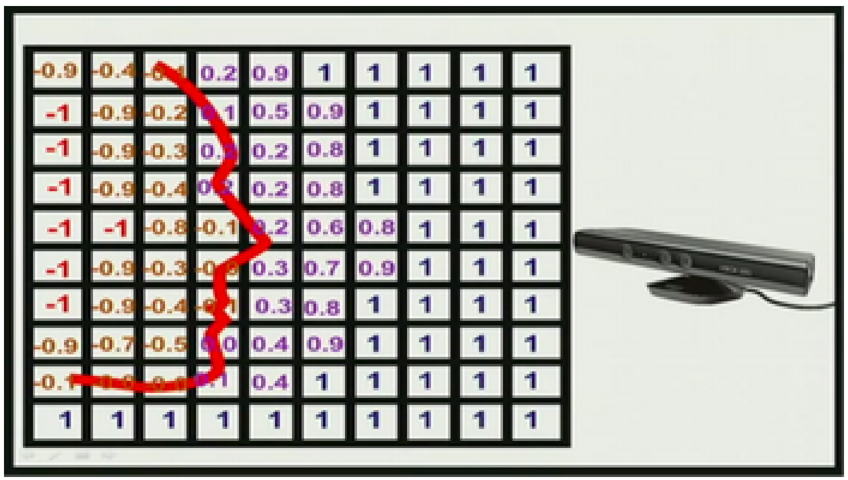
Volumetric Integration

Listing 2 Projective TSDF integration leveraging coalesced memory access.

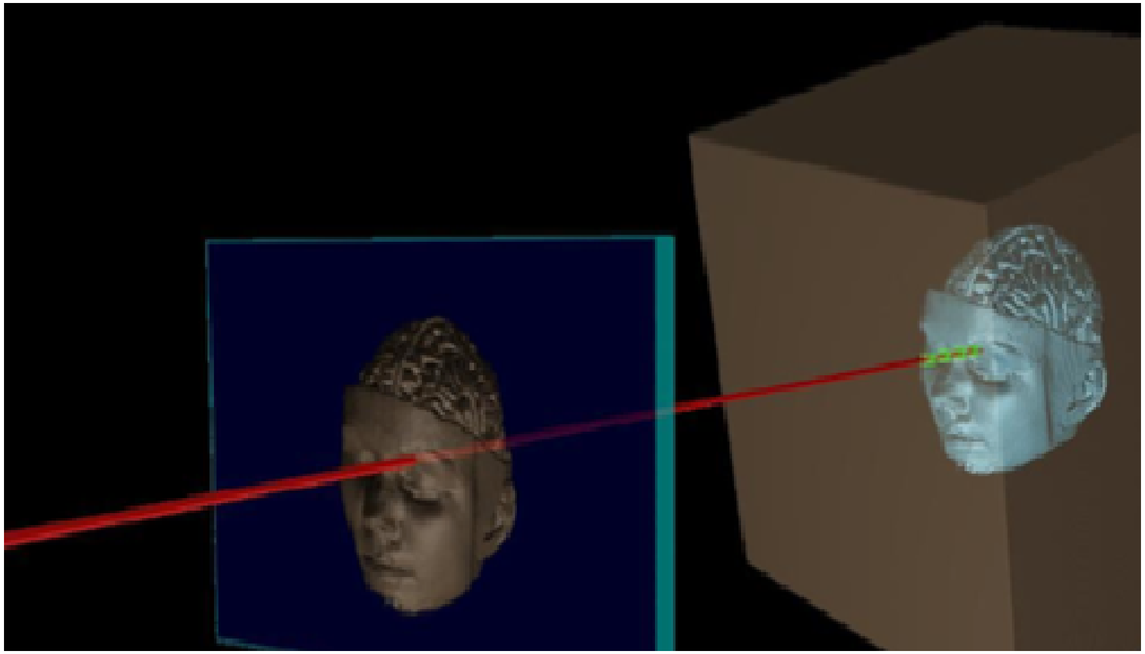
for each voxel **g** in x,y volume slice **in parallel do**
    while sweeping from front slice to back **do**
        $\mathbf{v}^{\mathrm{g}} \leftarrow$ convert **g** from grid to global 3D position
        $\mathbf{v} \leftarrow \mathbf{T}_i^{-1}\mathbf{v}^{\mathrm{g}}$
        $\mathbf{p} \leftarrow$ perspective project vertex $\mathbf{v}$
        if $\mathbf{v}$ in camera view frustum **then**
            $\mathbf{sdf}_i \leftarrow ||\mathbf{t}_i - \mathbf{v}^{\mathrm{g}}|| - \mathbf{D}_i(\mathbf{p})$
            if $(\mathbf{sdf}_i > 0)$ **then**
                $\mathbf{tsdf}_i \leftarrow min(1, \mathbf{sdf}_i/$ max truncation$)$
            else
                $\mathbf{tsdf}_i \leftarrow max(-1, \mathbf{sdf}_i/$ min truncation$)$
            $\mathbf{w}_i \leftarrow$ min(max weight, $\mathbf{w}_{i-1} + 1$)
            $\mathbf{tsdf}^{\mathrm{avg}} \leftarrow (\mathbf{tsdf}_{i-1}\mathbf{w}_{i-1} + \mathbf{tsdf}_i\mathbf{w}_i)/\mathbf{w}_i$
            store $\mathbf{w}_i$ and $\mathbf{tsdf}^{\mathrm{avg}}$ at voxel **g**

Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera (2011)


Y
Z
X


-0.9 -0.4 0.2 0.9 1 1 1 1 1
-1 -0.9 -0.2 0.5 0.9 1 1 1 1
-1 -0.9 -0.3 0.2 0.8 1 1 1 1
-1 -0.9 -0.4 0.2 0.8 1 1 1 1
-1 -1 -0.8 -0.1 0.6 0.8 1 1 1
-1 -0.9 -0.3 0.3 0.7 0.9 1 1 1
-1 -0.9 -0.4 0.3 0.8 1 1 1 1
-0.9 -0.7 -0.5 0.4 0.9 1 1 1 1
0.4 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1

Raycasting



Video

3D Reconstruction with Deep Learning

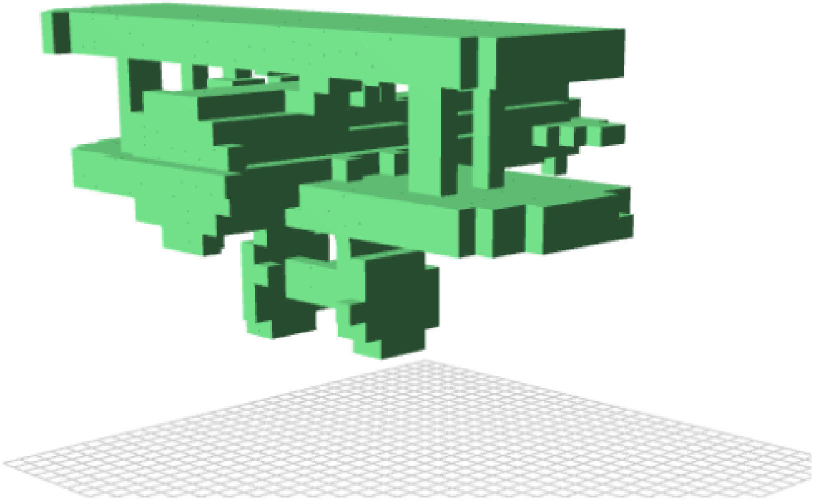
- Dataset
- 3D Shape Representations
- 3D CNN for Feature Extraction
- Deep Generative Models for 3D Reconstruction

ShapeNet
https://www.shapenet.org

- 55 common object categories with about 51,300 unique 3D models.

Input:

ground truth:

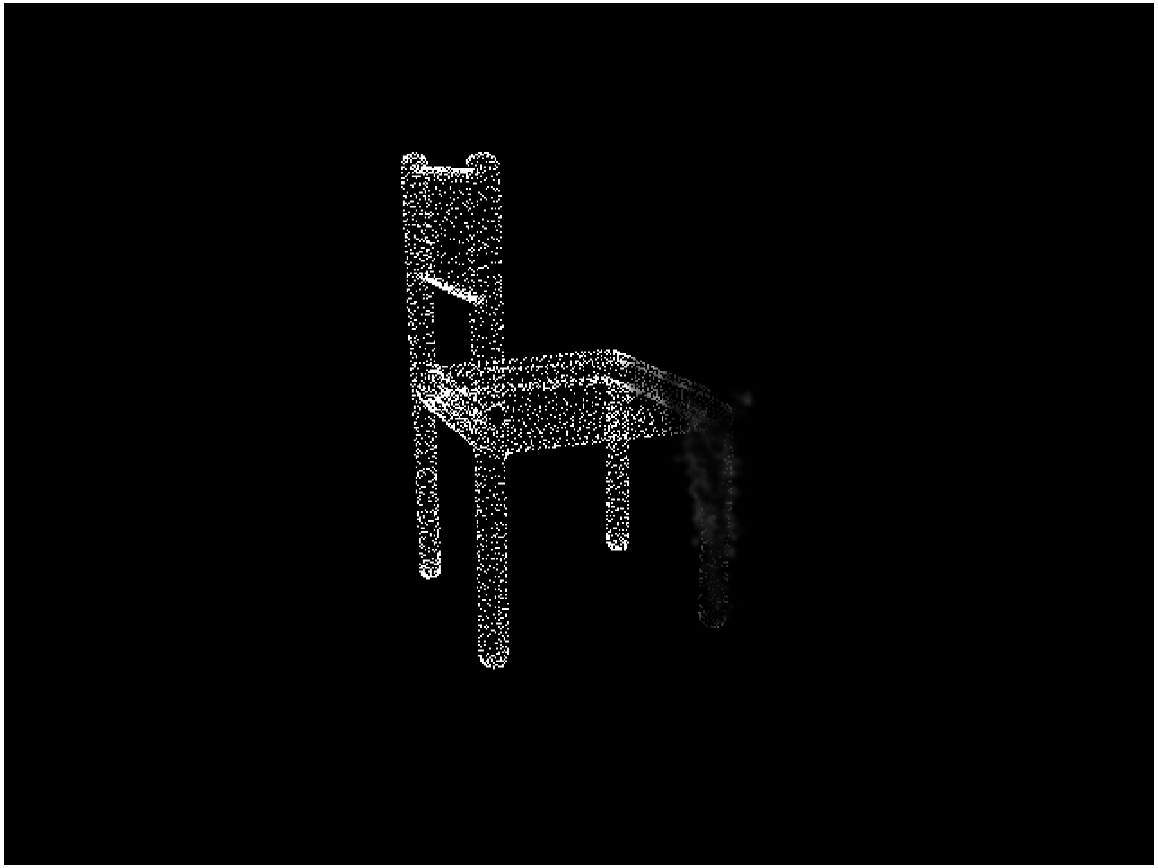


ScanNet
http://www.scan-net.org

- ScanNet is an RGB-D video dataset containing 2.5 million views in more than 1500 scans, annotated with 3D camera poses, surface reconstructions, and instance-level semantic segmentations.

Princeton ModelNet
http://modelnet.cs.princeton.edu/

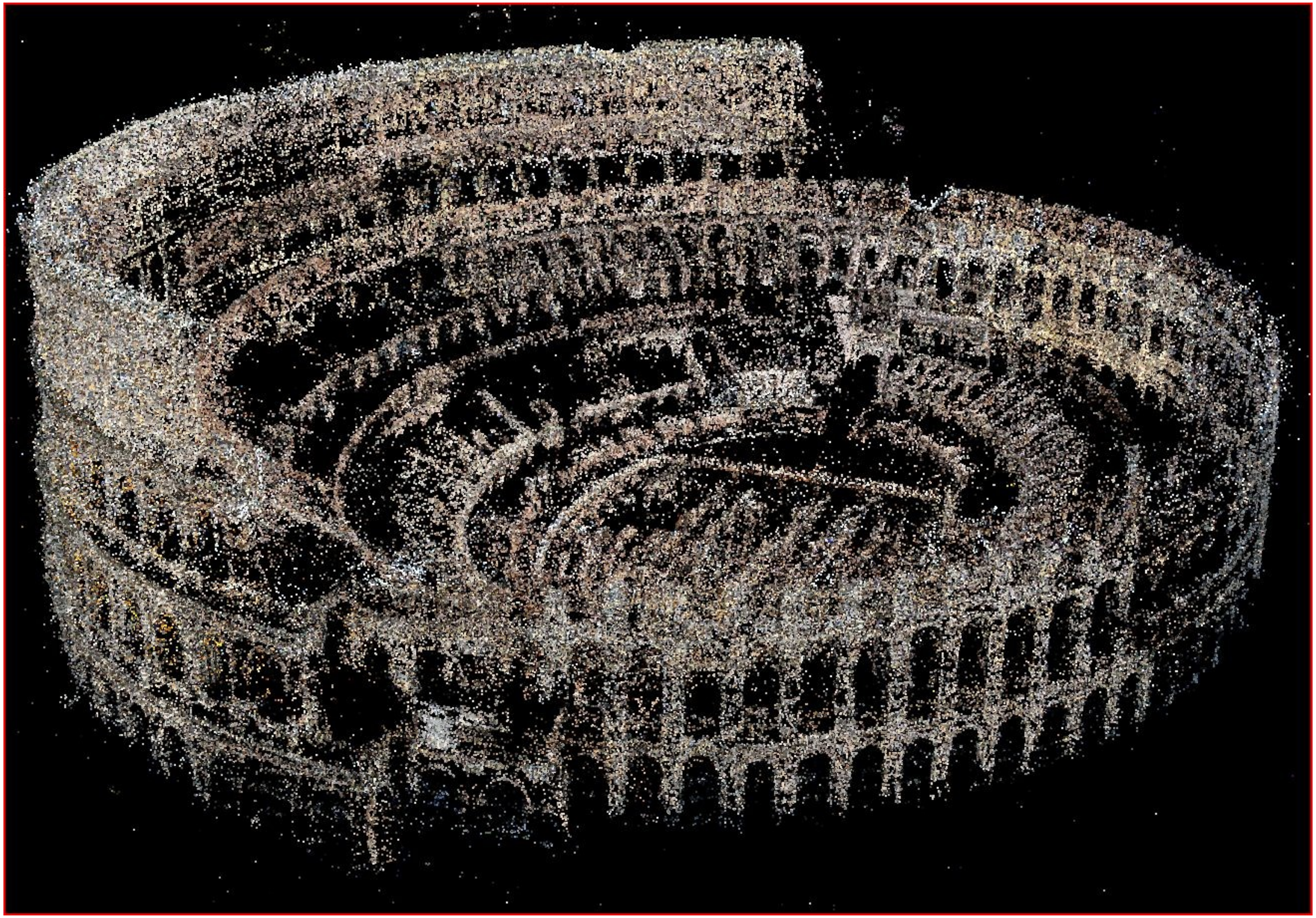
- 127,915 CAD Models
- 662 Object Categories
- 10 Categories with Annotated Orientation



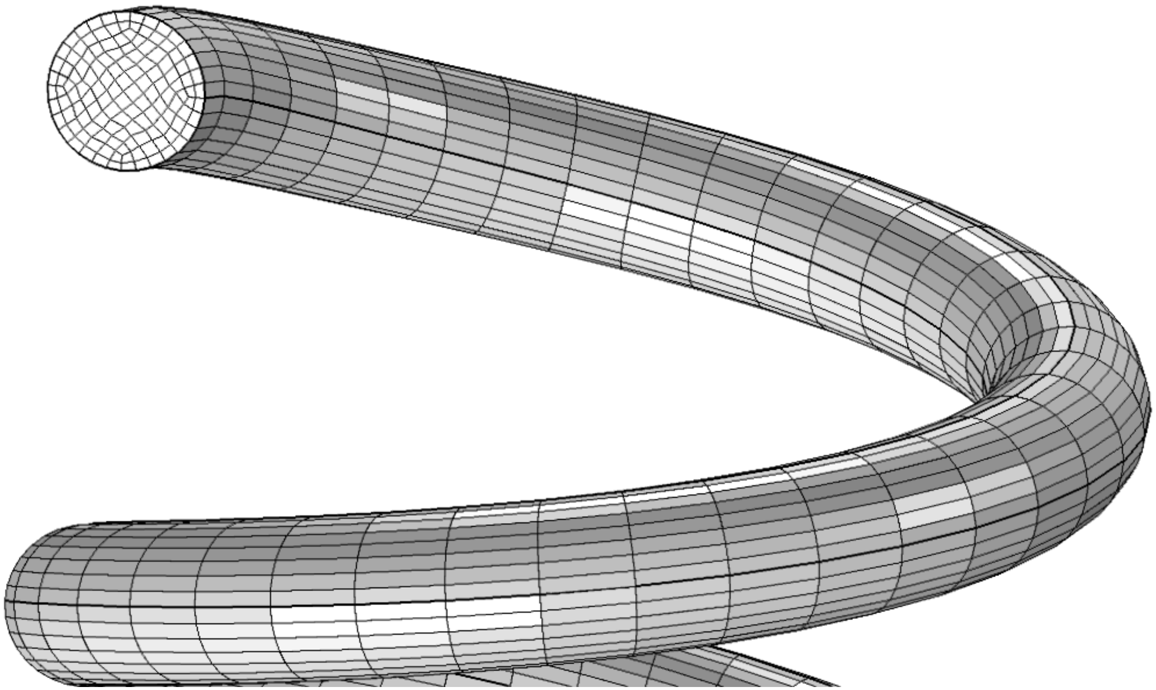
3D Shape Representations

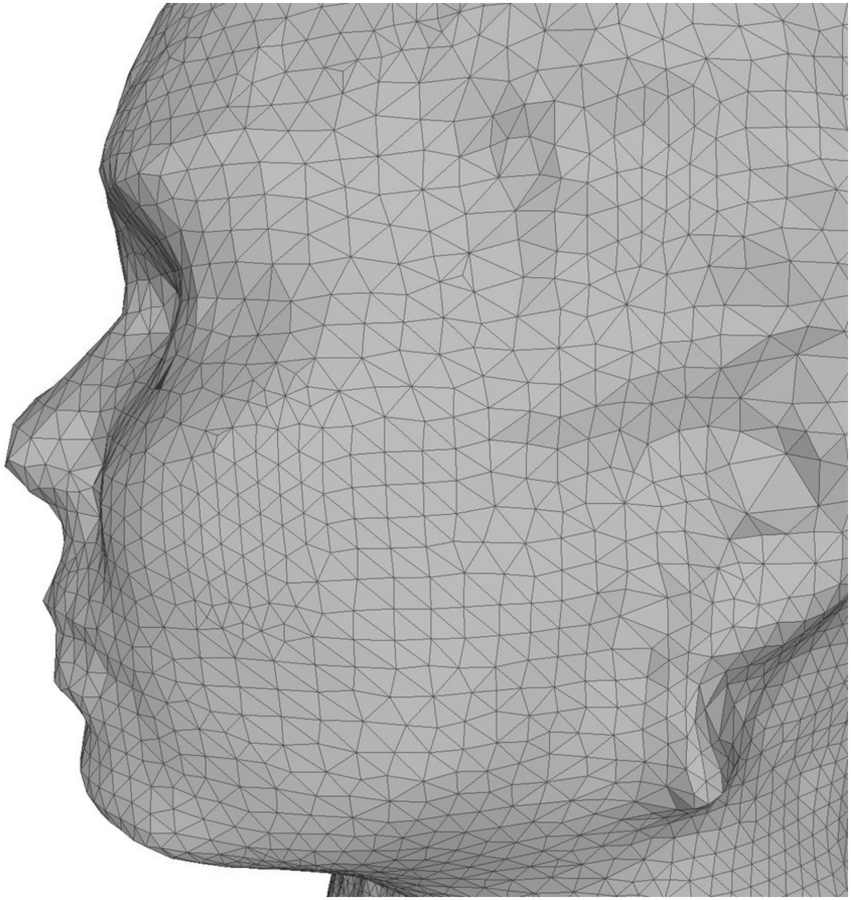
- Goal: Describe 3D shape in efficient way.

3D shape representations: point clouds

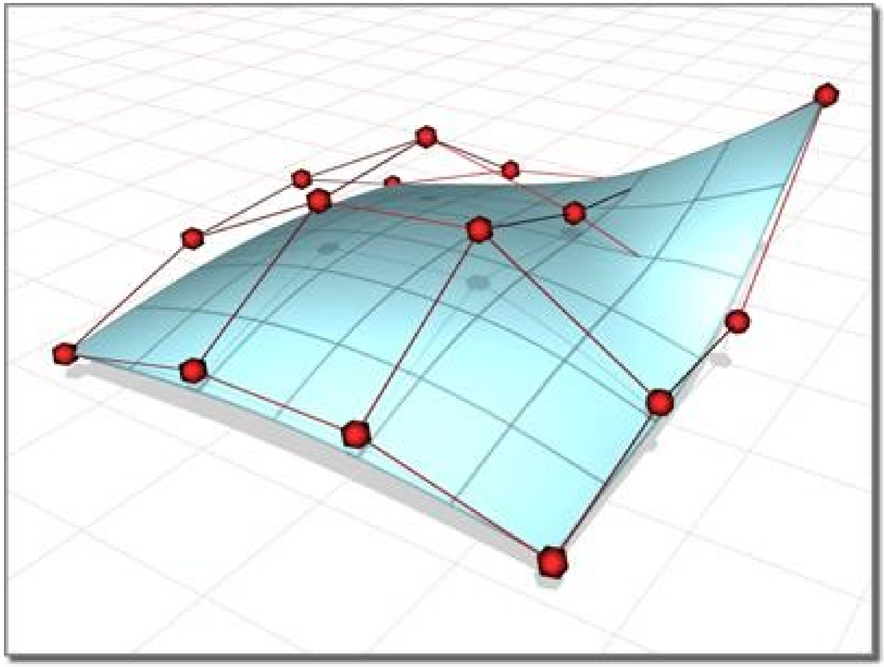


3D shape representations: mesh

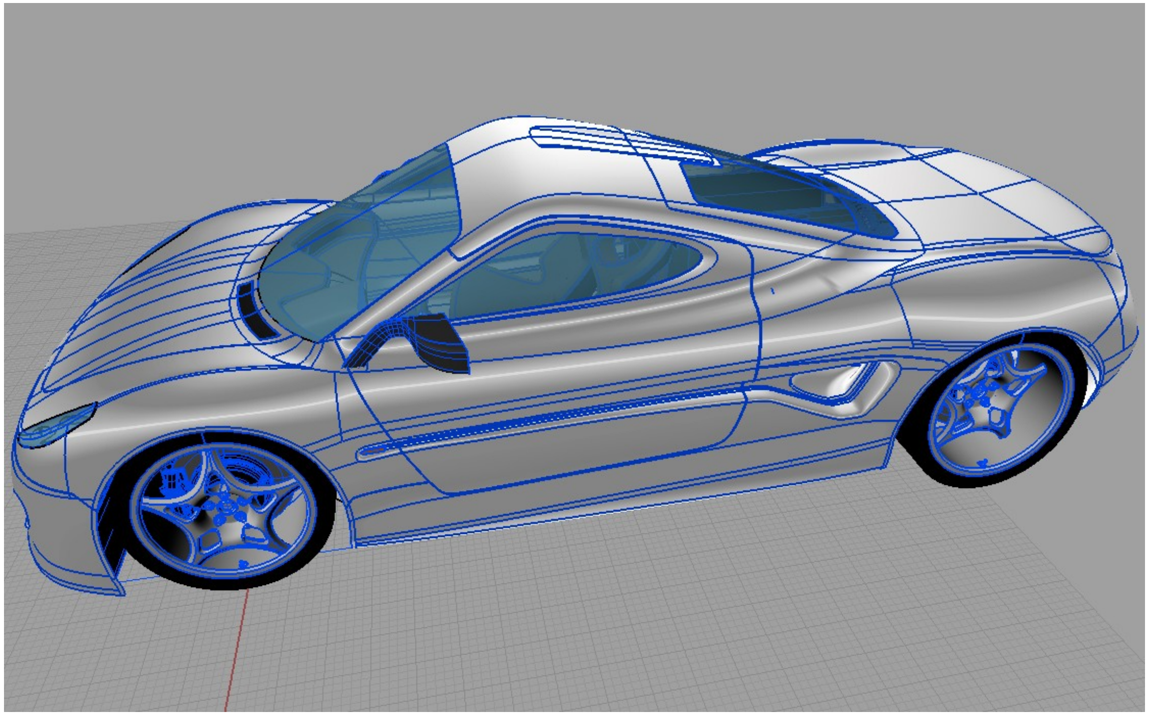




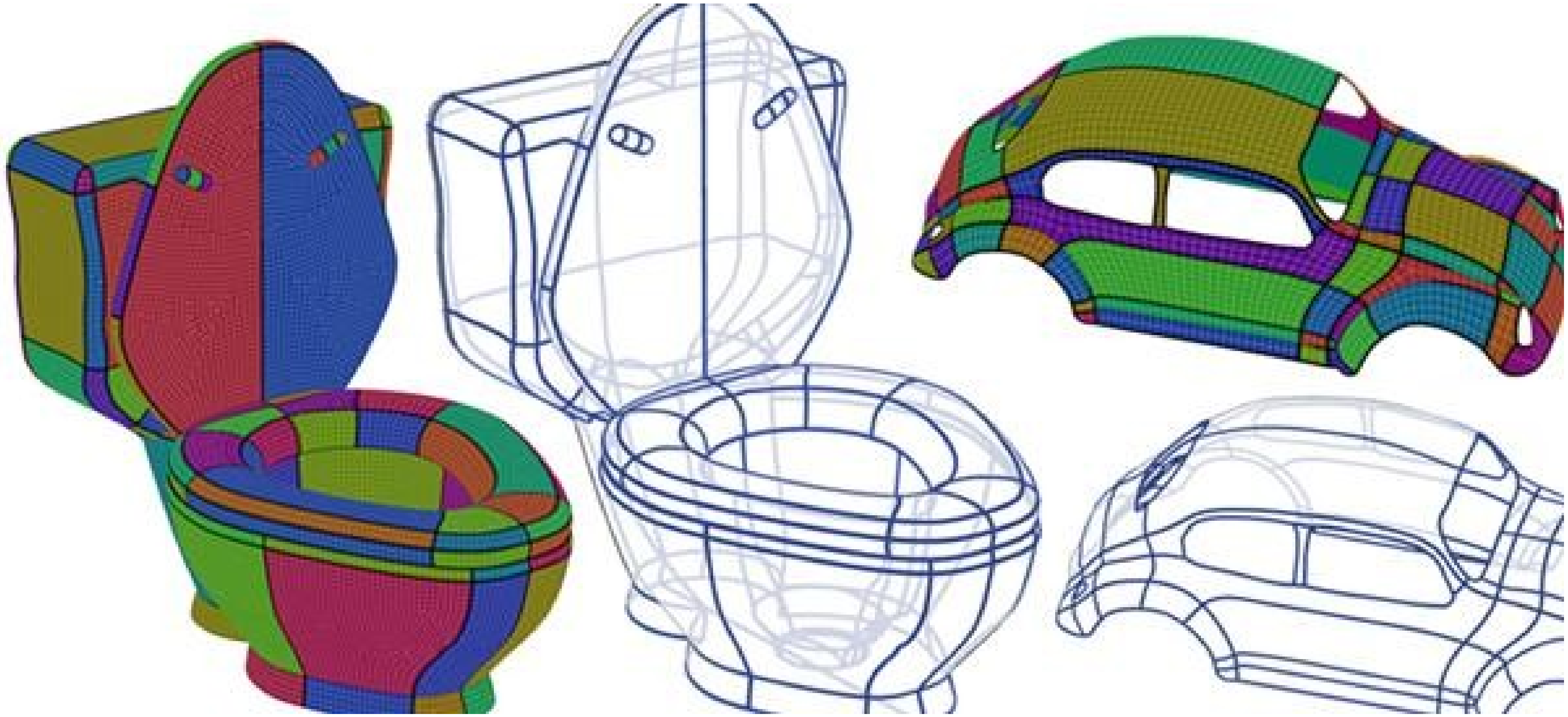
3D shape representations: NURBS Surface



3D shape representations: NURBS Surface

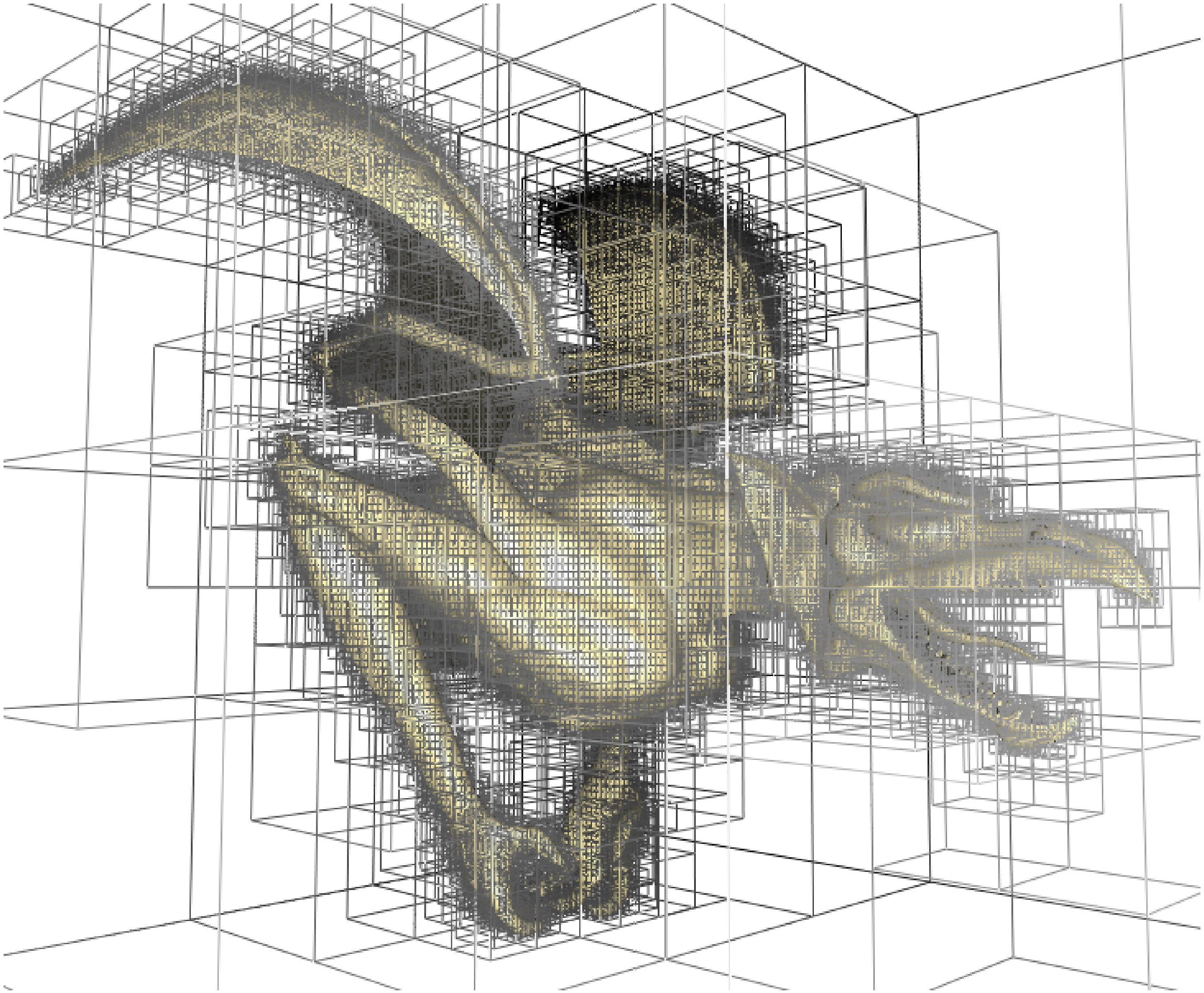


3D shape representations: curve networks



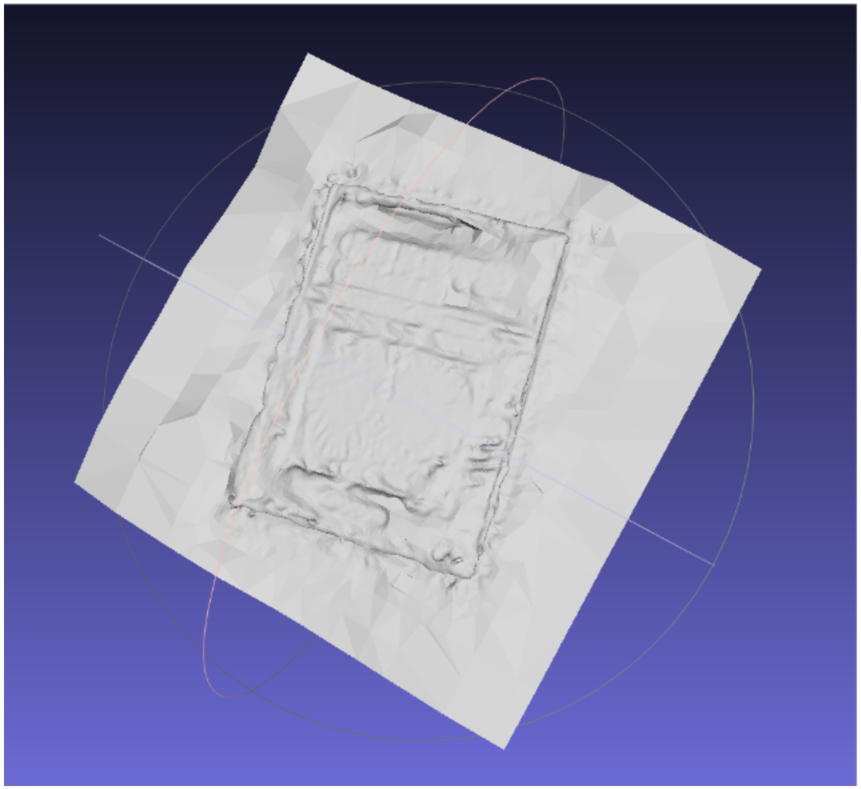
3D shape representations: voxel

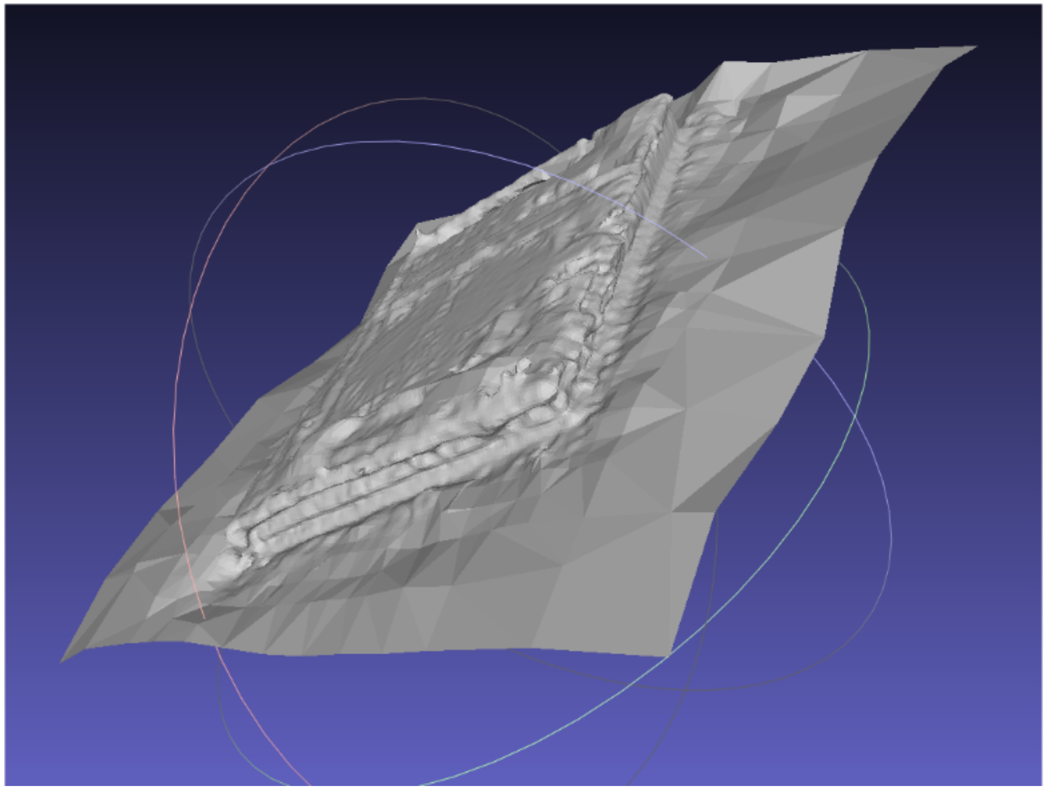
3D shape representations: octree

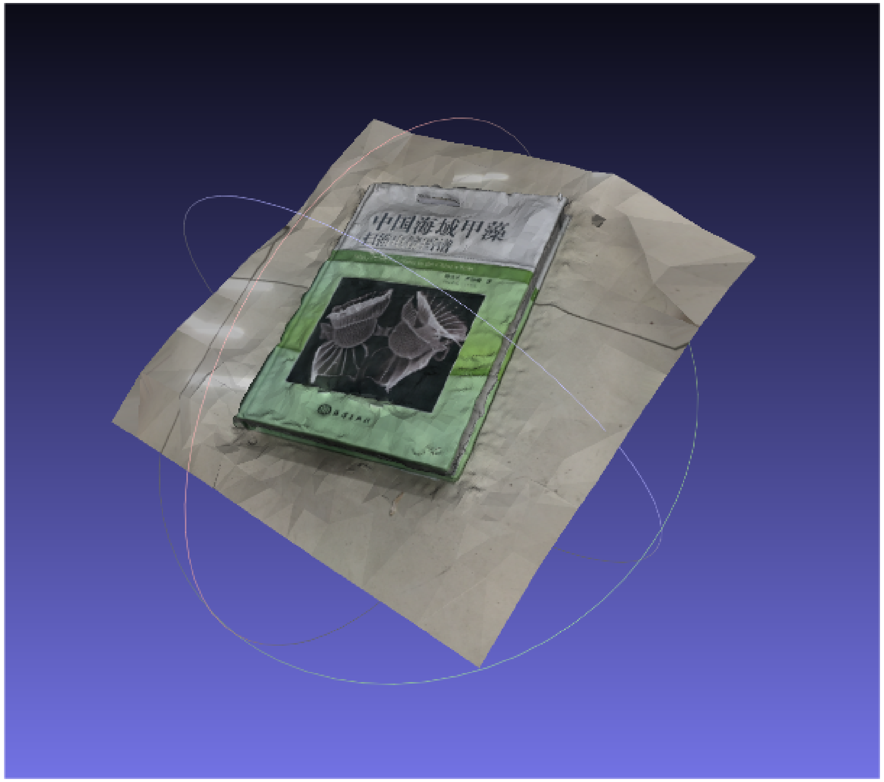
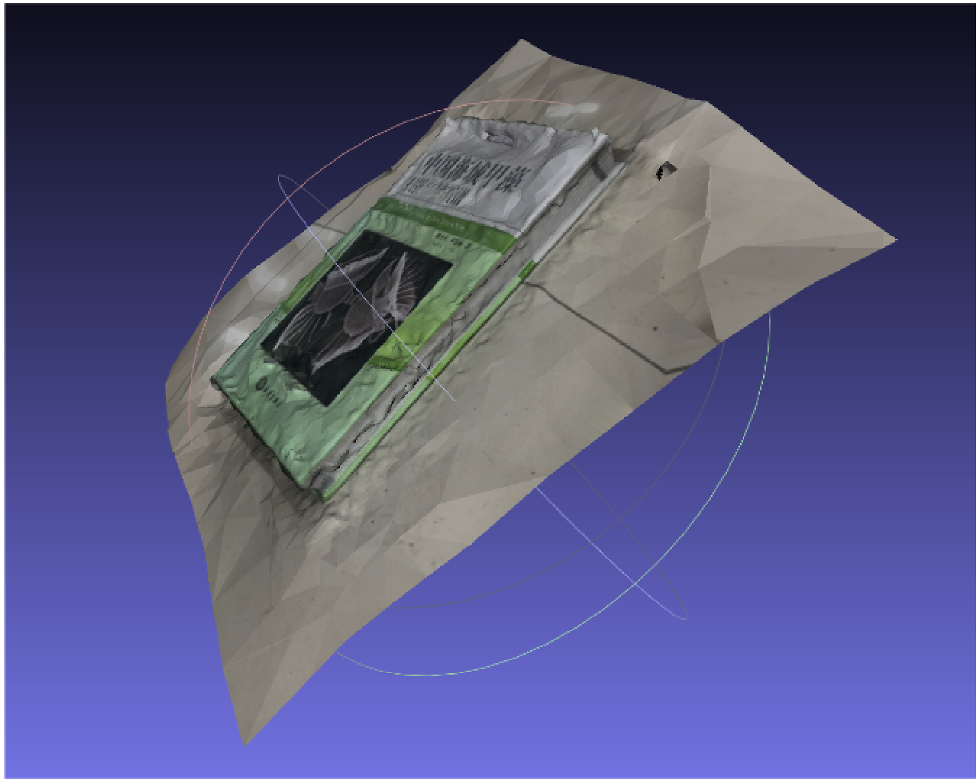


3D shape representations: depth maps

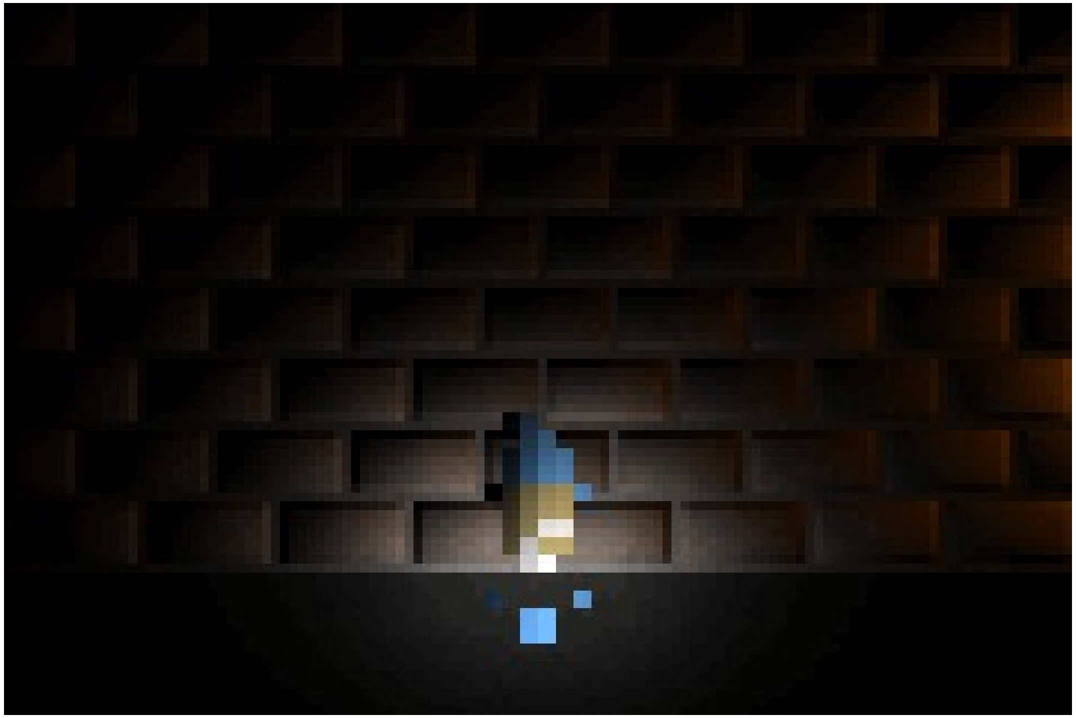
3D shape representations: normal maps





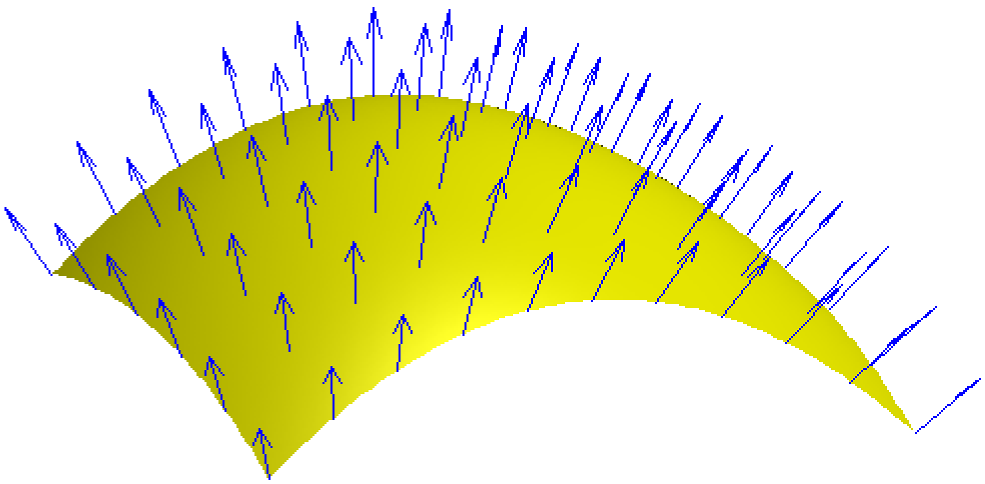

中国海域甲藻

中国海域甲藻

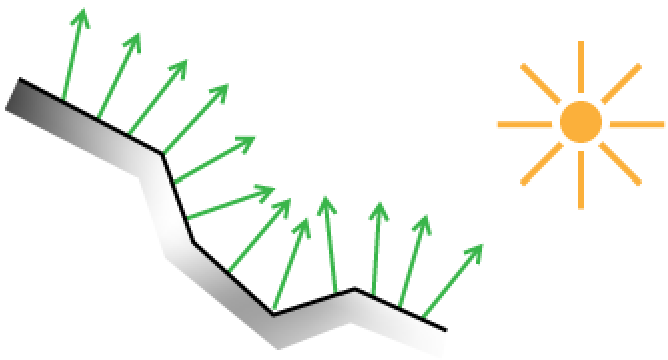
3D shape representations: normal maps



https://www.cnblogs.com/freeblues/p/5742956.html

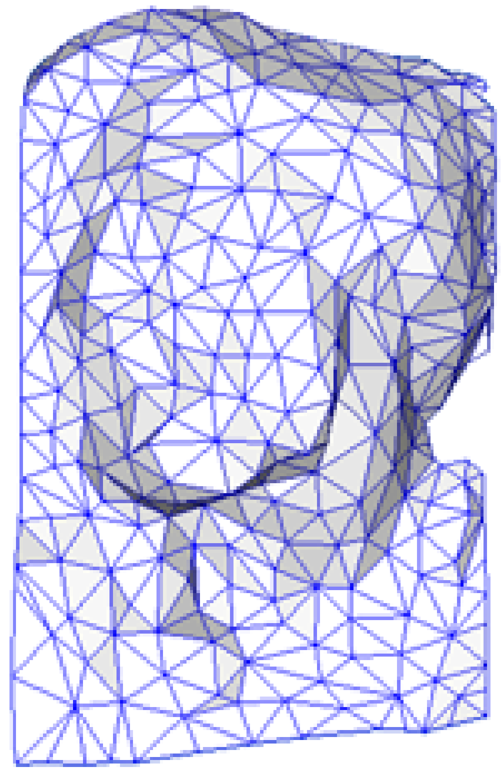
3D shape representations: normal maps



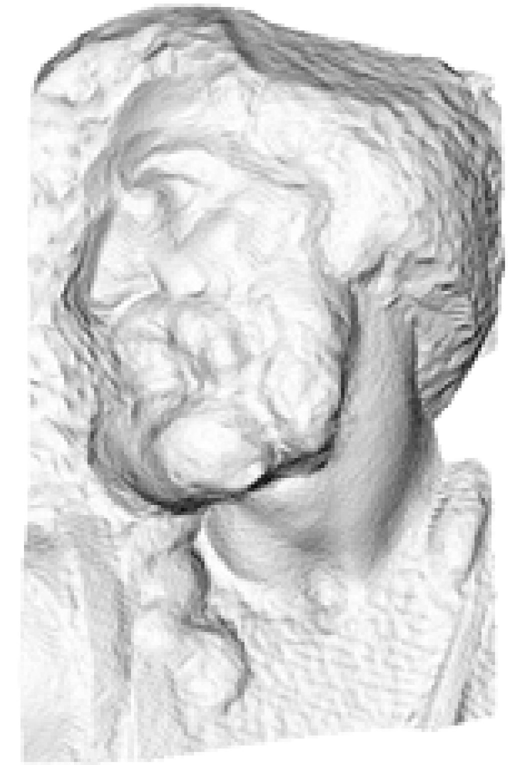


https://www.cnblogs.com/freeblues/p/5742956.html

Original mesh 4M triangles



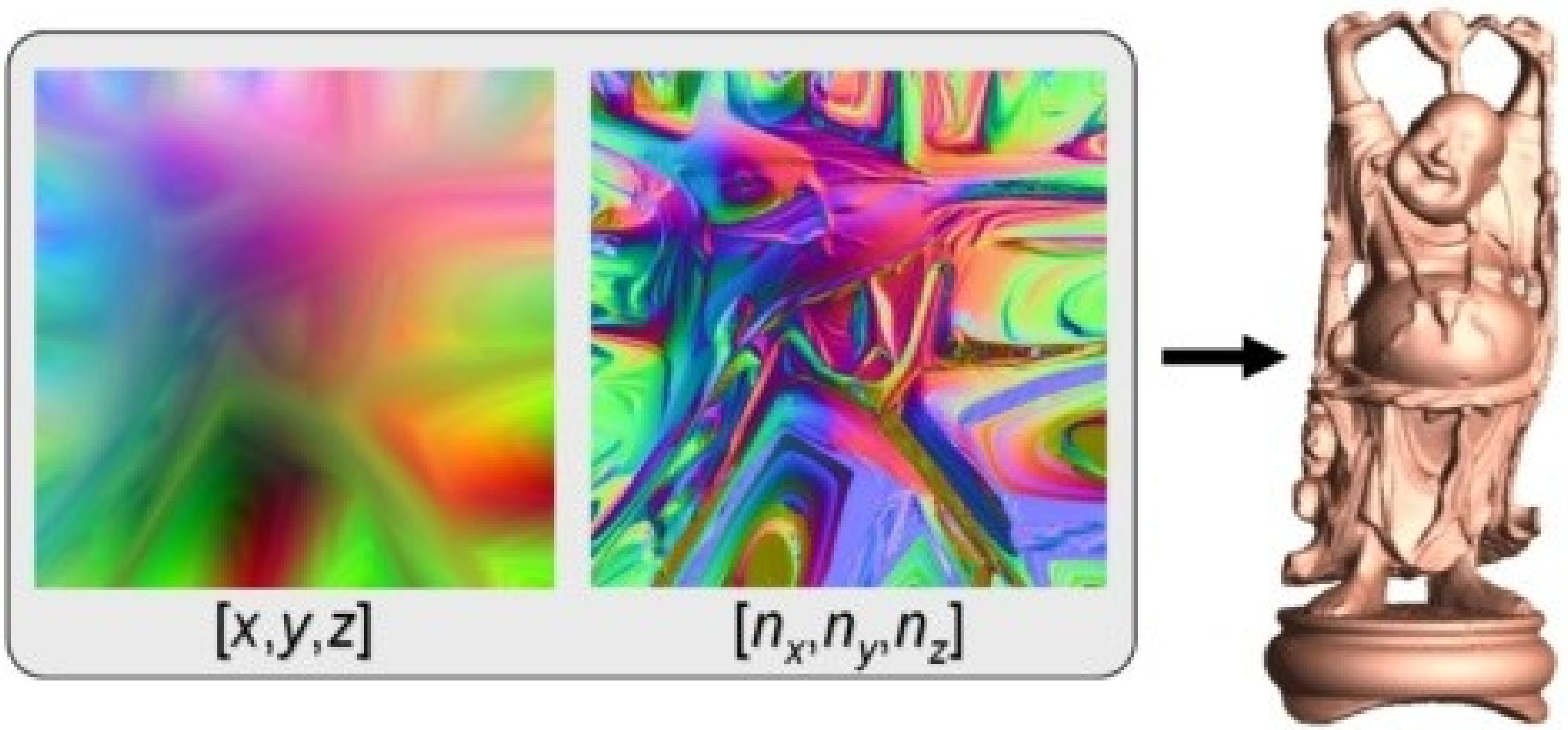
simplified mesh 500 triangles

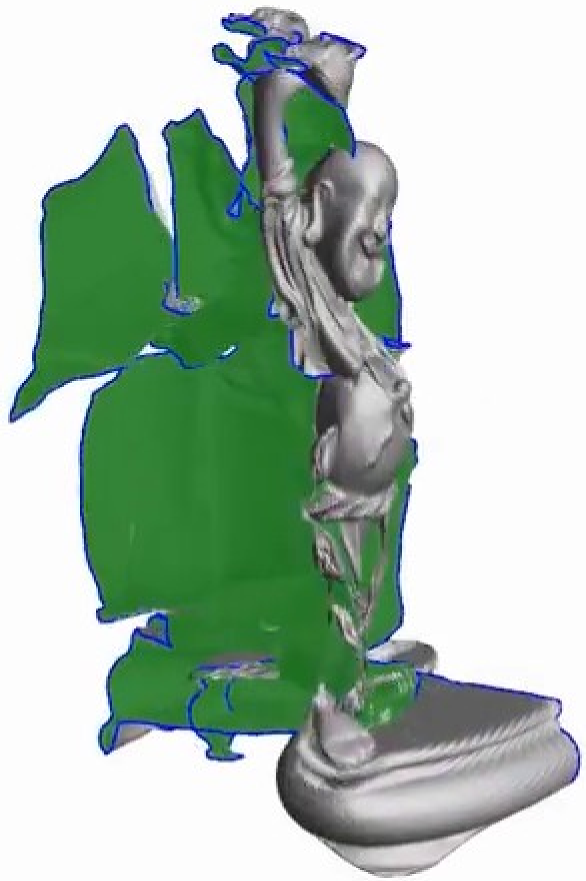


simplified mesh and normal mapping 500 triangles

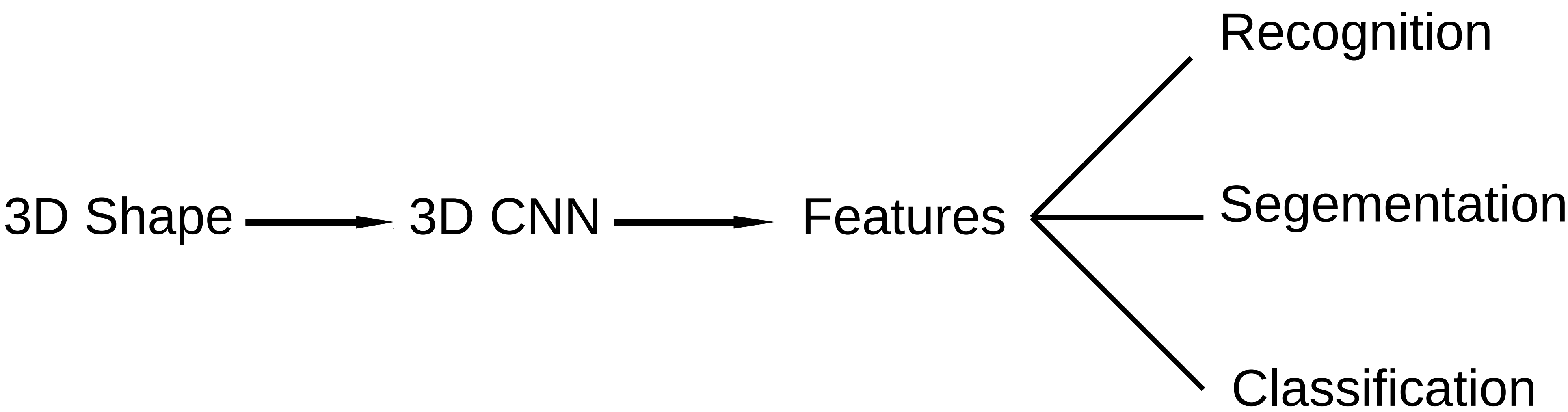
https://www.cnblogs.com/freeblues/p/5742956.html

3D shape representations: geometry image

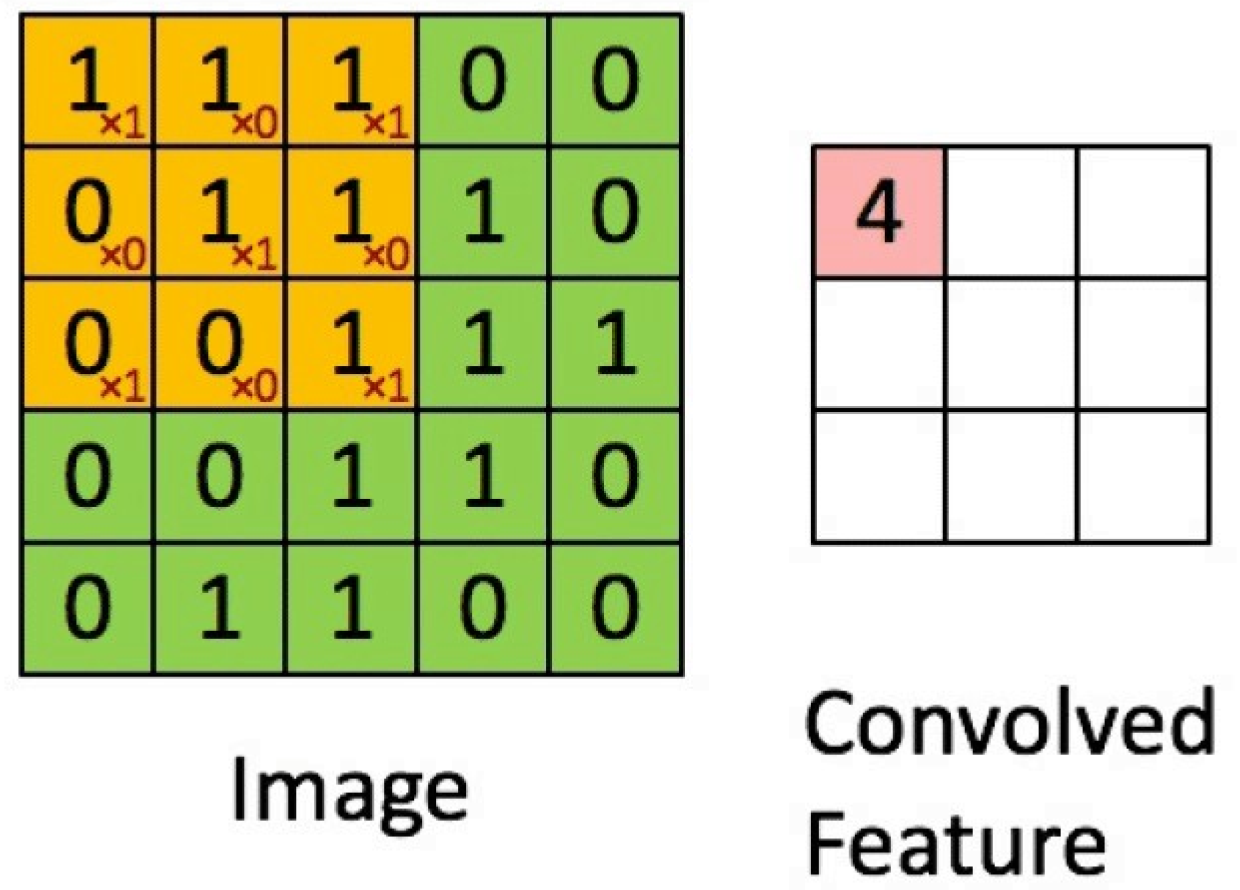




3D Convolutional Neural Network for Feature Extraction

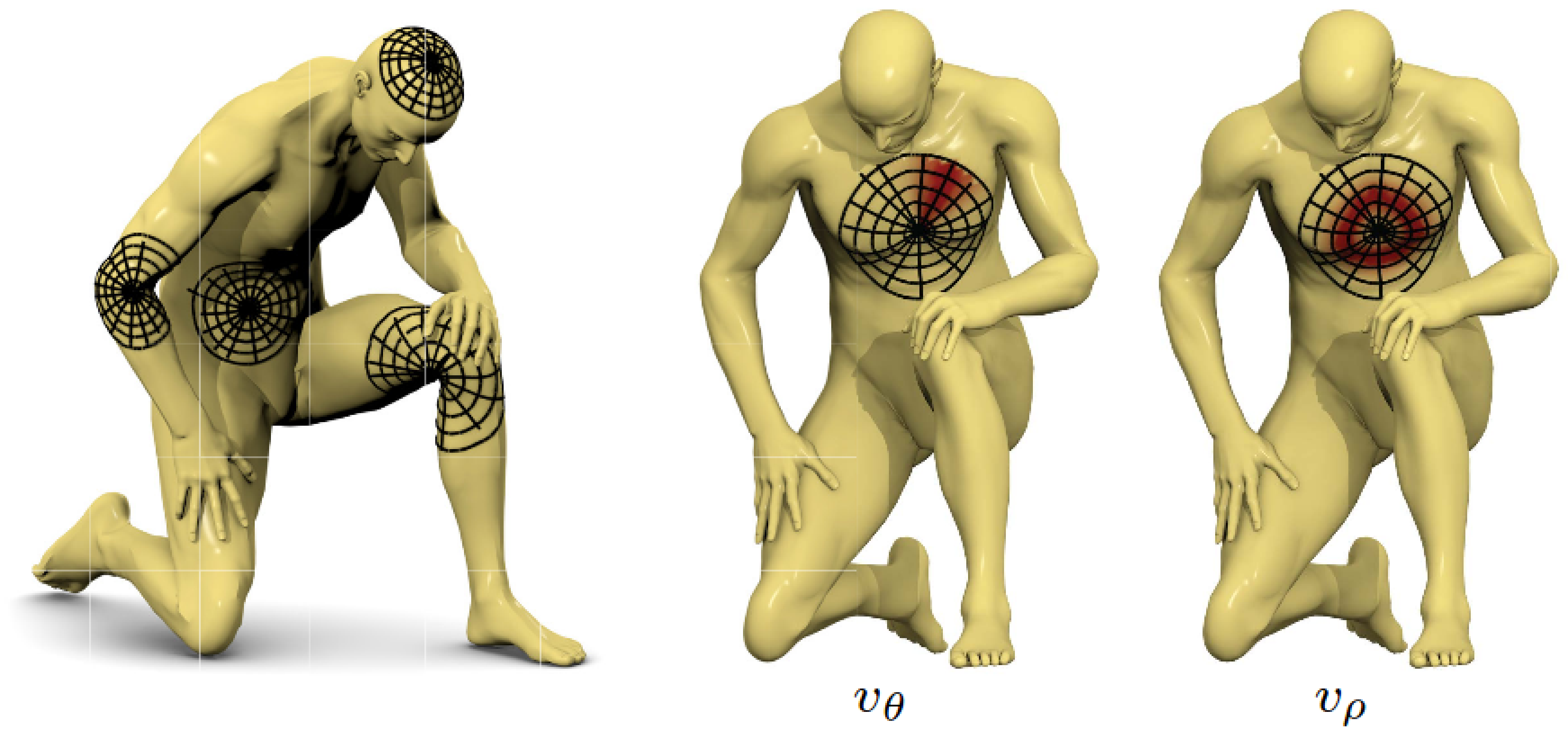


Review: 2D CNN



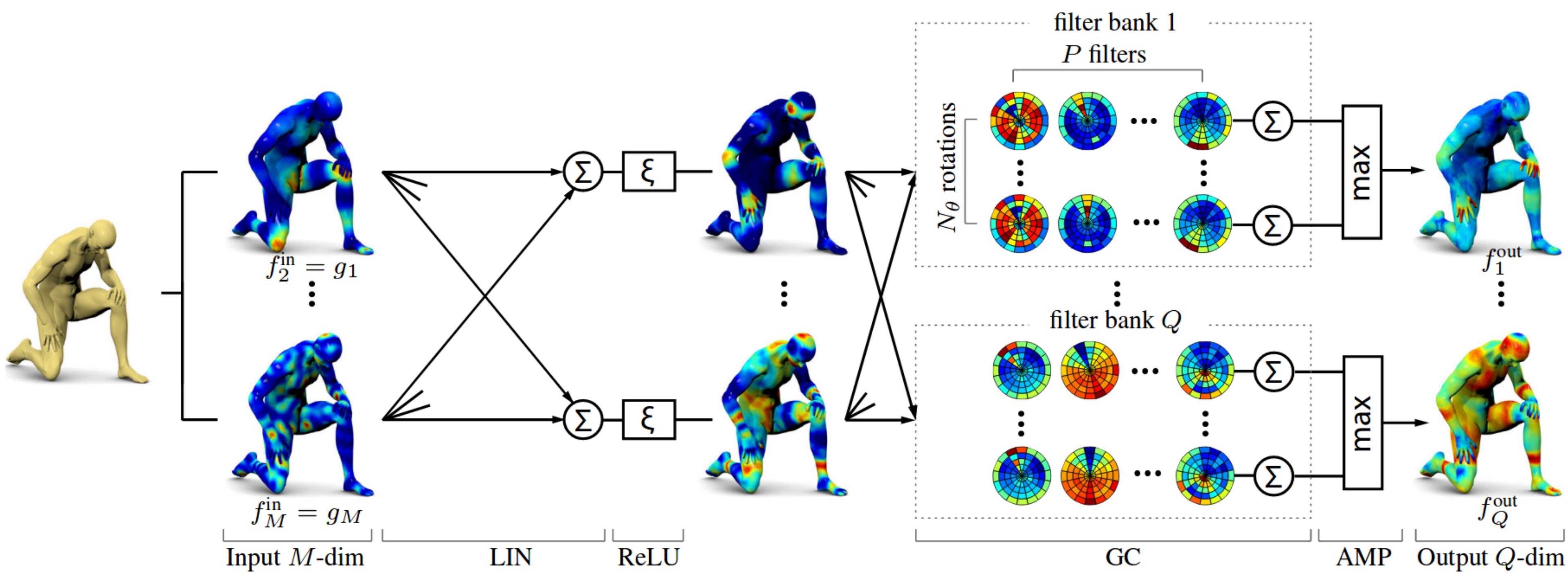
http://i-systems.github.io/HSE545/machine%20learning%20all/KIMM/07_KIMM_CNN.html

3D CNN: surface



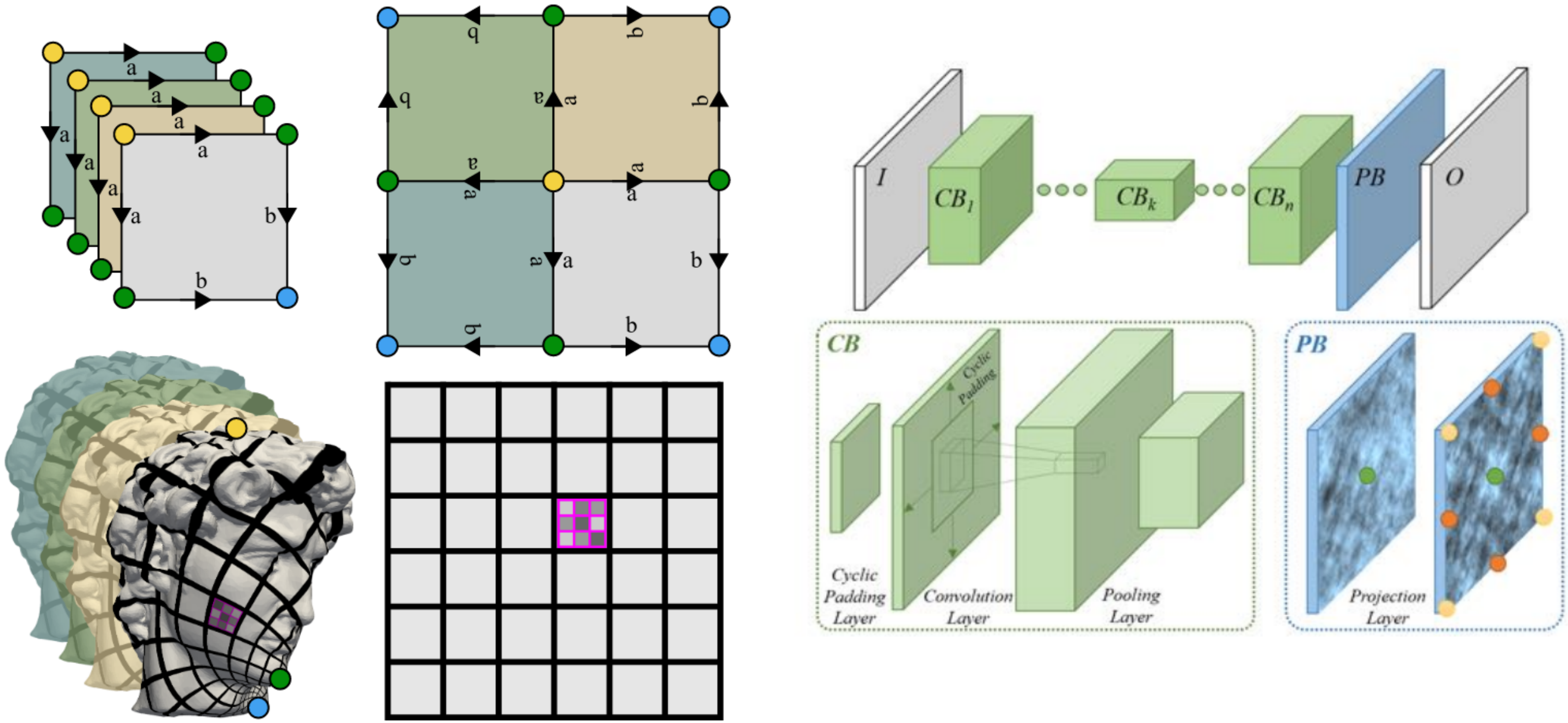
[3dRR 2015] Geodesic convolutional neural networks on Riemannian mainfolds

3D CNN: surface



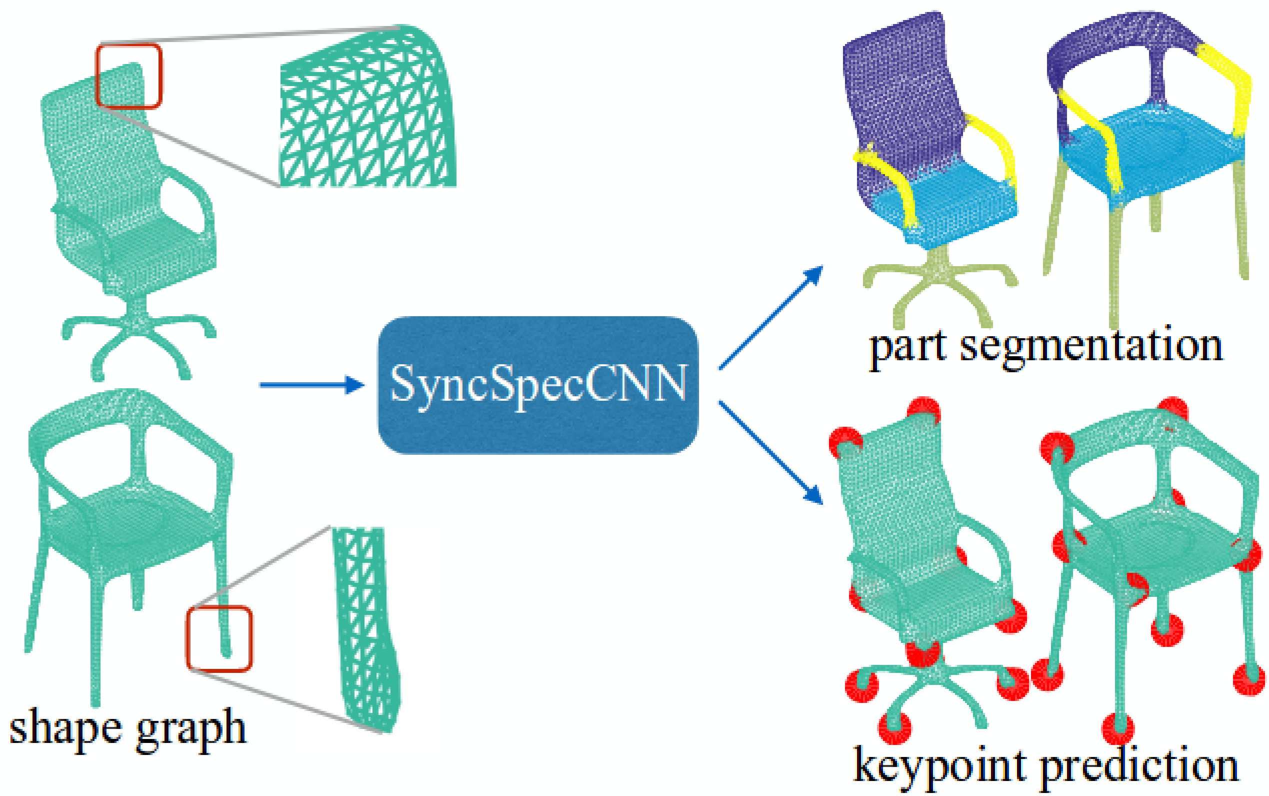
[3dRR 2015] Geodesic convolutional neural networks on Riemannian mainfolds

3D CNN: surface



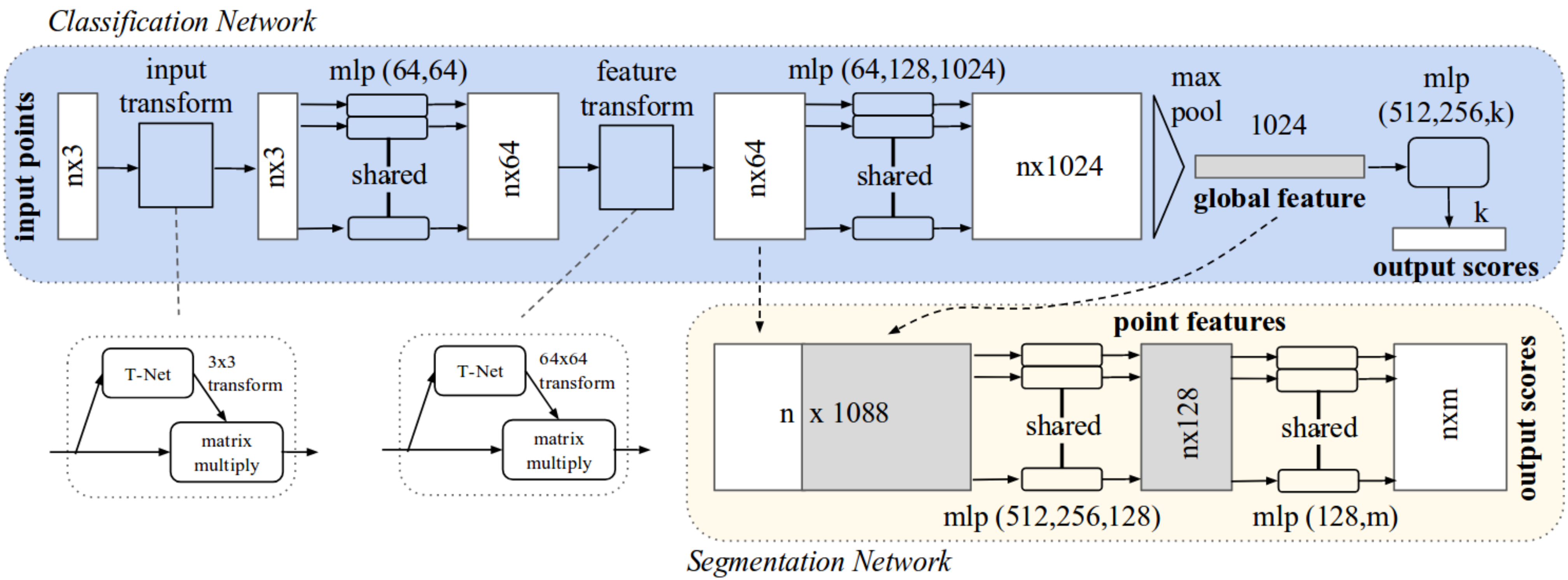
[SIGGRAPH 2017] Conwolutional Neural Networks on Surfaces via Seamless Toric Covers

3D CNN: mesh



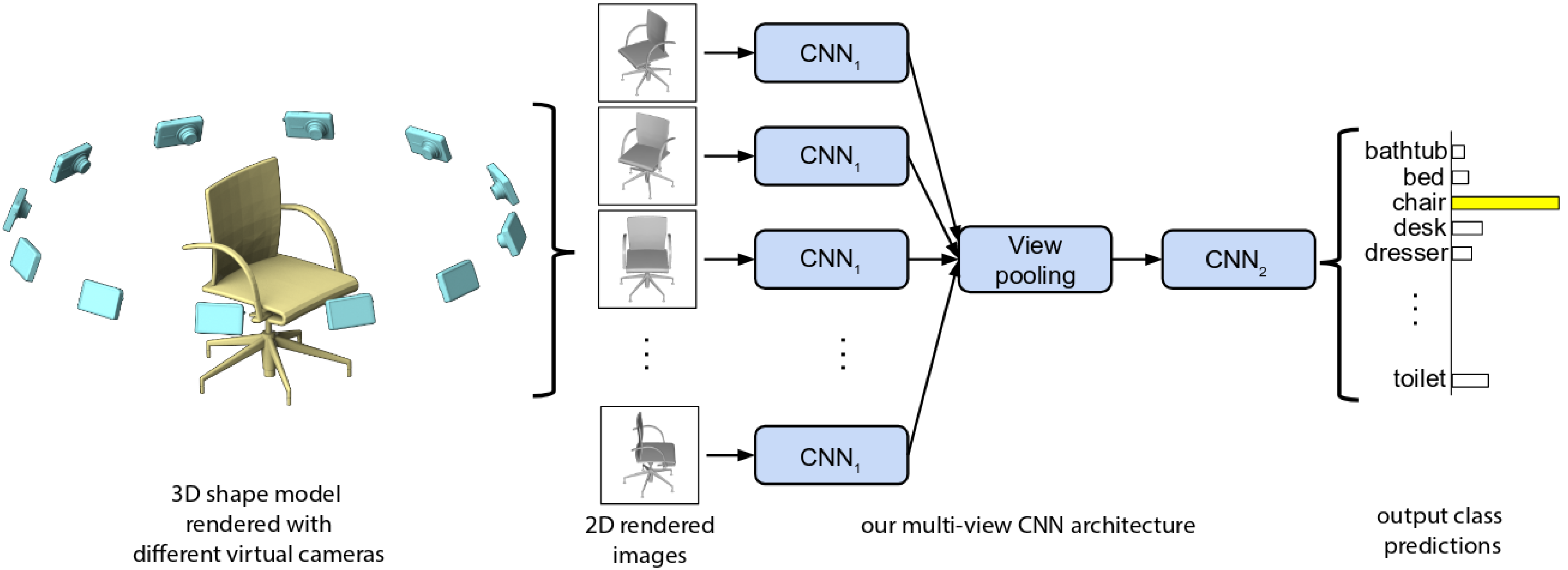
[CVPR 2017 Spotlight] SyncSpecCNN: Synchronized Specral CNN for 3D Shape Segmentation

3D CNN: point cloud



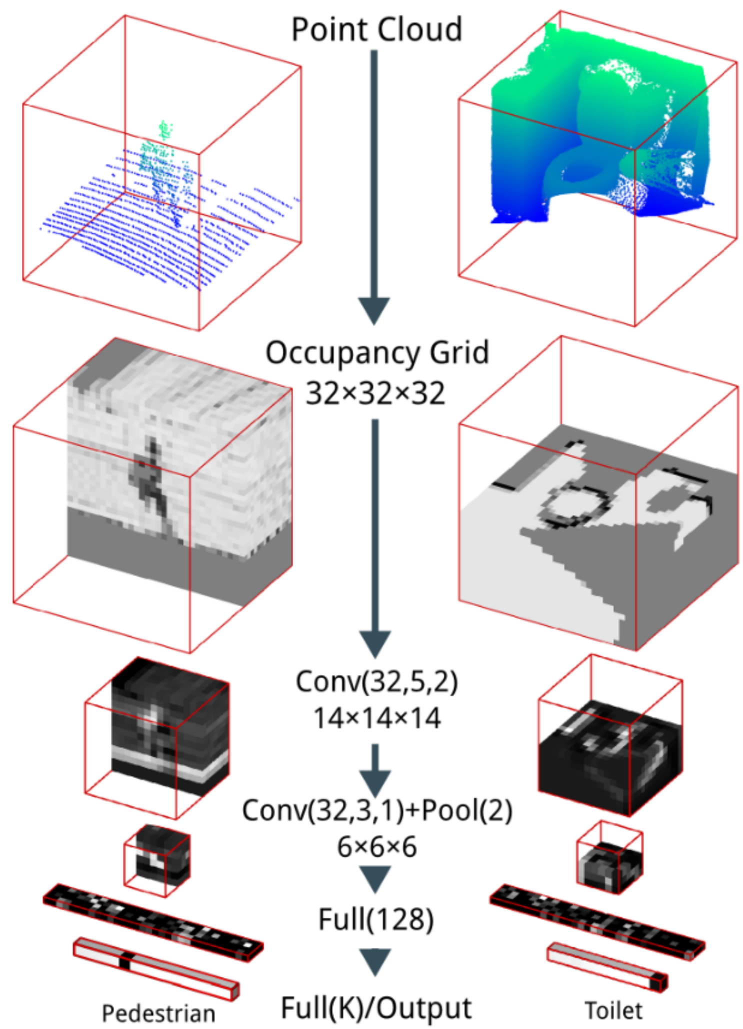
[CVPR 2017 Oral] PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation

3D CNN: multi-view images



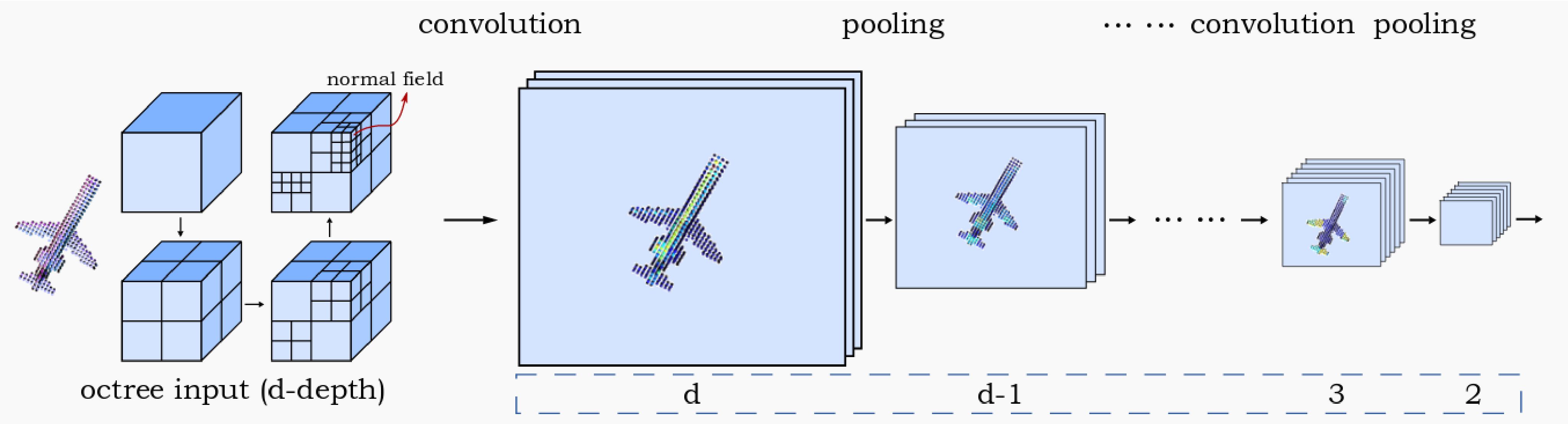
[ICCV 2015] Multi-view Convolutional Neural Networks for 3D Shape Recognition

3D CNN: voxel



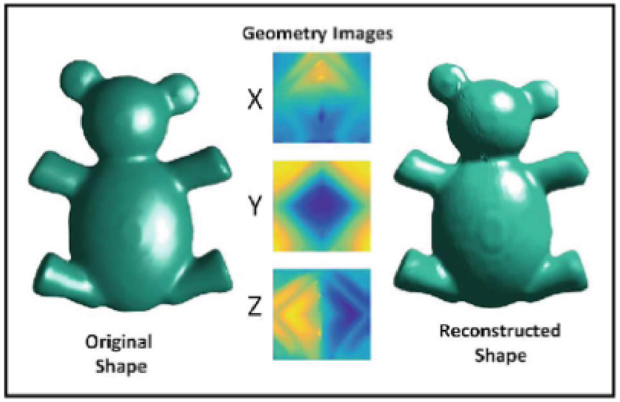
[IROS 2015] VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition

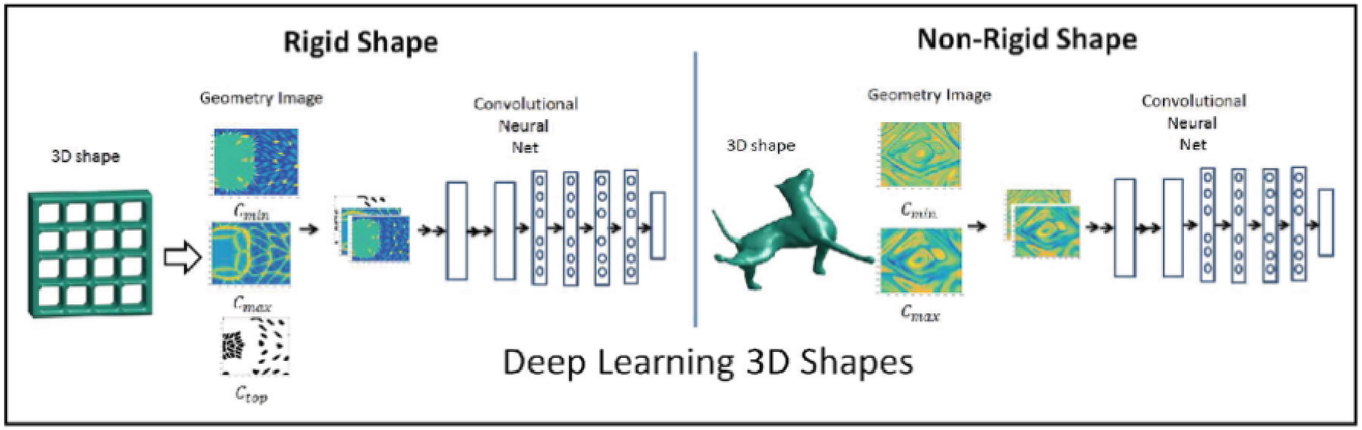
3D CNN: octree



[SIGGRAPH 2017] O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis

3D CNN: geometry image



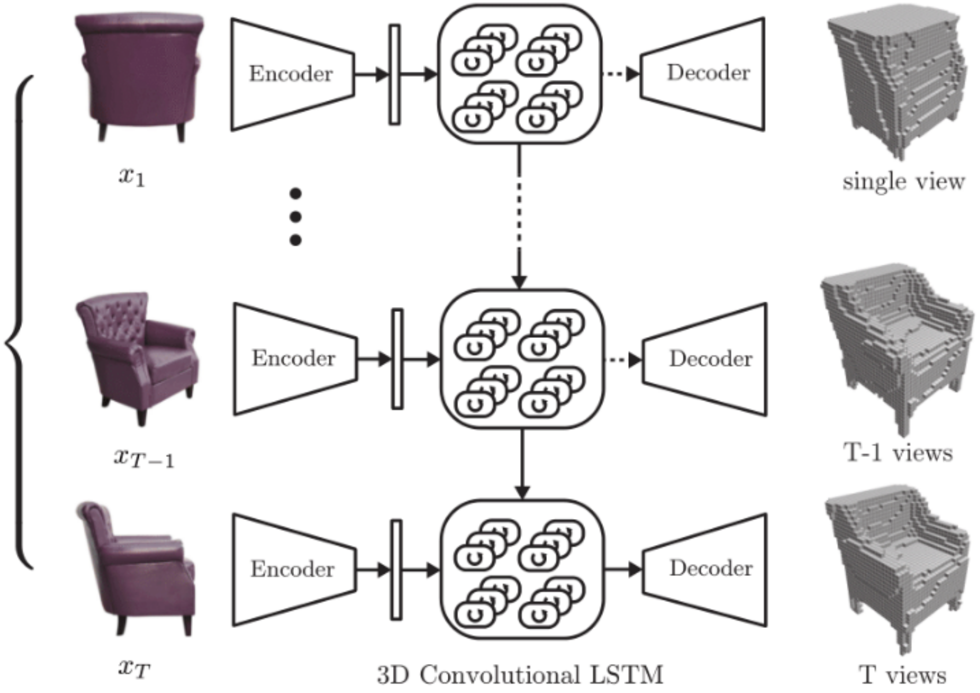


[ECCV 2016] Deep Learning 3D Shape Surfaces Using Geometry Images

Deep Generative Models for 3D Reconstruction

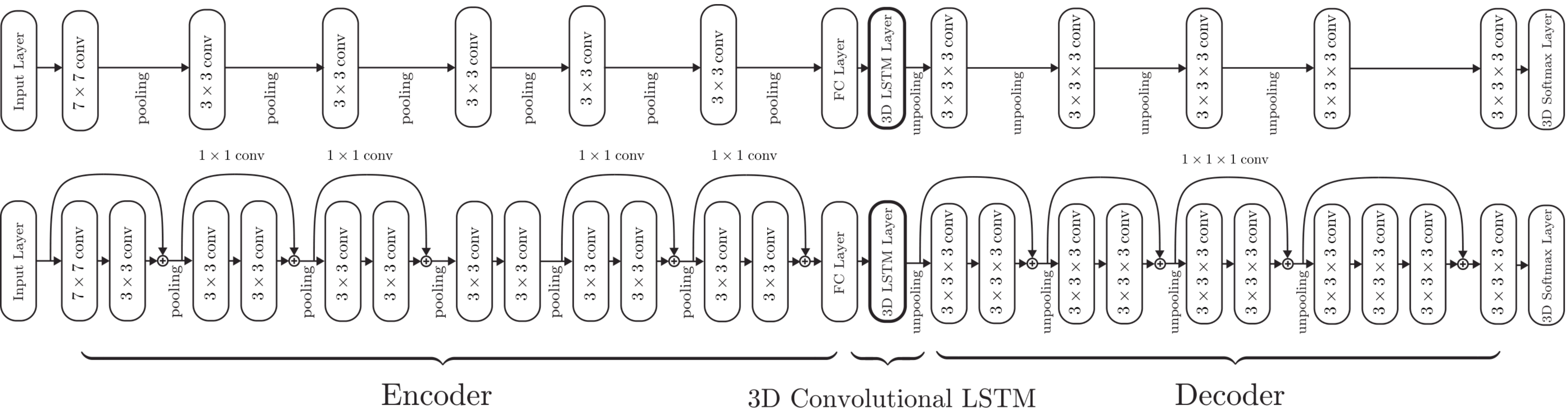
2D Images → Deep Generative Models → 3D Shape

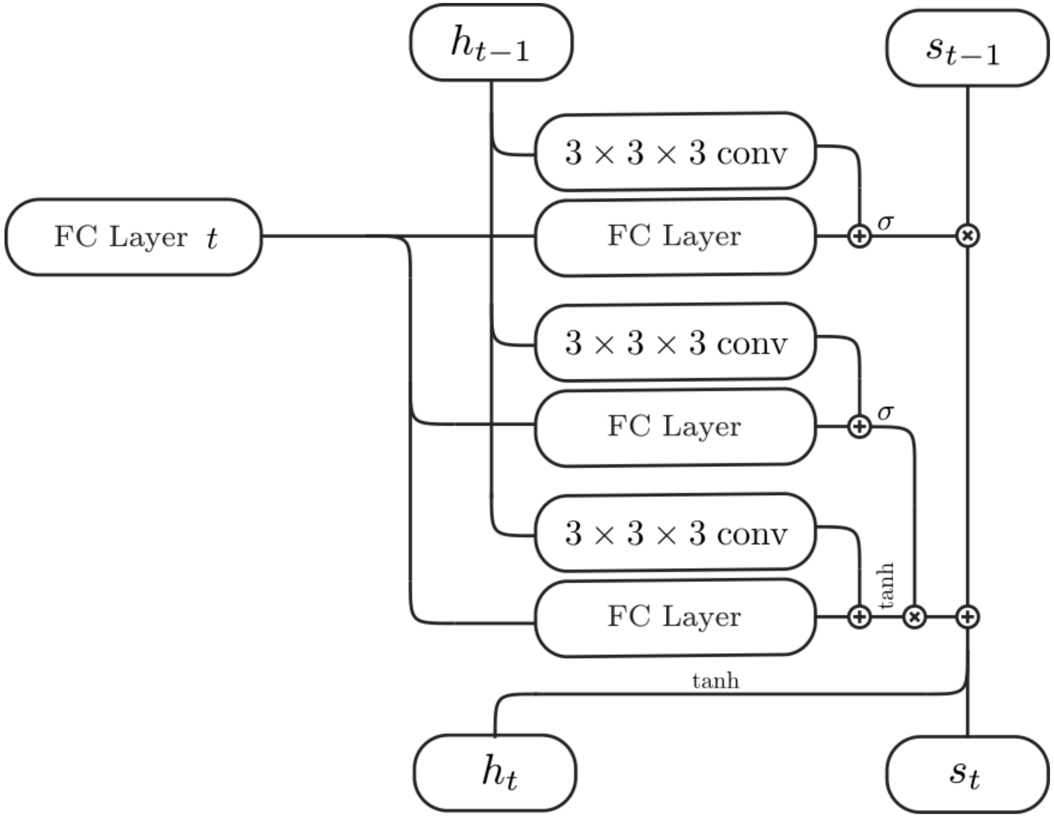
3D Generative Model: voxel



[ECCV 2016] 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction

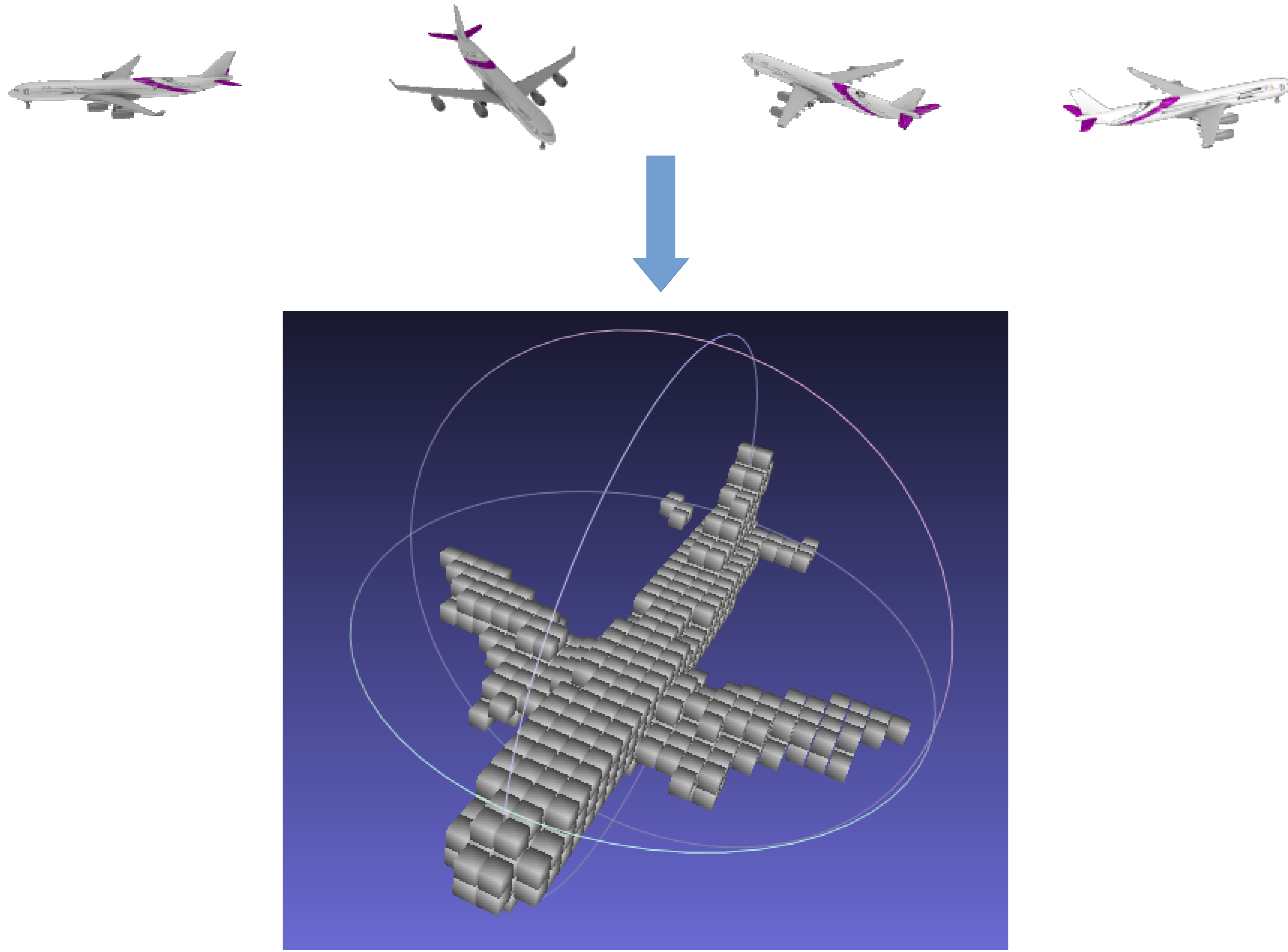
Network Architecture





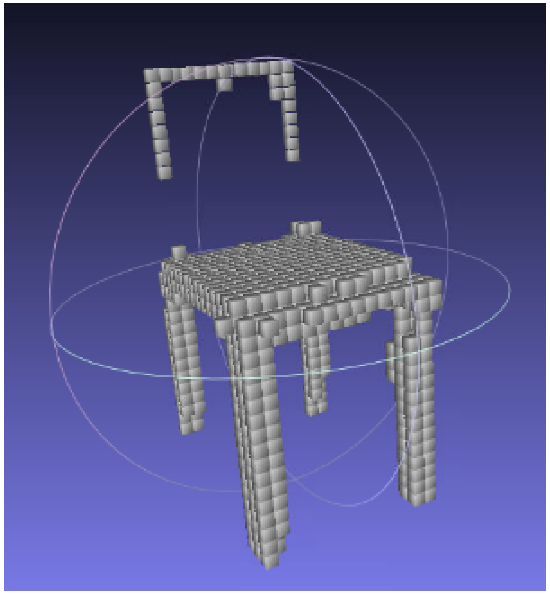
3D Convolutional LSTM

$$f_t = \sigma(W_f \mathcal{T}(x_t) + U_f * h_{t-1} + b_f)$$
$$i_t = \sigma(W_i \mathcal{T}(x_t) + U_i * h_{t-1} + b_i)$$
$$s_t = f_t \odot s_{t-1} + i_t \odot \tanh(W_s \mathcal{T}(x_t) + U_s * h_{t-1} + b_s)$$
$$h_t = \tanh(s_t)$$





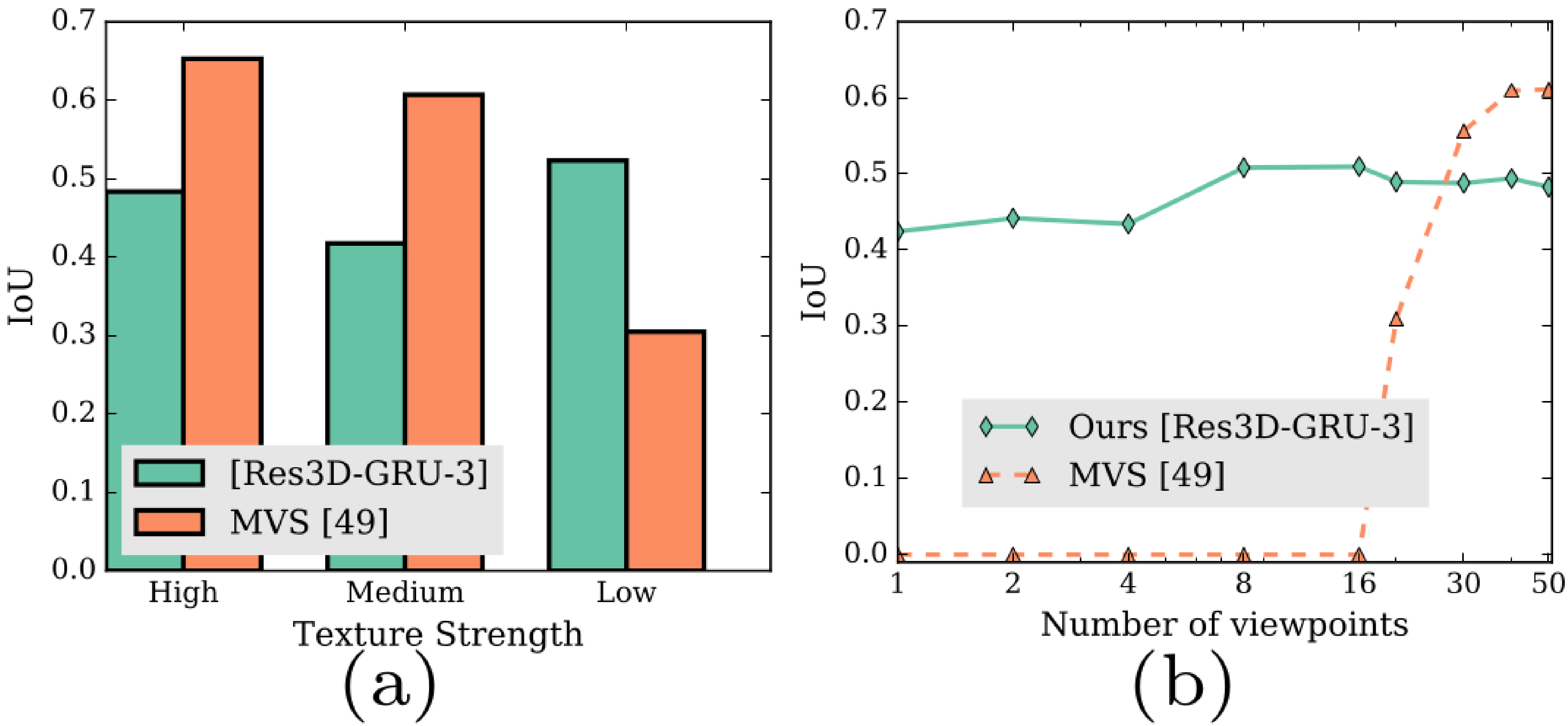


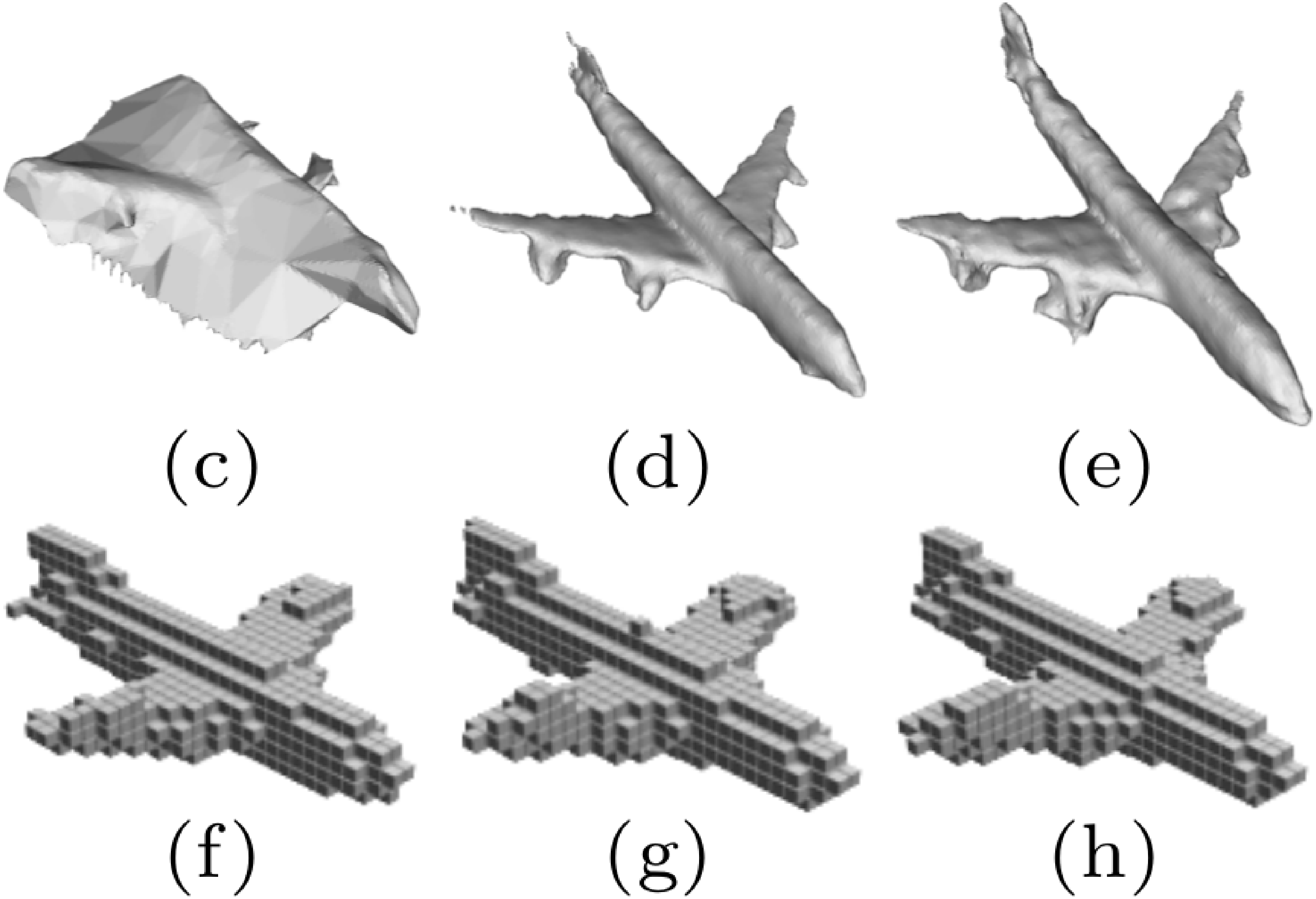


Intersection-over-Union (IoU)

$$IoU = \sum_{i,j,k} \left[I(p_{(i,j,k)} > t) I(y_{(i,j,k)}) \right] / \sum_{i,j,k} \left[I \left(I(p_{(i,j,k)} > t) + I(y_{(i,j,k)}) \right) \right]$$

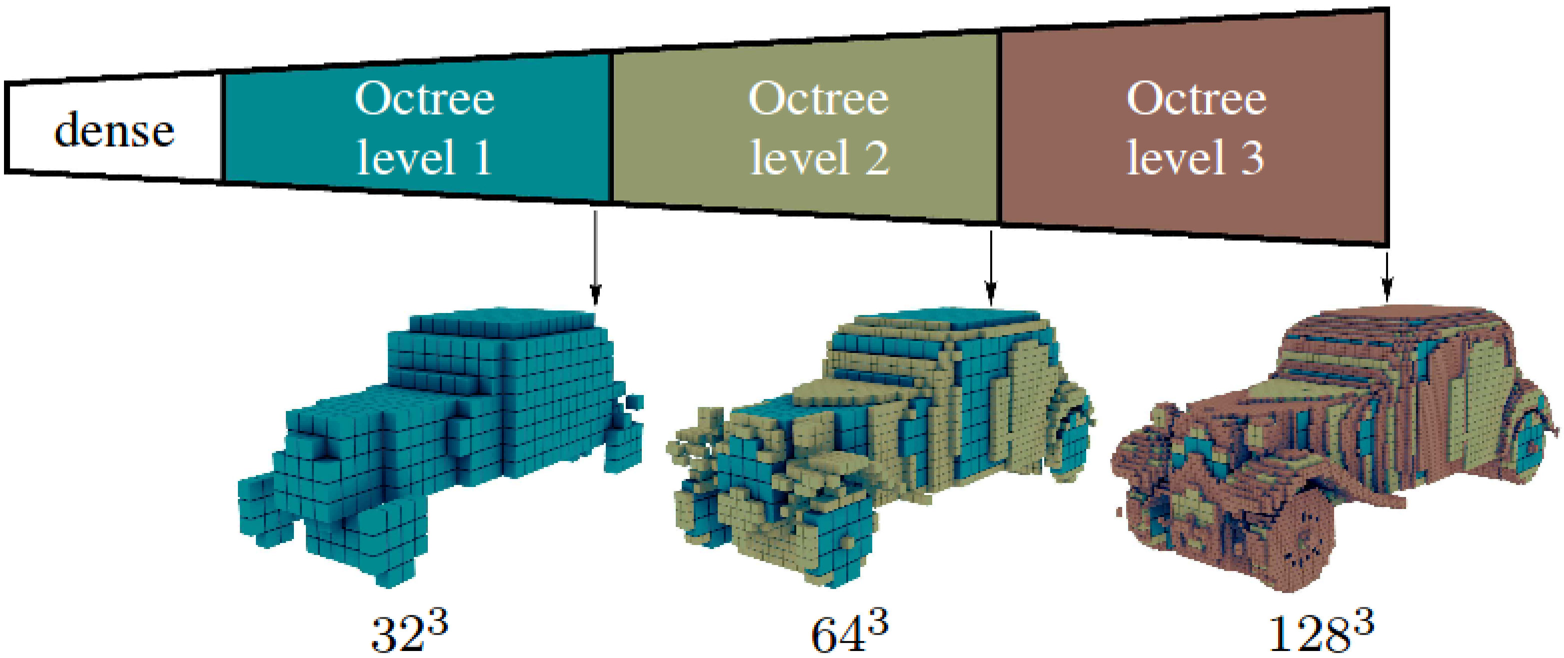
Multi View Stereo(MVS) vs. 3D-R2N2





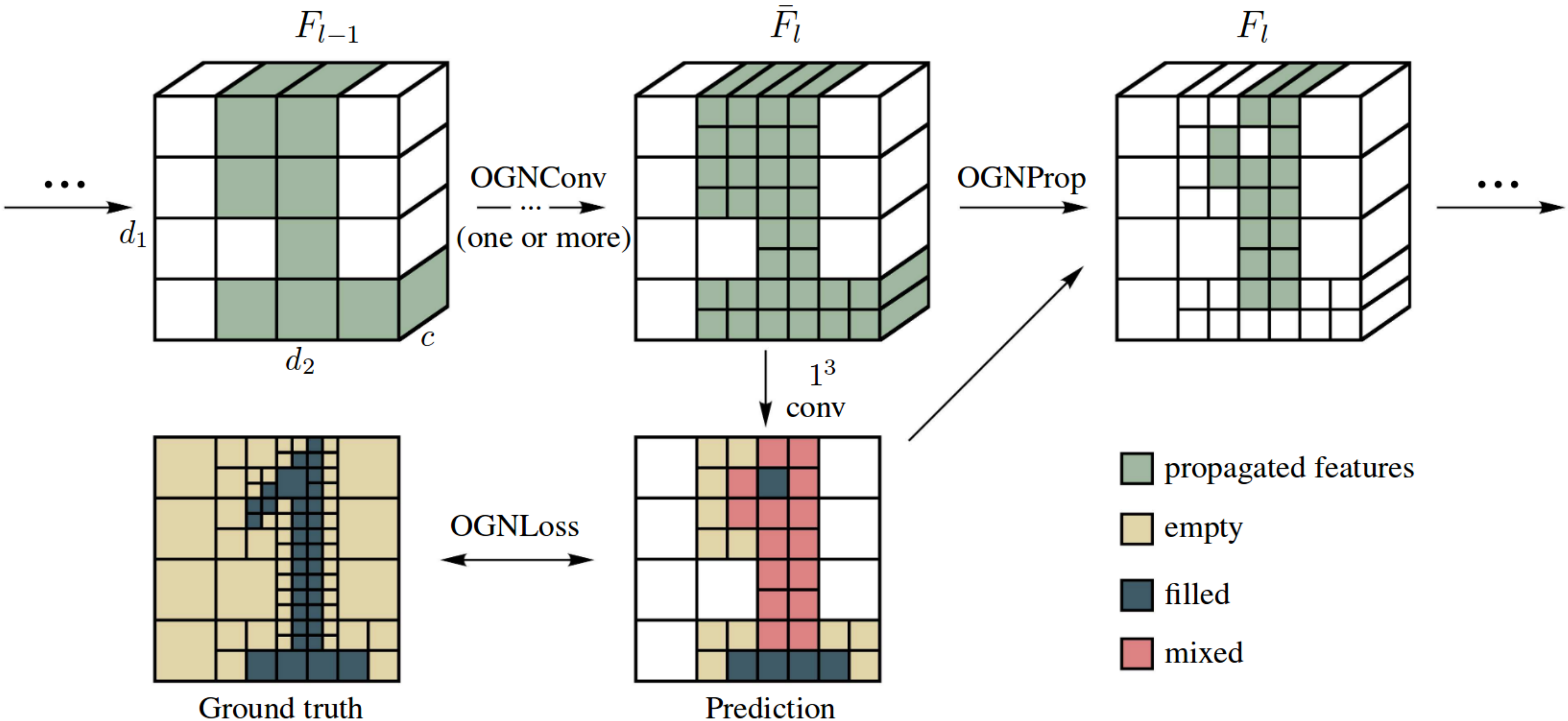
20 views 30 views 40 views

3D Generative Model: octree



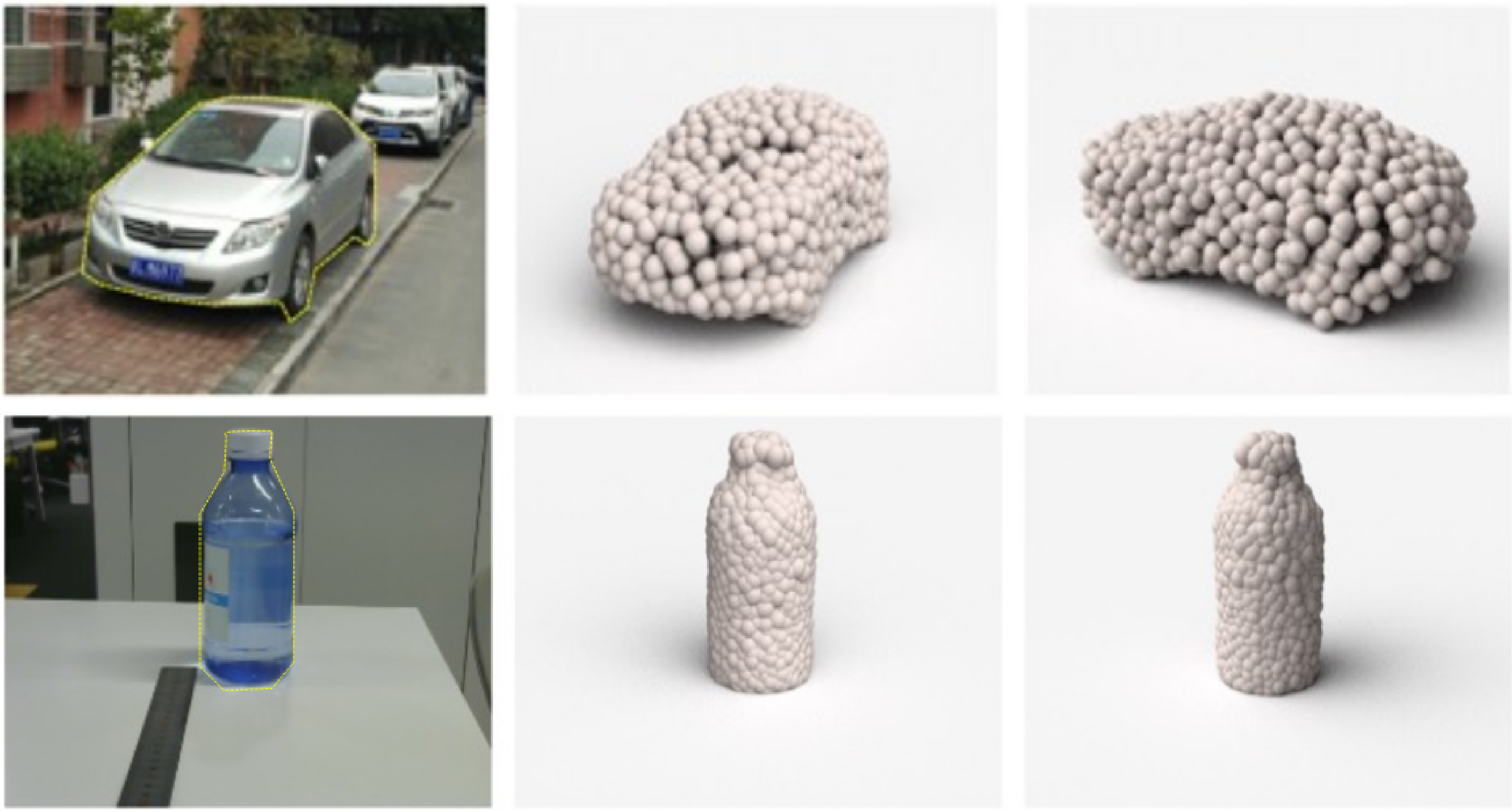
[Arxiv 2017] Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs

3D Generative Model: octree



[Arxiv 2017] Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs

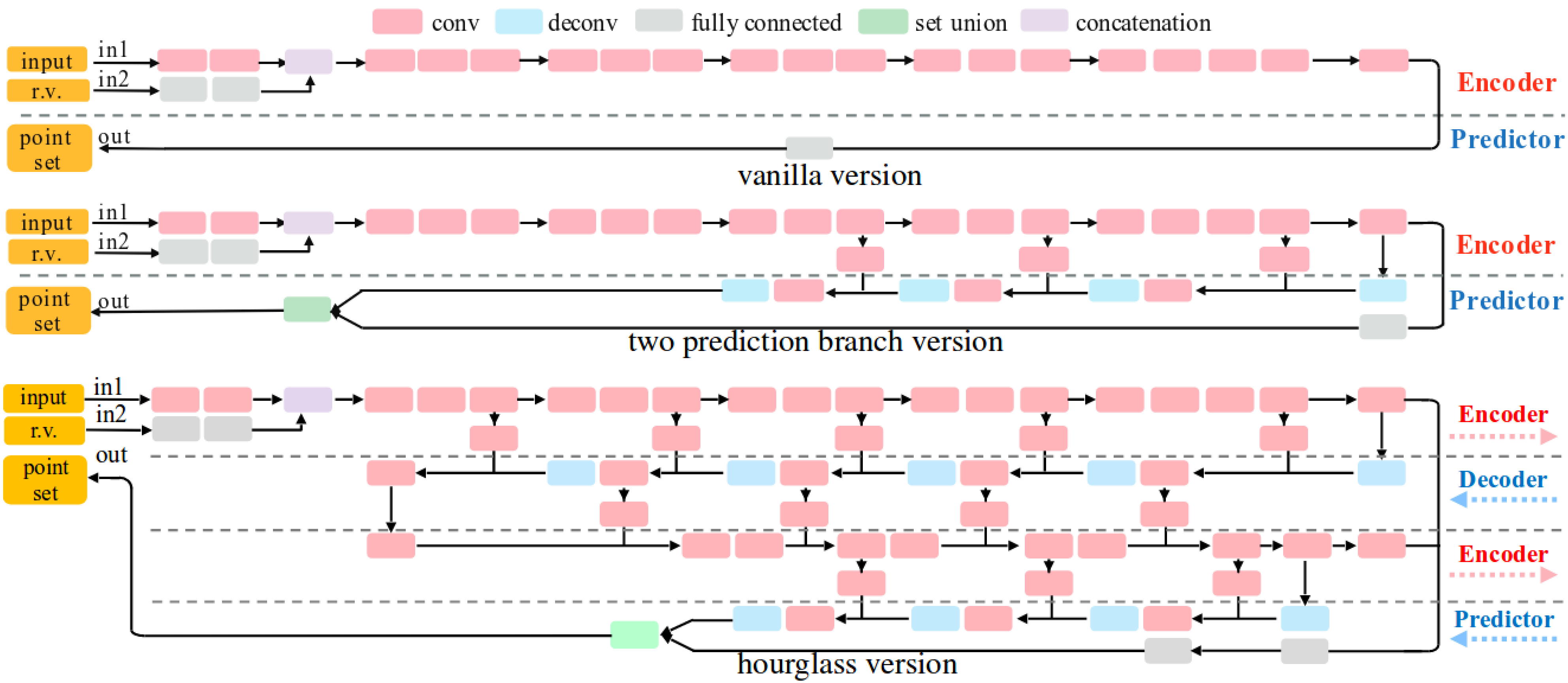
3D Generative Model: point cloud



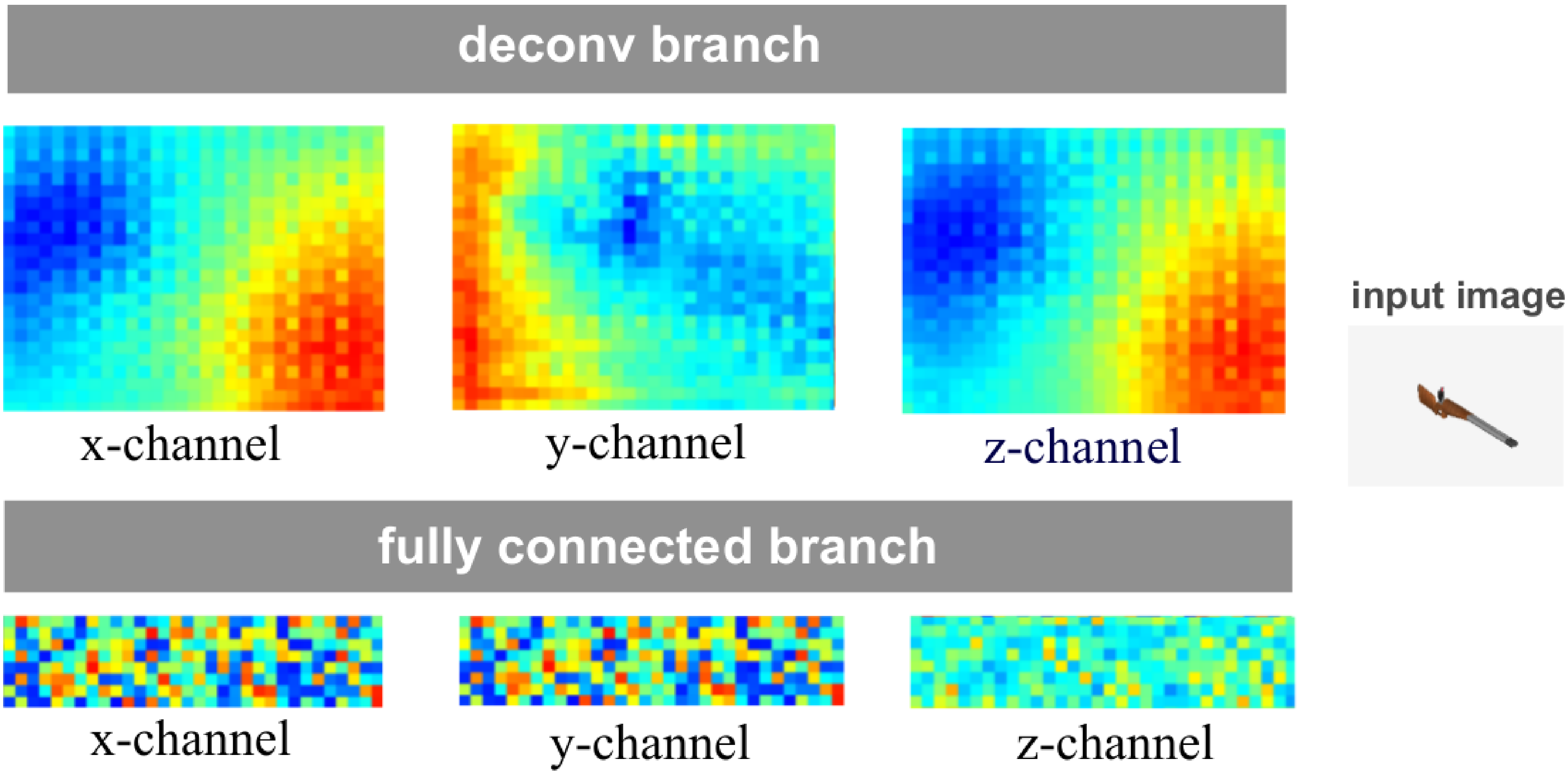
Input Reconstructed 3D point cloud

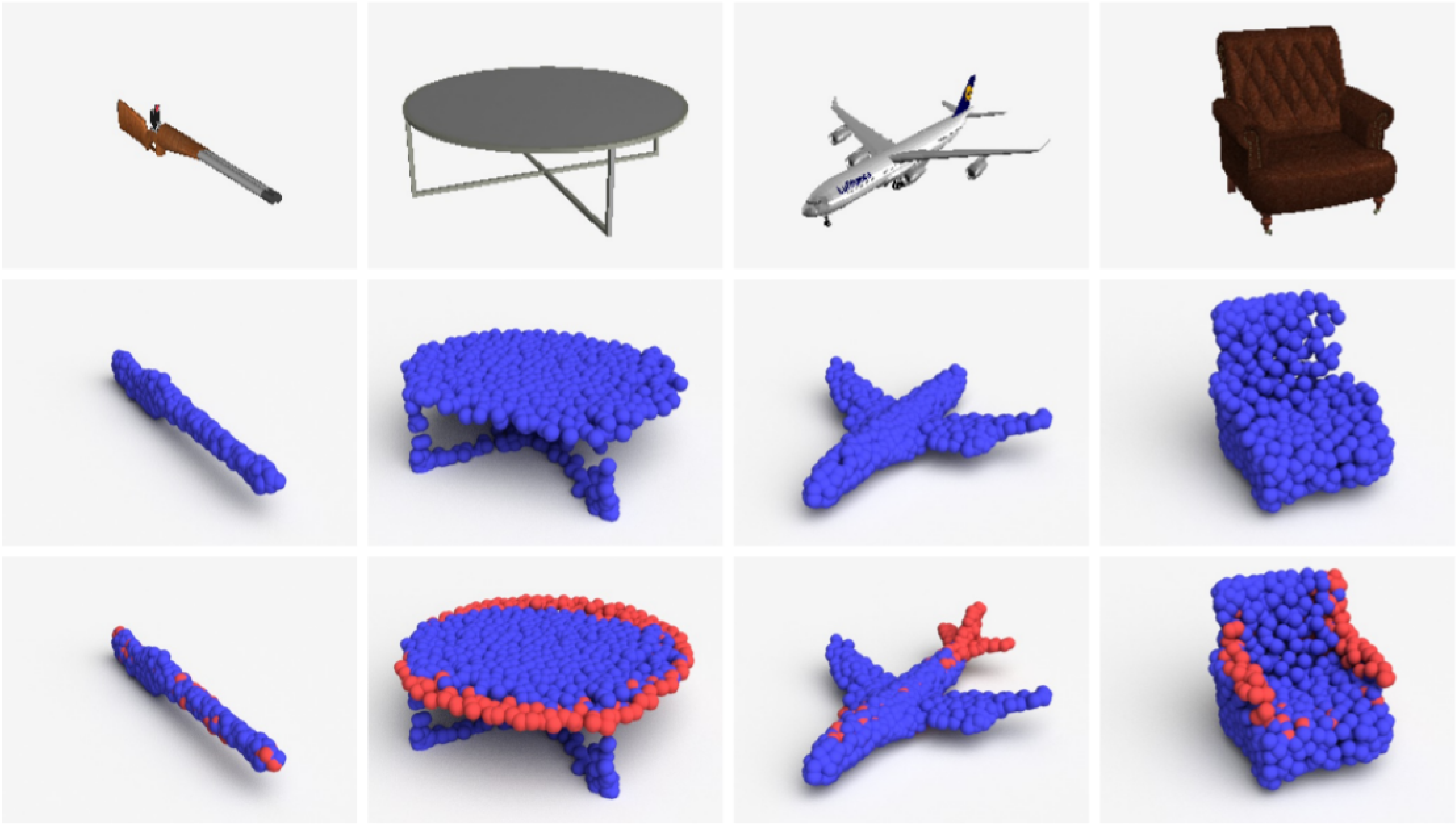
[CVPR 2017 Oral] A Point Set Generation Network for 3D Object Reconstruction from a Single Image

3D Generative Model: point cloud

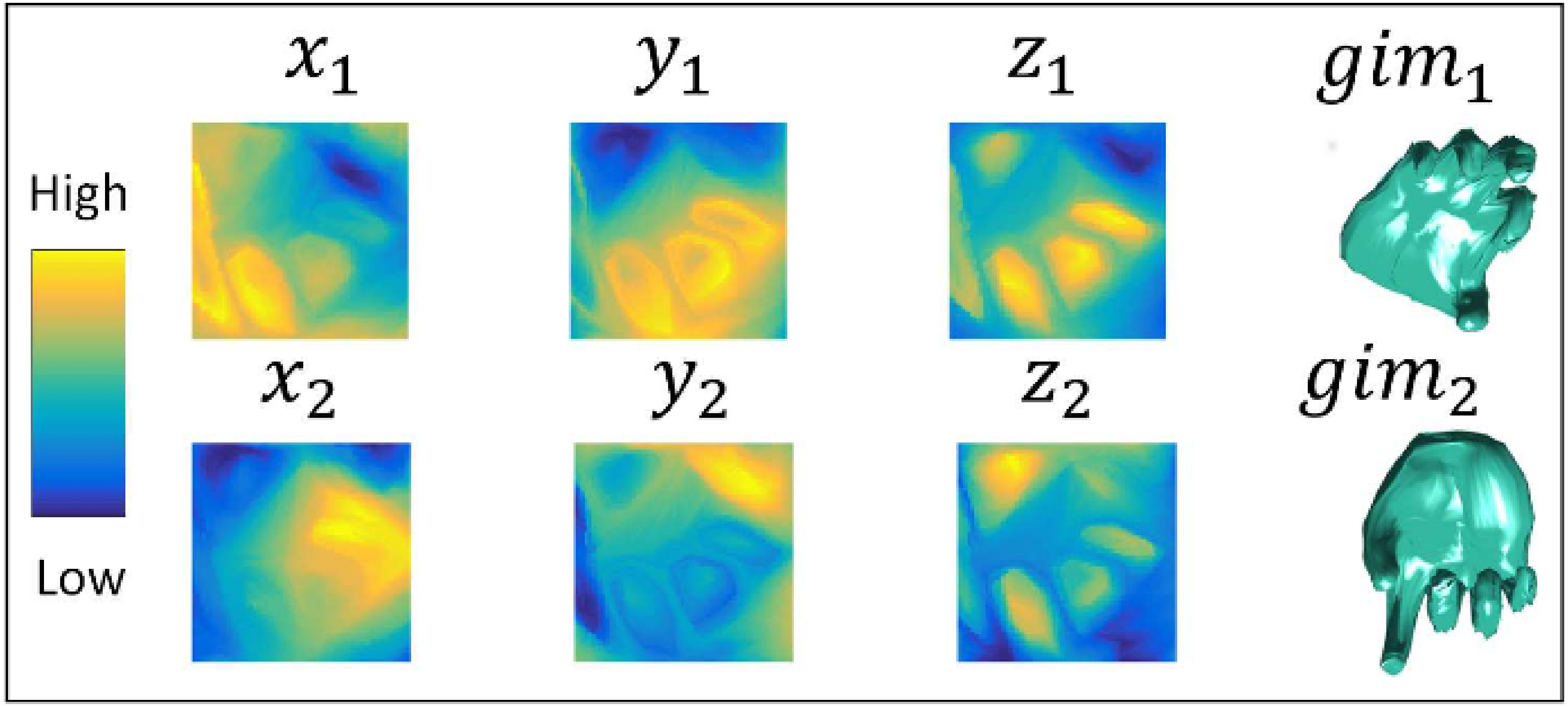


[CVPR 2017 Oral] A Point Set Generation Network for 3D Object Reconstruction from a Single Image


deconv branch
x-channel
y-channel
z-channel
input image
fully connected branch
x-channel
y-channel
z-channel

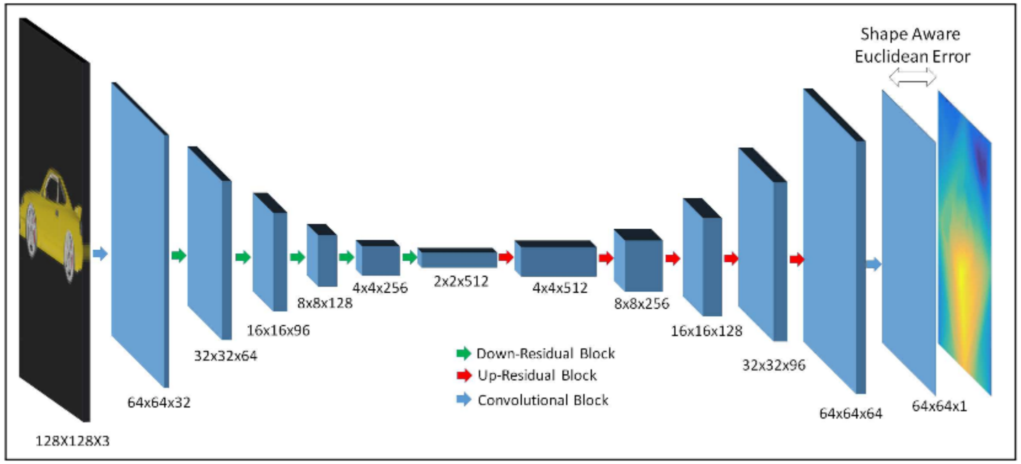


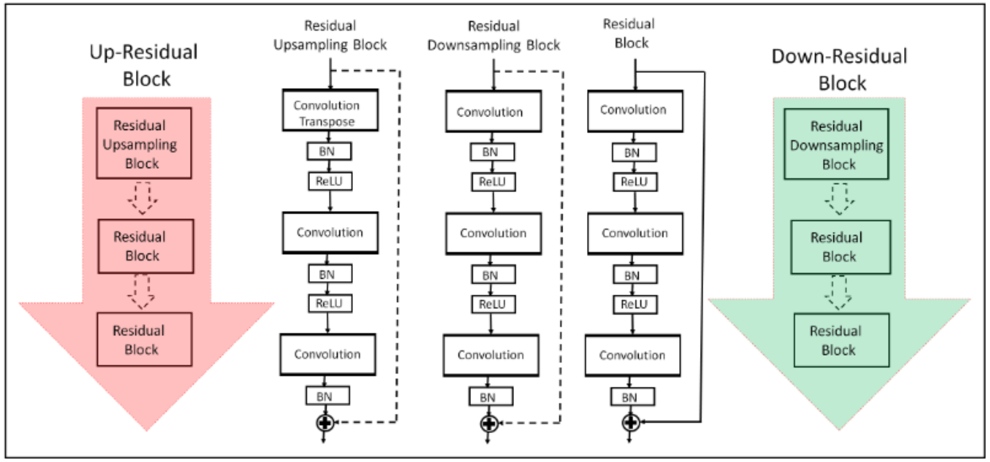
3D Generative Model: geometry image



[CVPR 2017] SurfNet: Generating 3D shape surfaces using deep residual networks

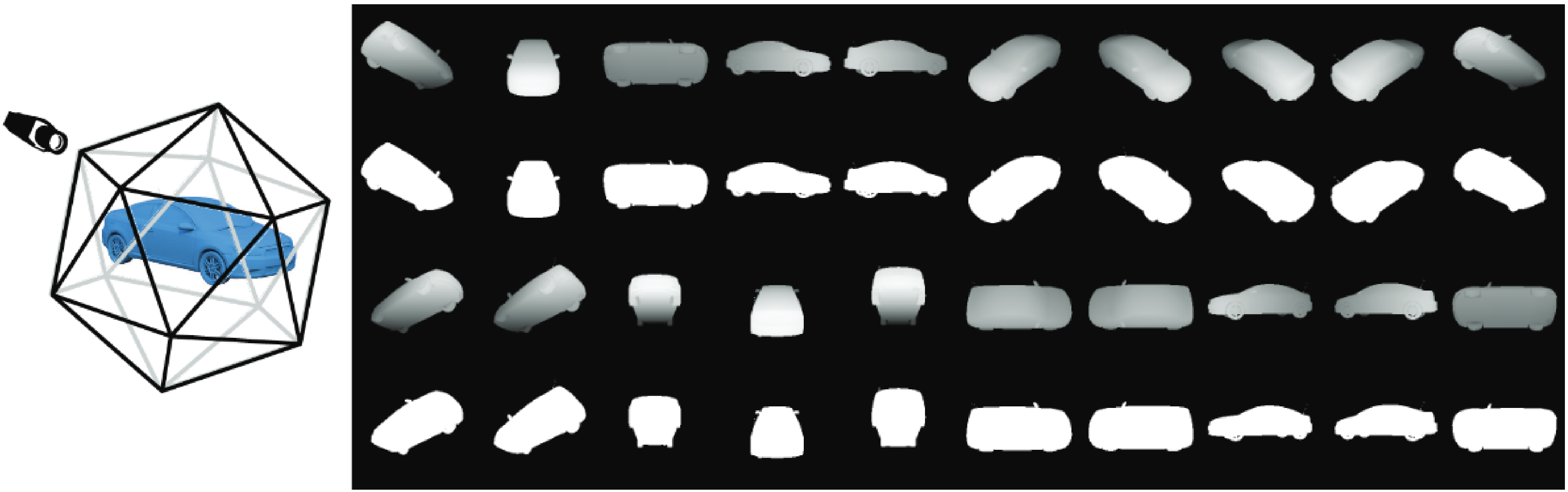
3D Generative Model: geometry image





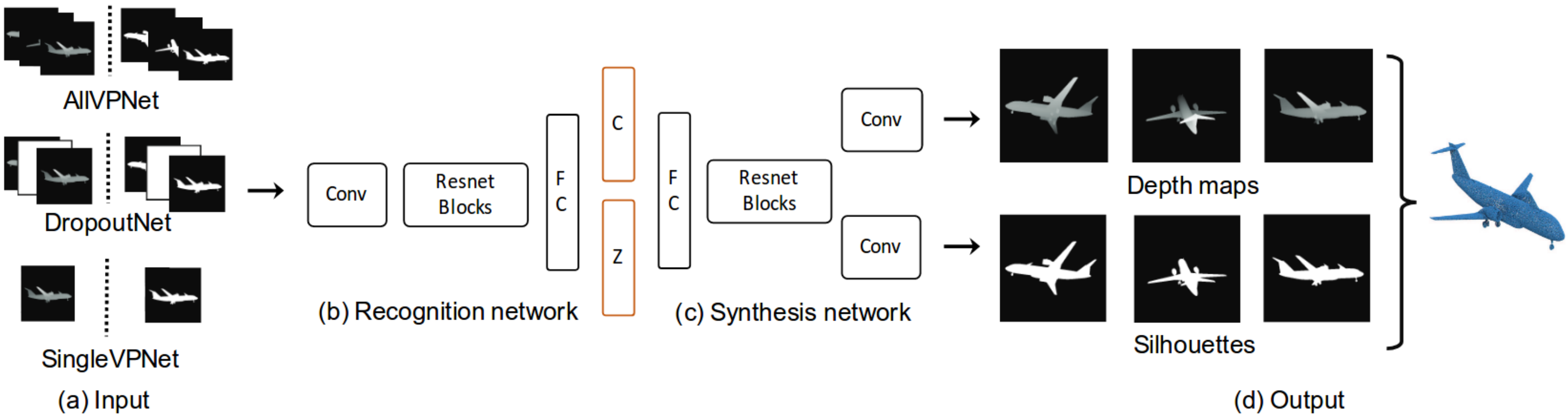
[CVPR 2017] SurfNet: Generating 3D shape surfaces using deep residual networks

3D Generative Model: depth map



[CVPR 2017] Synthesizing 3D Shapes via Modeling Multi-View Depth Maps and Silhouettes with Deep Generative Networks

3D Generative Model: depth map



[CVPR 2017] Synthesizing 3D Shapes via Modeling Multi-View Depth Maps and Silhouettes with Deep Generative Networks