Shape from X

Haoqiang Fan
fhq@megvii.com

Some figures adapted from http://cvg.ethz.ch/teaching/2012spring/3dphoto/Slides/3dphoto12_shapeFromX.pdf

Perception / Measurement of 3D

3D is vital for survival

How to reconstruct / perceive 3D

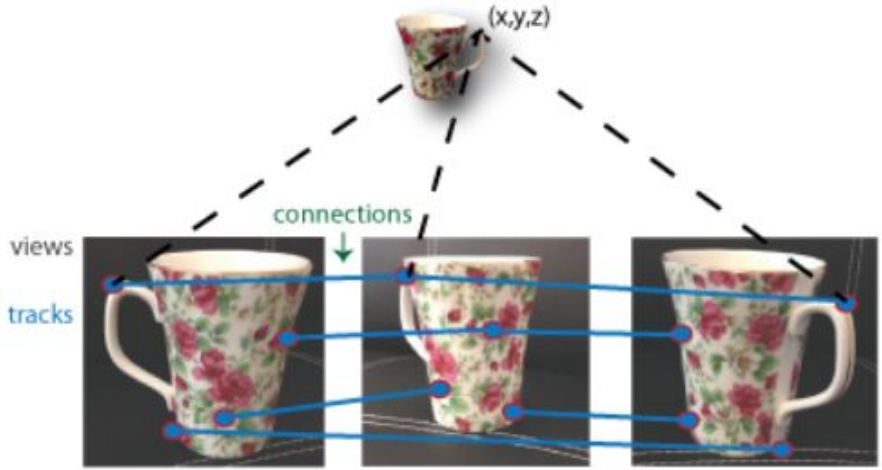
By means of visual information

-> optical, 2D array of input

Structure from Motion

The most easy-to-understand approach

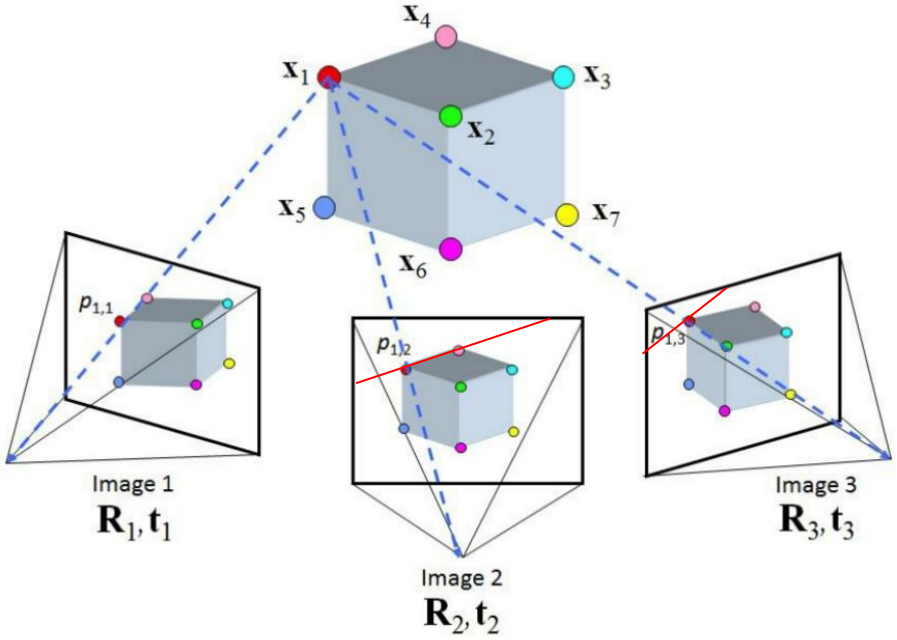
Triangulation



https://cn.mathworks.com/help/vision/ug/structure-from-motion.html

Triangulation

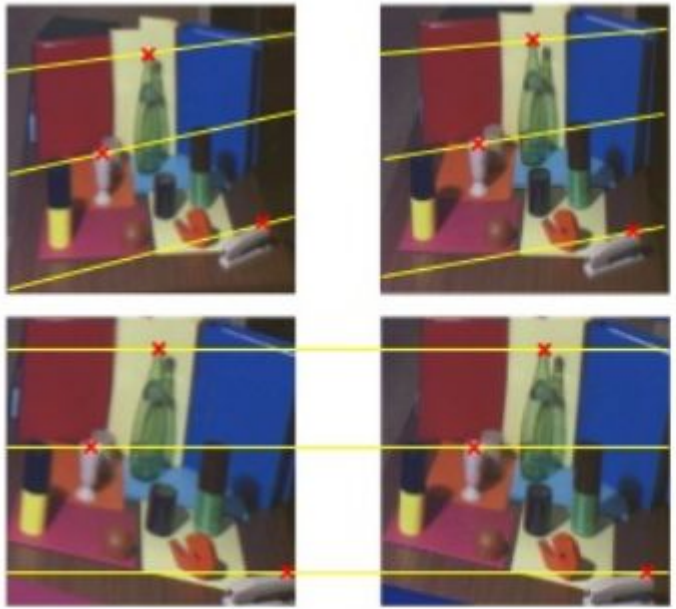
The epipolar constraint

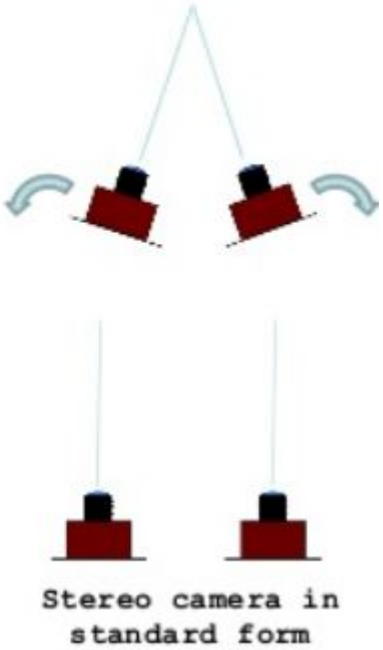


Stereo and kinect fusion for continuous 3D reconstruction and visual odometry

Stereo, rectification, disparity

row-to-row
correspondence



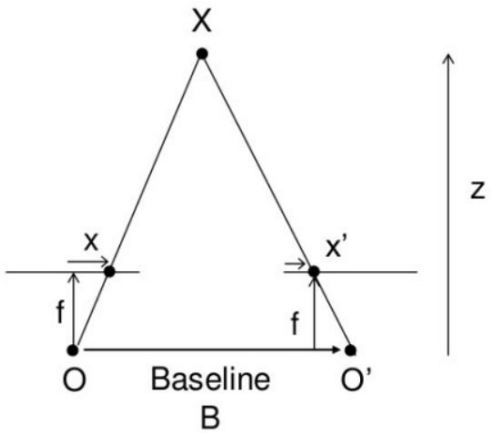


https://www.slideshare.net/DngNguyn43/stereo-vision-42147593

Disparity, depth

d=y_right - y_left

z=B*F/d



[OpenCV: Depth Map from Stereo Images]

Middlebury Stereo Evaluation

3D Point Cloud

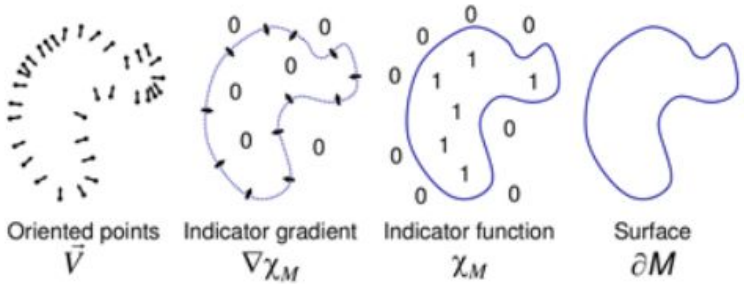
x=x_screen/F*z

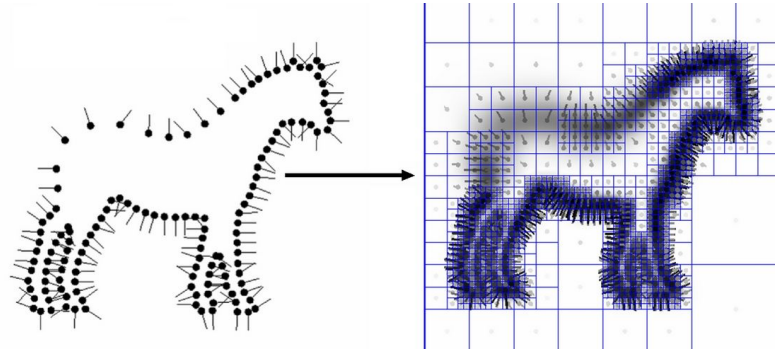
y=y_screen/F*z

Bundler: Structure from Motion (SfM) for Unordered Image Collections

Surface Reconstruction

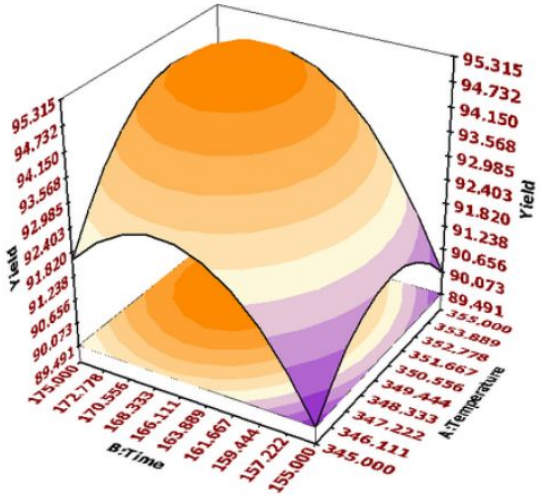
Integration of oriented point

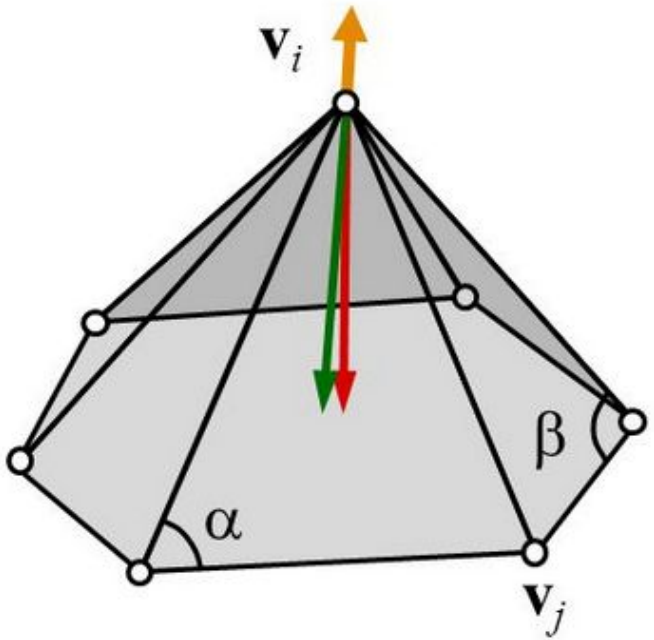




Laplacian and Normal

Laplacian = Normal * Mean Curvature





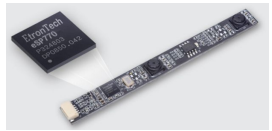
SfM Scanning

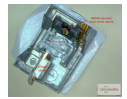
SLAM based positioning

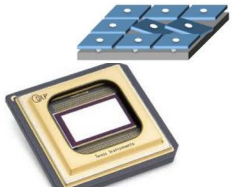
Depth Sensing: Active Sensors

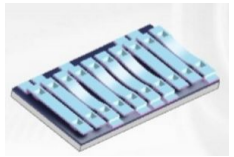
Structured Light

Time of Flight(ToF)



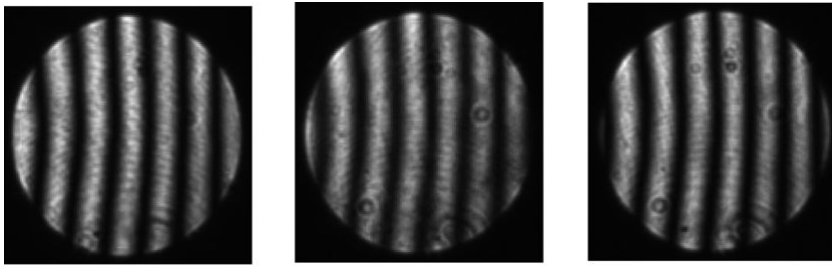






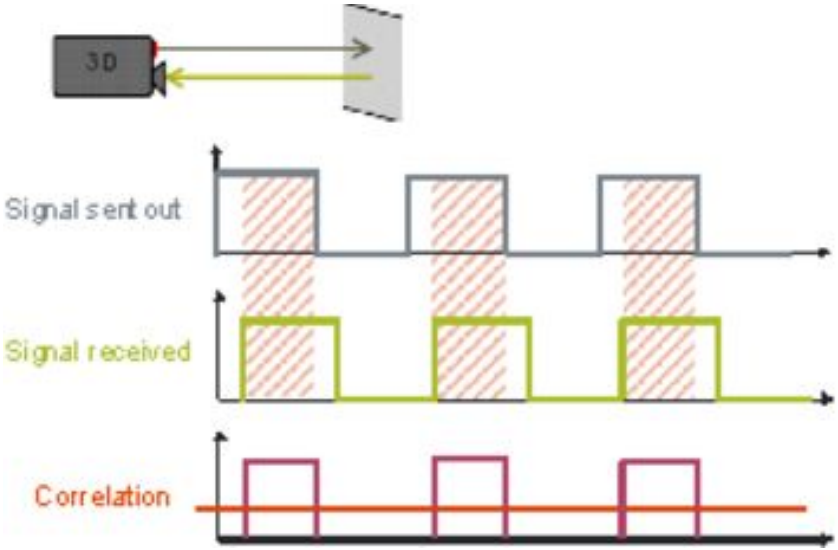
Structured Light

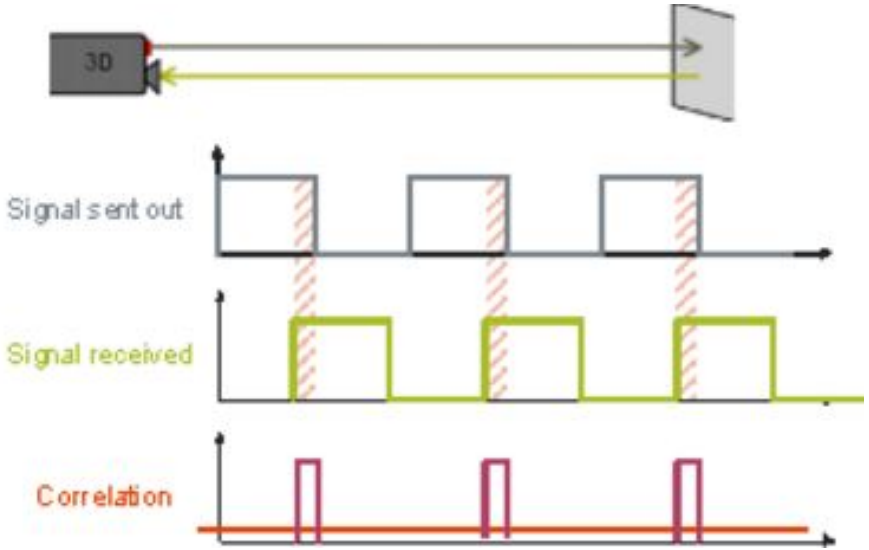
Static pattern & dynamic pattern



Time of Flight (ToF)

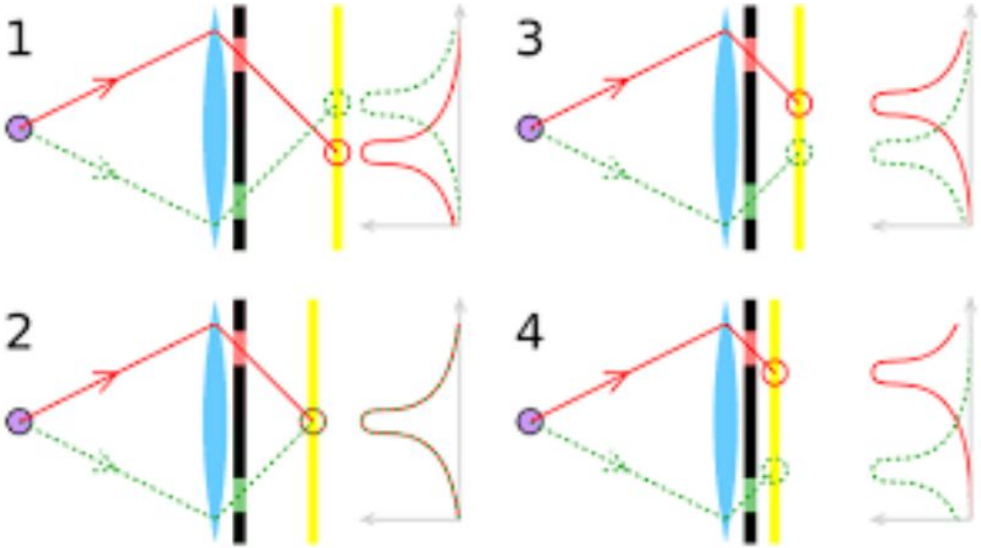
Pulsed modulation

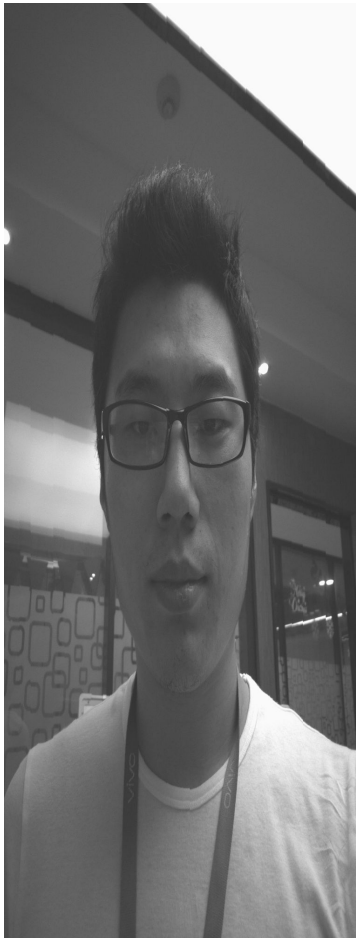




Short Baseline Stereo

Phase Detection Autofocus





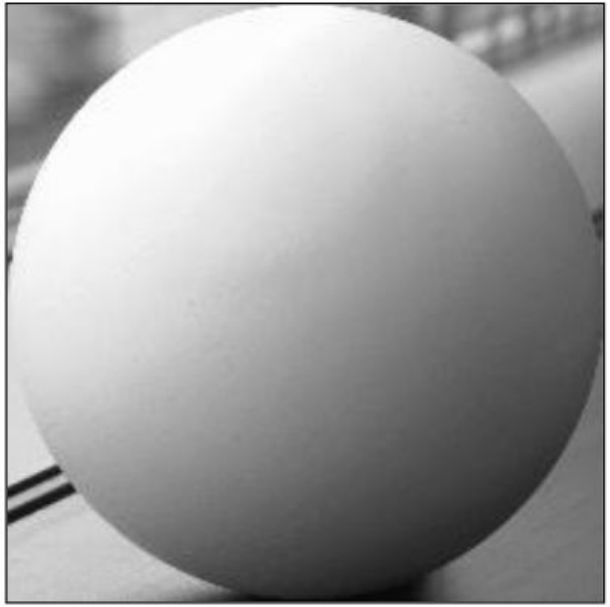
Shape from X

Structure from Motion: 3D geometry

Are there other possibilities?

Shape from Shading

Shading as a cue of 3D shape



The Lambertian Law

k : source brightness

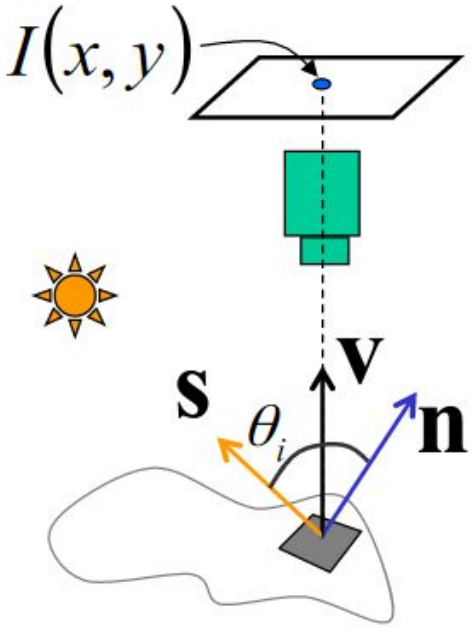
ρ : surface albedo (reflectance)

c : constant (optical system)

Image irradiance:

$$I = \frac{\rho}{\pi} kc \cos\theta_i = \frac{\rho}{\pi} kc\mathbf{n}\cdot\mathbf{s}$$

$$I = \cos\theta_i = \mathbf{n}\cdot\mathbf{s}$$



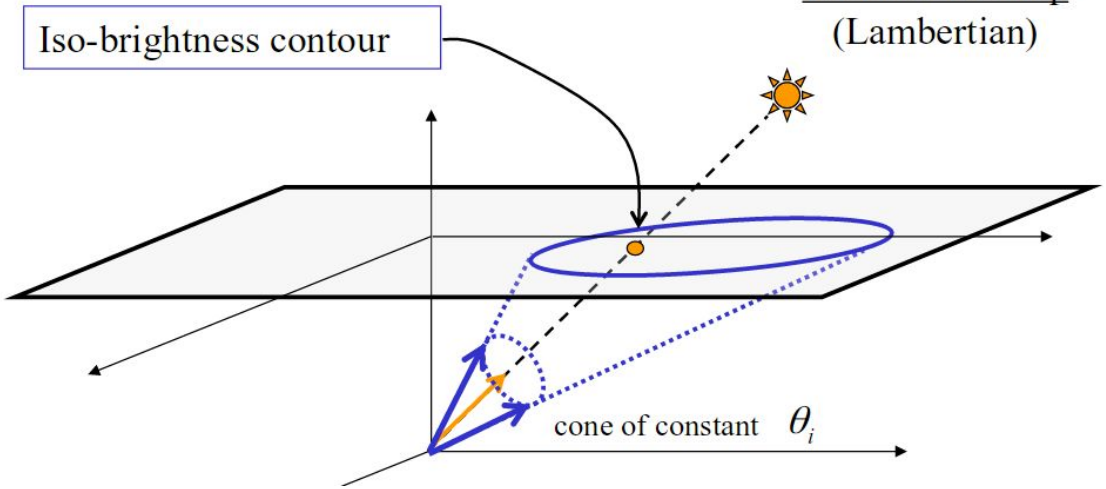
Shape from Shading

Solve for gradient

Assuming constant albedo

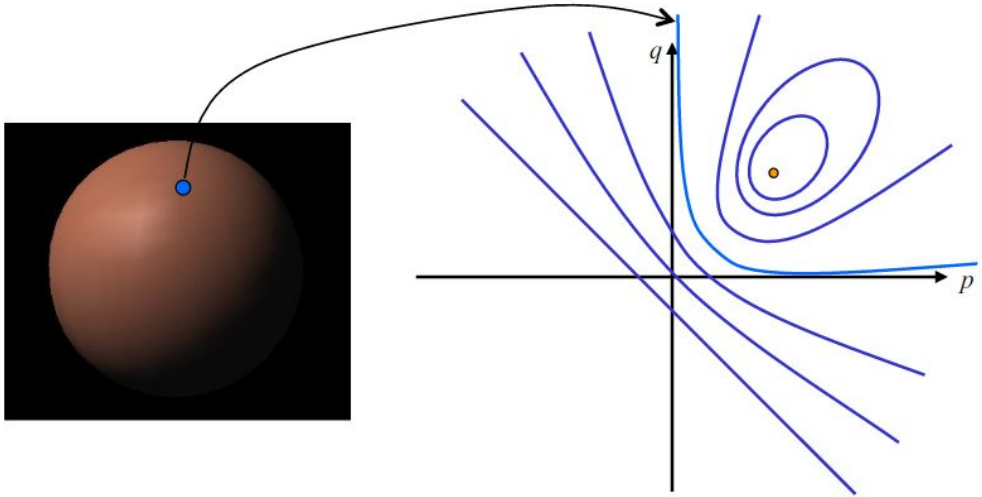
$$I = \cos\theta_i = \mathbf{n}\cdot\mathbf{s} = \frac{(pp_s + qq_s + 1)}{\sqrt{p^2 + q^2 + 1}\sqrt{p_S^2 + q_S^2 + 1}} = R(p,q)$$

Reflectance Map
(Lambertian)



Is Shape Uniquely Determined?

bas-relief ambiguity



Shape from Shading

Data term + Prior

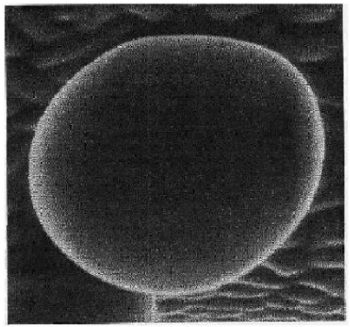
$$e_i = \iint_{\text{image}} \left(I(x,y) - R(f,g)\right)^2 dxdy$$

$$e_s = \iint_{\text{image}} \left(f_x^2 + f_y^2\right) + \left(g_x^2 + g_y^2\right) dxdy$$

$$e = e_s + \lambda e_i$$

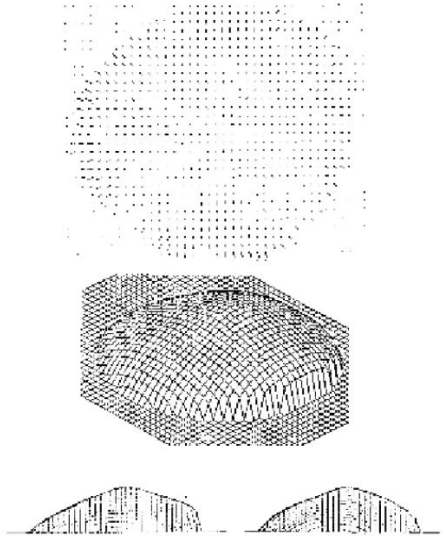
Shape from Shading

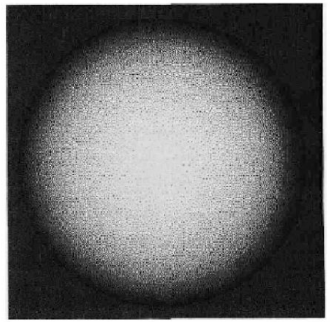
Example

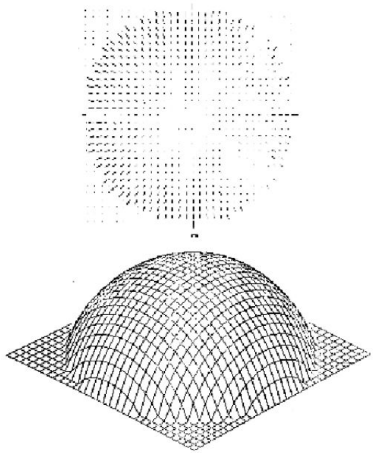


Scanning Electron Microscope image
(inverse intensity)

by Ikeuchi and Horn







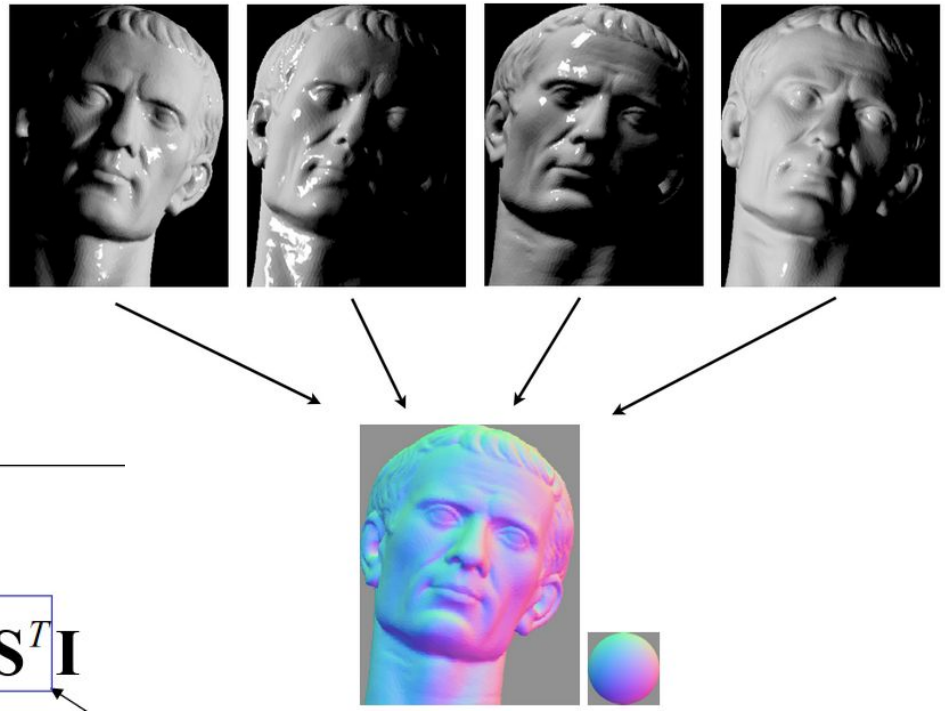
by Ikeuchi and Horn

Photometric Stereo

$$I_1 = \rho \mathbf{n} \cdot \mathbf{s}_1$$
$$I_2 = \rho \mathbf{n} \cdot \mathbf{s}_2$$
$$I_3 = \rho \mathbf{n} \cdot \mathbf{s}_3$$

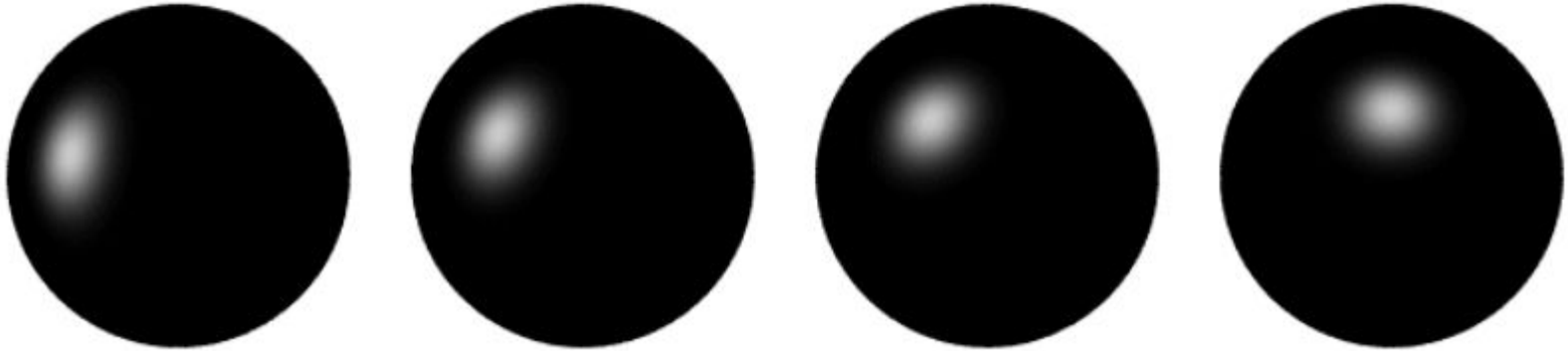
$$\begin{bmatrix} I_1 \\ \vdots \\ I_N \end{bmatrix} = \begin{bmatrix} \mathbf{s}_1^T \\ \vdots \\ \mathbf{s}_N^T \end{bmatrix} \rho \mathbf{n}$$

$$\mathbf{I} = \mathbf{S}\tilde{\mathbf{n}}$$
$$\mathbf{S}^T\mathbf{I} = \mathbf{S}^T\mathbf{S}\tilde{\mathbf{n}}$$
$$\tilde{\mathbf{n}} = \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T\mathbf{I}$$

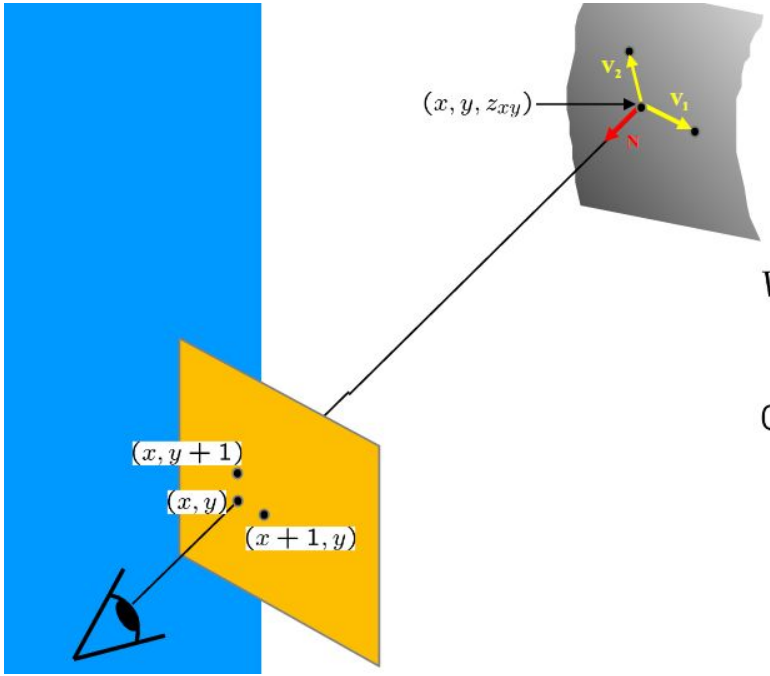


Photometric Stereo

Measure the normal direction: the chrome sphere



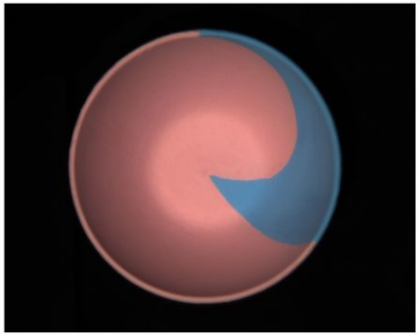
Depth from Normals

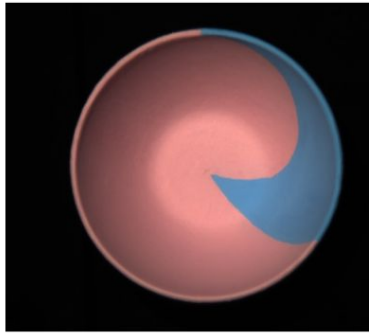


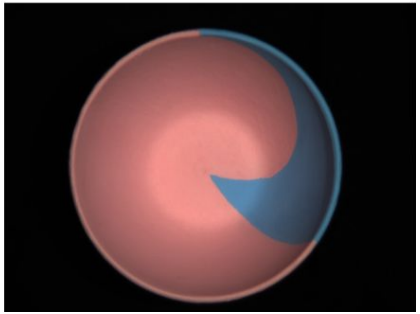
$$
\begin{aligned}
V_1 &= (x+1, y, z_{x+1,y}) - (x, y, z_{xy}) \\
&= (1, 0, z_{x+1,y} - z_{xy}) \\
0 &= N \cdot V_1 \\
&= (n_x, n_y, n_z) \cdot (1, 0, z_{x+1,y} - z_{xy}) \\
&= n_x + n_z(z_{x+1,y} - z_{xy})
\end{aligned}
$$

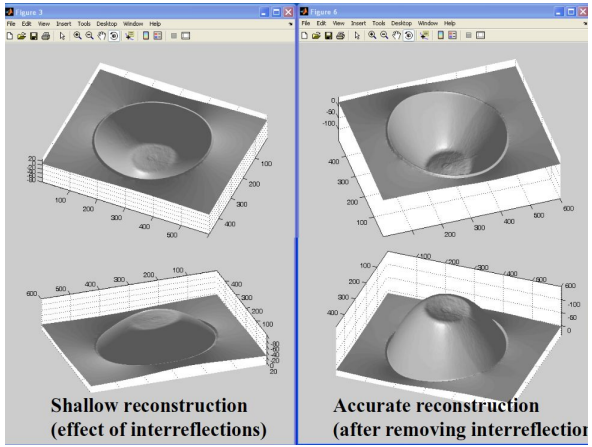
Example

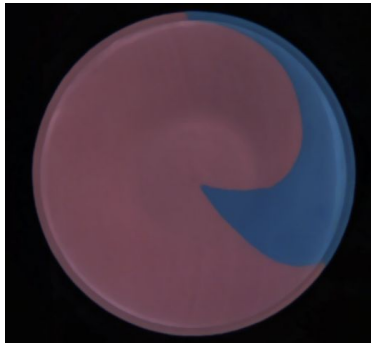
Good for near Lambertian material





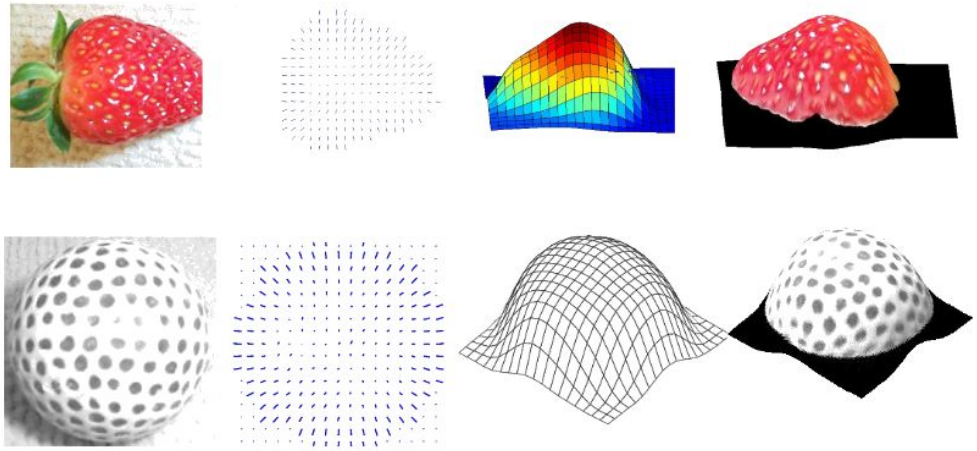






Shape from Texture

Solving normal from texture



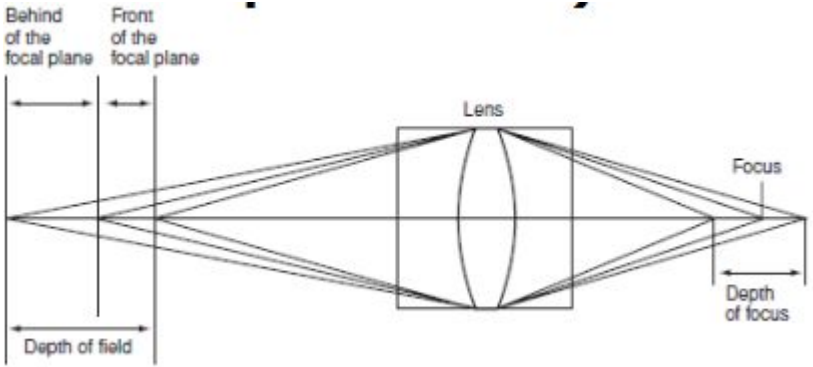
Examples from Angie Loh

Depth from Focus

Focus sweep

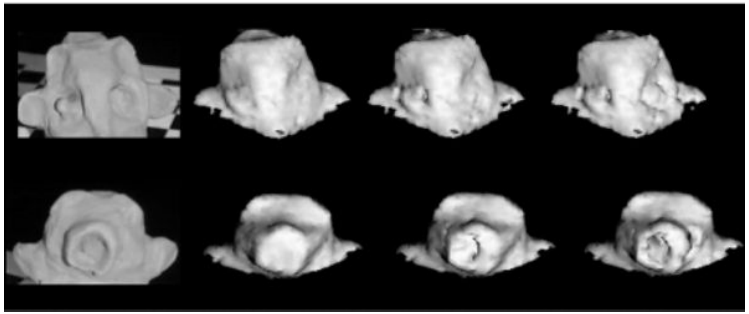
Depth from Defocus

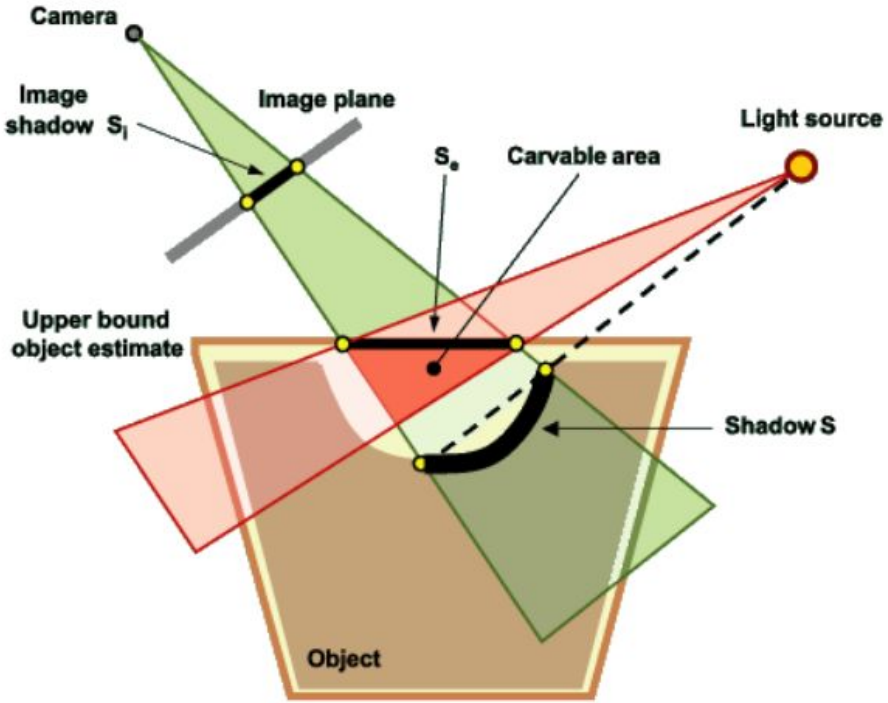
Measure blur, solve depth



Shape from Shadows

Shadow carving

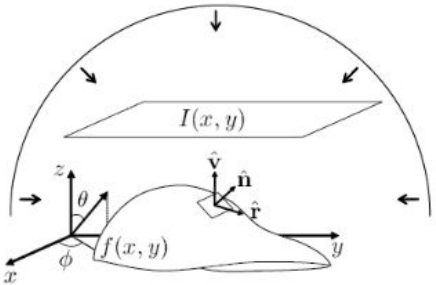


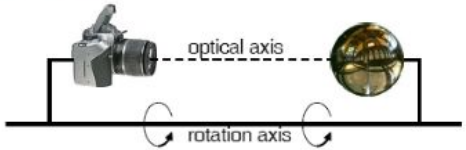


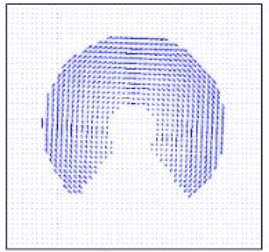
3D Reconstruction by Shadow Carving: Theory and Practical Evaluation”

Shape from Specularities

Solve deformation of mirrors.







Toward a Theory of Shape from Specular Flow

Shape from ?

Shape from Nothing?

Object priors!

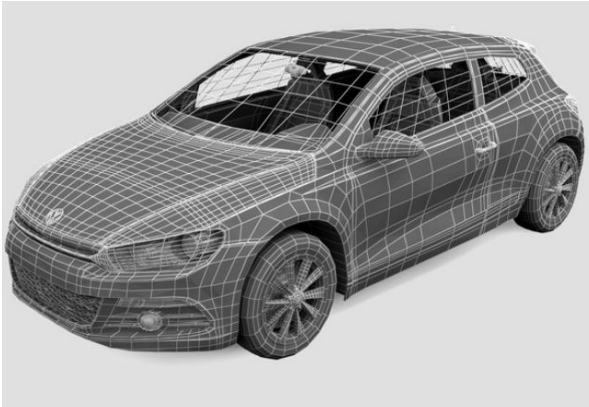
A **Point Set Generation Network** for
3D Object Reconstruction from a Single Image

Haoqiang Fan*[3,4] Hao Su*[1,2] Leonidas Guibas[1]

* indicates equal contribution

3D Reconstruction from Single Image

infer a whole shape, from a single image



3D Reconstruction from Single Image

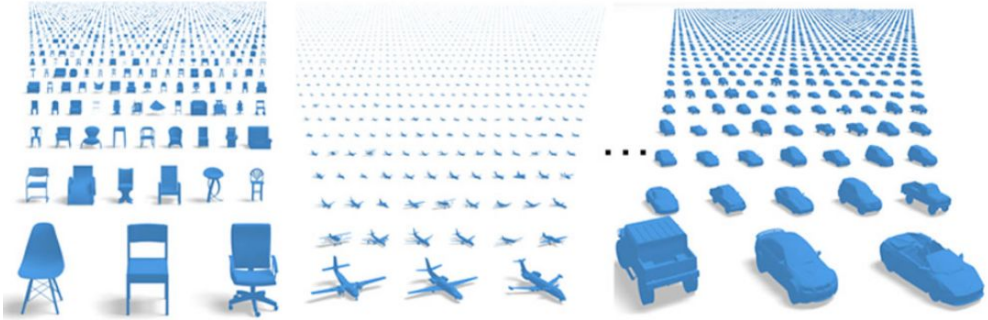
The ShapeNet Dataset

Google earth

SketchUp

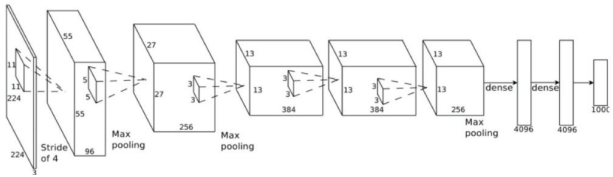
3D Warehouse

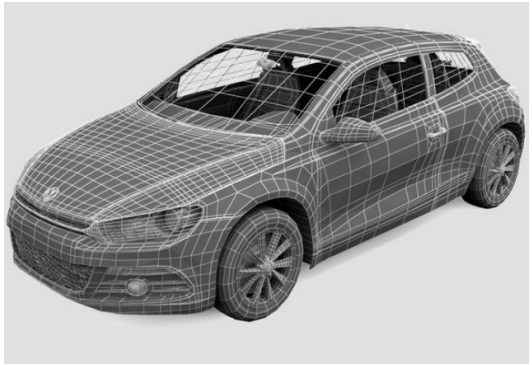
SHAPENET



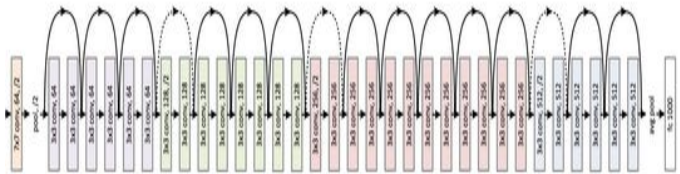
Face++ 旷视

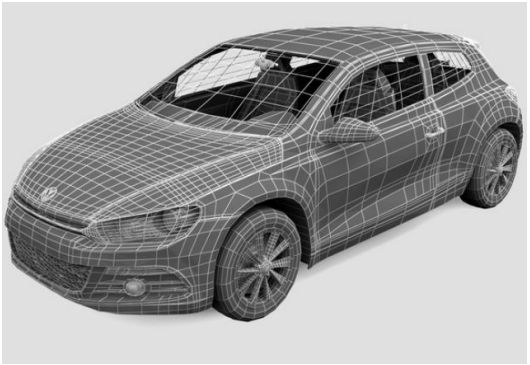
3D Reconstruction from Single Image



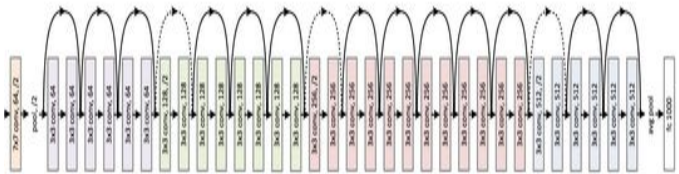


3D Reconstruction from Single Image

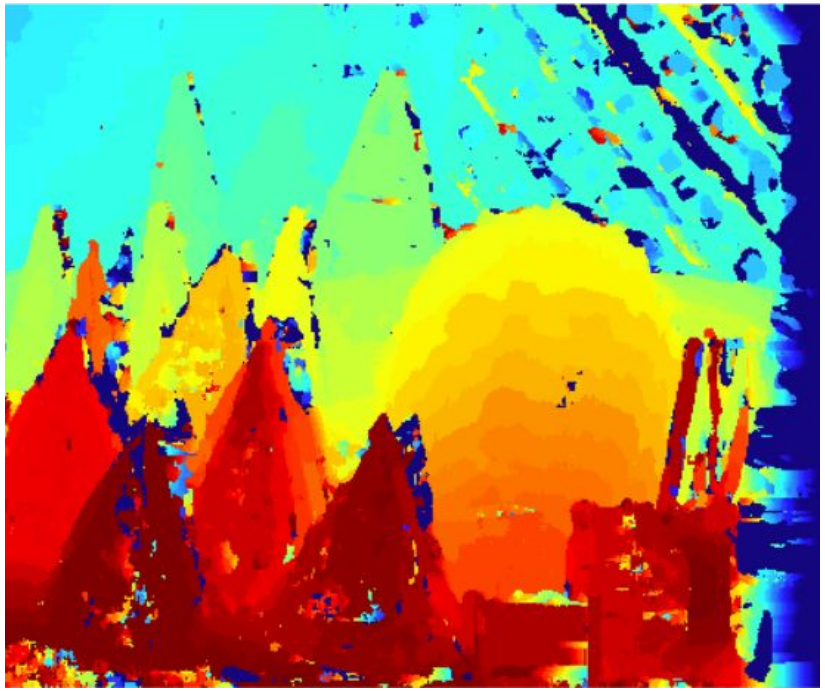




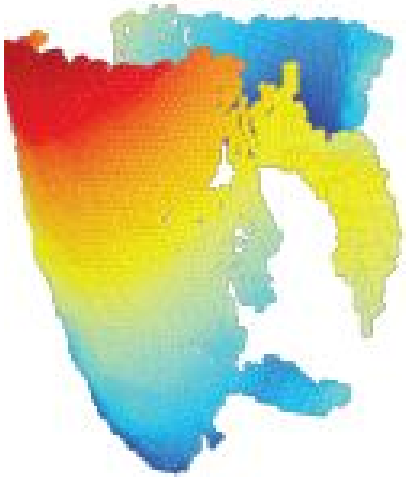
The issue of representation



Depth map



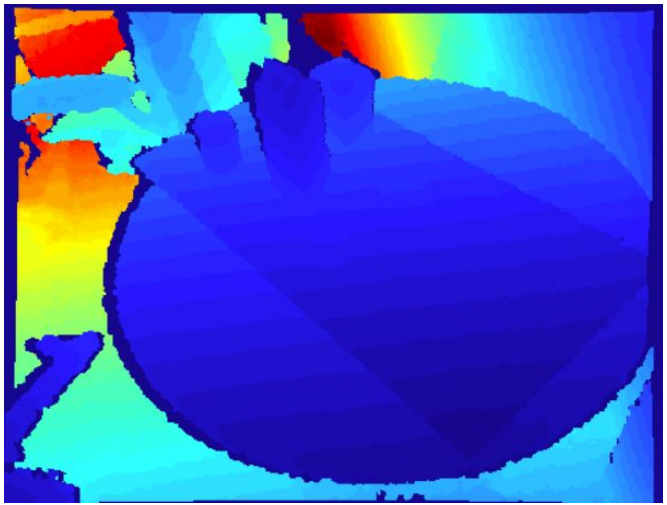
Depth map

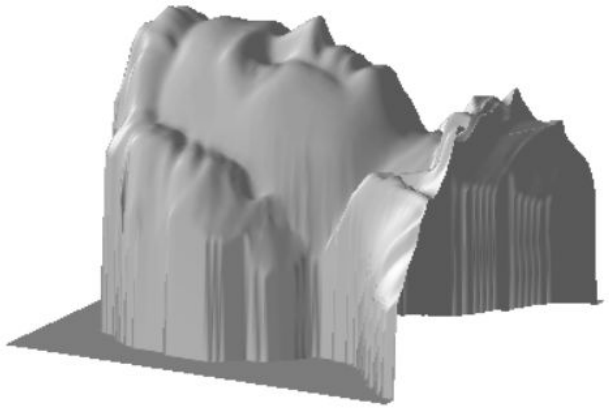


Second depth map

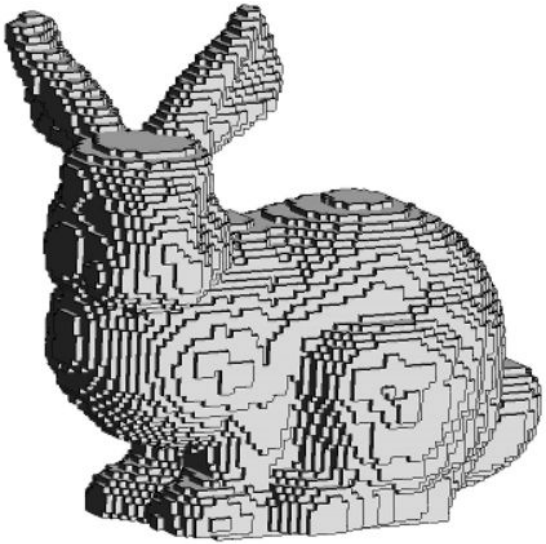
Second depth map

The problem of discontinuity

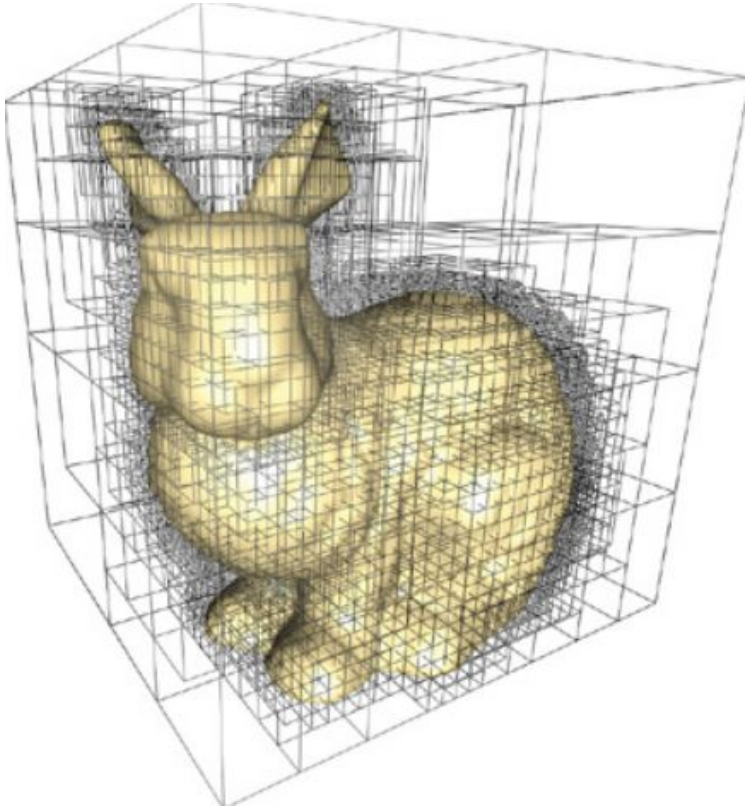




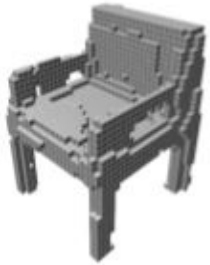
Volumetric Occupancy



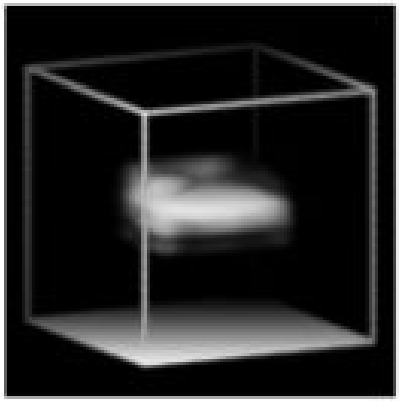
Problem of viewpoint

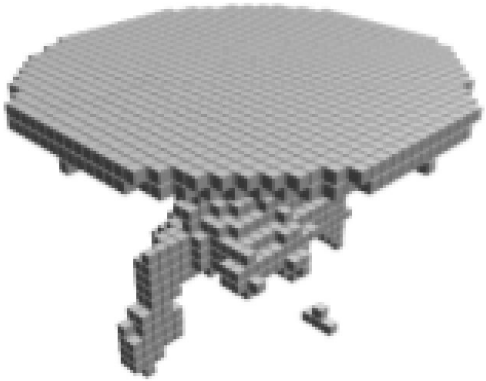


Canonical View



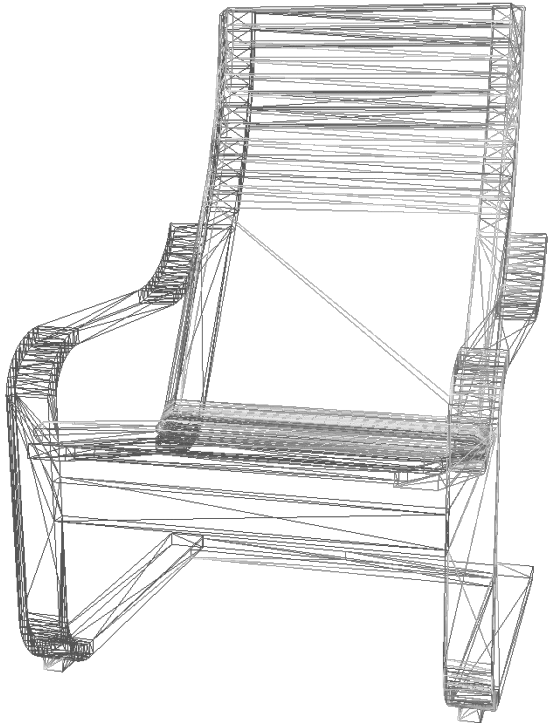
Volumetric Occupancy





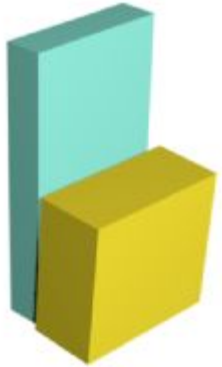
XML file

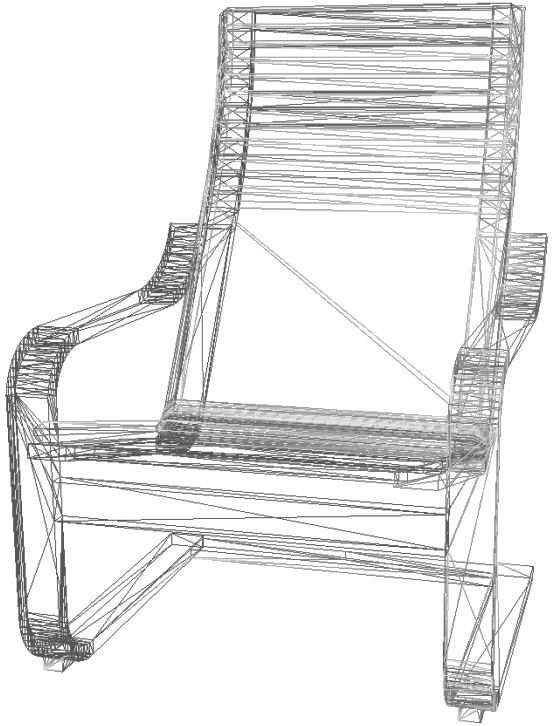
```
<scene>
  <Group>
    <Sphere>
      <position>0 128 0</position>
      <radius>112</radius>
      <numTheta>32</numTheta>
      <numPhi>32</numPhi>
      <material>
        <code>"MetallicPaint"</code>
        <parameters>
          <float name="eta">1.45</float>
          <float3 name="glitterColor">0.1 0.1 0.1</float3>
          <float name="glitterSpread">0.01</float>
          <float3 name="shadeColor">0.4 0.4 0.4</float3>
        </parameters>
      </material>
    </Sphere>
    <Sphere>
      <position>120 248 -120</position>
      <radius>8</radius>
      <numTheta>16</numTheta>
      <numPhi>16</numPhi>
      <material>
        <code>"Matte"</code>
        <parameters>
          <float3 name="reflectance">0.5 0.5 0.5</float3>
        </parameters>
      </material>
    </Sphere>
    <Sphere>
      <position>120 248 120</position>
      <radius>8</radius>
      <numTheta>16</numTheta>
      <numPhi>16</numPhi>
      <material>
        <code>"Matte"</code>
        <parameters>
          <float3 name="reflectance">0.5 0.5 0.5</float3>
        </parameters>
      </material>
    </Sphere>
```



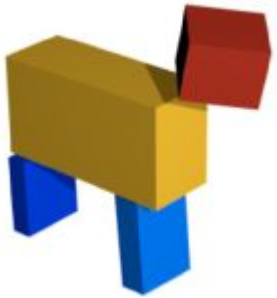
Face++ 旷视

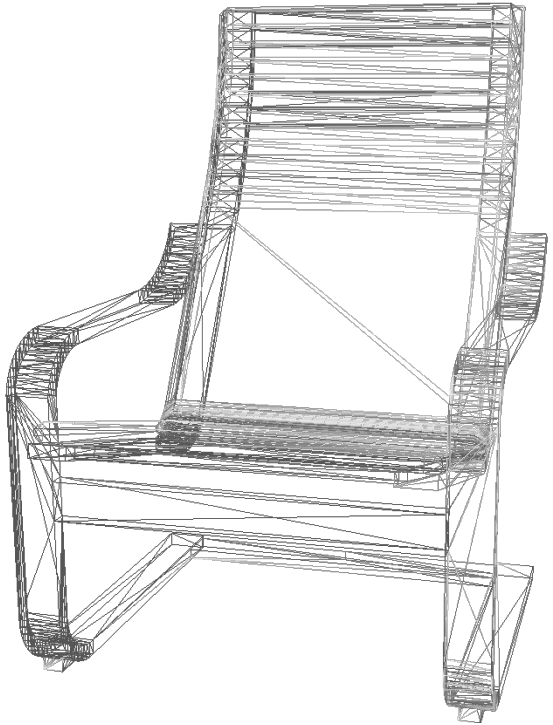
XML file



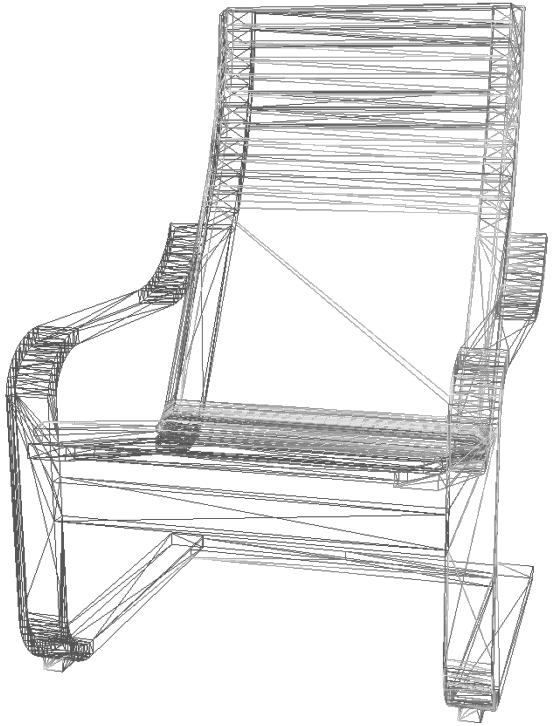


XML file





XML file



Can we find a representation that is..

flexible

structural

natural

A Point Set Generation Network for 3D Object Reconstruction from a Single Image

Input

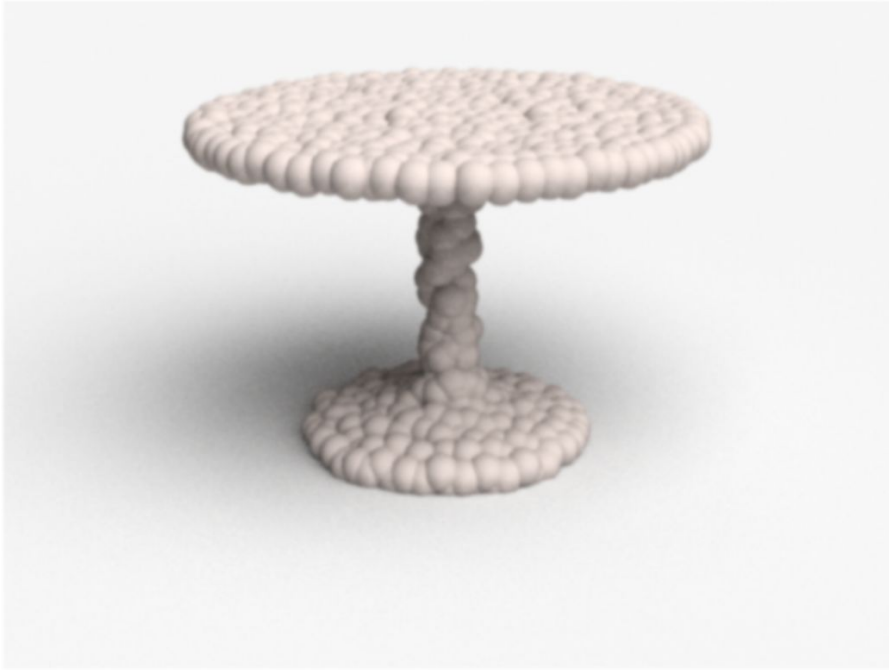
Reconstructed 3D point cloud

Point-based representation

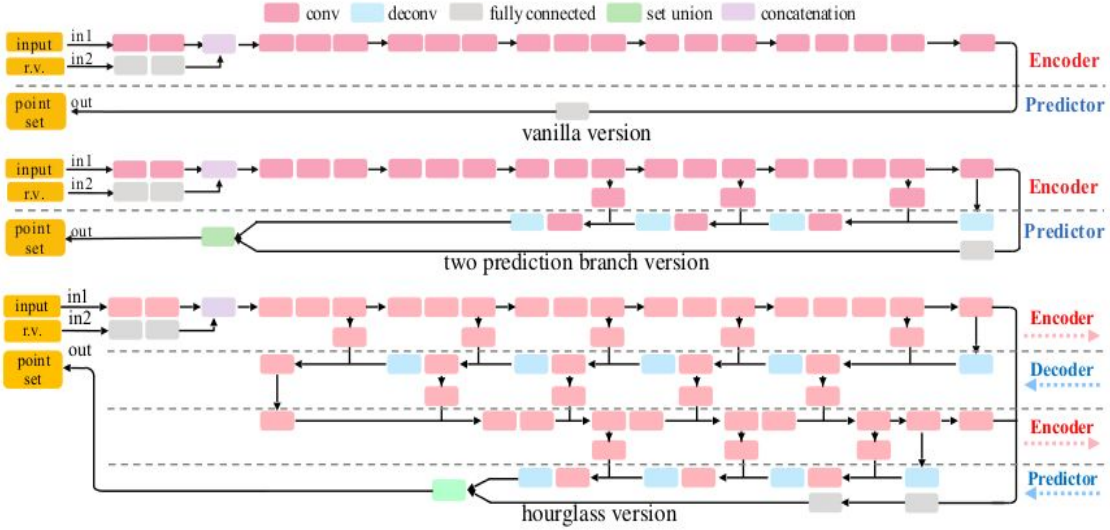
flexible

structural

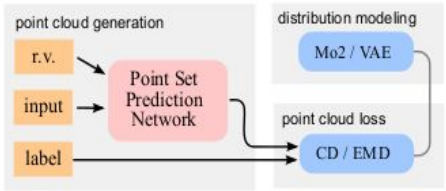
natural

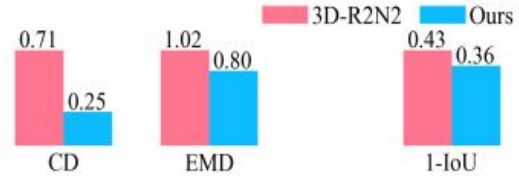


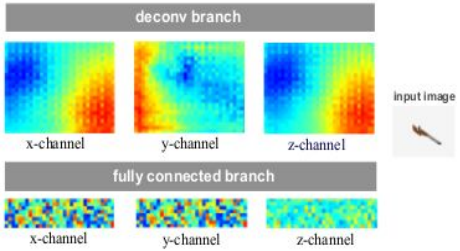
Implementation details



category	Ours	3D-R2N2		
	1 view	1 view	3 views	5 views
plane	**0.601**	0.513	0.549	0.561
bench	**0.550**	0.421	0.502	0.527
cabinet	0.771	0.716	0.763	**0.772**
car	0.831	0.798	0.829	**0.836**
chair	0.544	0.466	0.533	**0.550**
monitor	0.552	0.468	0.545	**0.565**
lamp	**0.462**	0.381	0.415	0.421
speaker	**0.737**	0.662	0.708	0.717
firearm	**0.604**	0.544	0.593	0.600
couch	**0.708**	0.628	0.690	0.706
table	**0.606**	0.513	0.564	0.580
cellphone	0.749	0.661	0.732	**0.754**
watercraft	**0.611**	0.513	0.596	0.610
mean	**0.640**	0.560	0.617	0.631

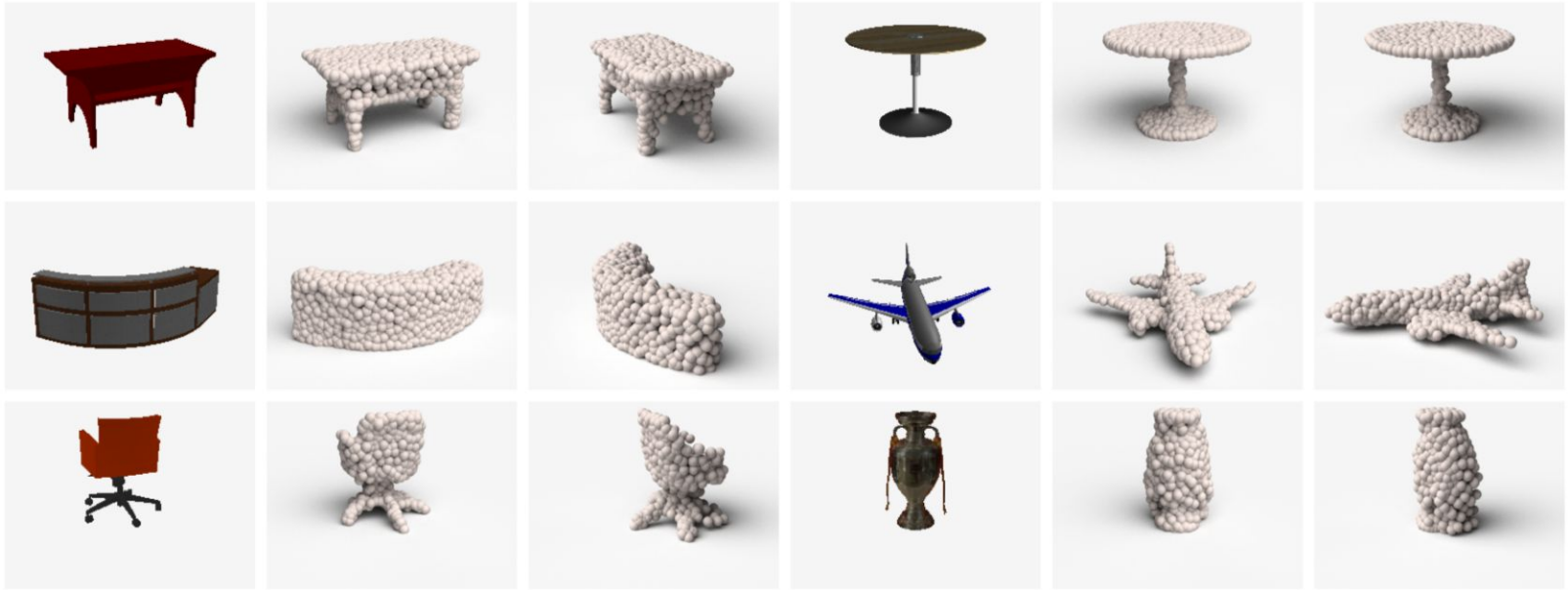




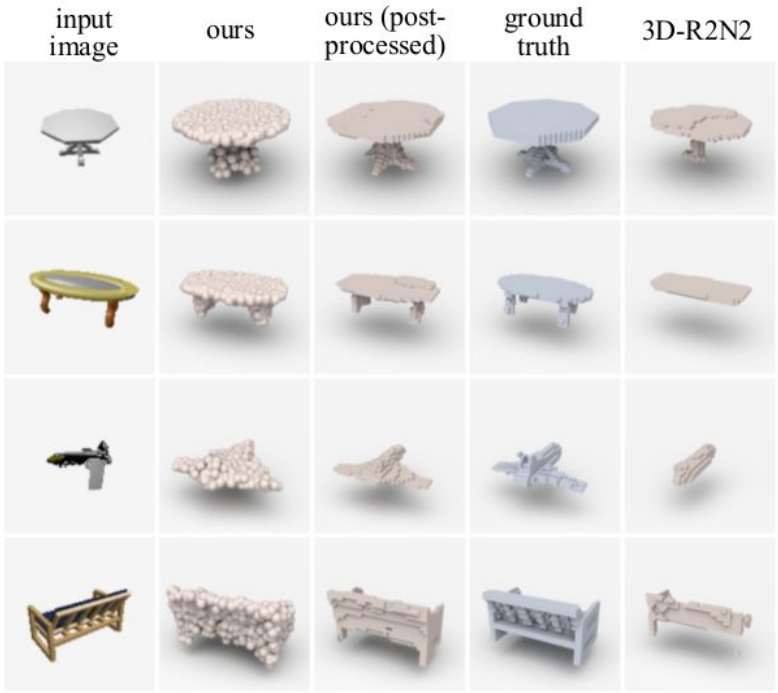


Face++ 旷视

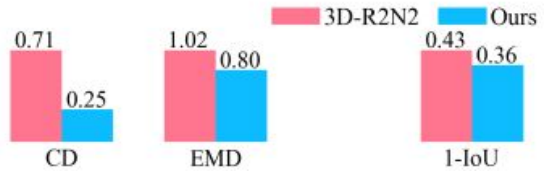
Results



Results

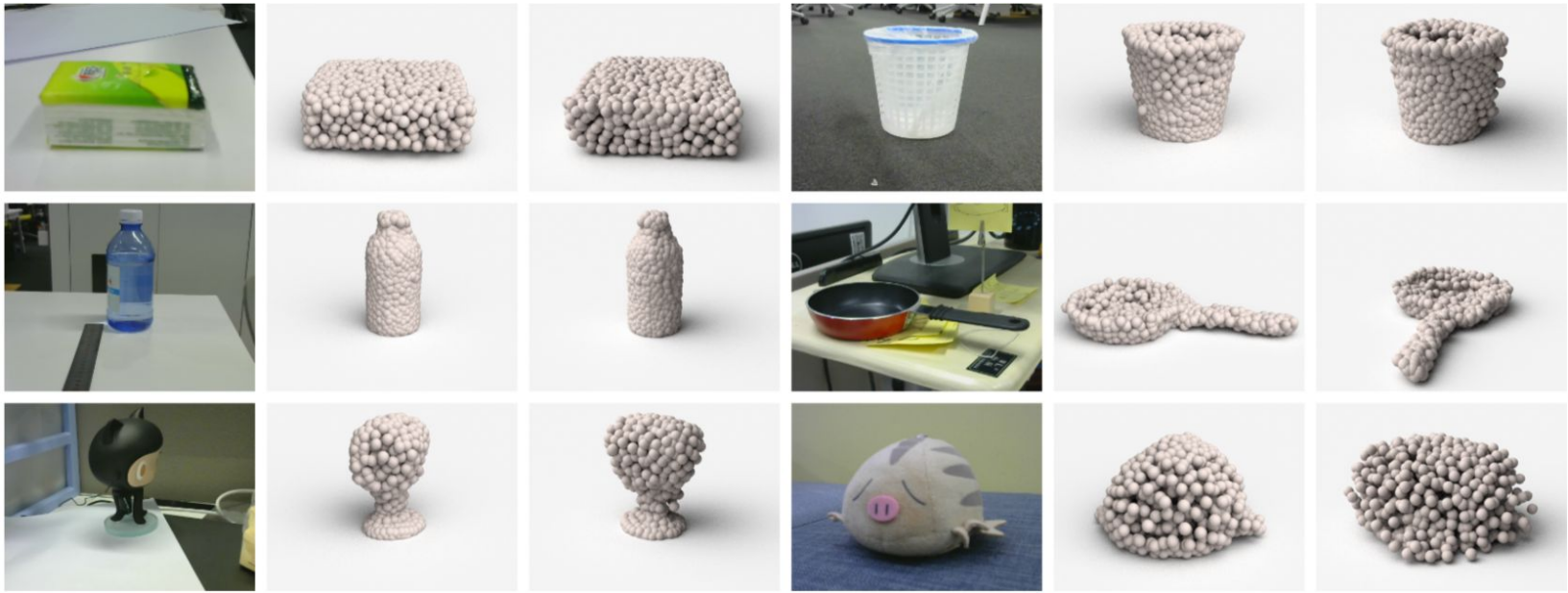


category	Ours	3D-R2N2		
	1 view	1 view	3 views	5 views
plane	**0.601**	0.513	0.549	0.561
bench	**0.550**	0.421	0.502	0.527
cabinet	0.771	0.716	0.763	**0.772**
car	0.831	0.798	0.829	**0.836**
chair	0.544	0.466	0.533	**0.550**
monitor	0.552	0.468	0.545	**0.565**
lamp	**0.462**	0.381	0.415	0.421
speaker	**0.737**	0.662	0.708	0.717
firearm	**0.604**	0.544	0.593	0.600
couch	**0.708**	0.628	0.690	0.706
table	**0.606**	0.513	0.564	0.580
cellphone	0.749	0.661	0.732	**0.754**
watercraft	**0.611**	0.513	0.596	0.610
mean	**0.640**	0.560	0.617	0.631

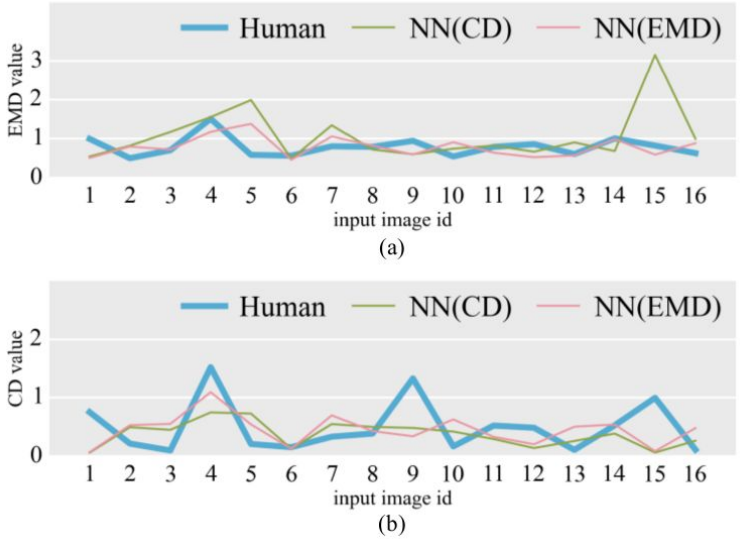


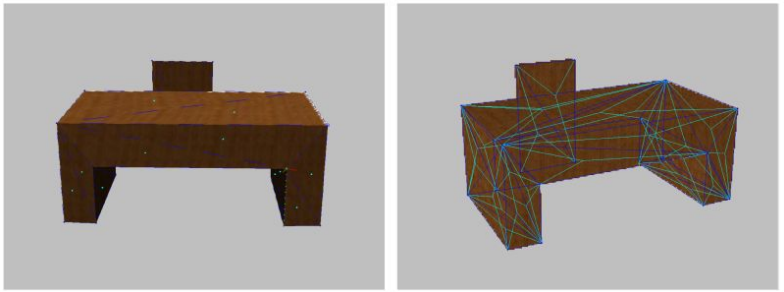
Face++ 旷视

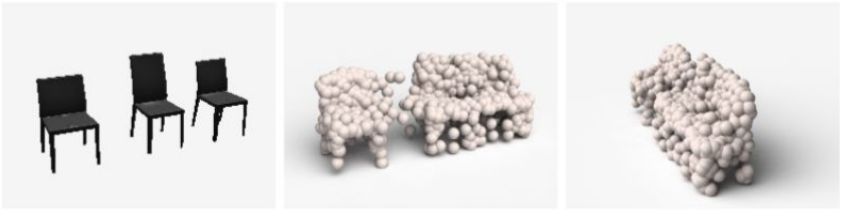
Results



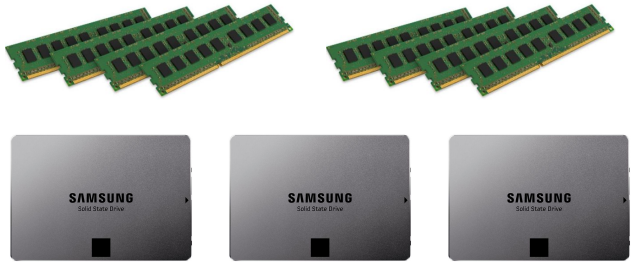
Human Performance

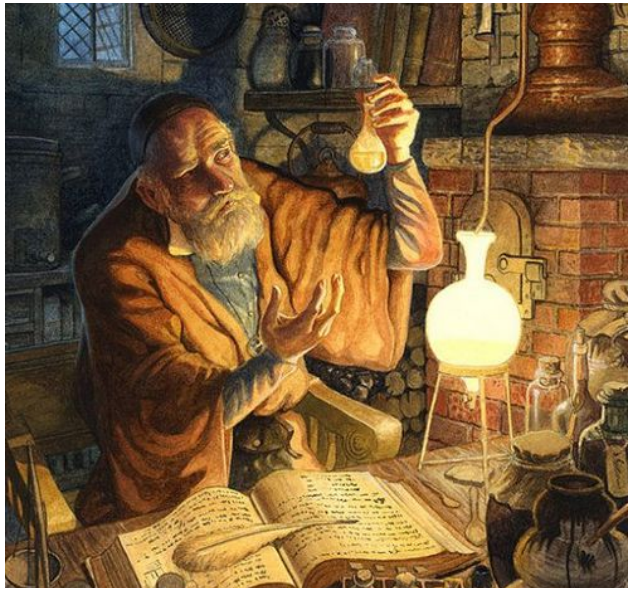







TITAN
TITAN
TITAN
TITAN
TITAN
TITAN
TITAN
TITAN


SAMSUNG
SAMSUNG
SAMSUNG



Face++ 旷视


EXIT
EXIT

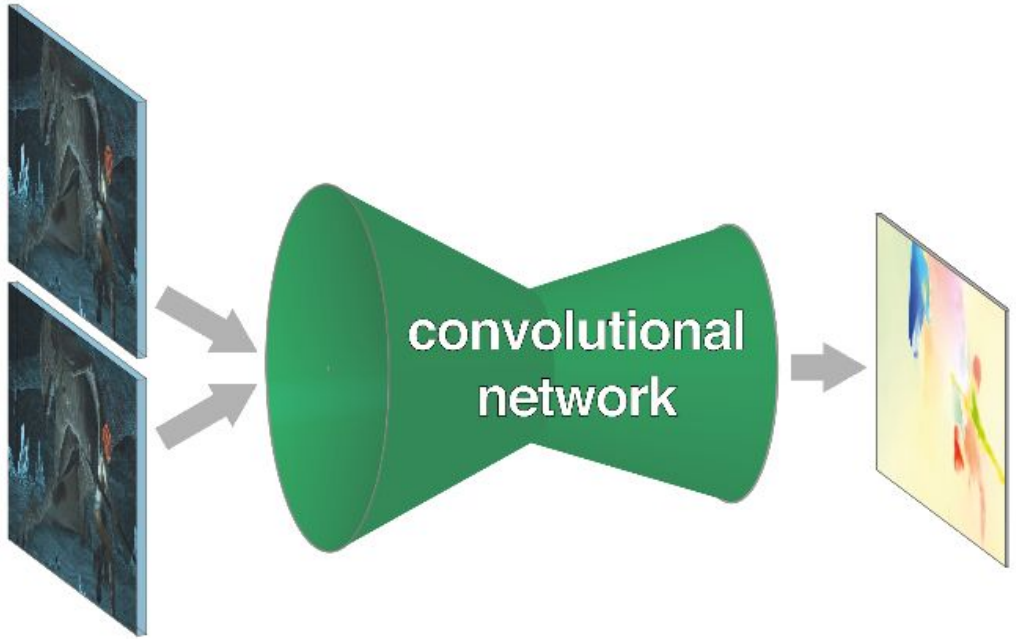
A Neural Method to Stereo Matching

Flownet & Dispnet

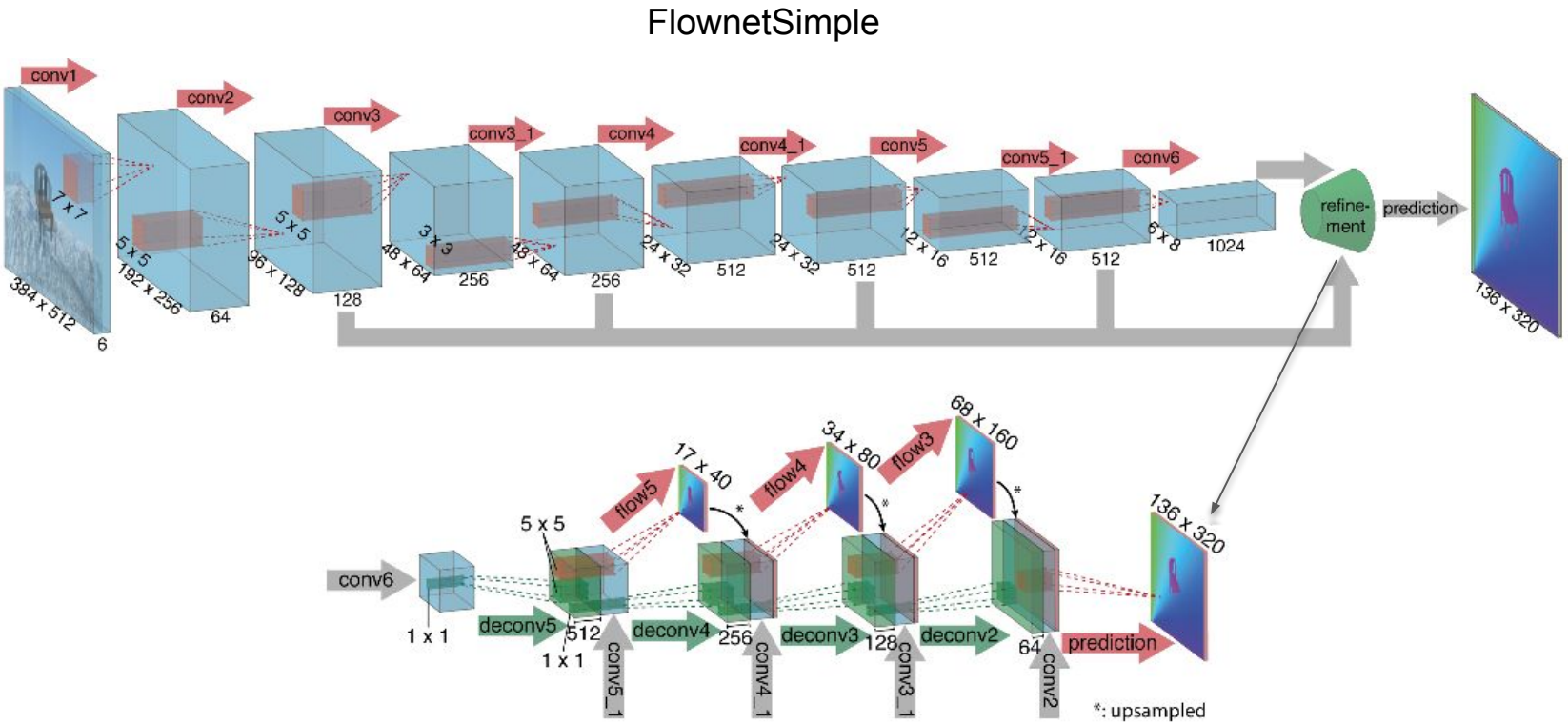
Using raw left and right images as input

Output disparity map

End-to-End training



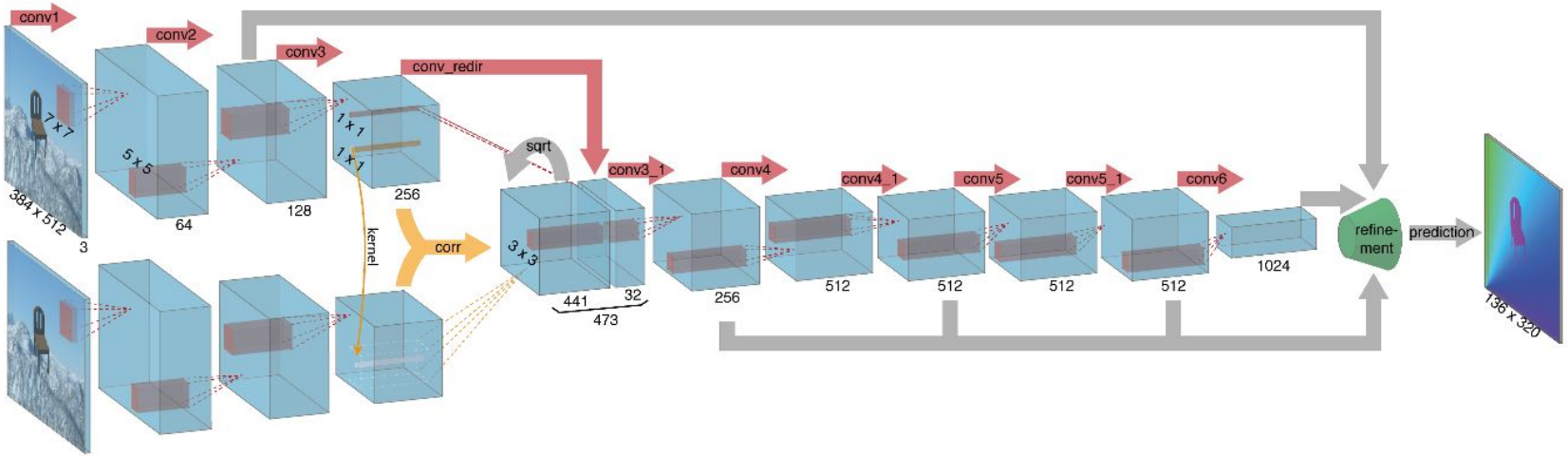
Using two stacked images as input



Adding Correlation Layer

Using correlation layer to explicitly provide cross view communication ability

FlownetCorr



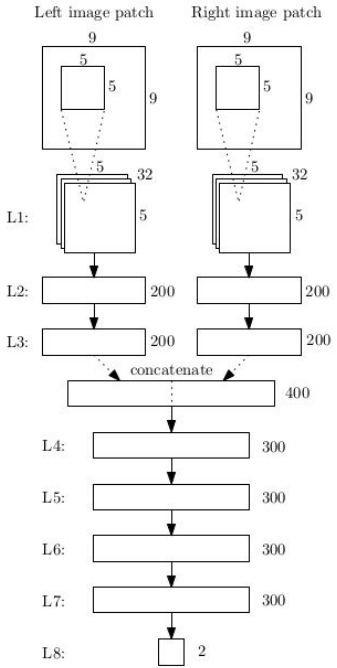
Stereo Matching Cost Convolutional Neural Network

Using CNN to calculate stereo matching cost between patches from different view

Following with several post-process:

- Cross-based cost aggregation
- Semiglobal matching
- Left-right consistency check

Disparity <-> Depth $z = \frac{fB}{d}$



MRF Stereo methods

$f_p = (a_p, b_p, c_p) \in \mathcal{L}$ for every pixel p

We estimate f by minimizing the following energy function based on pairwise MRF

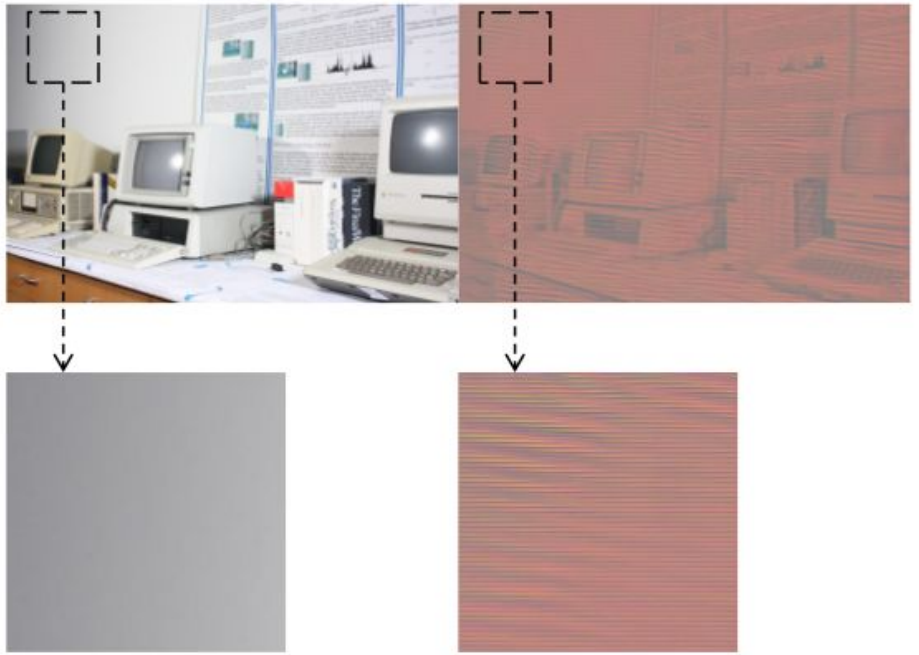
$$E(f) = \sum_{p \in \Omega} \phi_p(f_p) + \lambda \sum_{(p,q) \in \mathcal{N}} \psi_{pq}(f_p, f_q).$$

Data term $\phi_p(f_p)$

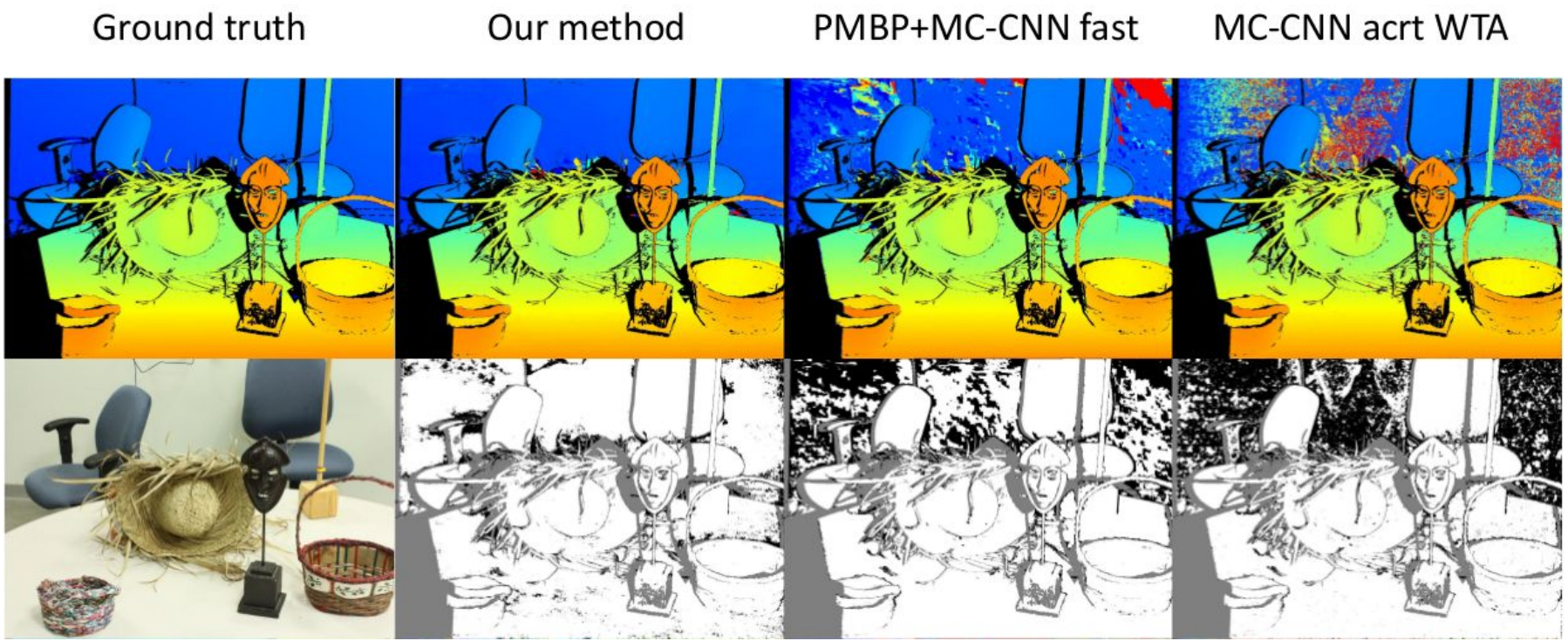
Smoothness term $\psi_{pq}(f_p, f_q)$

Global Local Stereo Neural Network

Feature visualization

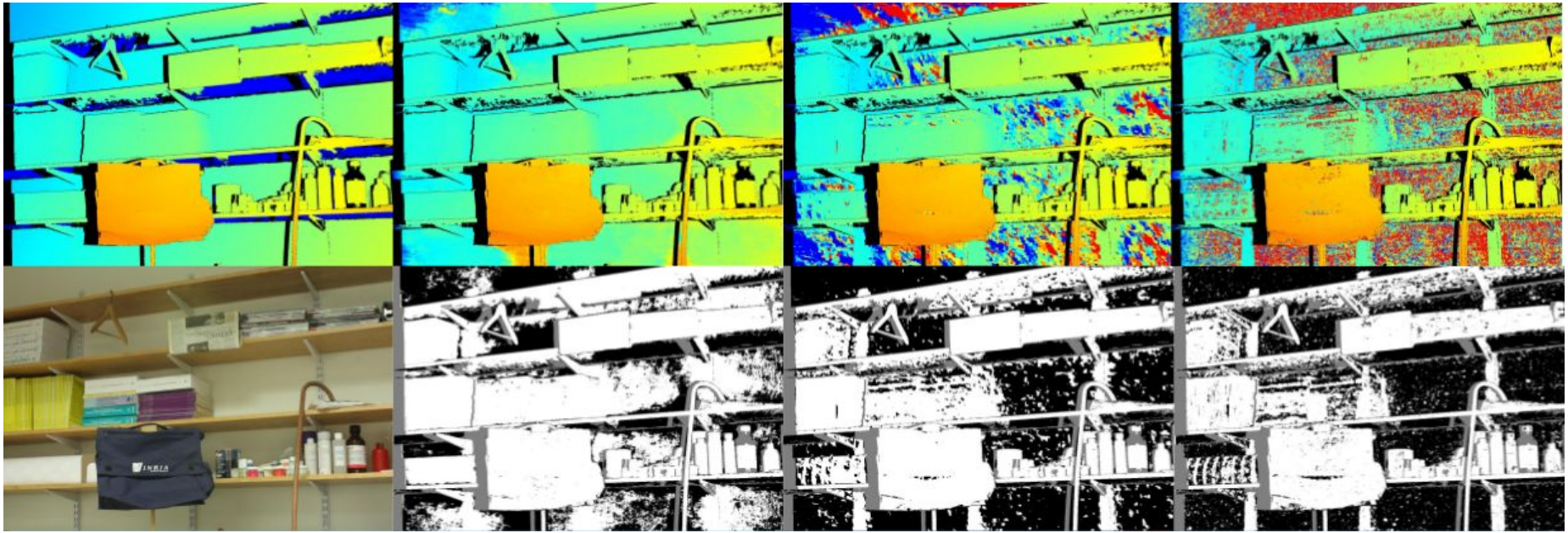


results

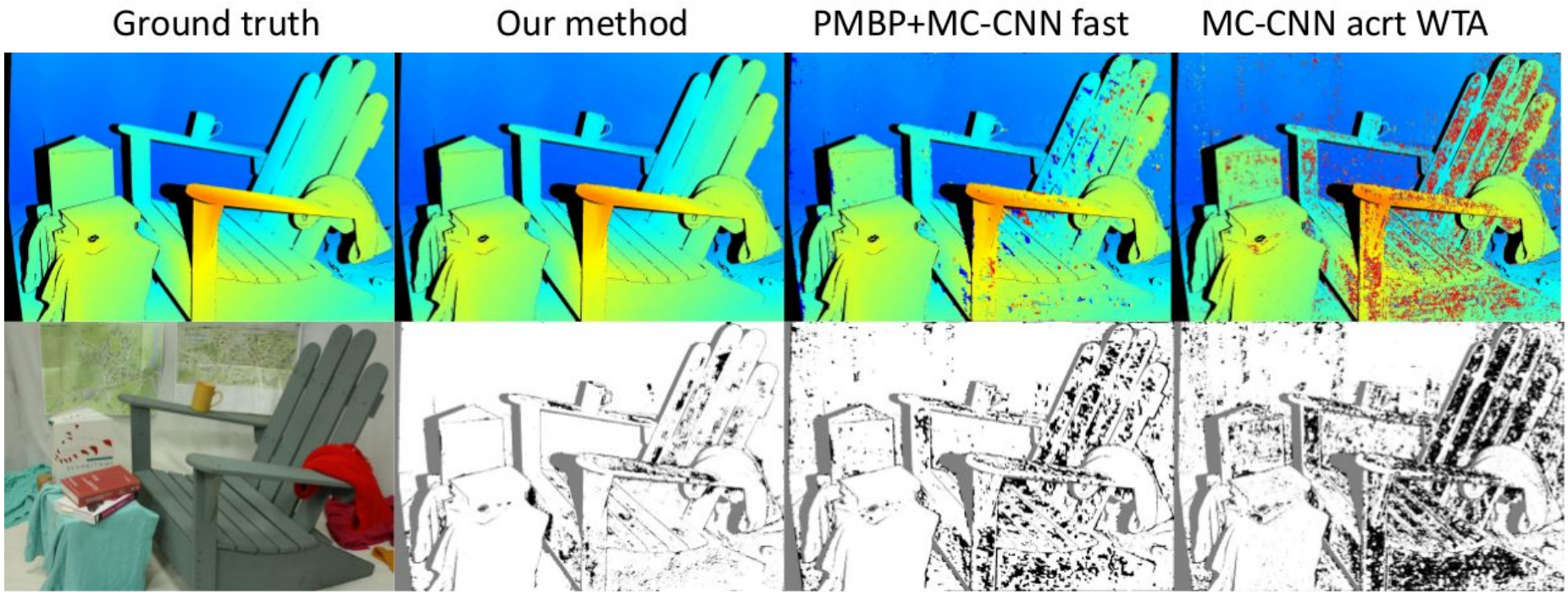


results

Ground truth | Our method | PMBP+MC-CNN fast | MC-CNN acrt WTA

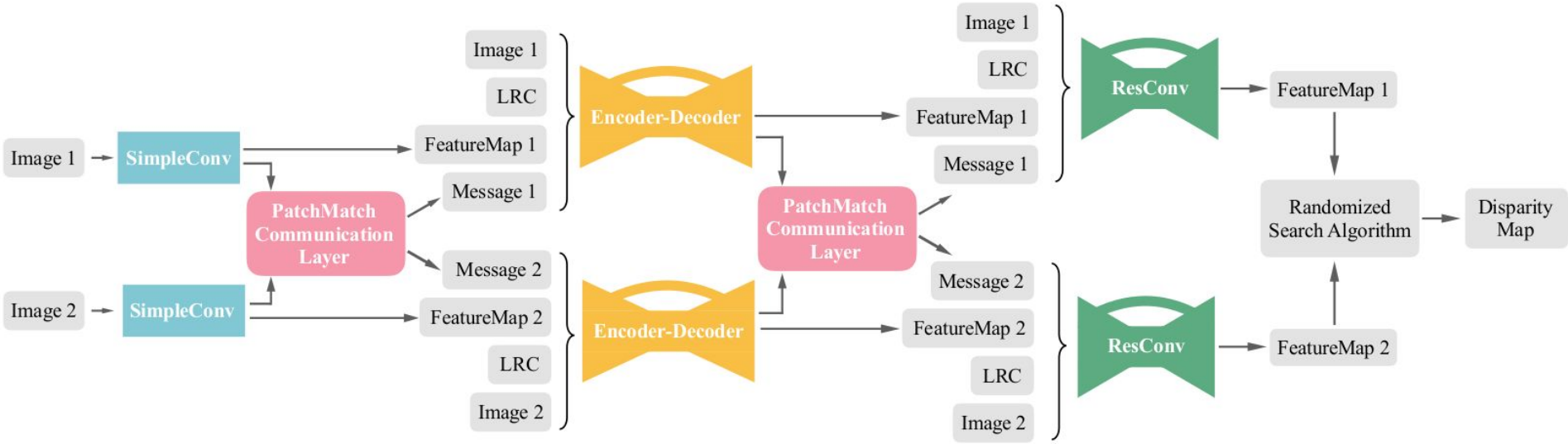


results



Implementation details

Entangle two view feature inside network.



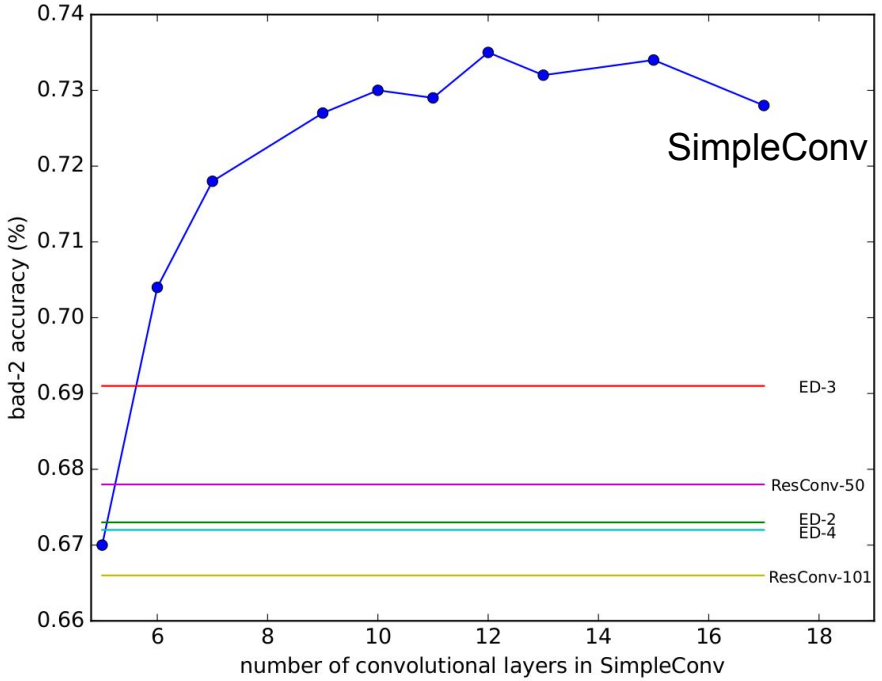
Large Receptive Field Neural Network

SimpleConv

Encoder-Decoder

ResConv

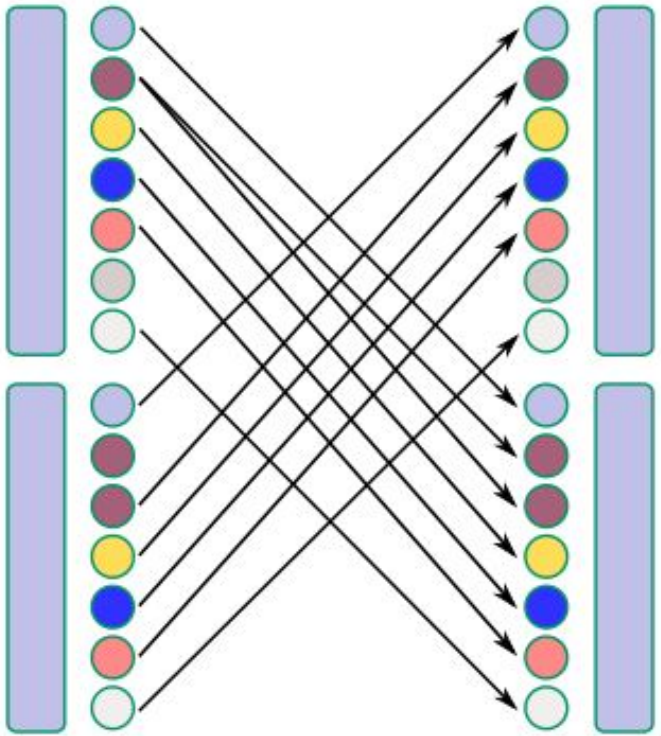
blindingly increasing the receptive field of feature networks may not Improve the performance



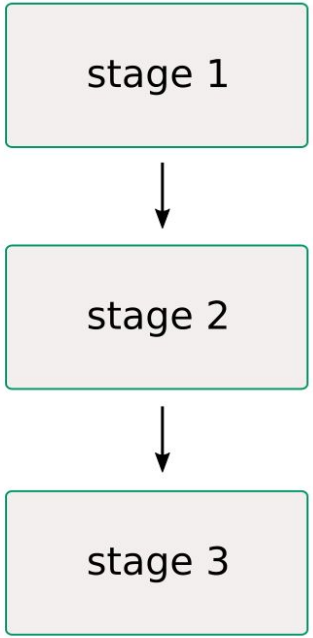
PatchMatch Communication Layer

Directly provide the ability of

communicating across two views



Multi-staged Cascade



Combinations	bad-2 error	bad-2 error with PCL
SS	0.334	0.332
SE	0.249	0.231
SSS	0.278	0.250
SSE	0.246	0.242
SSR	0.252	0.243
EES	0.327	0.295
ESE	0.247	0.242
SEE	0.248	0.218
SER	**0.245**	**0.211**

Thanks

Q/A

单击以结束放映

SemiGlobal Matching

we define an energy function E(D) that depends on the disparity map D

NP-Hard !!! But we can solve it through each directions to get an approximate solution by using Dynamic Programming(DP)

$$E(D) = \sum_{\mathbf{p}} \Bigg(C^4_{\text{CBCA}}(\mathbf{p}, D(\mathbf{p})) + \sum_{\mathbf{q} \in \mathcal{N}_\mathbf{p}} P_1 \times 1\{|D(\mathbf{p}) - D(\mathbf{q})| = 1\} + \sum_{\mathbf{q} \in \mathcal{N}_\mathbf{p}} P_2 \times 1\{|D(\mathbf{p}) - D(\mathbf{q})| > 1\} \Bigg)$$

$$C_\mathbf{r}(\mathbf{p}, d) = C^4_{\text{CBCA}}(\mathbf{p}, d) - \min_k C_r(\mathbf{p} - \mathbf{r}, k) + \min \Big\{ C_r(\mathbf{p} - \mathbf{r}, d), C_r(\mathbf{p} - \mathbf{r}, d-1) + P_1, C_r(\mathbf{p} - \mathbf{r}, d+1) + P_1, \min_k C_\mathbf{r}(\mathbf{p} - \mathbf{r}, k) + P_2 \Big\}.$$

Slanted patch matching

The disparity d_p of each pixel p is over-parameterized by a local disparity plane

$$d_p = a_p u + b_p v + c_p$$

Each pixels in the same plane has the same parameter (a_p, b_p, c_p)

The true disparity maps are approximately piecewise linear

We can estimate (a_p, b_p, c_p) for each pixel p instead of directly estimate d_p