Learning 3D Vision

Computer Vision
Fall 2018
Columbia University

Project FAQ

- **“What happens if my proposal is different from my final project?”**
  - We want to see that you have a clear idea and plausible path to the solution
  - My first attempt almost never works.
  - Just do a cool project!

Project FAQ

- **“I don’t have 500 GPUs”**
  - In my last paper, I only used 1 GPU, no parameter tuning
  - Isolate your idea and focus on it

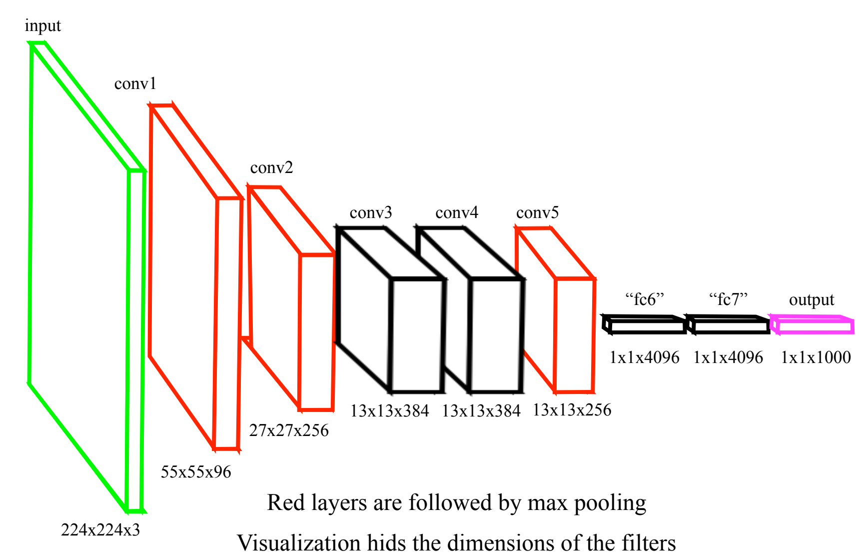
Binocular Stereo

Key Idea: use feature motion to understand shape

Photometric Stereo

Key Idea: use pixel brightness to understand shape

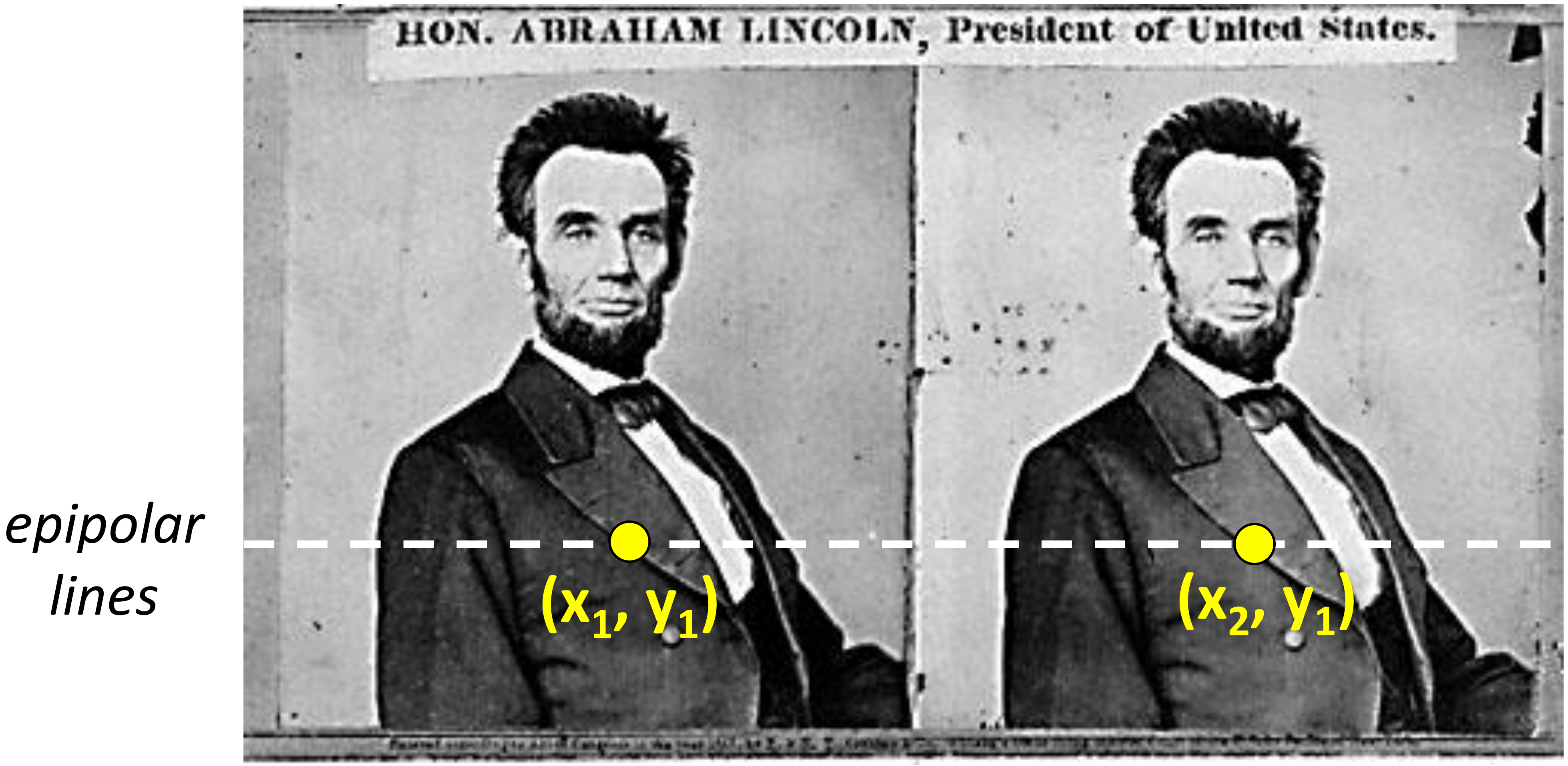
Learning-based 3D



Key Idea: learn it from data

Why learn 3D vision?

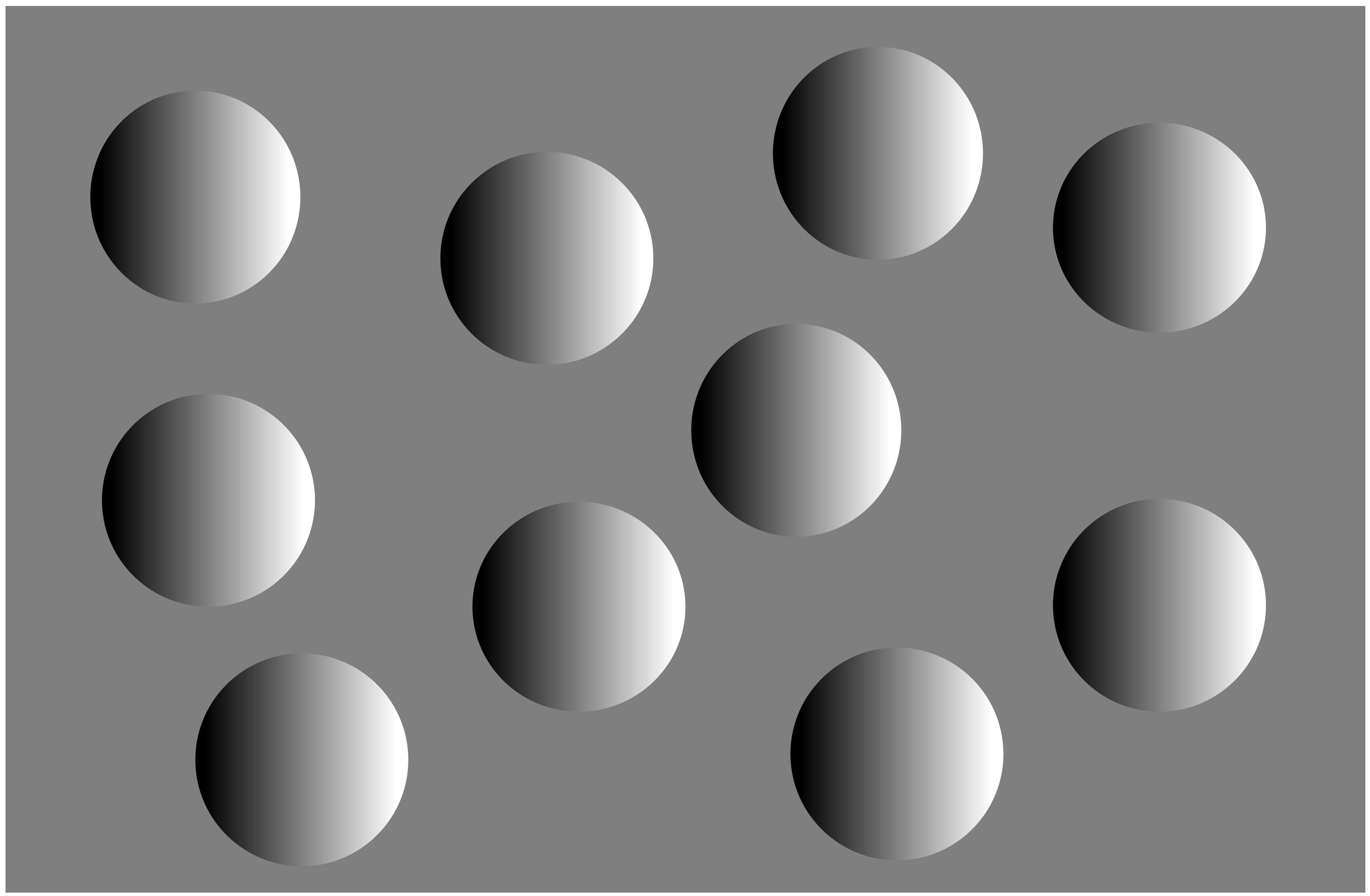
Why learn 3D vision?



Matching is hard

Why learn 3D vision?

 ?



 ?

Structure depends on priors

Why learn 3D vision?

The world is not ideal diffuse

Why learn 3D vision?

We want to recognize objects in 3D

Single image
Multiple images
Depth image
...
3D voxels

Neural Net

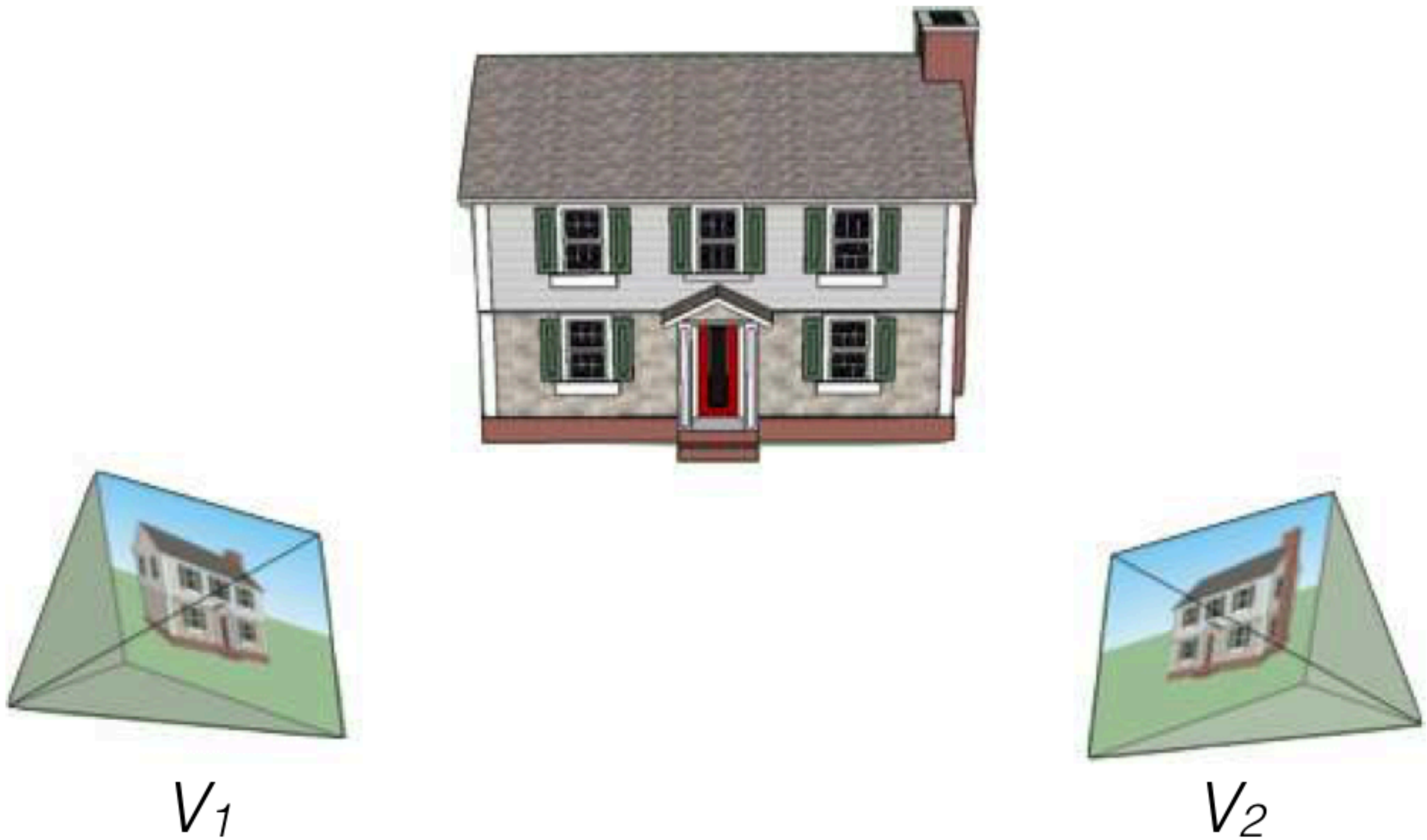
Depth image
Object recognition
View synthesis
...
3D Models

Single image
Multiple images
Depth image
...
3D voxels

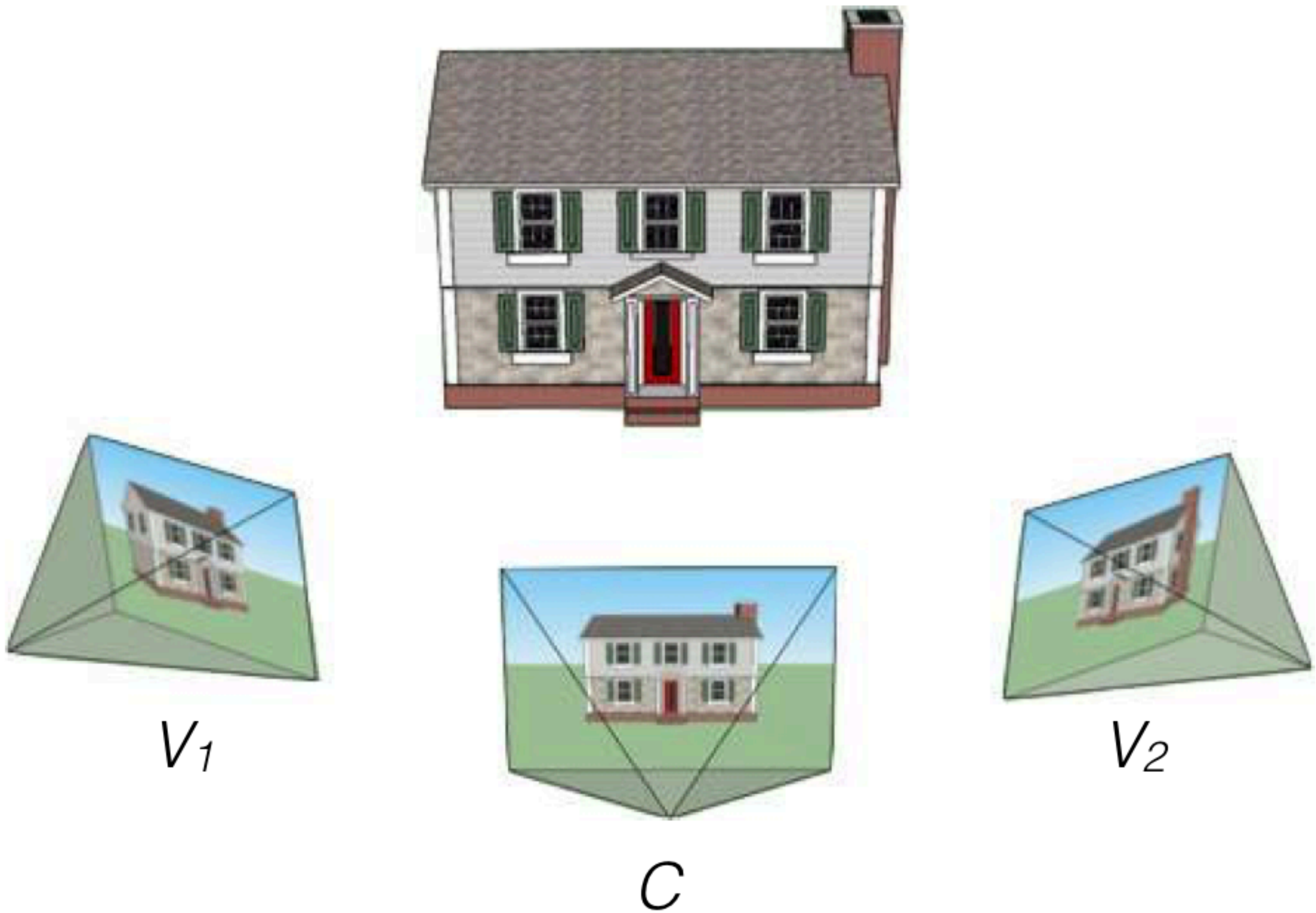
Neural Net

Depth image
Object recognition
View synthesis
...
3D Models

John Flynn et al. DeepStereo

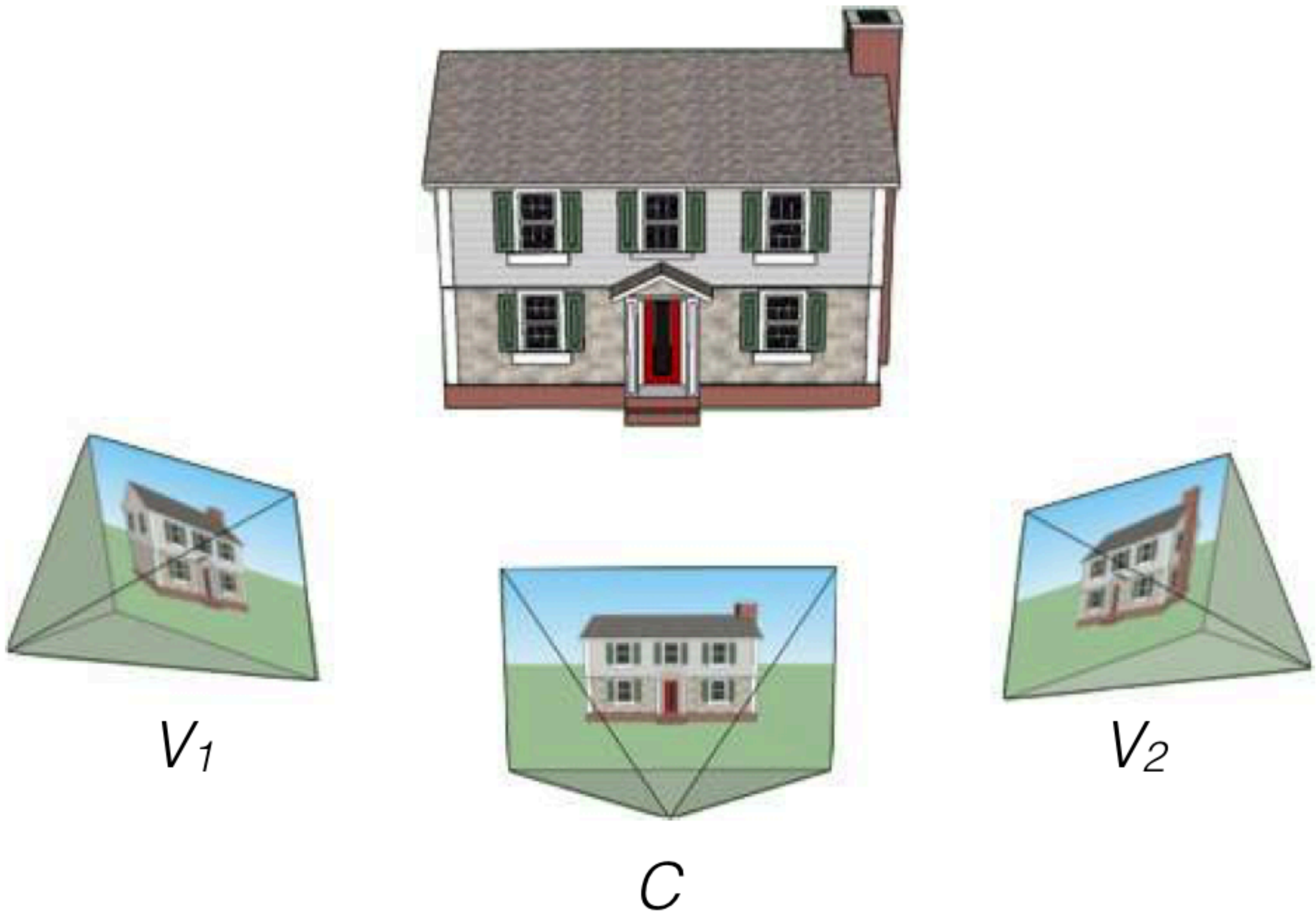


John Flynn et al. DeepStereo



Given v1 and v2, reconstruct the new view C

John Flynn et al. DeepStereo



Stereo with three cameras?

John Flynn et al. DeepStereo

(

Multi-view Stereo

Point Grey's Bumblebee XB3

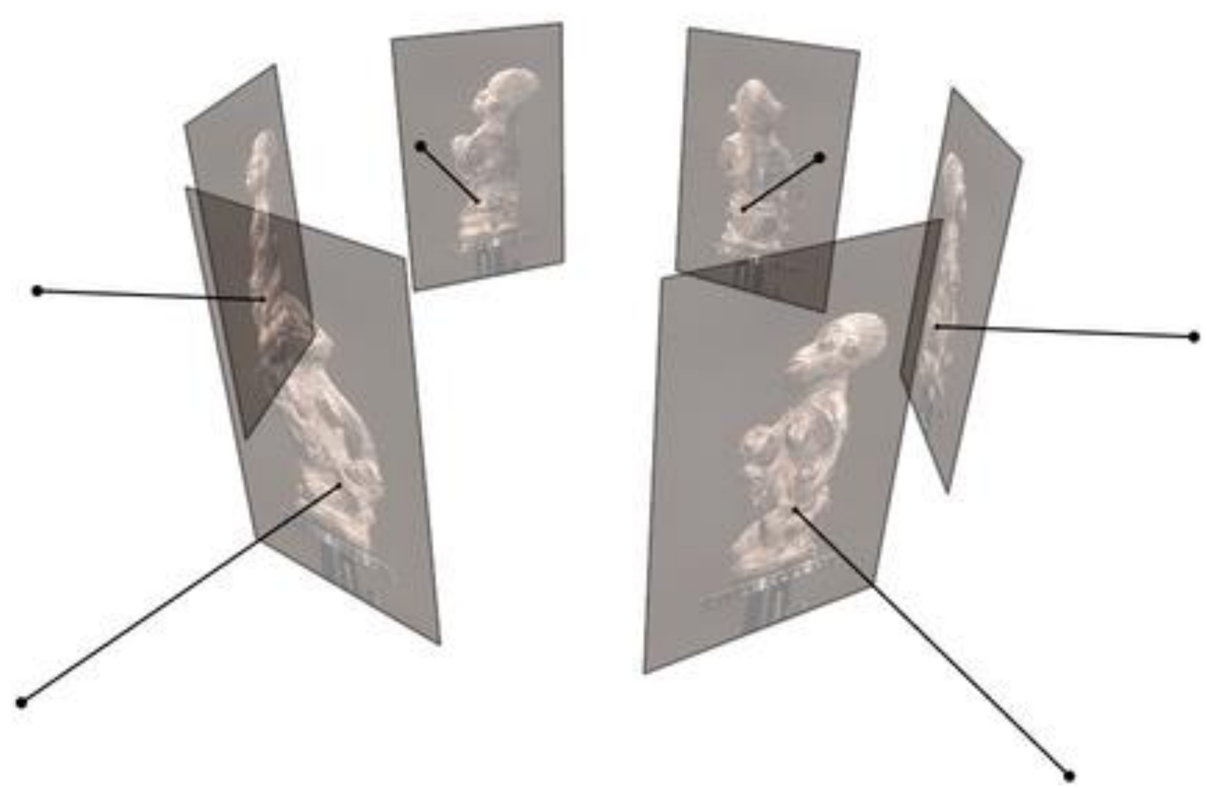
Point Grey's ProFusion 25

CMU's 3D Room

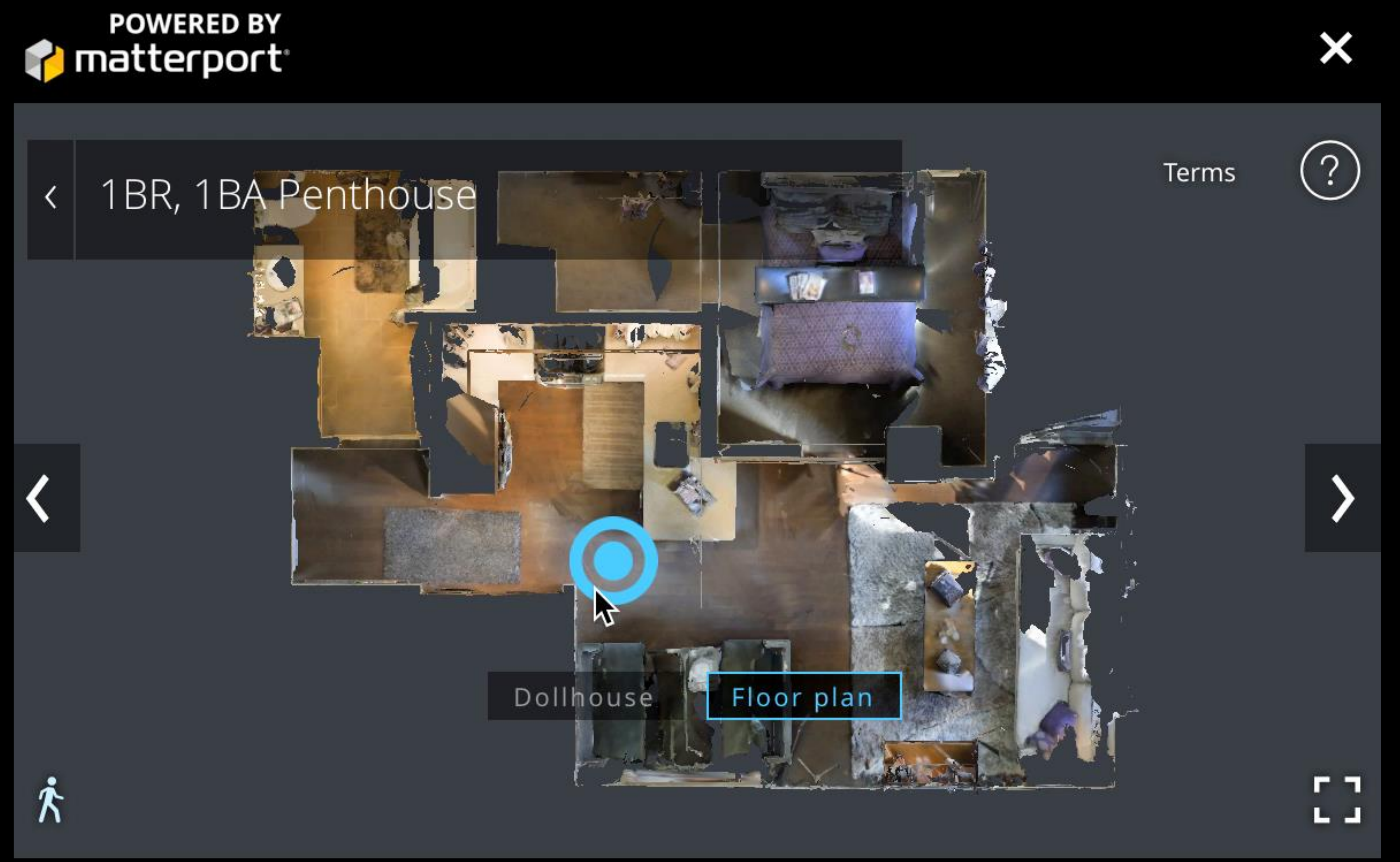
Multi-view Stereo

Input: calibrated images from several viewpoints

Output: 3D object model

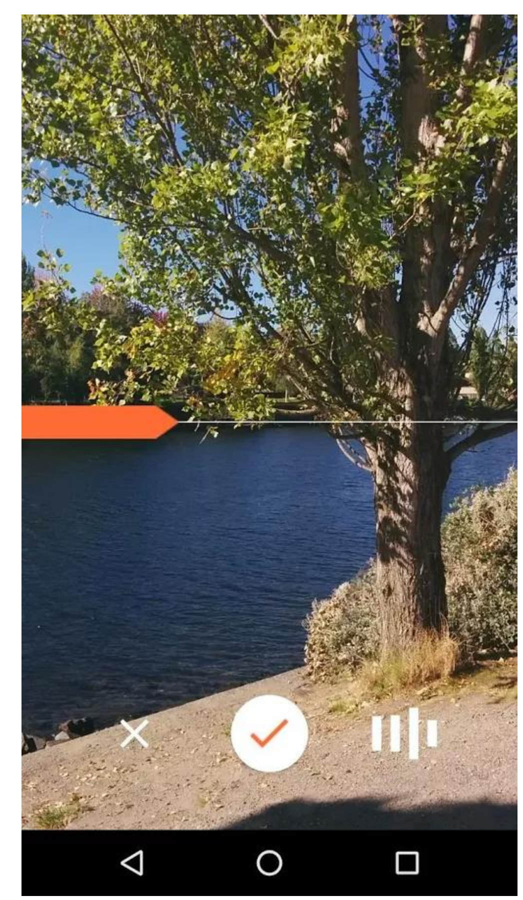


Figures by Carlos Hernandez

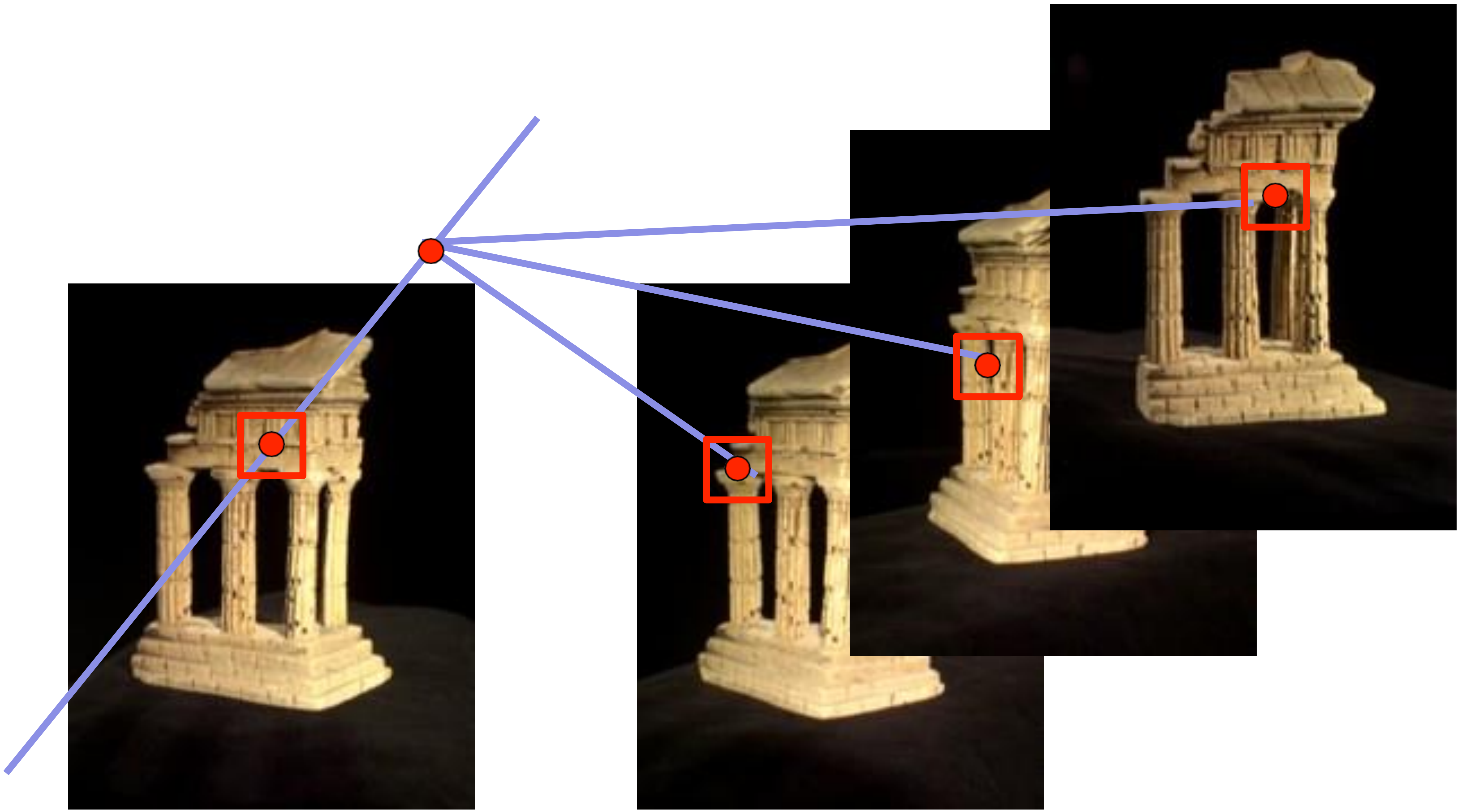

POWERED BY
matterport
1BR, 1BA Penthouse
Terms
Dollhouse
Floor plan

JUMP

Google

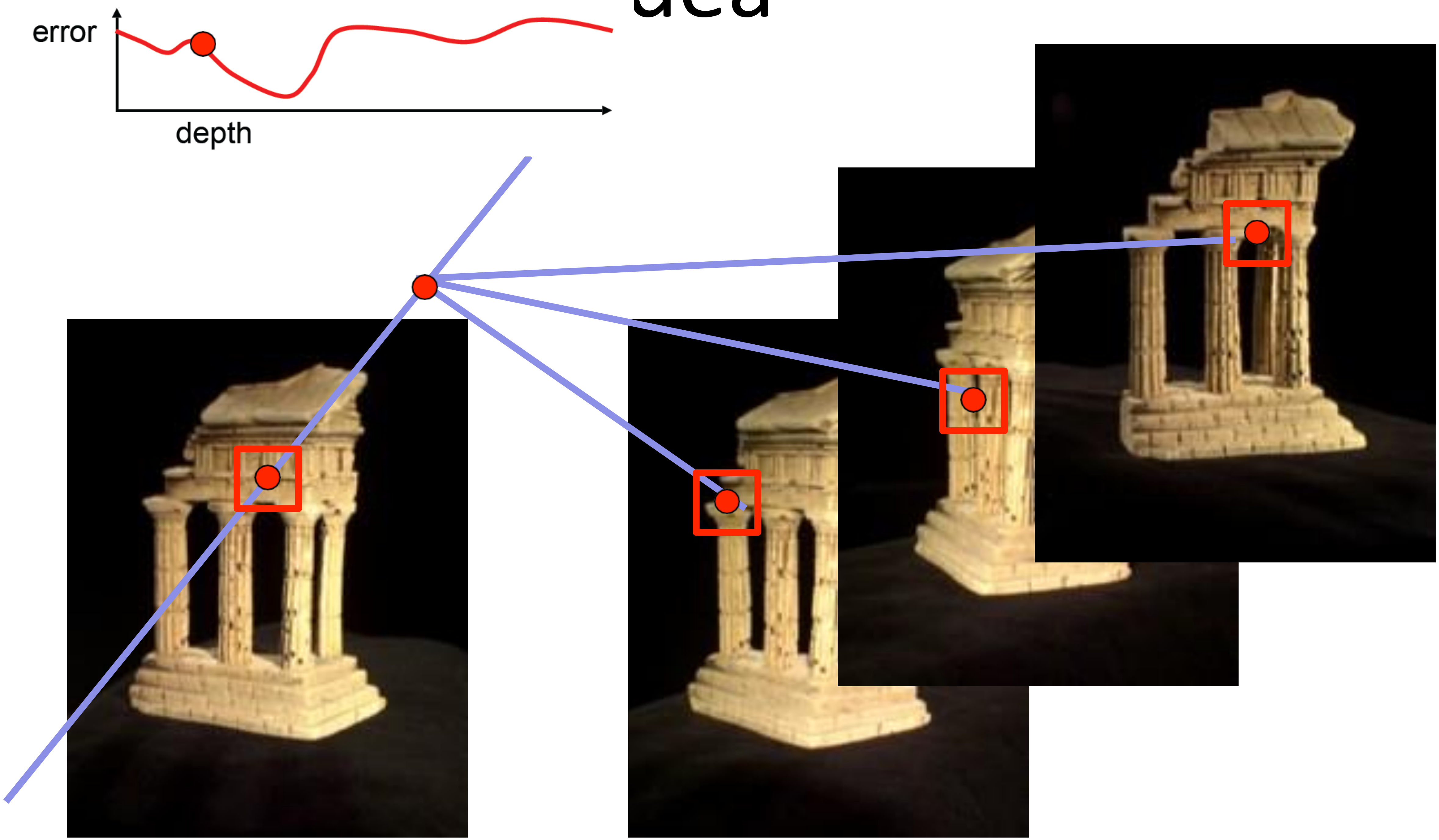


Multi-view stereo: Basic idea



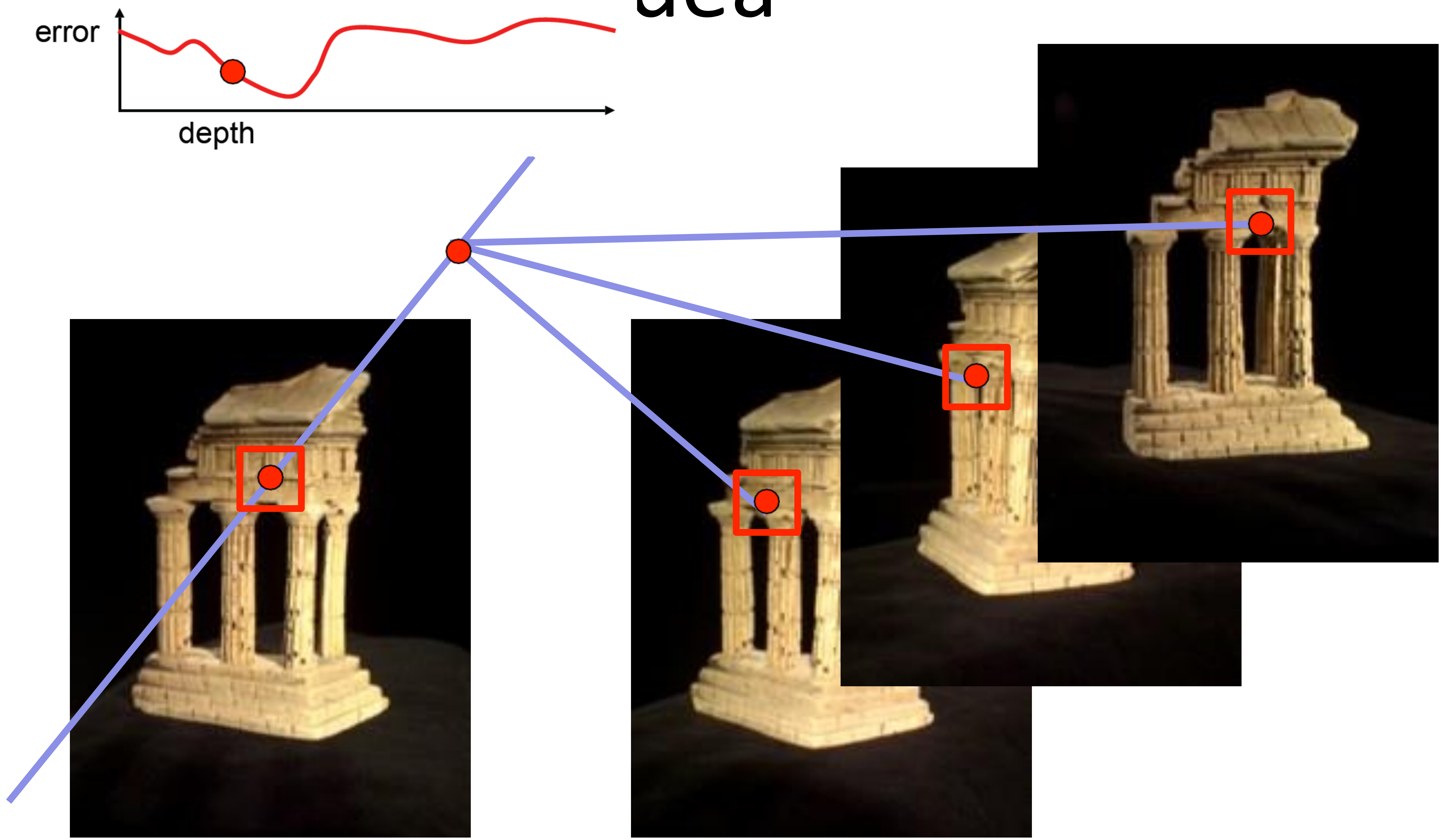
Source: Y. Furukawa

Multi-view stereo: Basic idea



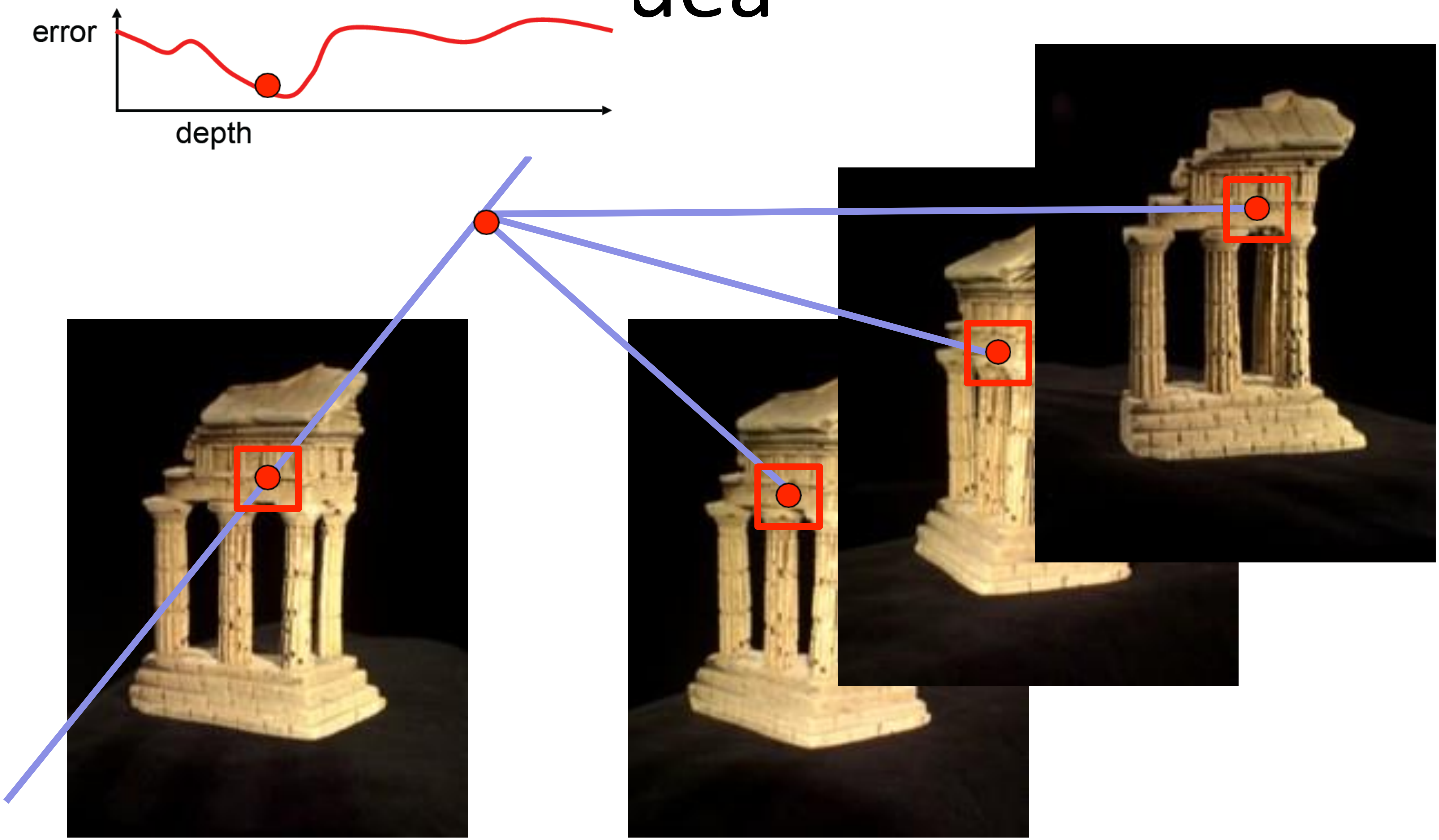
Source: Y. Furukawa

Multi-view stereo: Basic idea



Source: Y. Furukawa

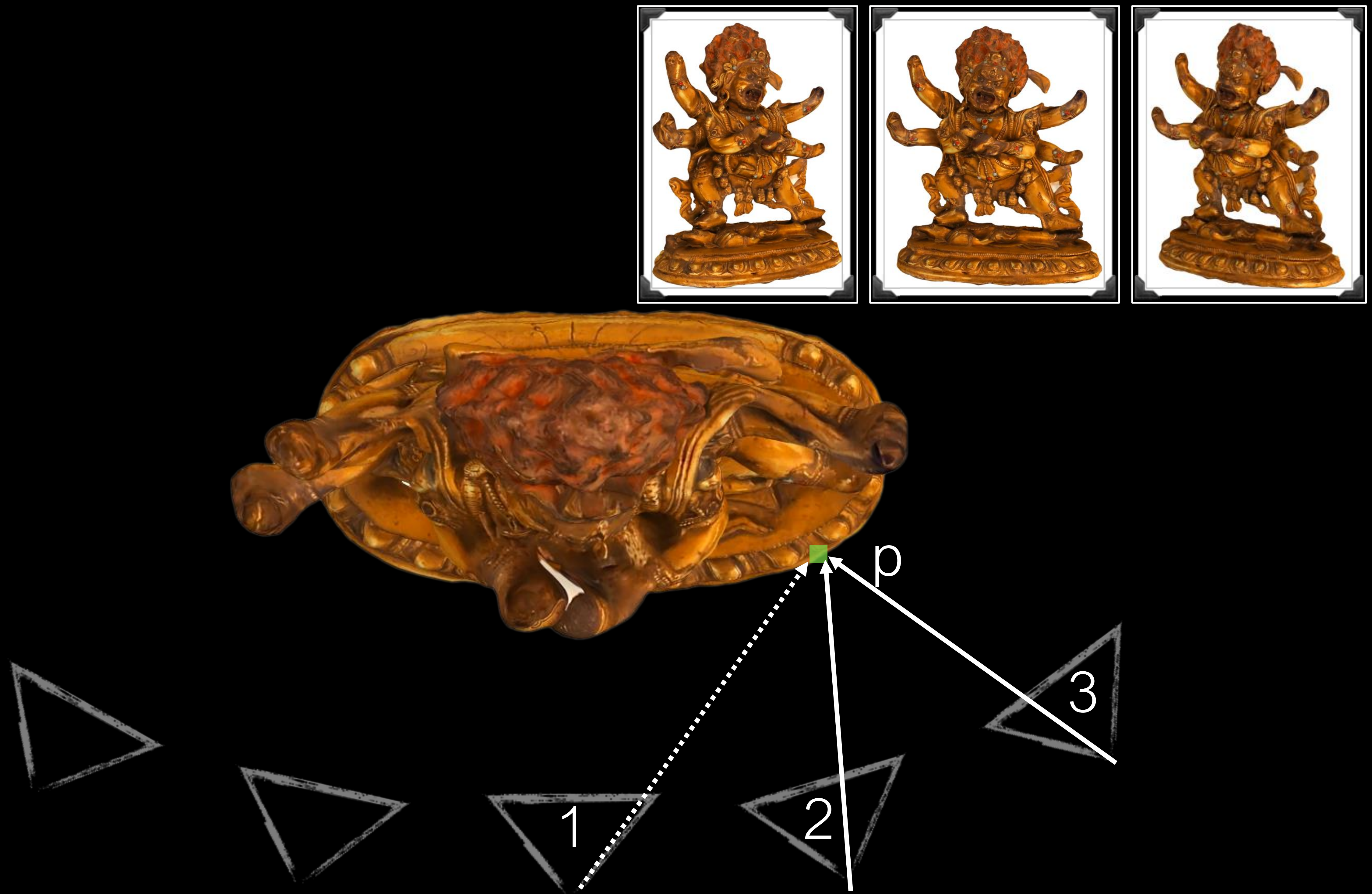
Multi-view stereo: Basic idea



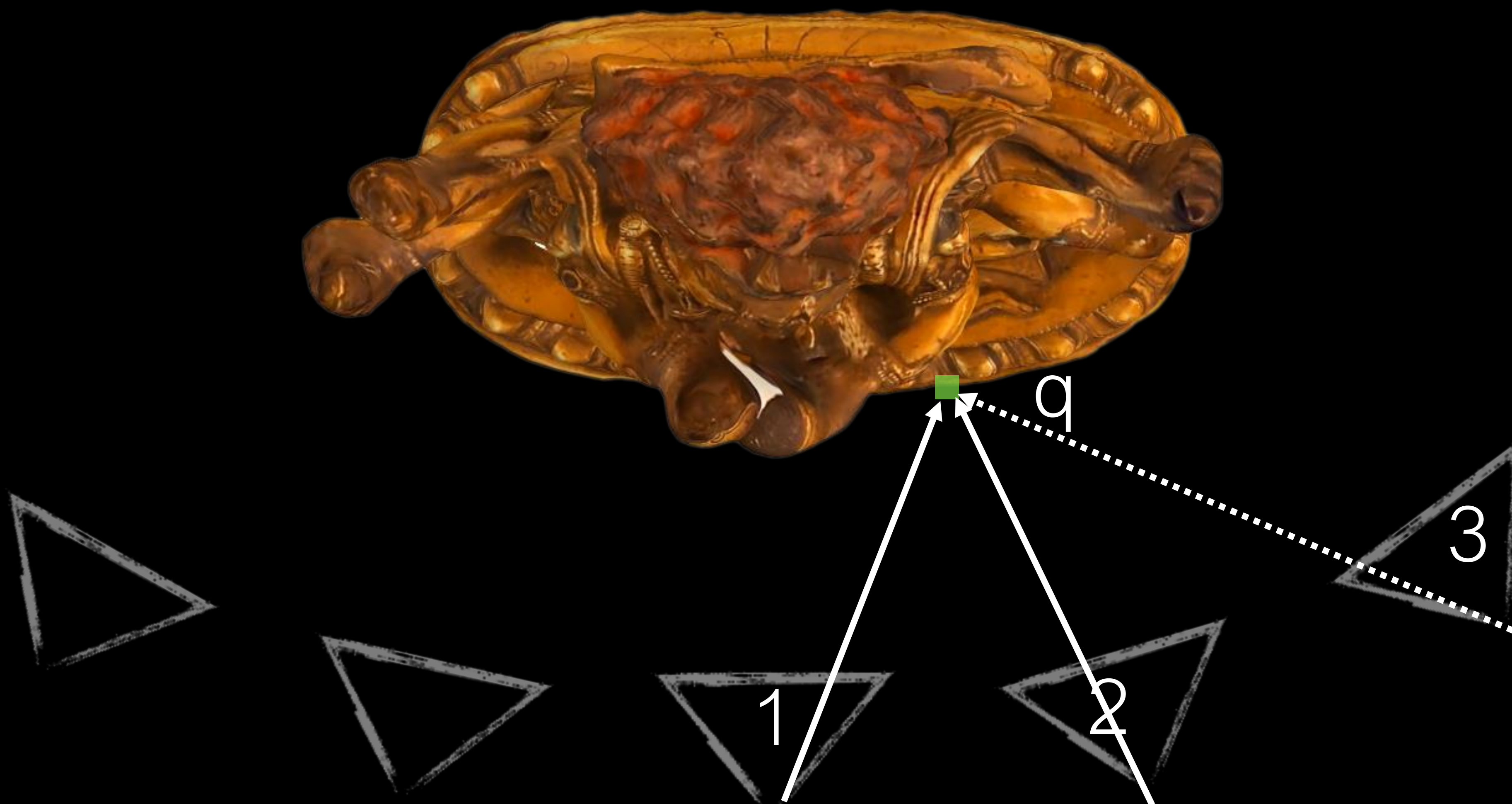
Source: Y. Furukawa

Why multi-view stereo?

1. Some cameras have a closer view



Cameras 2 and 3 can more clearly see point p.

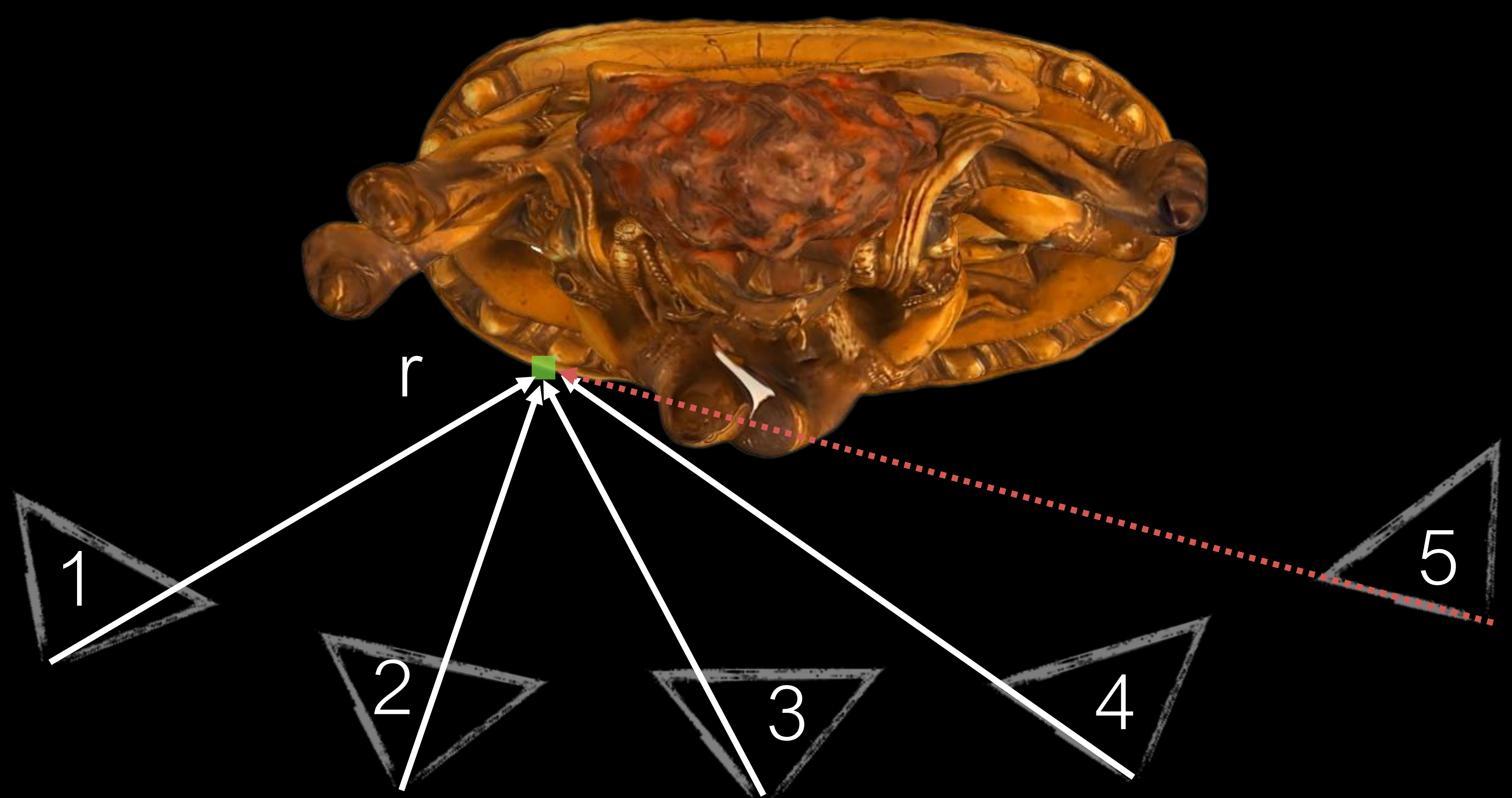


Cameras 1 and 2 can more clearly see point q.

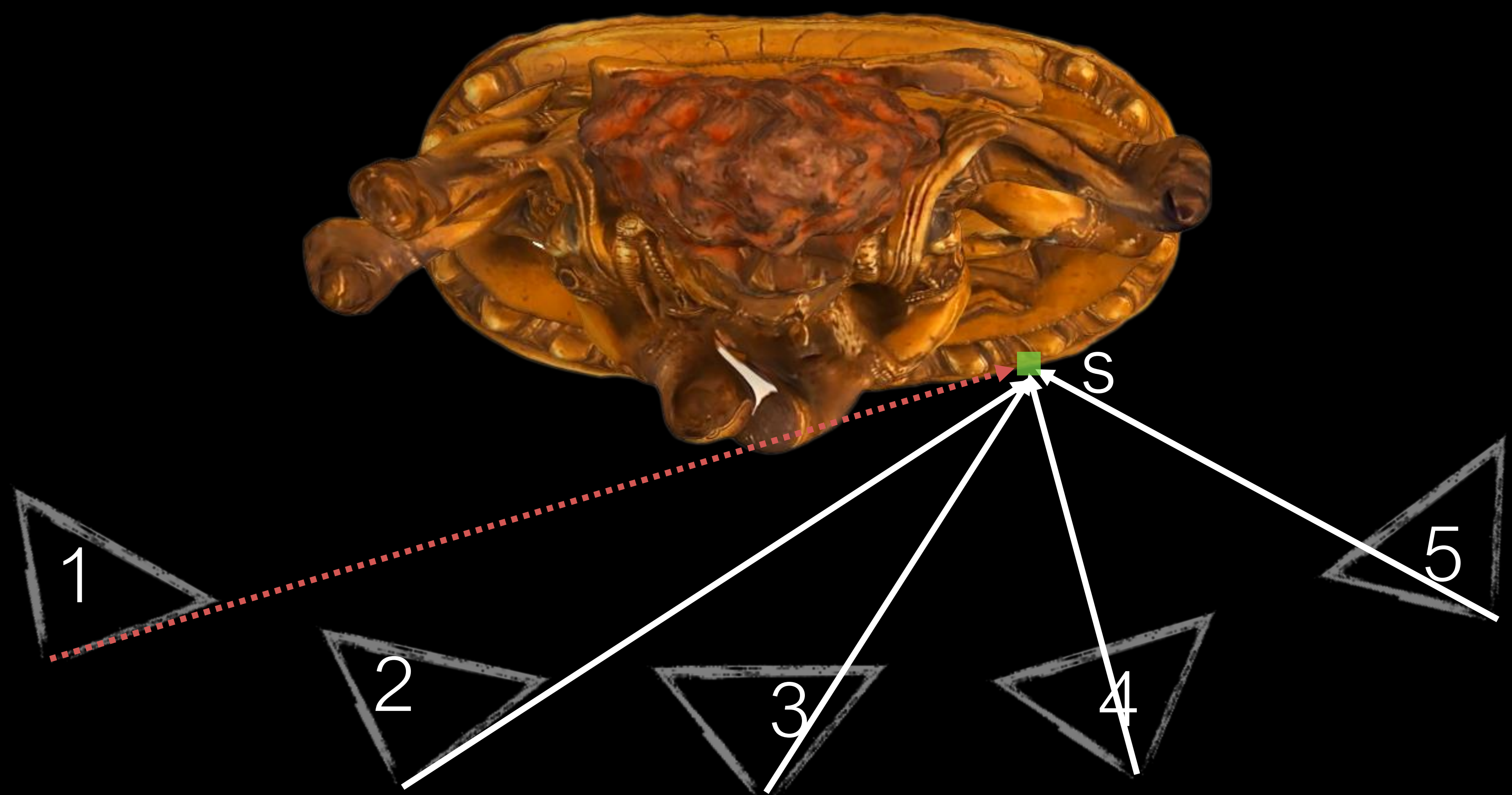
Why multi-view stereo?

1. Some cameras have a closer view

2. Cameras can’t see everything



Camera 5 can’t see point r.



Camera 1 can't see point s.

Why multi-view stereo?

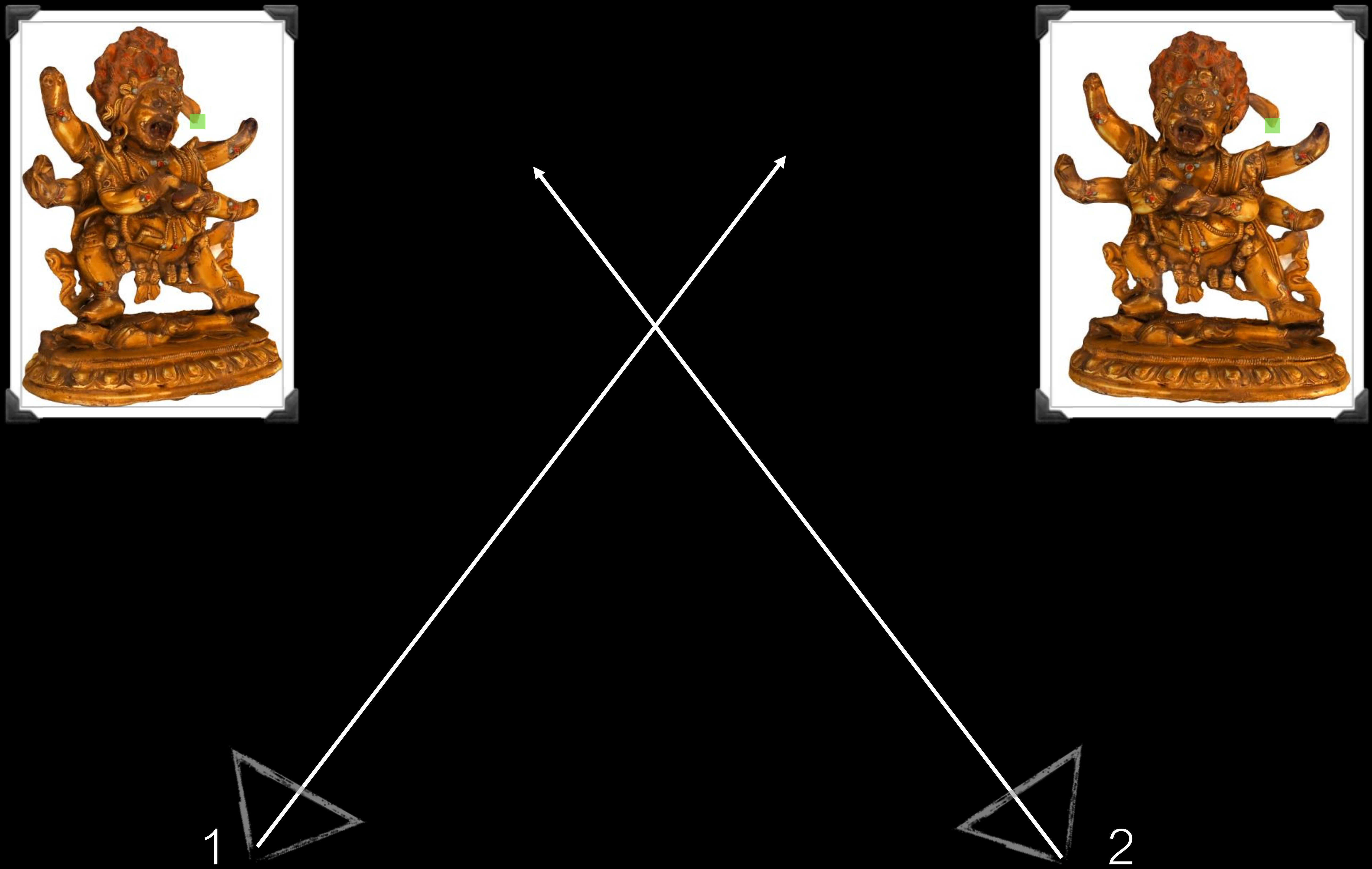
1. Some cameras have a closer view

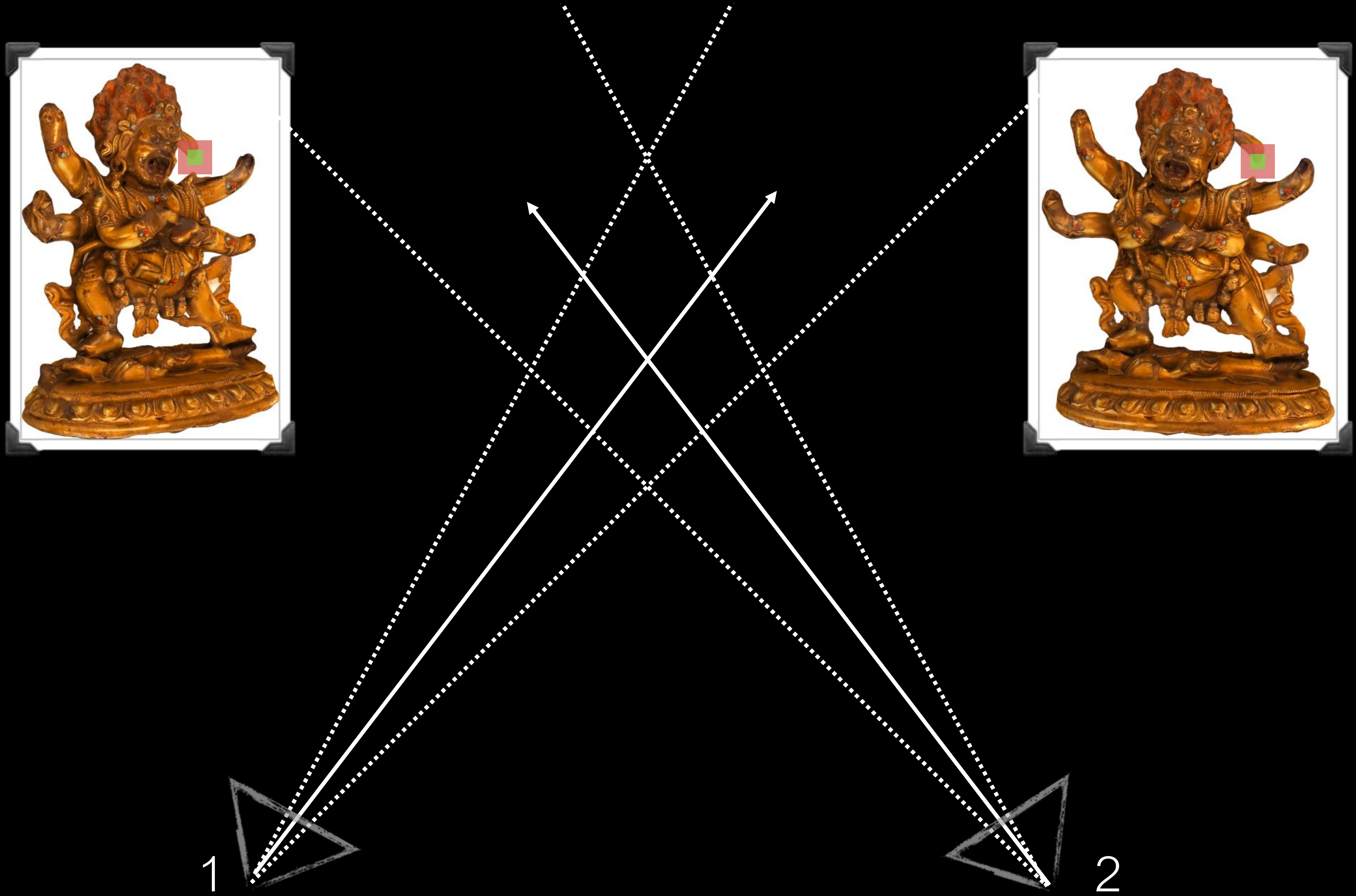
2. Cameras can’t see everything

3. Multiple cameras can reduce measurement error

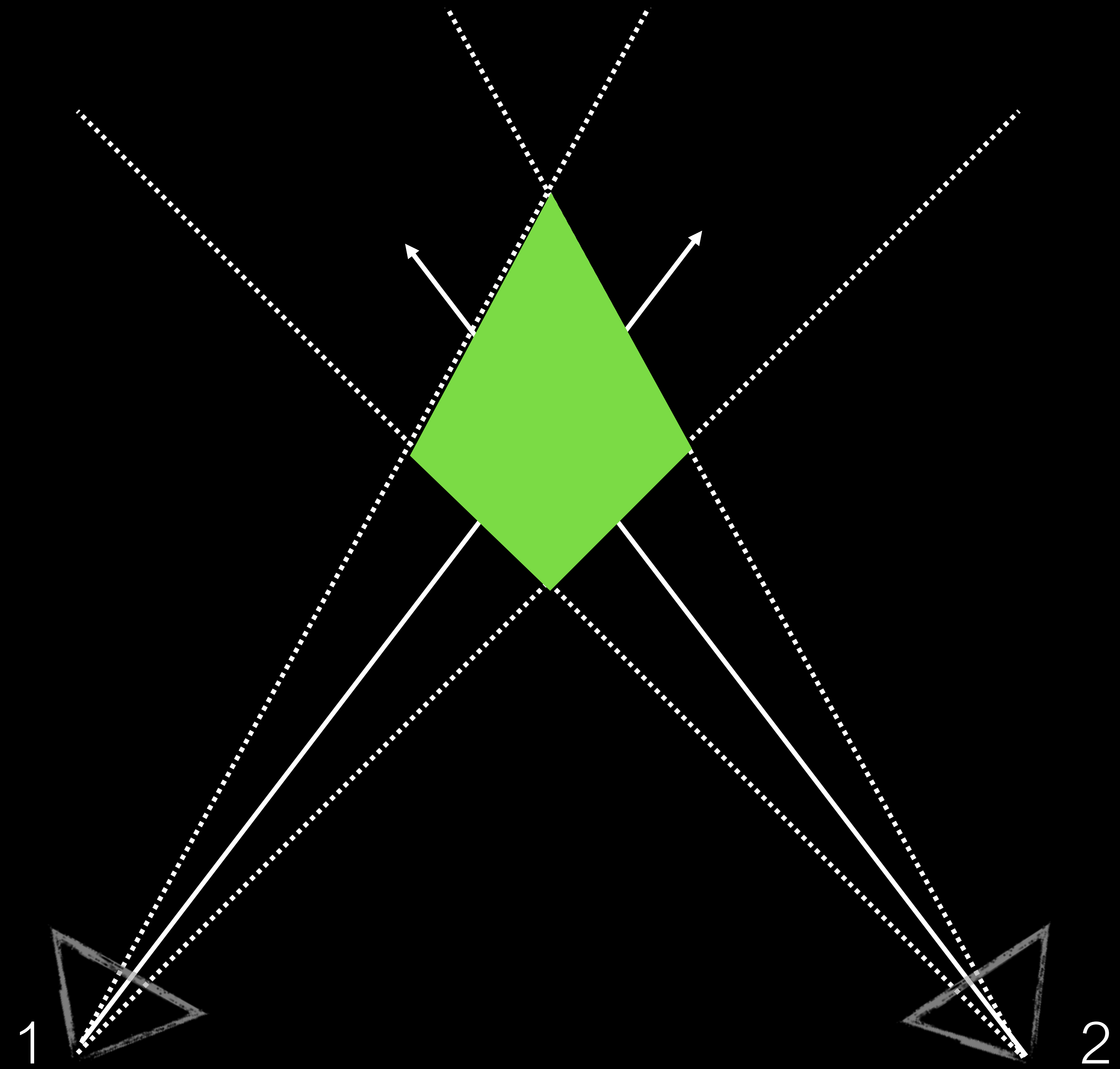
1

2

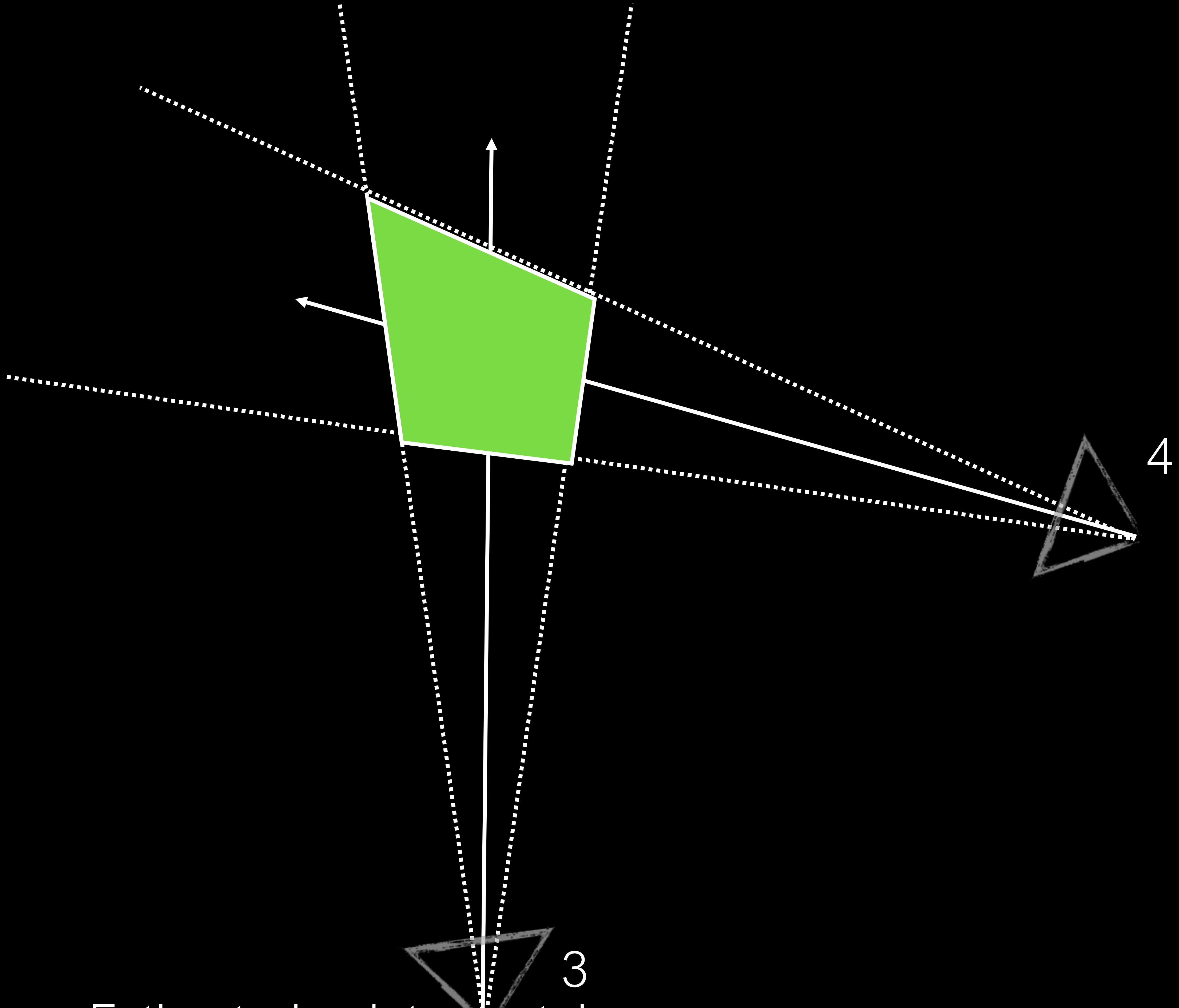

1
2



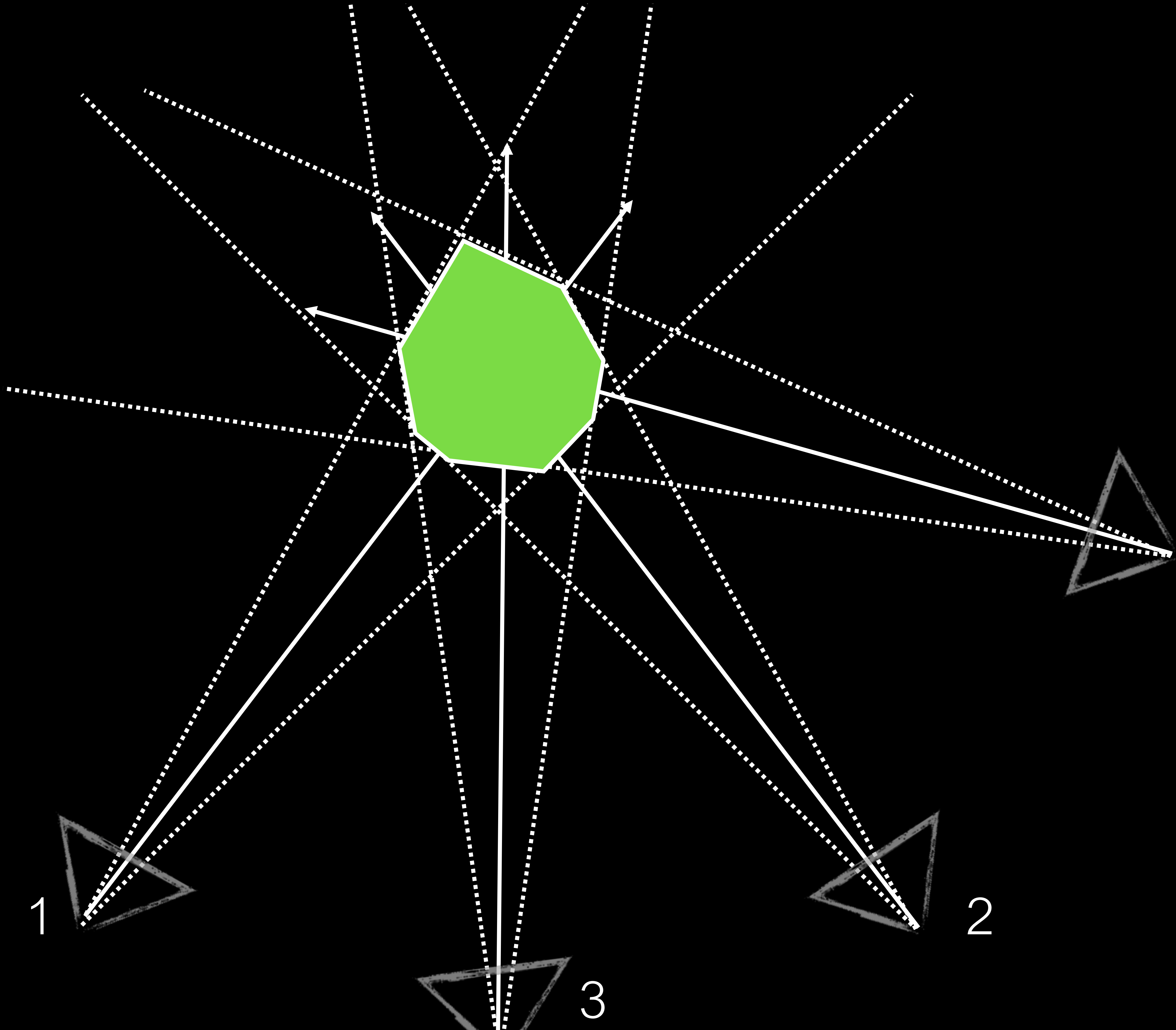
Estimated points contain some error.



Estimated points contain some error.



Estimated points contain some error.

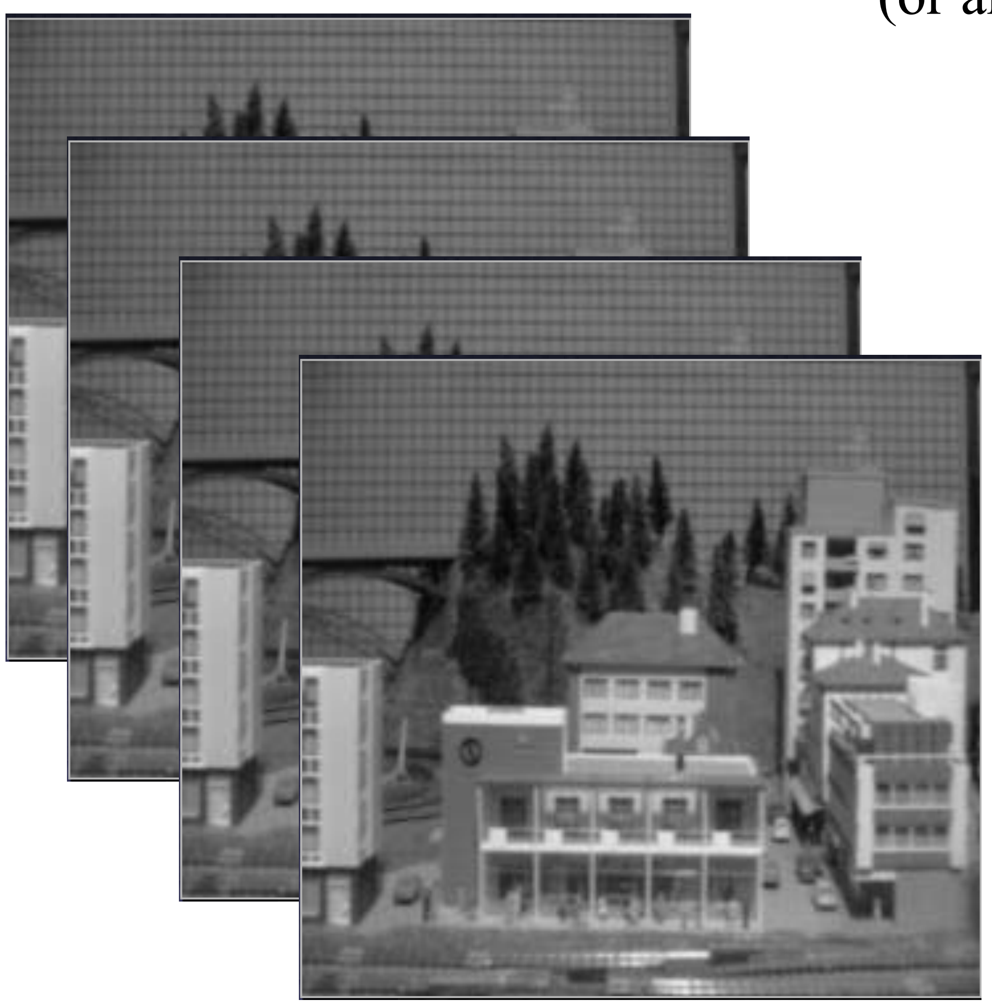


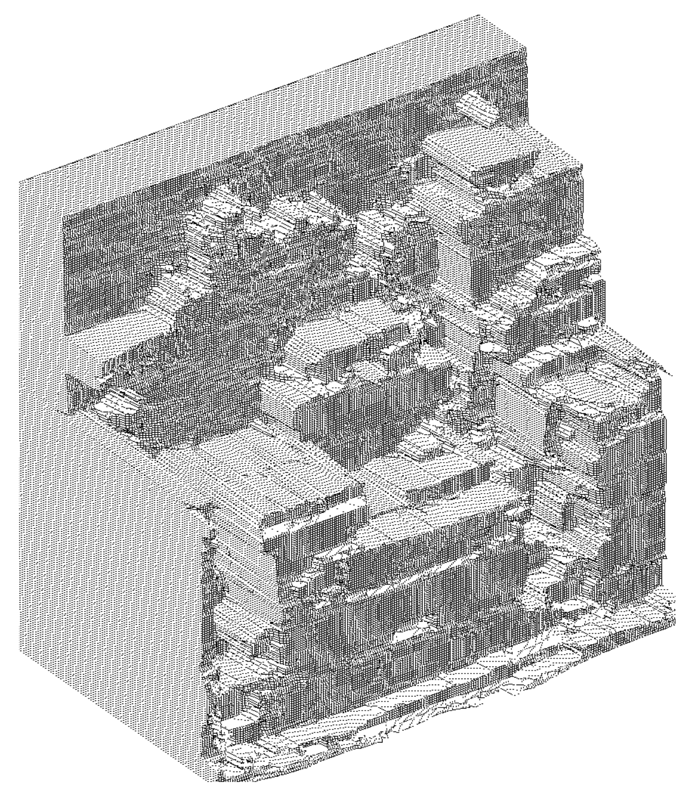
Additional views reduce error.

Multiview Stereo (version 0)

-Pick one reference view
-For each point and for each candidate depth

- keep depths with low SSD error in all other views
  (or any *photoconsistency* measure)



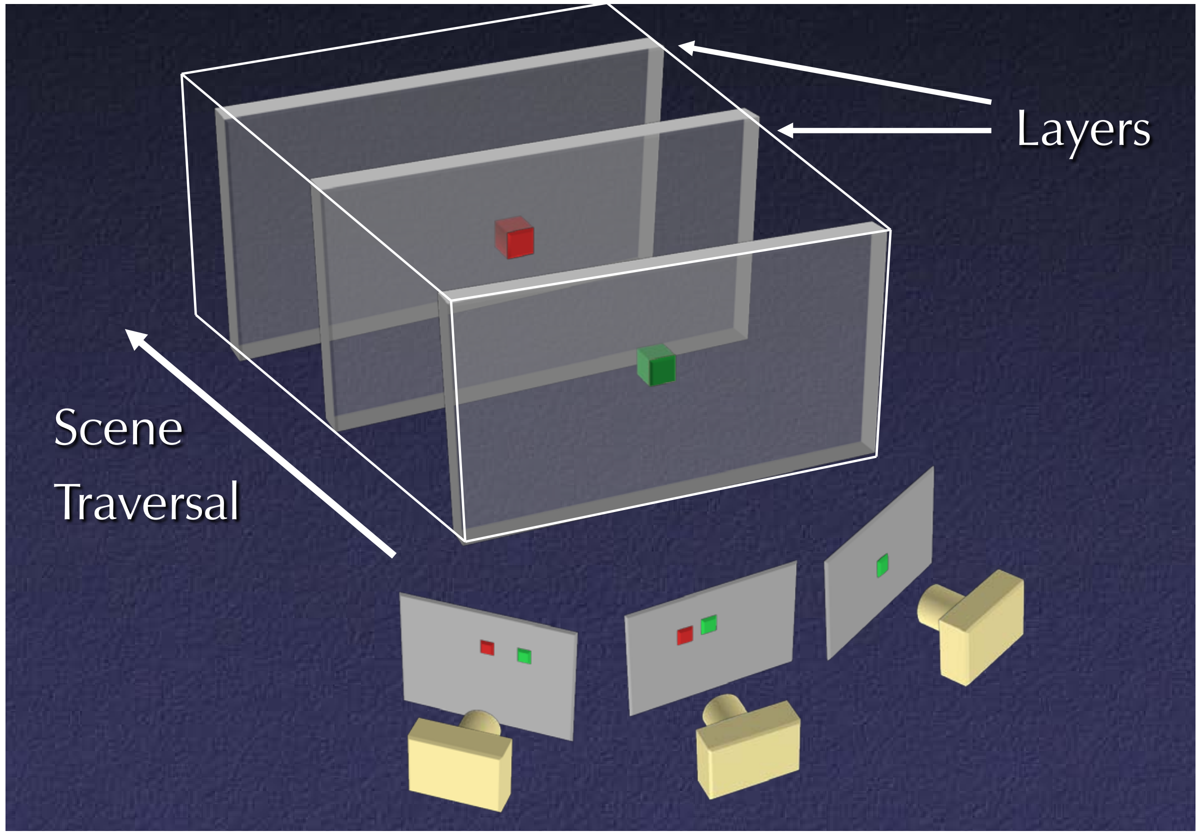


Problem: not all points are visible in all other views (occlusion and visibility major nuisance!)

Plane-sweep Stereo

Sweep over voxel plane-by-plane, starting closest-to-front

Quickly validate voxels in a plane by computing their appearance in a virtual view using all N cameras



Store photoconsistent color in a 3D voxel grid (don't need a reference image)

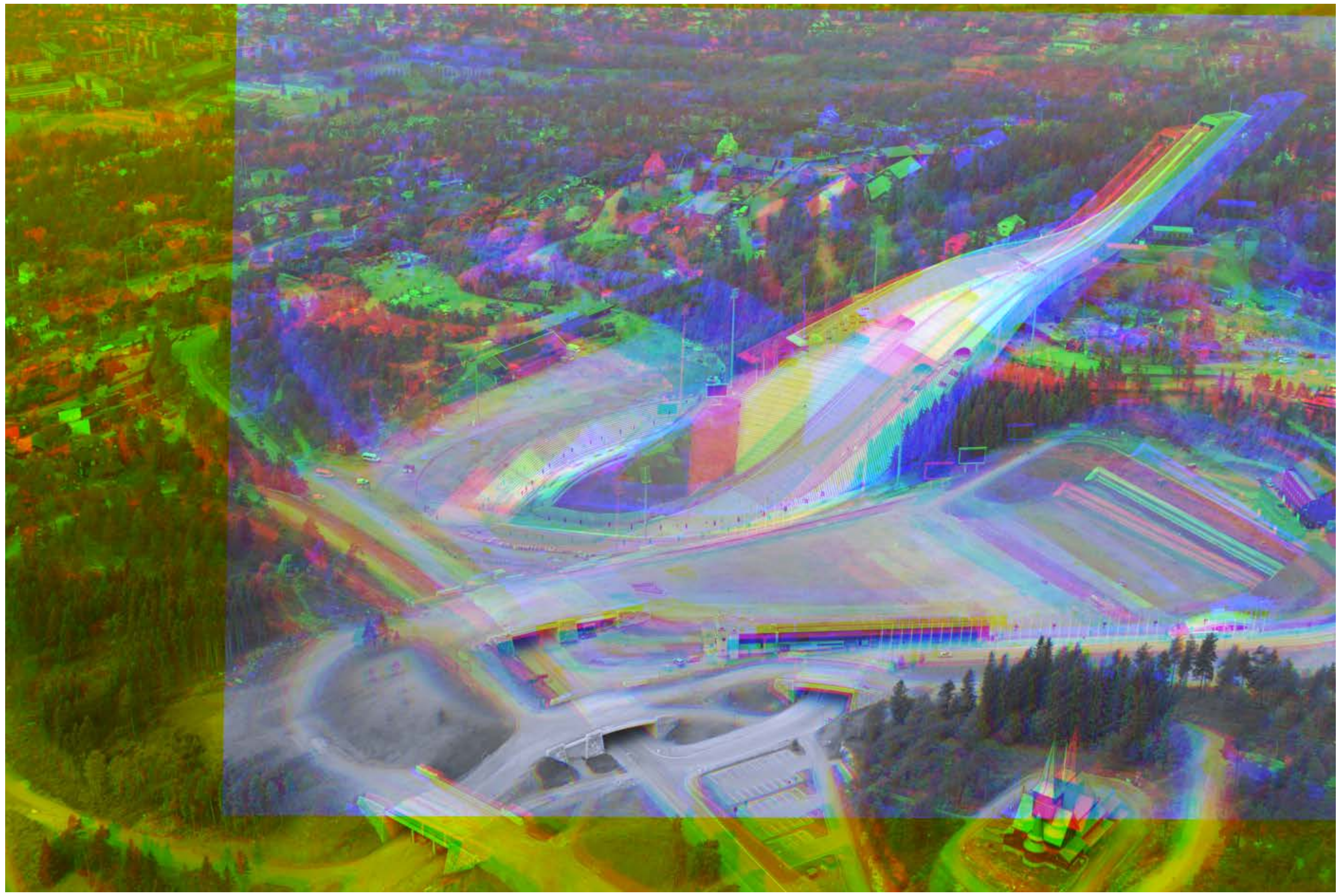
Reconstuct shape *and* appearance

Plane sweep for multiple views

Red:

Green:

Blue:



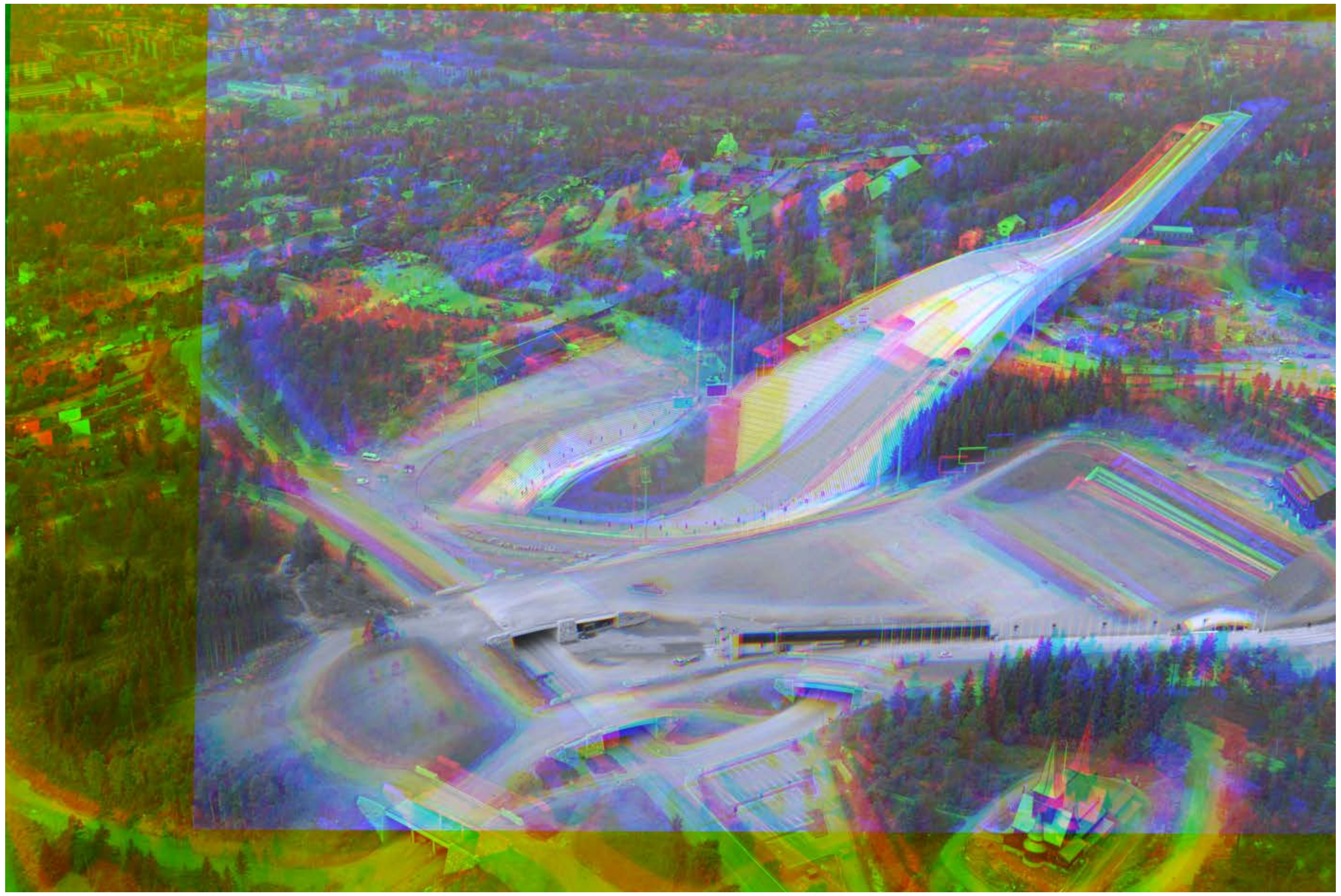
d_m = 520 meter

Plane sweep for multiple views

Red:

Green:

Blue:



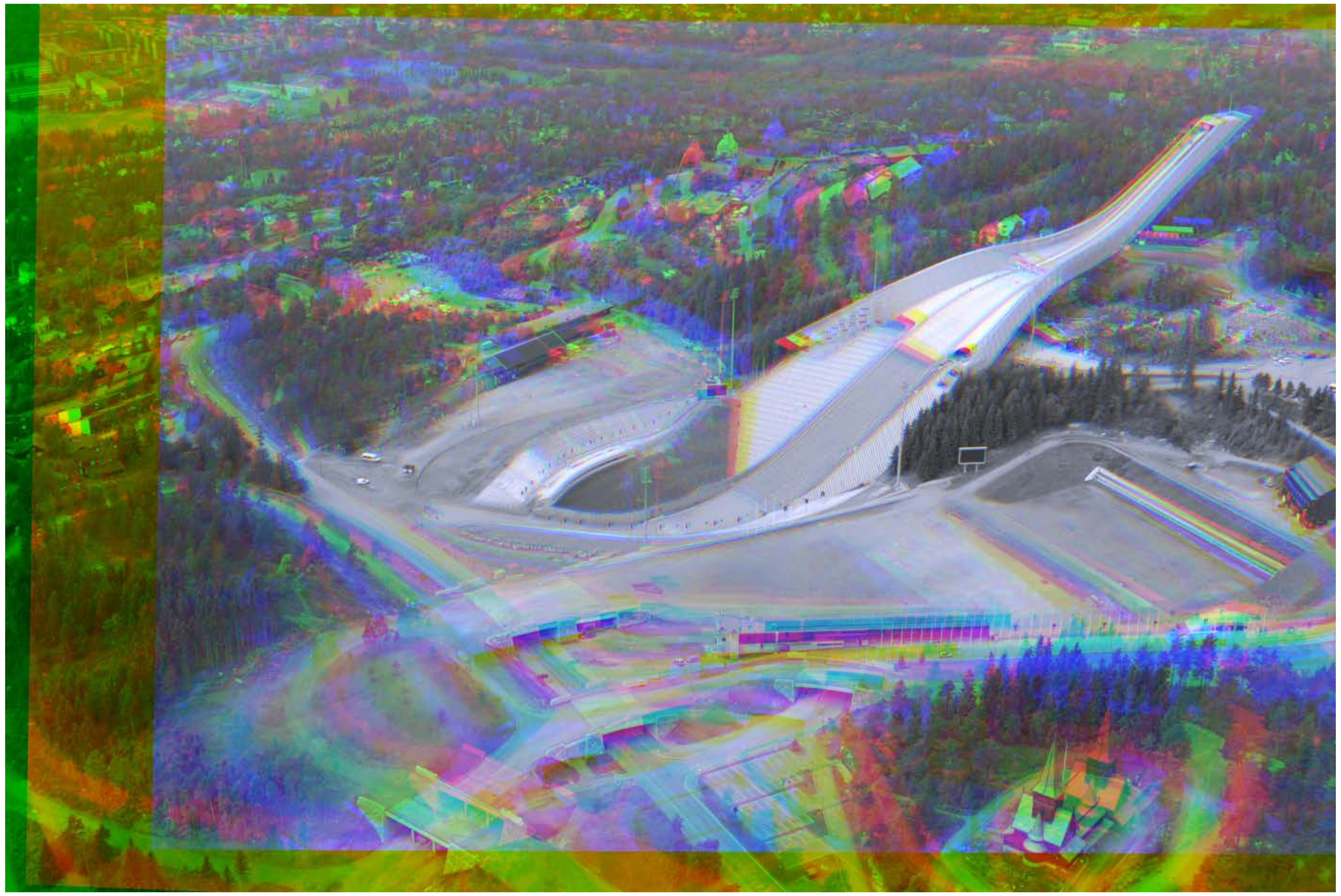
d_m = 583 meter

Plane sweep for multiple views

Red:

Green:

Blue:



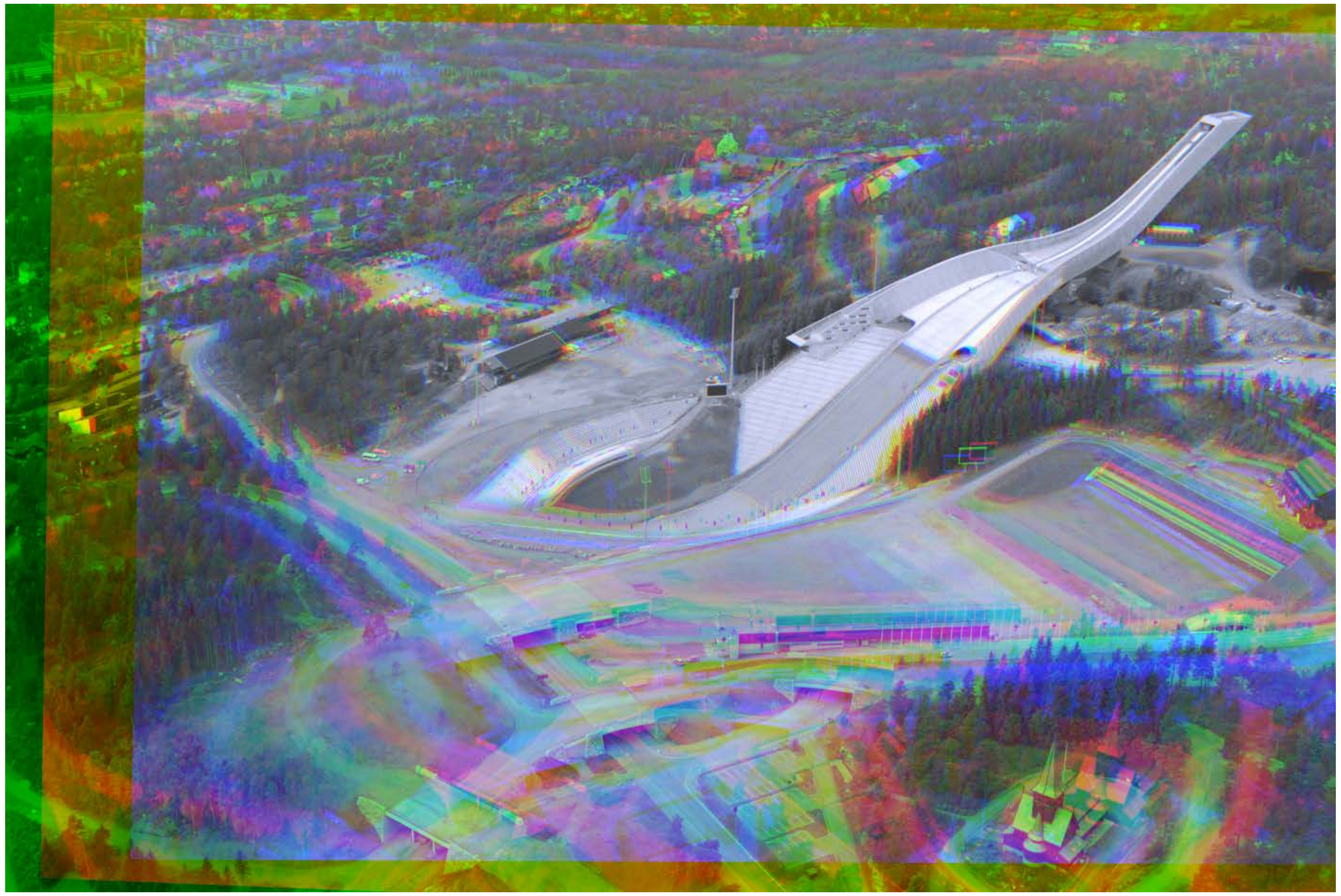
d_m = 706 meter

Plane sweep for multiple views

Red:

Green:

Blue:



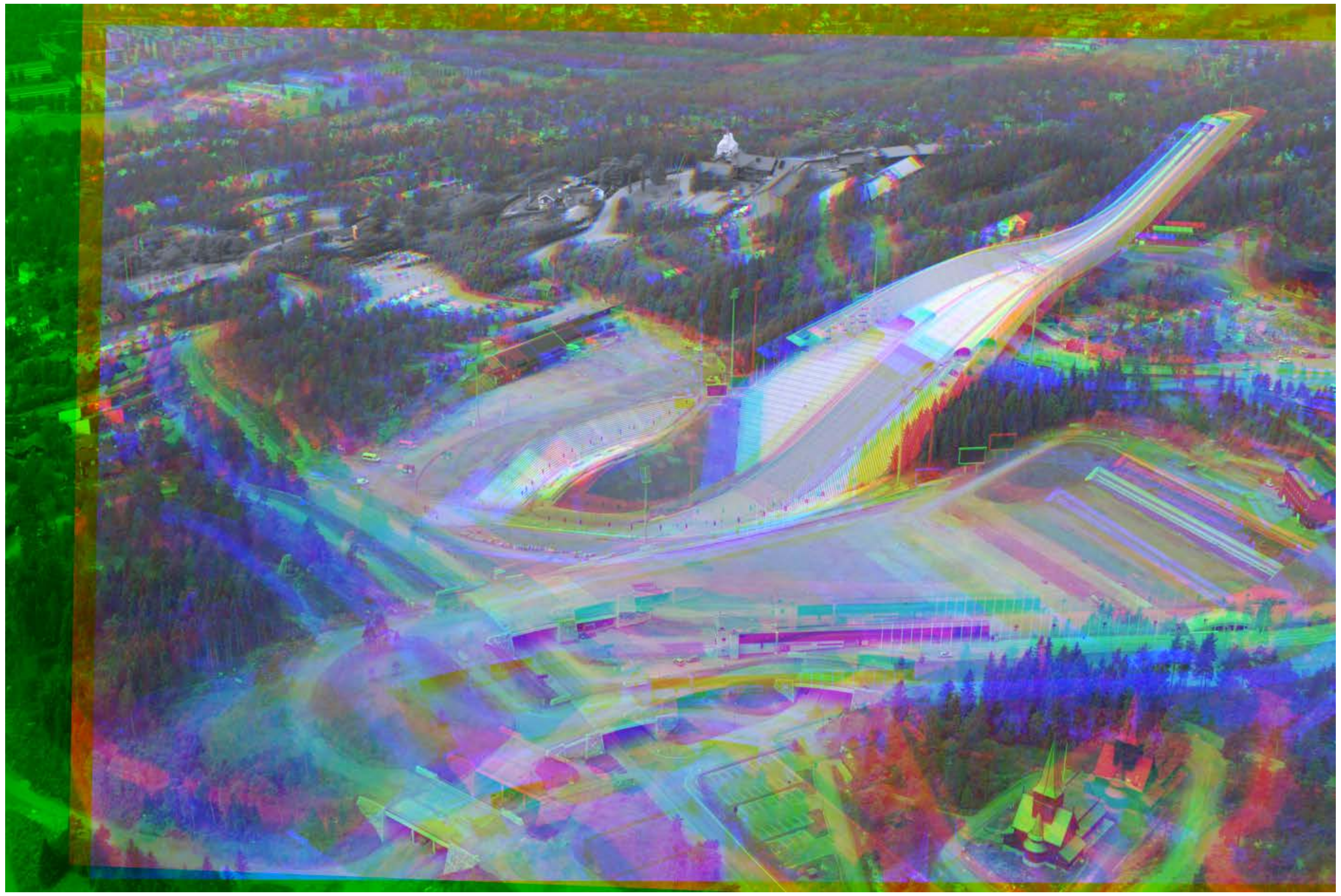
d_m = 790 meter

Plane sweep for multiple views

Red:

Green:

Blue:



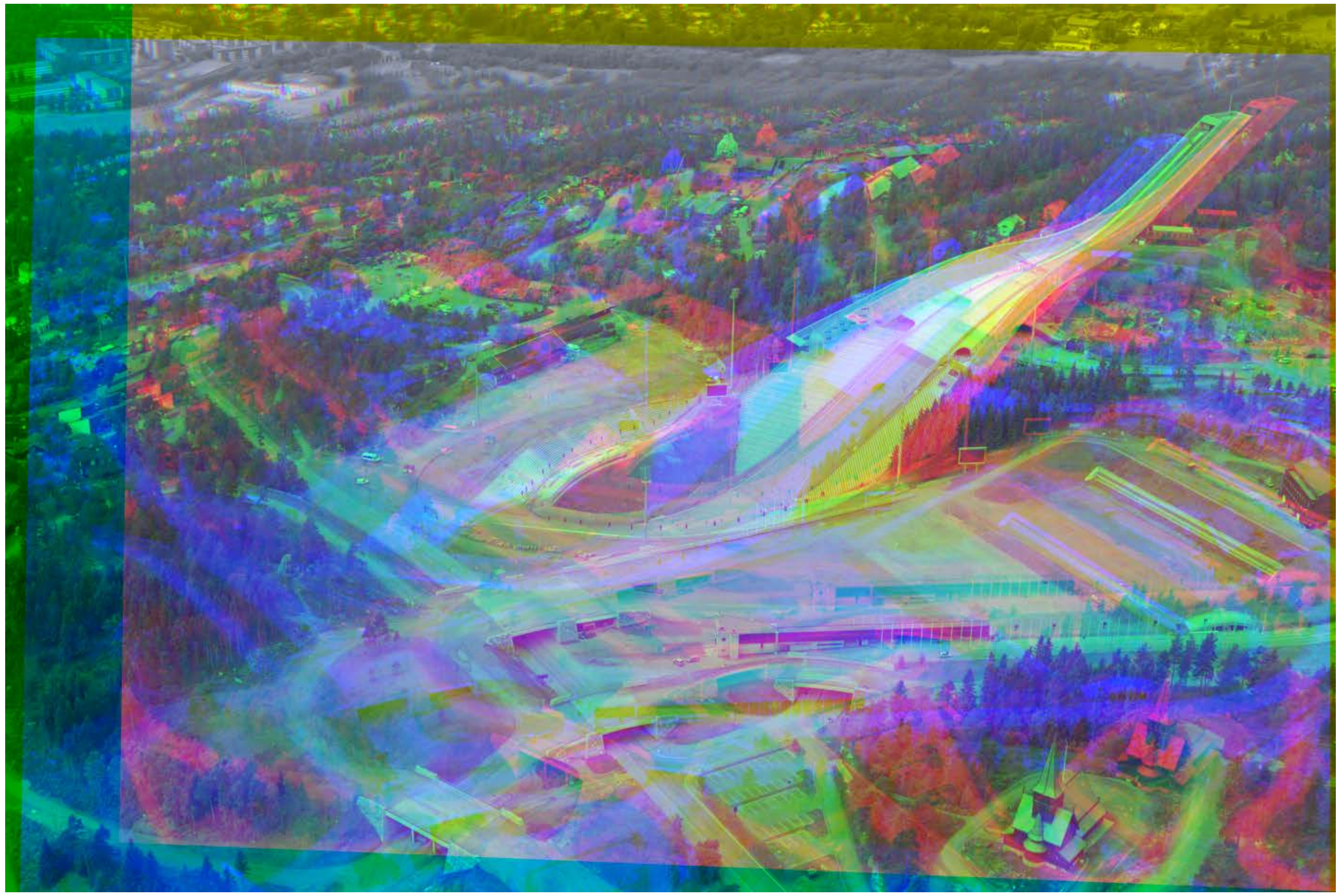
d_m = 1026 meter

Plane sweep for multiple views

Red:

Green:

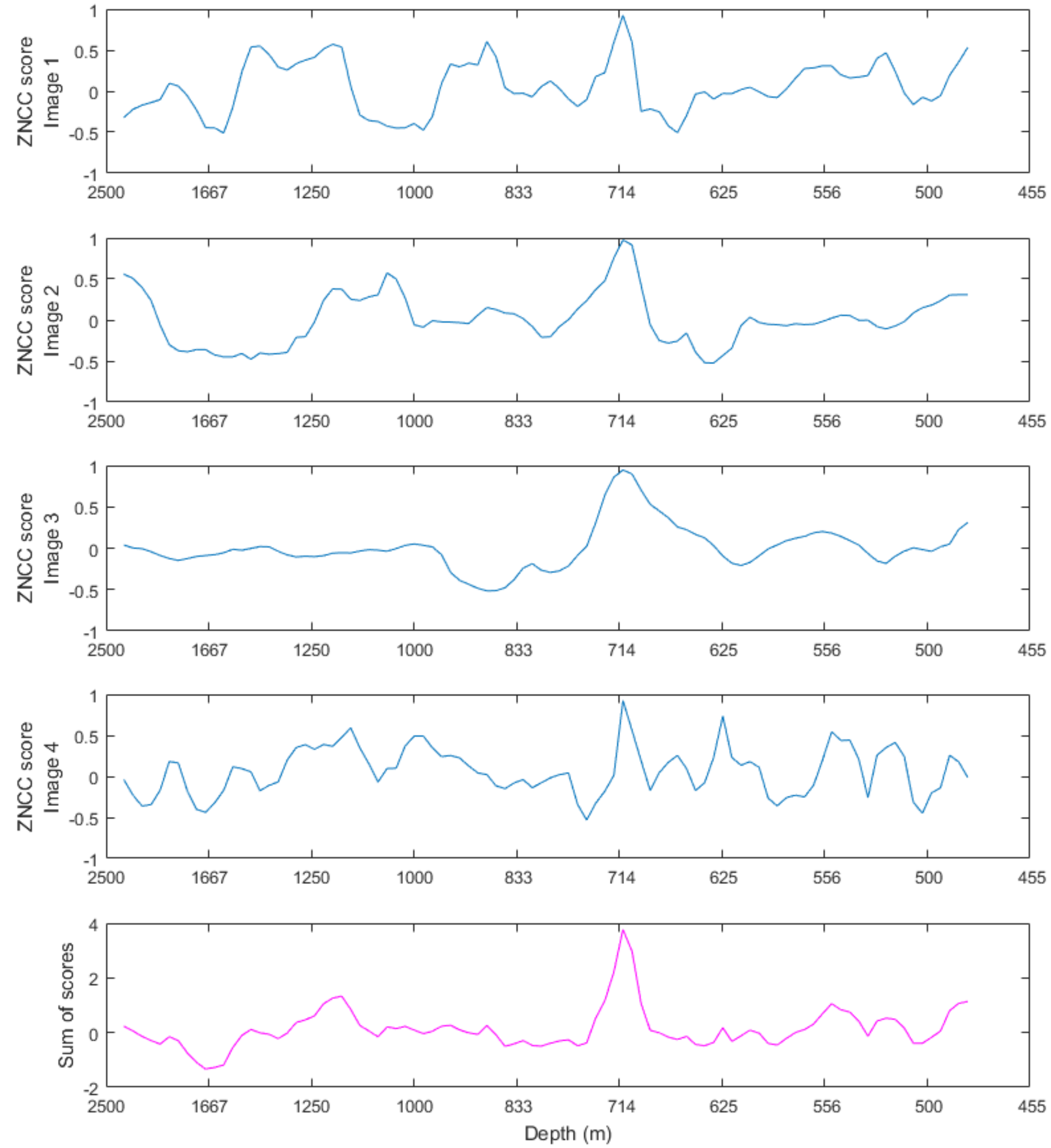
Blue:



d_m = 2168 meter

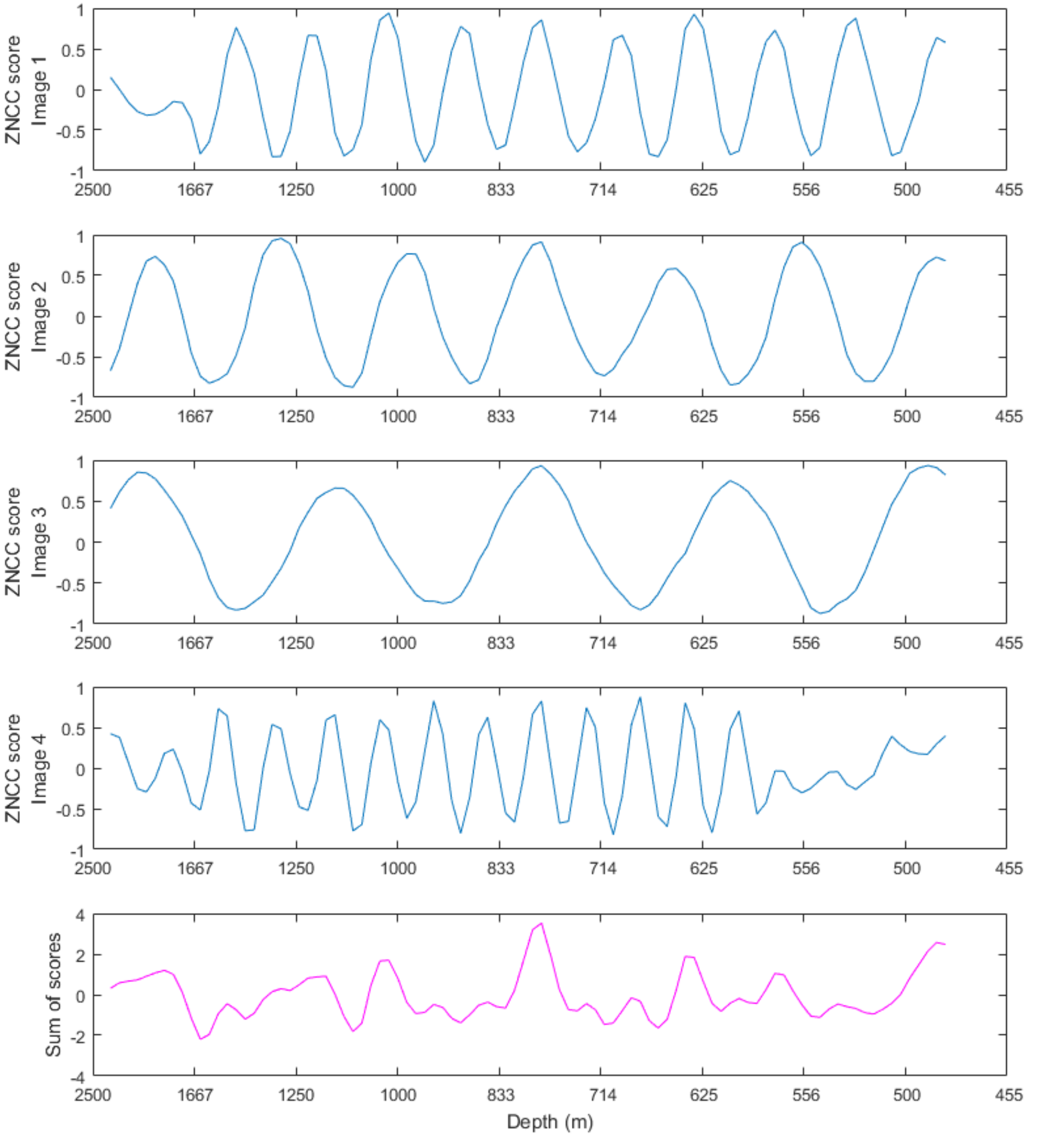
Plane sweep stereo example

- ZNCC scores for different depths and *k*

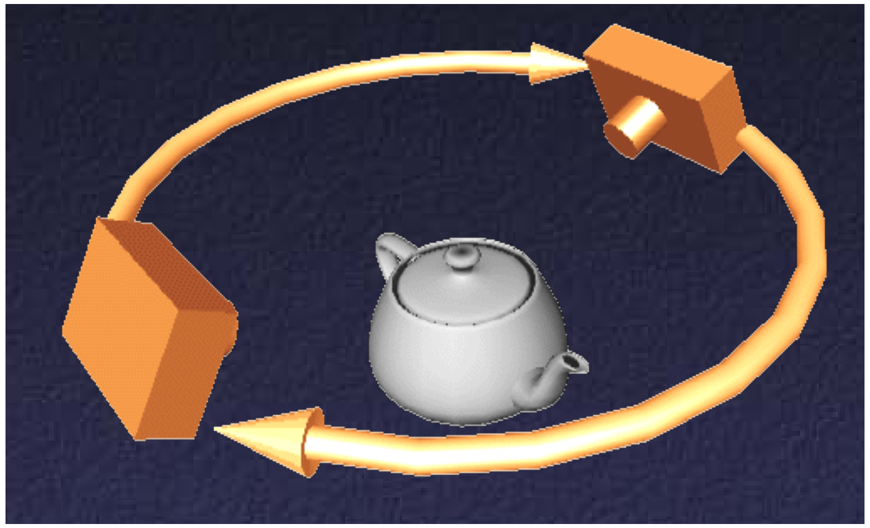


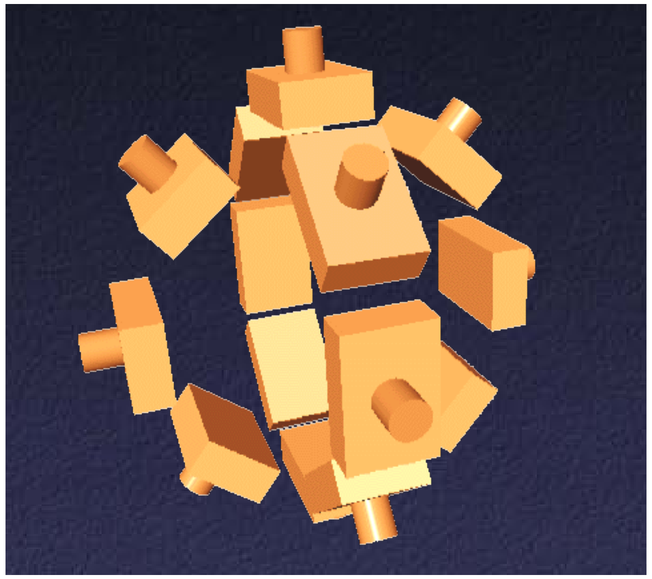
Plane sweep and ambiguities

- Multiple views can resolve ambiguities in difficult areas!



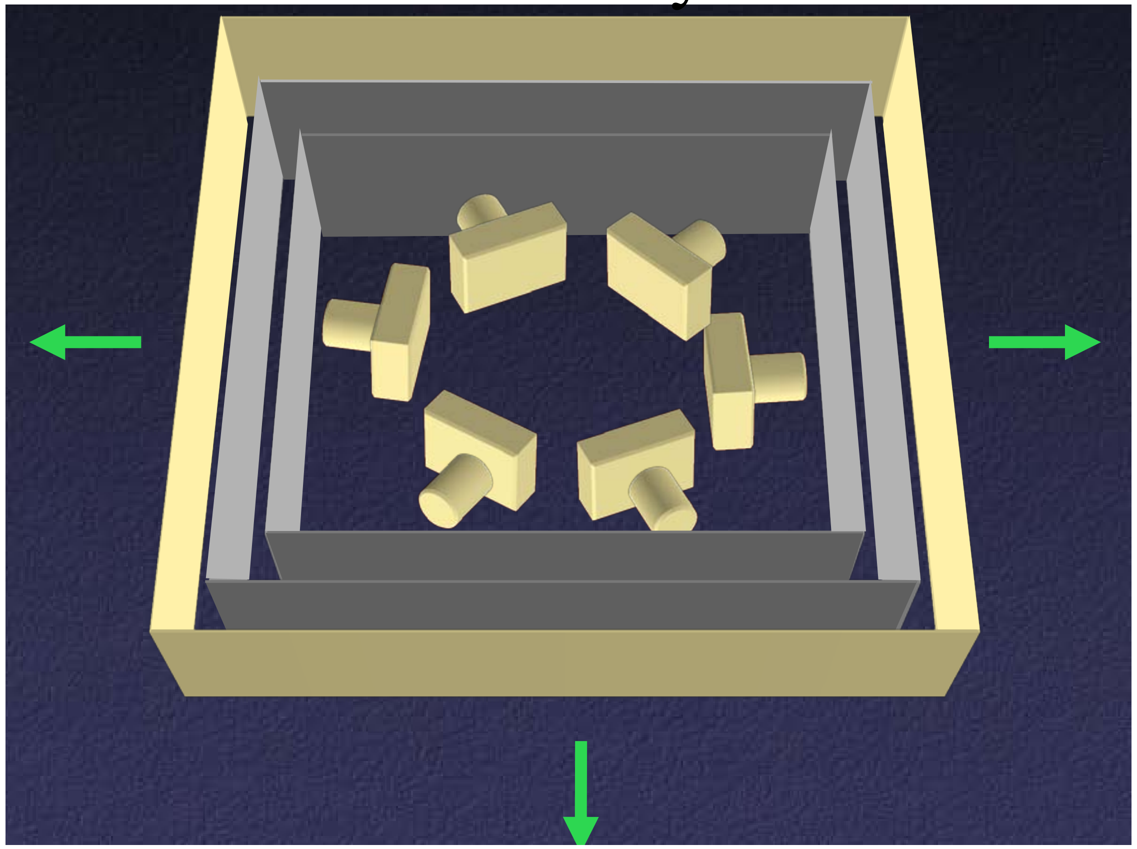
What about other camera steups?





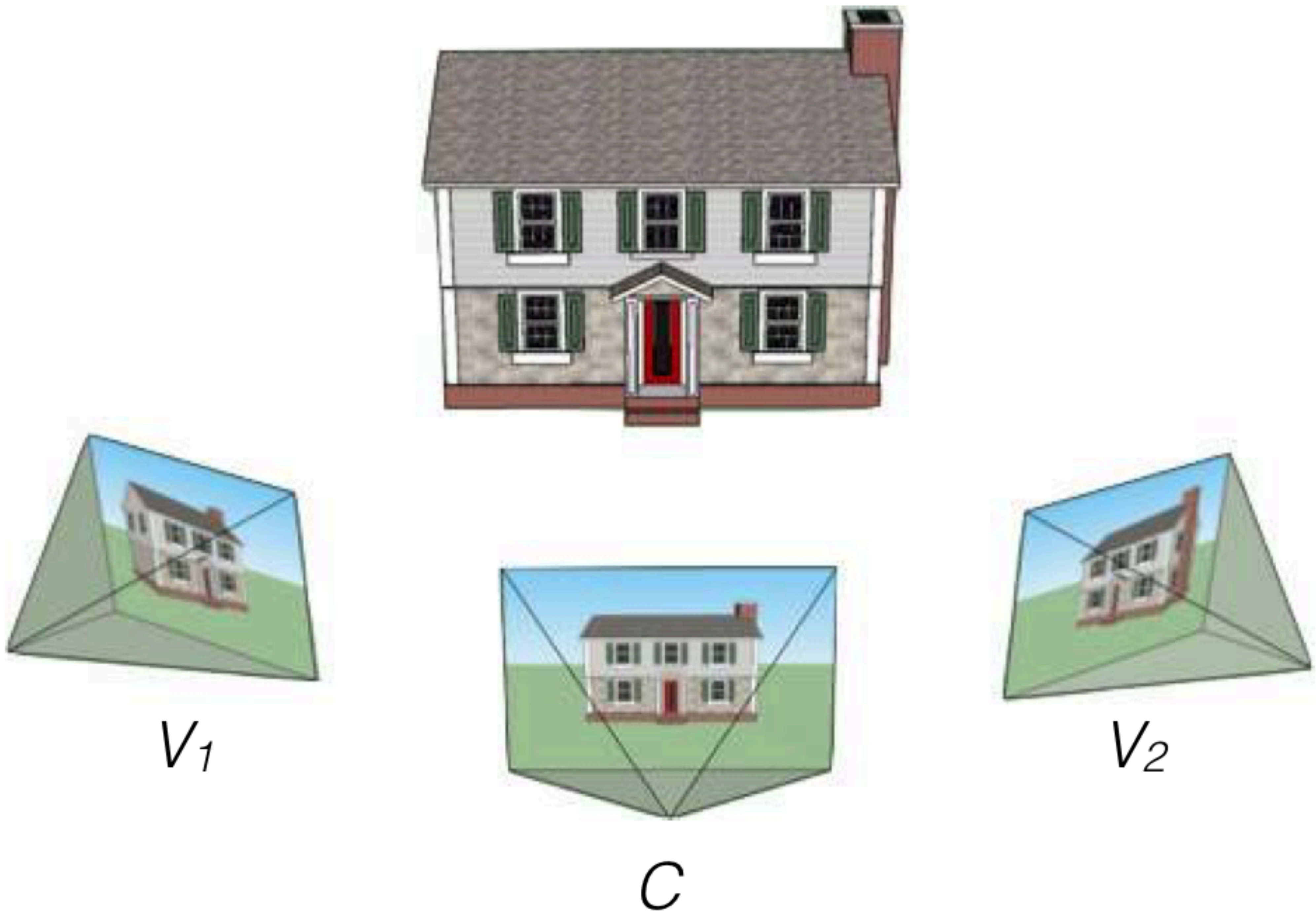
Panoramic depth ordering

Seitz & Dyer



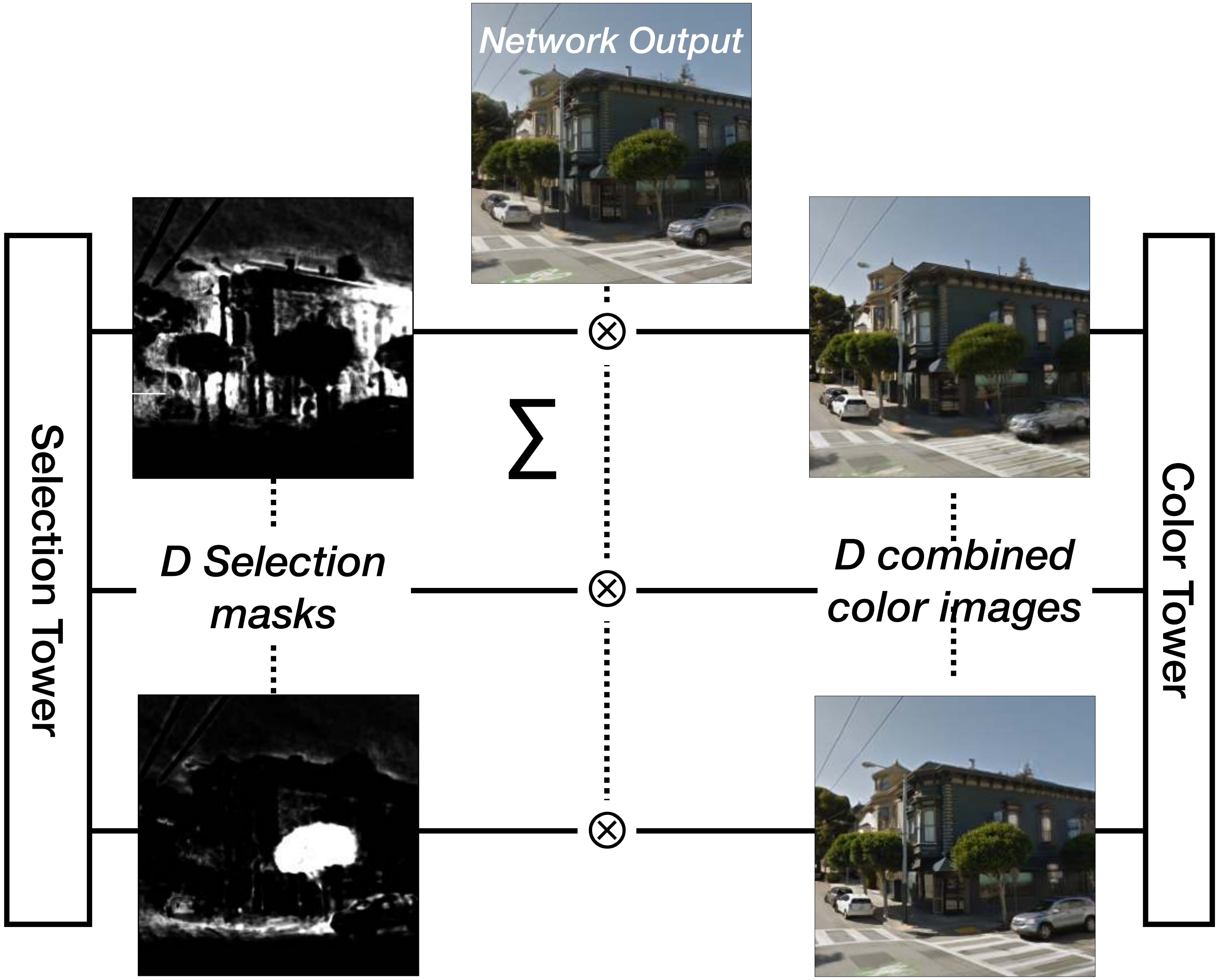
Layers radiate inwardly/outwardly

)



Synthesizing C from V1 and V2

John Flynn et al. DeepStereo



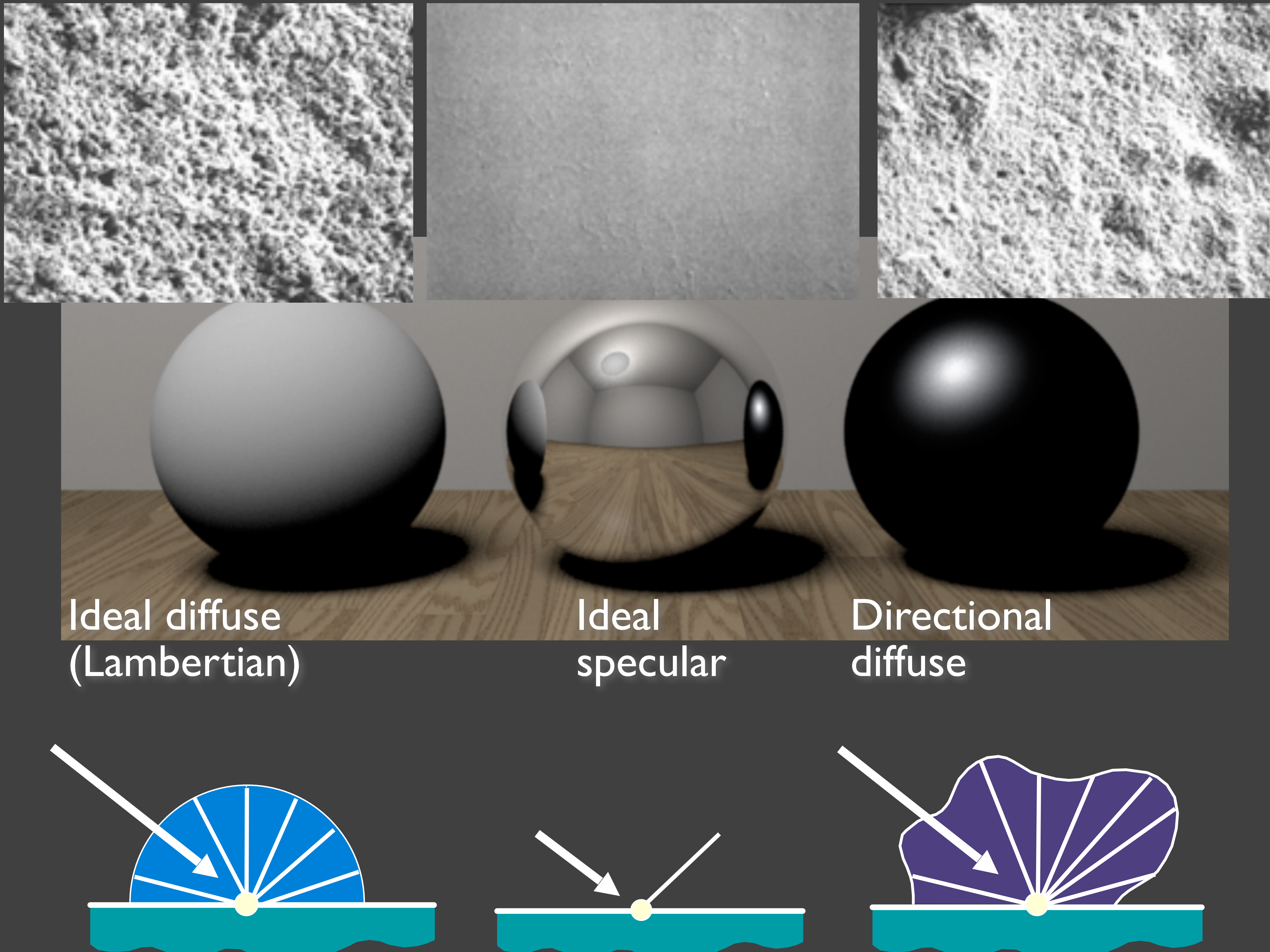
Train deep network select pixel from 1 of K depth planes

(At each pixel, output 1 of K classes)

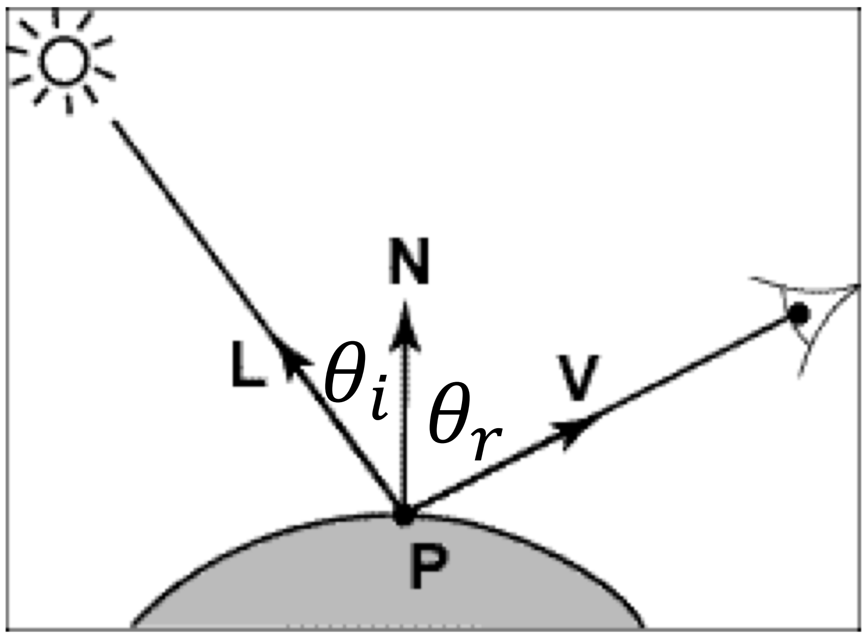
Single image
Multiple images
Depth image
...
3D voxels

Neural Net

Depth image
Material properties
View synthesis
...
3D Models


Ideal diffuse
(Lambertian)
Ideal
specular
Directional
diffuse

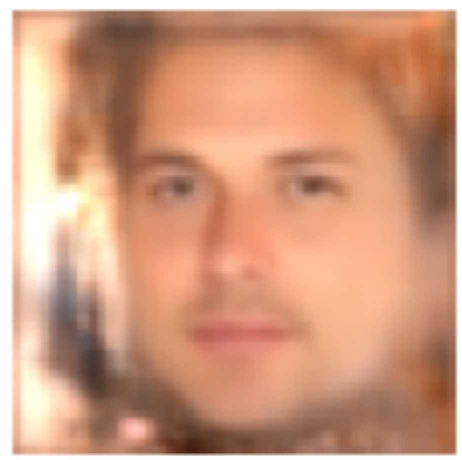
Light rays interacting with a surface



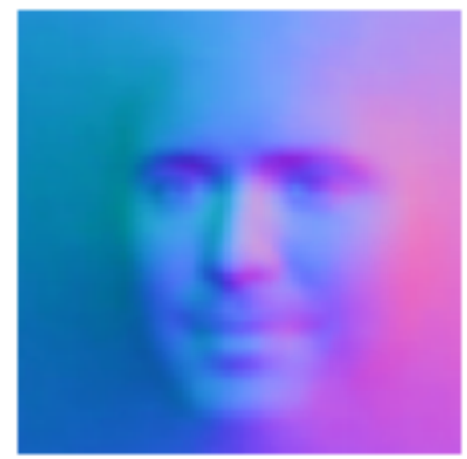
$$L_r = \rho(\theta_i, \theta_r) L_i \cos\theta_i$$

- Special case 1: Perfect mirror
  - $\rho(\theta_i, \theta_r)$ = 0 unless $\theta_i = \theta_r$
- Special case 2: Matte surface
  - $\rho(\theta_i, \theta_r)$ = ρ_0 (constant)

Image



Reflectance



Shape

Shading

[Neural Face Editing with Intrinsic Image Disentangling]

Slide Credit: Wei-Chiu Ma

Applications

- Object insertion

[Rendering Synthetic Objects into Legacy Photographs]

Slide Credit: Wei-Chiu Ma

Applications

- Material editing

[Learning Non-Lambertian Object Intrinsics…]

Slide Credit: Wei-Chiu Ma

Applications

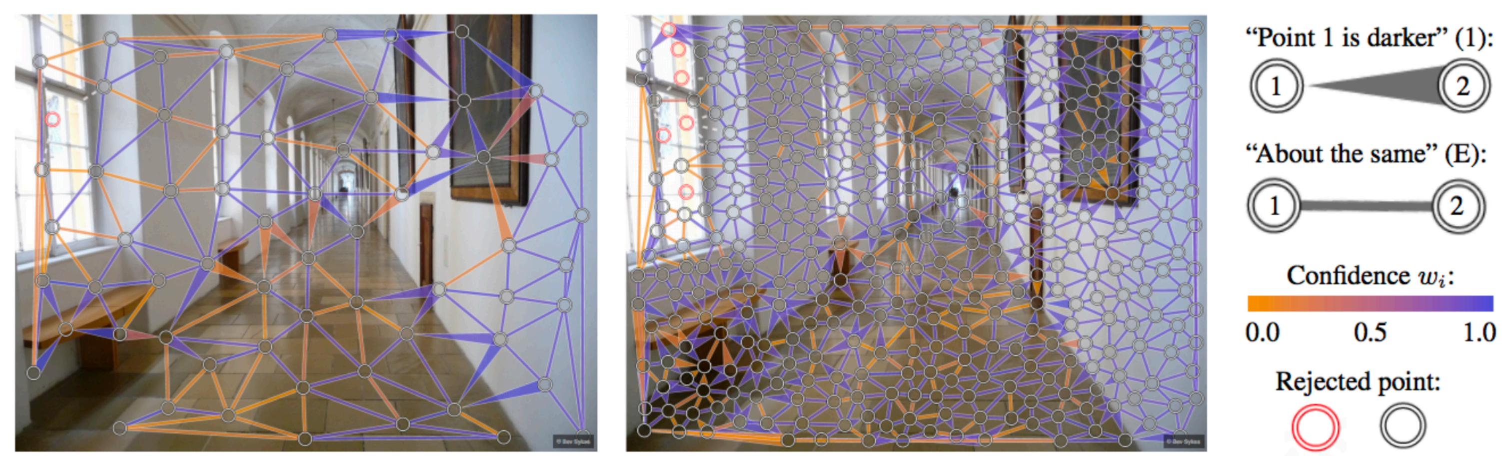
- Shadow/shading removal

[Removing shadows from images using retinex]

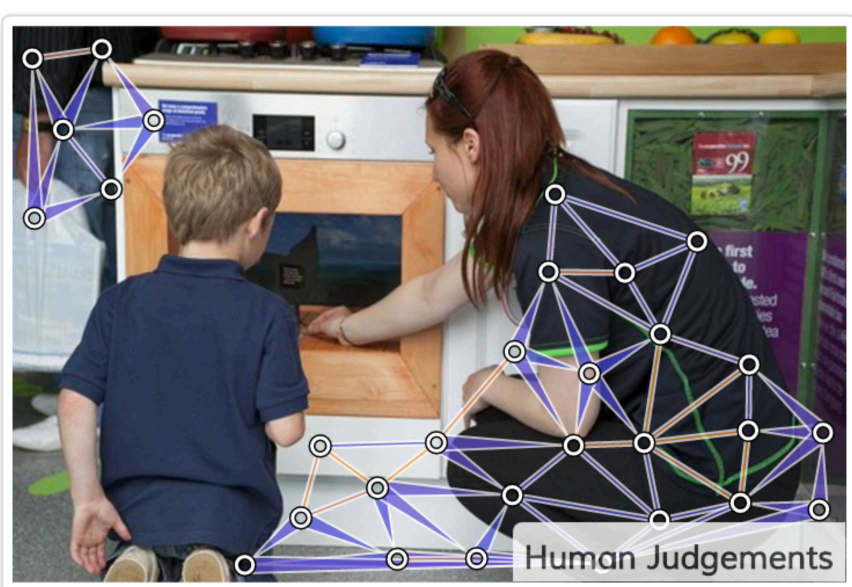
Slide Credit: Wei-Chiu Ma

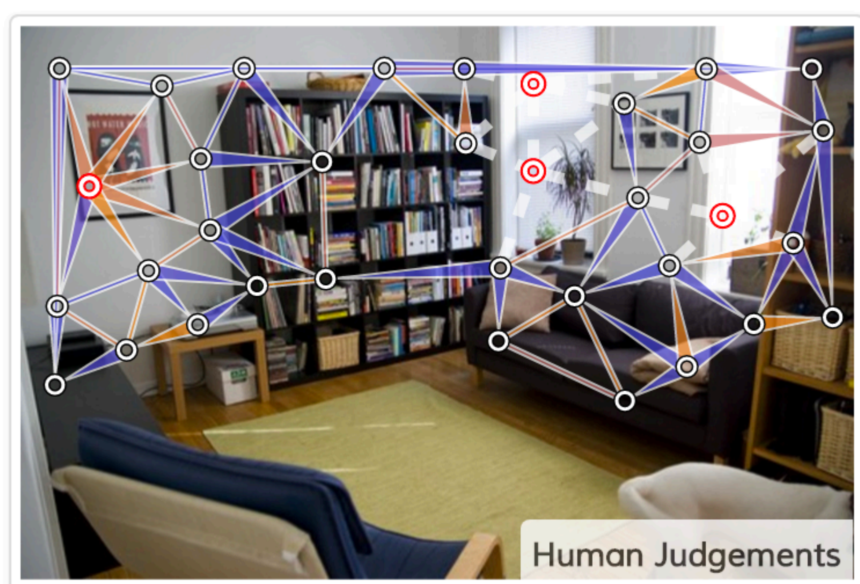
Intrinsic Images In the Wild

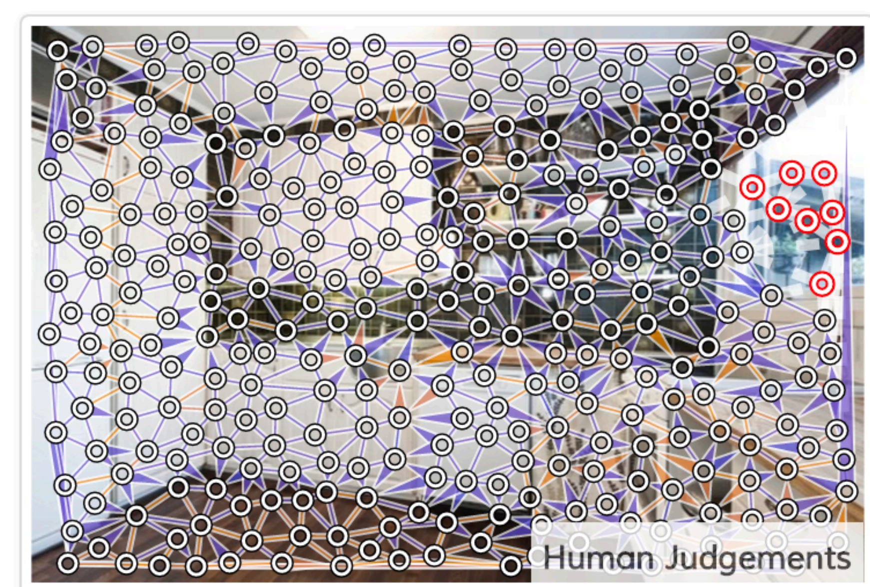
- Thousands of real world images
- Relative annotation (which is darker)
- Sparsely annotated



(a) *Sparse sampling.* **(b)** *Dense sampling.*

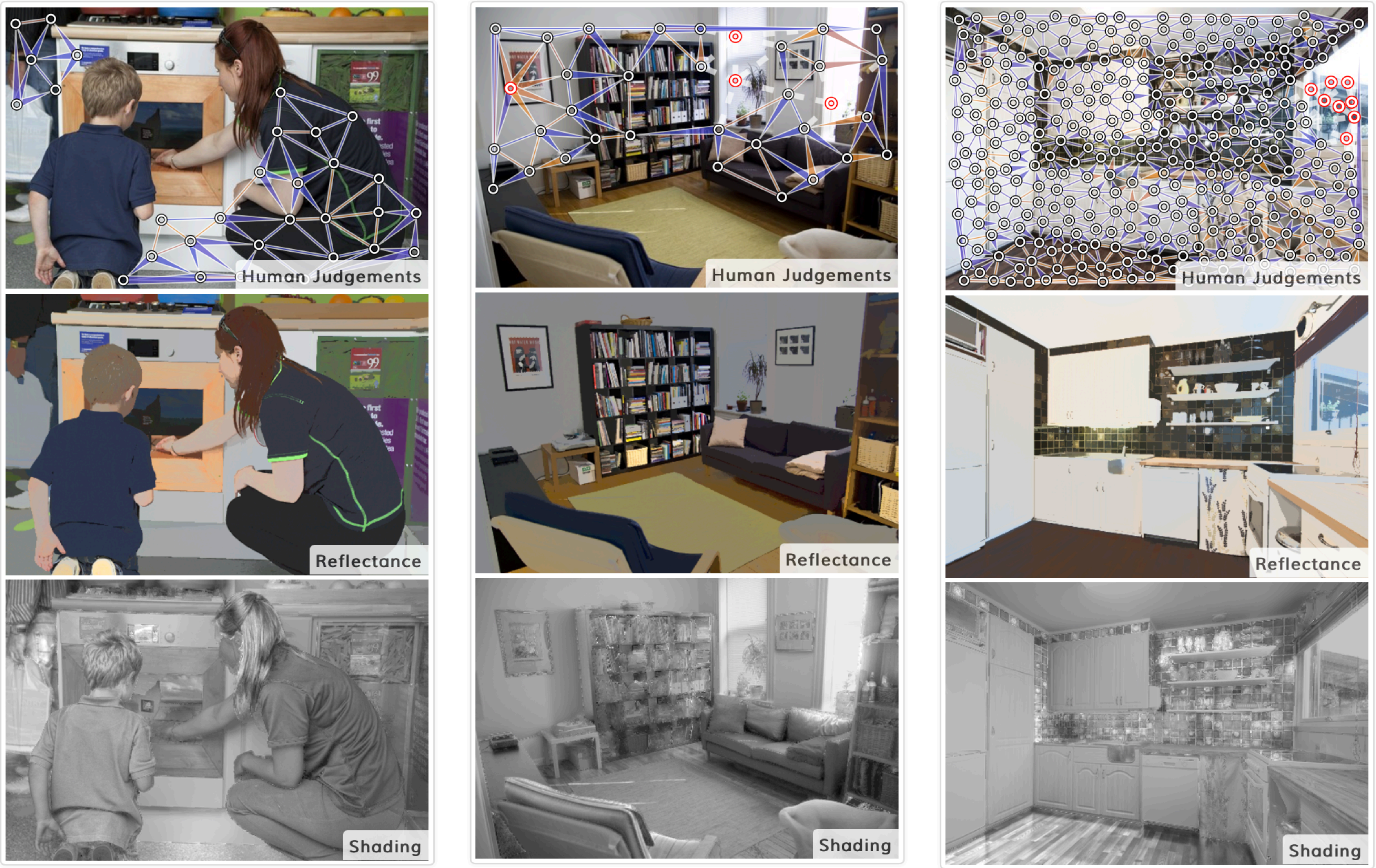






[Intrinsic Images in the Wild]

Intrinsic Images In the Wild



[Intrinsic Images in the Wild]

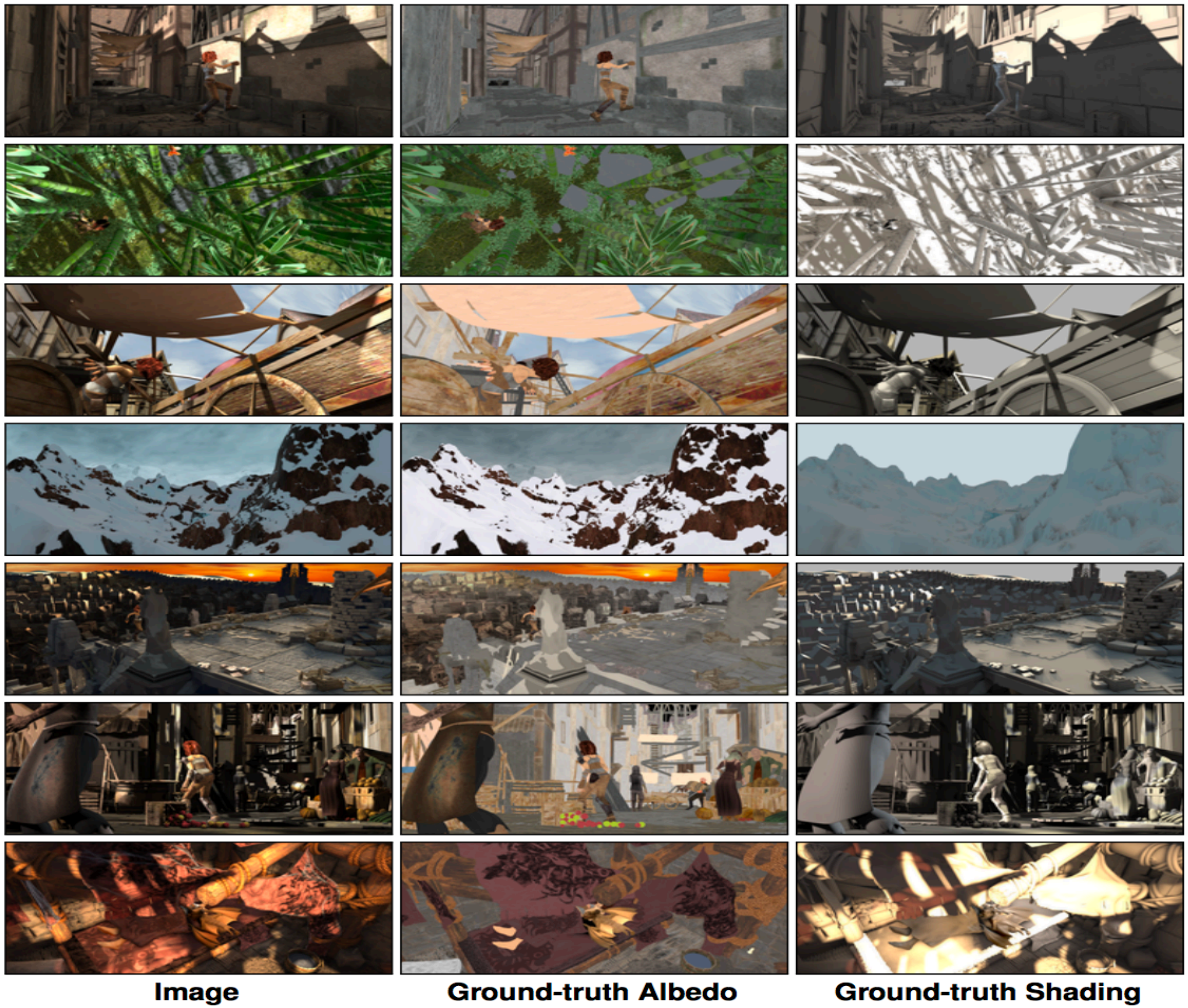
Intrinsic Images In the Wild

(a) Living room (b) Bedroom 1 (c) Kitchen 1 (d) Bedroom 2 (e) Kitchen 2

Automatic intrinsic image decomposition

[Intrinsic Images in the Wild]

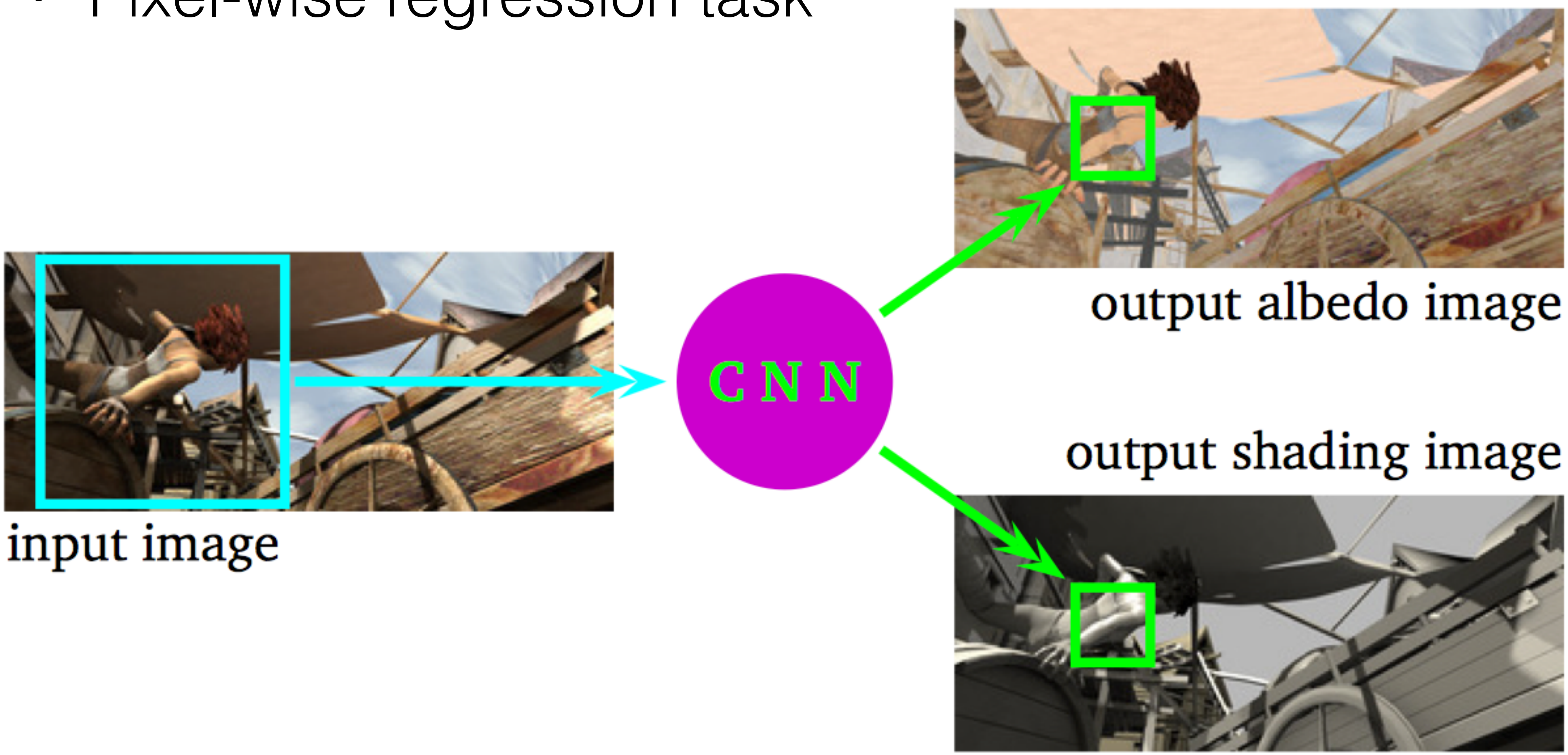
MPI Sintel Dataset



[A naturalistic open source movie for optical flow evaluation]

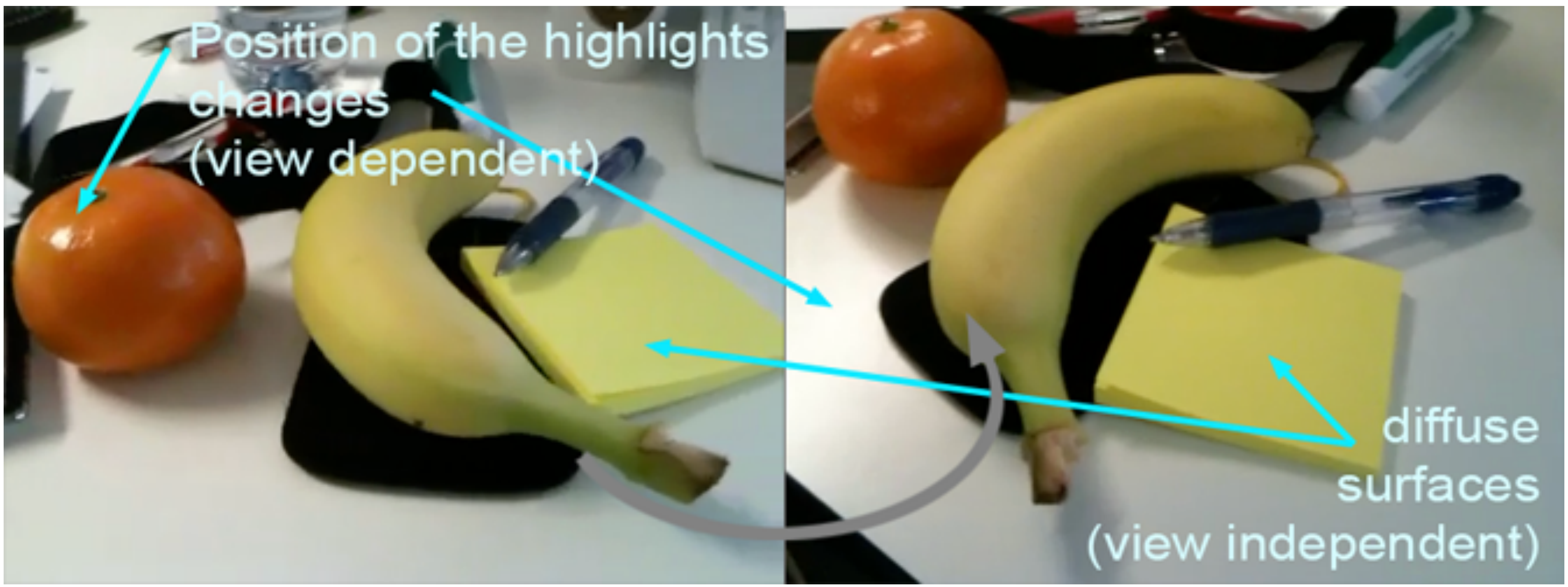
Direct Intrinsics

- Pixel-wise regression task



[Narihira et al. 2015]

What changes across views?



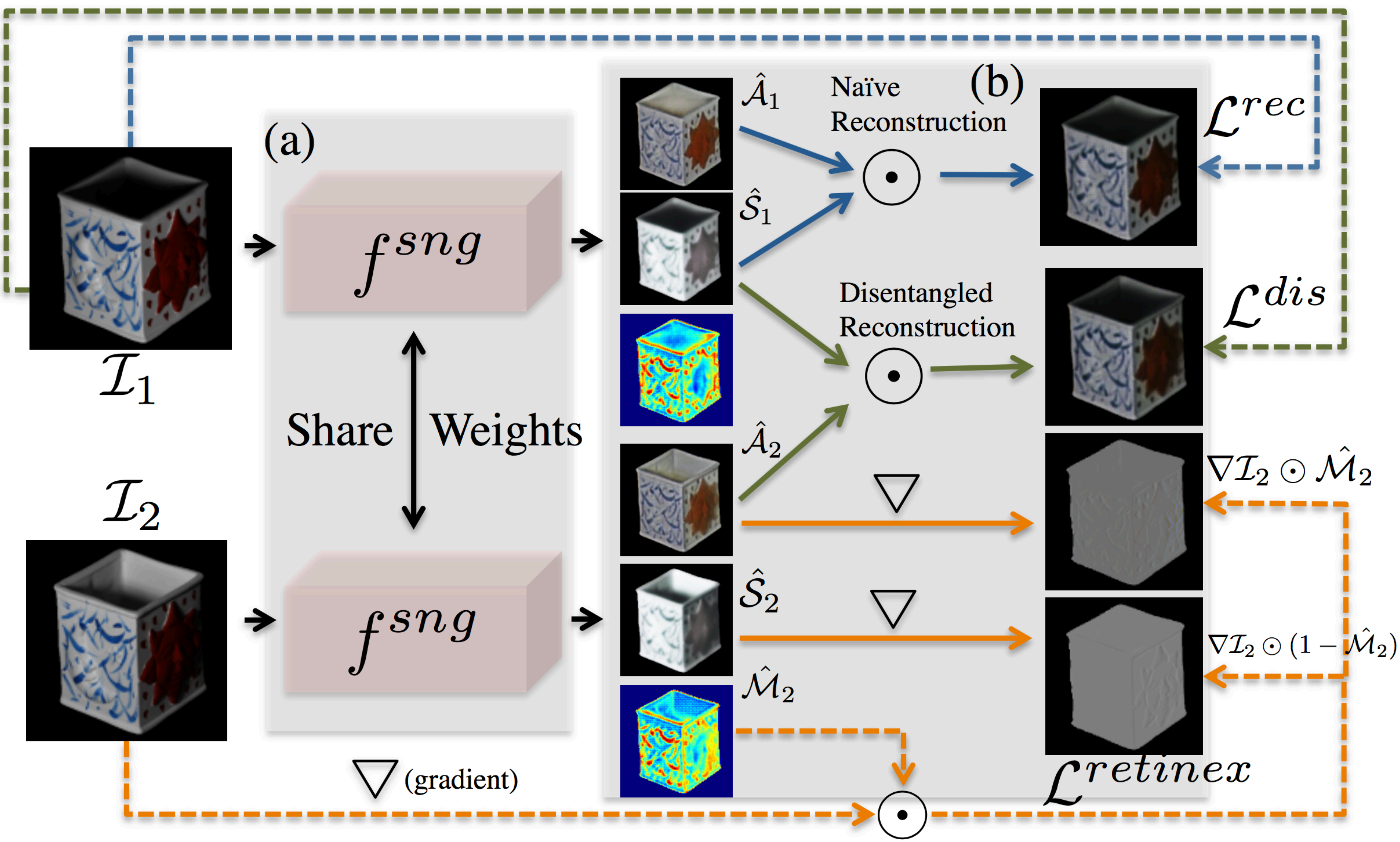
Slide Credit: Wei-Chiu Ma

What changes across time?

[Deriving intrinsic images from image sequences]

Slide Credit: Wei-Chiu Ma

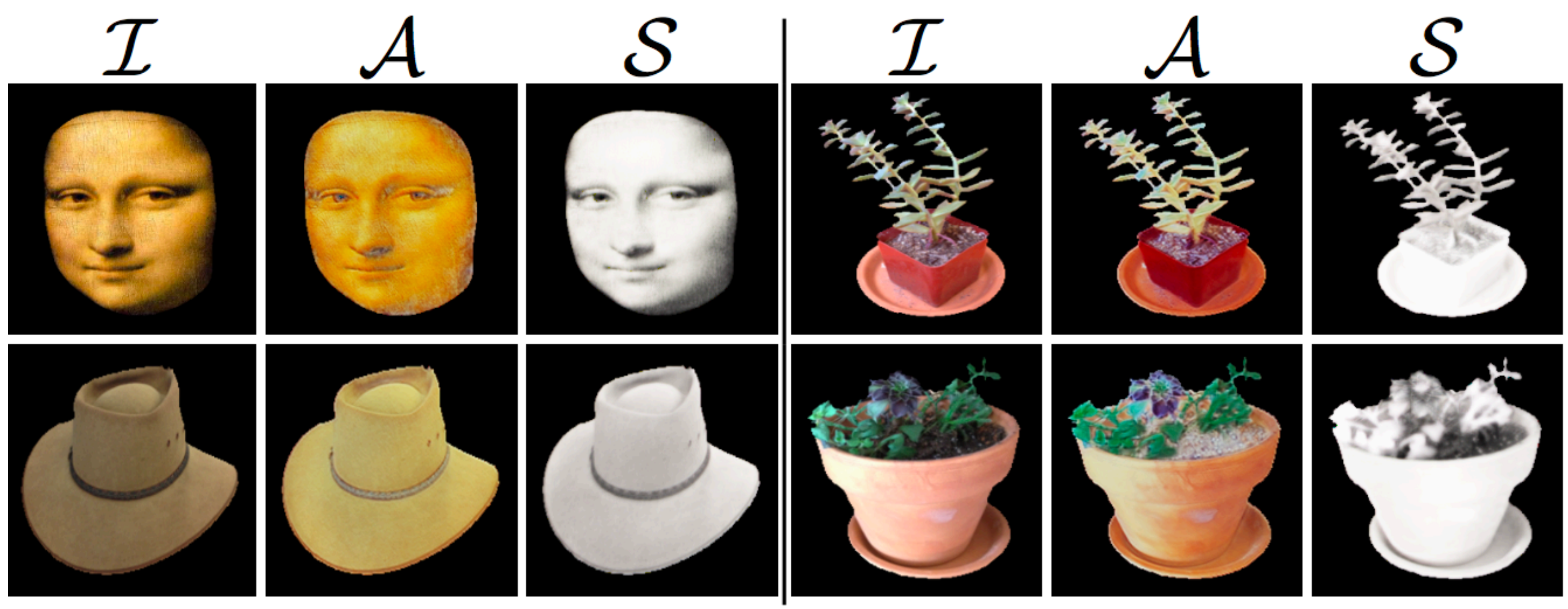
Intrinsic Image Decomposition



[Single Image Intrinsic Decomposition without…]

Slide Credit: Wei-Chiu Ma

Intrinsic Image Decomposition



Slide Credit: Wei-Chiu Ma

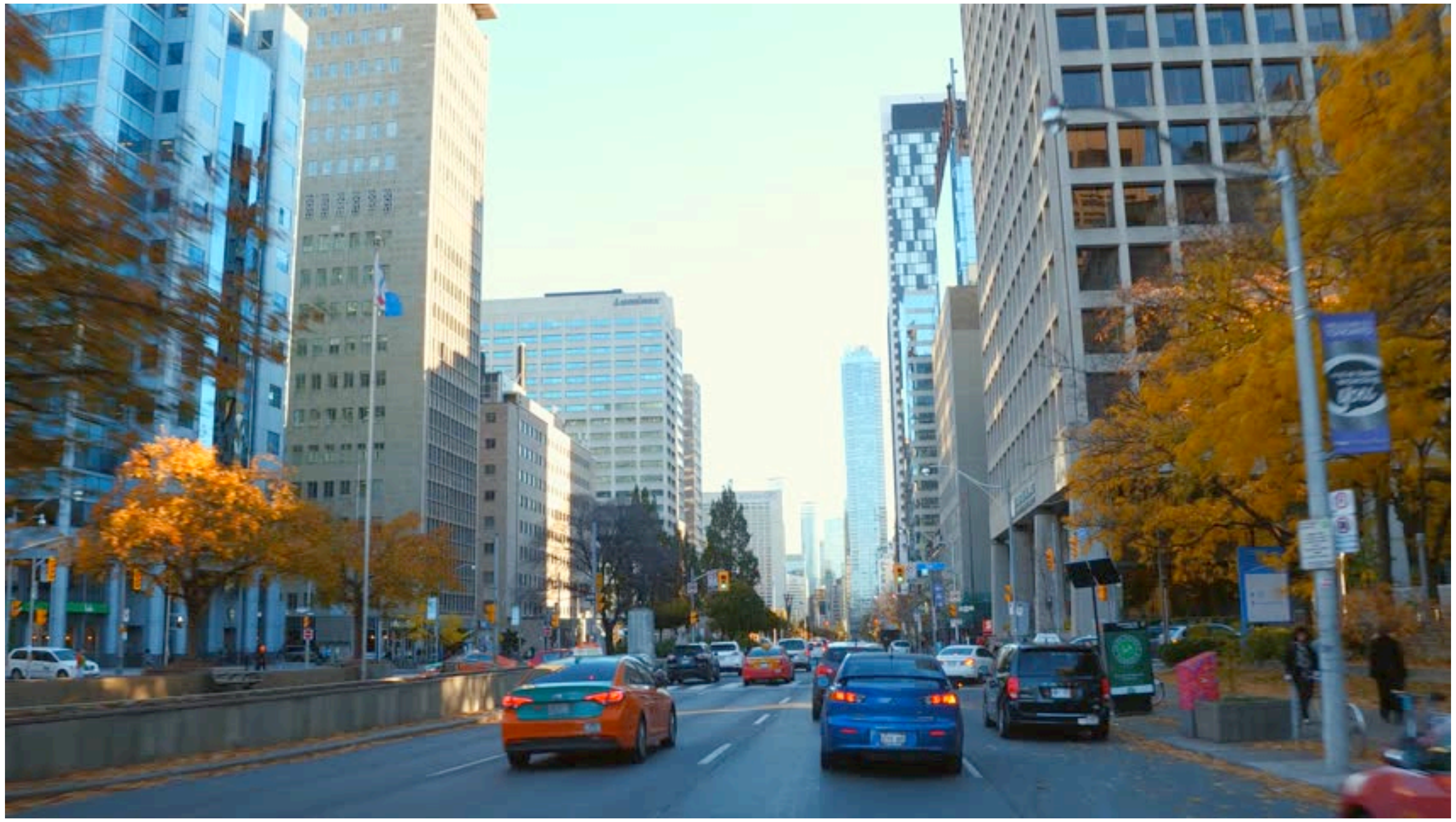
Single image
Multiple images
Depth image
...
3D voxels

Neural Net

Depth image
Material properties
View synthesis
...
3D Models

Estimating depth from a single image

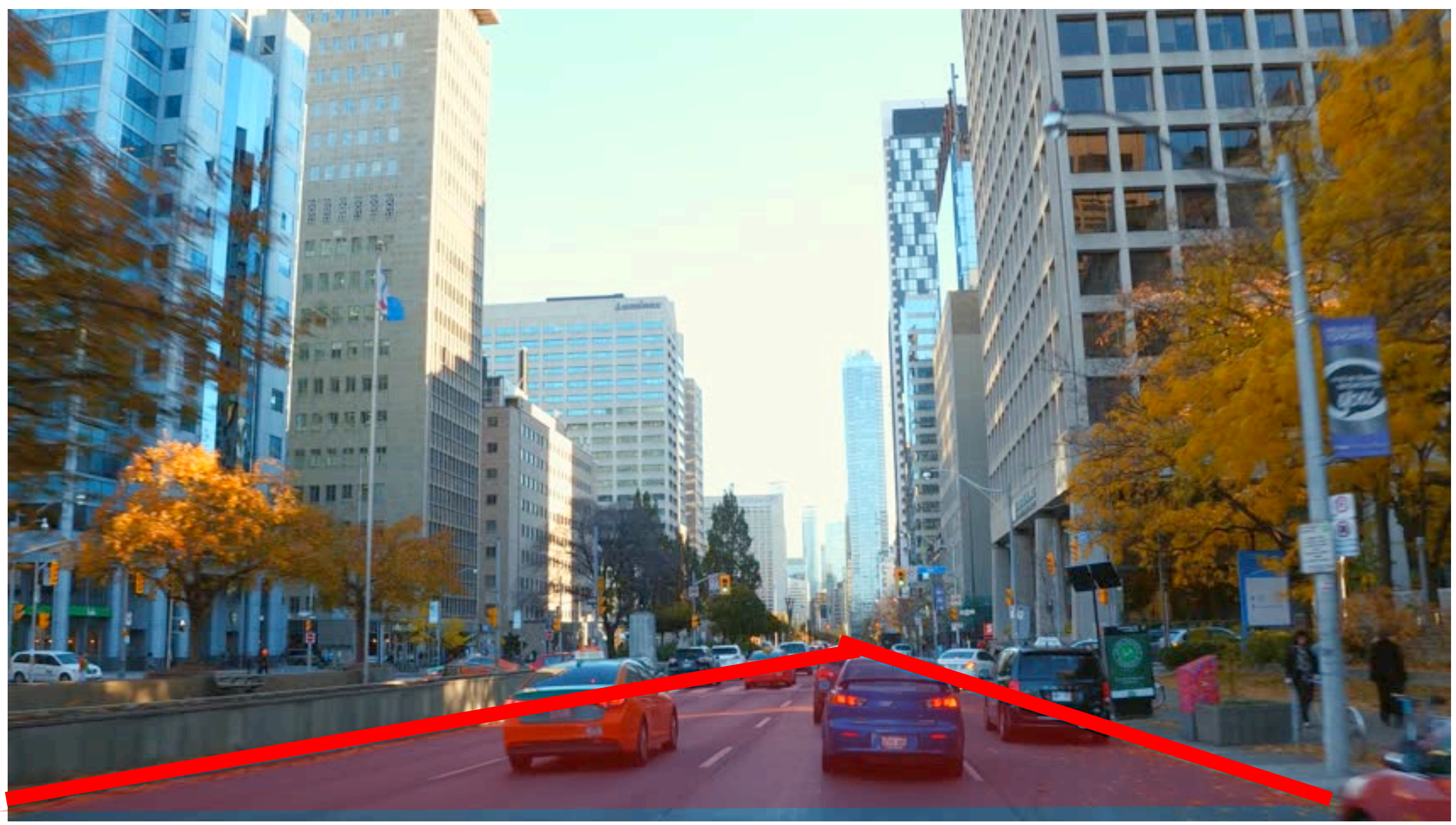
- Why is this even possible?



Slide Credit: Bharath Hariharan

Estimating depth from a single image

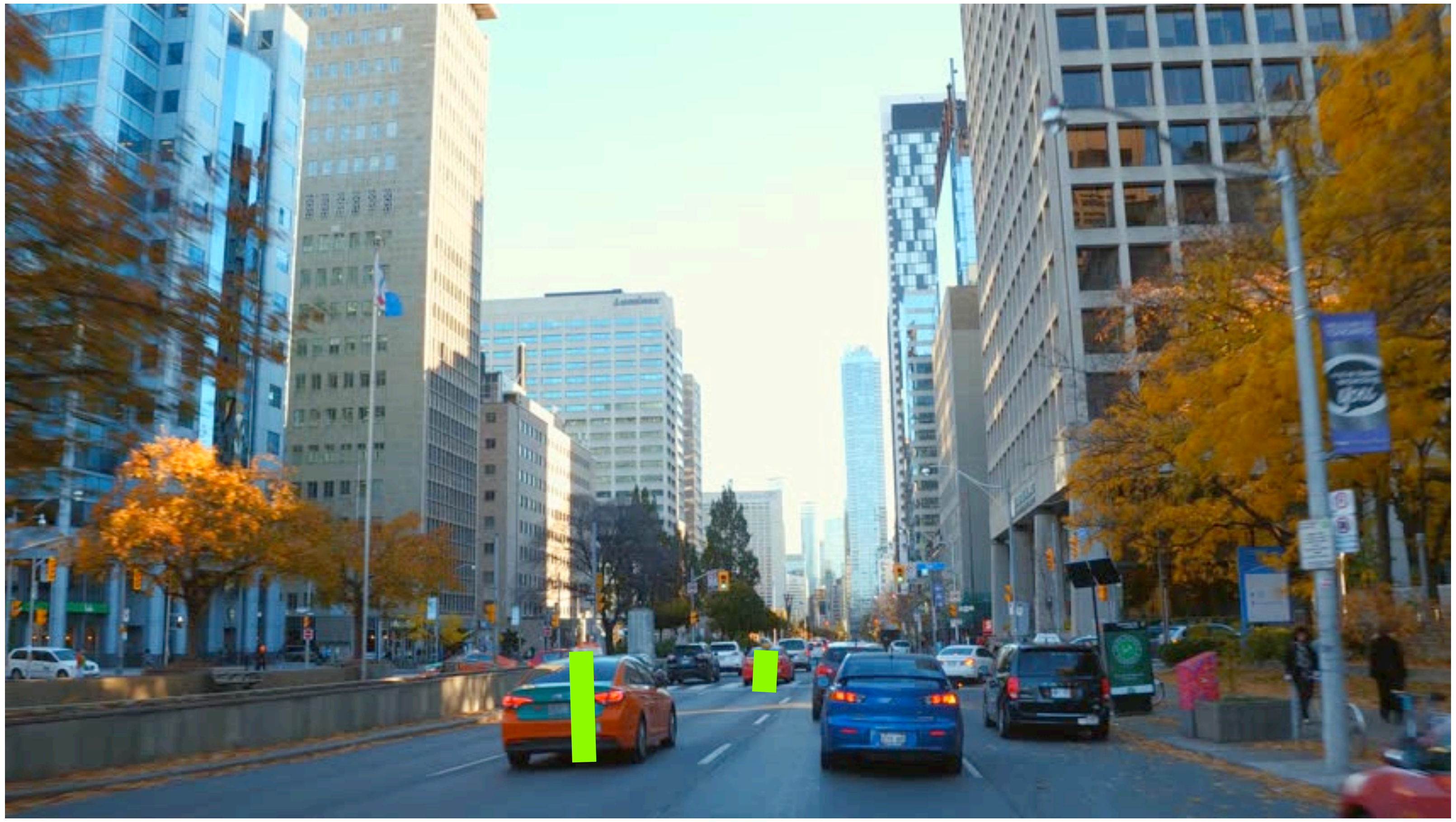
- Why is this even possible?



Slide Credit: Bharath Hariharan

Estimating depth from a single image

- Why is this even possible?

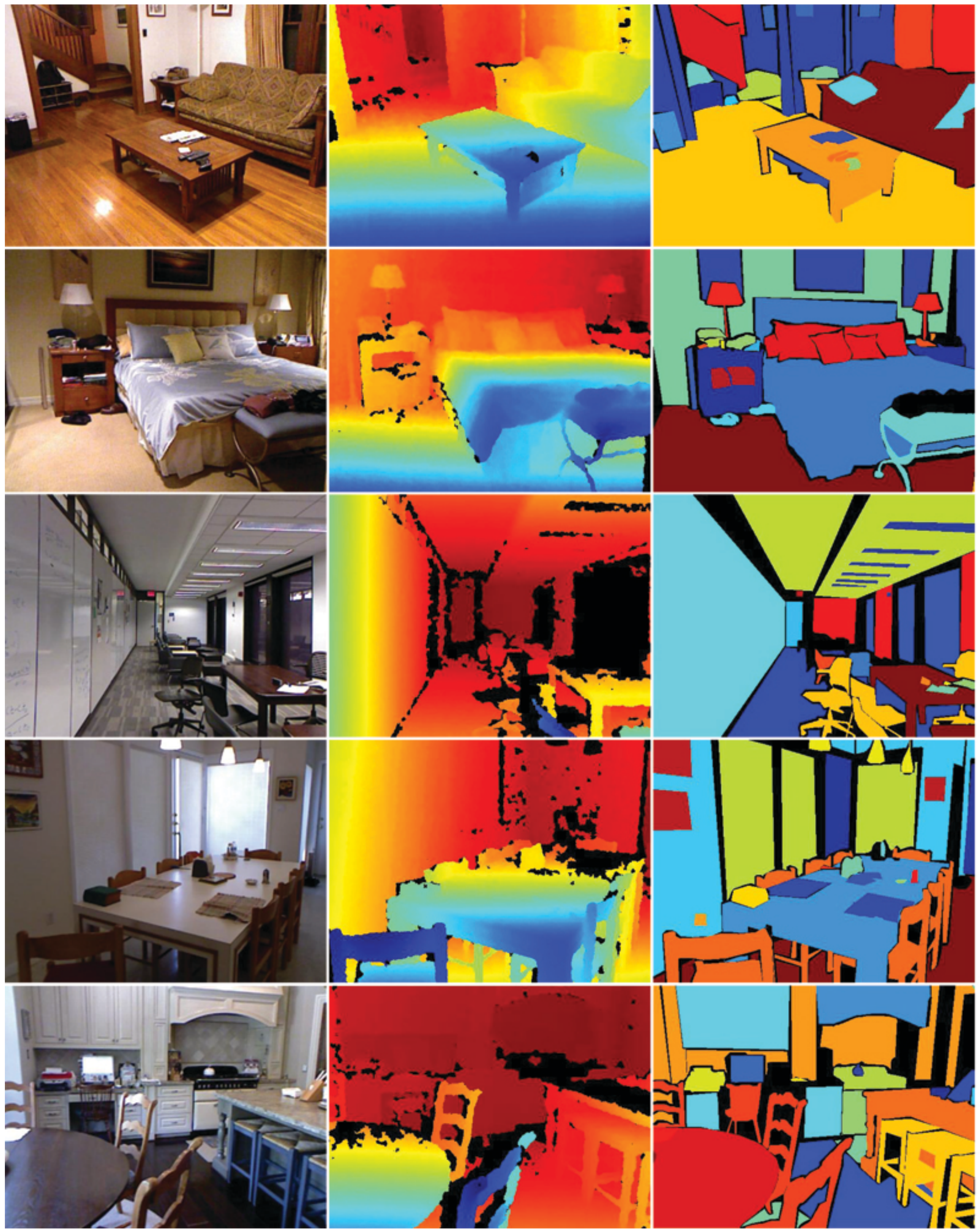


Slide Credit: Bharath Hariharan

First we need lots of data…

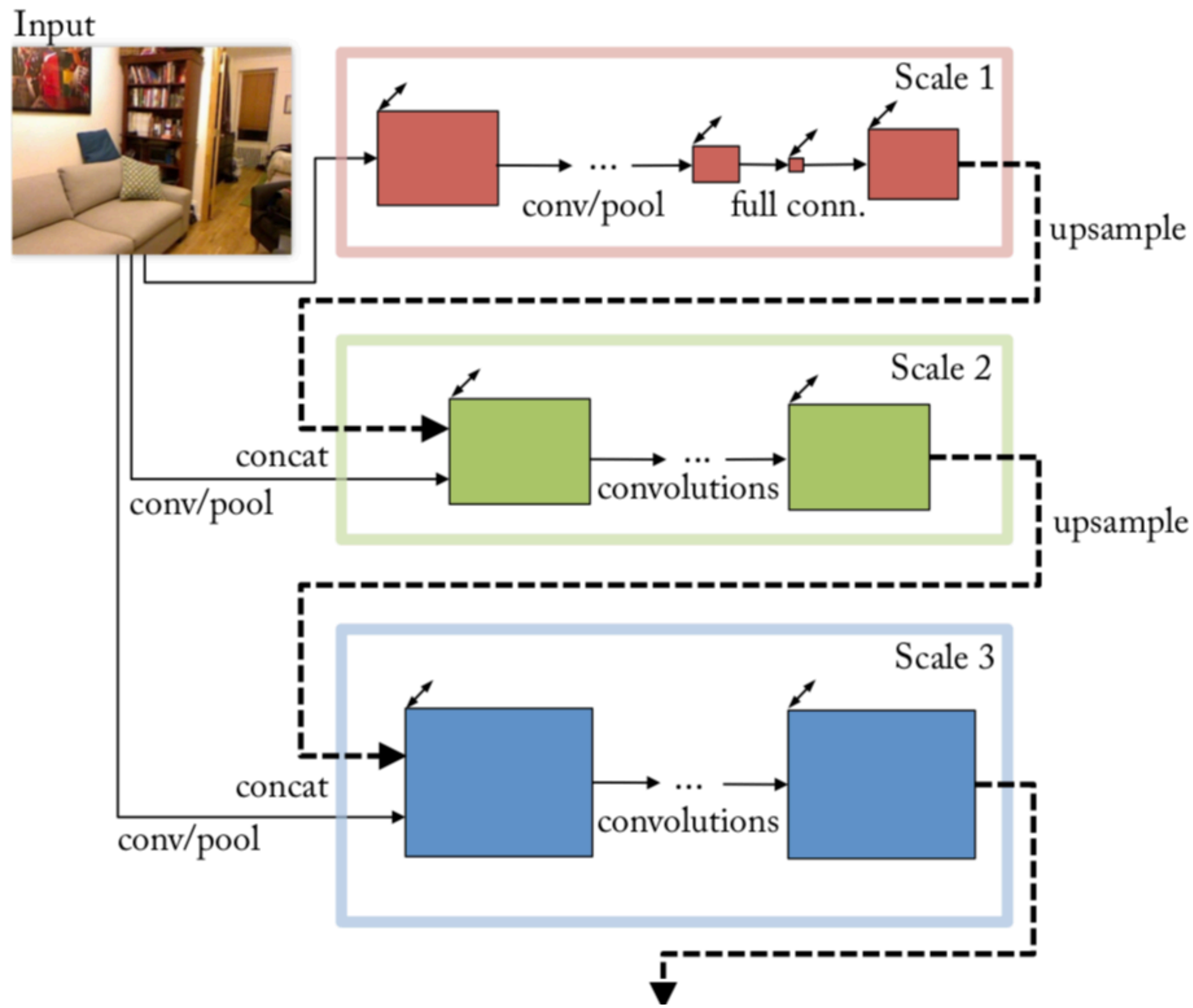
Find a gaming friend

Collect depth data



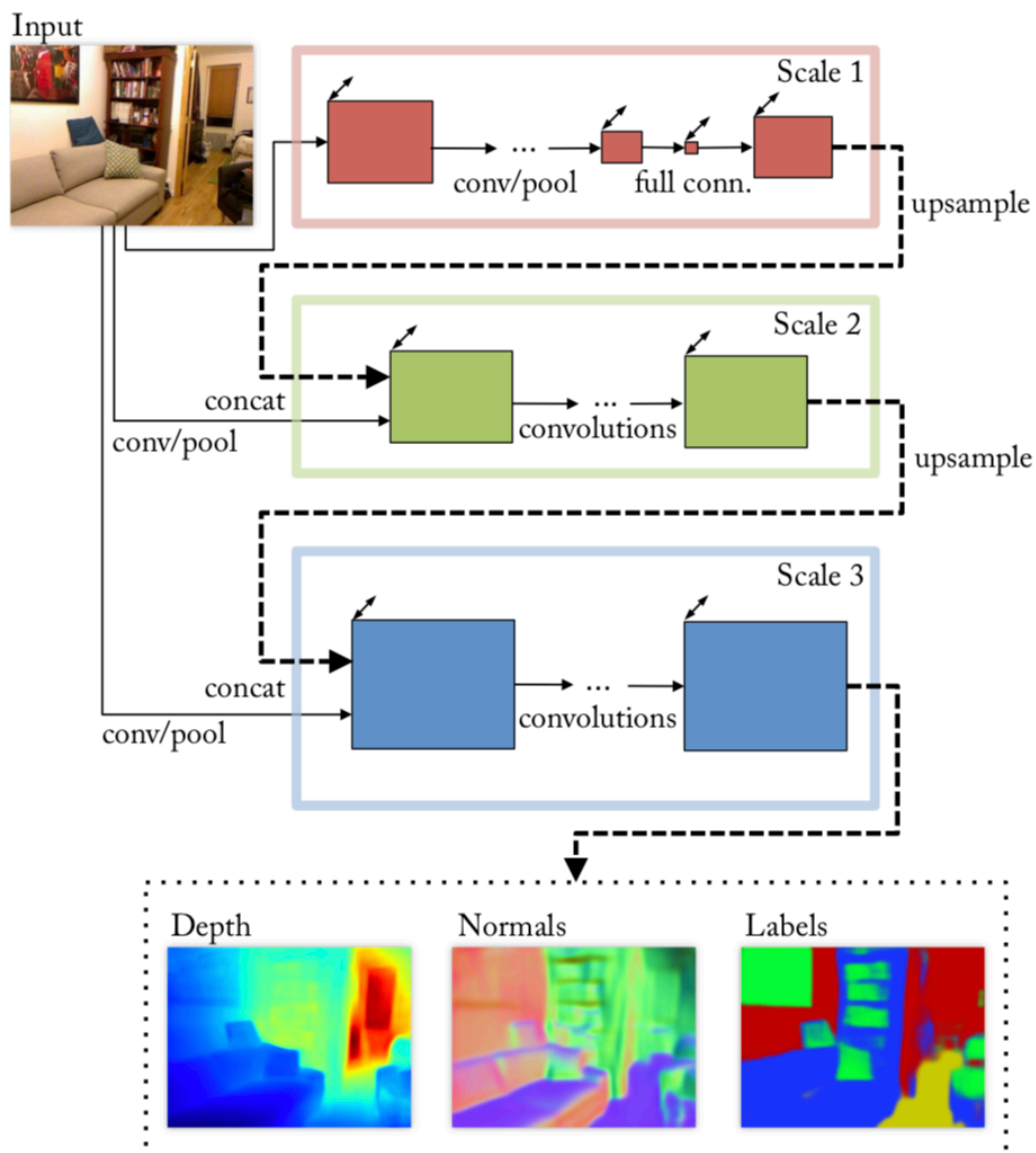
NYU Depth v2

Then create a neural net

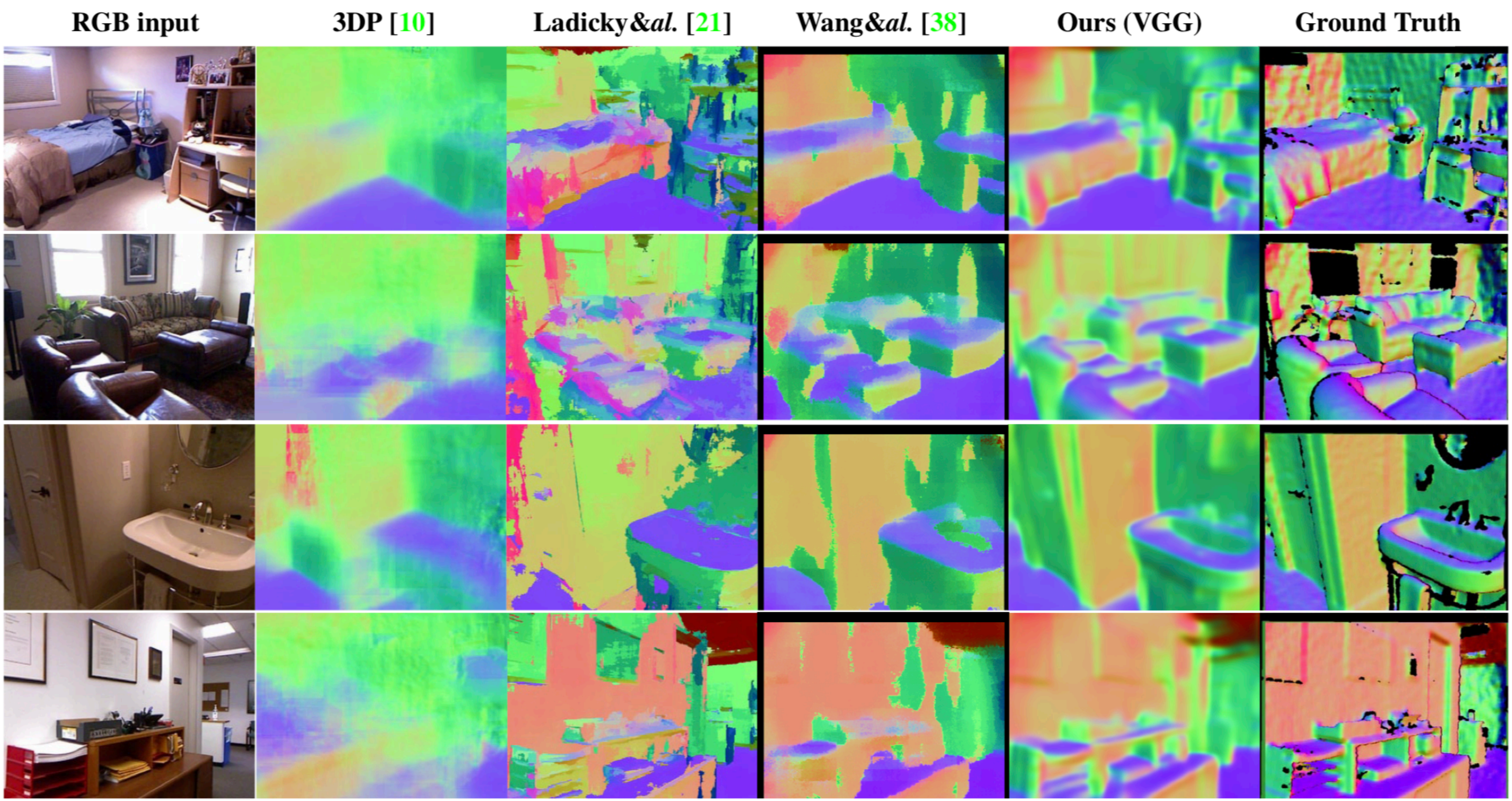


Go back to gaming friend

Predict depth & normals

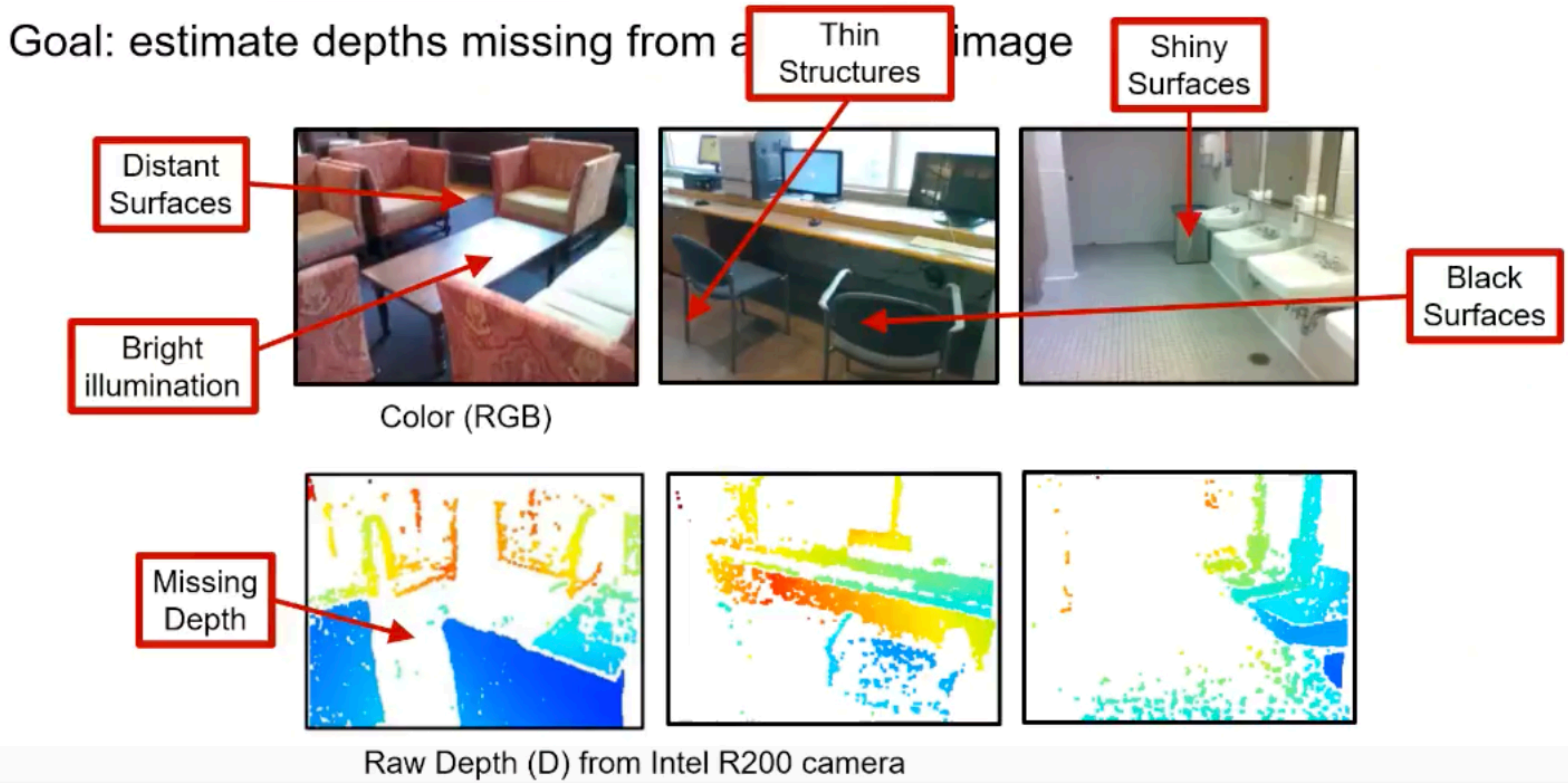


Predict depth & normals



David Eigen and Rob Fergus, NYU

Missing Depth



Slide from Thomas Funkhouser

Go back to gaming friend

Go back to gaming friend

Playing for Depth

Mohammad M. Haji-Esmaeili
mohammadhaji@modares.ac.ir

Gholamali Montazer
montazer@modares.ac.ir

Figure 1: Images and Depths extracted from the game Grand Theft Auto V

Depth predictions

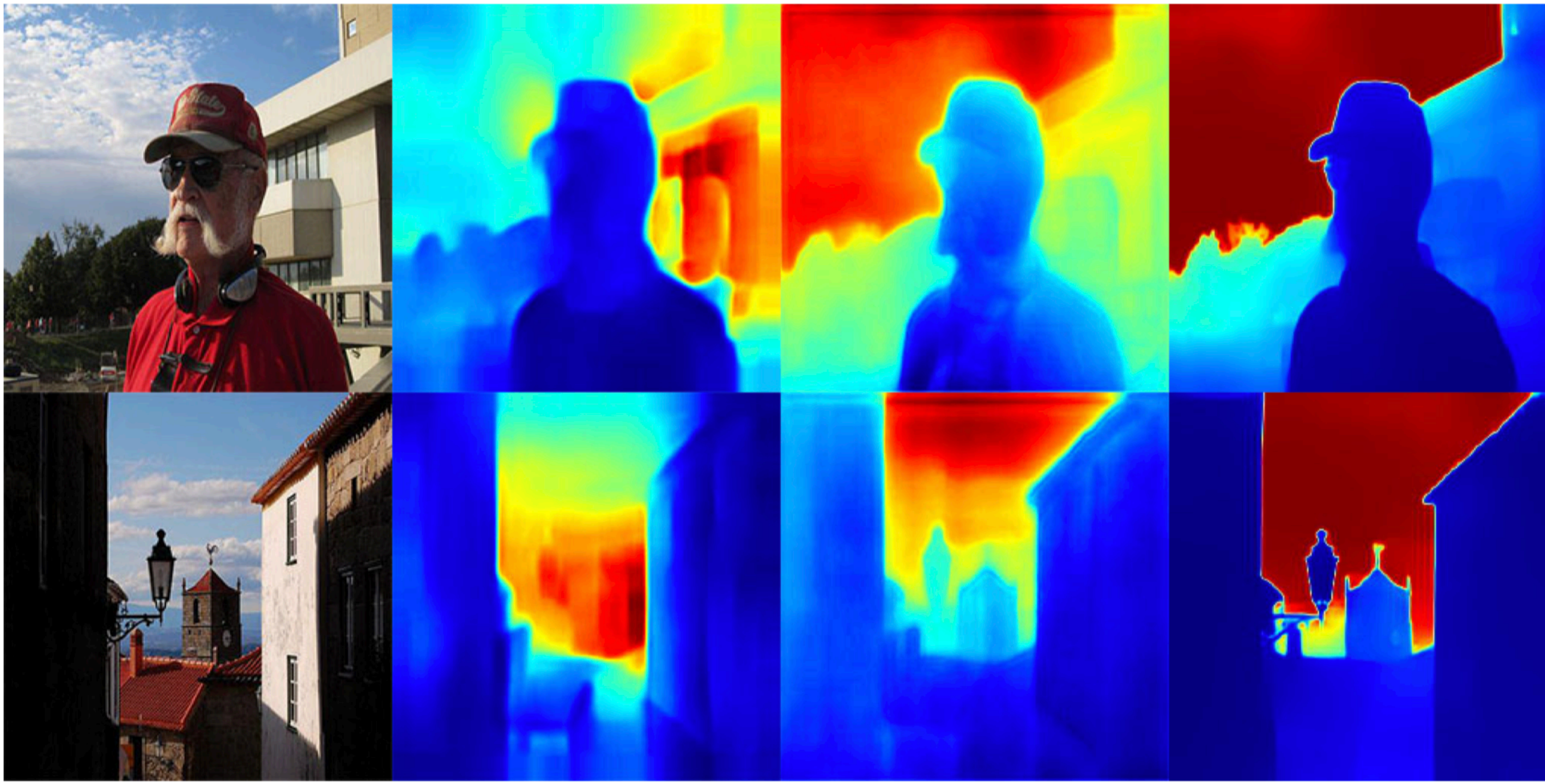


Figure 9: Qualitative Results. From left to right: Input Image, Eigen [8], DIW (Full) [10], Our Approach

Depth predictions

No flowers in Grand Theft Auto V?

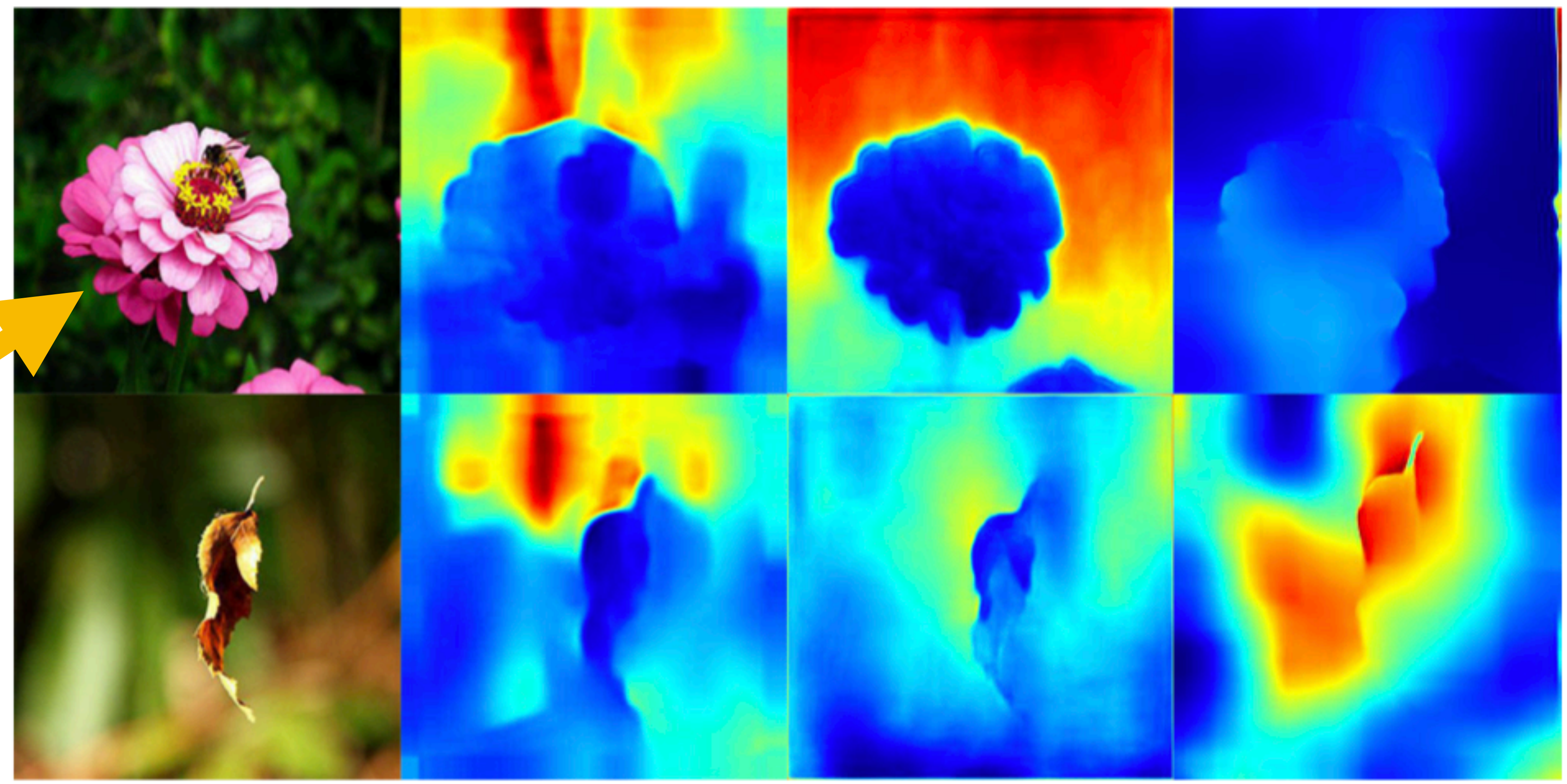
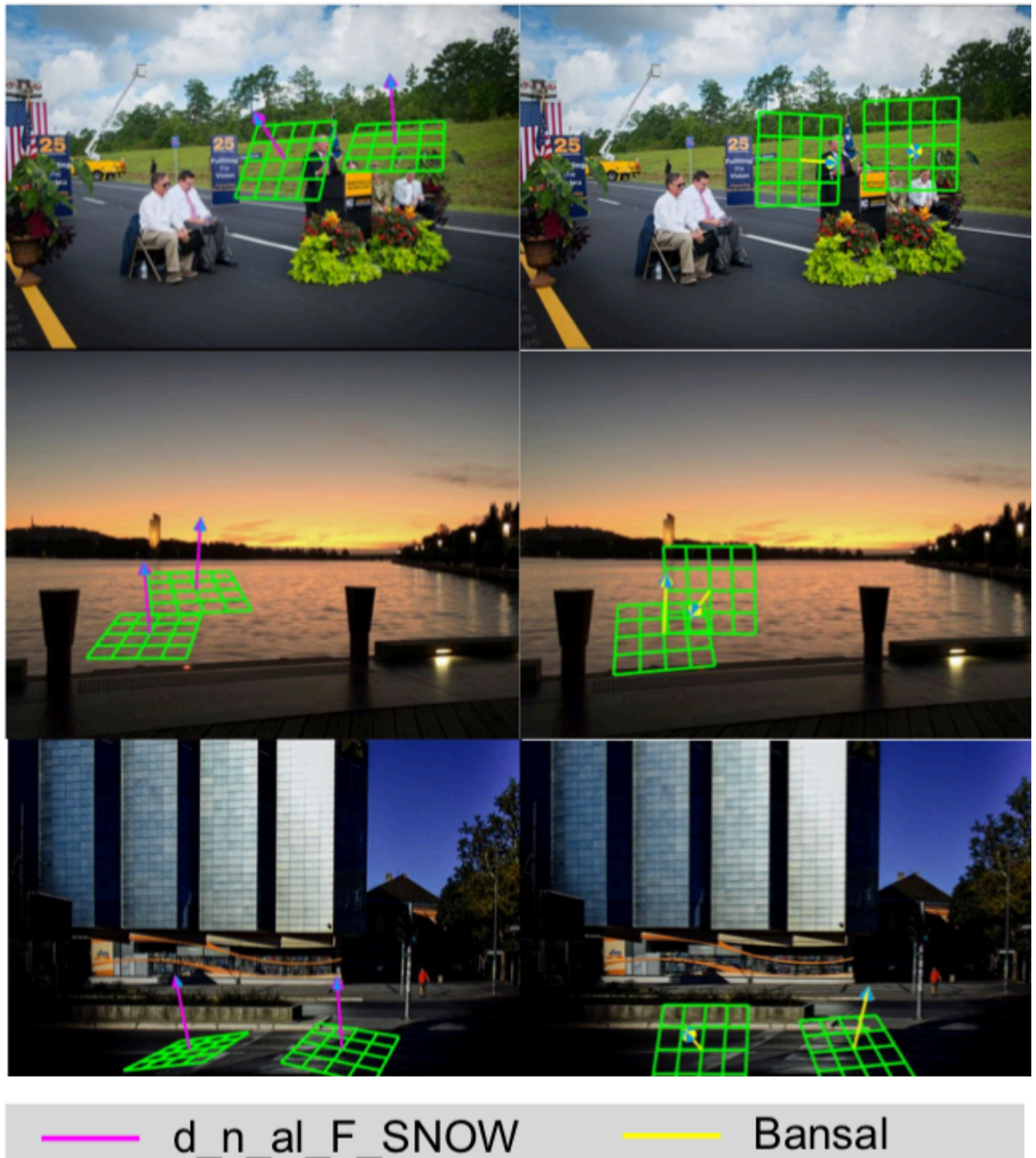


Figure 11: Failure Cases. From left to right: Input Image, Eigen [8], DIW (Full) [10], Our Approach

Go make a game

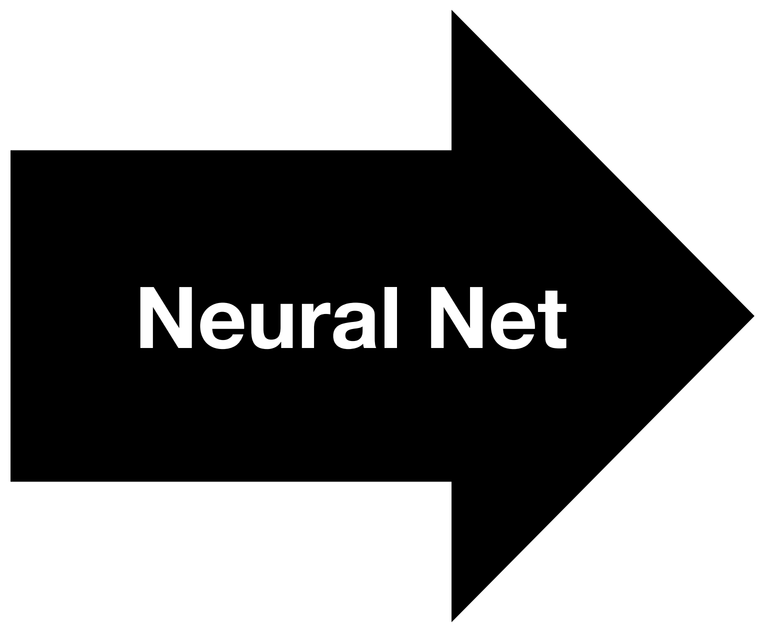
Weifeng Chen, Donglai Xiang, Jia Deng
Surface Normal Estimation in the Wild

Normal predictions



Weifeng Chen, Donglai Xiang, Jia Deng
Surface Normal Estimation in the Wild

Single image
Multiple images
Depth image
...
3D voxels



Depth image
Material properties
View synthesis
...
3D Models