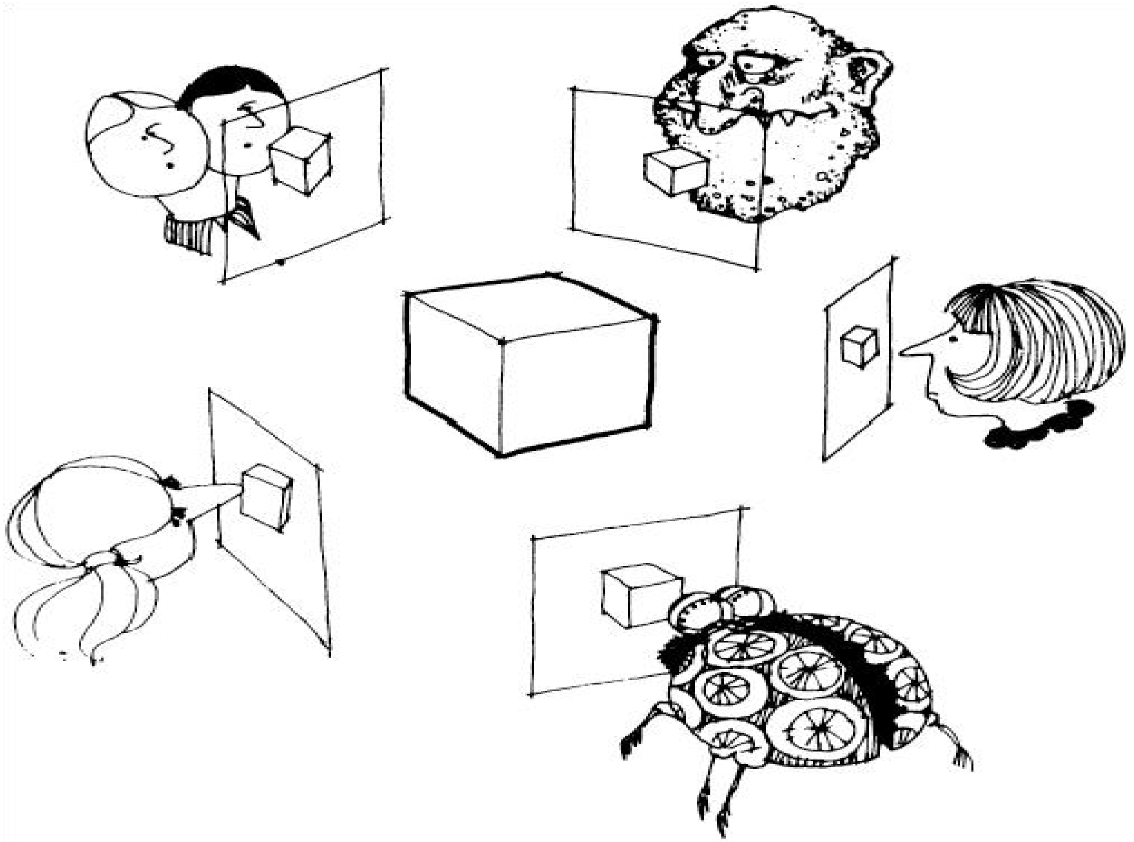
Project-5 3D Reconstruction, View Synthesis

Yujiao Shi
SIST, ShanghaiTech
Autumn, 2024

Multi-view stereo

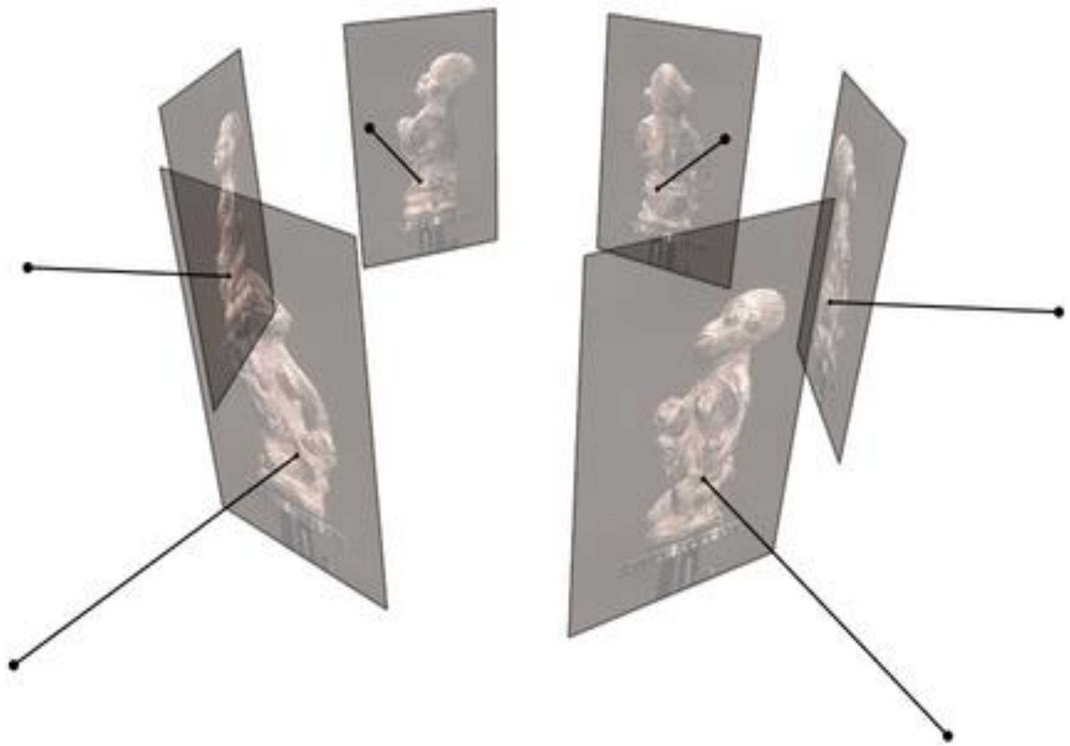


Many slides adapted from S. Seitz, Y. Furukawa, N. Snavely

Multi-view stereo

上海科技大学 ShanghaiTech University

- Goal: given several images of the same object or scene, compute a representation of its 3D shape



Source: C. Hernandez, N. Snavely

Multi-view stereo

上海科技大学
ShanghaiTech University

- Goal: given several images of the same object or scene, compute a representation of its 3D shape
- “Images of the same object or scene”
  - Arbitrary number of images (from two to thousands)
  - Arbitrary camera positions (special rig, camera network or video)
  - Calibration may be known or unknown

Multi-view stereo

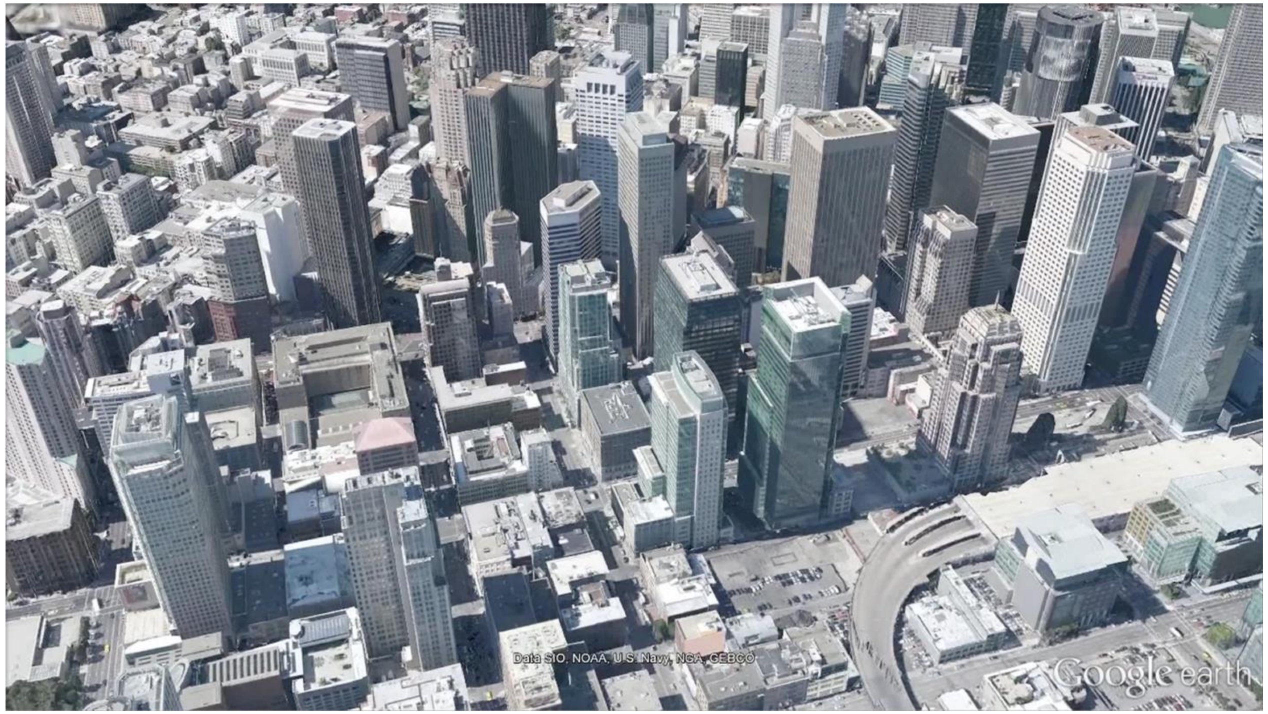
上海科技大学 ShanghaiTech University

- Goal: given several images of the same object or scene, compute a representation of its 3D shape
- “Images of the same object or scene”
  - Arbitrary number of images (from two to thousands)
  - Arbitrary camera positions (special rig, camera network or video)
  - Calibration may be known or unknown
- “Representation of 3D shape”
  - Depth maps
  - Meshes
  - Point clouds
  - Patch clouds
  - Volumetric models
  - Neural radiance field
  - SDF
  - 3D Gaussians
  - ….

Applications

Source: N. Snavely

Applications



Source: N. Snavely

Applications

上海科技大学
ShanghaiTech University

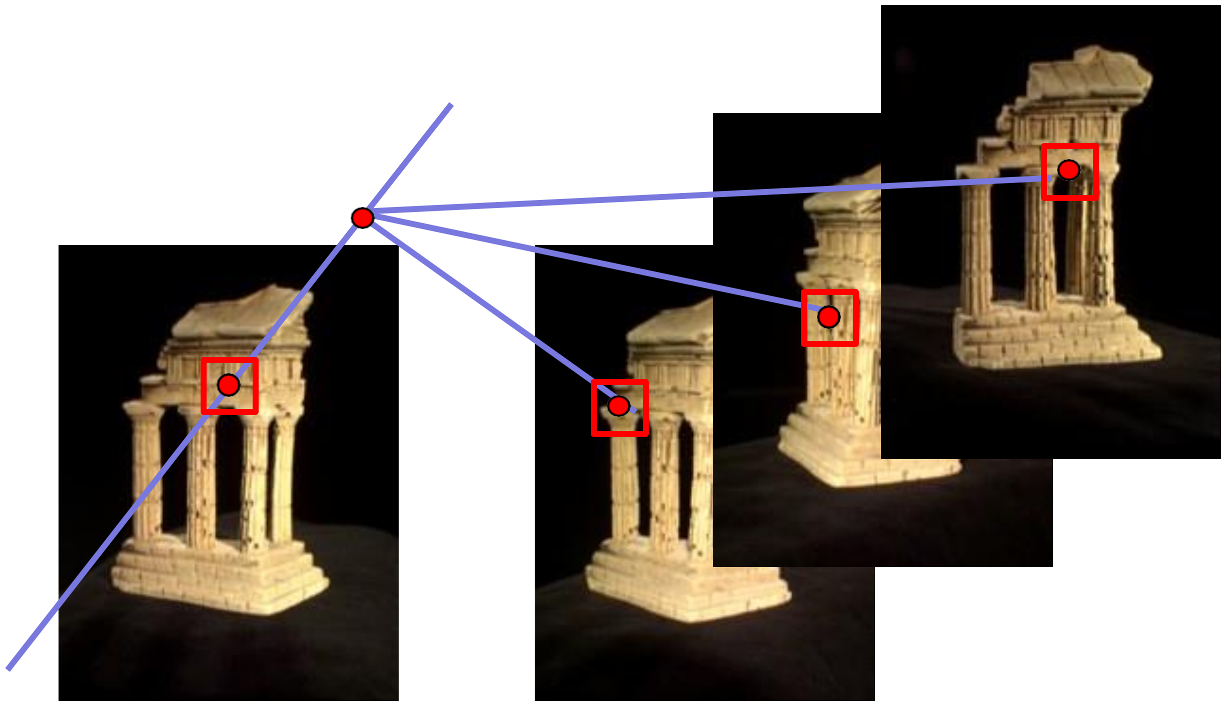
Source: N. Snavely

Applications

- Enable inspection in hard to reach areas with drone photos and 3D reconstruction
- Create 3D model from images
- Provide tools to inspect on images and map interactions to 3D

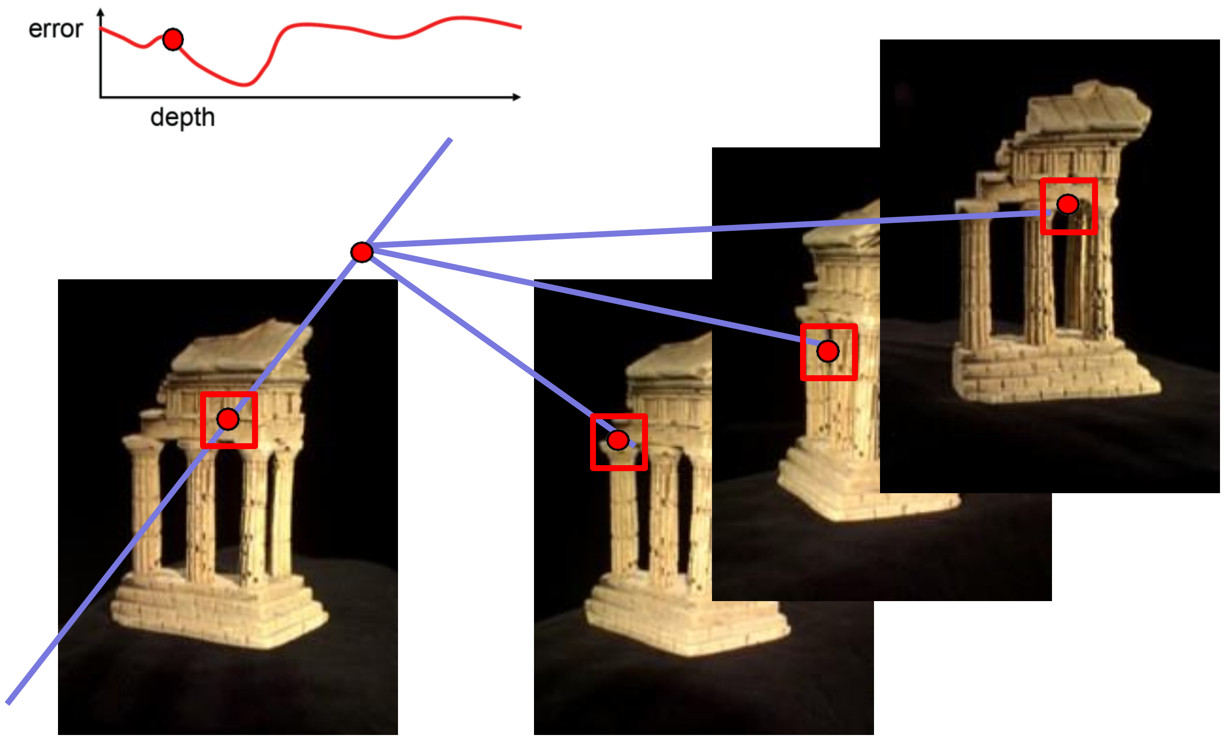
Source: D. Hoiem

Multi-view stereo: Basic idea



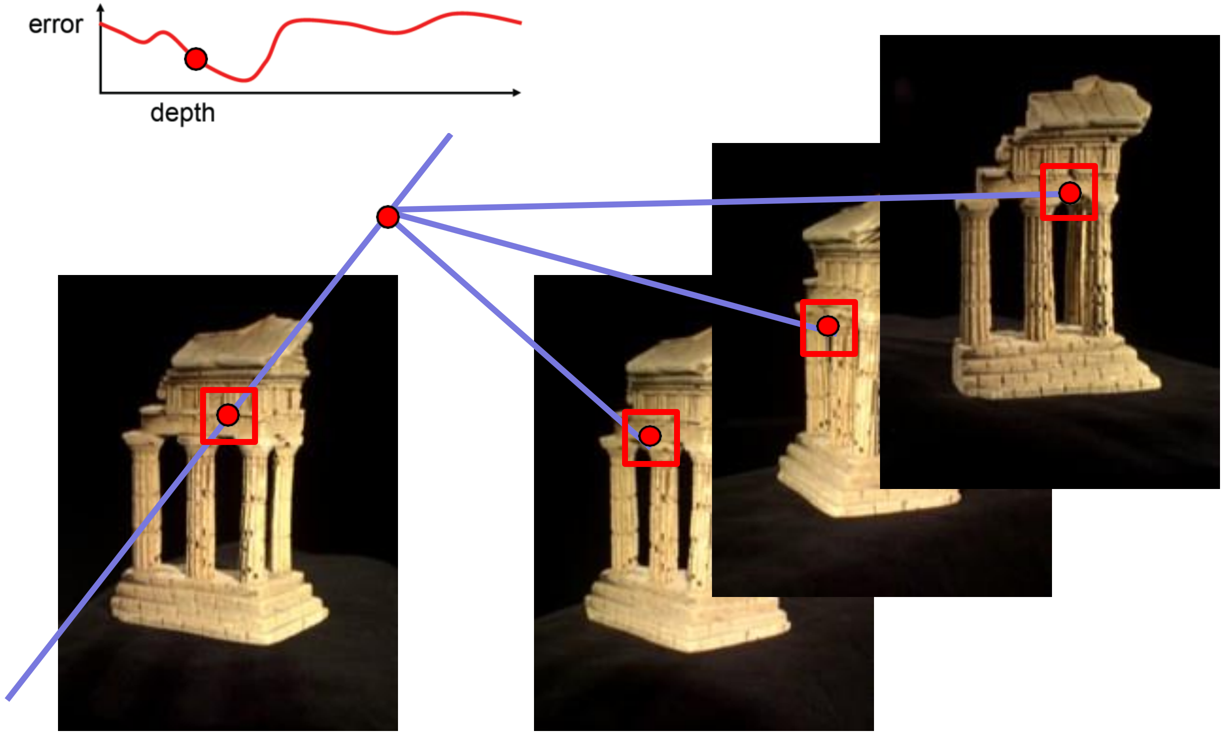
Source: Y. Furukawa

Multi-view stereo: Basic idea



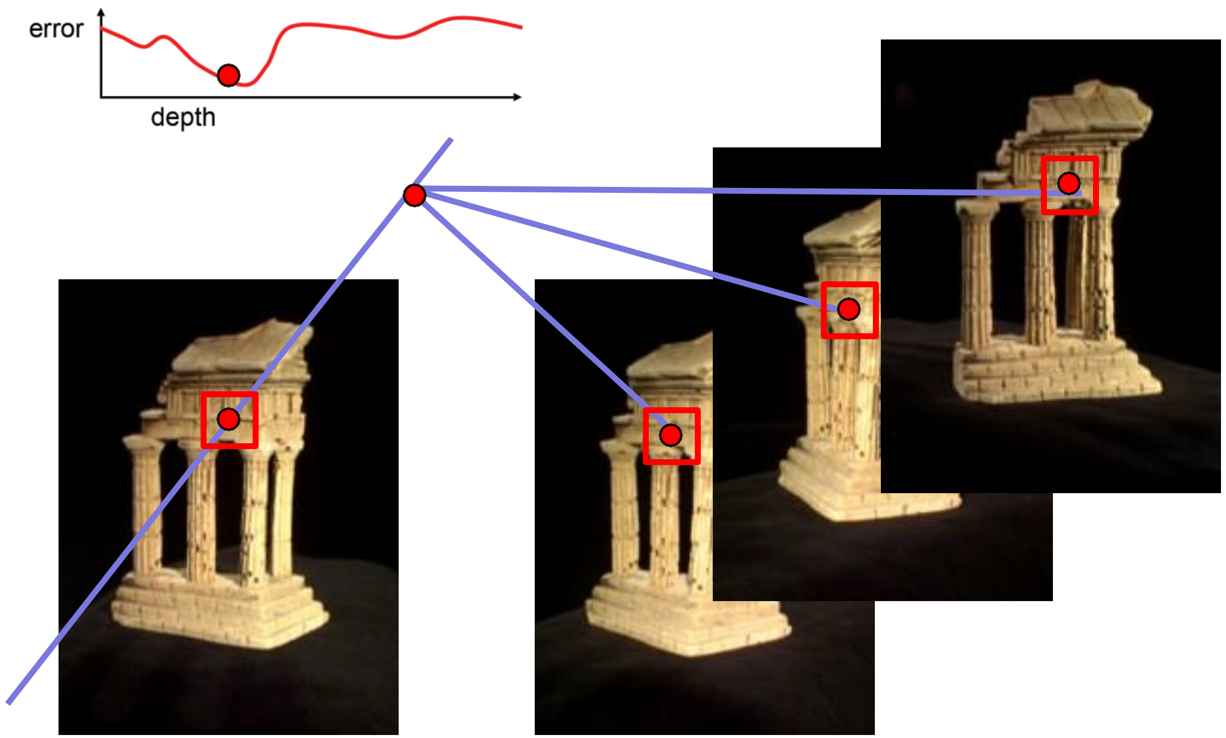
Source: Y. Furukawa

Multi-view stereo: Basic idea



Source: Y. Furukawa

Multi-view stereo: Basic idea

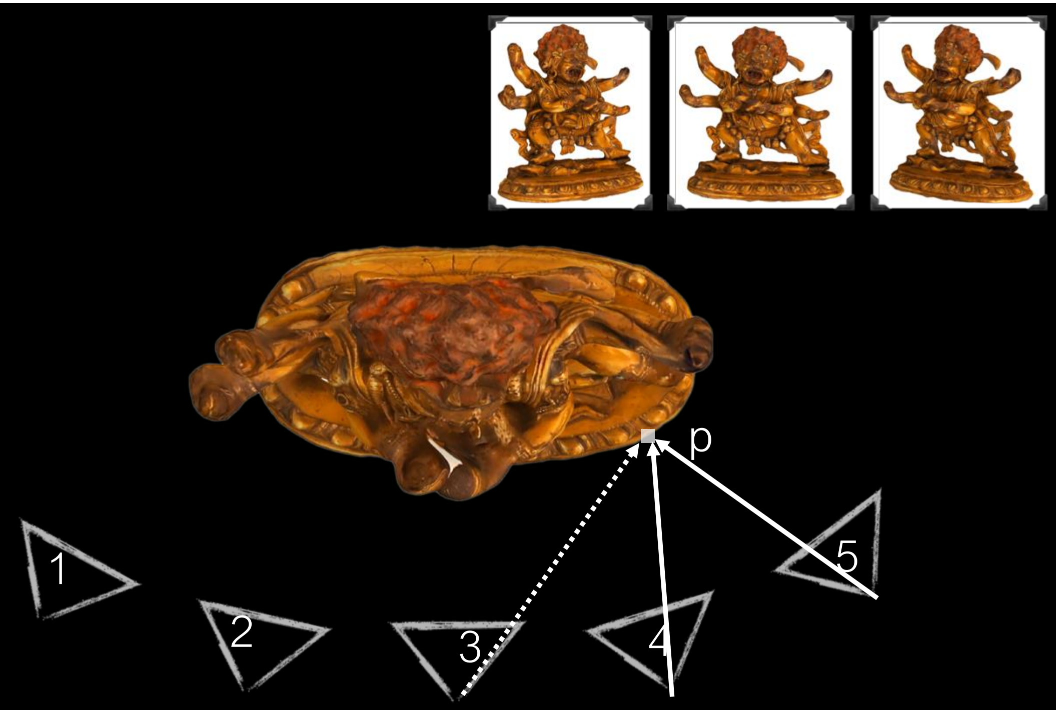


Source: Y. Furukawa

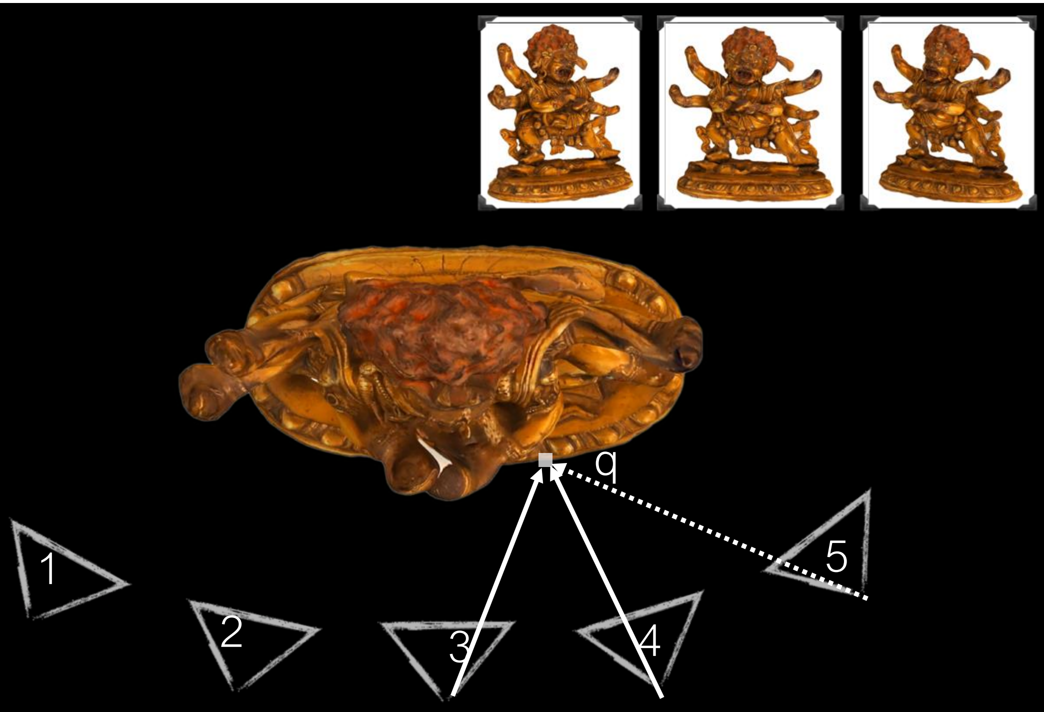
Why MVS?

- Different points on the object’s surface will be more clearly visible in some subset of cameras
  - Could have high-res closeups of some regions
  - Some surfaces are foreshortened from certain views
  - Some points may be occluded entirely in certain views

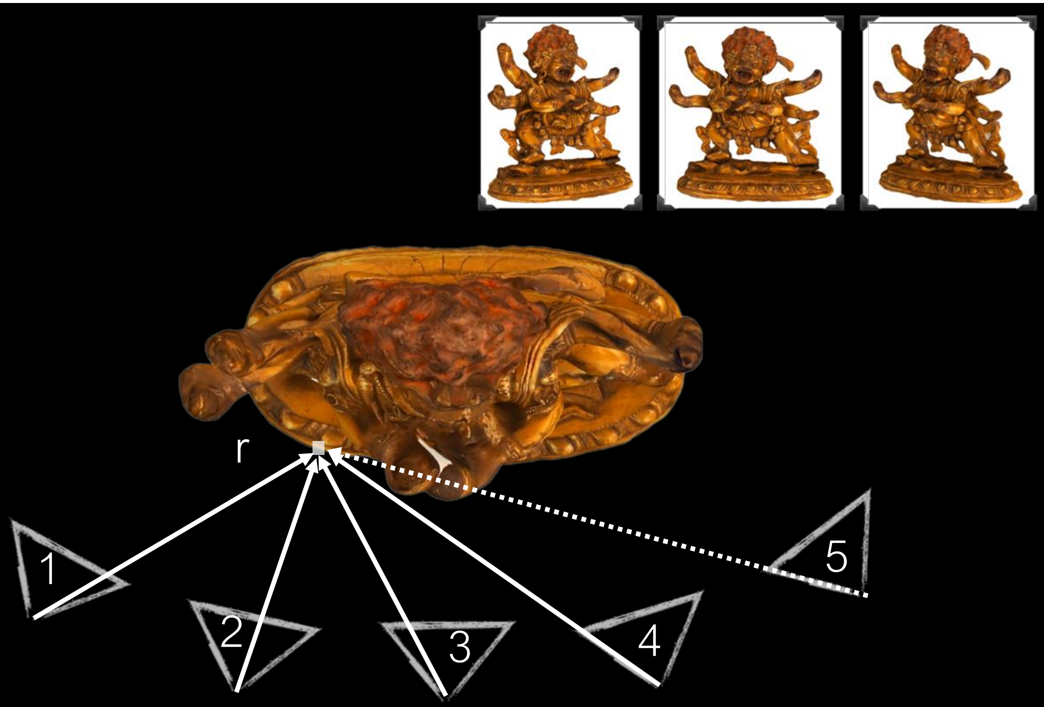
Source: N. Snavely



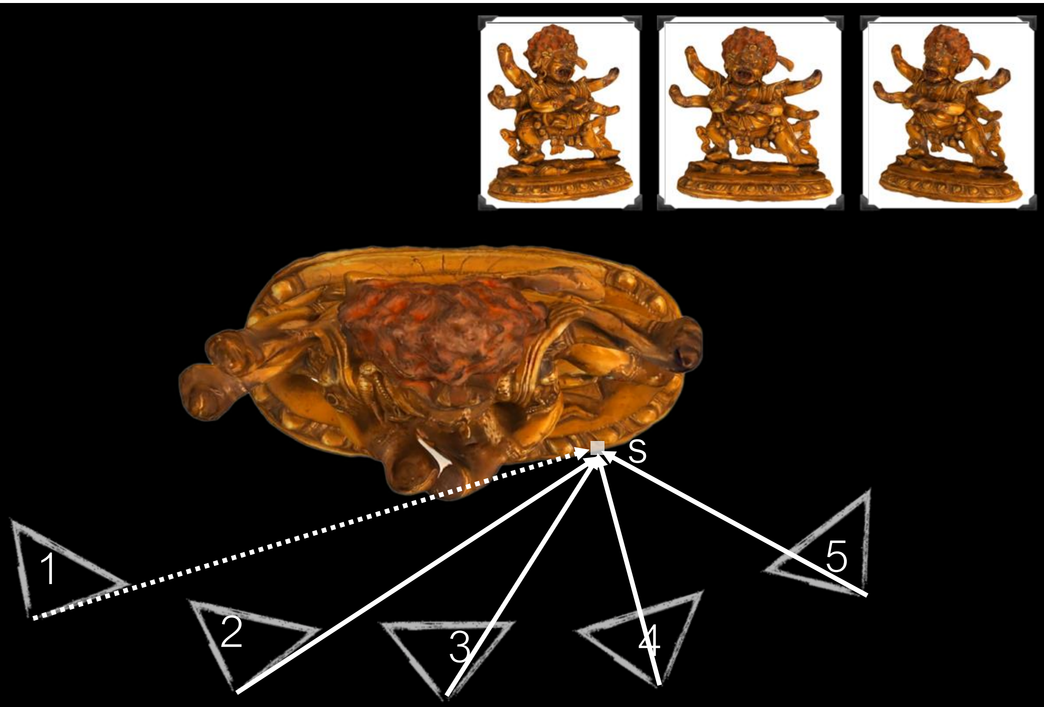
Cameras 4 and 5 can more clearly see point p



Cameras 3 and 4 can more clearly see point q



Camera 5 can't see point r



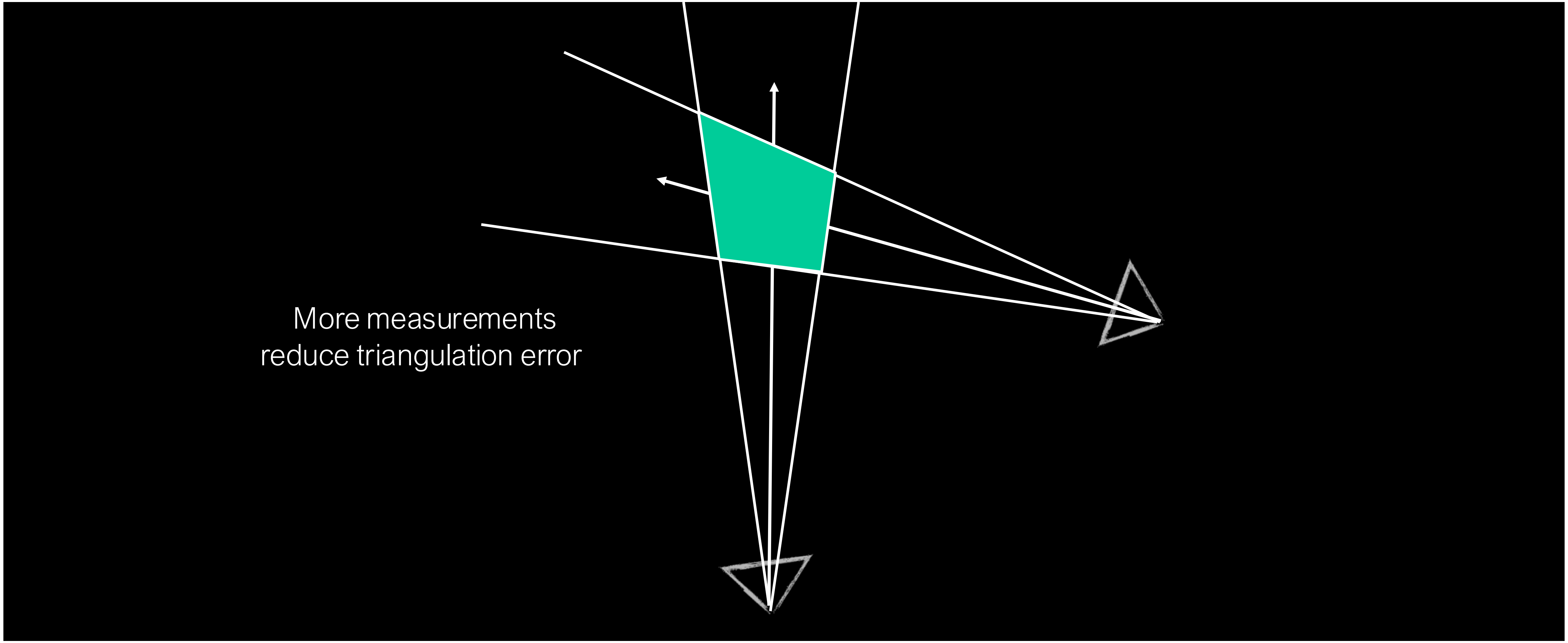
Camera 1 can’t see point s

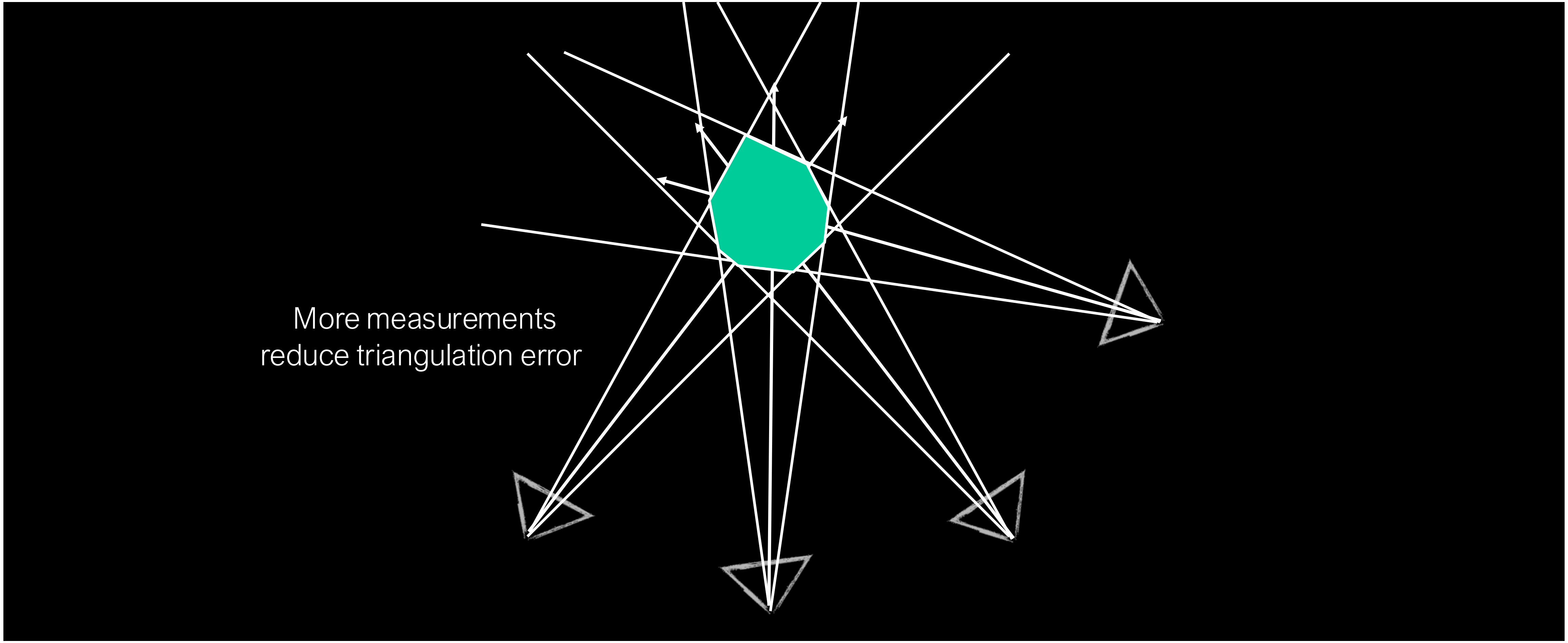
Why MVS?

上海科技大学
ShanghaiTech University

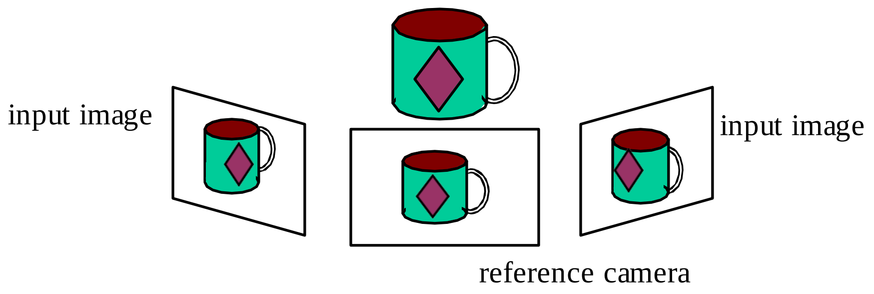
- Different points on the object's surface will be more clearly visible in some subset of cameras
    - Could have high-res closeups of some regions
    - Some surfaces are foreshortened from certain views
    - Some points may be occluded entirely in certain views
- More measurements per point can reduce error

Source: N. Snavely


More measurements
reduce triangulation error


More measurements
reduce triangulation error

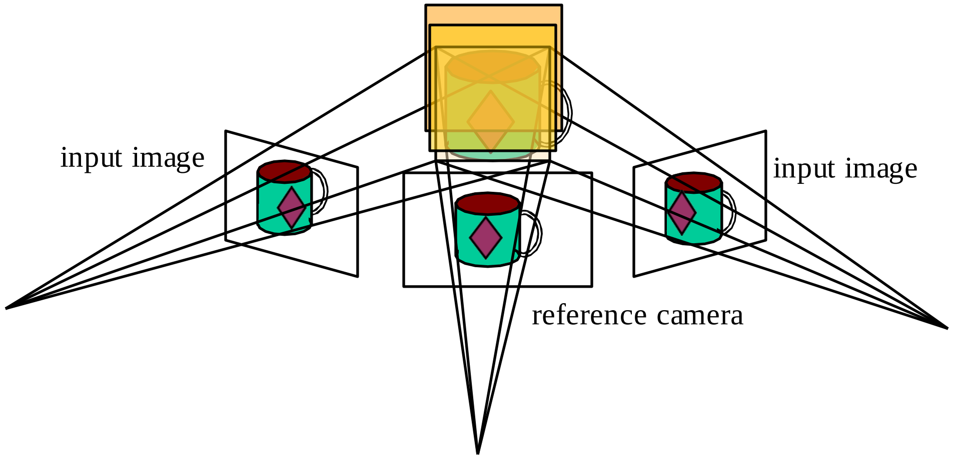
Plane sweep stereo



- Sweep plane across a range of depths w.r.t. a reference camera
- For each depth, project each input image onto that plane (homography) and compare the resulting stack of images

R. Collins, A space-sweep approach to true multi-image matching, CVPR 1996

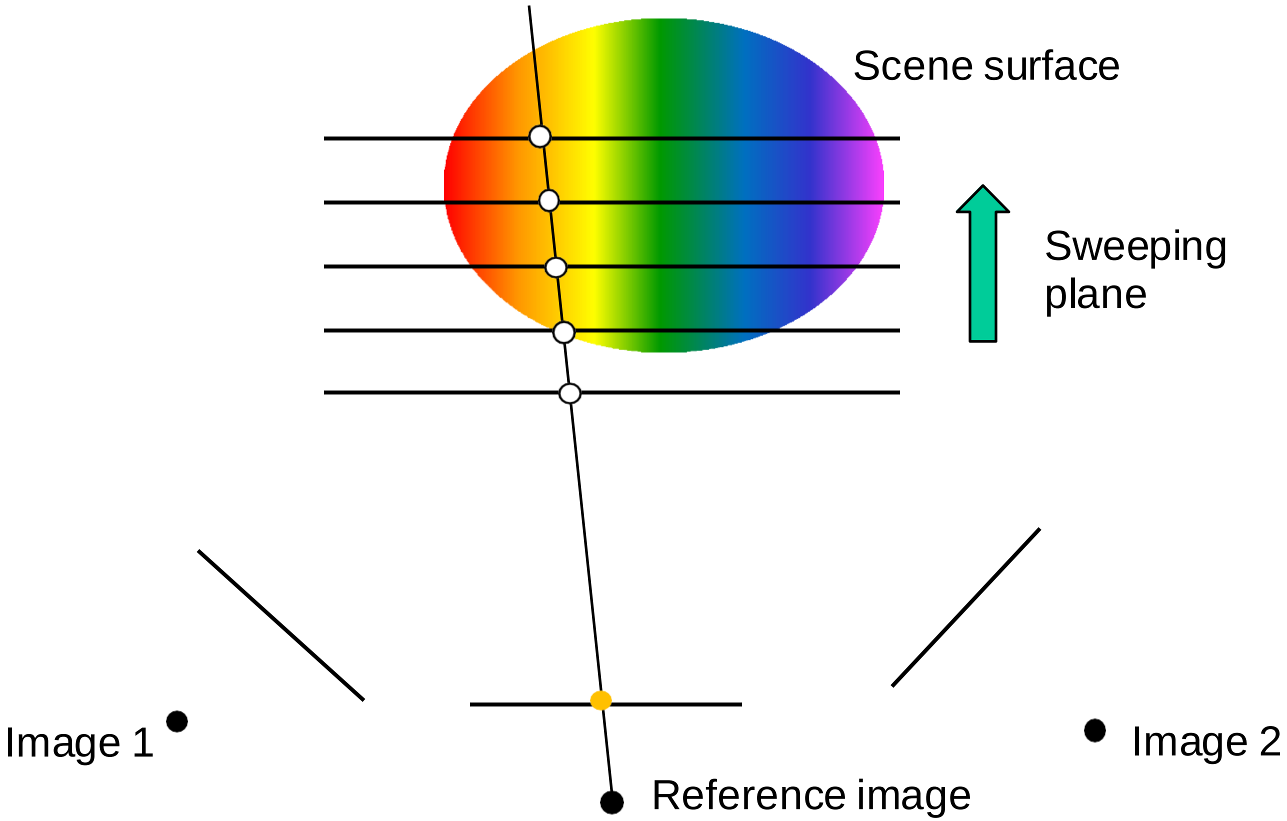
Plane sweep stereo



- Sweep plane across a range of depths w.r.t. a reference camera
- For each depth, project each input image onto that plane (homography) and compare the resulting stack of images

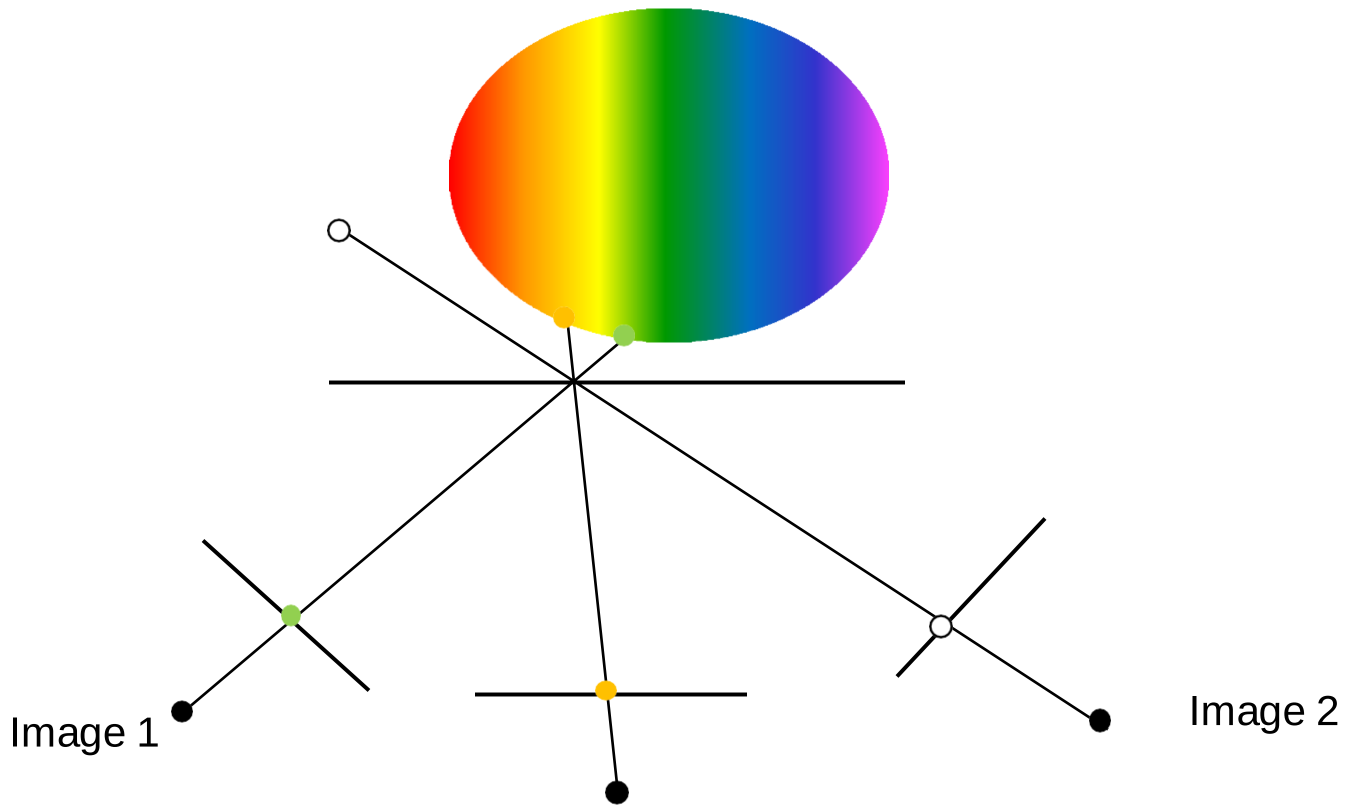
R. Collins, [A space-sweep approach to true multi-image matching](), CVPR 1996

Plane sweep stereo

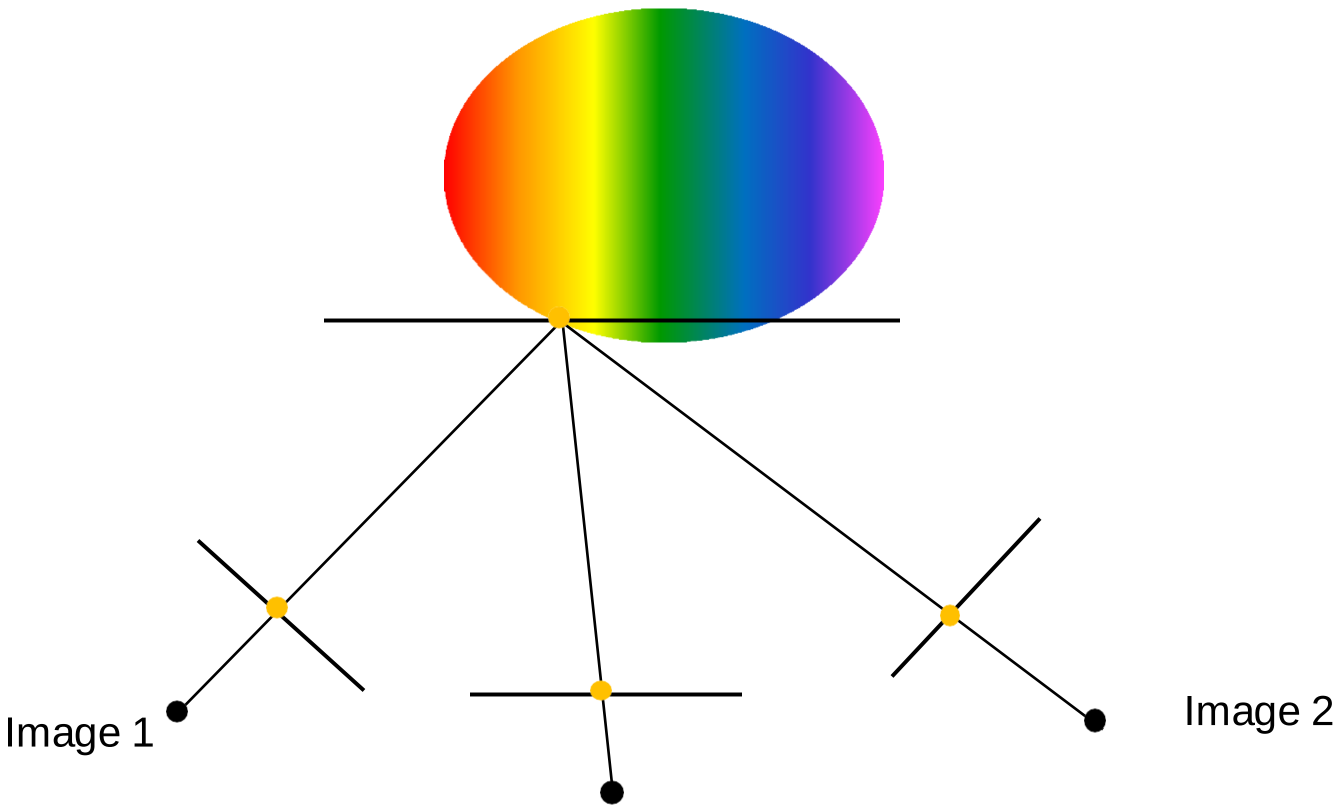


Plane sweep stereo: Key Idea

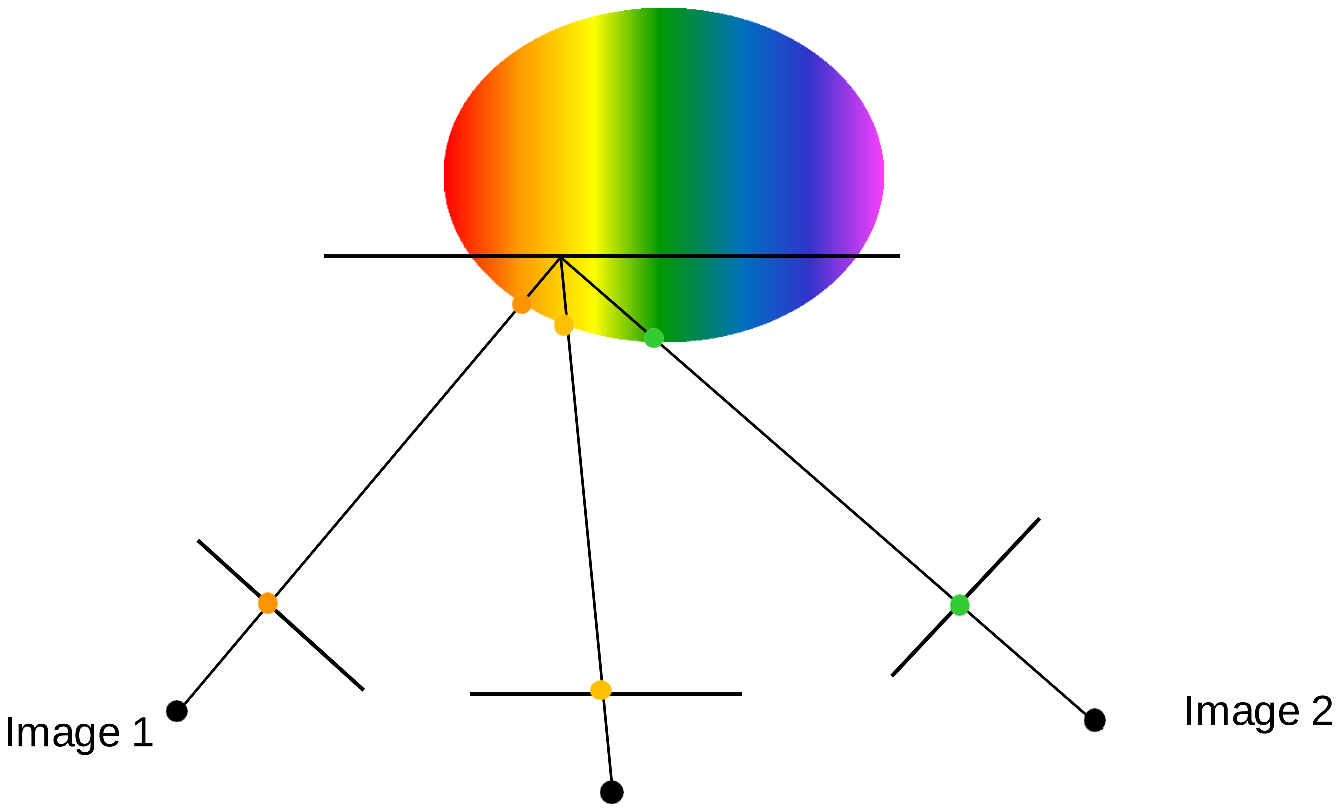
上海科技大学
ShanghaiTech University



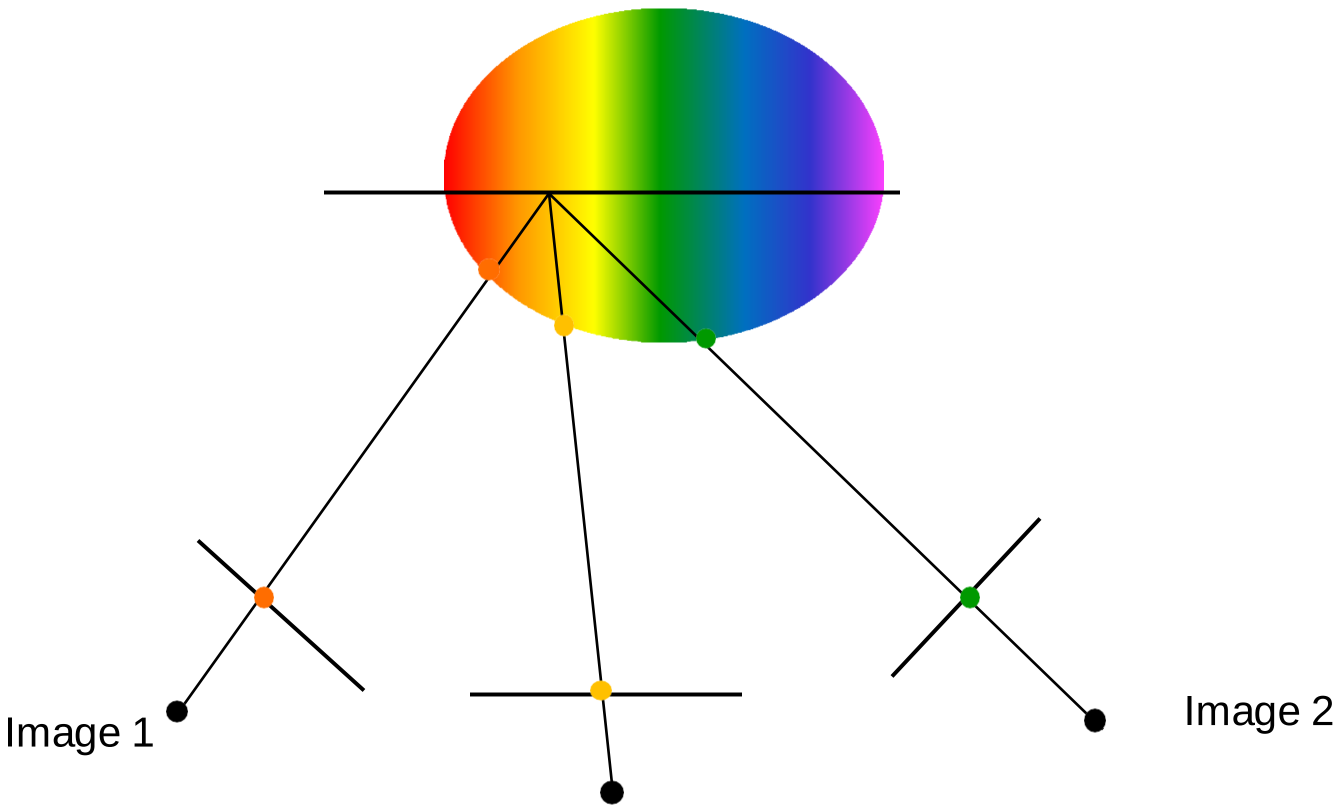
Plane sweep stereo: Key Idea



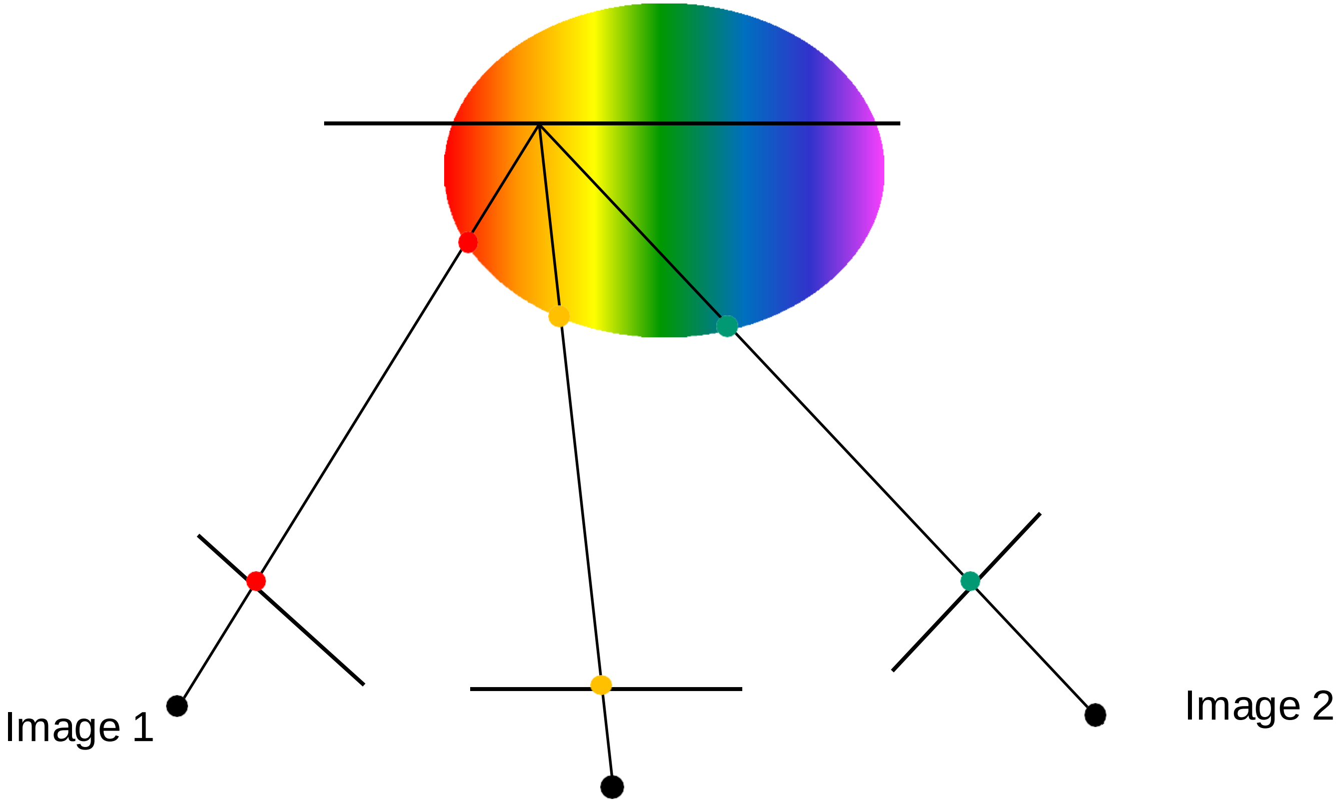
Plane sweep stereo: Key Idea



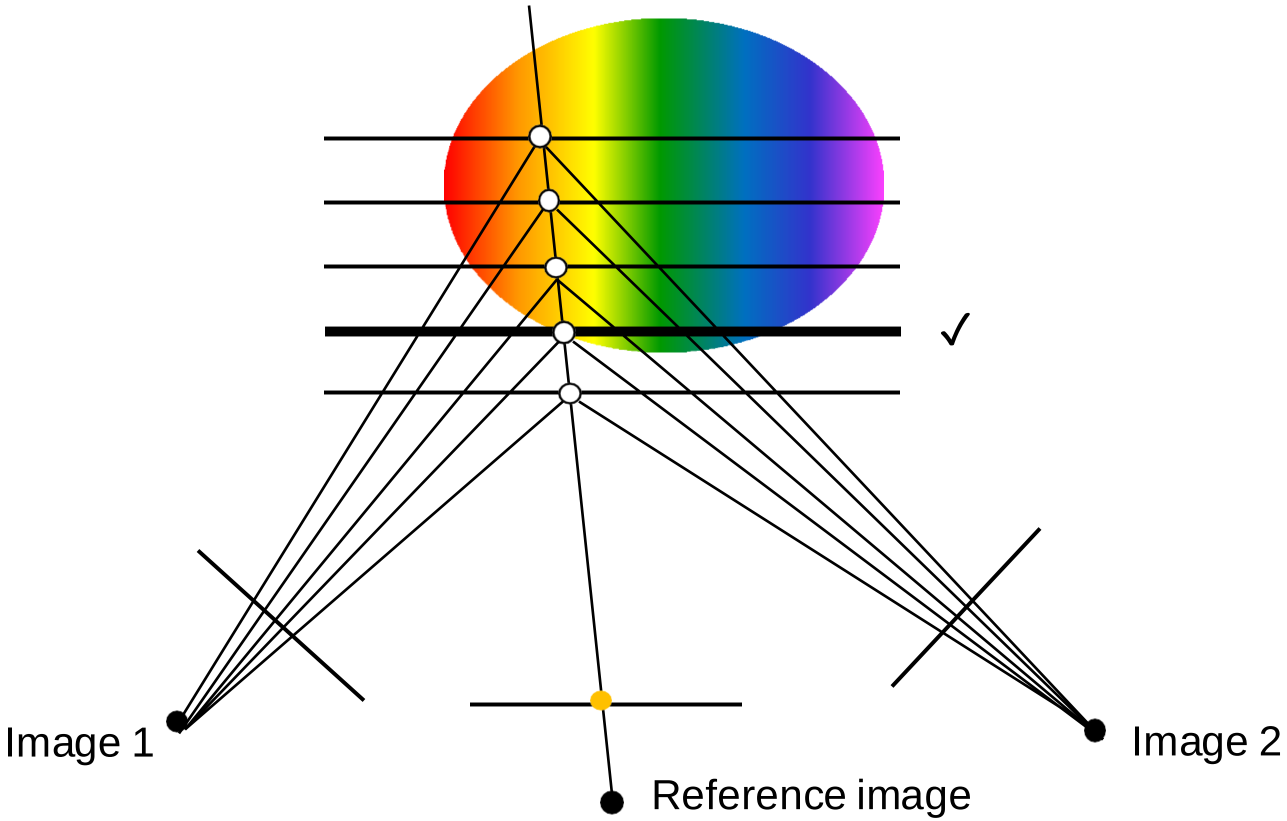
Plane sweep stereo: Key Idea



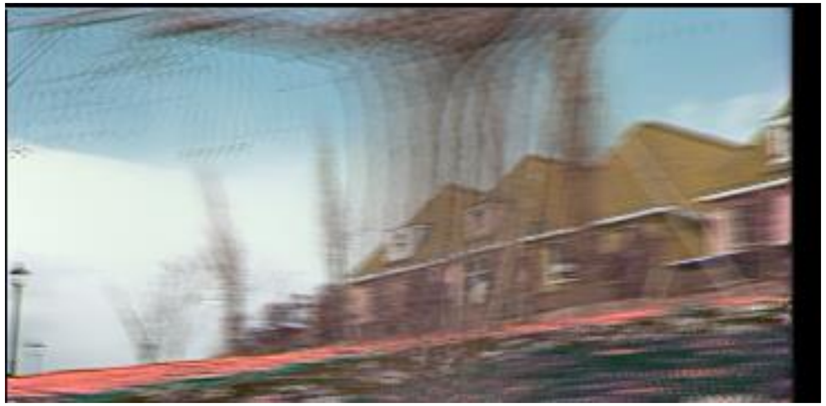
Plane sweep stereo: Key Idea



Plane sweep stereo: Key Idea



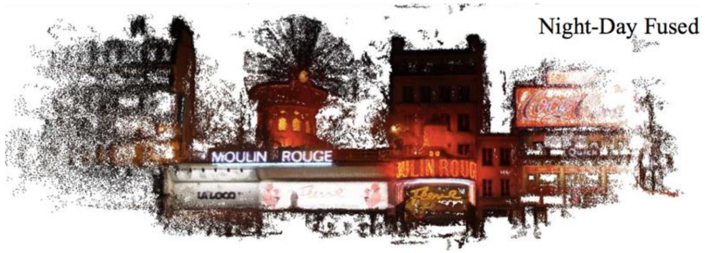
Plane sweep stereo: Fast implementation



- For each depth plane
  - Compute homographies projecting each image onto that depth plane
  - For each pixel in the composite image stack, compute the variance
- For each pixel, select the depth that gives the lowest variance

R. Yang and M. Pollefeys, Multi-Resolution Real-Time Stereo on Commodity Graphics Hardware, CVPR 2003

Ongoing research directions



[Challenging lighting conditions]

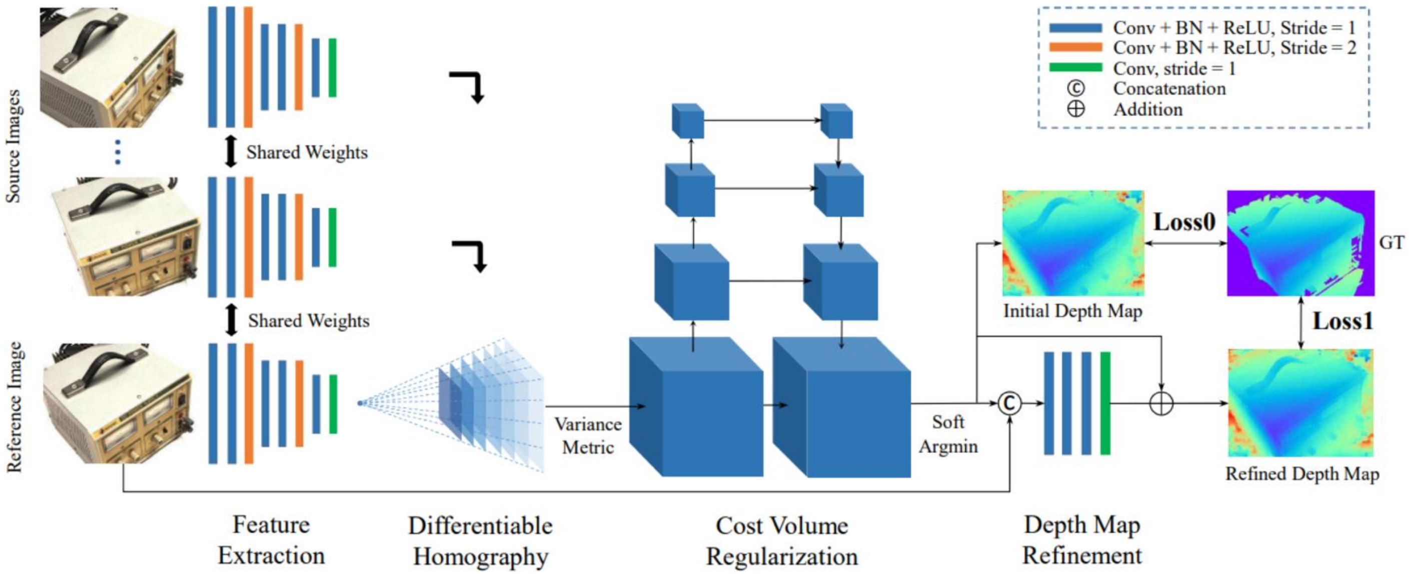
[Ground/aerial]

[Indoor modeling]

[Dynamic reconstruction]

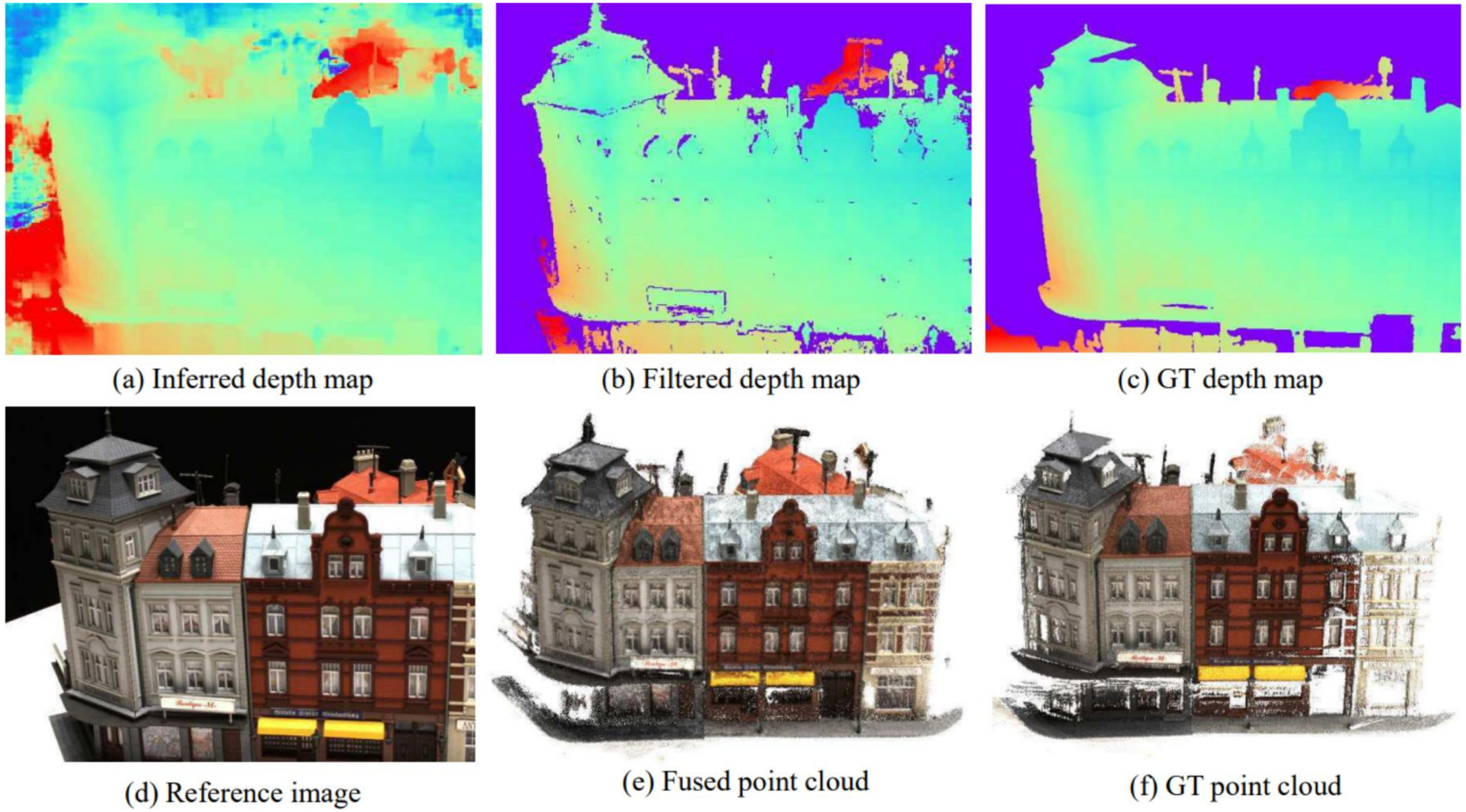
Deep learning for MVS

上海科技大学
ShanghaiTech University



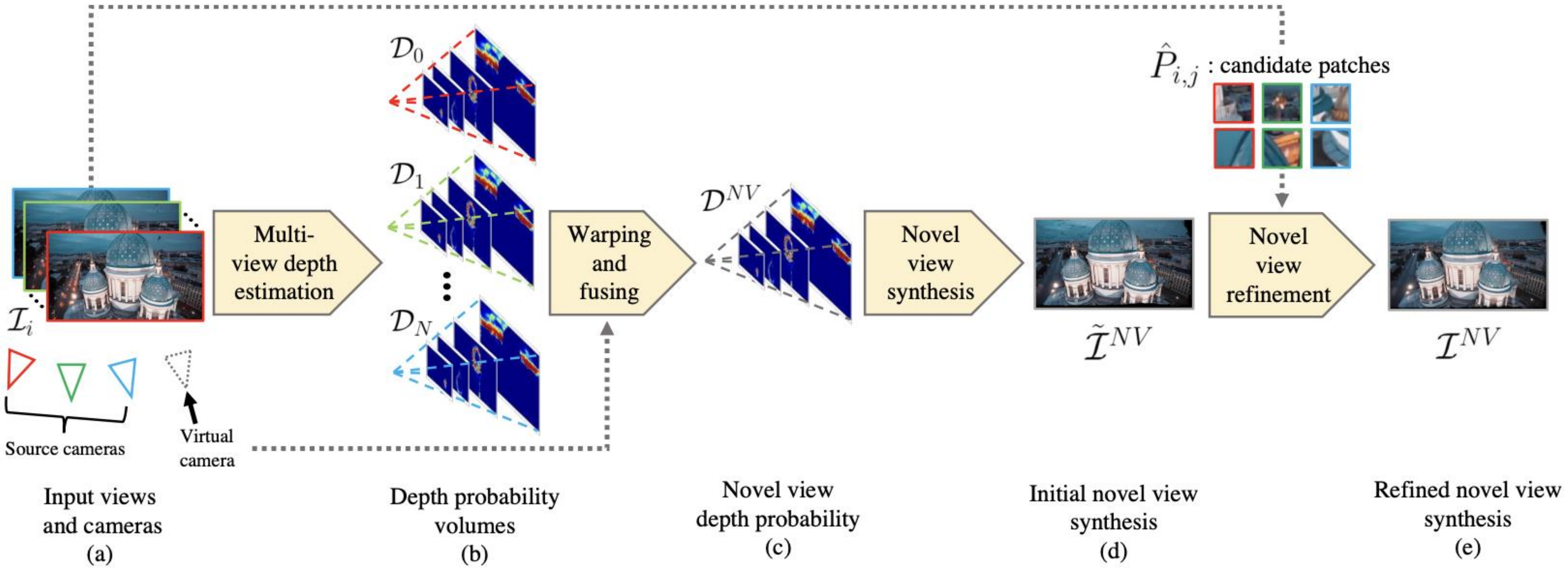
Y. Yao et al. [MVSNet: Depth Inference for Unstructured Multi-view Stereo](). ECCV 2018

Deep learning for MVS



Y. Yao et al. [MVSNet: Depth Inference for Unstructured Multi-view Stereo](). ECCV 2018

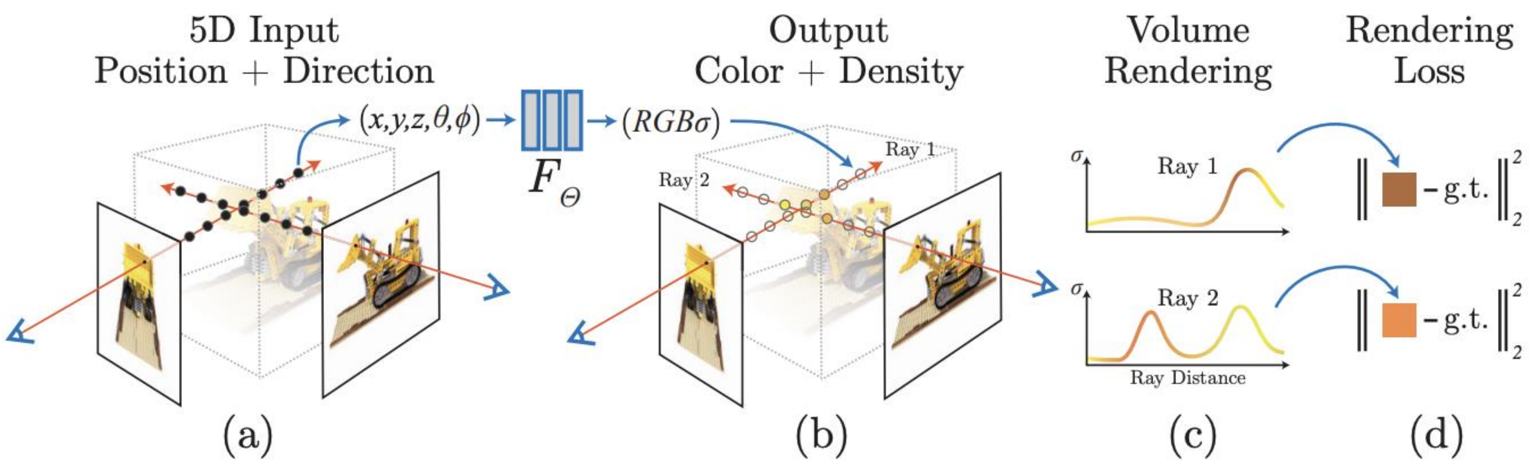
MVS for View Synthesis



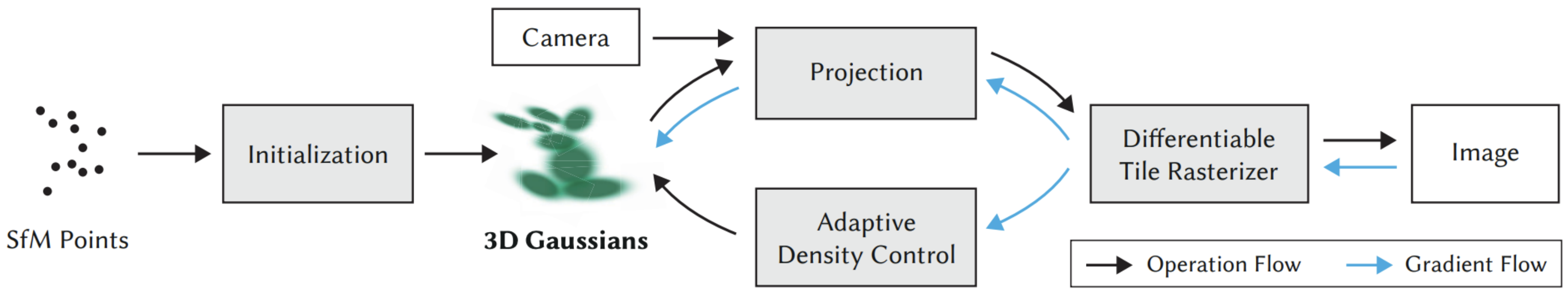
C. Inchang et al.-Extreme View Synthesis. ICCV 2019

NeRF & 3D Gaussian Splatting

上海科技大学
ShanghaiTech University



NeRF



3DGS

Project-5 Requirement (Basics)

上海科技大学
ShanghaiTech University

- Take multi-view pictures in your real life (or use drone-view images in MatrixCity)
  - Run MVS methods for Reconstruction or
  - NeRF/3DGS etc. for Novel View Synthesis
  - Note: Any test images should not appear in training
  - Analyze the performance, especially failure cases