3D Reconstruction/ Depth Sensing

簡韶逸 Shao-Yi Chien

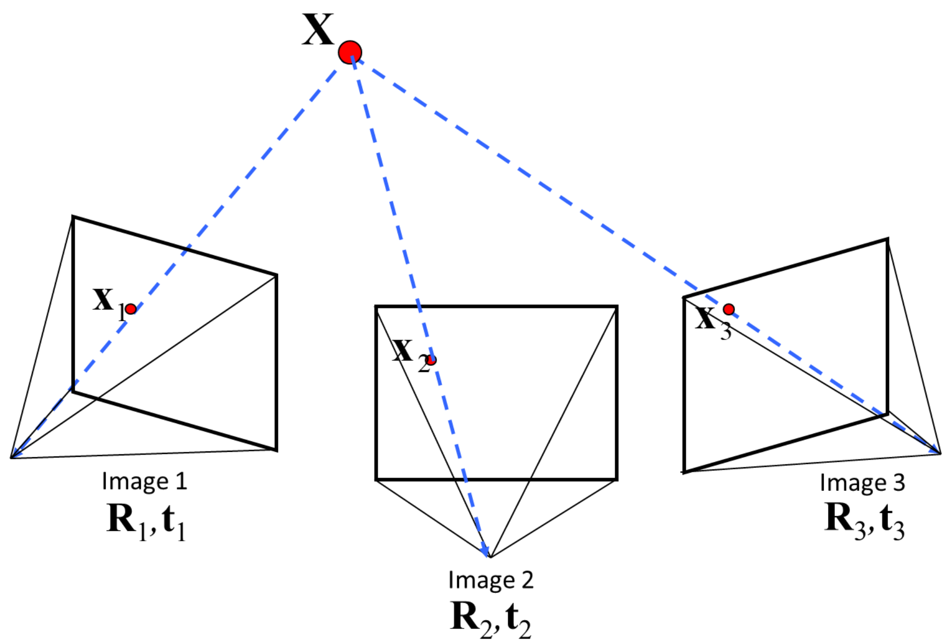
Department of Electrical Engineering

National Taiwan University

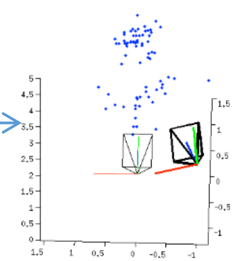
Outline

- Structure from Motion
  - Use slides from SFMedu
  - http://3dvision.princeton.edu/courses/SFMedu/
- Large Scale Reconstruction
- Depth Sensing

Structure from Motion



→ keypoints → match → fundamental matrix → essential matrix → [R|t] → triangulation →
→ keypoints ↗



Large Scale Reconstruction

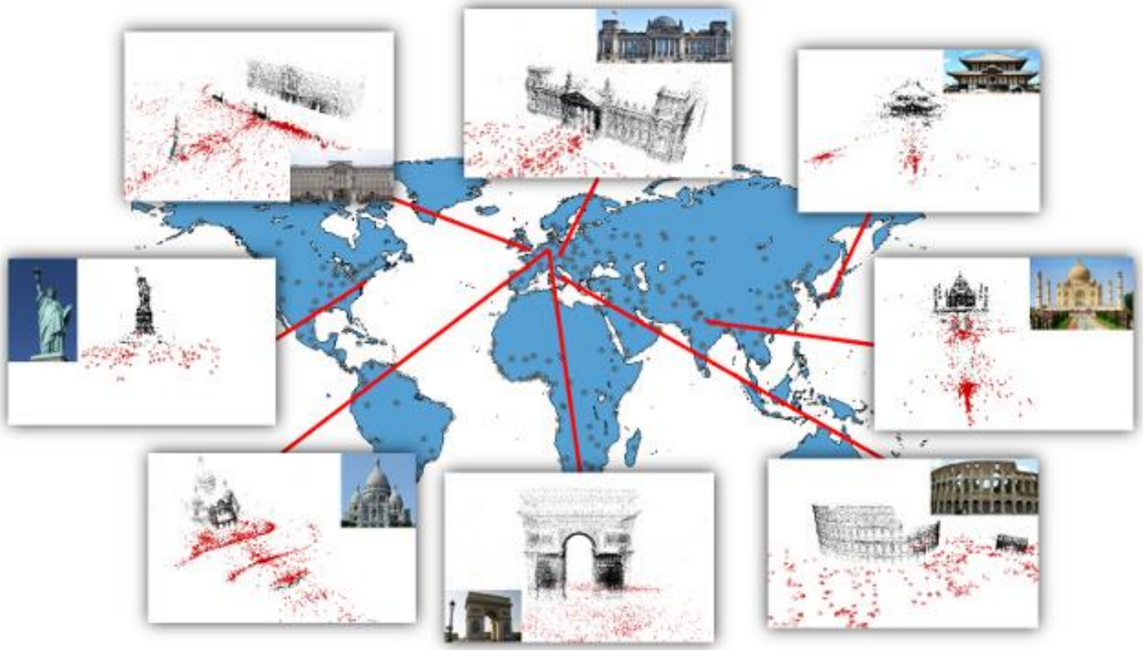
- Building Rome in a Day [ICCV 2009]
  - https://grail.cs.washington.edu/rome/

Large Scale Reconstruction

- Building Rome on a Cloudless Day [ECCV 2010]
  - https://www.youtube.com/watch?v=4cEQZreQ2zQ

Large Scale Reconstruction

- Reconstructing the World* in Six Days [CVPR 2015]
  - As captured by the Yahoo 100 million image dataset
  - http://www.cs.unc.edu/~jheinly/reconstructing_the_world.html
  - https://youtu.be/bRYqyoqUJuM



Depth Sensing with 3D Cameras

Range Acquisition Taxonomy -- Optical Methods

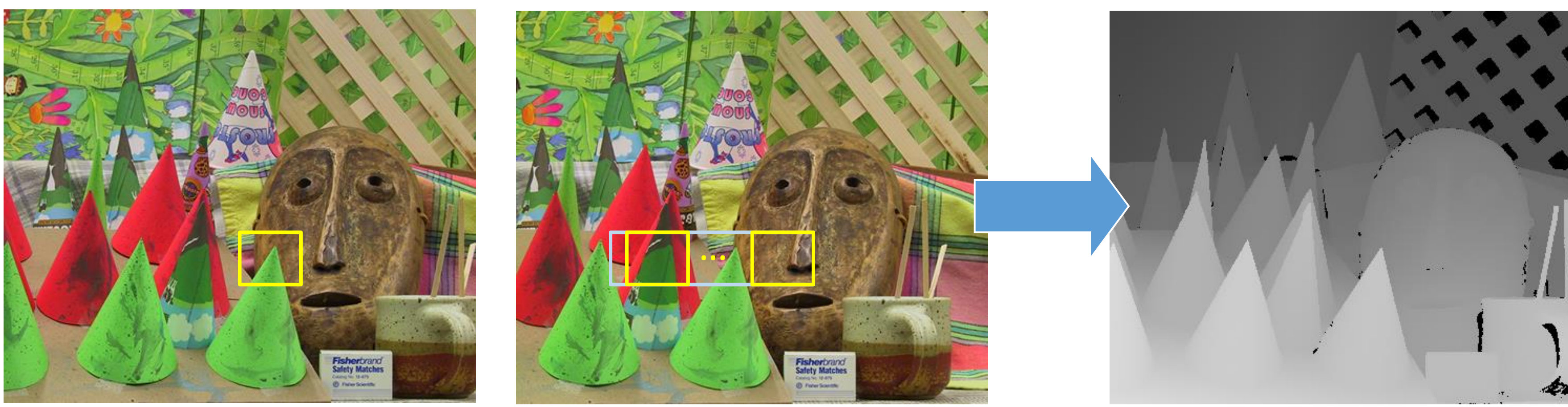
- optical methods
  - 2D-to-3D Conversion
    - depth from X:
      - focus
      - motion
      - color/texture/shading
      - geometry
    - Structure from motion
    - Depth Fusion
  - Active sensor
    - time of flight – Sharp Infra-red sensor
  - Passive sensor
    - Stereo/Multi-view triangulation
    - Light Field Cameras
    - Active+Passive Fusion

Acquisition (Off-the-shelf Products)

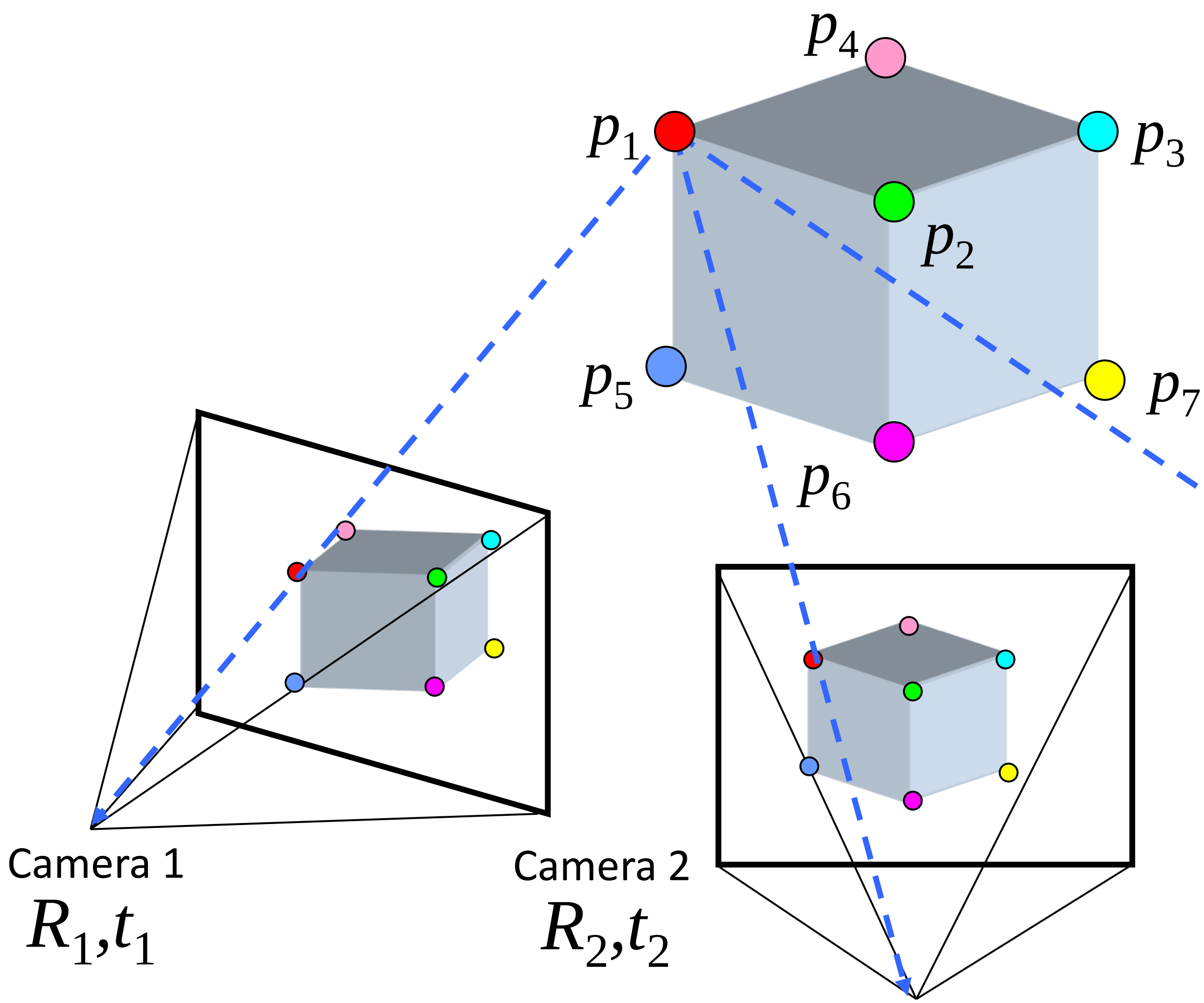
- Shape from stereo (Stereo Vision)
  - Leap Motion
- Structured light (Light coding)
  - Kinect
  - PrimeSense CARMINE 1.08 / 1.09 and Capri 1.25
  - Occipital Structure Sensor
  - Google Project Tango
  - Intel RealSense
  - Apple FaceID
- Time of flight
  - Kinect 2
  - SoftKinetic (acquired by Sony)

Shape from Stereo

- Two (or more) cameras concurrently capture the same scene
  - Find correspondence between stereo images



Shape from Stereo



minimize

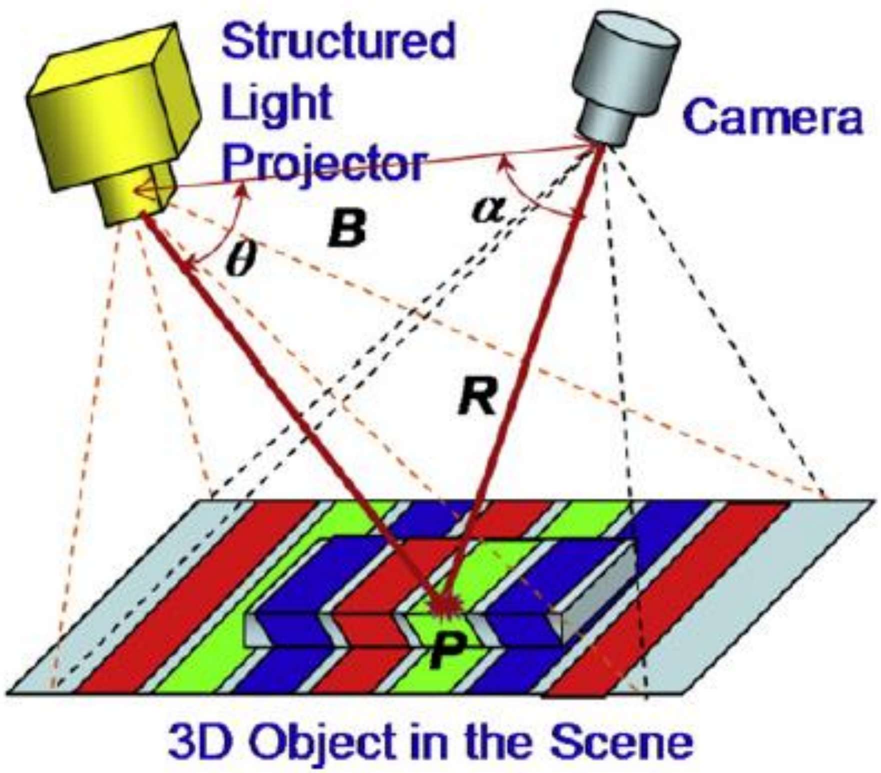
$$\mathrm{f}\,(\mathbf{R},\mathbf{T},\mathbf{P})$$

Shape from Stereo

- Problems
  - The identification of common points within the image pairs, the solution of the well-known correspondence problem
  - The quality depends on the sharpness of the surface texture (affected by variation in surface reflectance)

Structured Light

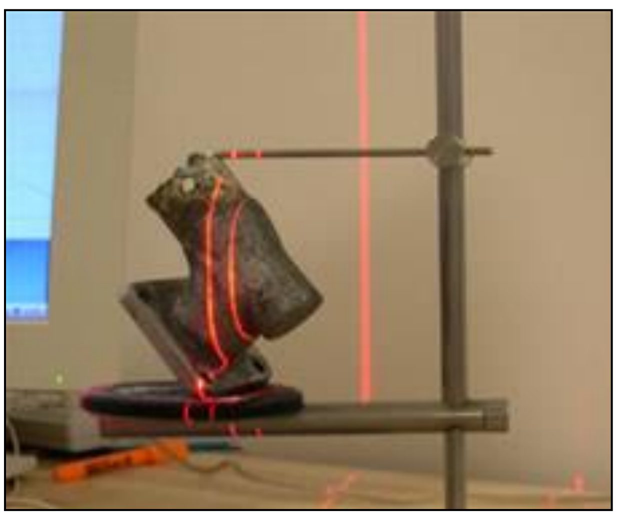
- Tri-angulation principle

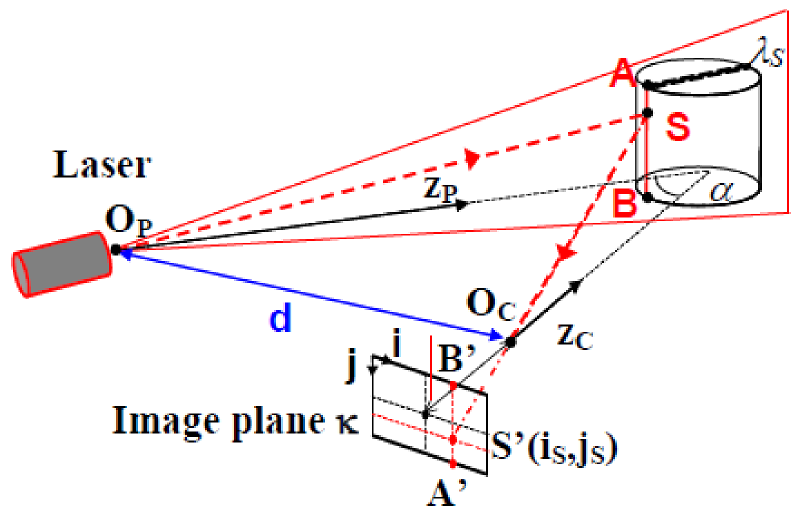


$$R = B\frac{\sin\theta}{\sin(\alpha+\theta)}$$

Structured Light

- Two types
  - Single-point triangulators
  - Laser stripes
- All based on the active triangulation principle





Structured Light

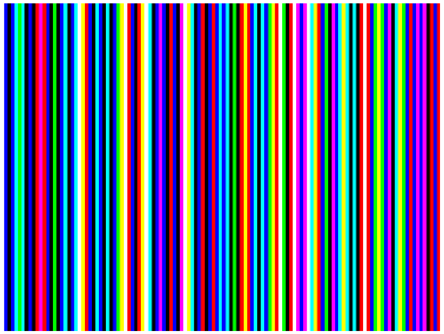
- Active triangulations
    - Active version of shape from stereo
- Project a spatially- and/or temporally-encoded image sequence using projector

Douglas Lanman and Gabriel Taubin, "Build your own 3D scanner: Optical triangulation for beginners," Siggraph 2009 and Siggraph Asia 2009 courses. http://mesh.brown.edu/byo3d/index.html

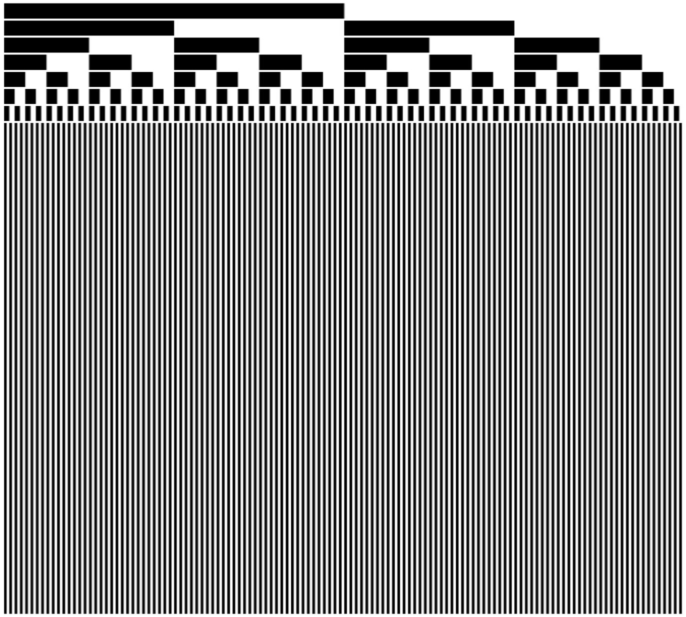
Structured Light

- Examples of projected patterns

Single-shot patterns (N-arrays, grids, random, etc.)



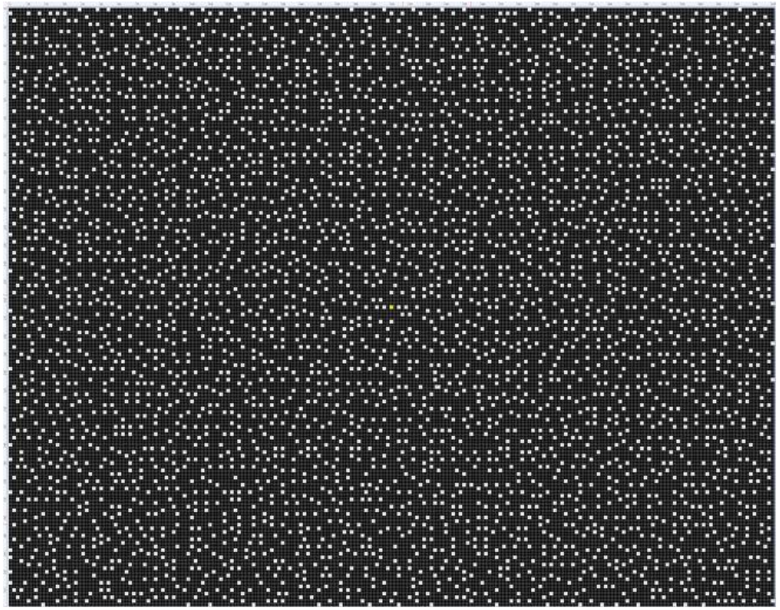
De Bruijn Sequence

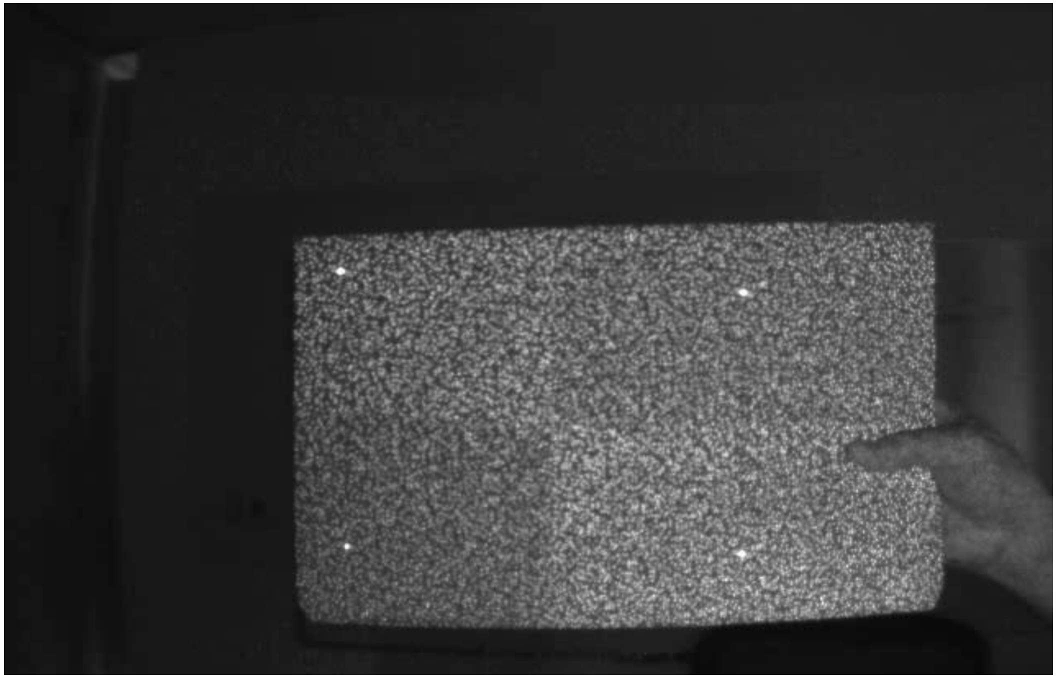


Binary Codes

Structured Light

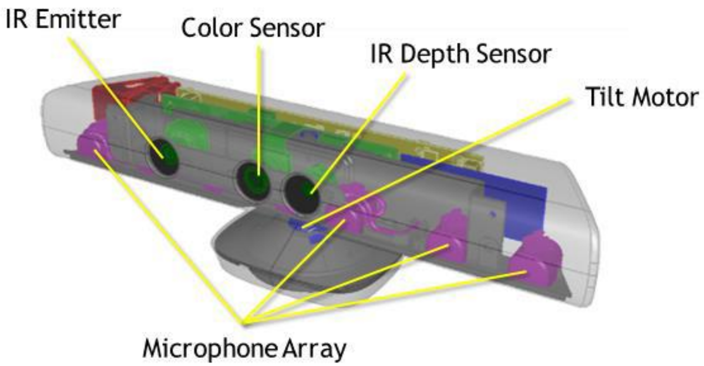
- Kinect projected pattern
  - https://www.youtube.com/watch?v=uq9SEJxZiUg





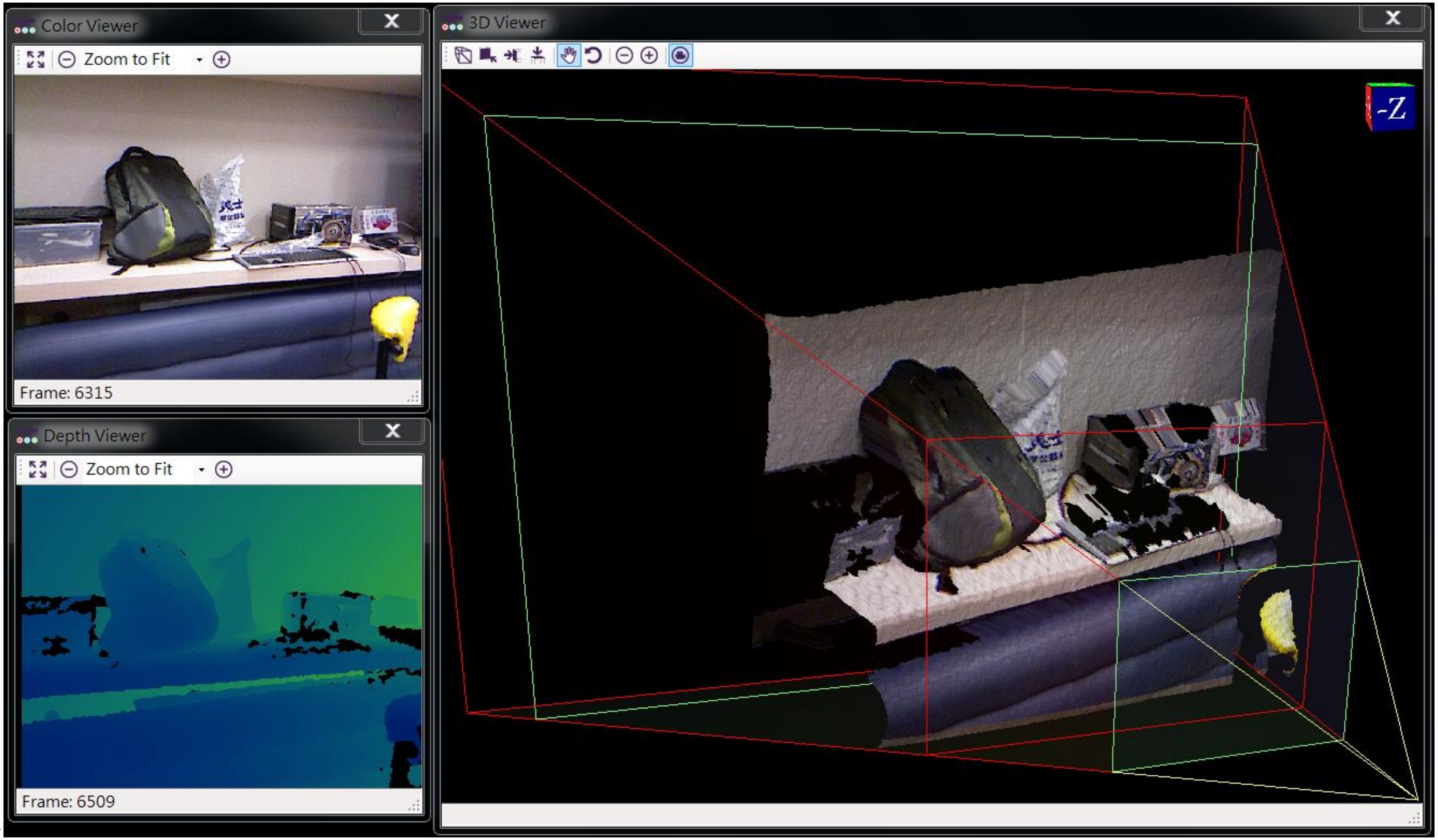
Kinect

- Distance: 0.8-4m
- Near mode: 0.4-3m (Kinect for Windows)
- FOV：57°H、43°V
- RGB: 1280x960@12FPS / 640x480@30FPS / 640x480@15FPS
- Depth: 640x480 / 320x240 / 80x60 (30FPS)



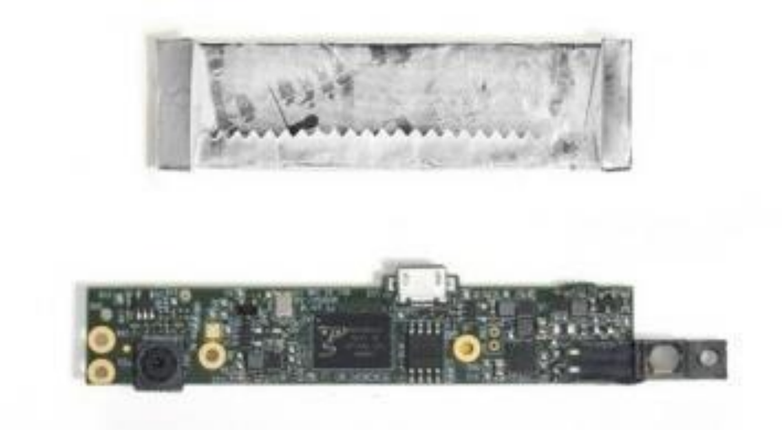
Kinect: RGB + Depth + Point Cloud

- Point cloud from Kinect



Primesense Capri 1.25

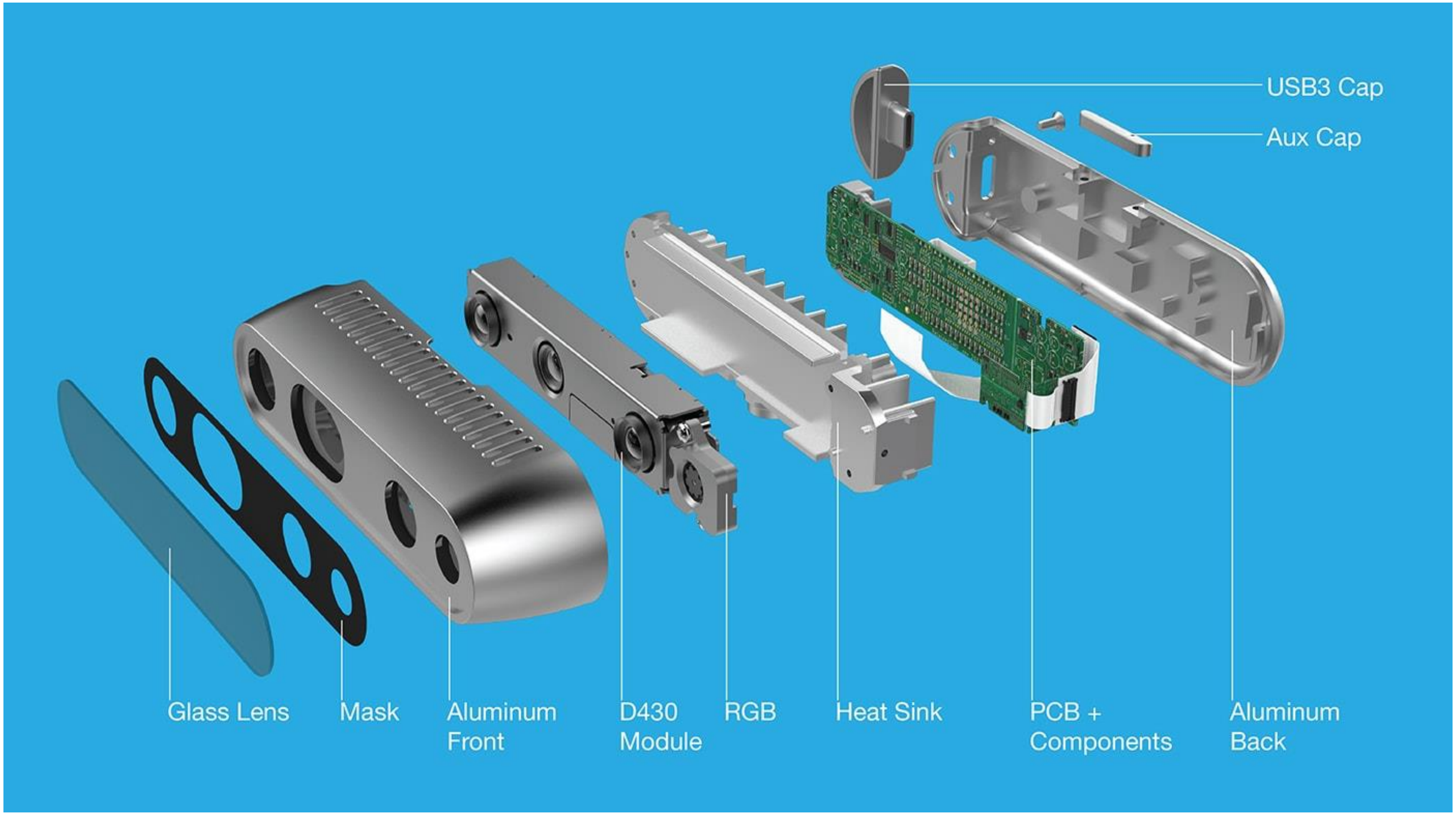
- The smallest 3D sensor right now
- For embedded system



Google Project Tango

- https://www.google.com/atap/projecttango/
  - https://www.youtube.com/watch?v=Qe10ExwzCqk
- PrimeSense PSX1200 Capri PS1200 3D sensor SoC
- With InvenSense MPU-9150 motion tracking device
- Depth: 320×180@5FPS (?)
- RGB: a 4MP rear-facing RGB/IR camera, a 180° field of view rear-facing fisheye camera

Intel RealSense

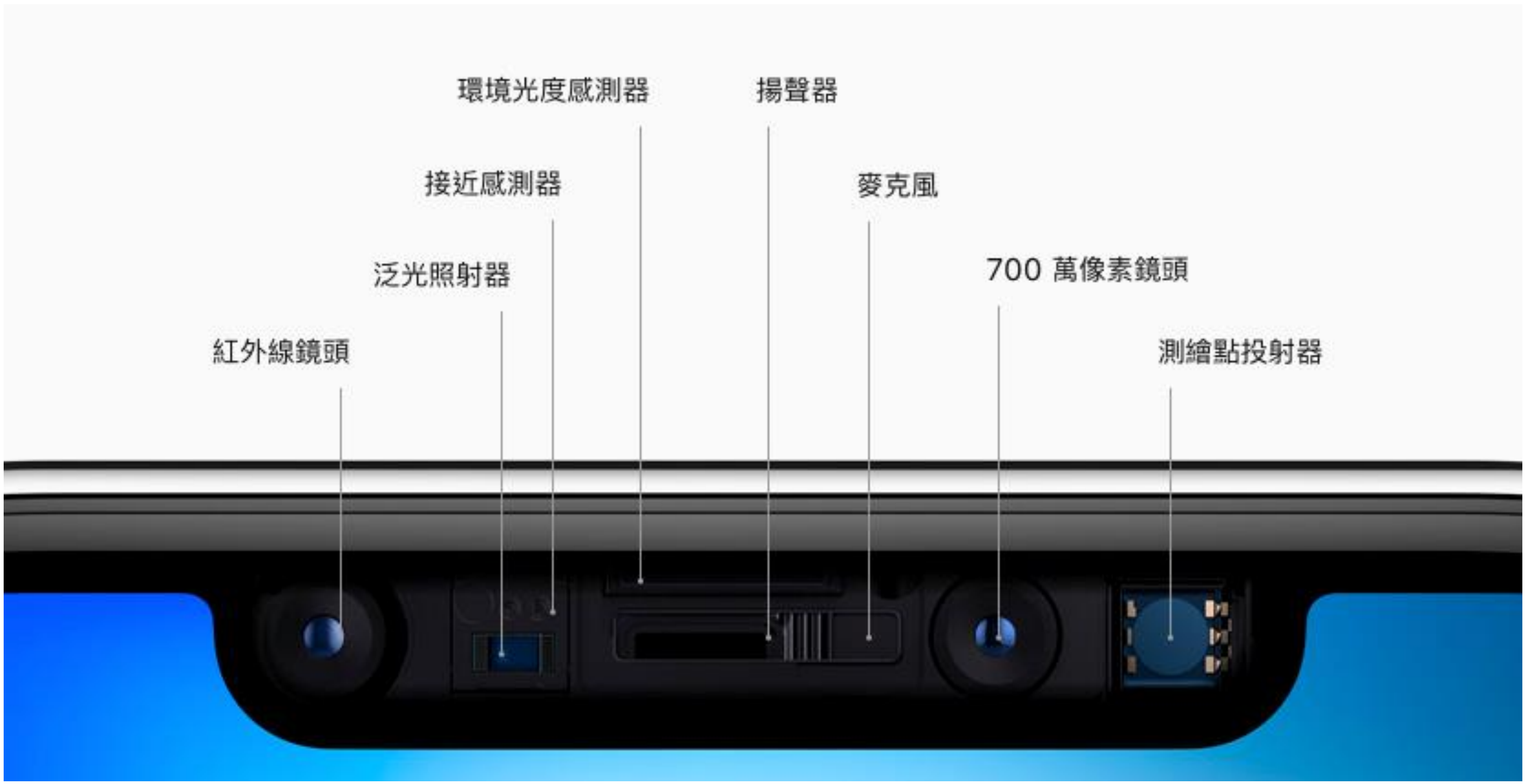


[Intel]

Intel RealSense: D435i

- Intel RealSense Module D430 + RGB CameraVision
- DepthDepth Technology: Active IR Stereo
- Minimum Depth Distance (Min-Z): 0.105 m
- Depth Output Resolution & Frame Rate: Up to 1280 x 720 active stereo depth resolution. Up to 90 fps.
- RGB Sensor Resolution: 1920 x 1080
- Processor: Intel RealSense Vision Processor D4

Apple FaceID

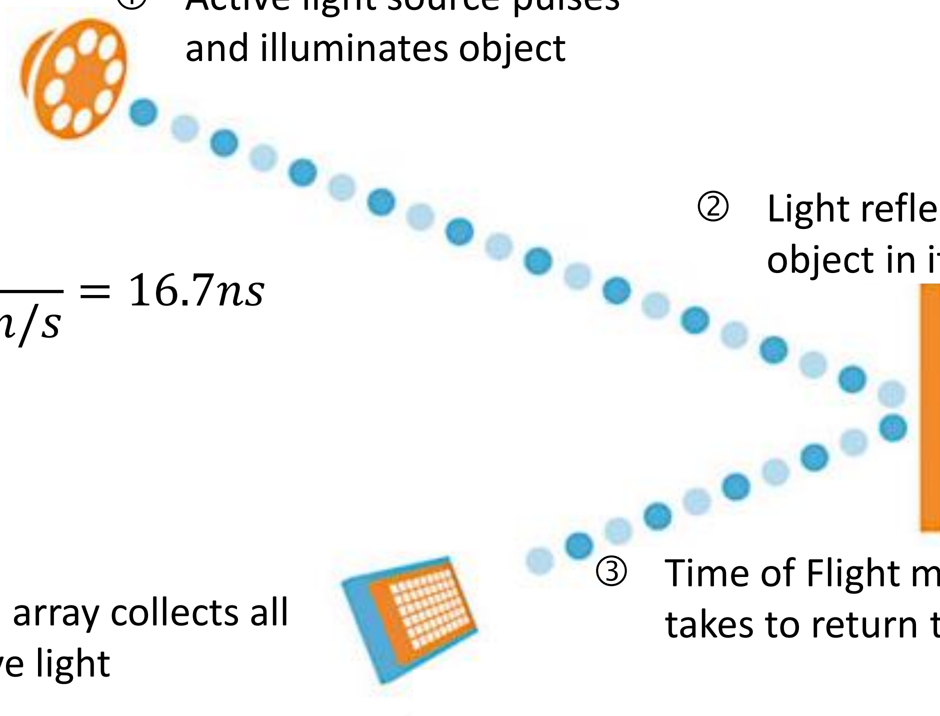


[Apple]

Apple FaceID

Time of Flight

① Active light source pulses and illuminates object



② Light reflects off of any object in its Field of View

$$t = \frac{2d}{c} = \frac{5.0m}{3 \times 10^8 m/s} = 16.7ns$$

③ Time of Flight measures the time it takes to return to the receiver

④ Pixel array collects all active light

Time of Flight

- The emitter unit generates a laser pulse
  - A receiver detects the reflected pulse, and suitable electronics measures the roundtrip travel time of the returning signal and its intensity
- The measurement resolutions vary with the range
  - Large measuring range, it gives excellent results
  - Not suitable for small objects
    - Requires very high speed timing circuitry

Kinect 2

- https://www.youtube.com/watch?v=Hi5kMNfgDS4
  - 3 times the fidelity over Kinect
  - 3DV Systems & Canesta (bought by MS in 2009)
  - Closer IR sensor and illuminator: Less shadow in depth image
- Distance: 0.5-4.5m
- FOV：70°H、60°V
- RGB: 1920x1080@30FPS
- Depth: 512x424@30FPS

SIGGRAPH Talks 2011

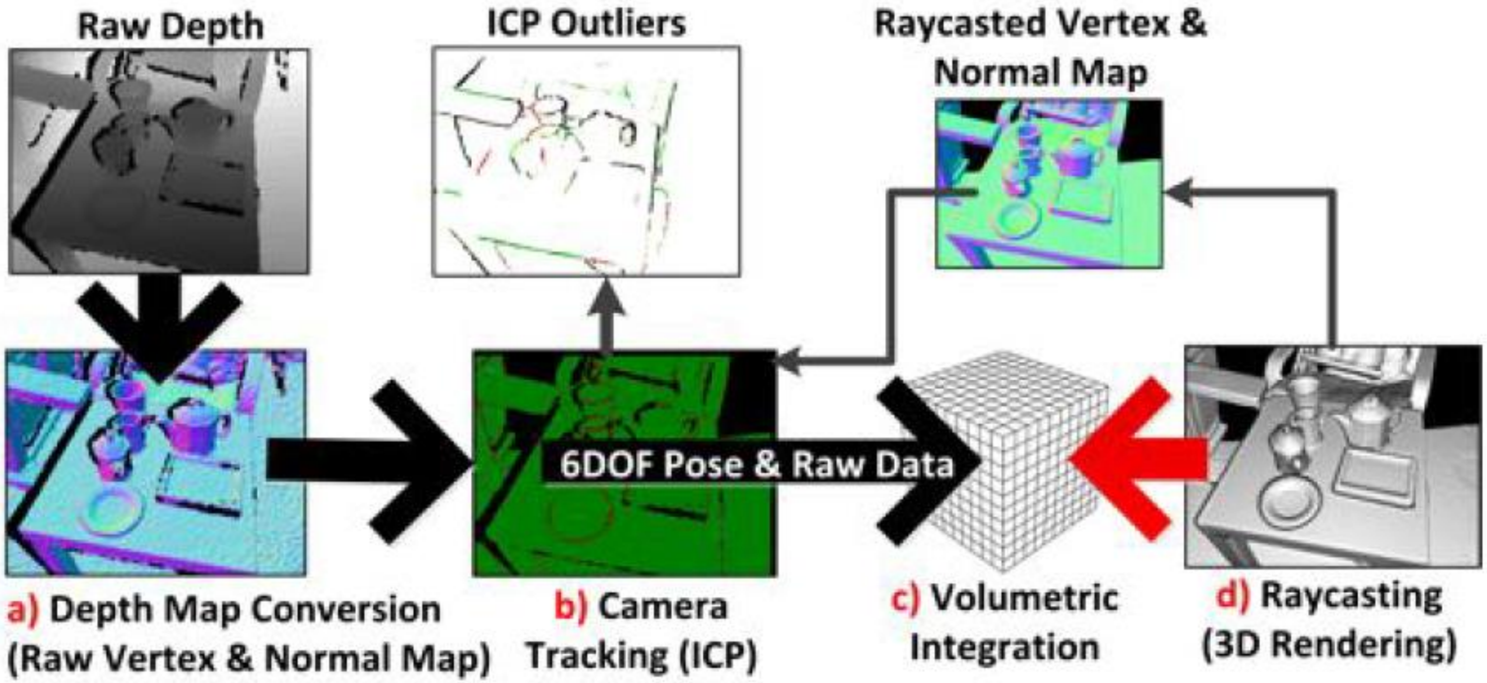
KinectFusion:

Real-Time Dynamic 3D Surface Reconstruction and Interaction

Shahram Izadi 1, Richard Newcombe 2, David Kim 1,3, Otmar Hilliges 1, David Molyneaux 1,4, Pushmeet Kohli 1, Jamie Shotton 1, Steve Hodges 1, Dustin Freeman 5, Andrew Davison 2, Andrew Fitzgibbon 1

1 Microsoft Research Cambridge 2 Imperial College London
3 Newcastle University 4 Lancaster University
5 University of Toronto

KinectFusion



Ref: Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon, "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera," in *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*.