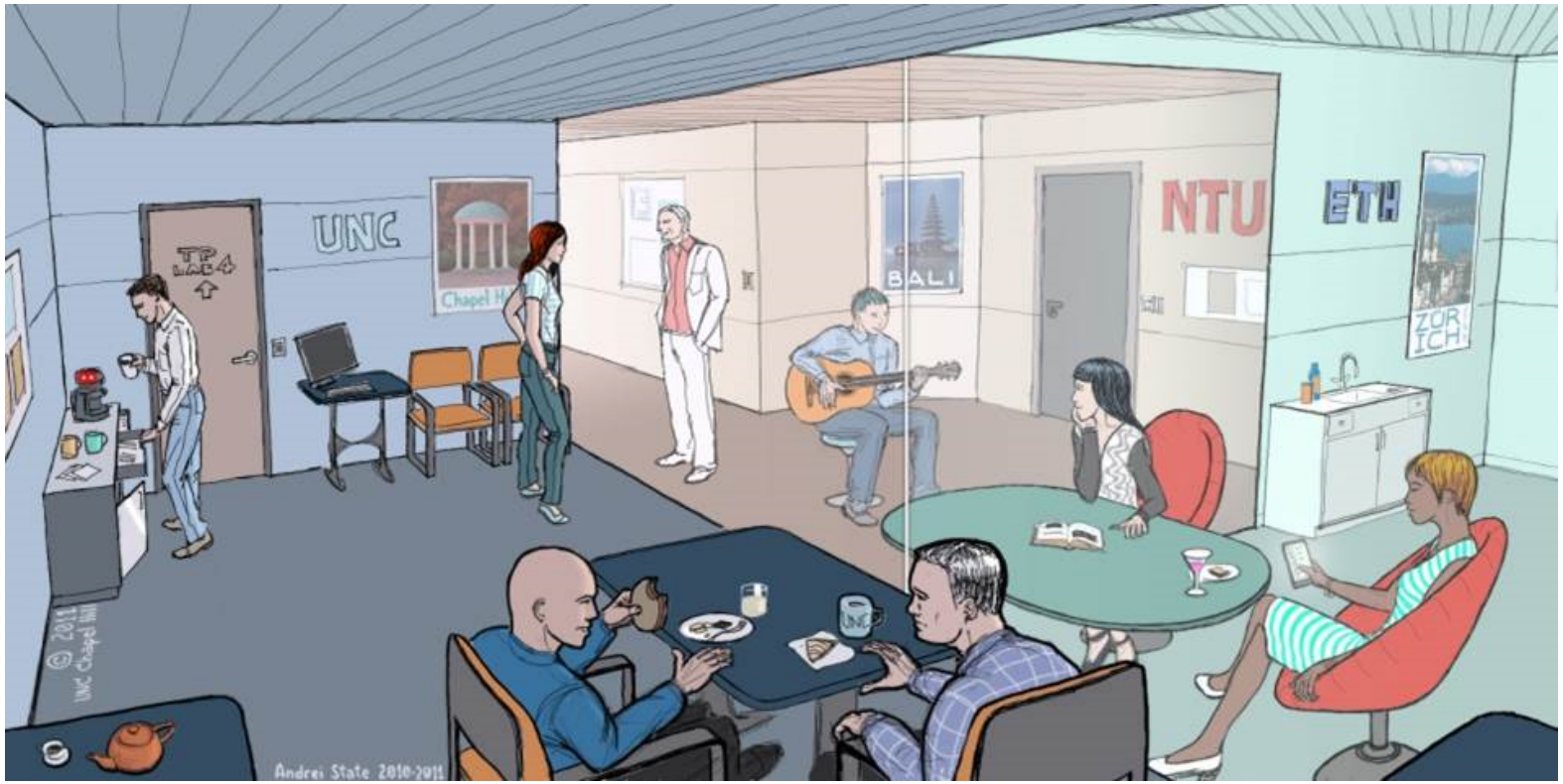
A Brief Introduction to 3D Capture Technology

Mingsong Dou

Google

12/05/2017

My experiences with 3D Capture



Autostereoscopic Display

+

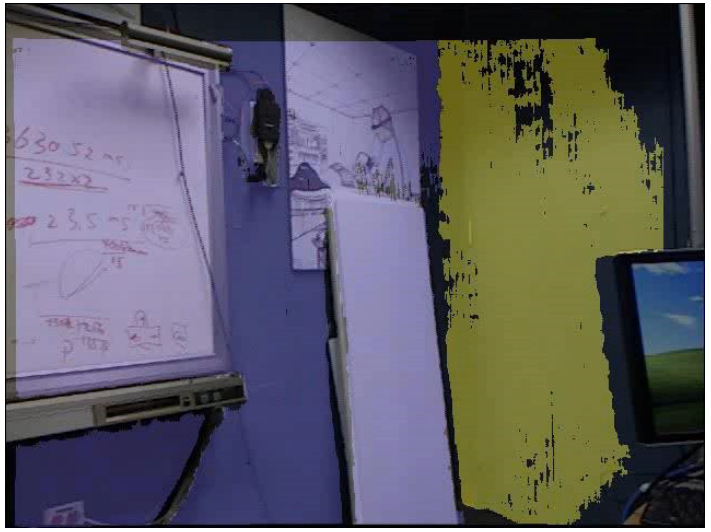
3D Data Acquisition

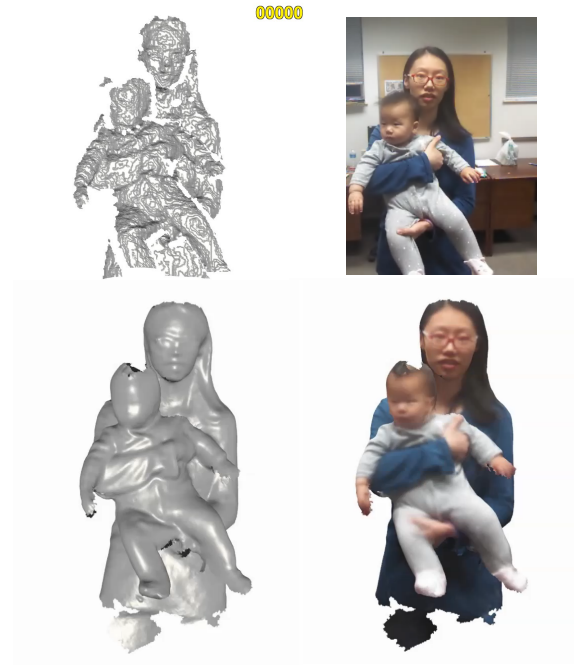
It starts with the Telepresence Project…

My experiences with 3D Capture

Gaze-correction for room-sized telepresence, IEEE VR 2012

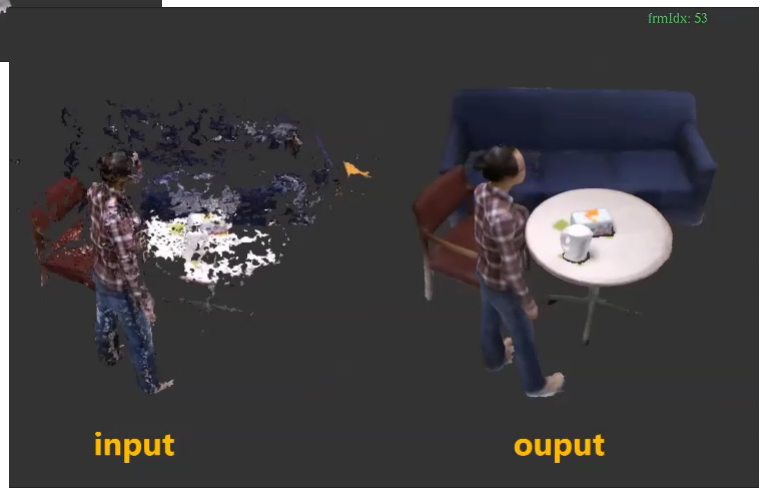
Room scanning with bundle adjustment of points and planes, ACCV 2012





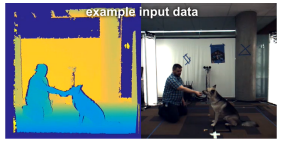
Non-rigid Surface Scanning with dense nonrigid bundle adjustment, CVPR 2015

Scanning and Tracking Dynamic Objects, ISMAR 2013



Room-sized Dynamic Scence Capture, IEEE VR 2014

Fusion4D
SIGGRAPH'16



real-time multi-view reconstruction

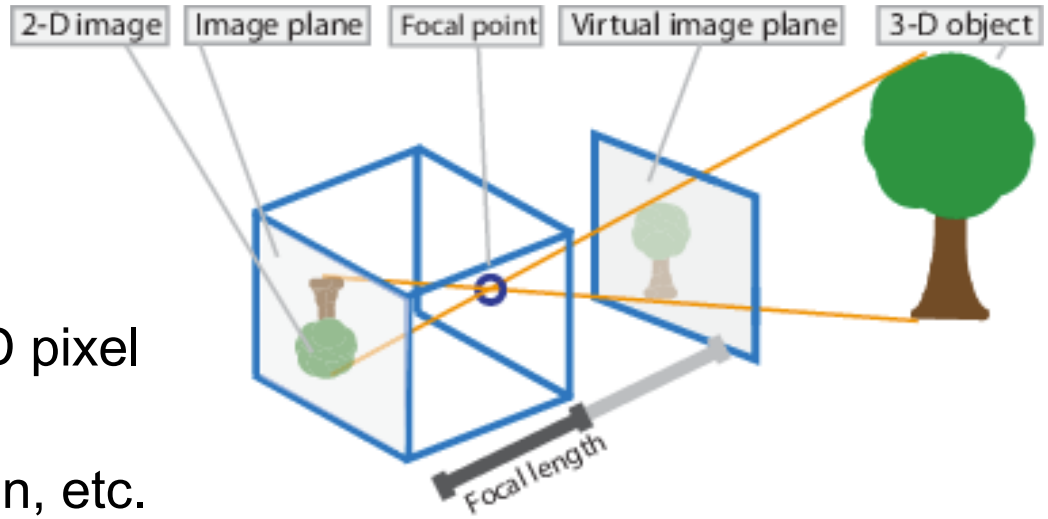
Holoportation
UIST'16

Motion2Fusion
SIGGRAPH Asia, 2017

Outline

- 3D Capture Sensors and Depth Estimation
  - Stereo
  - Structured Light
  - Time-of-Flight
  - Multive-view capture
- World Reconstruction
  - SLAM
  - Kinect Fusion
- People Reconstruction
  - Parametric Tracking
  - Non-Rigid Tracking and Fusion
- Applications in Mixed/Virtual Reality
  - Holoportation

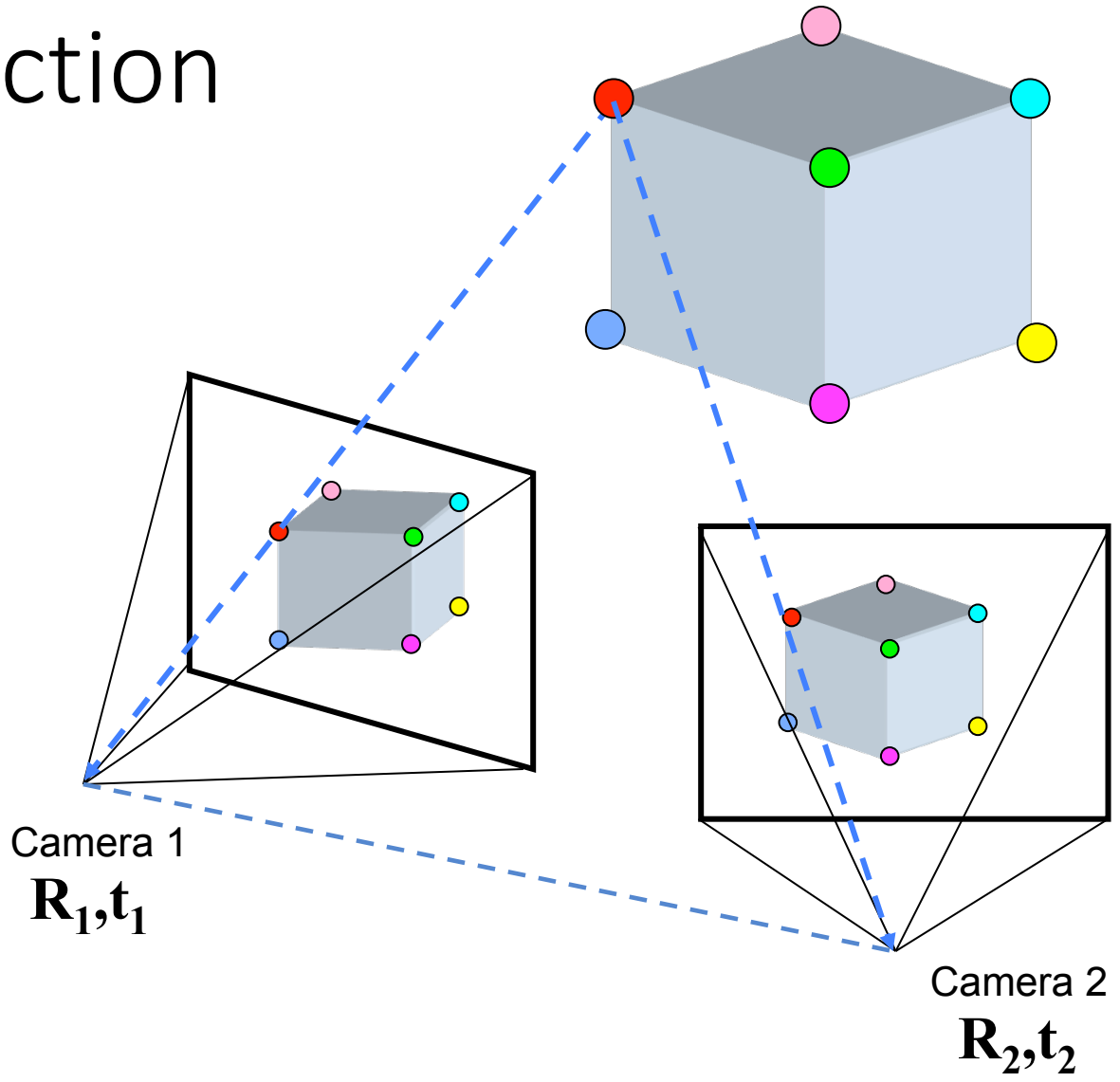
Camera Pinhole Model



- Camera model:
  - Map a 3D point $X=[x;y;z]$ in the world to a 2D pixel position on the image $x=[u;v]$
  - Intrinsics: focus length, pixel size, lens distortion, etc. Usually represented as a 3x3 matrix ***K***.
  - Extrinsics: camera position and orientation, represented by a rotation matrix ***R*** and translation vector ***t***.

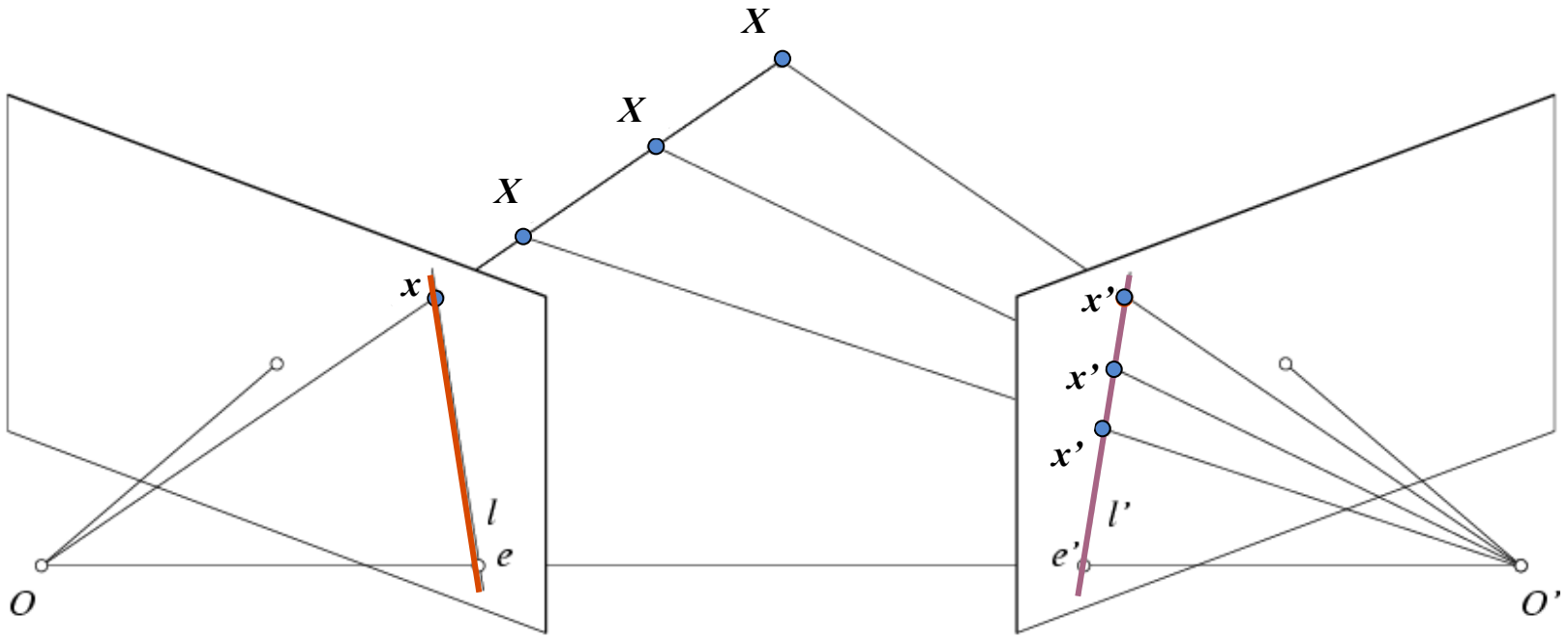
3D Capture/Reconstruction

- Reverse Rendering Problem
  - From 2d image point x to the corresponding 3D world point X
  - Triangulation



Stereo

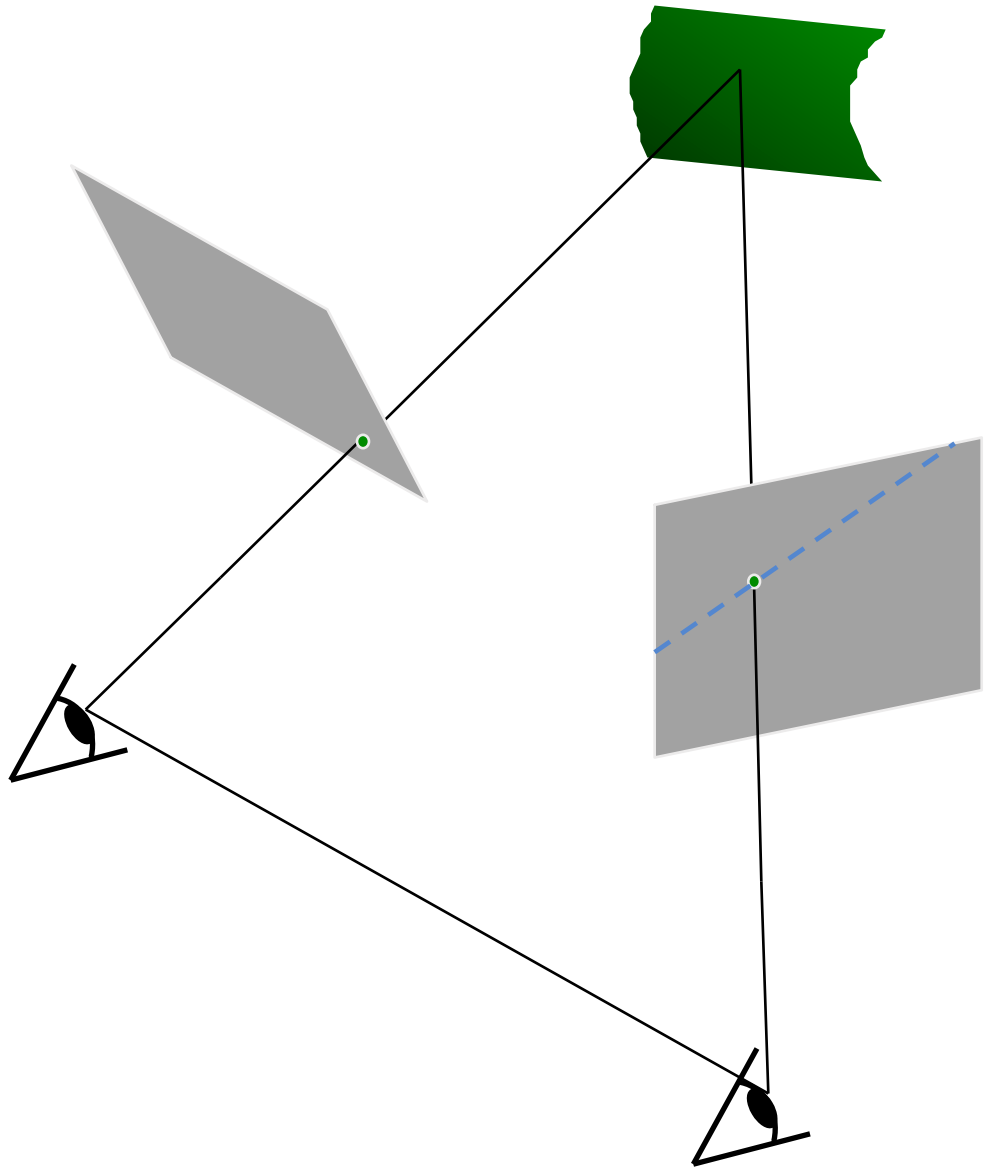
- 1D search problem:
  - search along the epipolar line



Epipolar Constraints: Potential matches for ***x*** have to lie on the corresponding epipolar line ***l'***.

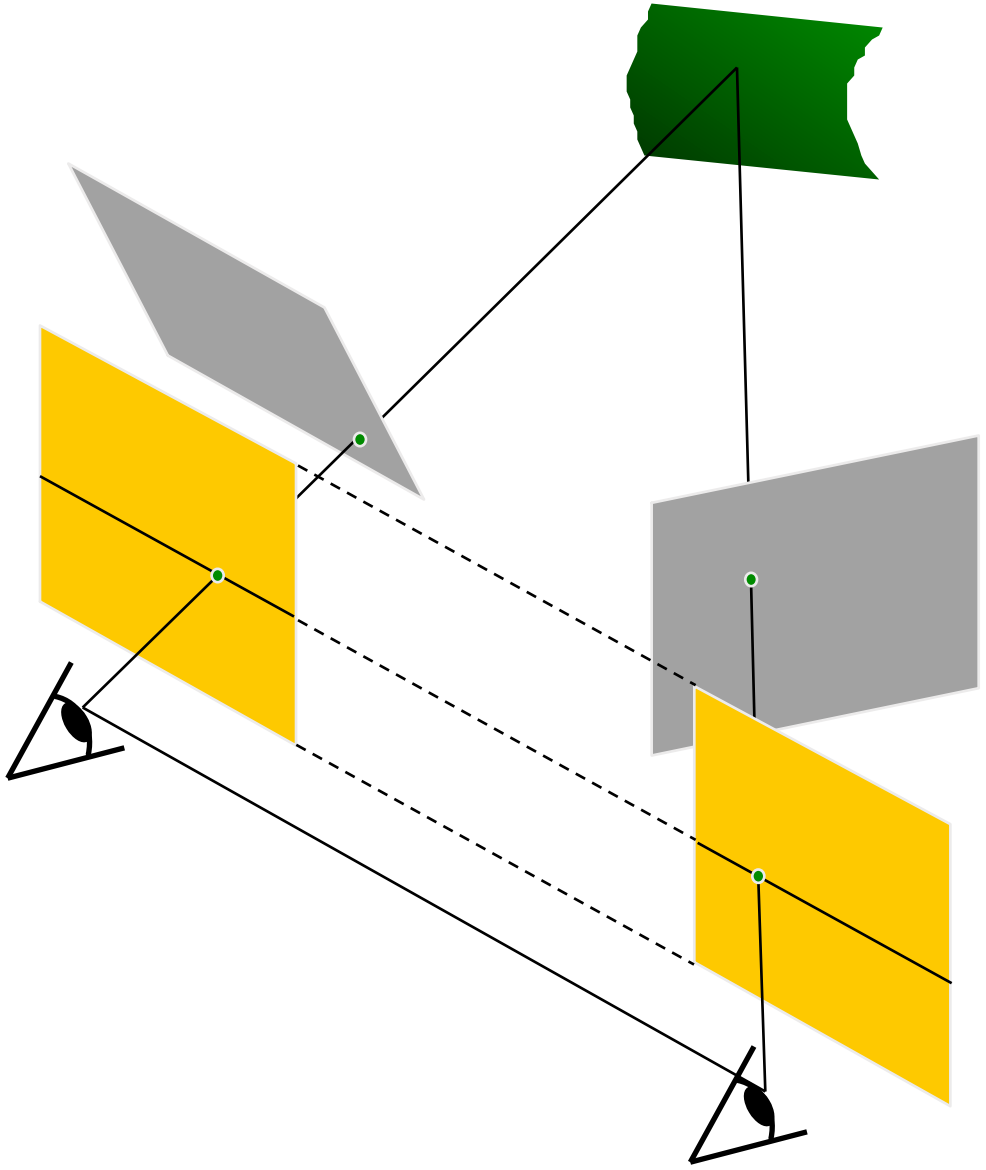
Stereo

- 1D search problem:
  - search along the epipolar line



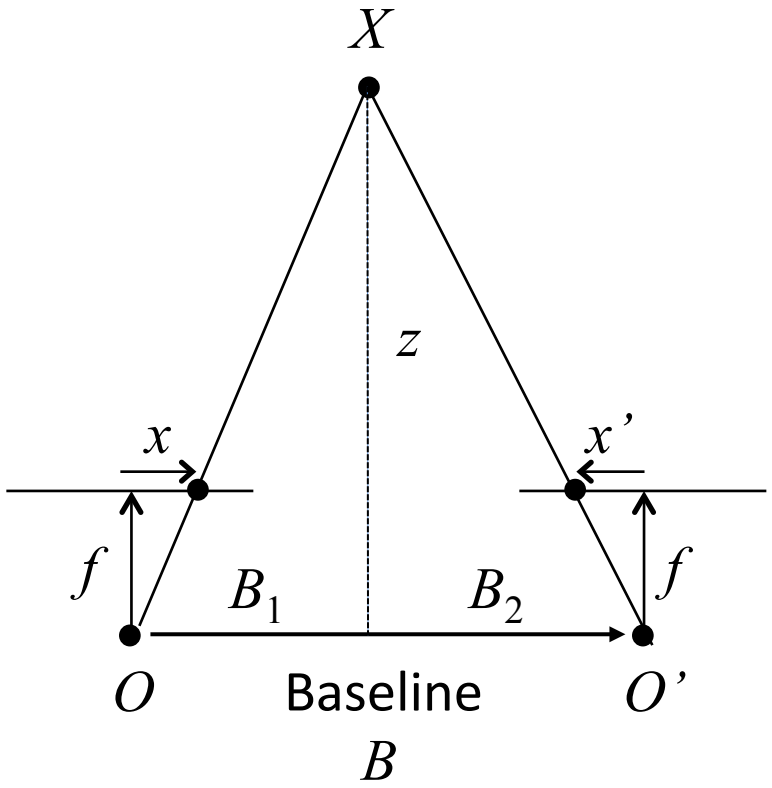
Stereo

- 1D search problem:
  - search along the epipolar line
- A trick to boost the performance
  - Image rectification



Stereo

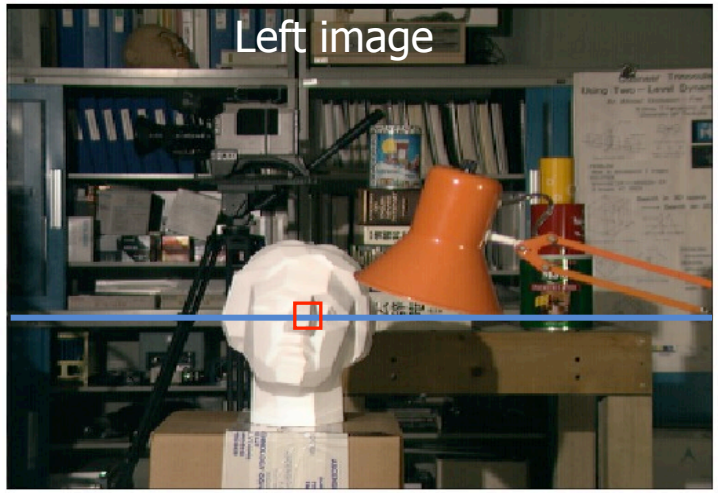
- 1D search problem:
  - search along the epipolar line
- A trick to boost the performance
  - Image rectification
- Disparity map
  - Disparity is inversely proportional to depth

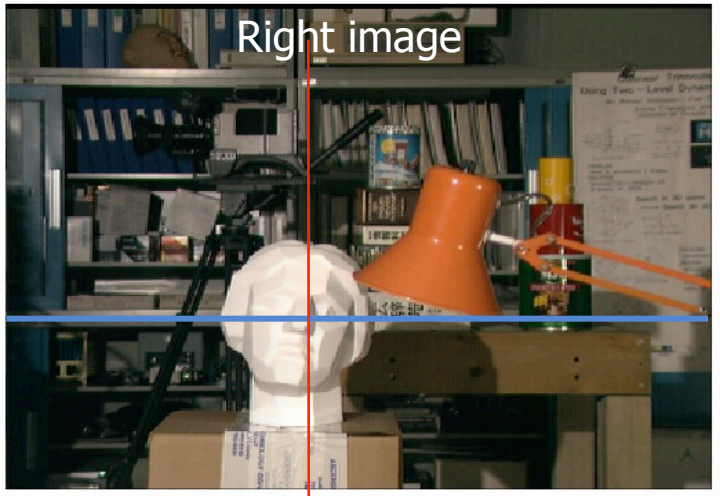


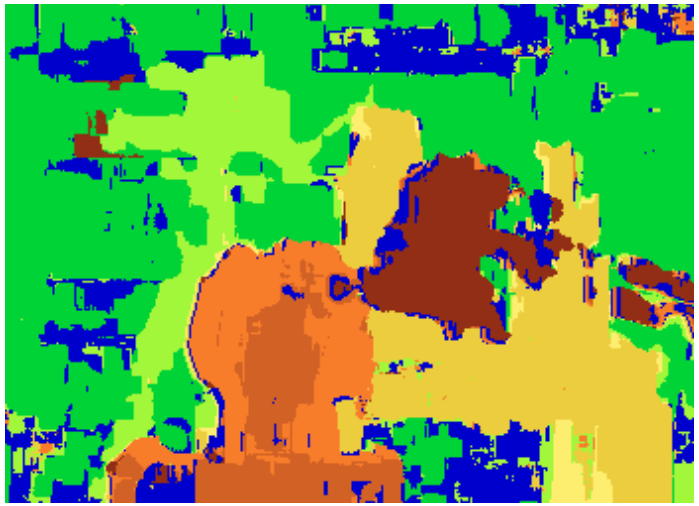
$$disparity = x - x' = \frac{B \cdot f}{z}$$

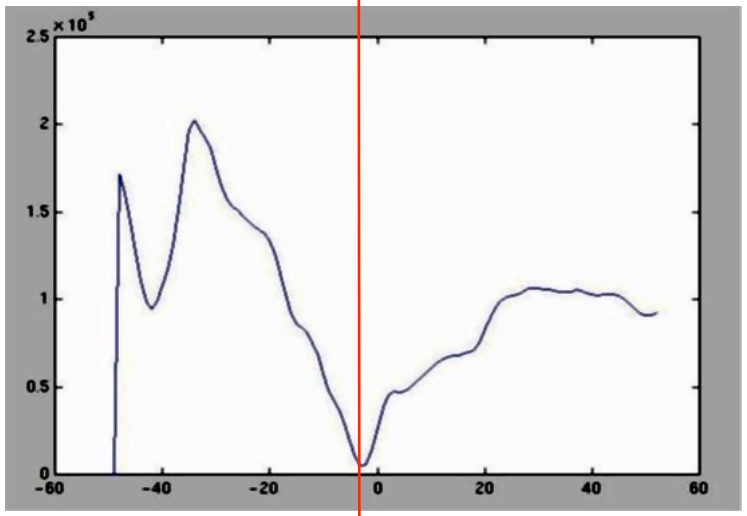
Stereo

- 1D search problem:
  - search along the epipolar line
- A trick to boost the performance
  - Image rectification
- Disparity map
  - Disparity is inversely proportional to depth

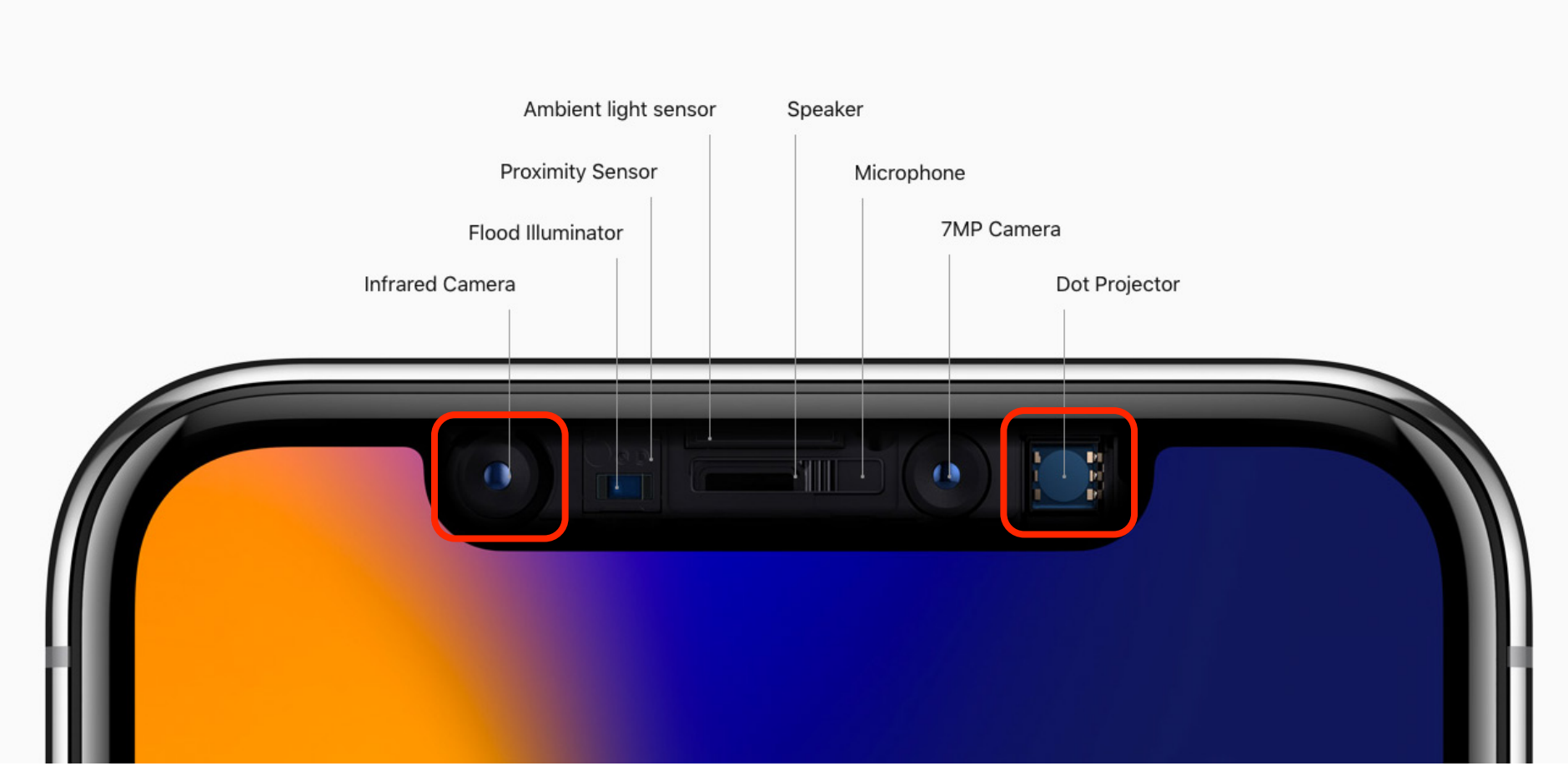




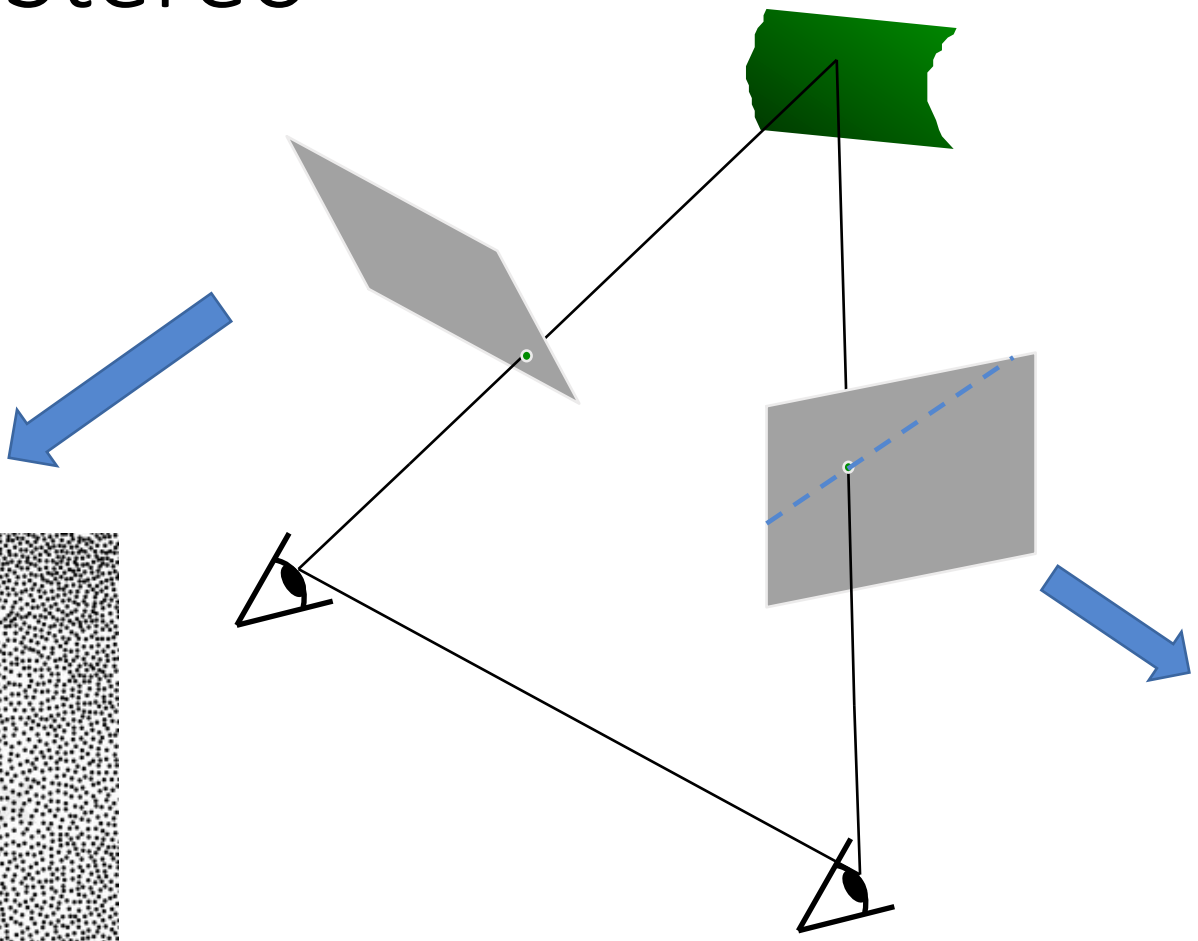


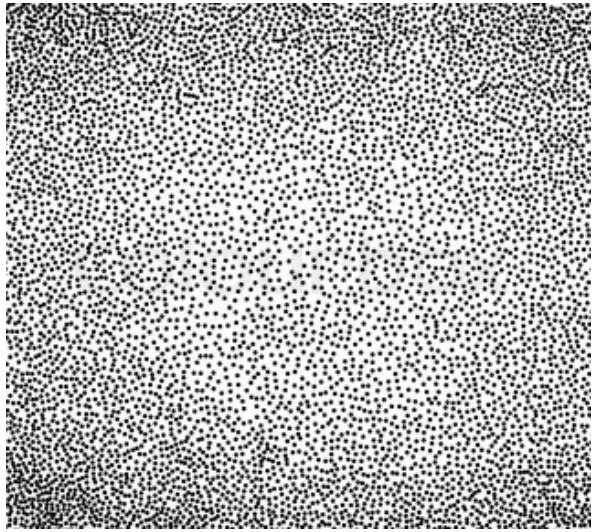


Active Stereo



Active Stereo



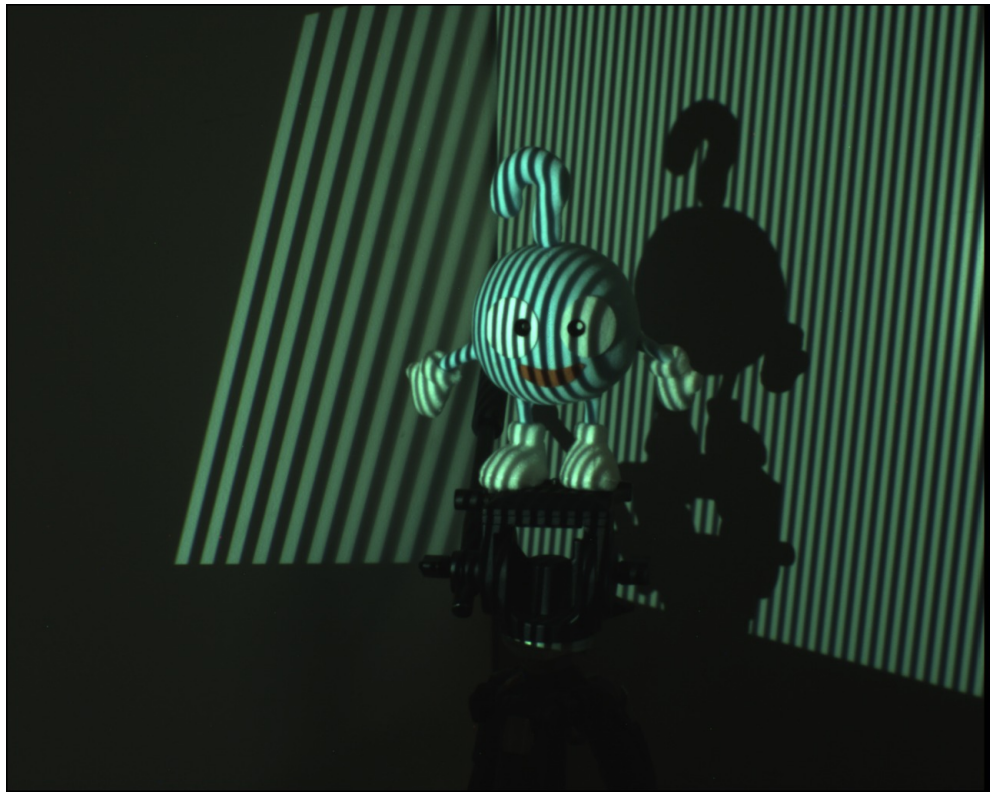


Active Stereo

- Other commercial sensors
  - Microsoft Kinect, Intel RealSense, …
- IR: doesn’t work well at outdoor enviroment

Other 3D capture techniques

- Structured Light



Other 3D capture techniques

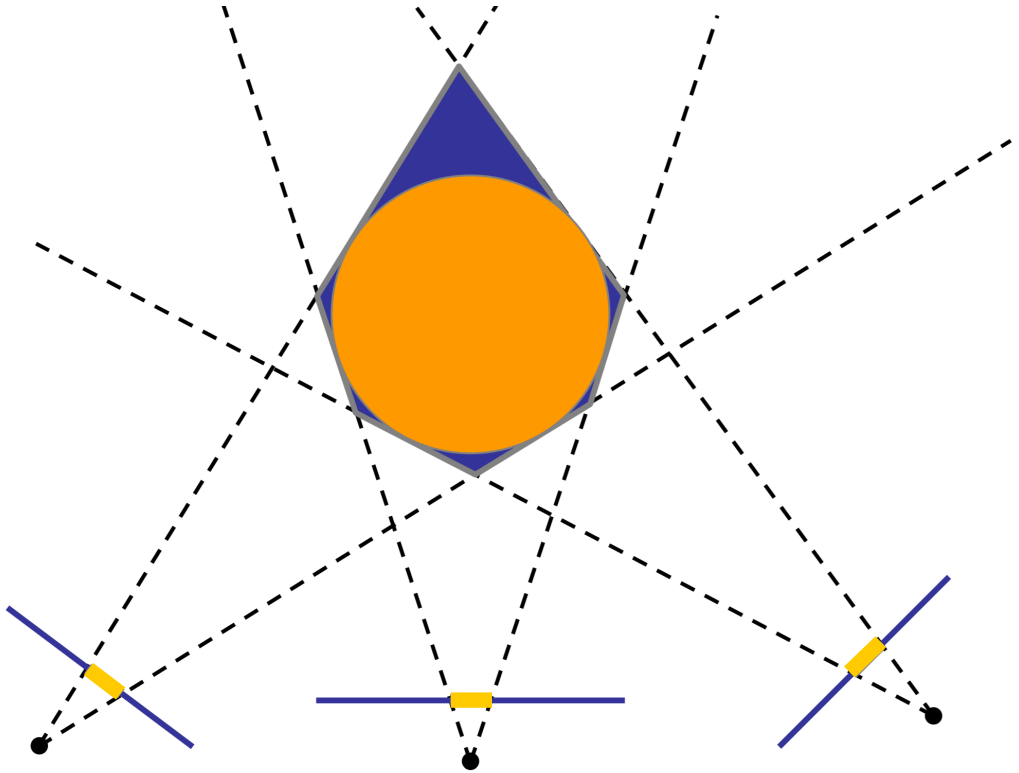
- Structured Light
- Time-of-flight

Microsoft Kinect V2

Other 3D capture techniques

- Structured Light
- Time-of-flight
- LIDAR

Shape-from-silhouette

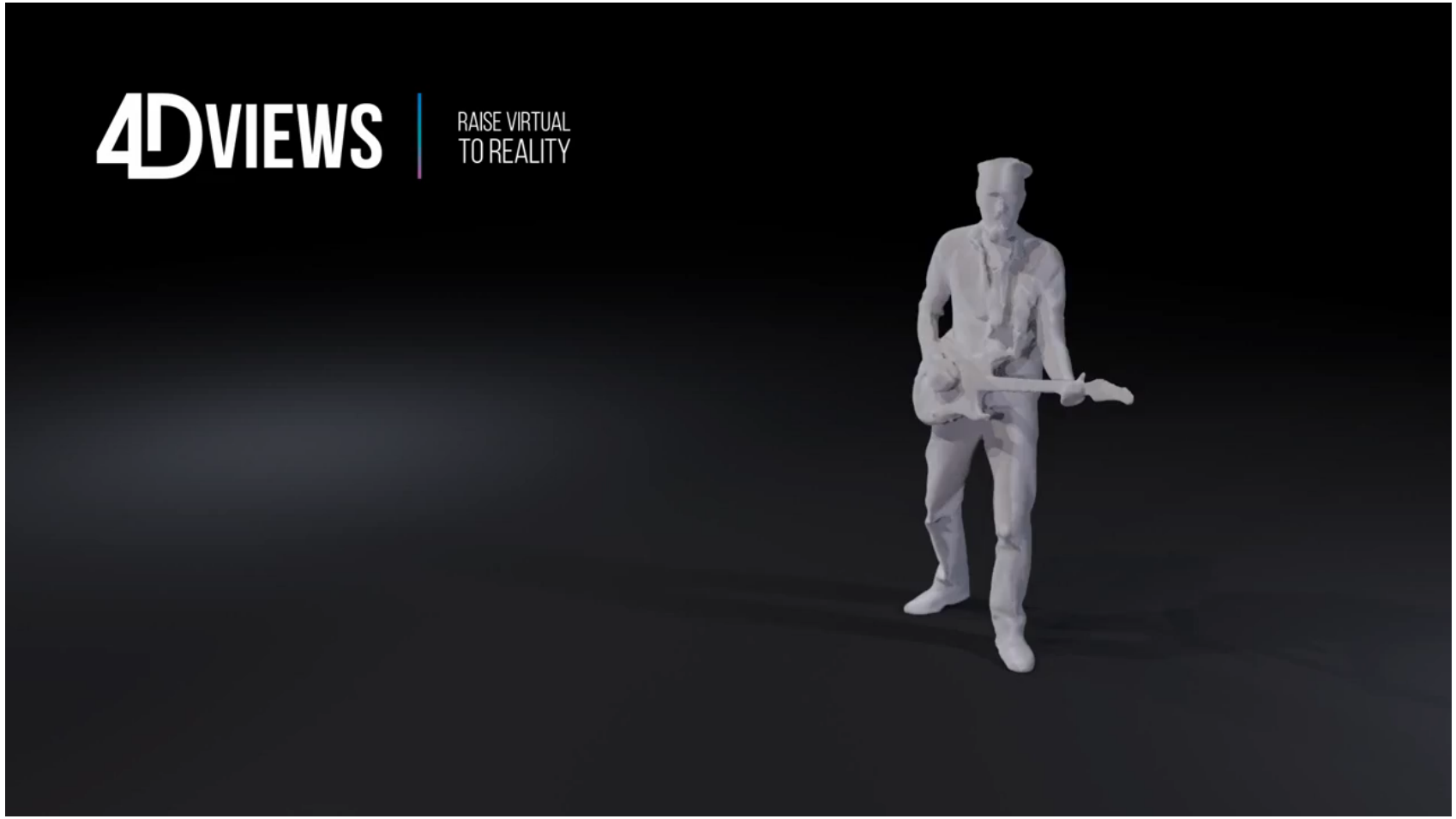


Intersection of foreground cones

Shape-from-silhouette

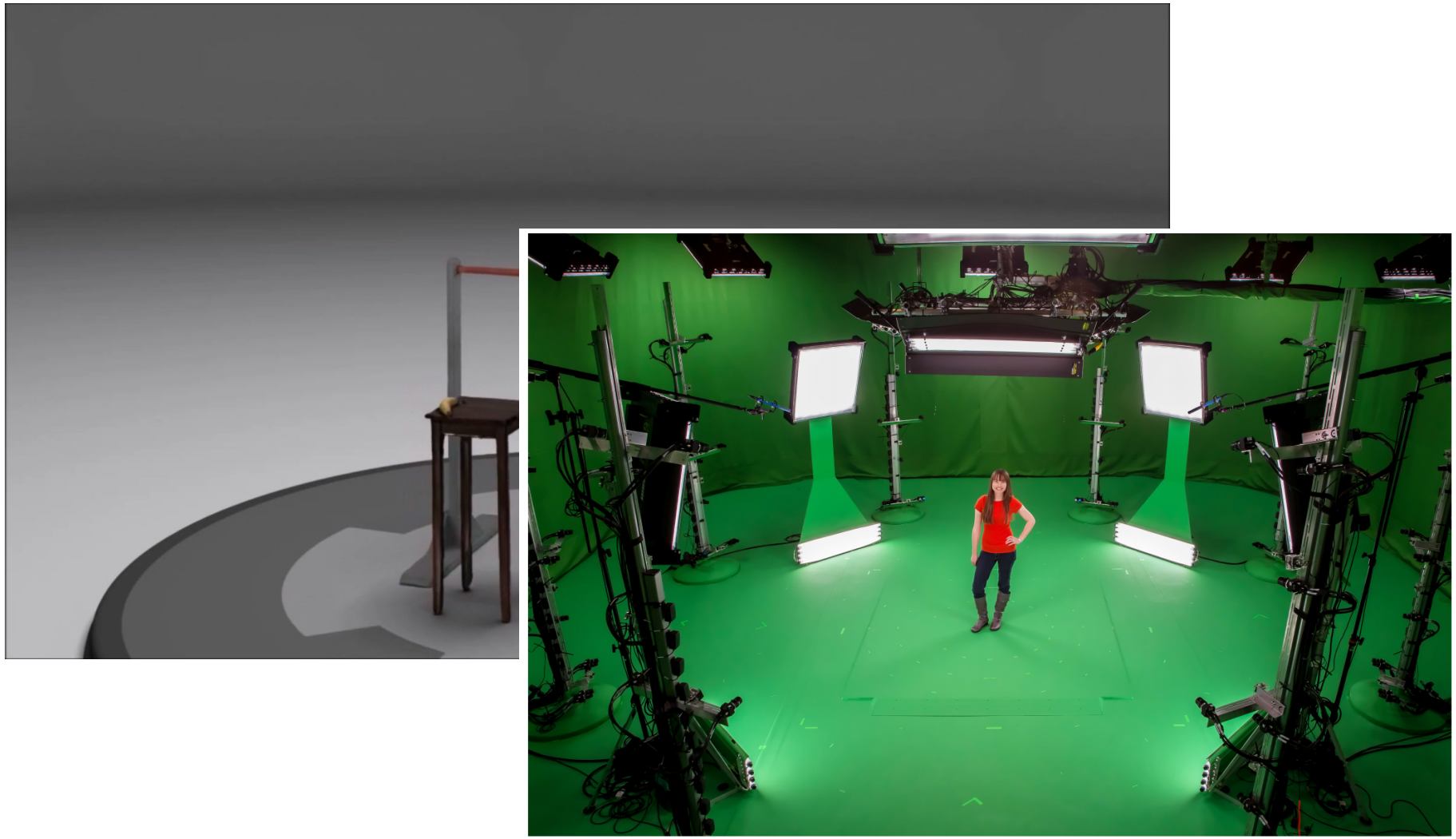
Offline, Controlled environment
Geometry quality is low, but sharp edges

Shape-from-silhouette



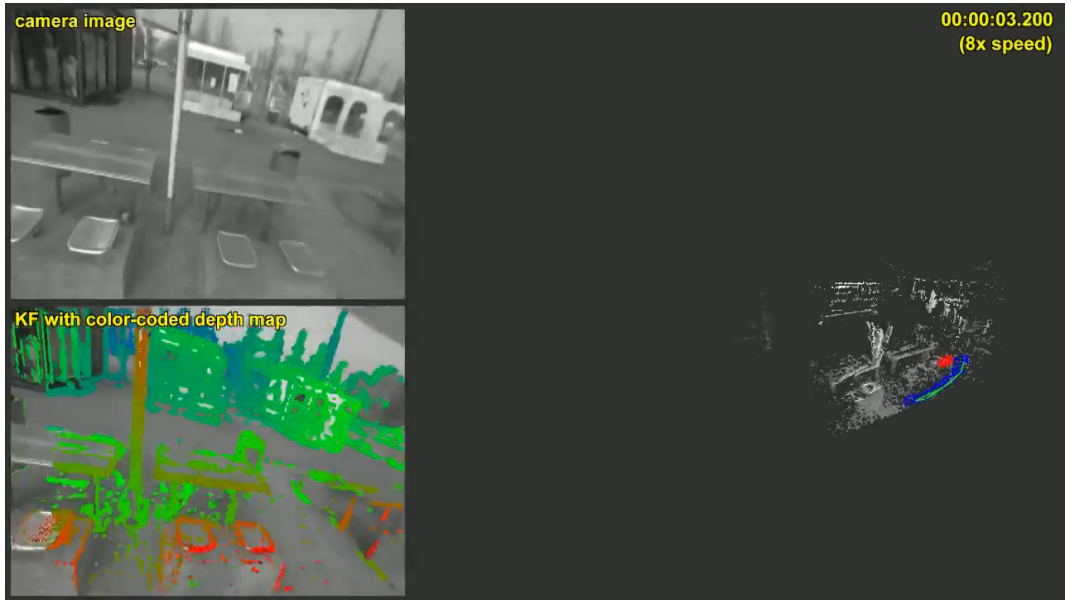
Offline, Controlled environment
Geometry quality is low, but sharp edges

Microsoft Free Viewpoint Video (H-Cap)



Reconstruct the world

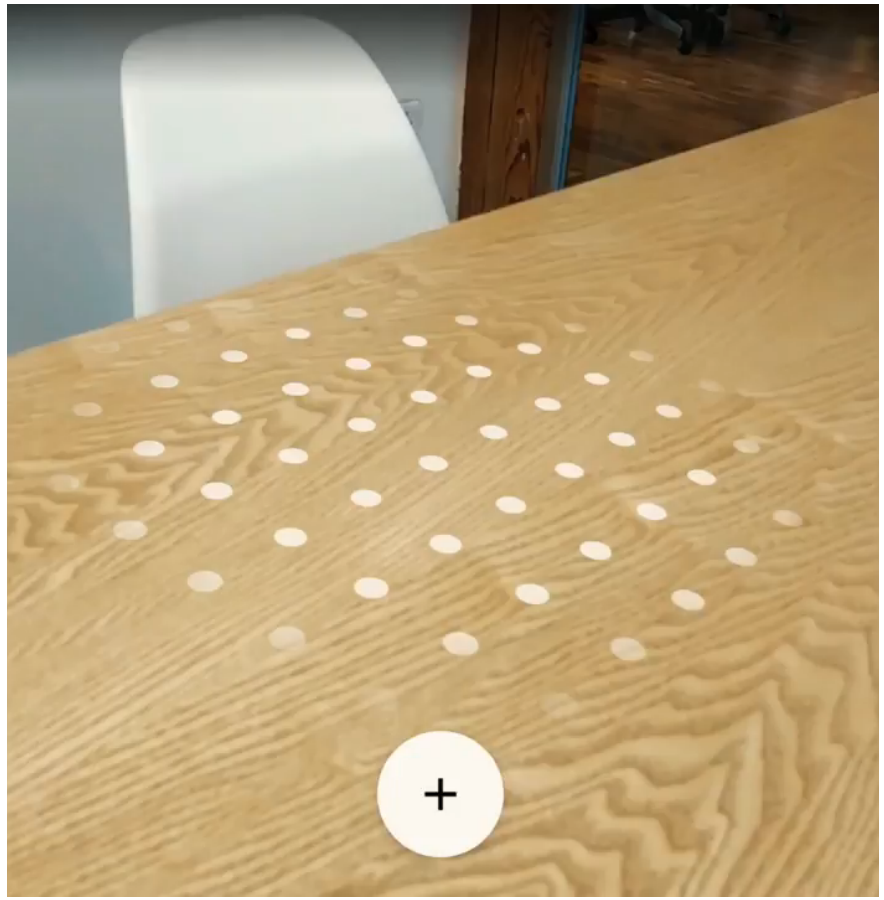
- Simultaneous localization and mapping (SLAM)
  - One moving camera



"LSD-SLAM: Large-Scale Direct Monocular SLAM", J. Engel, T. Schöps, D. Cremers, *ECCV*, 2014

Reconstruct the world

- Simultaneous localization and mapping (SLAM)
  - One moving camera
- Google ARCore/Apple ARKit
  - mobile camera + IMU



Google ARCore

Reconstruct the world

- Simultaneous localization and mapping (SLAM)
  - One moving camera
- Google ARCore/Apple ARKit
  - mobile camera + IMU
- Bundle Adjustment
  - a technique to simultaneously optimize both geometry and camera poses

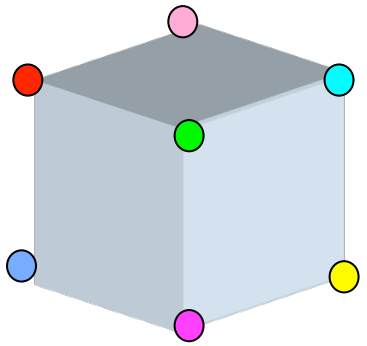
"***Building Rome in a Day***", Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski. ICCV 2009

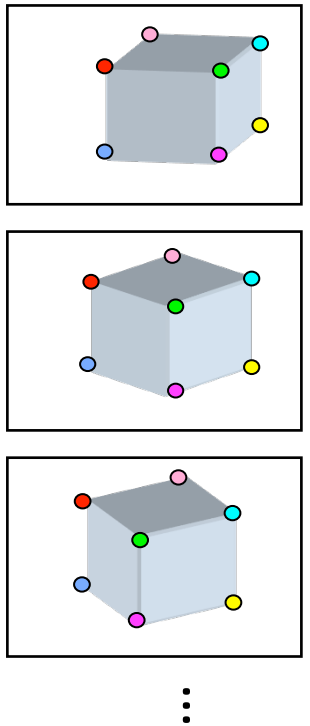
Reconstruct the world

- Simultaneous localization and mapping (SLAM)
  - One moving camera
- Google ARCore/Apple ARKit
  - mobile camera + IMU
- Bundle Adjustment
  - a technique to simultaneously optimize both geometry and camera poses
- KinectFusion
  - One depth camera, detailed geometry

"***KinectFusion: Real-time dense surface mapping and tracking.***" Newcombe, Richard A., Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. ISMAR 2011

Bundle Adjustment



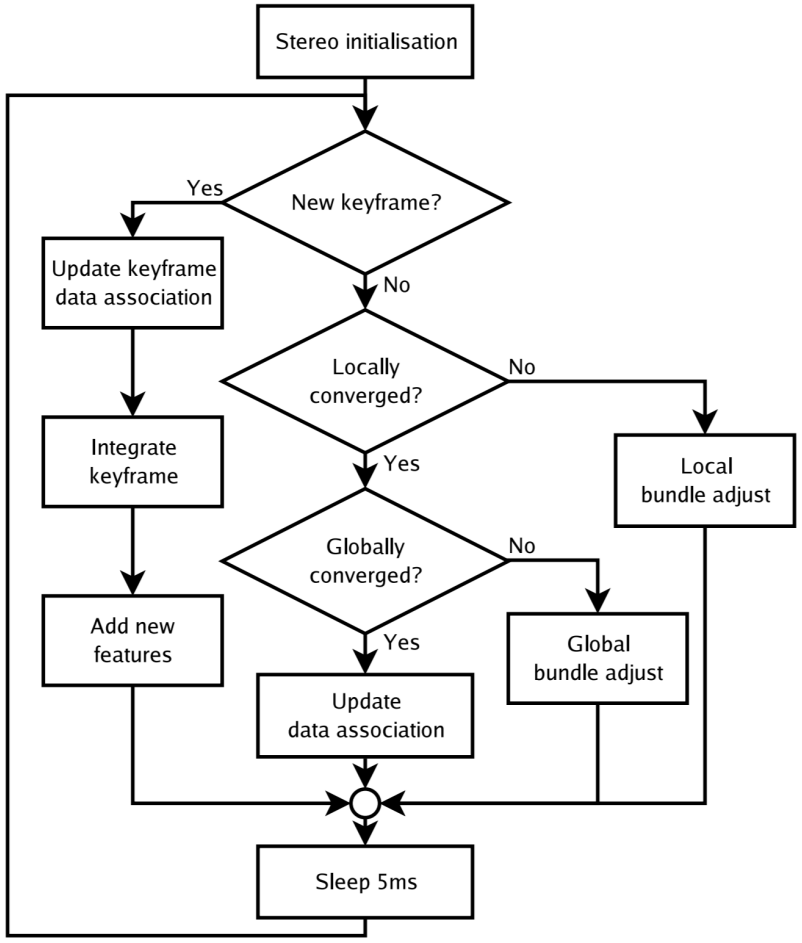


Observation:
Sparse features

$$\min_{\{X_k\},\{R_i,t_i,K_i\}} \sum_{i,k} \| x_{i,k} - Proj(X_k; R_i, t_i, K_i) \|^2$$

SLAM

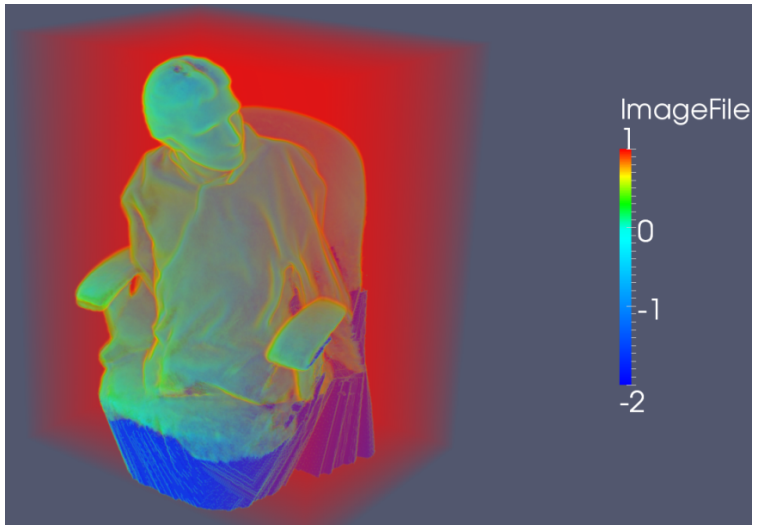
- detect sparse feature points, eg. *SIFT*
- Initial the system with two-view-geometry
- Estimate camera poses for later frames by matching 2D features with 3D points

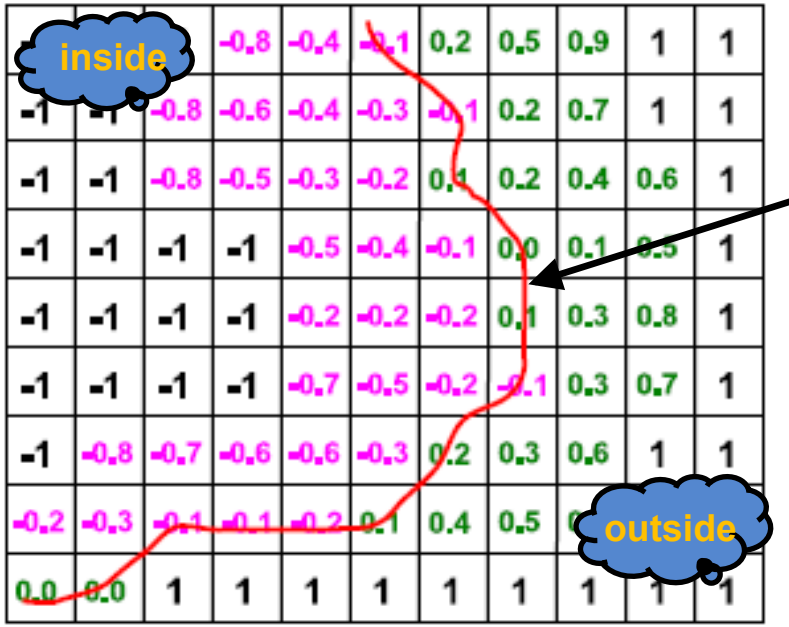


"*Parallel Tracking and Mapping for Small AR Workspaces*", Georg Klein and David Murray, ISMAR'07

KinectFusion

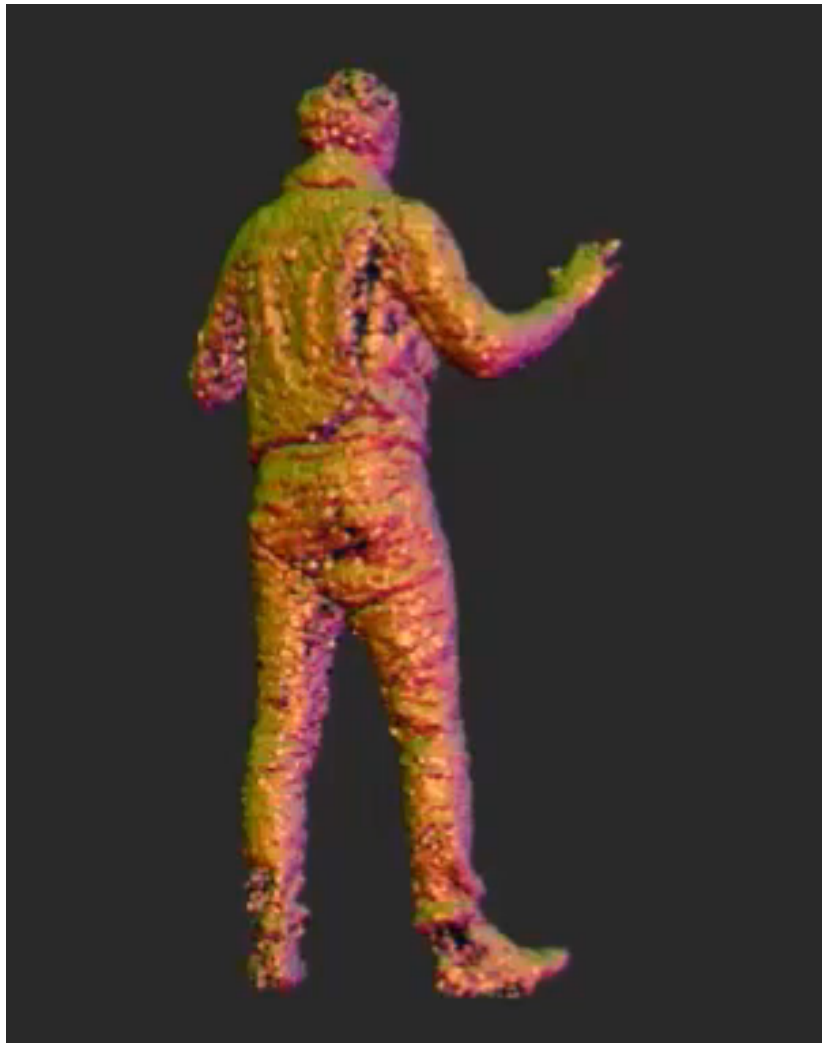
- Data accumulation in a TSDF grid
- Camera pose tracking with ***Iterative Closest Point (ICP)***





People Reconstruction

The need for temporal consistency

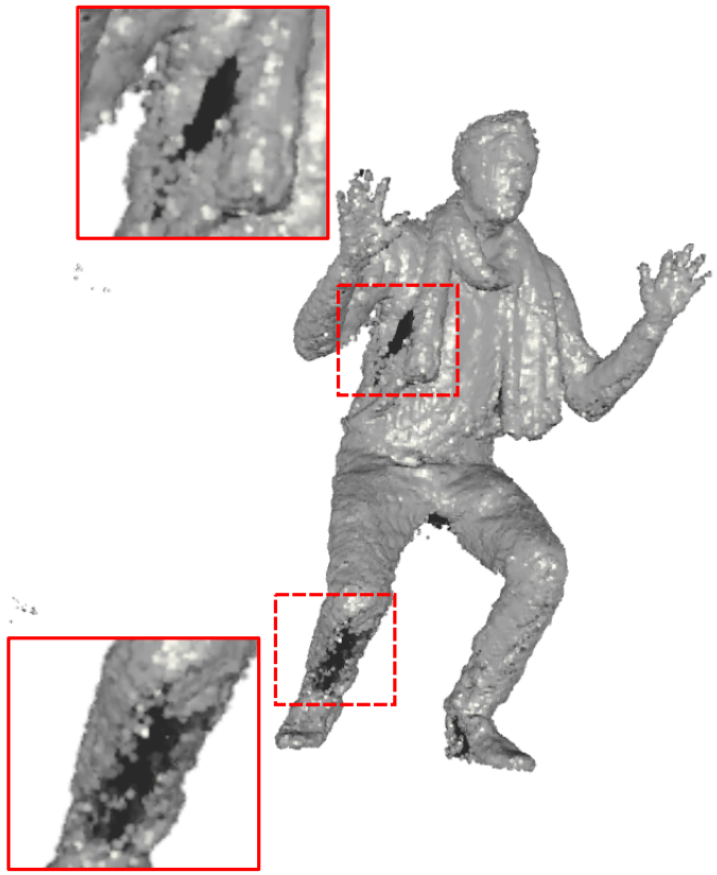


Live Data

Temporally Consistent Model

The need for temporal consistency

Multiple Point Clouds
(Bilaterally-Smoothed)

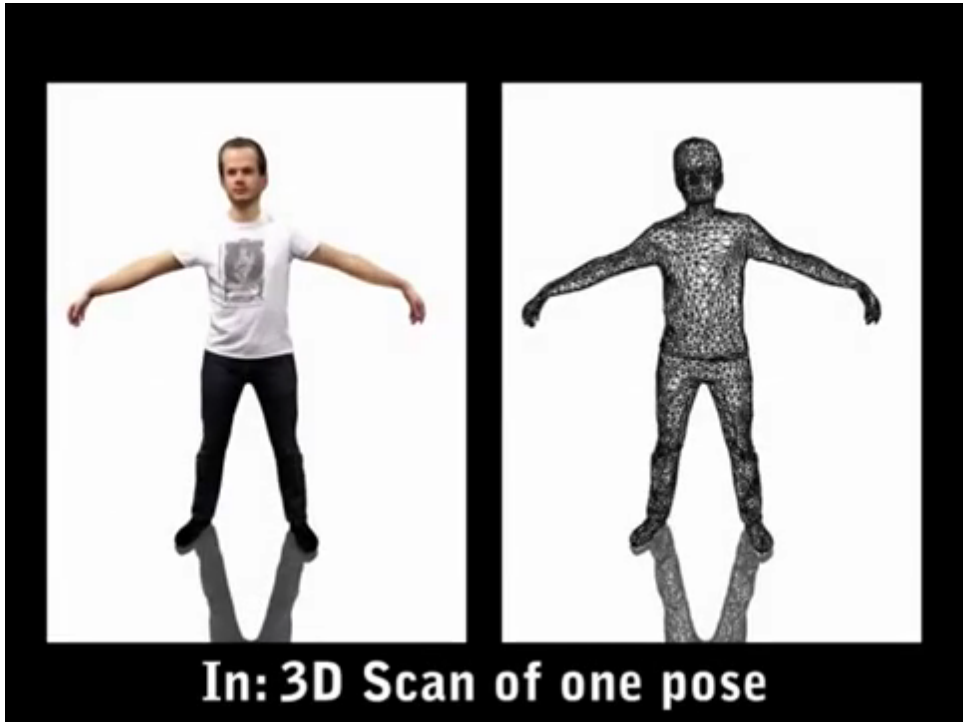


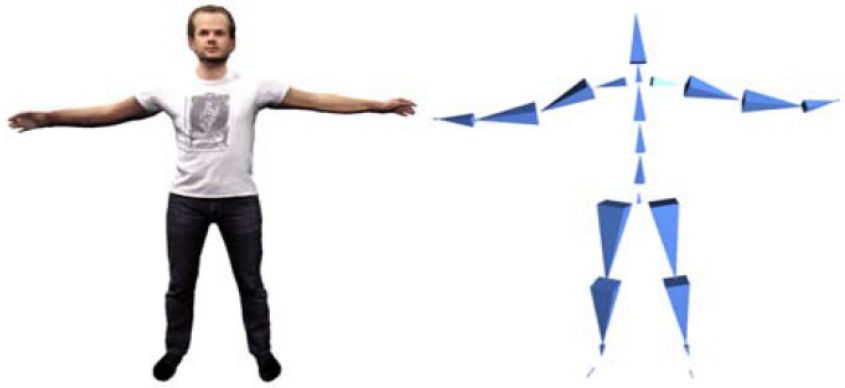
Fused Live Data
(Kinect Fusion)

Temporally consistent model

Reconstruct People

- Articulated Body tracking
  - Human Body tracking



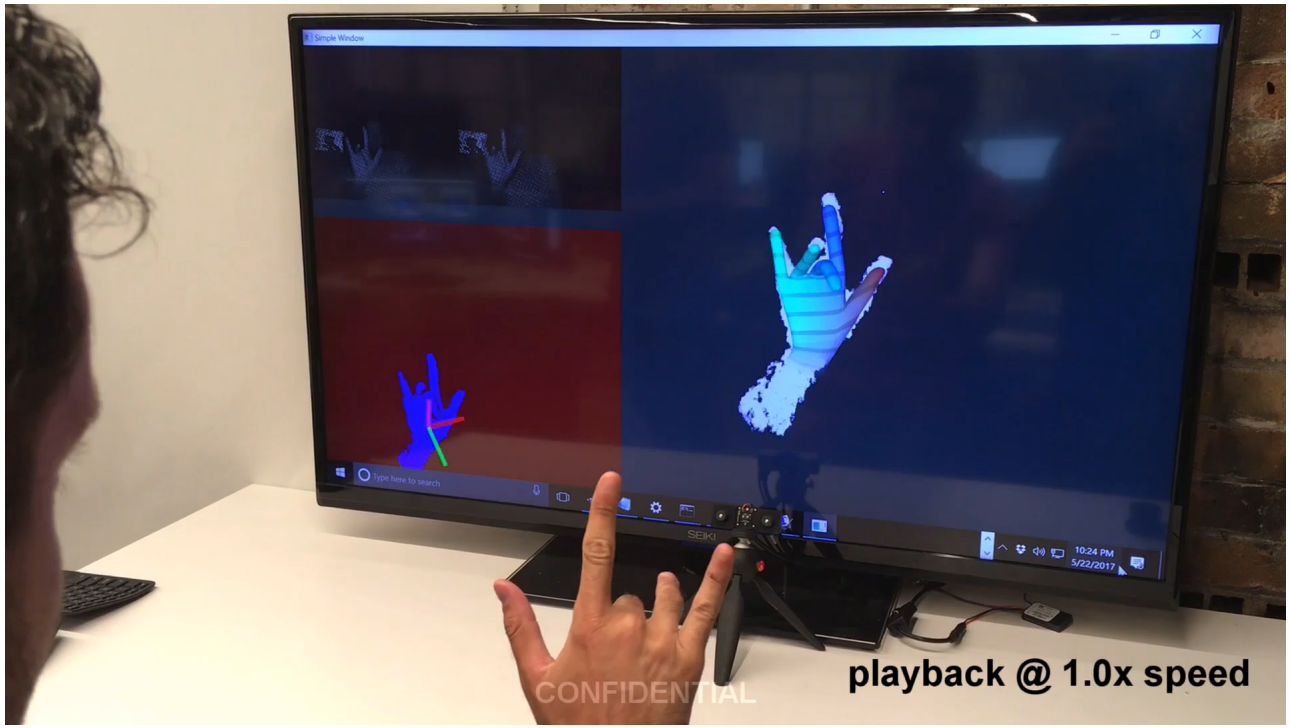


Prescan + Skeleton

"***Marker-less Motion Capture of Skinned Models in a Four Camera Set-up using Optical Flow and Silhouettes***", L. Ballan and G. M. Cortelazzo, 3DPVT 2008

Reconstruct People

- Articulated Body tracking
  - Human Body tracking
  - Hand Tracking



"Articulated Distance Fields for Ultra-Fast Tracking of Hands Interacting", Jonathan Taylor, Vladimir Tankovich, Danhang Tang, Cem Keskin, David Kim, Philip Davidson, Adarsh Kowdle, Shahram Izadi. SIGGRAPH Asia 2017

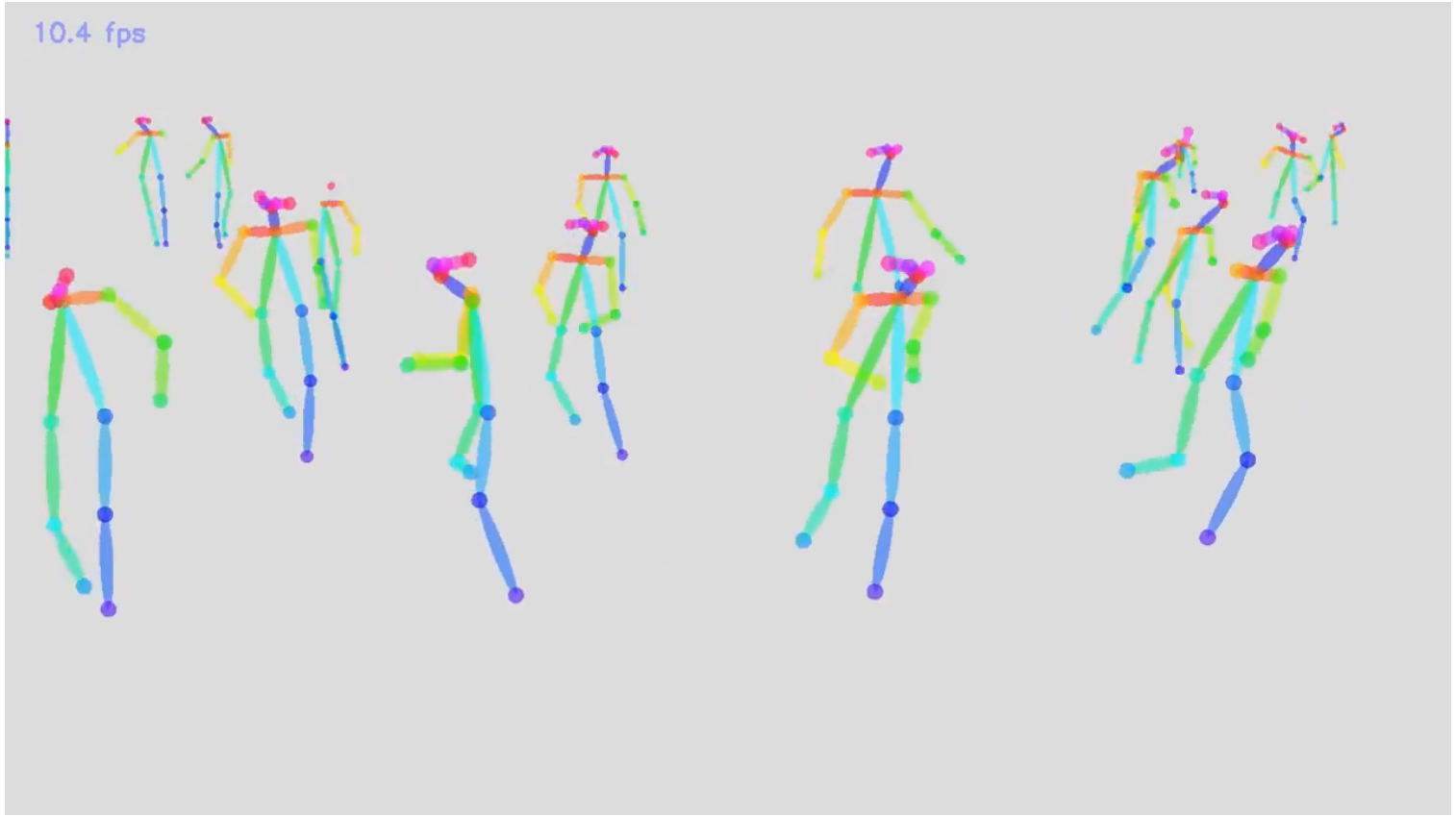
Reconstruct People

- Articulated Body tracking
  - Human Body tracking
  - Hand Tracking

"Articulated Distance Fields for Ultra-Fast Tracking of Hands Interacting", Jonathan Taylor, Vladimir Tankovich, Danhang Tang, Cem Keskin, David Kim, Philip Davidson, Adarsh Kowdle, Shahram Izadi. SIGGRAPH Asia 2017

Reconstruct People

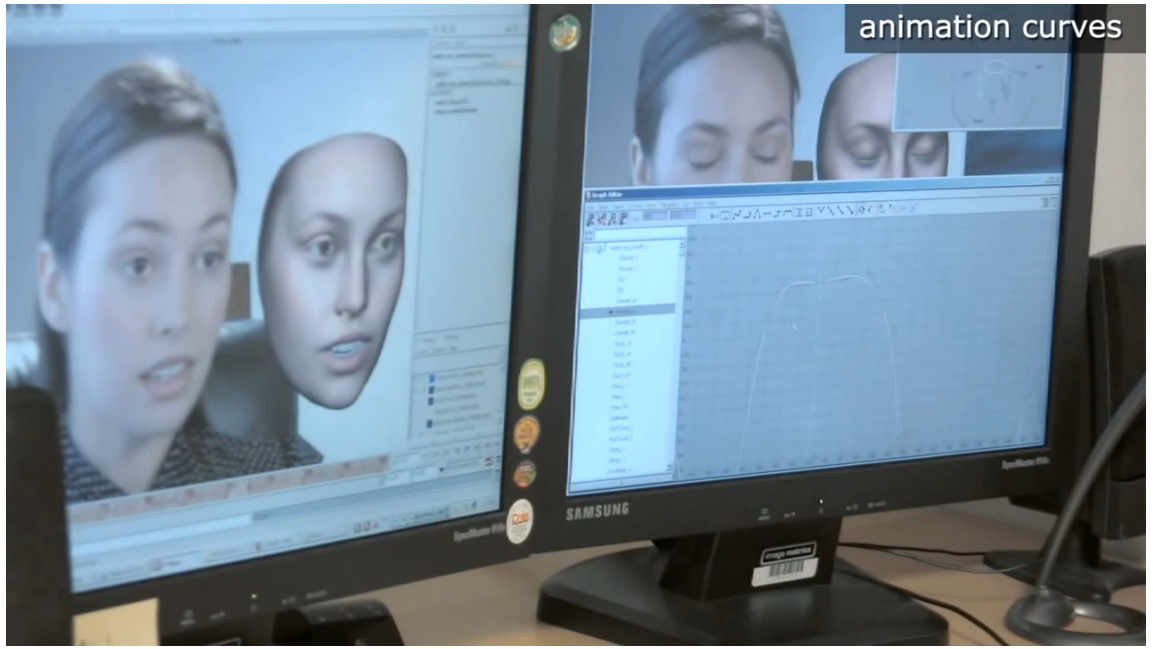
- Articulated Body tracking
  - Human Body tracking
  - Hand Tracking
- Discriminative method
  - Kinect body tracker
  - deep learning



"*Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*", Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh, CVPR 2017.

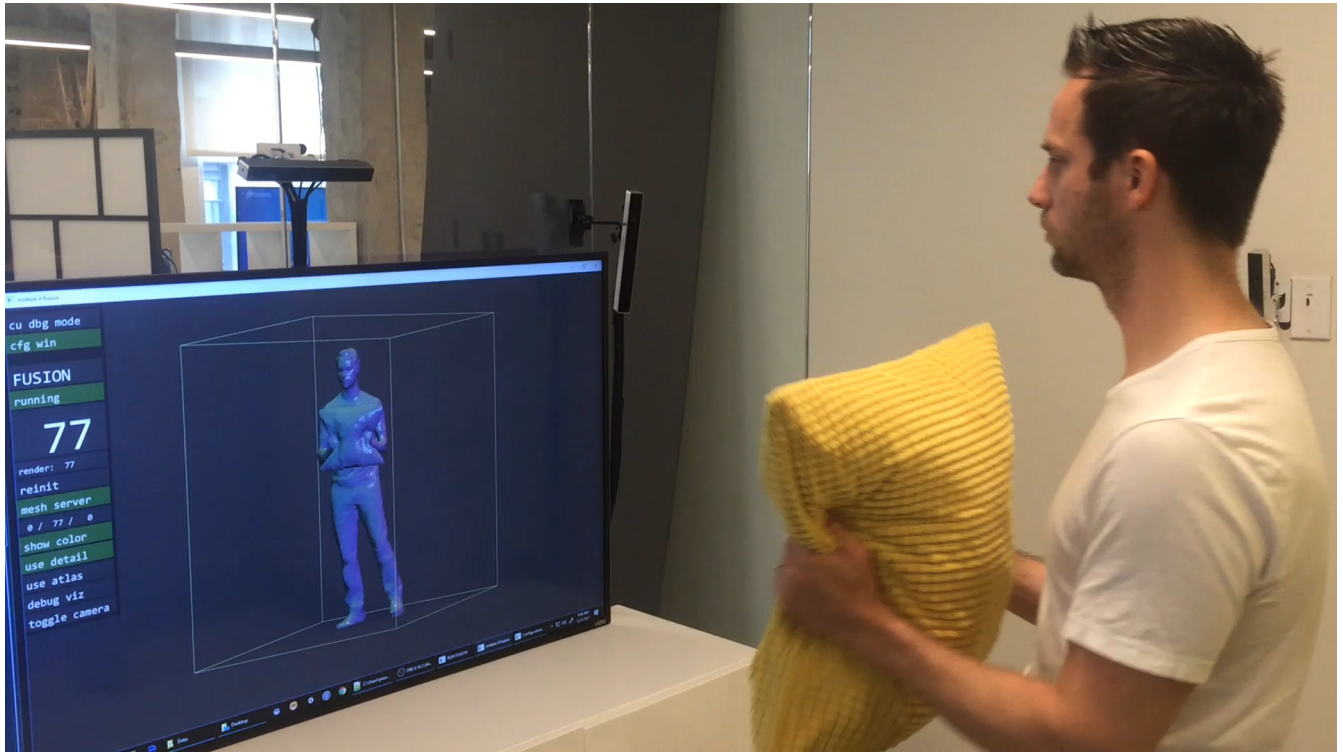
Reconstruct People

- Articulated Body tracking
  - Human Body tracking
  - Hand Tracking
- Discriminative method
  - Kinect body tracker
  - deep learning
- Face Tracking



Digital Emily Project, Paul Debevec

General Nonrigid Surface Tracking and Fusion



- Less infra-structure, no green screen.
- No pre-scan, works for general surface

Non-rigid Alignment

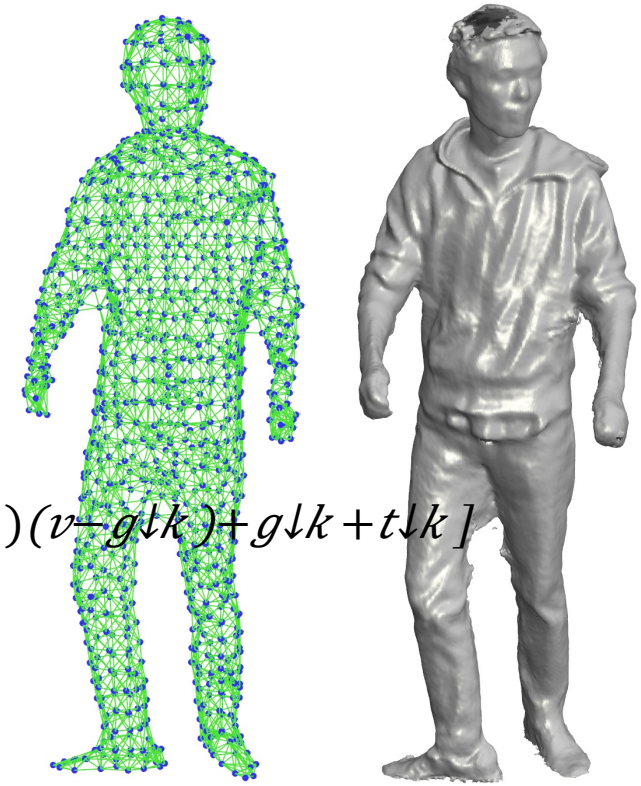
Parameterize motion field as ***Embedded Deformation Graphs***

- One affine transform per node
- Enforce smoothness over edges.

Skinning: attach vertices to neighboring nodes

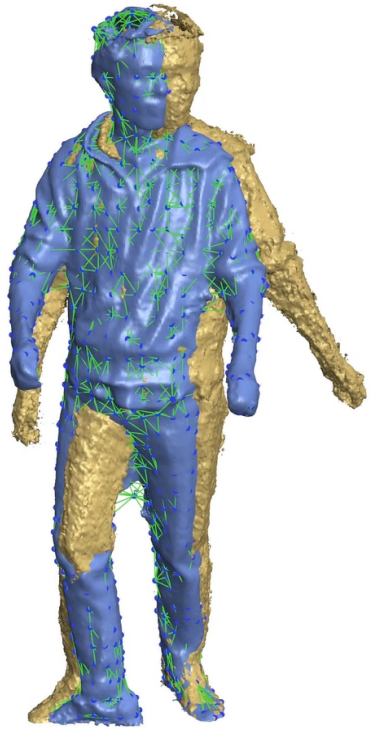
- Based on ***geodesic distance***;
- linear blending

$$v = \sum_{k \in N} w_k \left[A(g_k)(v - g_k) + g_k + t_k \right]$$



Deformation Graph

Reference



$$\min \lambda_{data} E_{data}(G) + \lambda_{reg} E_{reg}(G)$$

geometry alignment

learned correspondences

texture/color consistency

free-space penalization

smooth motion field

as rigid as possible: minimum distortion

Data

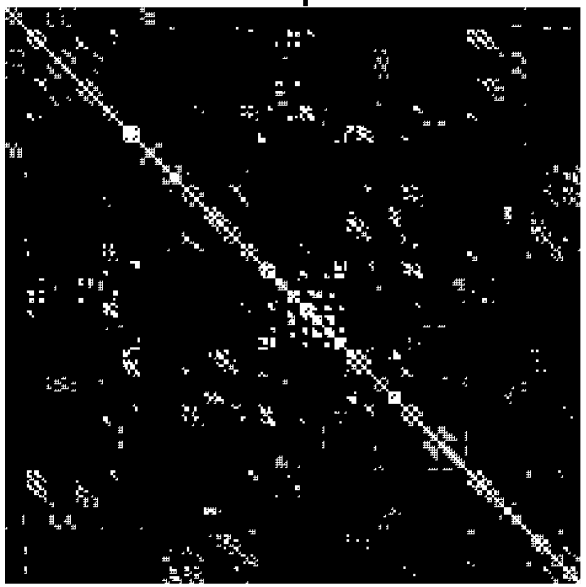
Sumner, Robert W., Johannes Schmid, and Mark Pauly. "Embedded deformation for shape manipulation." ACM Transactions on Graphics (TOG). Vol. 26. No. 3. ACM, 2007.

Solve

$$\min: \lambda_{data} E_{data}(G) + \lambda_{reg} E_{reg}(G)$$

Non-linear least square problem

- **Custom/efficient** sparse Levenberg-Marquardt solver
- PC GPU ~1ms per iteration
- solve the normal equation at each step: $(\mathbf{J}^T \mathbf{J} + \delta I)\mathbf{h} = -\mathbf{J}^T \mathbf{f}$



$\mathbf{J}^T \mathbf{J}$ Non-Zero Pattern

A sparse block matrix, each pixel represents a 7x7 matrix.

Data Term

$$E_{data}(G) = \sum_{m=1}^{M} \sum_{n=1}^{N} (\Psi(v_m; \mathcal{D}_n))^2$$

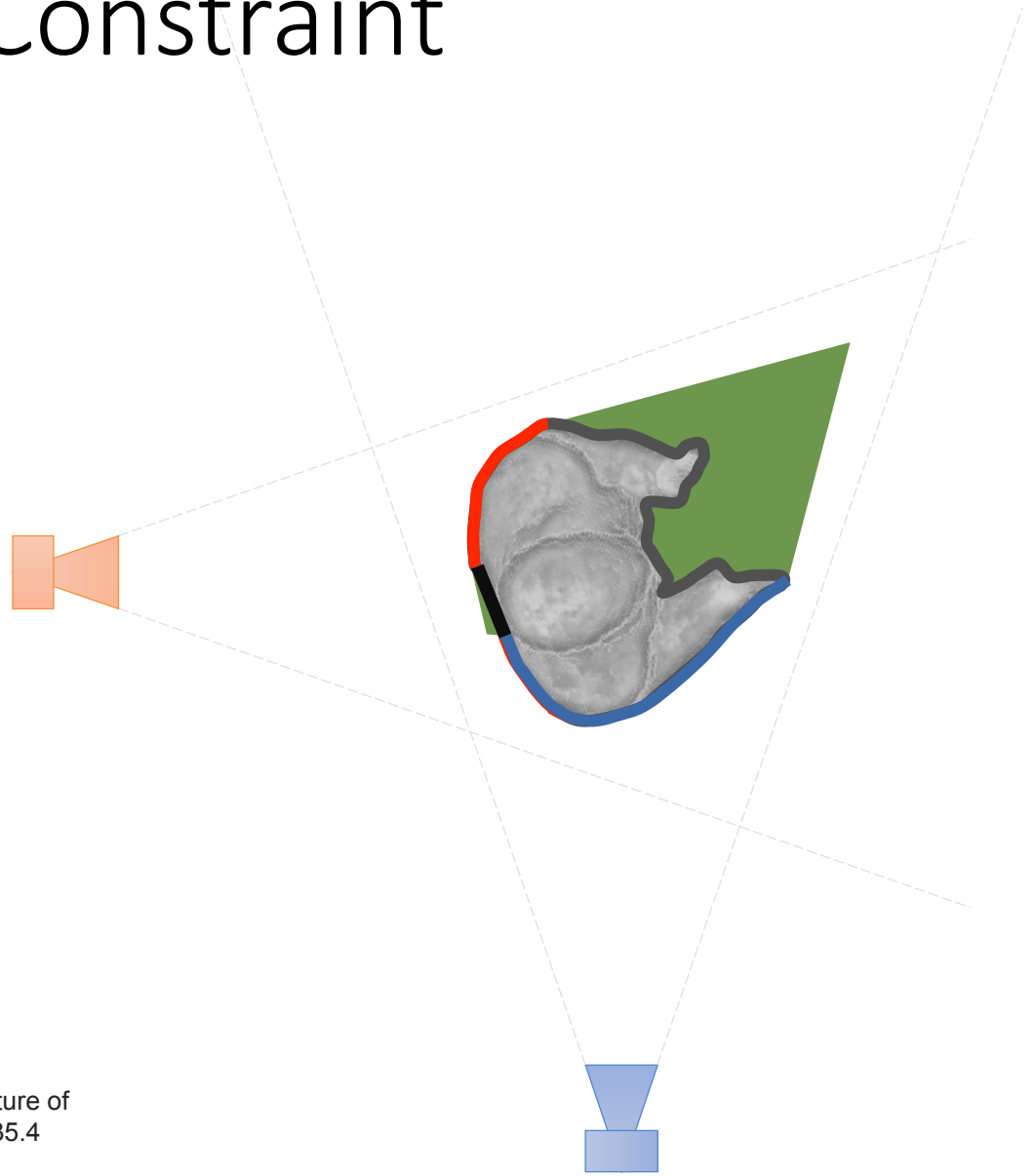
- Alignment residual between model and data

$$\Psi(v_m; \mathcal{D}_n) = \delta_{mn} n_m T(v_m - \Gamma_n(v_m))$$

$\Gamma_n(v_m)$: projective correspondence of v_m in depth map $\mathcal{D}_n$

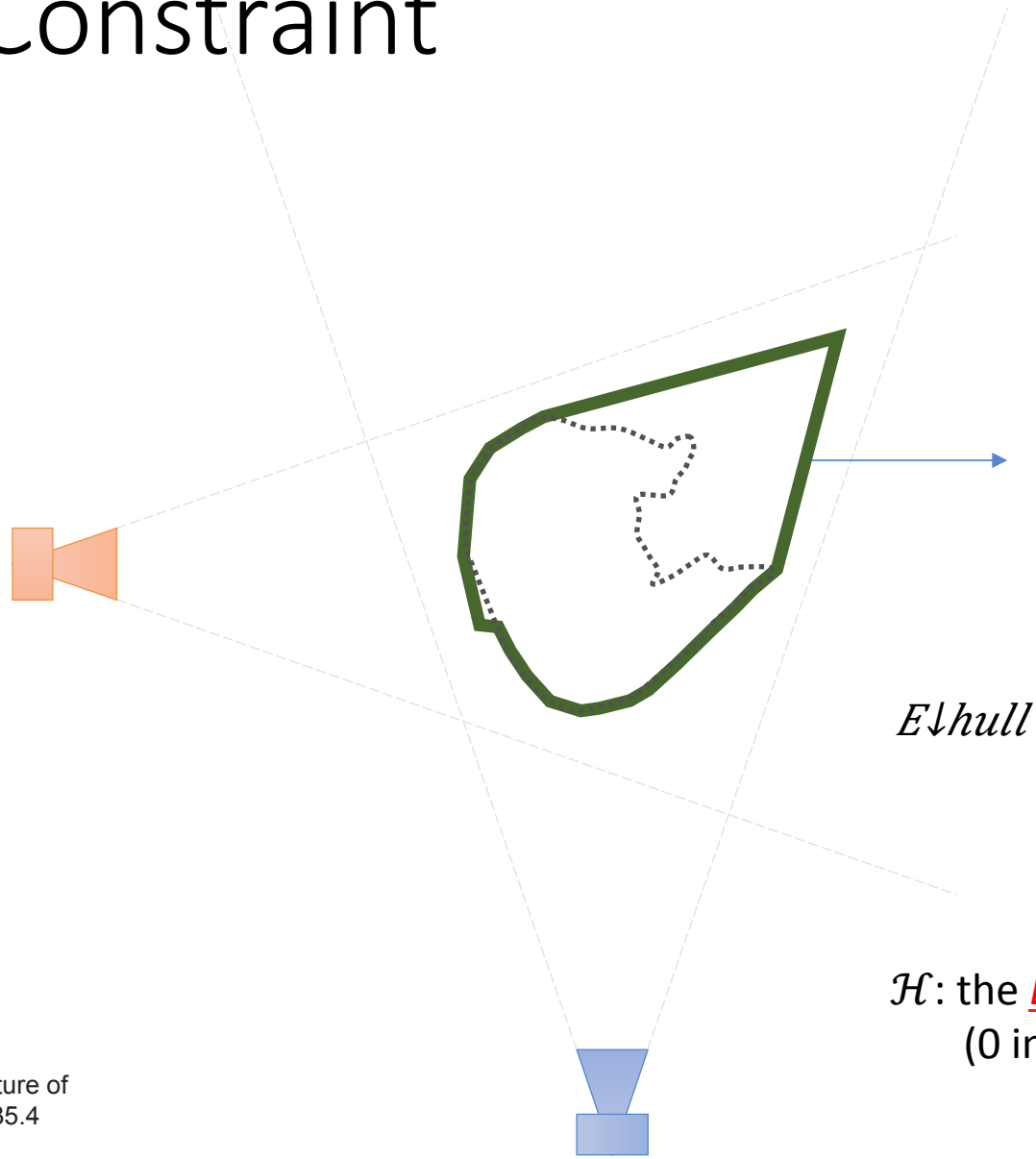
$$\delta_{mn} = \begin{cases} 1 & \text{if } v_m \text{ is visible to } \mathcal{D}_n \\ 0 & \text{otherwise} \end{cases}$$

Visual Hull Constraint



Dou, Mingsong, et al. "Fusion4d: Real-time performance capture of challenging scenes." *ACM Transactions on Graphics (TOG)* 35.4 (2016): 114.

Visual Hull Constraint



$$E\downarrow hull\ (G)=\sum m\uparrow M \blacksquare \mathcal{H}\uparrow 2\ (v\downarrow m\ )$$

$\mathcal{H}$: the *Distance Transform* to the visual hull
(0 inside visual hull)

Dou, Mingsong, et al. "Fusion4d: Real-time performance capture of challenging scenes." *ACM Transactions on Graphics (TOG)* 35.4 (2016): 114.

Color consistency constraint

Measure the difference between the observed color images $\{I_k\}_{k=1}^{K}$ and the reconstructed image by projecting the per-vertex-colored mesh:

$$E_{clr} = \sum_i \sum_{k=1}^{K} \delta_{ik} \| I_k(\Pi_k(v_i)) - c_i \|^2$$

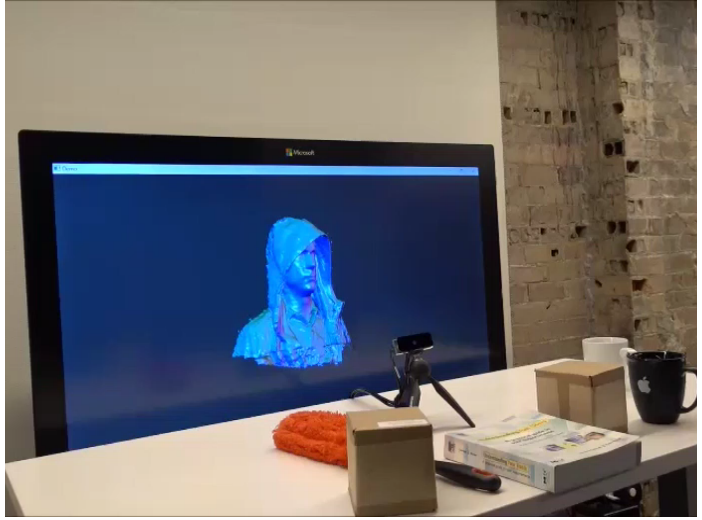
visibility test; camera project function; per-vertex color

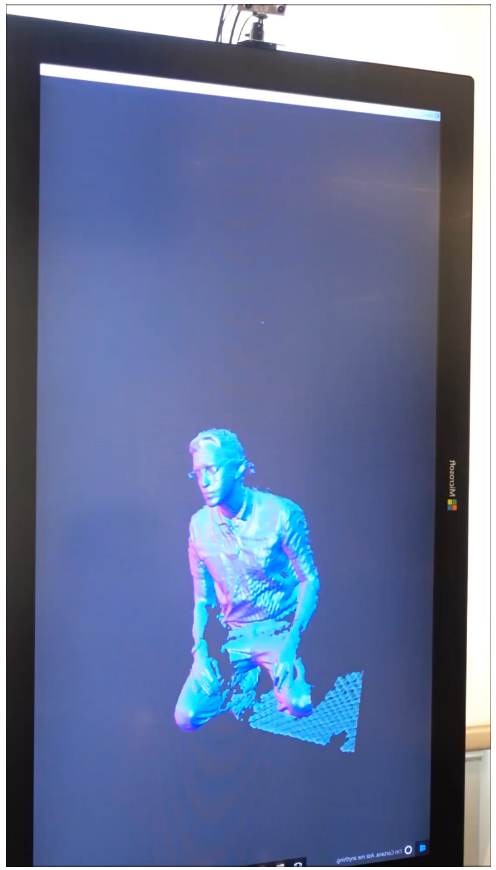
with color constraint

w/o color constraint

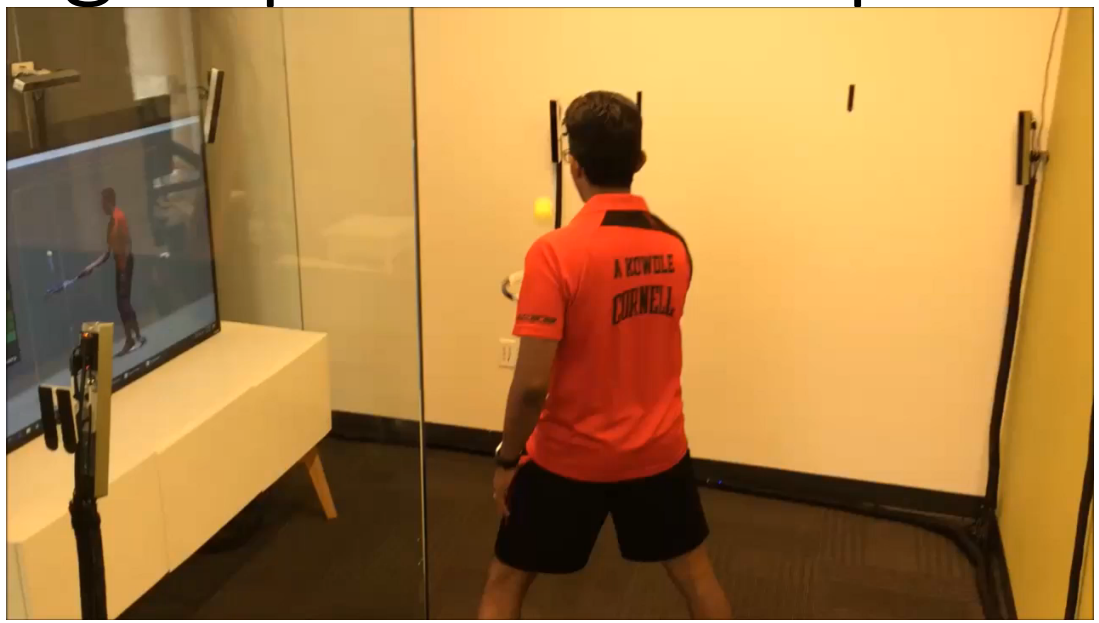
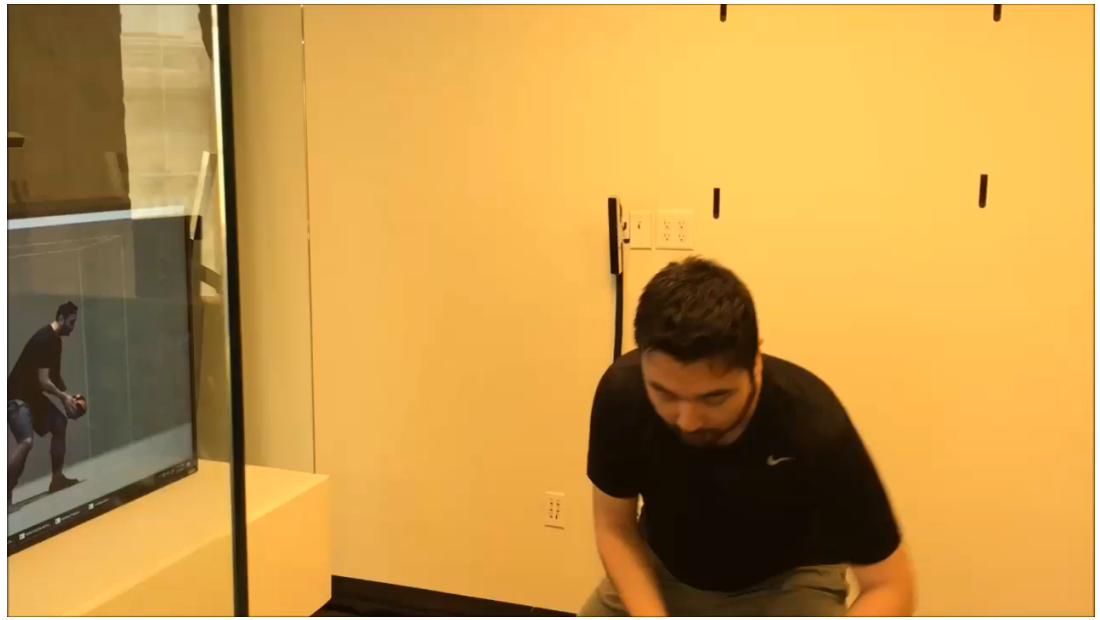
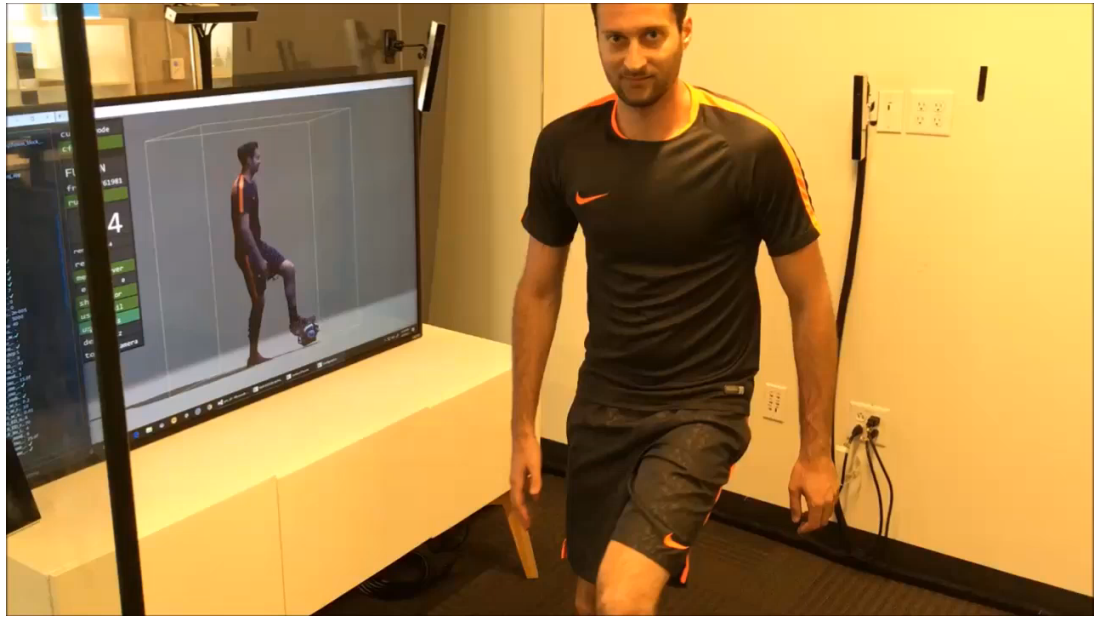
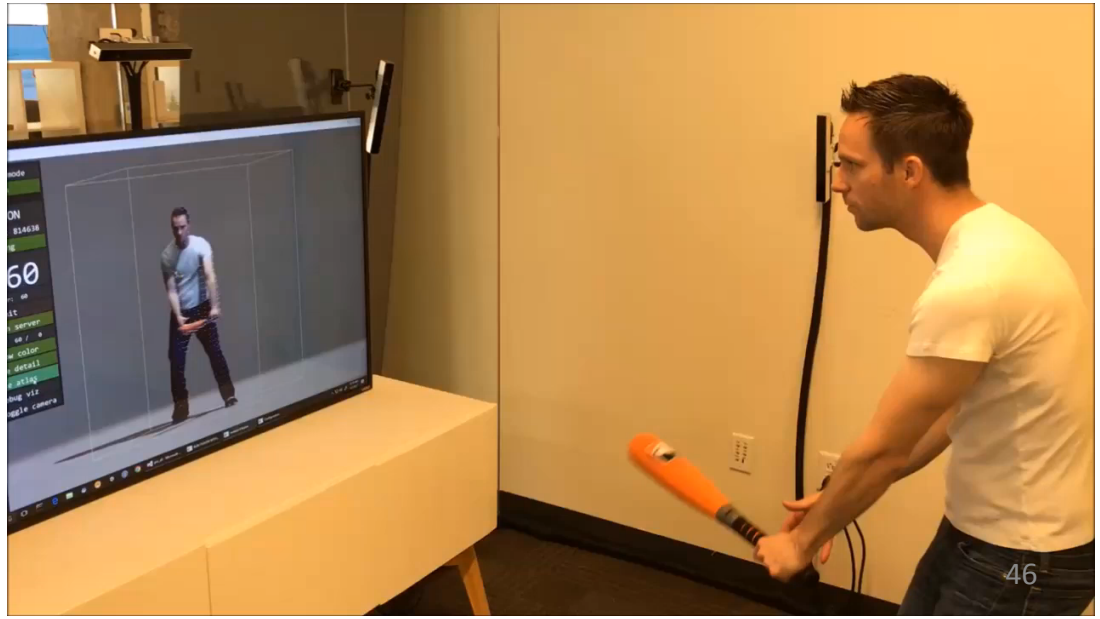
Raw Depth

Single Sensor Results





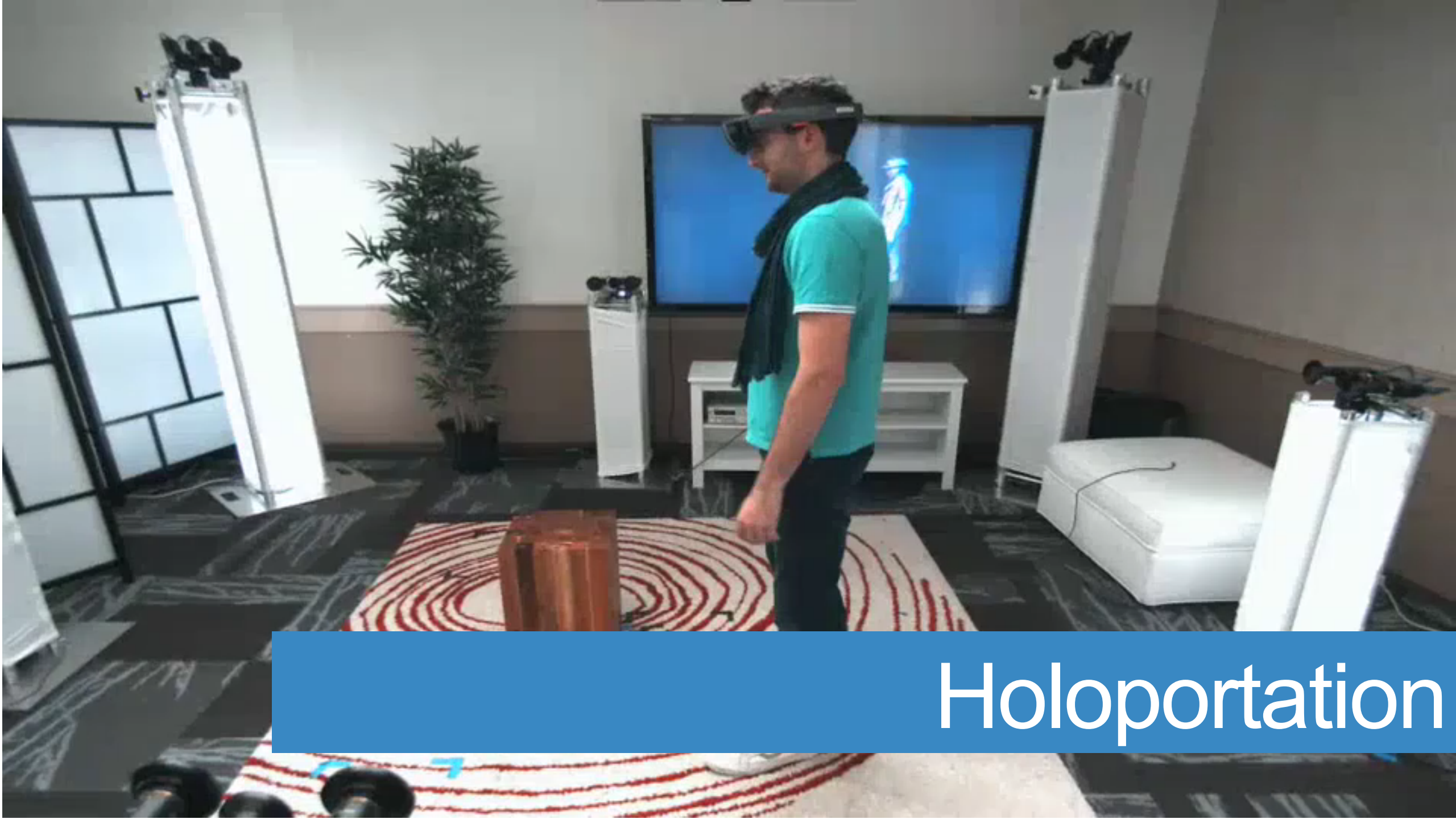
High Speed 360 Capture

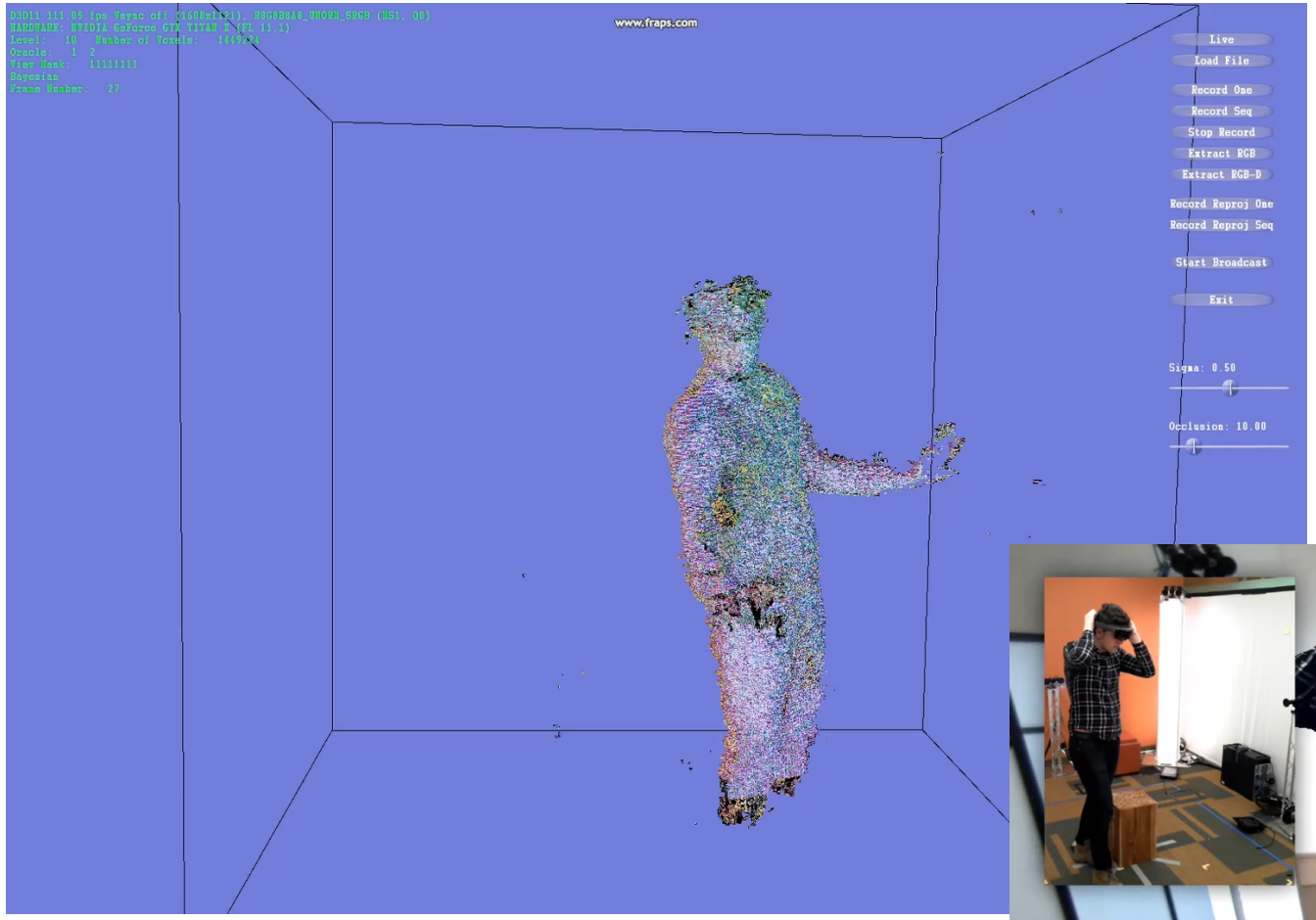
Holoportation

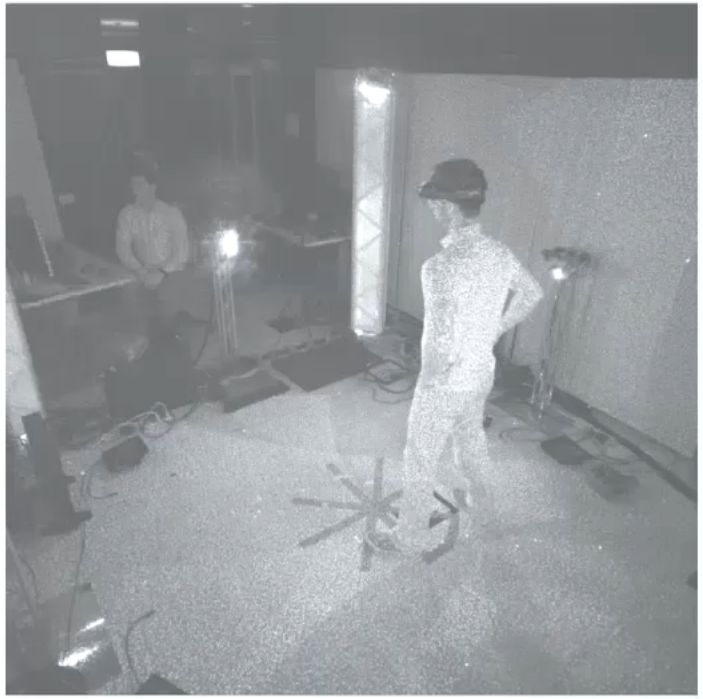

Normal environment
24 cameras
Real-time
Holoportation

Holoportation

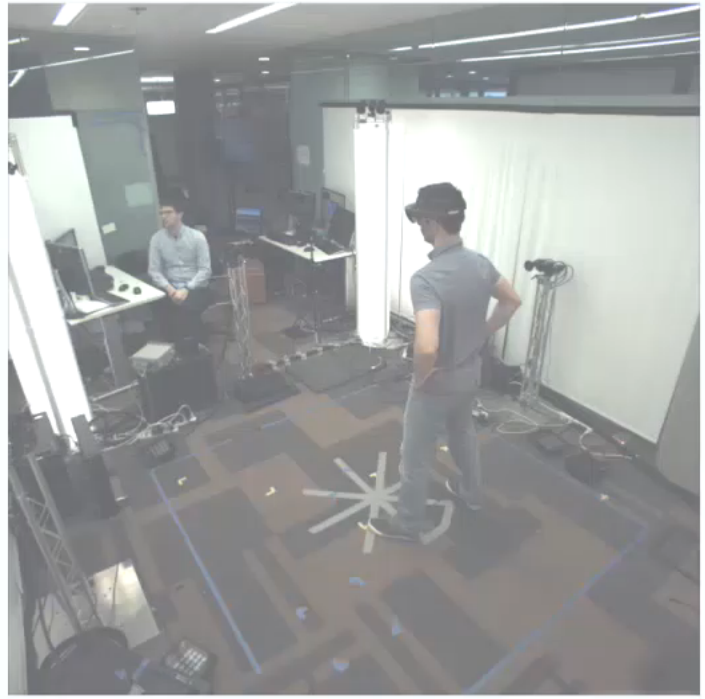


Holoportation


www.fraps.com
Live
Load File
Record One
Record Seq
Stop Record
Extract RGB
Extract RGB-D
Record Reproj One
Record Reproj Seq
Start Broadcast
Exit
Sigma: 0.50
Occlusion: 10.00

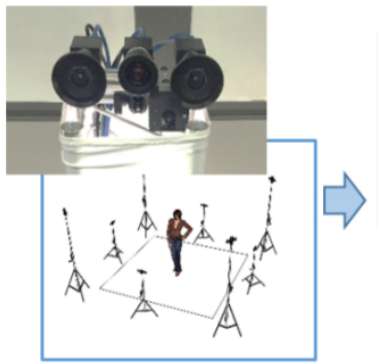


SITE A



SITE B

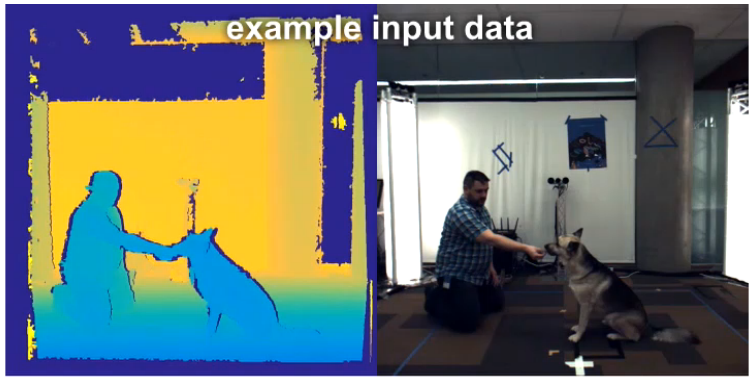
8 Pods

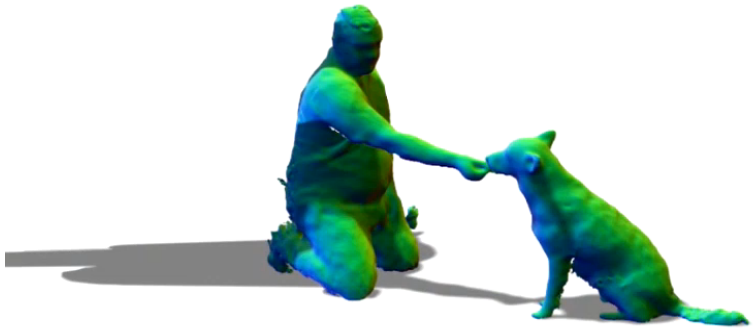


Capture

Results

Non-human examples





real-time multi-view reconstruction

Usage Scenarios

Challenges for the future: FoV

Challenges: FoV

Challenges: Infrastructure

Challenges: Headset removal

Compression

- Raw Data from Cameras = 23 Gb/s
- Current HoloPort Compression = 1 Gb/s
- HD video = 10 Mb/s
- Exploit temporally consistency and texture atlas


Thank you!