

The Real-World Applications of **3D Reconstruction** in the **Panoptic Studio**

Hanbyul (Han) Joo

The Robotics Institute
Carnegie Mellon University

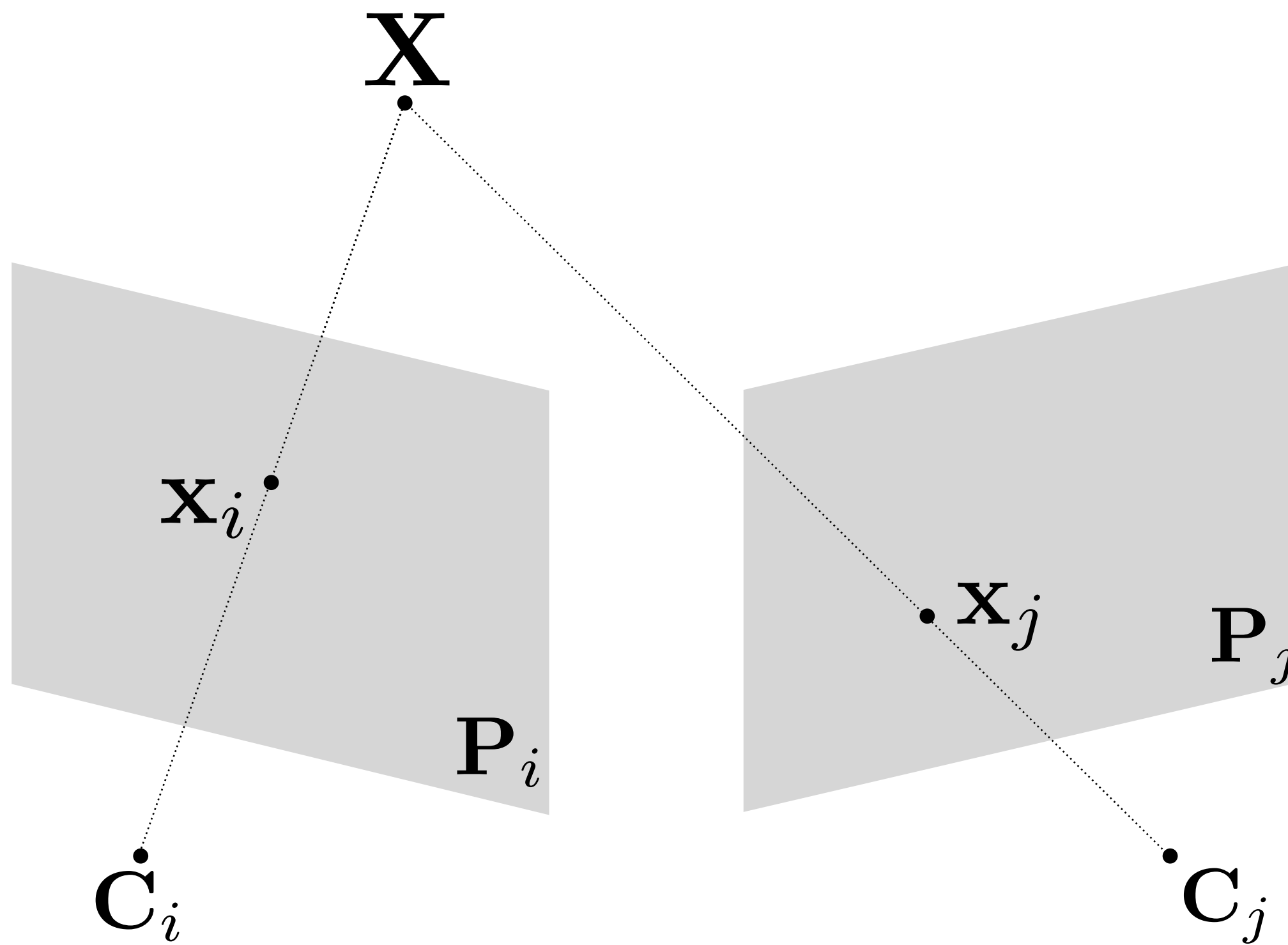
Basic Knowledge for 3D Reconstruction

You Should Be Familiar with

- Camera Matrix
- Triangulation
- Stereo and Structured Light
- Fundamental Matrix and Essential Matrix
- Camera Pose Estimation (Perspective-n-Point)
- DLT and SVD (Homogeneous Least Square Problem)

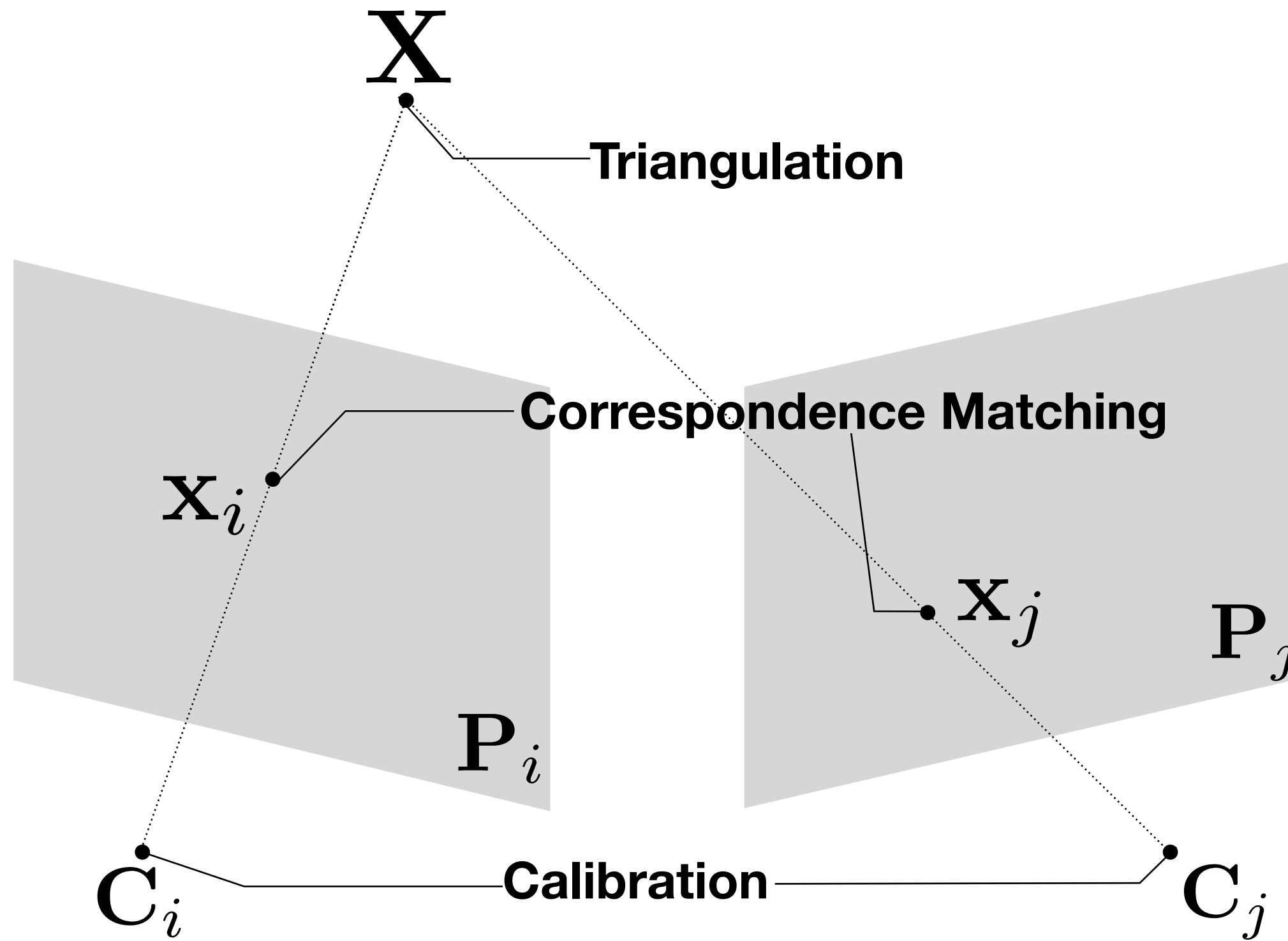
Building A System For 3D Reconstruction

An Example with Two Static Cameras



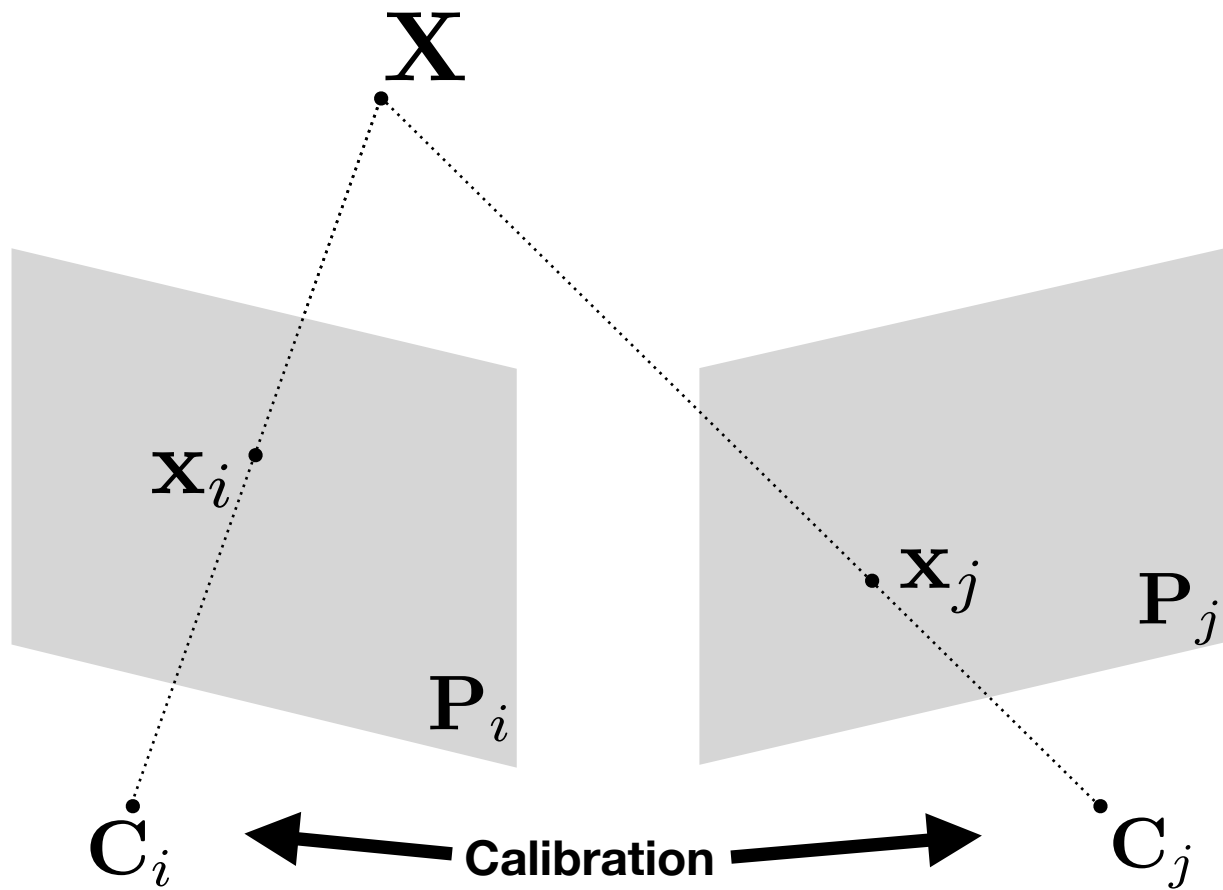
Building A System For 3D Reconstruction

An Example with Two Static Cameras



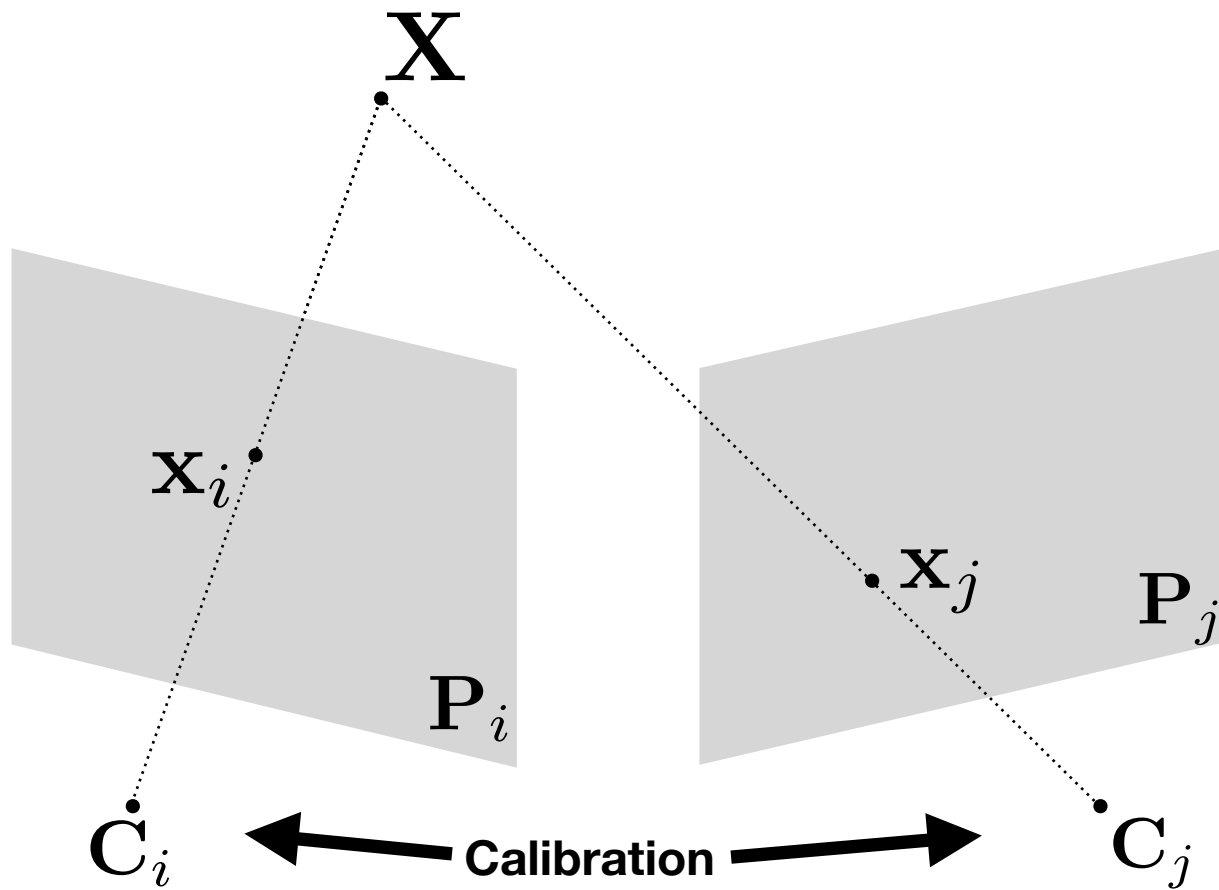
Calibration

Computing Geometrical Relation Between Cameras (**K**, **R**, **t**)



Calibration

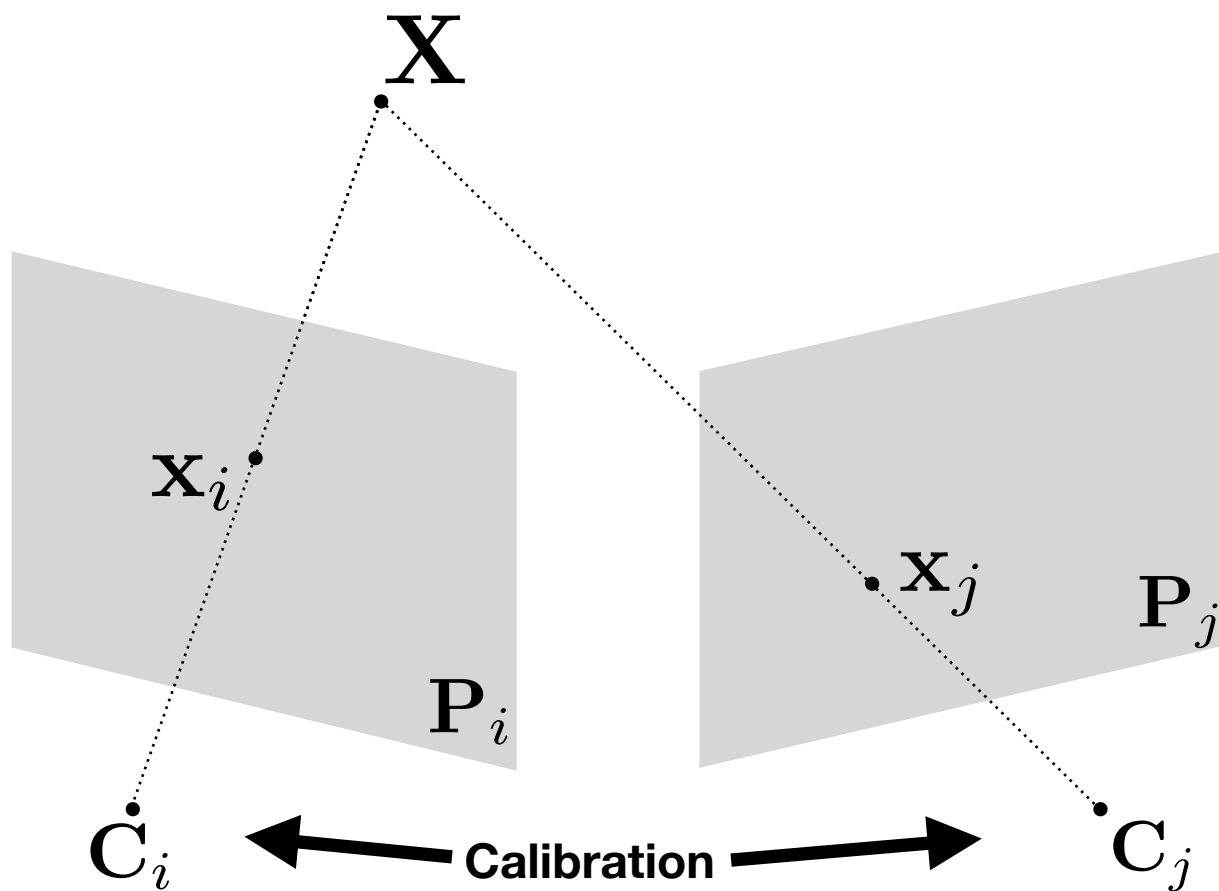
Computing Geometrical Relation Between Cameras (\mathbf{K} , \mathbf{R} , \mathbf{t})



- Read H&Z book
- Read vision/geometry lecture slides
- Open Matlab
- Compute Essential matrix
-

Calibration

Computing Geometrical Relation Between Cameras (\mathbf{K} , \mathbf{R} , \mathbf{t})

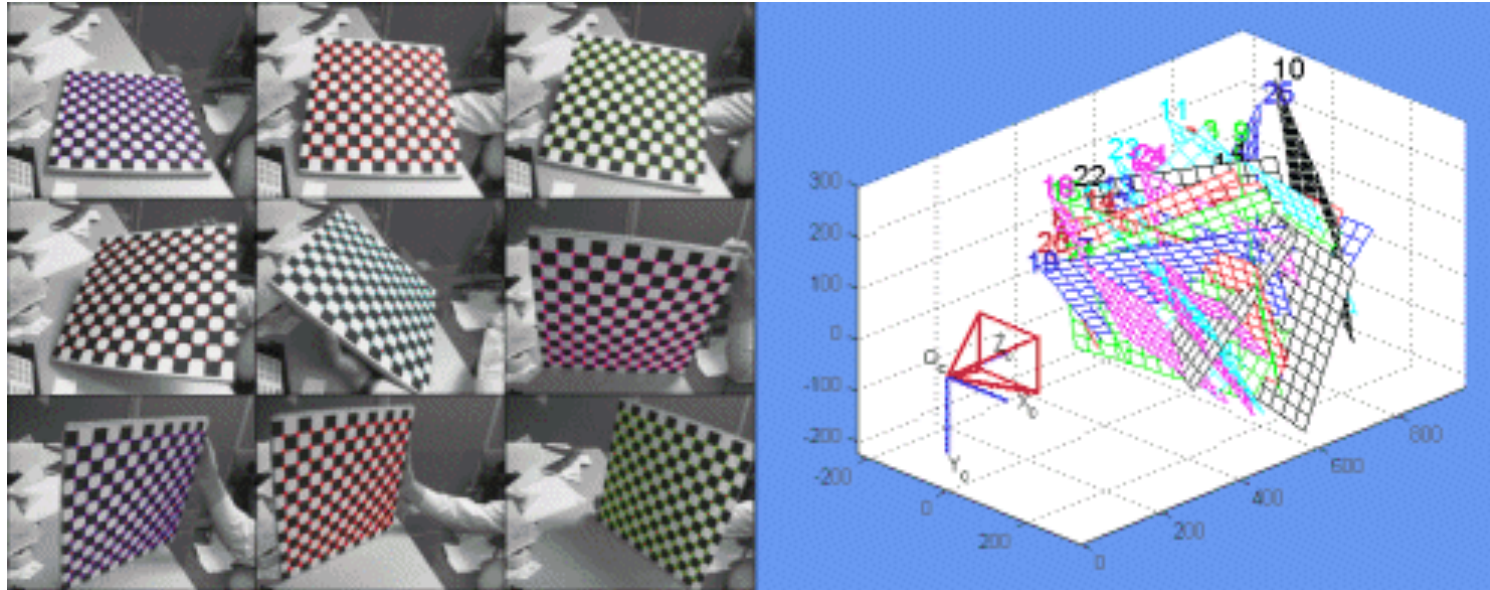


- Read H&Z book
- Read 16-720 lecture slides
- Open Matlab
- Compute Essential matrix
-



Calibration

Use Calibration Softwares



e.g., Caltech Calibration Toolbox

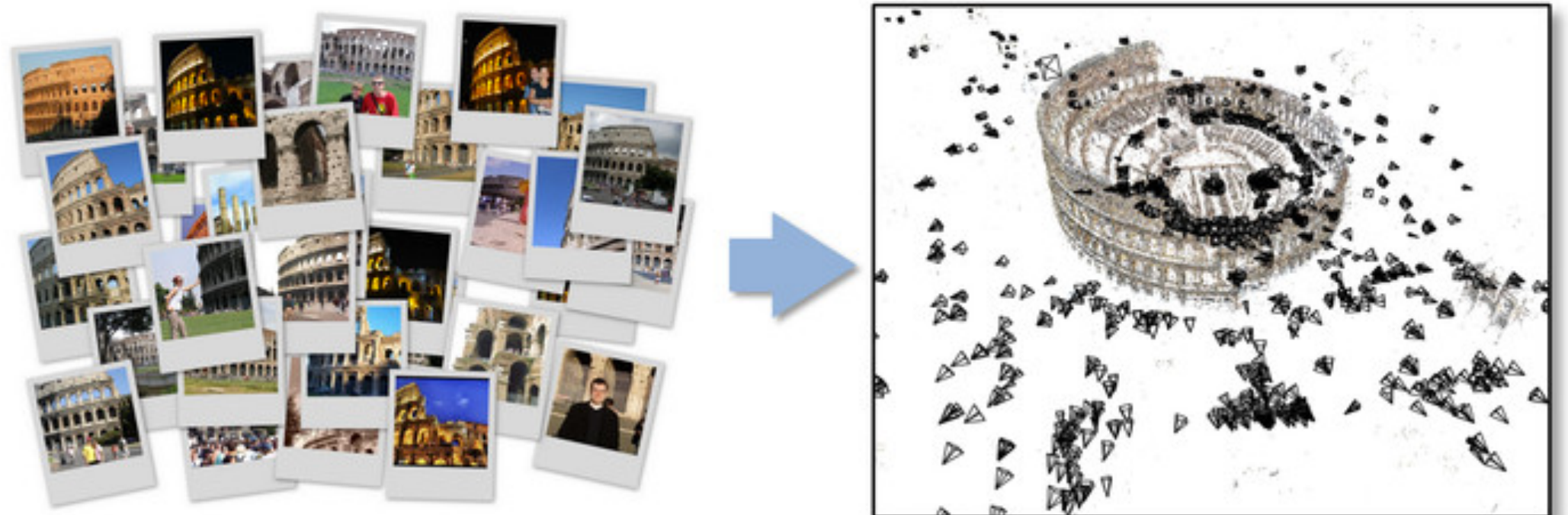
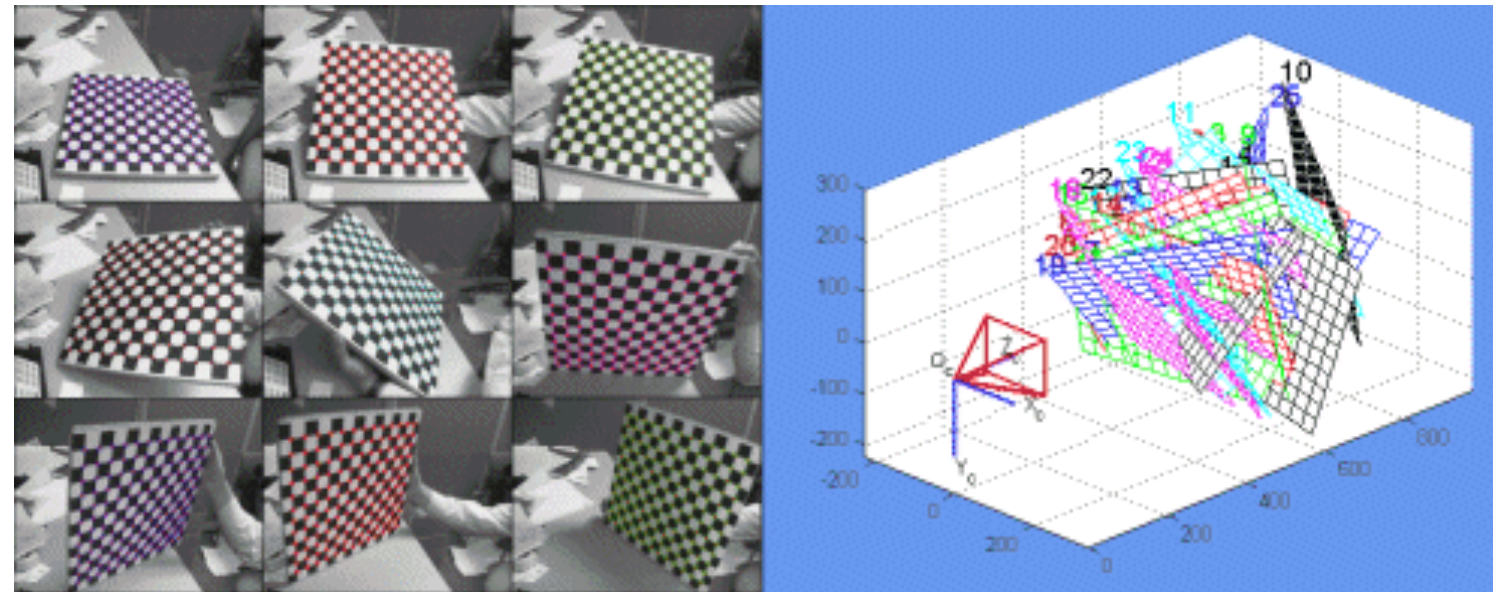
(http://www.vision.caltech.edu/bouguetj/calib_doc/)

- Print a checkerboard
- Capture multiple images
 - ✓ Patterns should be static or use synchronized cameras
- Run the calibration toolbox
 - ✓ Input: Patterns captured at the same time
 - ✓ Output: **K,R,t** for each camera

Calibration

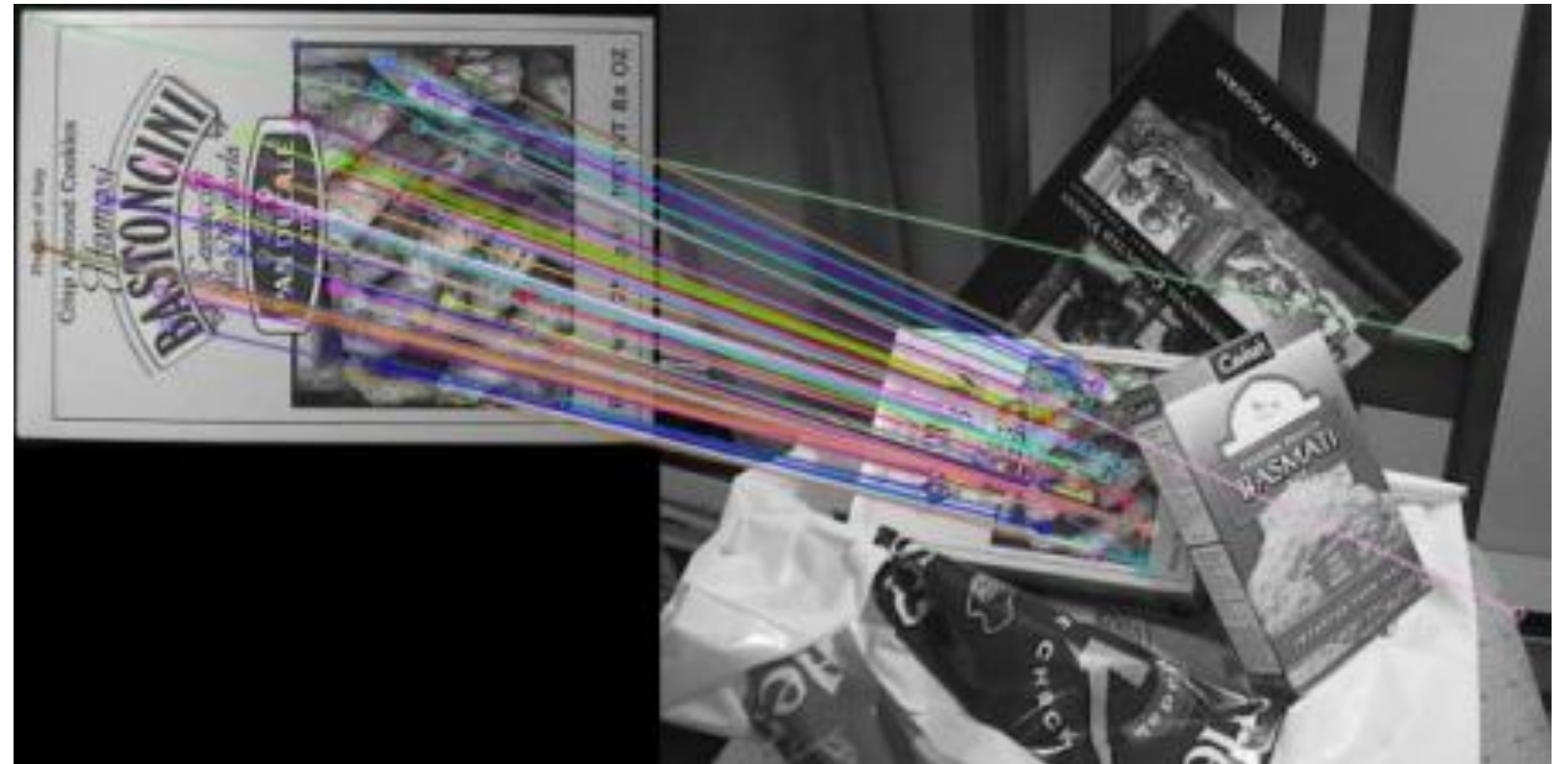
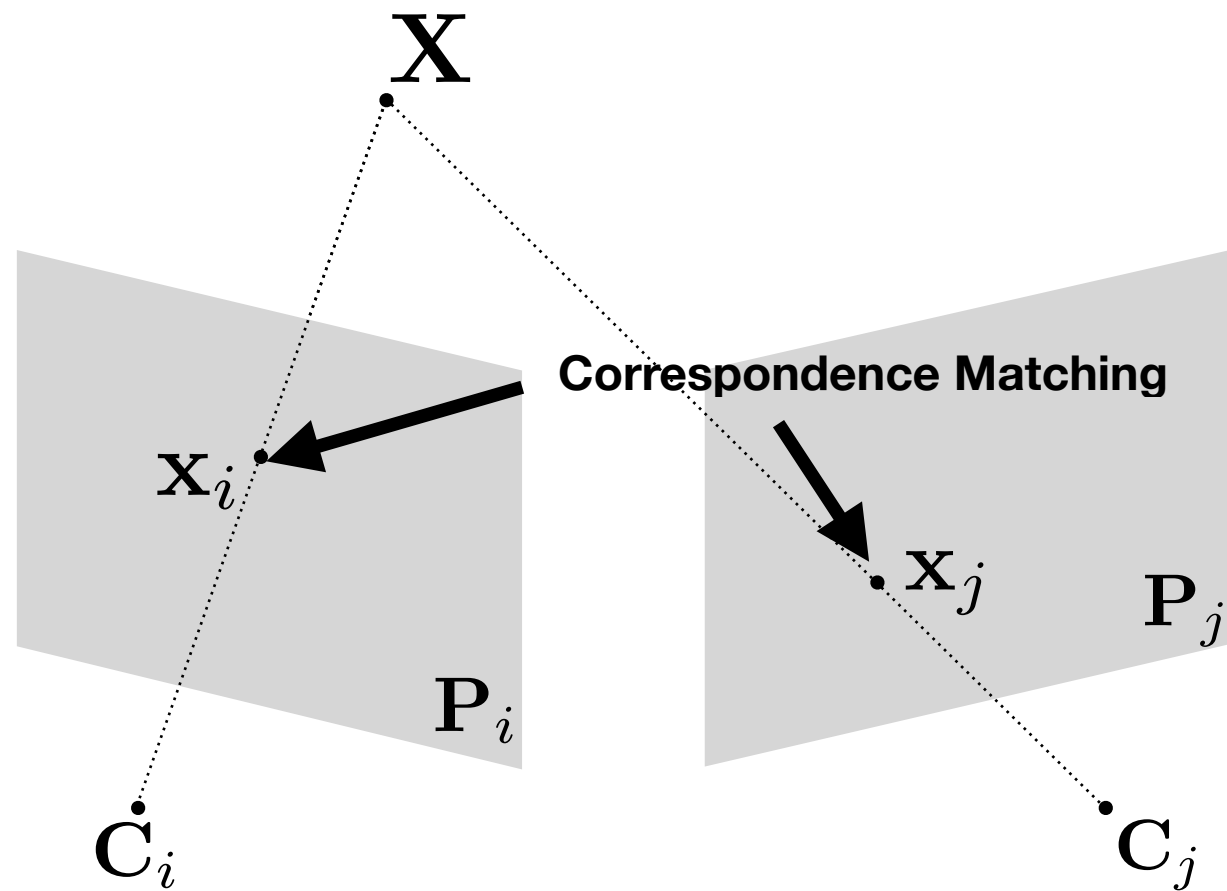
Why Checkerboard?

For more accurate correspondence matching!



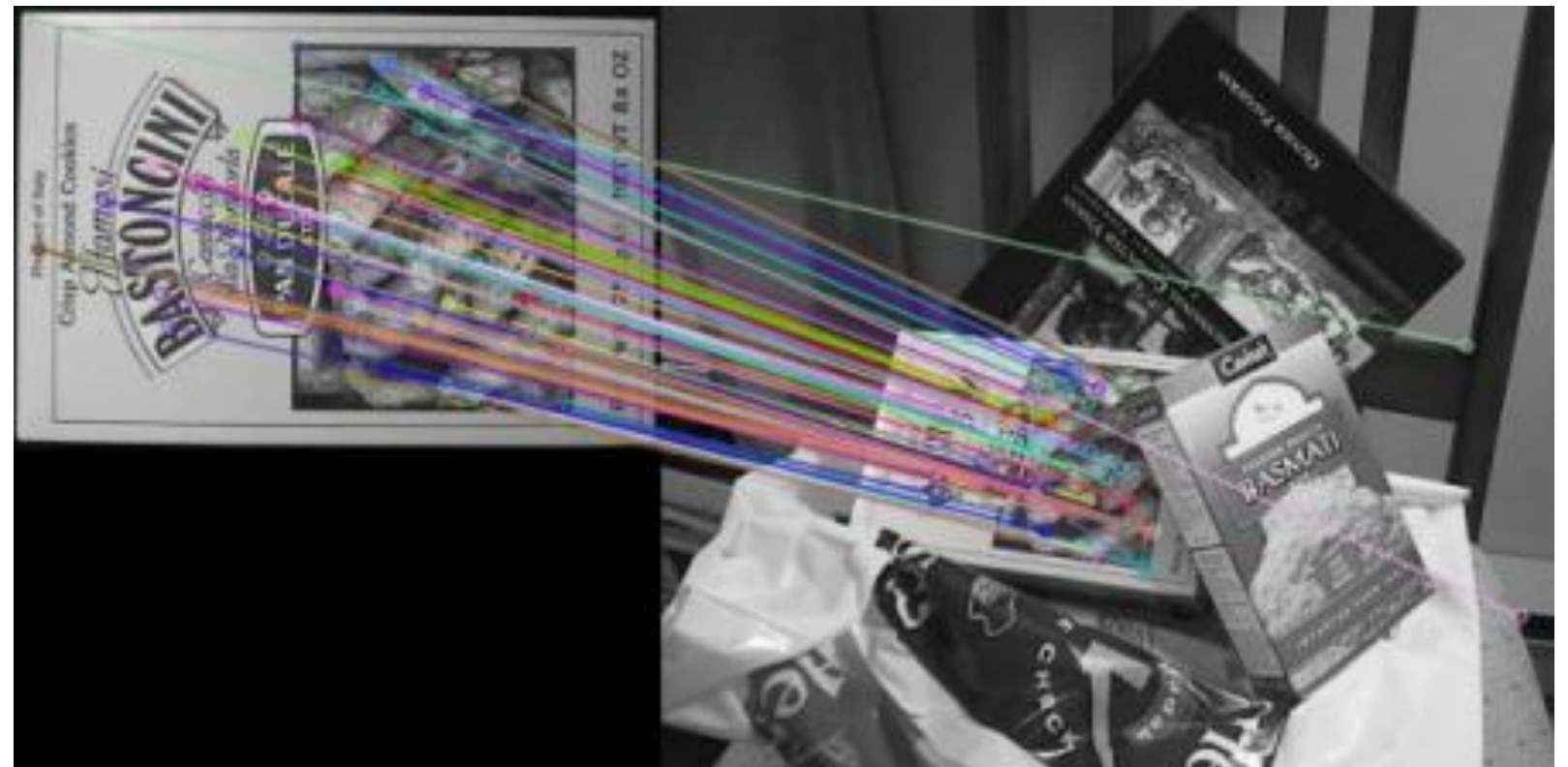
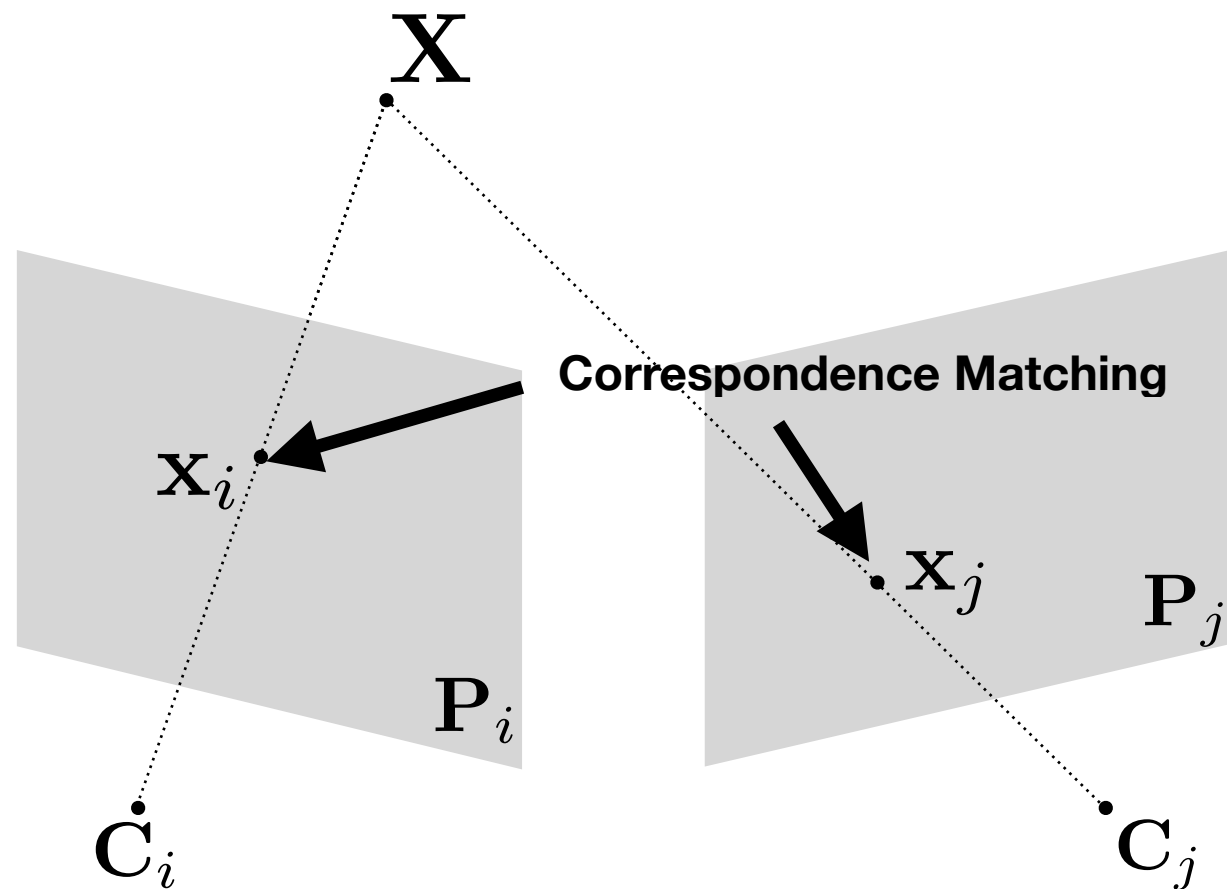
Correspondence Matching

SSD, NCC, SIFT, SURF, or Recent Deep Learning Methods



Correspondence Matching

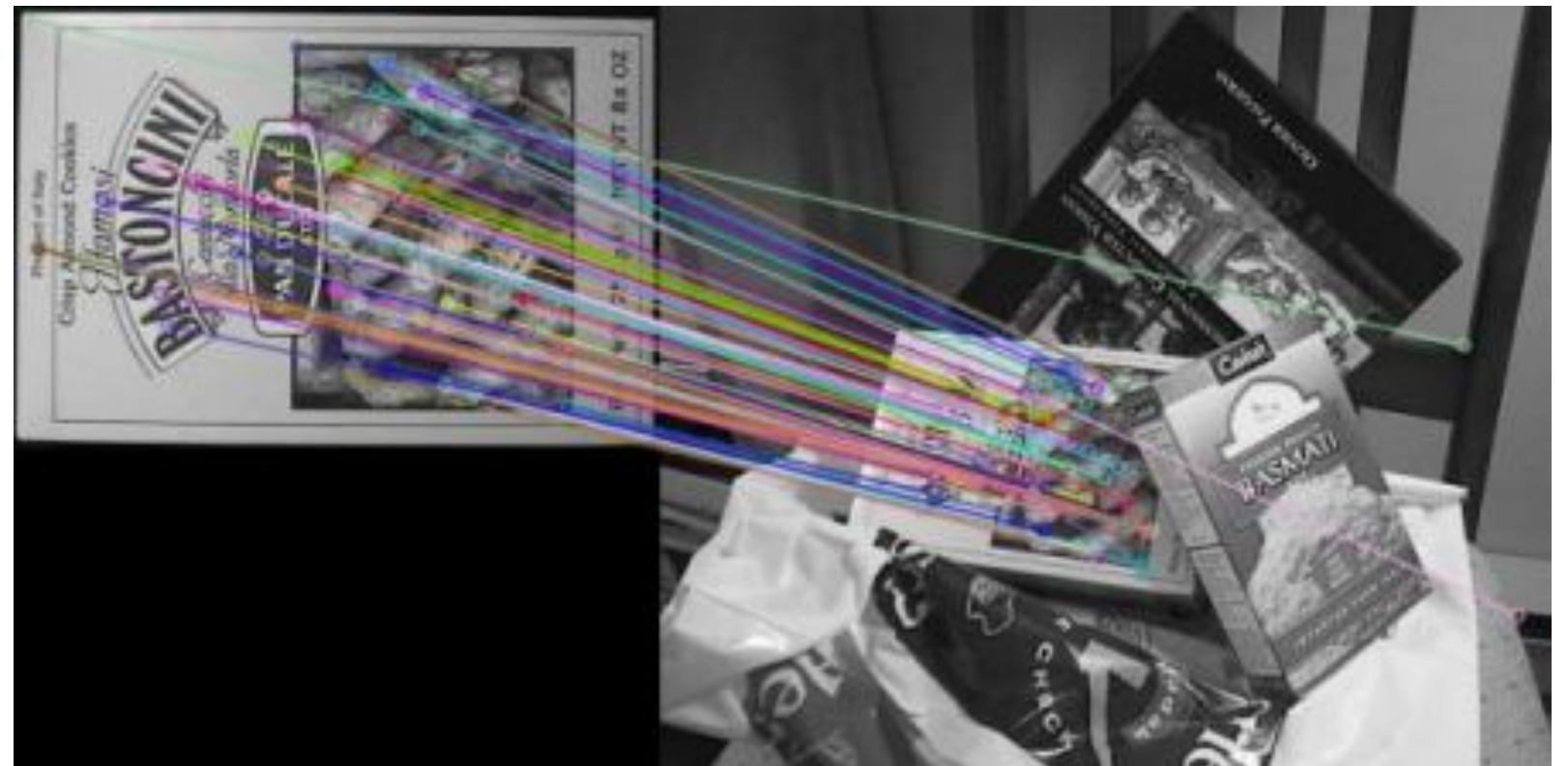
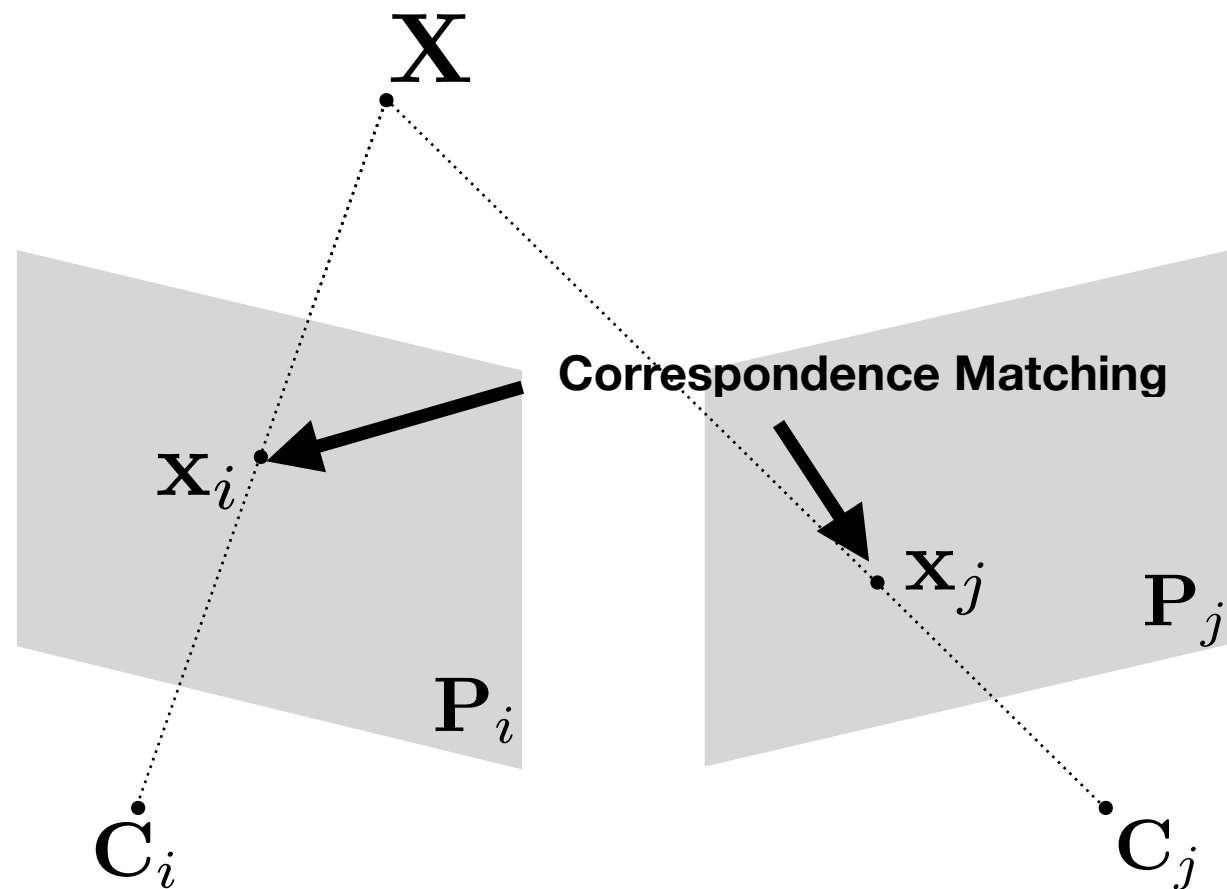
SSD, NCC, SIFT, SURF, or Recent Deep Learning Methods



How can we handle outliers?

Correspondence Matching

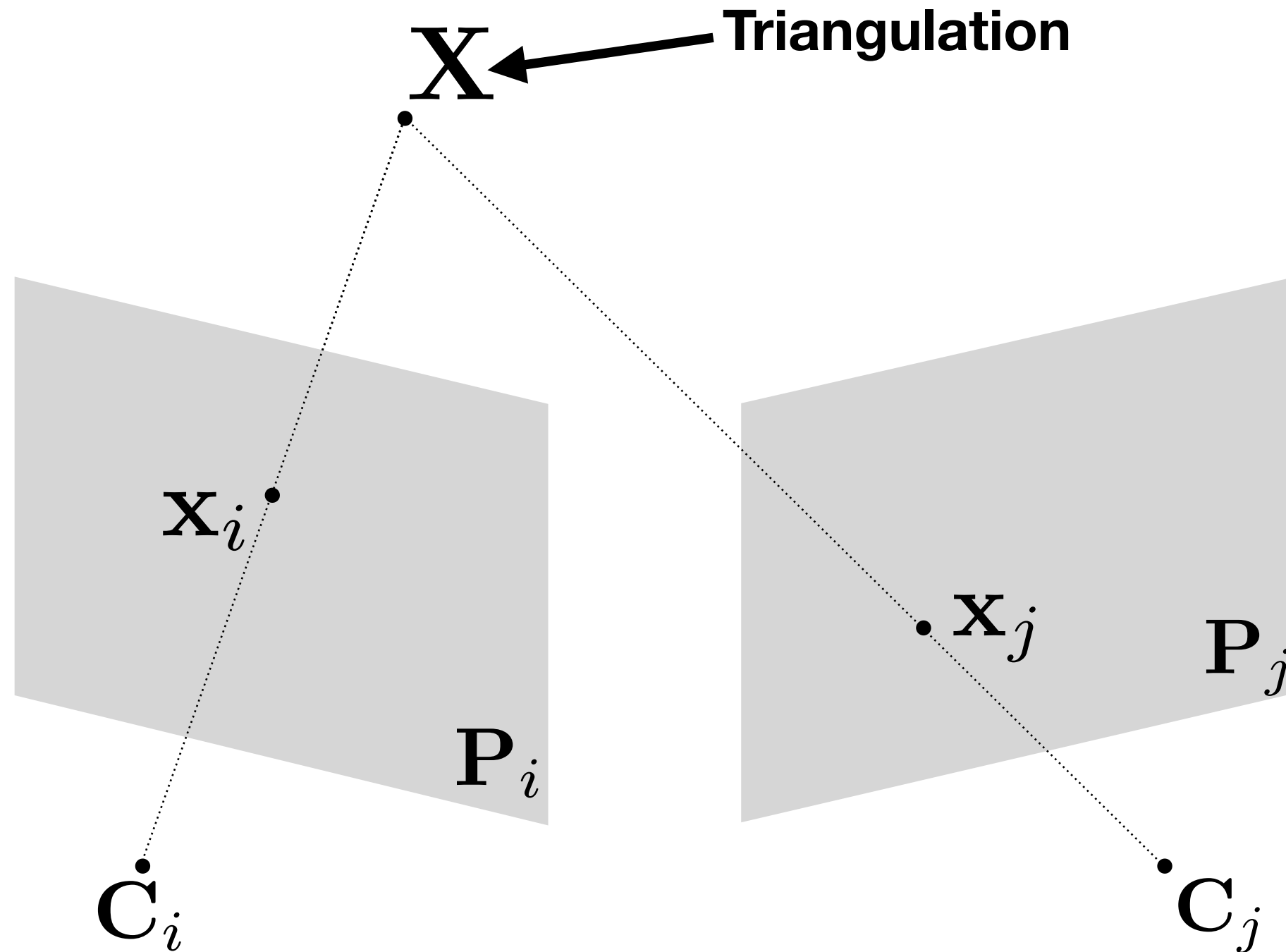
SSD, NCC, SIFT, SURF, or Recent Deep Learning Methods



How can we handle outliers?
RANSAC (will talk later)

Triangulation

Reconstructing A 3D Point from Image Measurements



Given by matching

Given by calibration

Using the fact that the cross product should be zero

$$\mathbf{x} \times \mathbf{P} \mathbf{X} = \mathbf{0}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{p}_1^\top \mathbf{X} \\ \mathbf{p}_2^\top \mathbf{X} \\ \mathbf{p}_3^\top \mathbf{X} \end{bmatrix} = \begin{bmatrix} y \mathbf{p}_3^\top \mathbf{X} - \mathbf{p}_2^\top \mathbf{X} \\ \mathbf{p}_1^\top \mathbf{X} - x \mathbf{p}_3^\top \mathbf{X} \\ x \mathbf{p}_2^\top \mathbf{X} - y \mathbf{p}_1^\top \mathbf{X} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Third line is a linear combination of the first and second lines.
(x times the first line plus y times the second line)

One 2D to 3D point correspondence give you 2 equations

$$\begin{bmatrix} yp_3^\top X - p_2^\top X \\ p_1^\top X - xp_3^\top X \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} yp_3^\top - p_2^\top \\ p_1^\top - xp_3^\top \end{bmatrix} \boxed{X} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$A_i \boxed{X} = \mathbf{0}$$

Now we can make a system of linear equations
(two lines for each 2D point correspondence)

Concatenate the 2D points from both images

$$\begin{bmatrix} y\mathbf{p}_3^\top - \mathbf{p}_2^\top \\ \mathbf{p}_1^\top - x\mathbf{p}_3^\top \\ y'\mathbf{p}'_3{}^\top - \mathbf{p}'_2{}^\top \\ \mathbf{p}'_1{}^\top - x'\mathbf{p}'_3{}^\top \end{bmatrix} \mathbf{X} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{A} \mathbf{X} = \mathbf{0}$$

How do we solve homogeneous linear system?

S V D !

Recall: Total least squares

(Warning: change of notation. \mathbf{x} is a vector of parameters!)

$$\begin{aligned} E_{\text{TLS}} &= \sum_i (\mathbf{a}_i \mathbf{x})^2 \\ &= \|\mathbf{A}\mathbf{x}\|^2 && \text{(matrix form)} \\ \|\mathbf{x}\|^2 &= 1 && \text{constraint} \end{aligned}$$

$$\begin{array}{l} \text{minimize } \|\mathbf{A}\mathbf{x}\|^2 \\ \text{subject to } \|\mathbf{x}\|^2 = 1 \end{array} \quad \rightarrow \quad \begin{array}{l} \text{minimize } \frac{\|\mathbf{A}\mathbf{x}\|^2}{\|\mathbf{x}\|^2} \\ \text{(Rayleigh quotient)} \end{array}$$


Solution is the eigenvector
corresponding to smallest eigenvalue of

$$\mathbf{A}^\top \mathbf{A}$$

Recall: Total least squares

(Warning: change of notation. \mathbf{x} is a vector of parameters!)

$$\begin{aligned} E_{\text{TLS}} &= \sum_i (\mathbf{a}_i \mathbf{x})^2 \\ &= \|\mathbf{A}\mathbf{x}\|^2 && \text{(matrix form)} \\ \|\mathbf{x}\|^2 &= 1 && \text{constraint} \end{aligned}$$

minimize $\|\mathbf{A}\mathbf{x}\|^2$
subject to $\|\mathbf{x}\|^2 = 1$  minimize $\frac{\|\mathbf{A}\mathbf{x}\|^2}{\|\mathbf{x}\|^2}$
(Rayleigh quotient)

Solution is the eigenvector
corresponding to smallest eigenvalue of

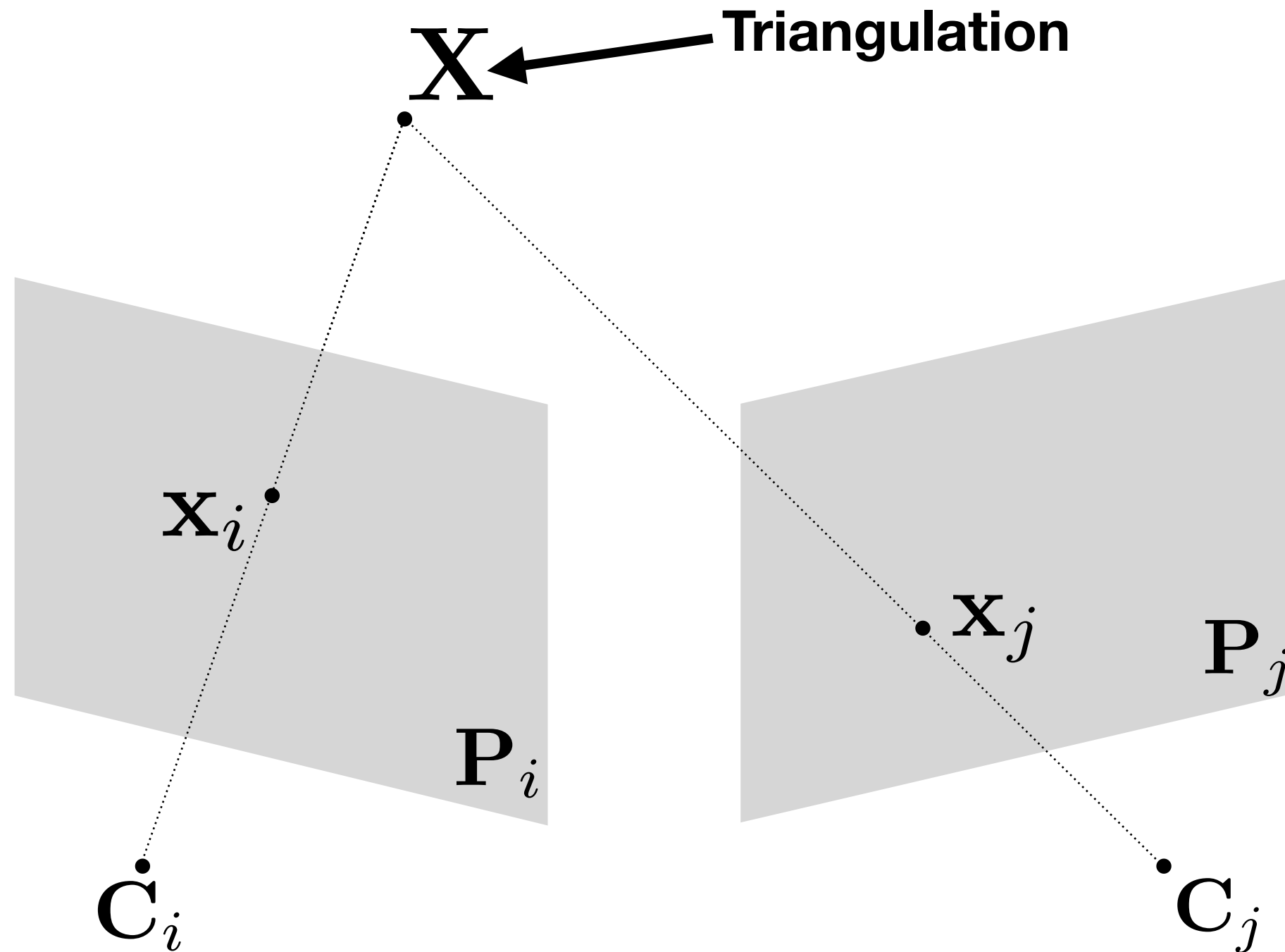
$$\mathbf{A}^\top \mathbf{A}$$

Analytical solution:

Best solution in this cost function,
but may not be the best **geometrically**

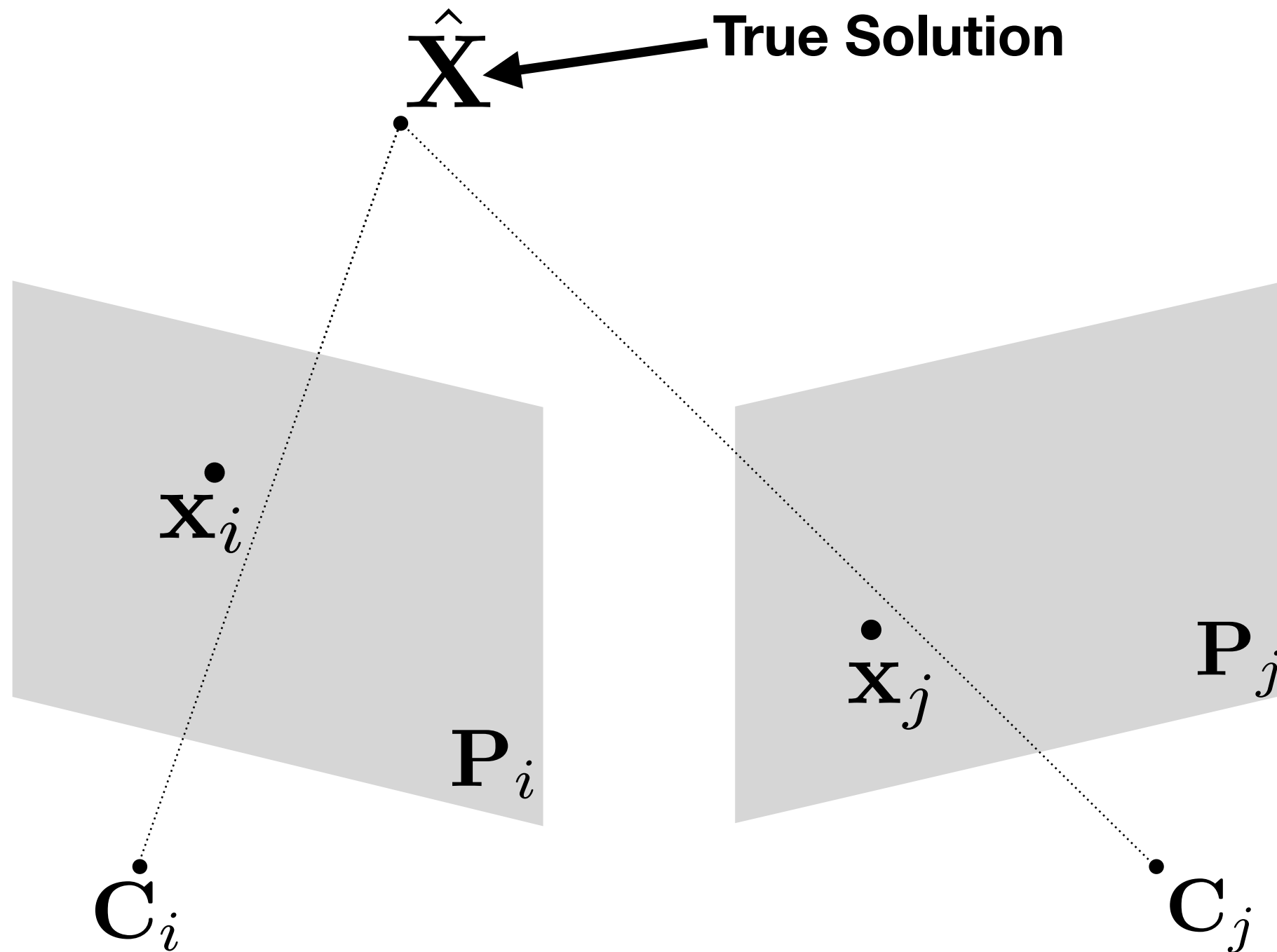
Triangulation

A Geometrical Solution



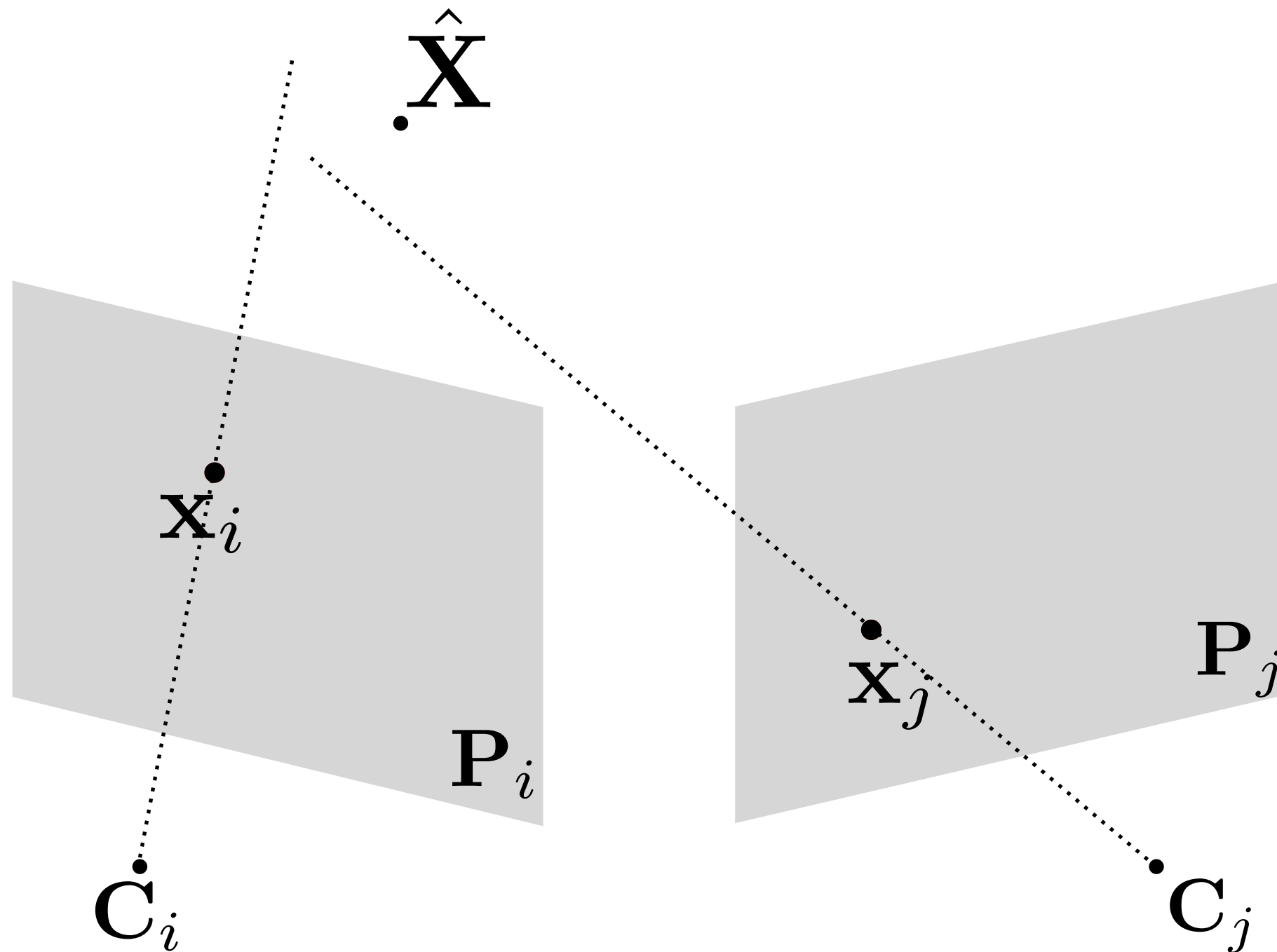
Triangulation

A Geometrical Solution



Triangulation

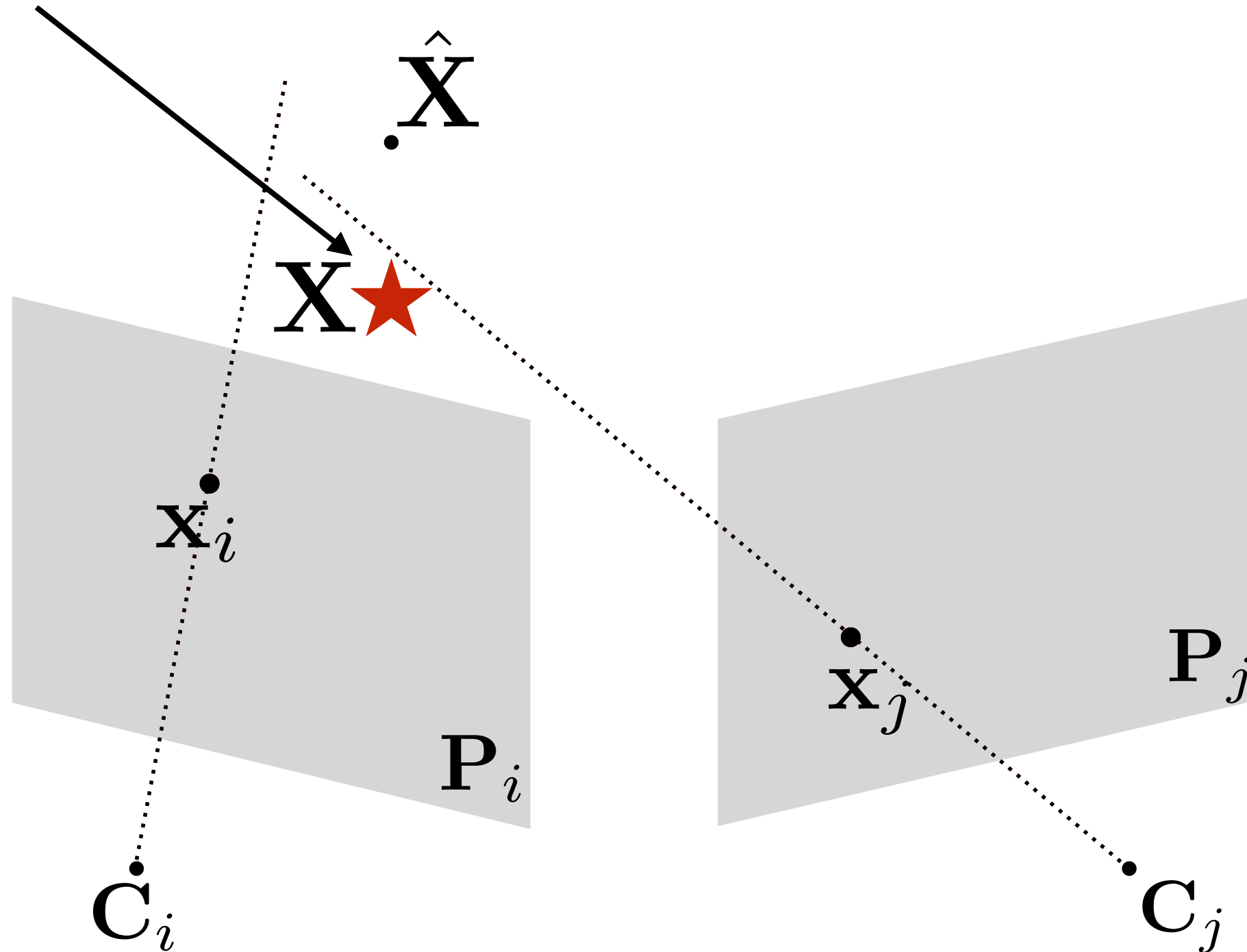
A Geometrical Solution



Triangulation

A Geometrical Solution

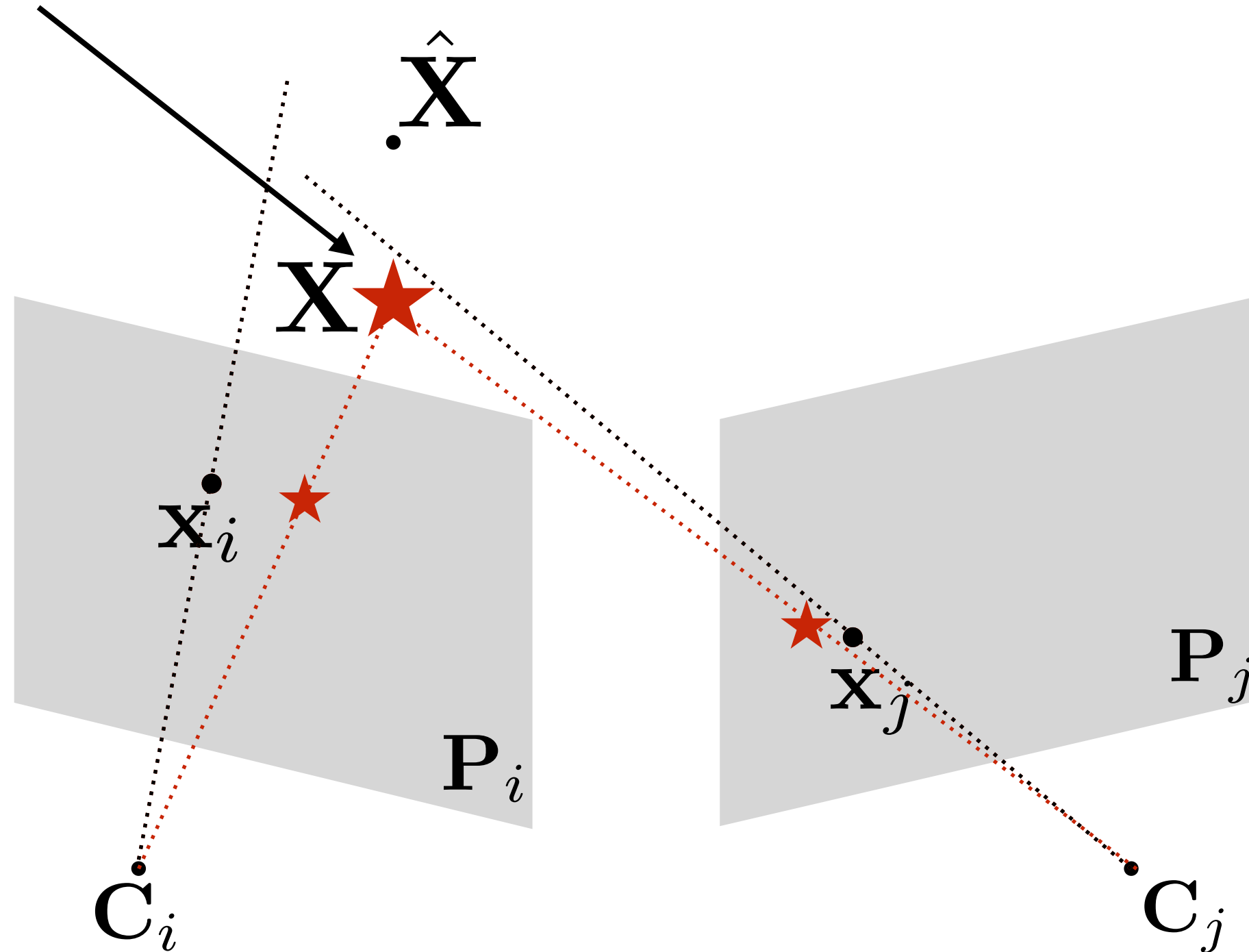
Triangulation



Triangulation

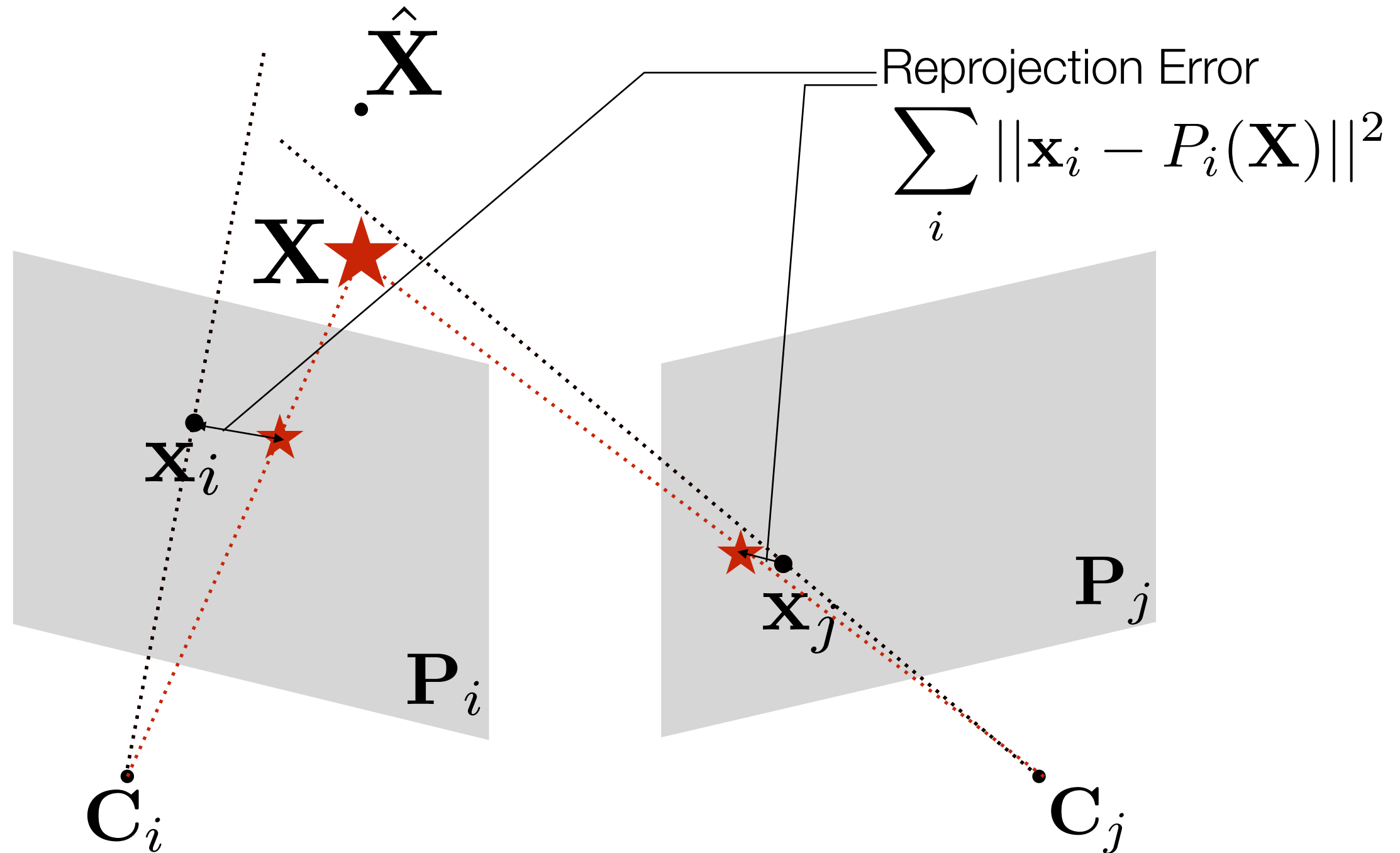
A Geometrical Solution

Triangulation



Triangulation

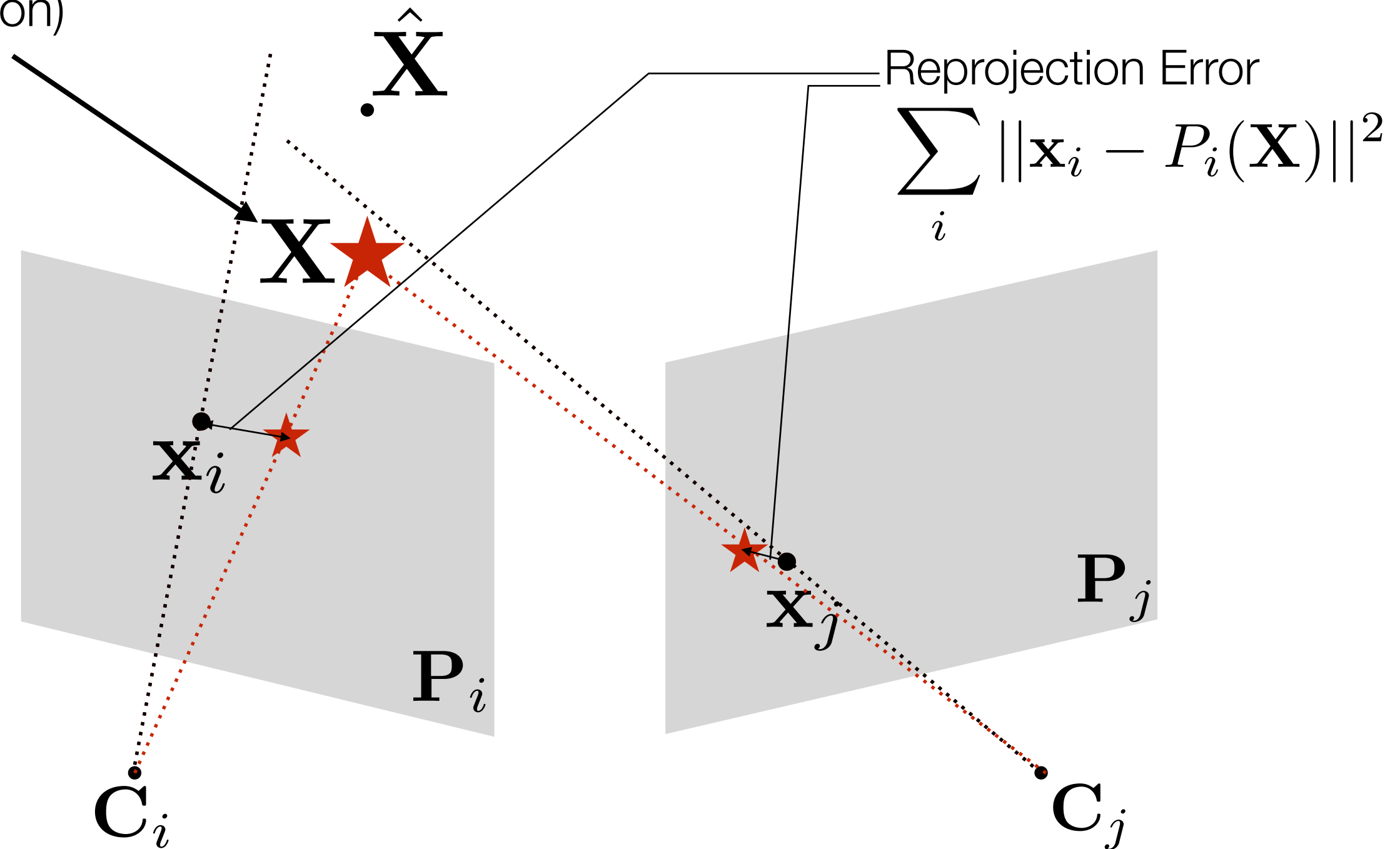
A Geometrical Solution



Triangulation

A Geometrical Solution

Move \mathbf{X} to minimize
reprojection error
(Nonlinear optimization)



Triangulation

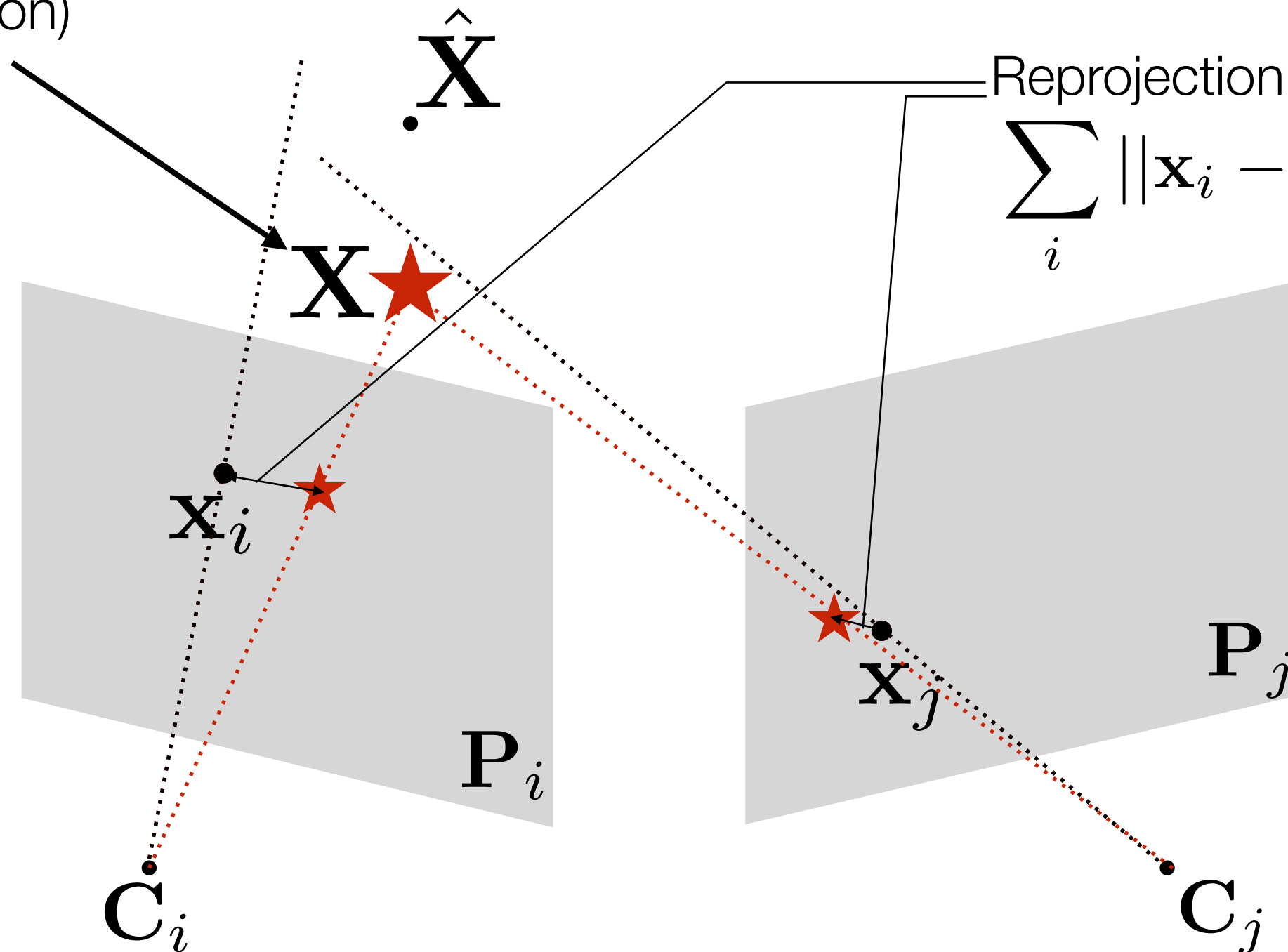
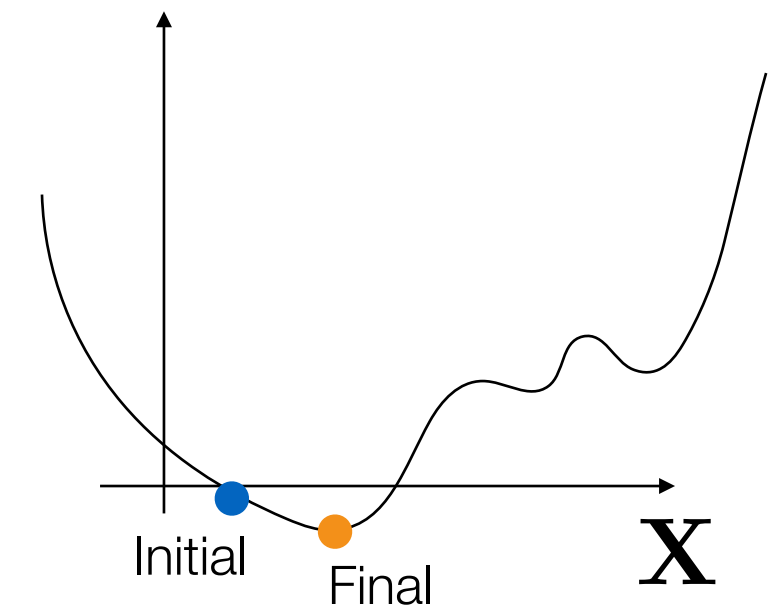
A Geometrical Solution

Move \mathbf{X} to minimize reprojection error
(Nonlinear optimization)

Reprojection Error

$$\sum_i \|\mathbf{x}_i - P_i(\mathbf{X})\|^2$$

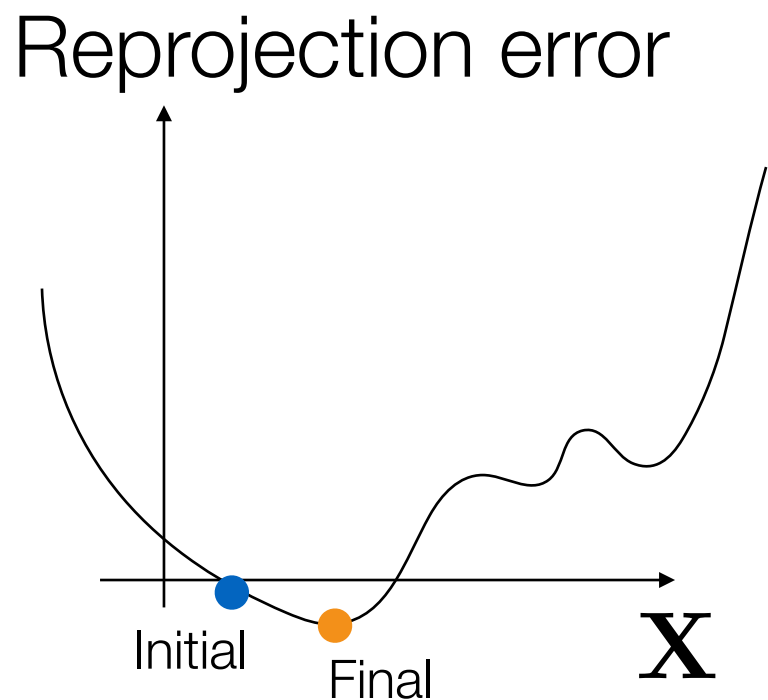
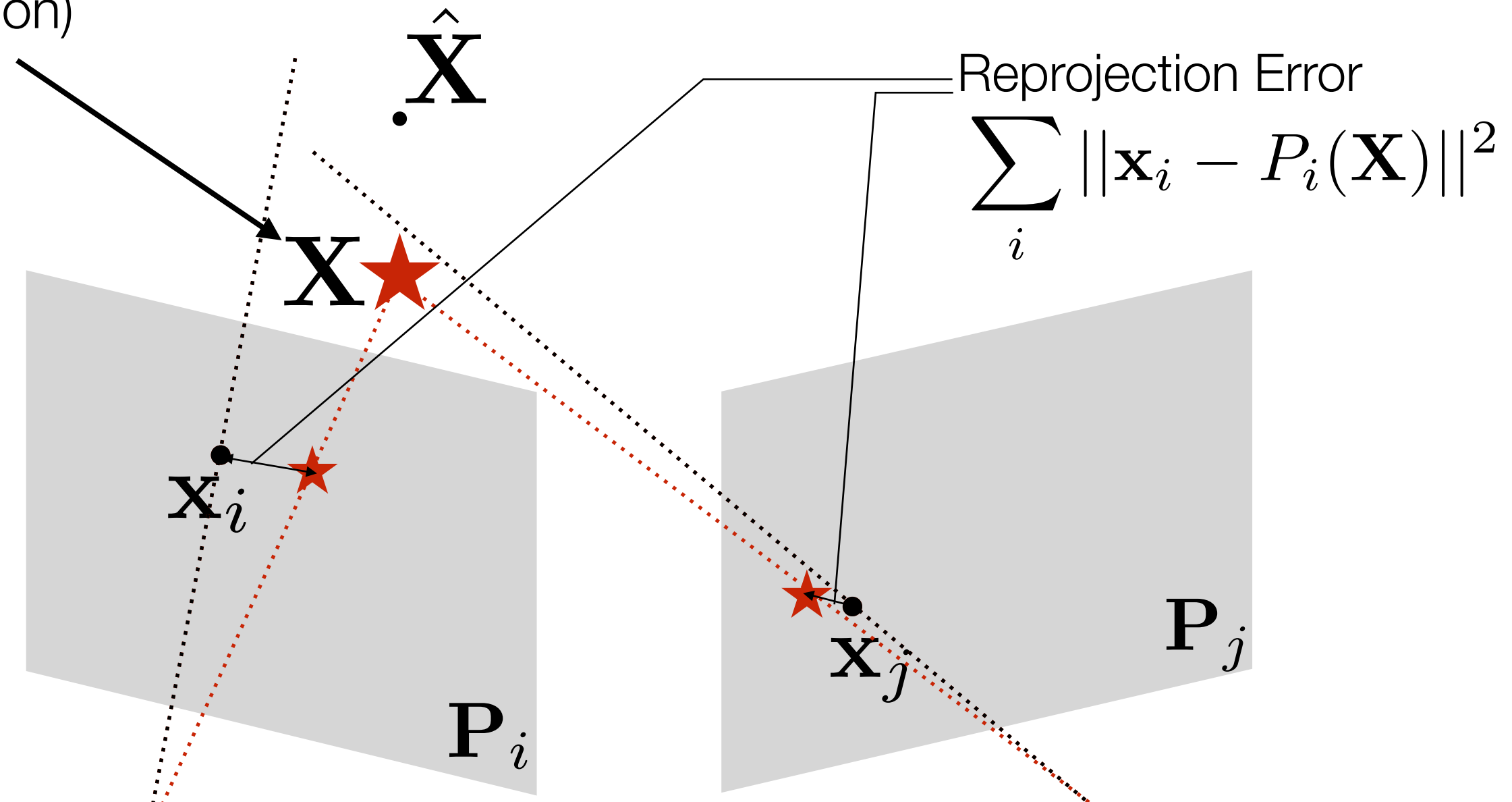
Reprojection error



Triangulation

A Geometrical Solution

Move \mathbf{X} to minimize reprojection error
(Nonlinear optimization)

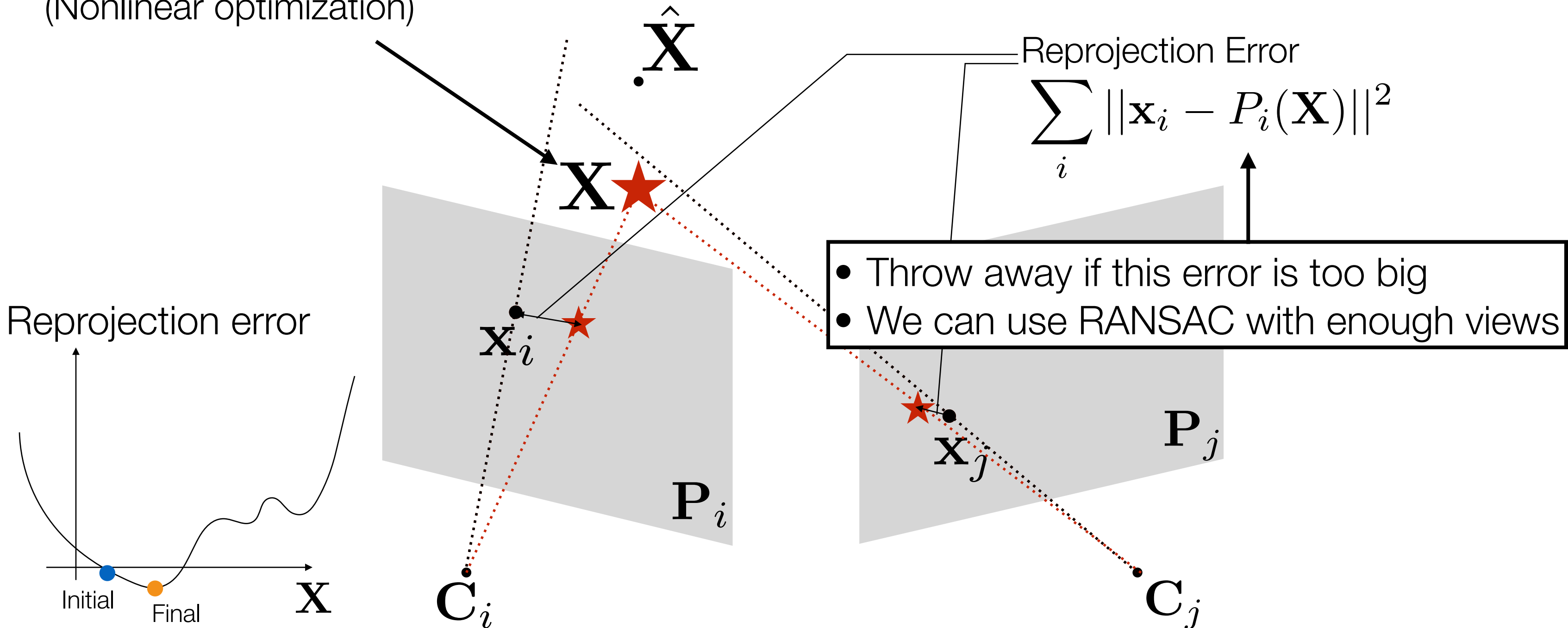


- Compute initial \mathbf{X} by SVD (fast)
- Find \mathbf{X} which minimizes the reprojection errors (slow)

Triangulation

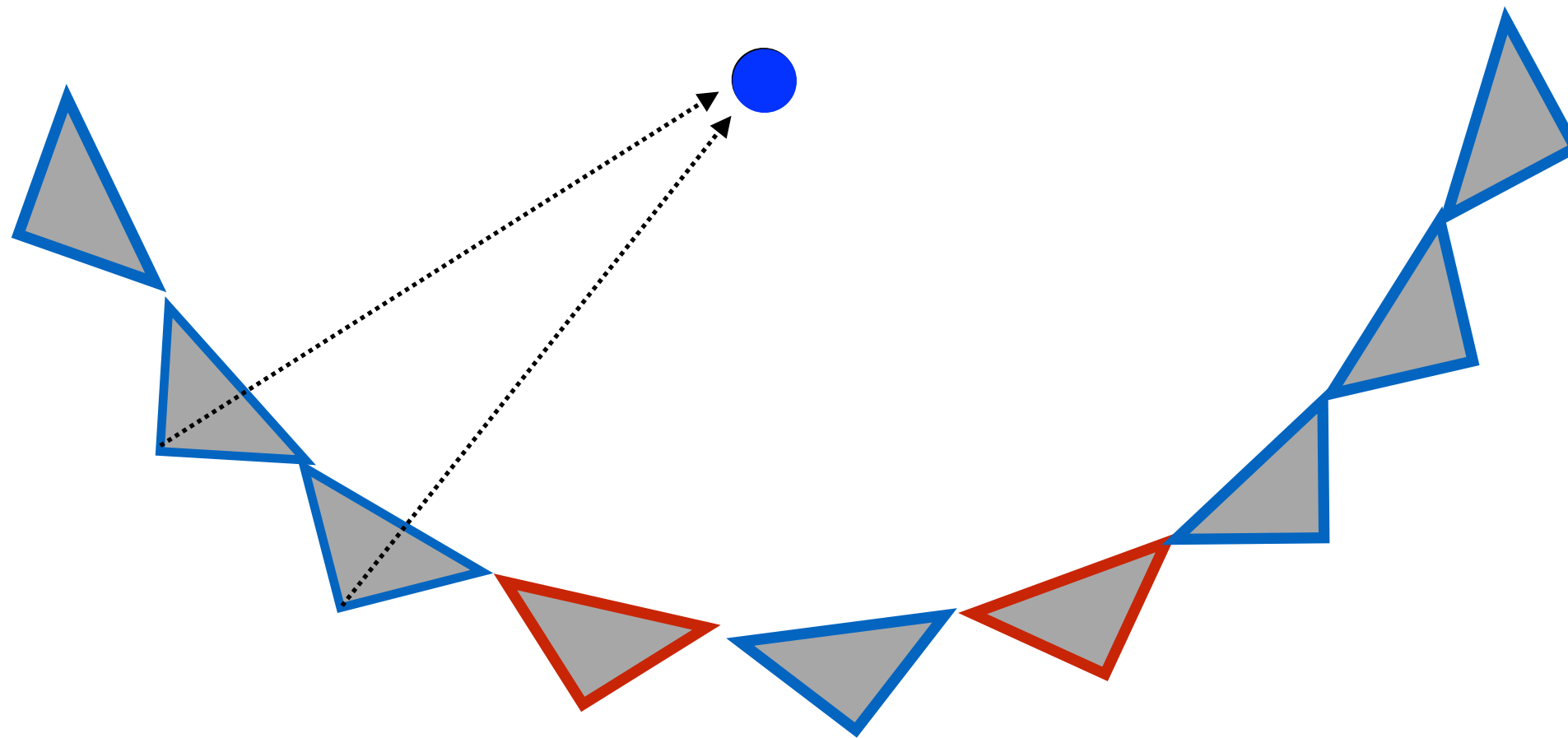
A Geometrical Solution

Move \mathbf{X} to minimize reprojection error
(Nonlinear optimization)



Triangulation with RANSAC

Outlier Filtering

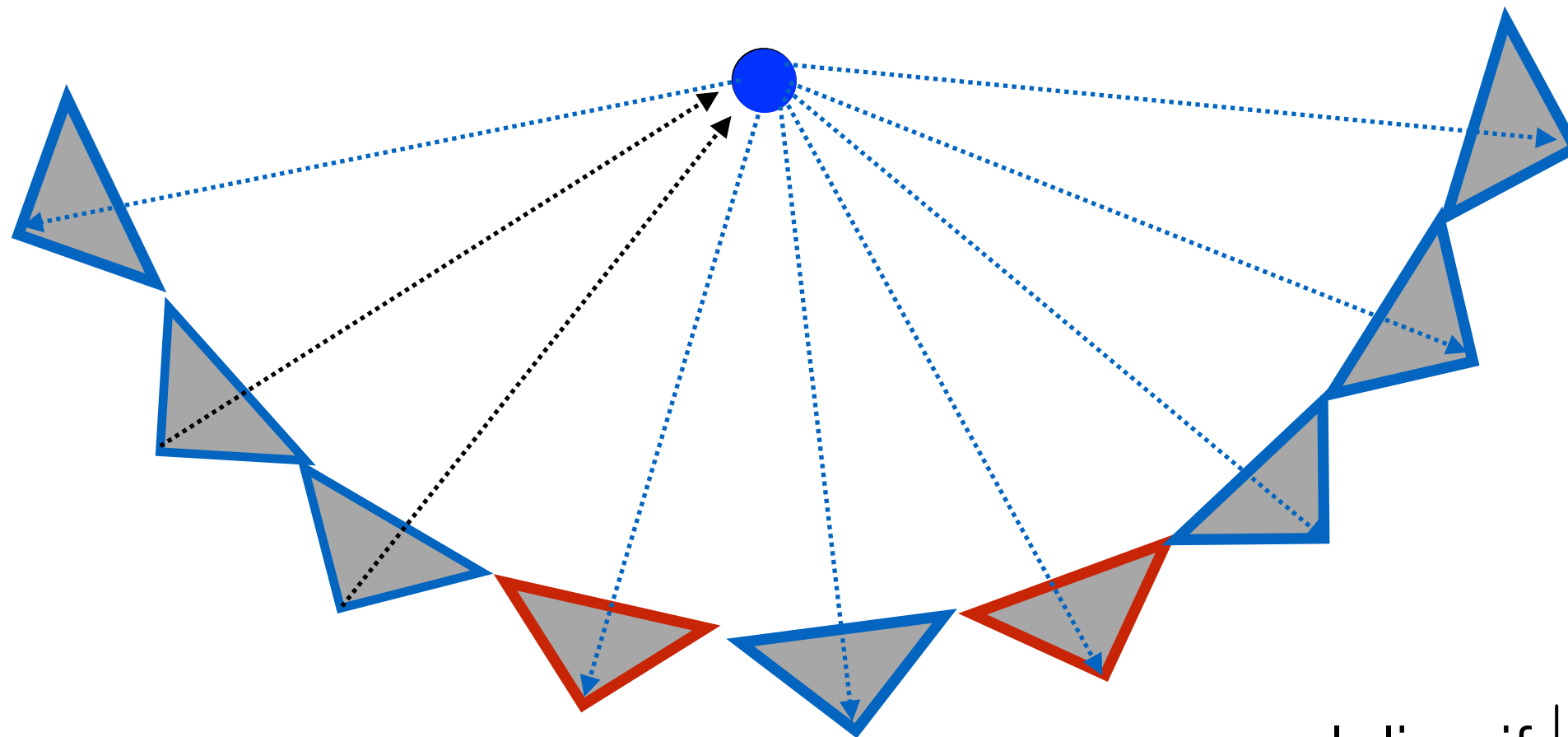


Blue: correct matching

Red: wrong matching

Triangulation with RANSAC

Outlier Filtering



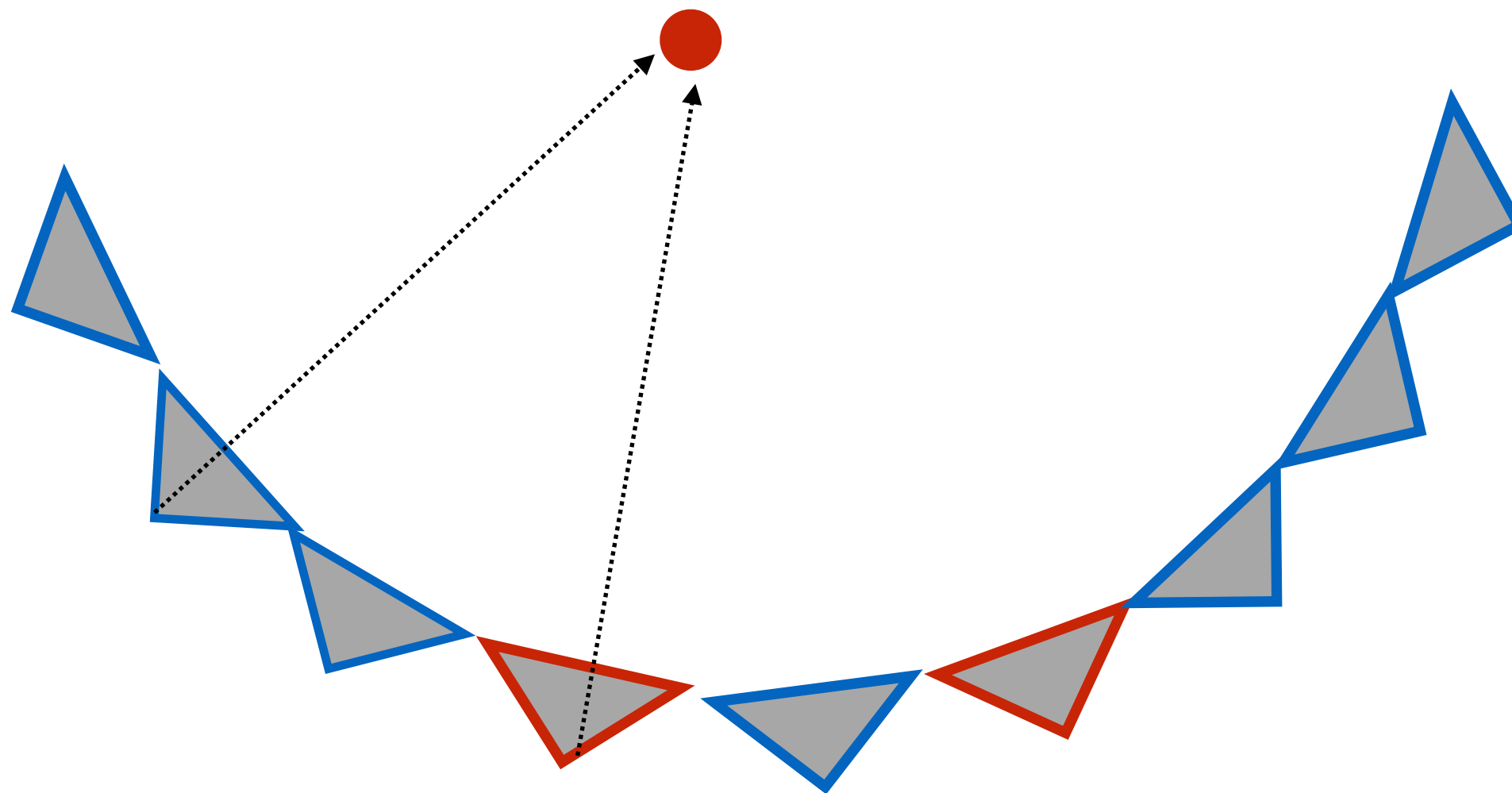
Blue: correct matching
Red: wrong matching

Inlier, if $\|\mathbf{x}_i - P_i(\mathbf{X})\|^2 < \tau$

We expect many inliers

Triangulation with RANSAC

Outlier Filtering with RANSAC

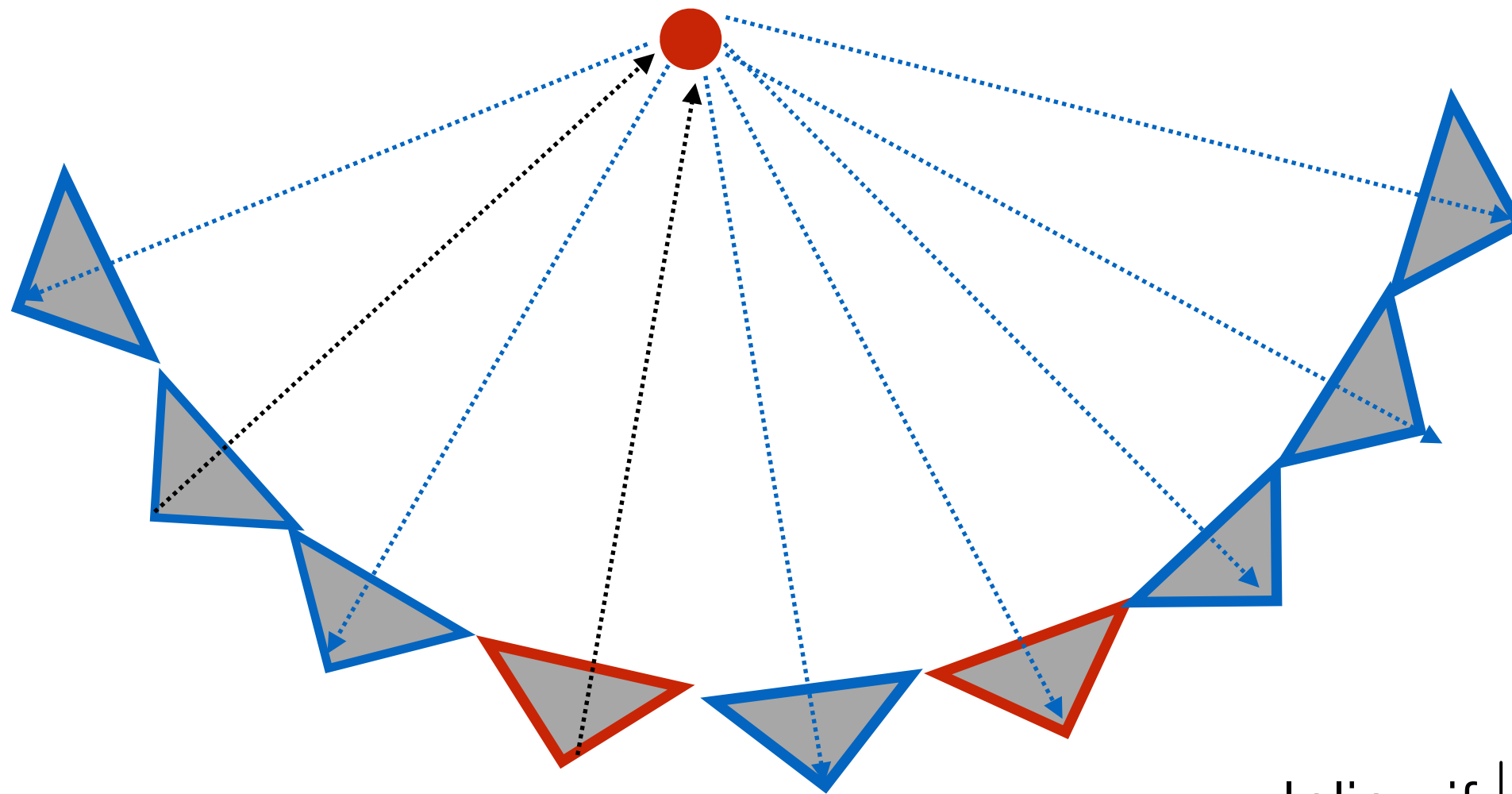


Blue: correct matching

Red: wrong matching

Triangulation with RANSAC

Outlier Filtering with RANSAC

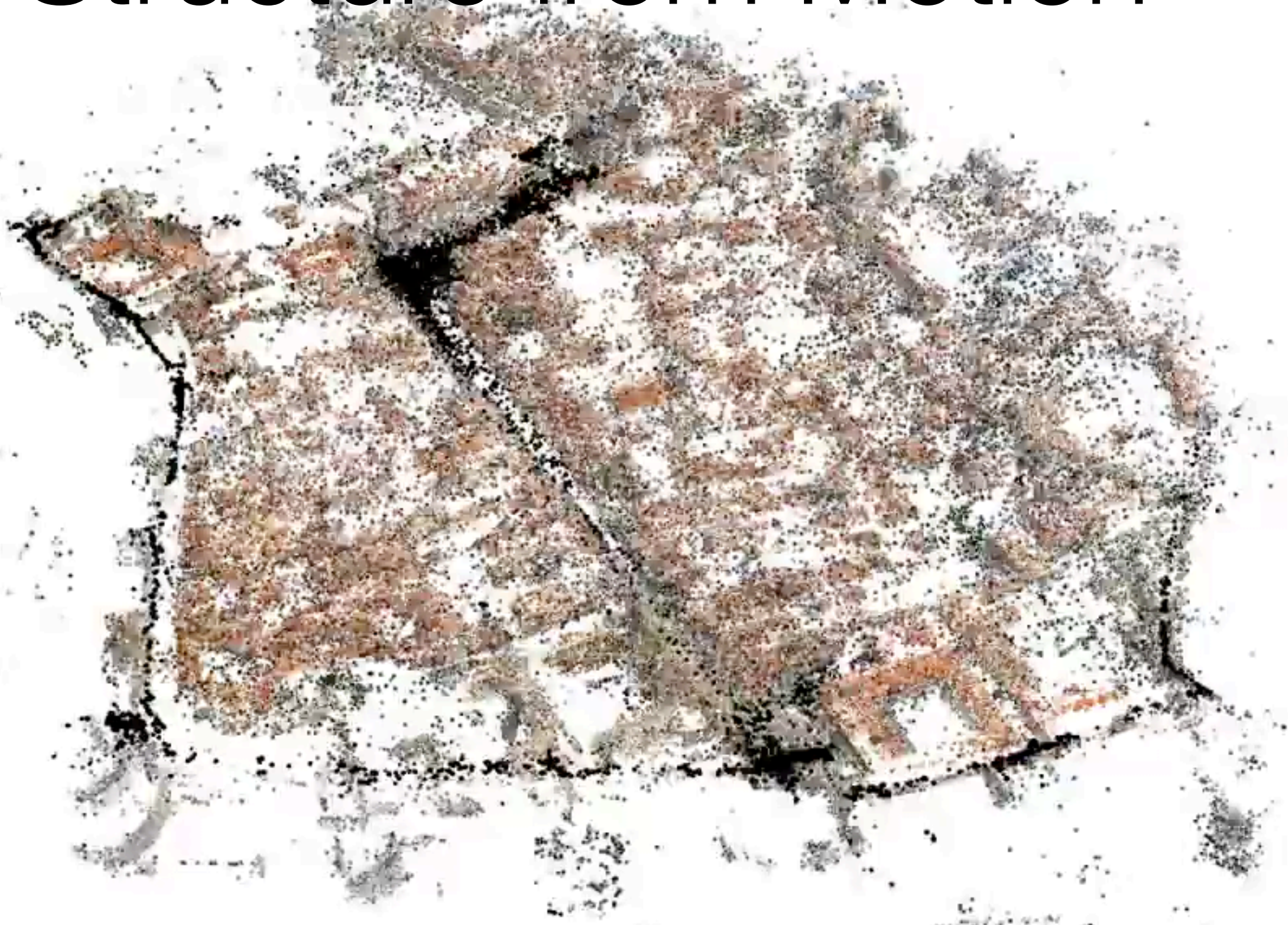


Blue: correct matching
Red: wrong matching

Inlier, if $\|\mathbf{x}_i - P_i(\mathbf{X})\|^2 < \tau$

We expect few inliers

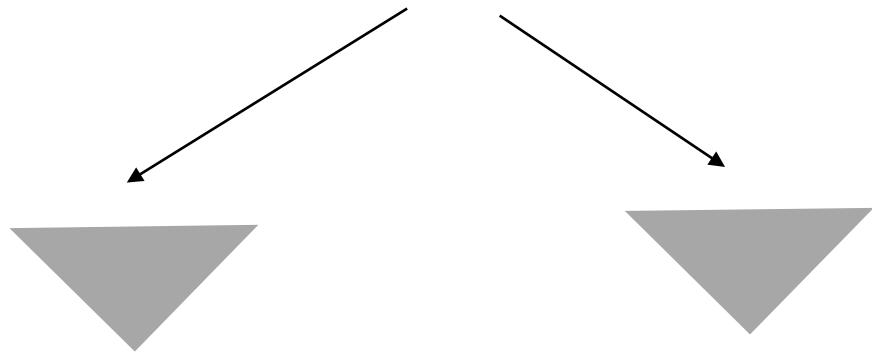
Structure from Motion



Structure from Motion

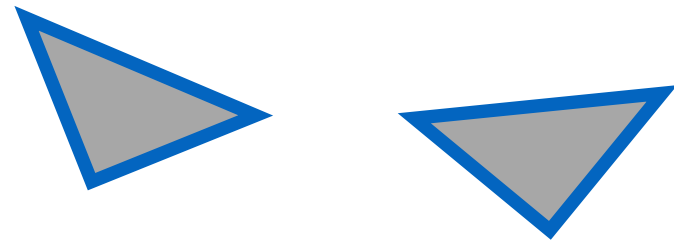
The Basic Idea

Find corresponding 2D points



Structure from Motion

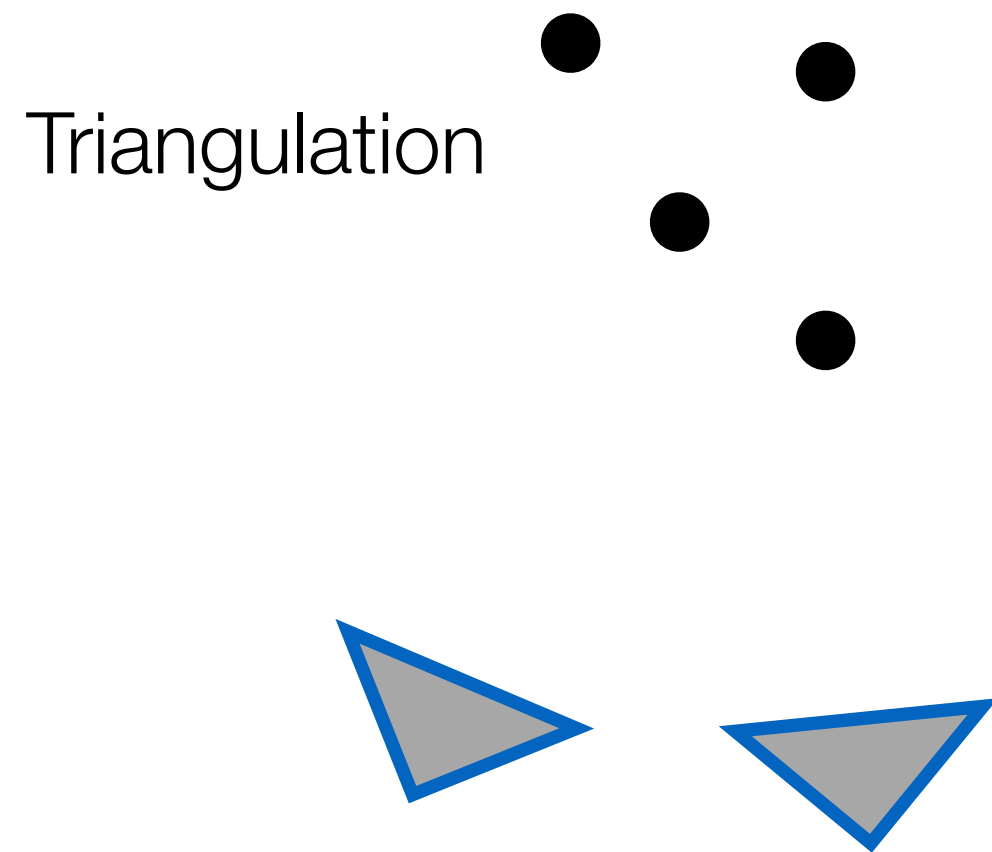
The Basic Idea



Camera pose estimation
by two-view geometry

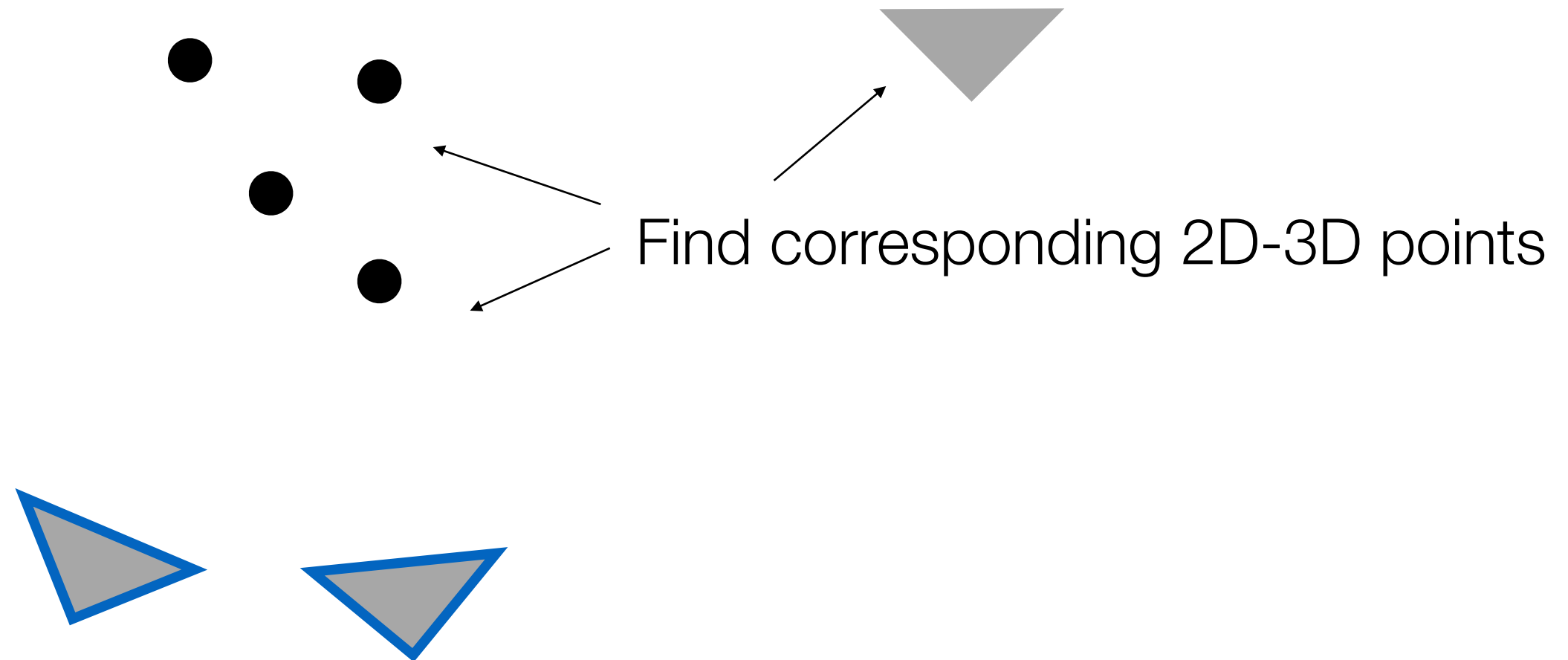
Structure from Motion

The Basic Idea



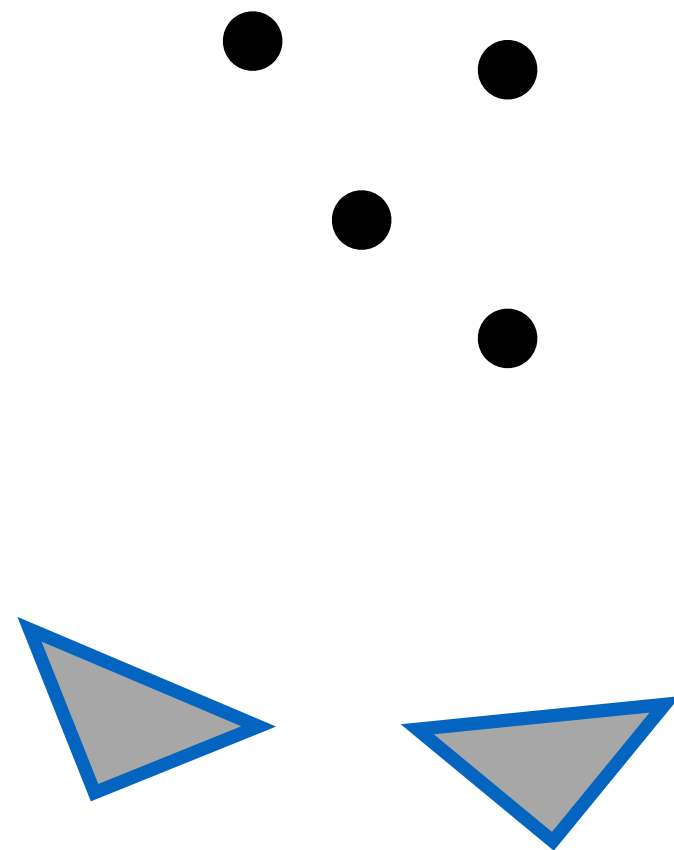
Structure from Motion

The Basic Idea



Structure from Motion

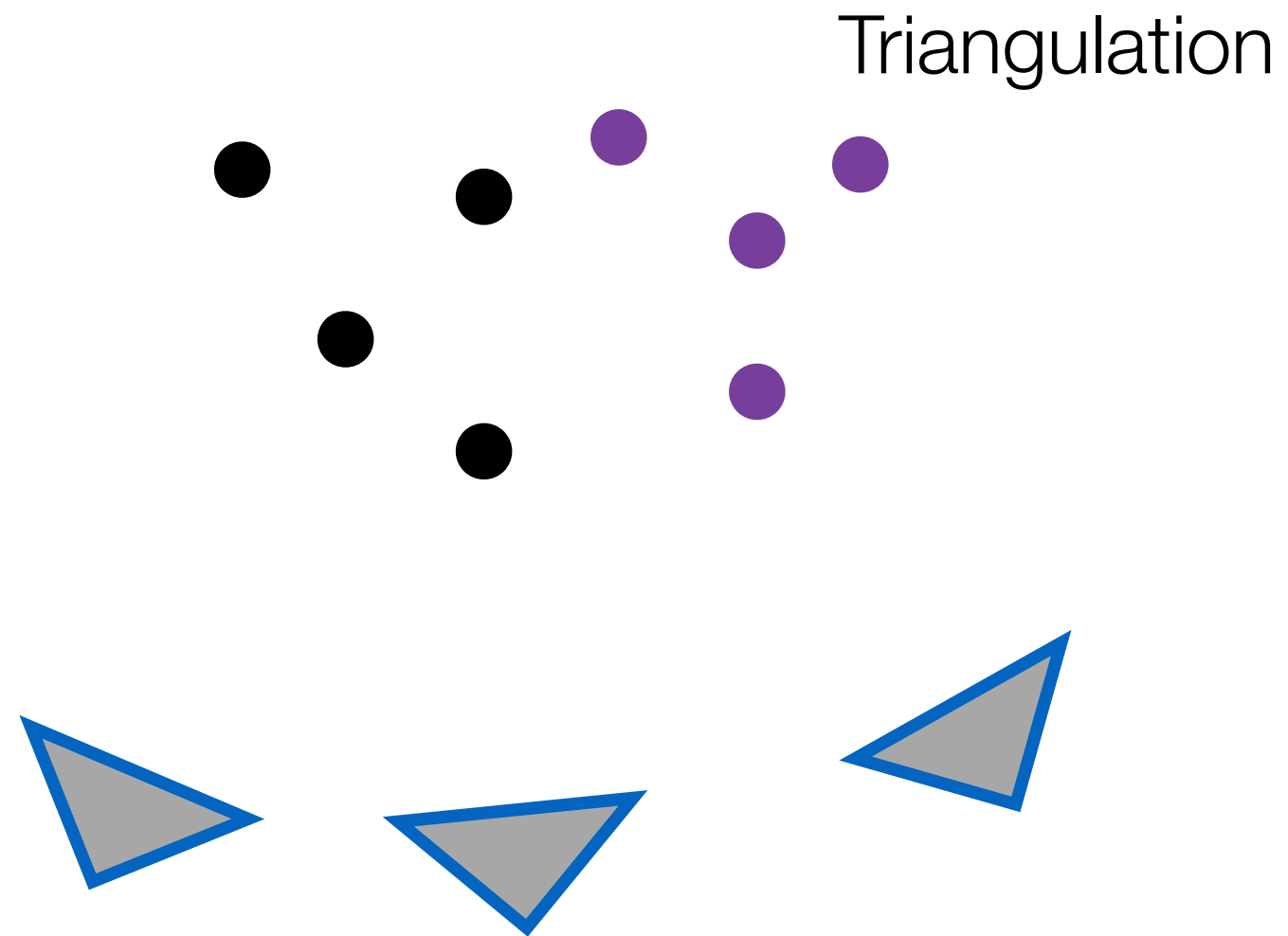
The Basic Idea



Camera pose estimation
by Perspective-n-Point Algorithm

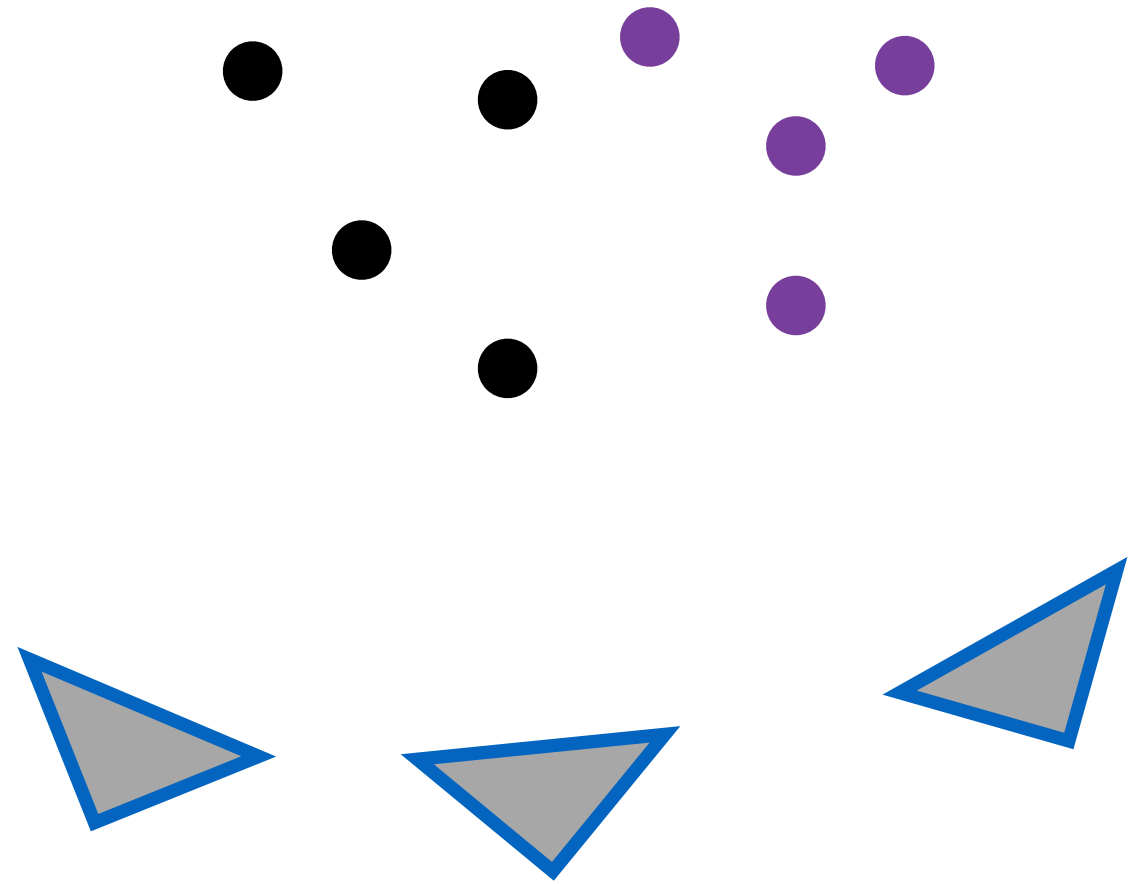
Structure from Motion

The Basic Idea



Structure from Motion

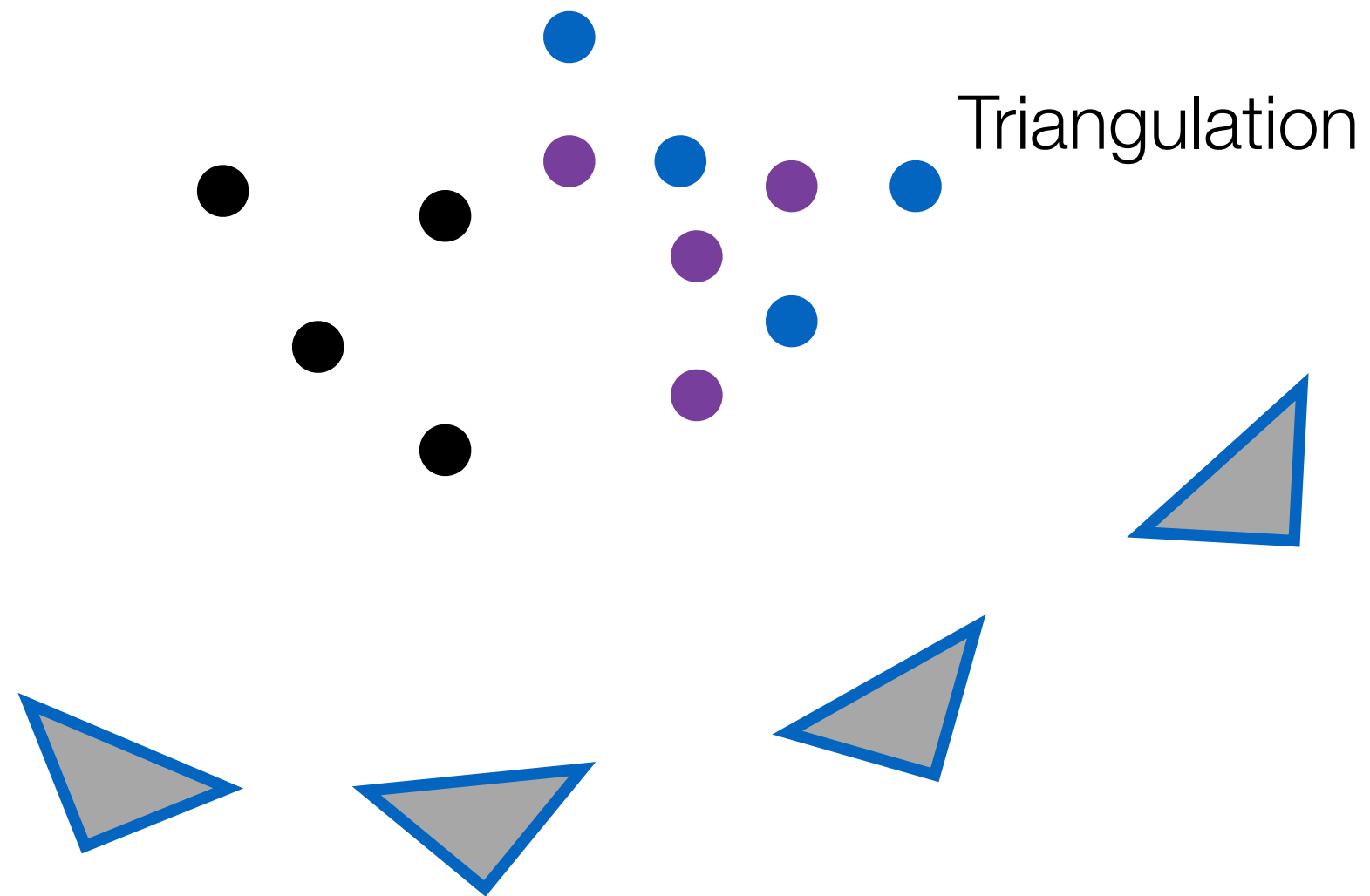
The Basic Idea



Camera pose estimation
by Perspective-n-Point Algorithm

Structure from Motion

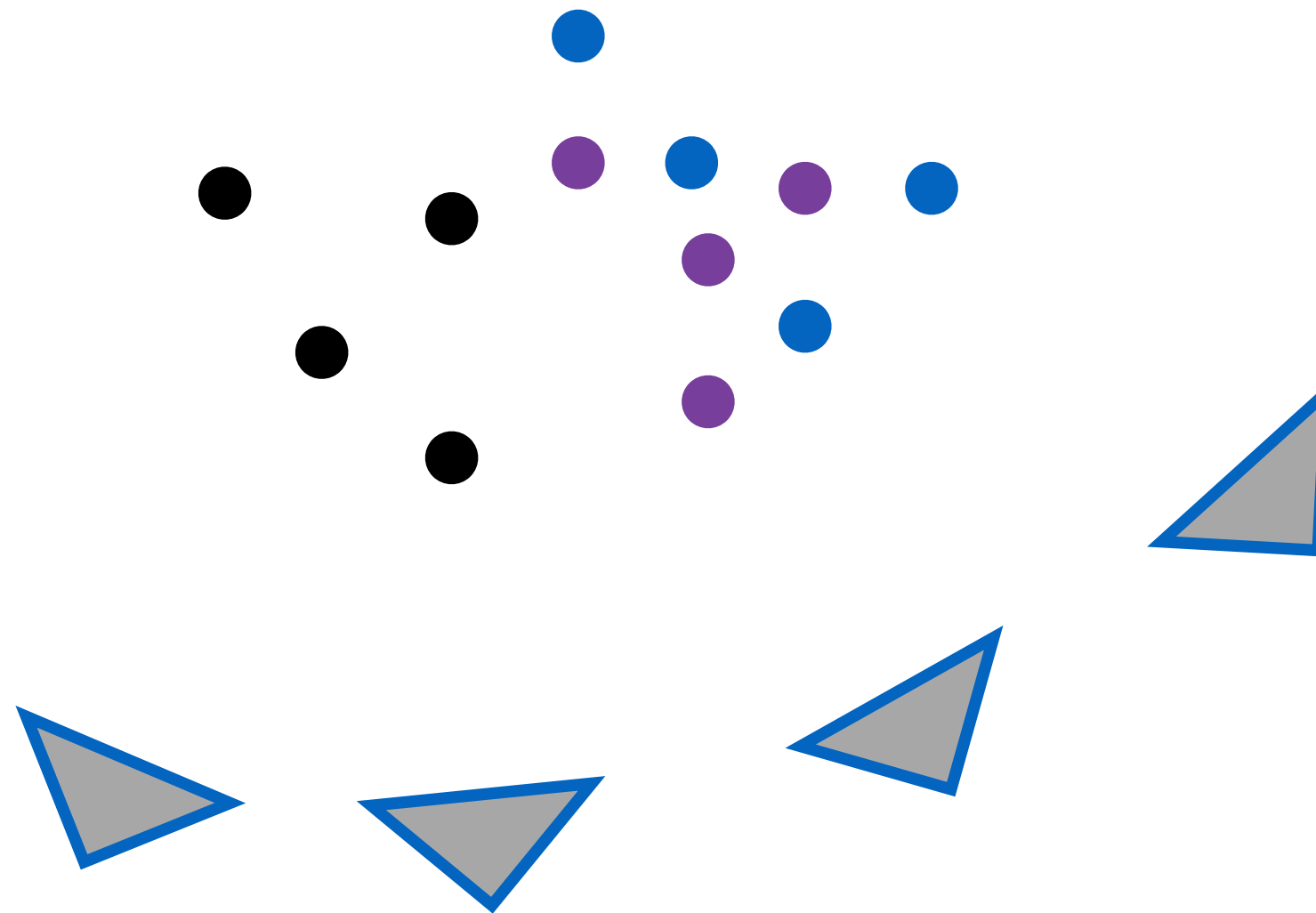
The Basic Idea



Structure from Motion

The Problem of Sequential Method

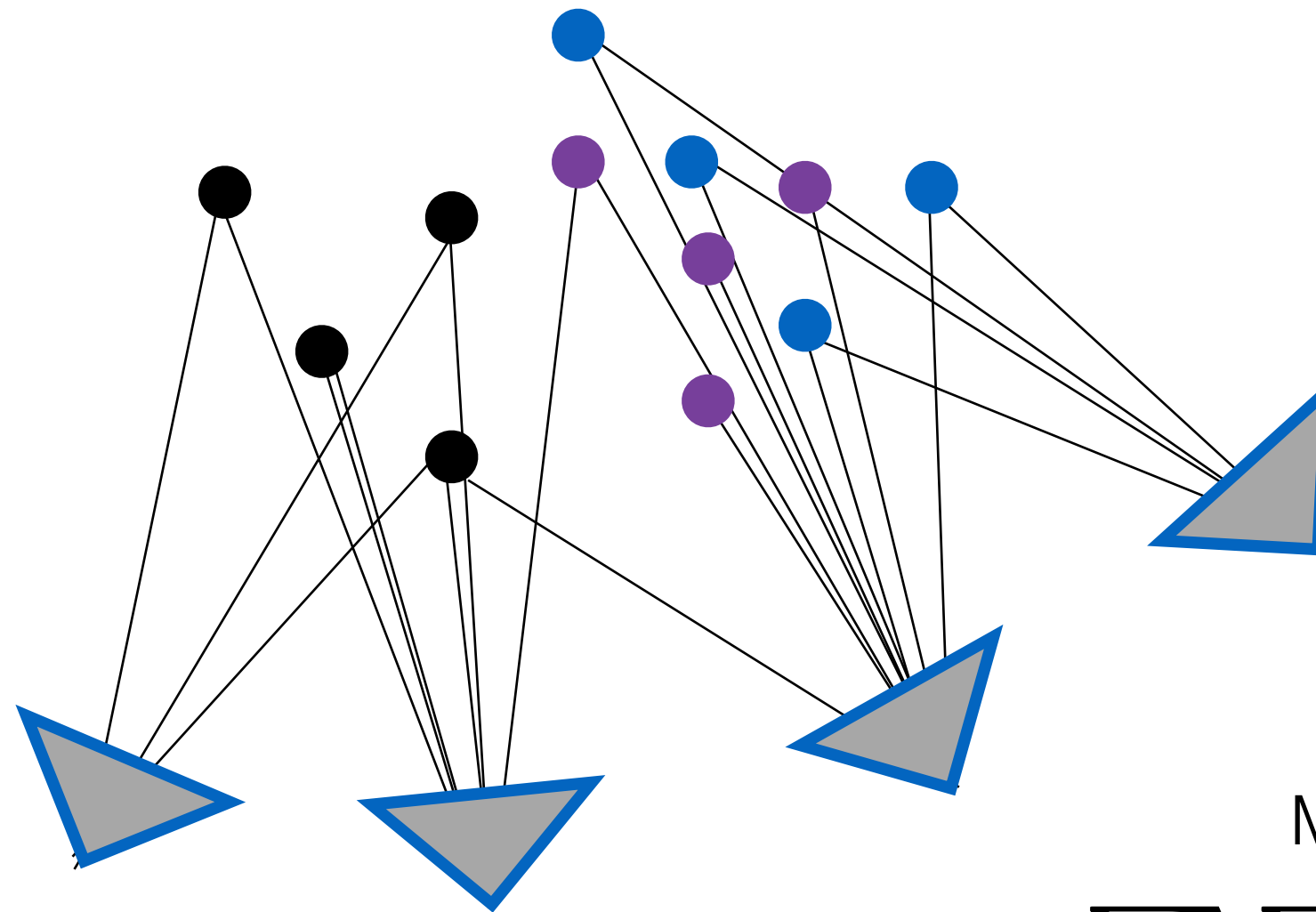
Error accumulation for \mathbf{X} , P



Structure from Motion

Bundle Adjustment

Error accumulation for \mathbf{X} , P



Minimizing reprojection error:

$$\min_{P, \mathbf{X}} \sum_k \sum_i ||\mathbf{x}_i^k - P_k(\mathbf{X}_i)||^2$$

Cool!... But Something Is Missing



Cool!... But Something Is Missing



Cool!... But Something Is Missing



No human in the city, why?

- No multiple views
- Few correspondences



The Panoptic Studio

A Massively Multiview System for Social Interaction Capture

Projector

HD Camera

VGA Camera

480 VGA Cameras
31 HD Cameras
10 Kinects
23 Microphones
5 Projectors
48 Machines
2 Petabyte Storage

[Joo et al., ICCV 2015]
[Joo et al., TPAMI 2017]

The Panoptic Studio

A Massively Multiview System for Social Interaction Capture



The Panoptic Studio

A Massively Multiview System for Social Interaction Capture



The Panoptic Studio

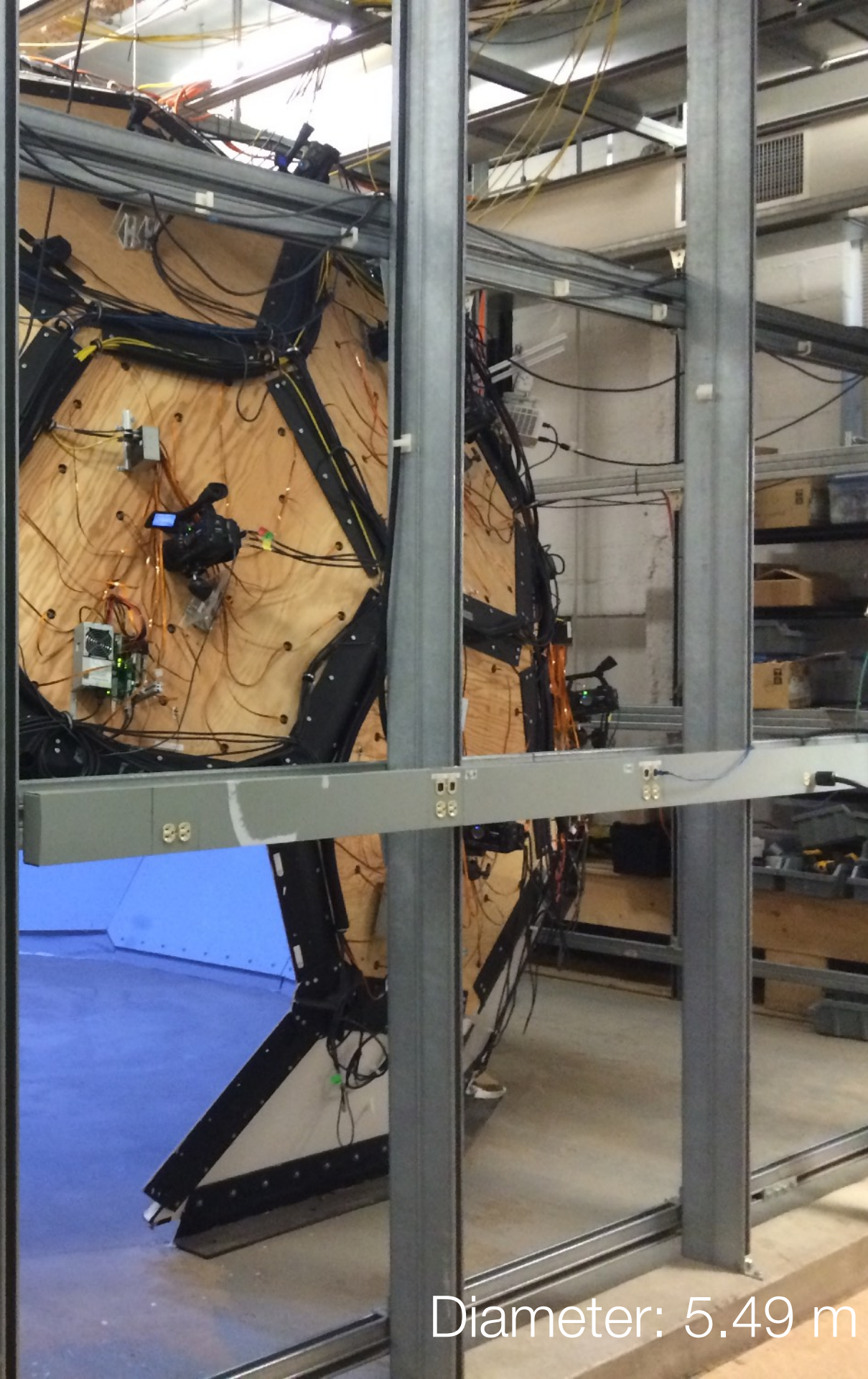
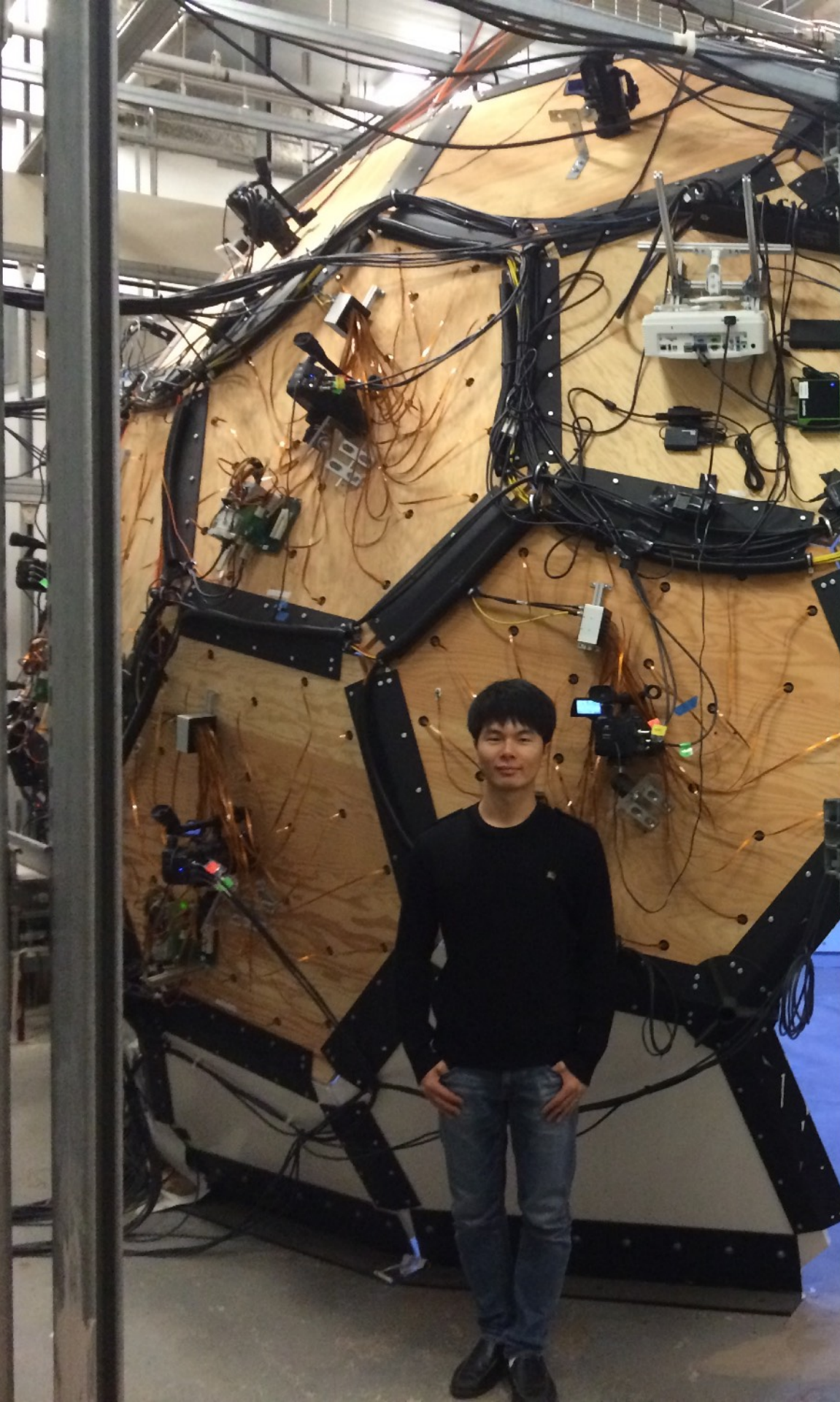
A Massively Multiview System for Social Interaction Capture

The image shows a top-down view of a circular studio floor with a hexagonal grid pattern. Several black circular markers are placed at various points on the grid. Three white arrows point from text labels to specific markers: one to a marker in the upper left, one to a marker in the center, and one to a marker in the lower right. A silver ribbon cable is visible at the bottom of the frame.

HD Camera

Kinect

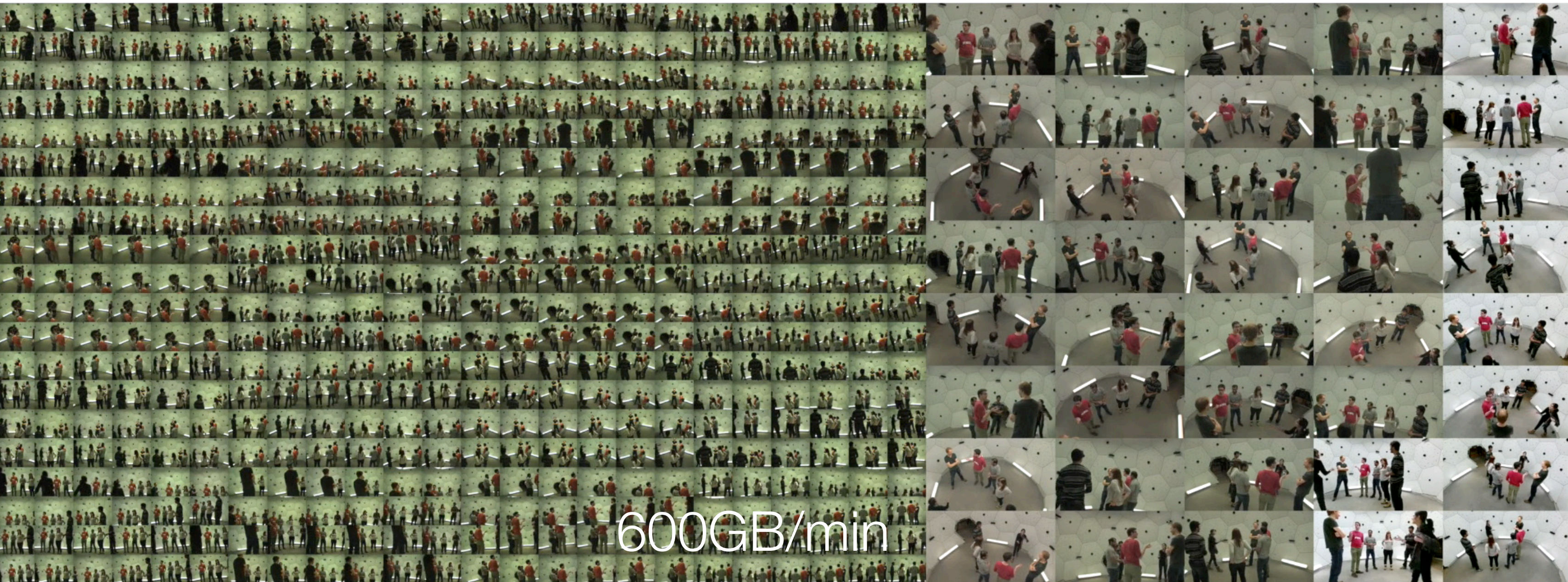
VGA Camera



Diameter: 5.49 m

Synchronized Videos from Unique 521 Views

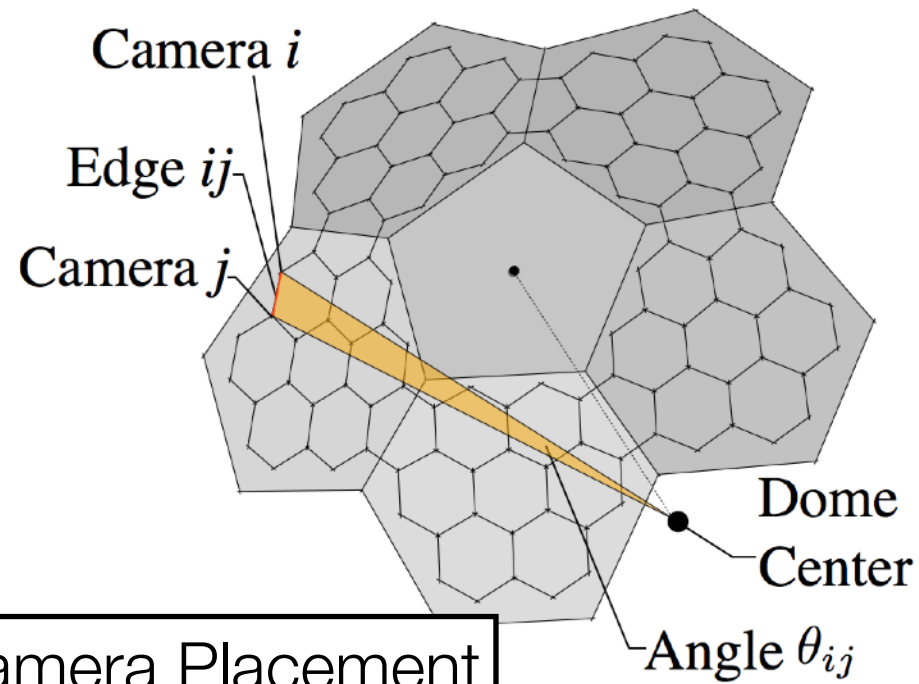
480 VGAs, 31 HDs, and 10 RGB+Ds



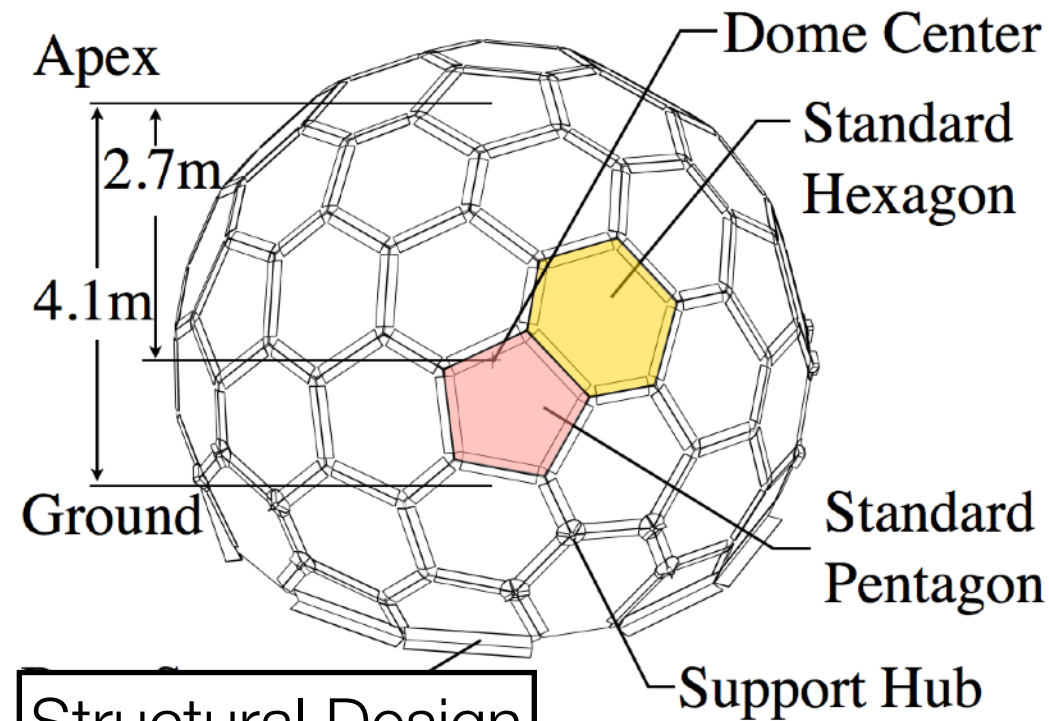
600GB/min

Software and Hardware Challenges

How to Build the Panoptic Studio



Camera Placement

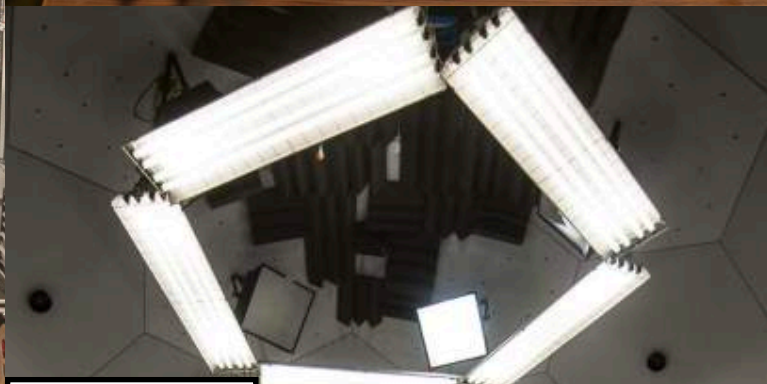


Structural Design

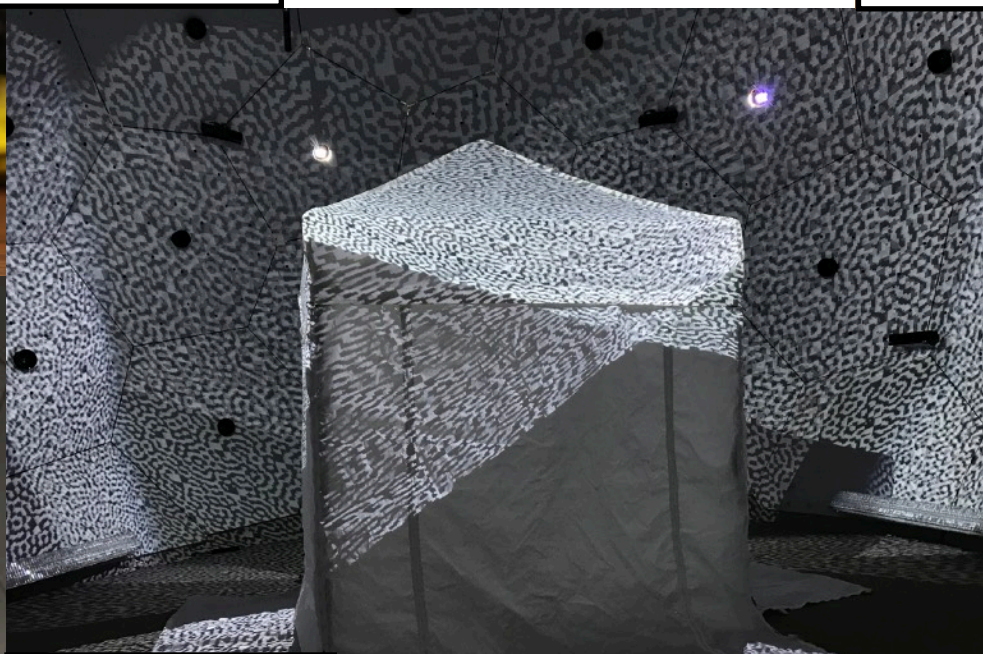


Sensor Types, Networking

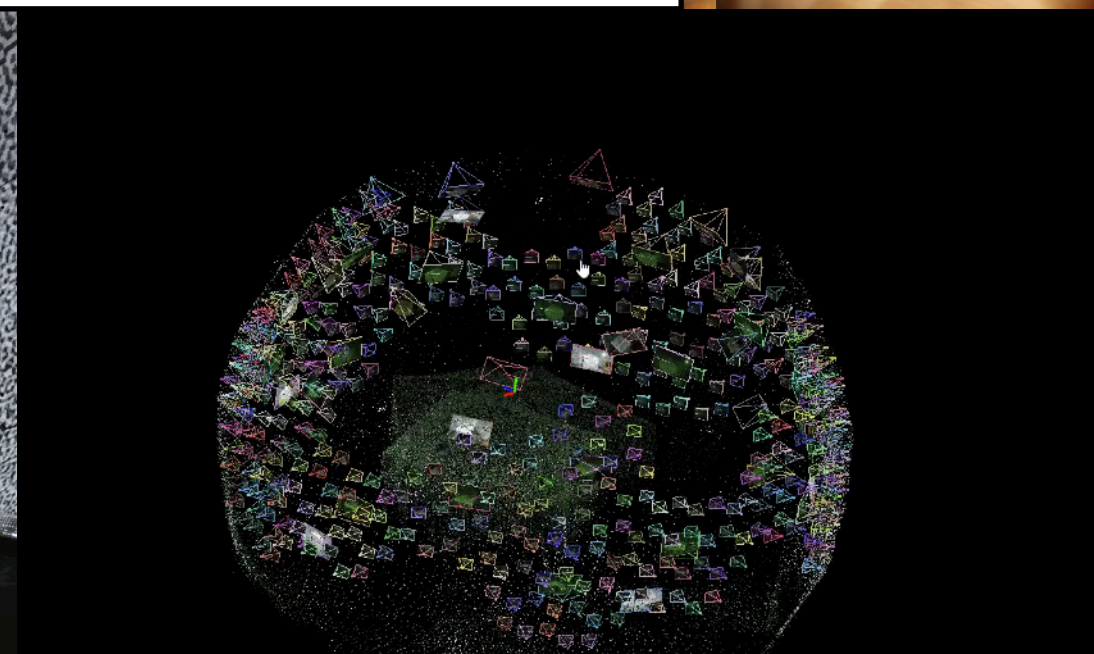
Synchronization



Lighting



Calibration



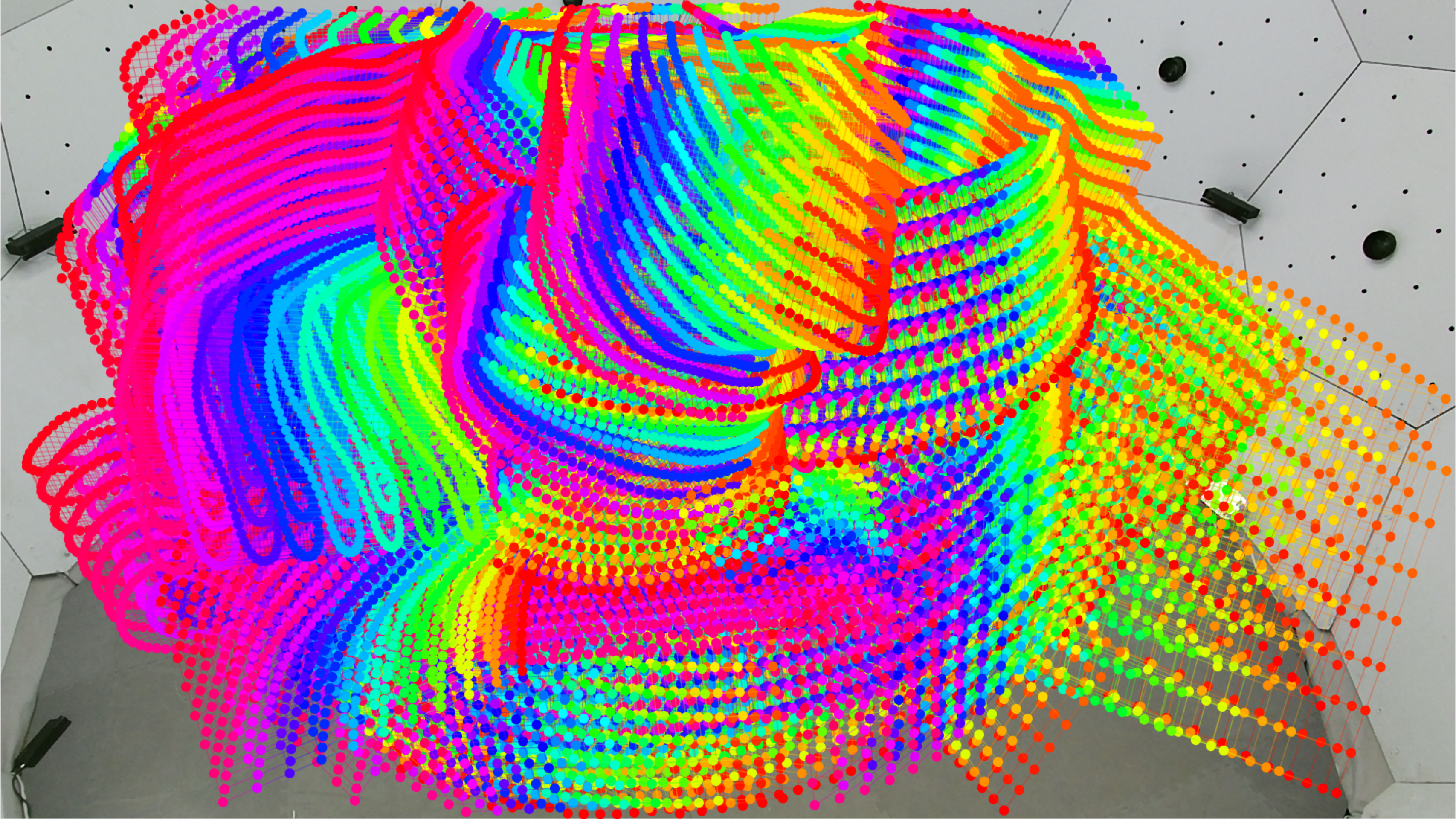
Storage





Calibration?





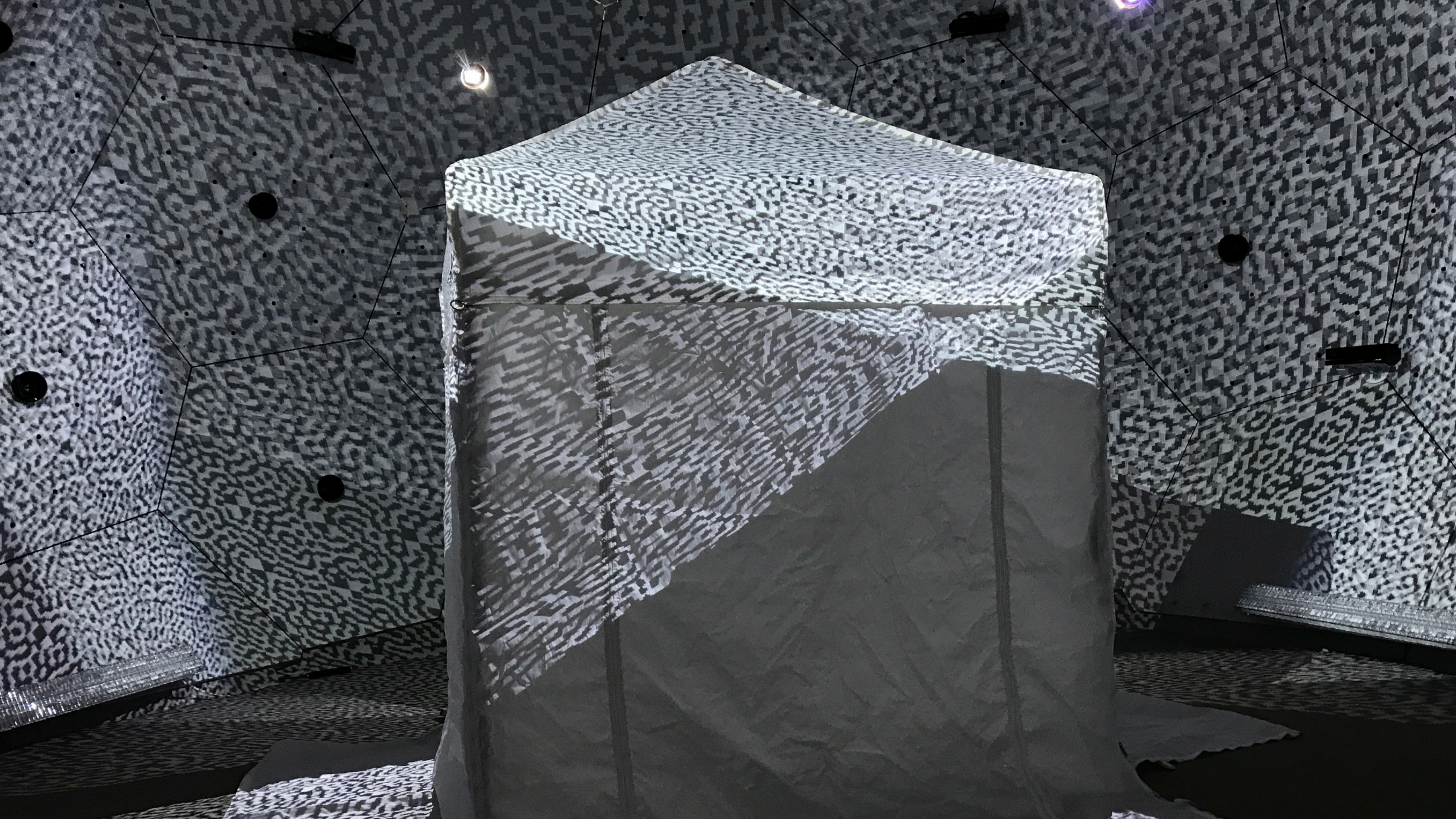


YOU MISSED A SPOT →

- **Painful..**
- **Wasting (1.2TB for 2min)**
- **Need to Extract frames from videos**



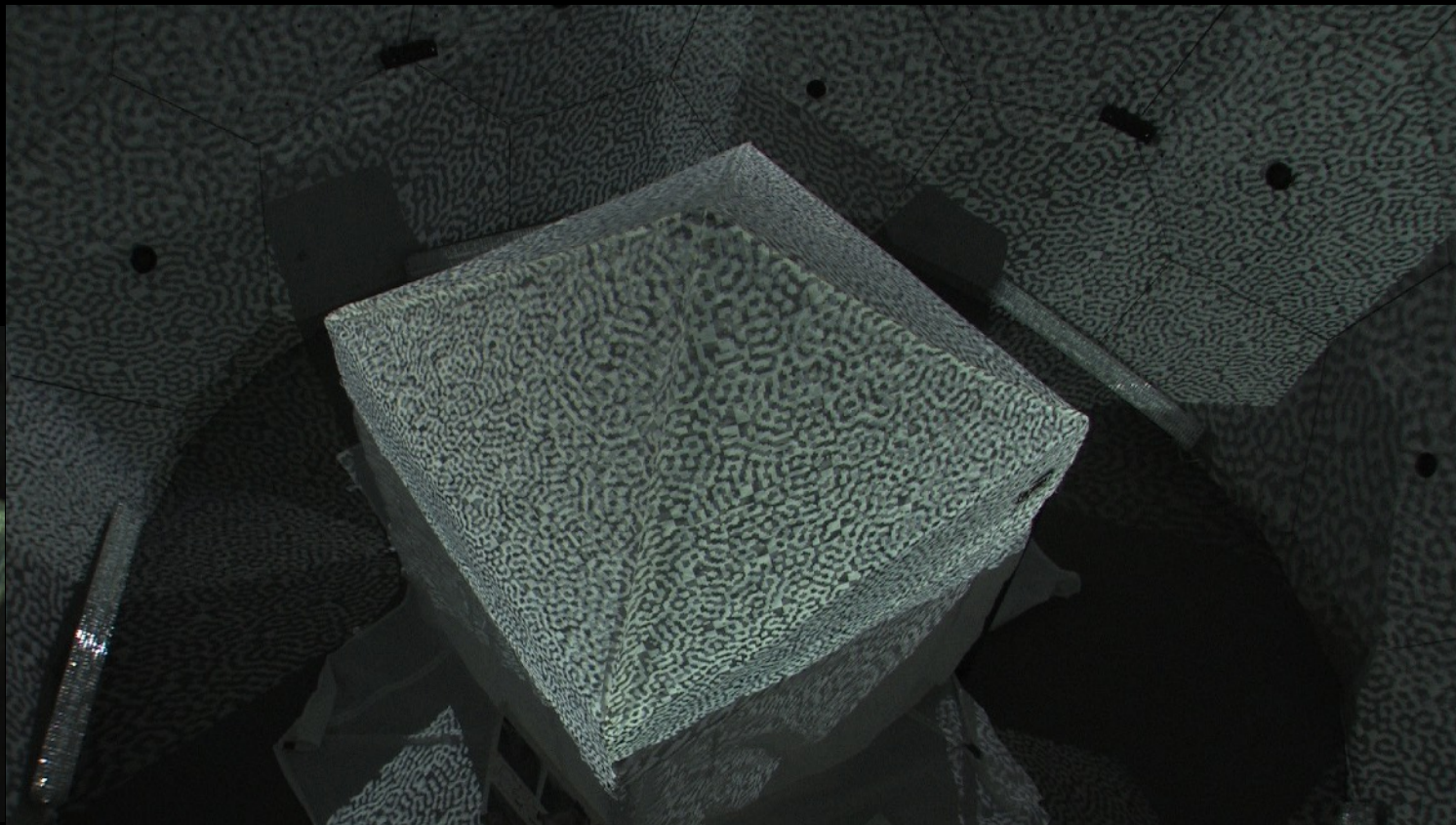




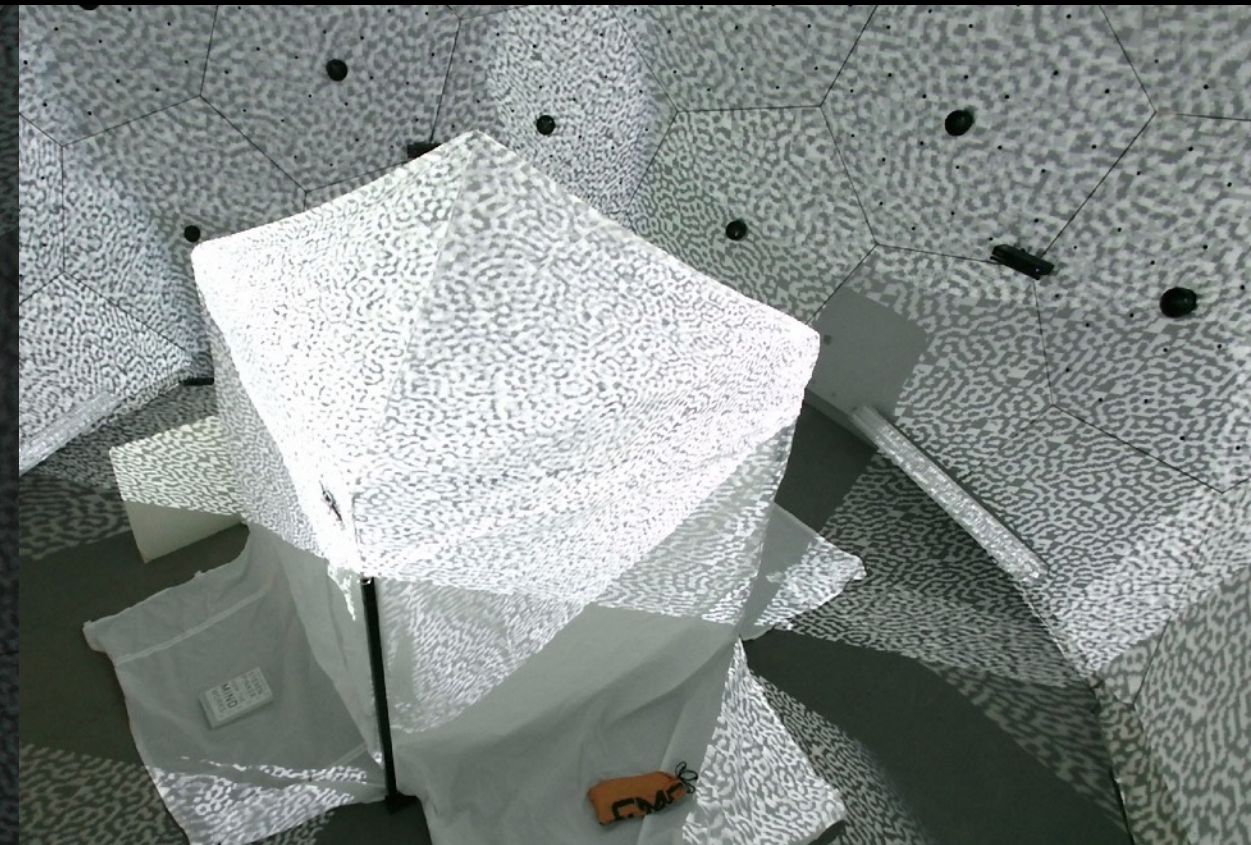
Panoptic Studio Camera Calibration



VGA



HD

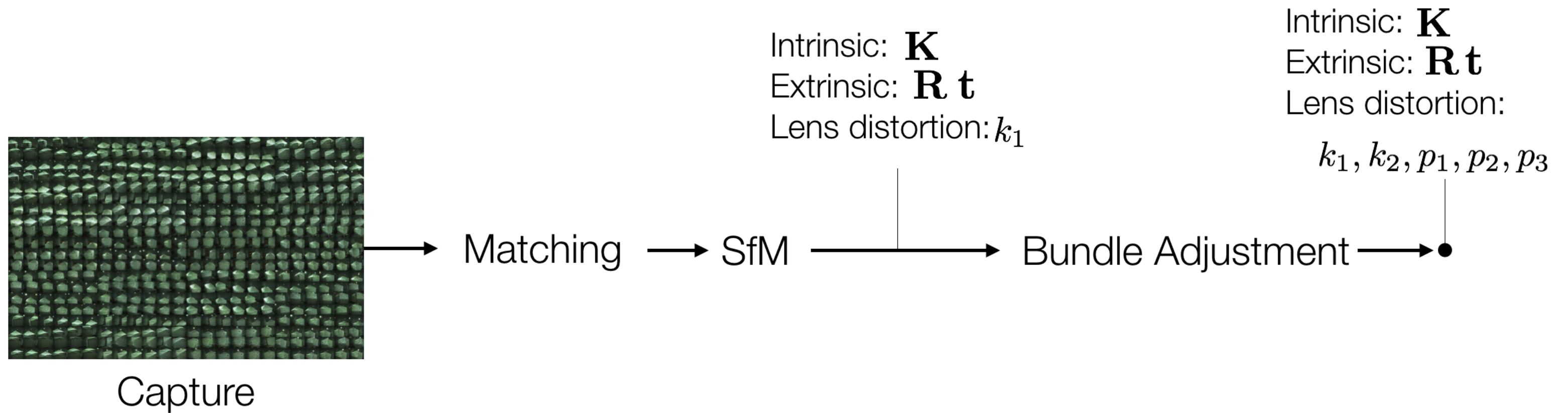


Kinect Color

Run **Structure from Motion** to get calibration parameters!

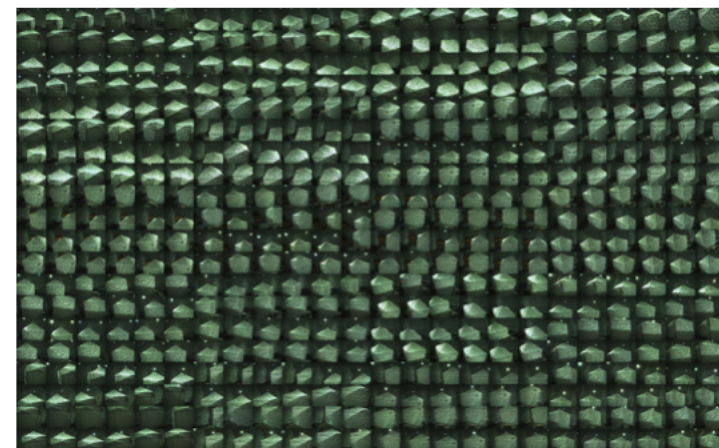
Calibration for Panoptic Studio

Based on Structure-from-Motion



Calibration for Panoptic Studio

Based on Structure-from-Motion



Capture

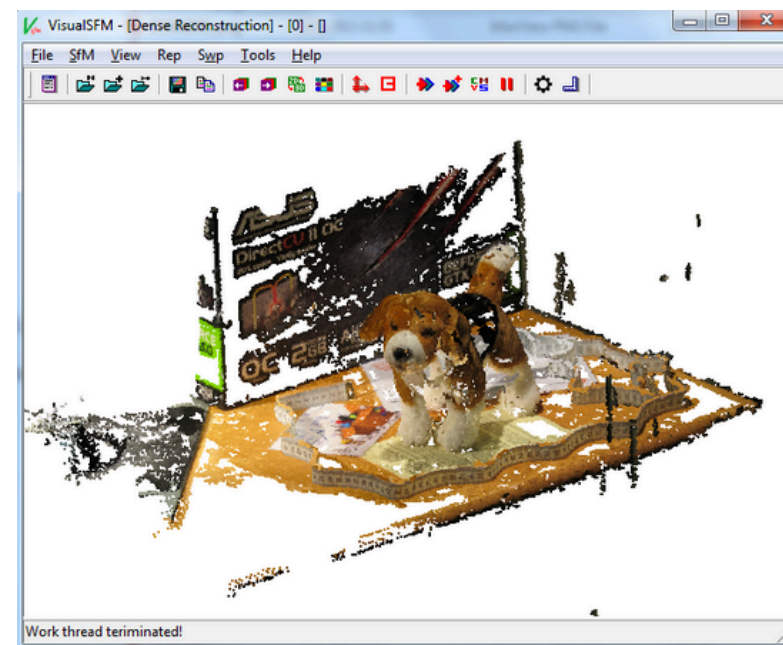
Matching

SfM

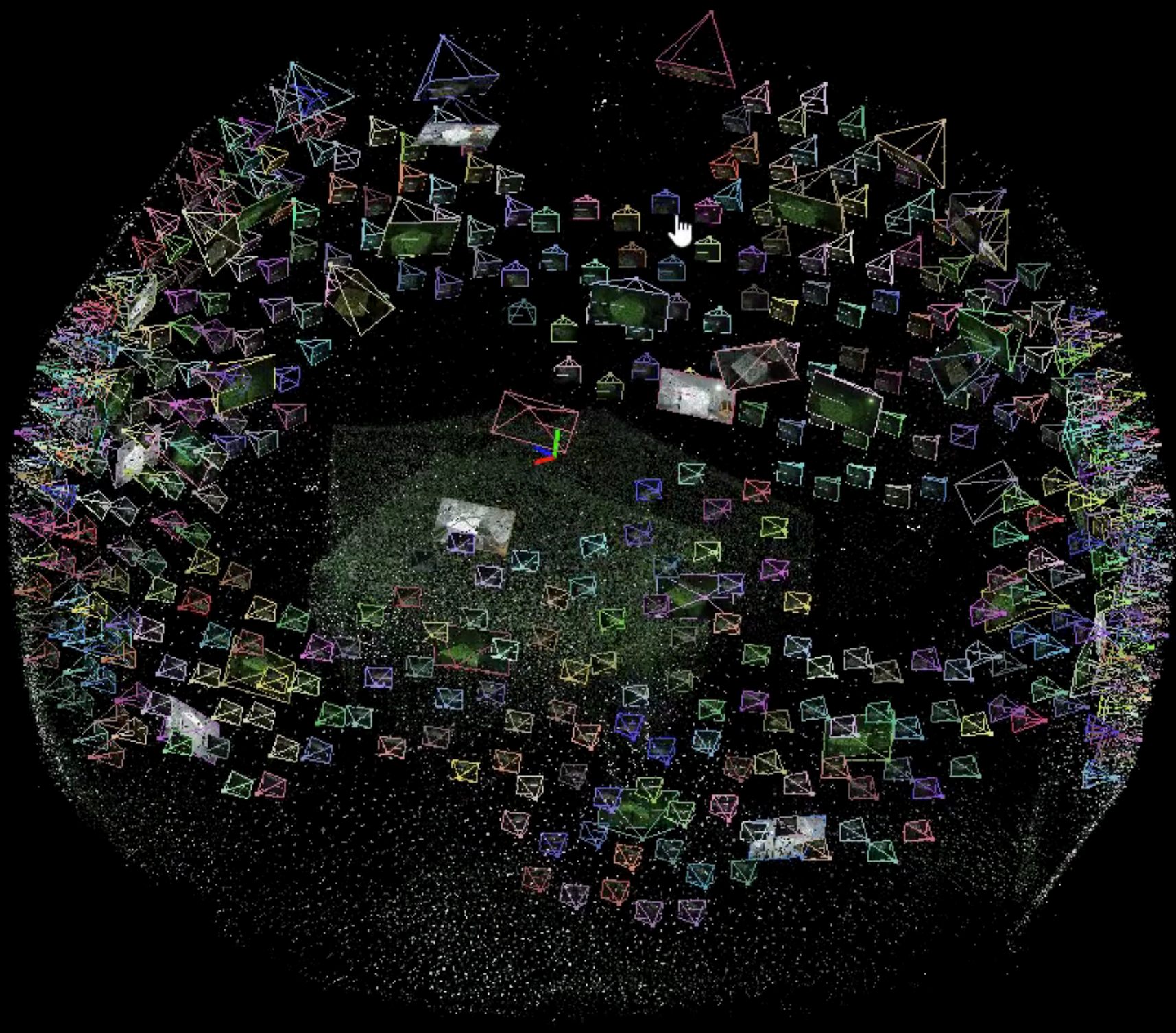
Bundle Adjustment

Intrinsic: \mathbf{K}
Extrinsic: $\mathbf{R} \mathbf{t}$
Lens distortion: k_1

Intrinsic: \mathbf{K}
Extrinsic: $\mathbf{R} \mathbf{t}$
Lens distortion:
 k_1, k_2, p_1, p_2, p_3

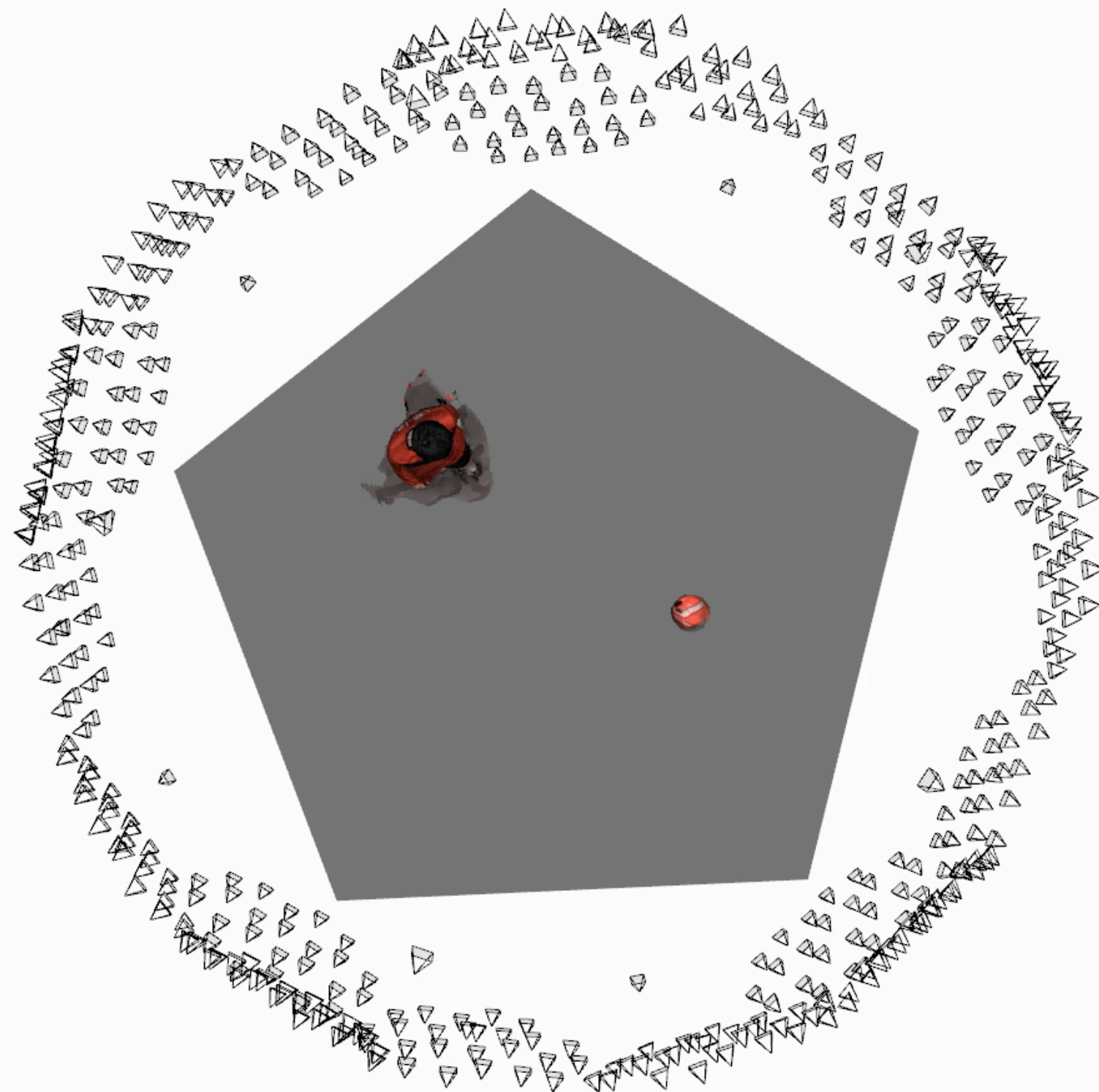


VisualSfM by Changchang Wu



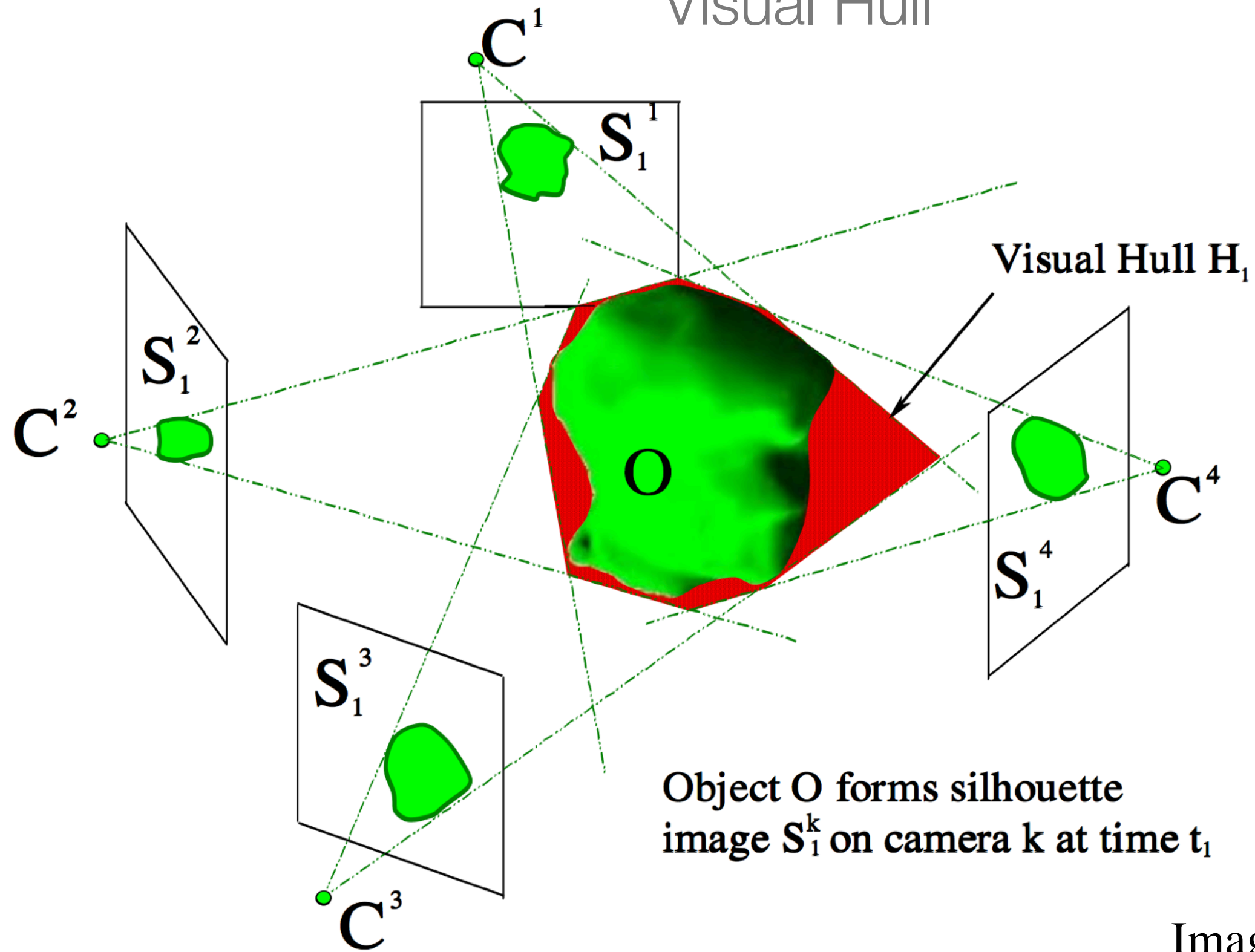
Measuring 3D Volume

Visual Hull



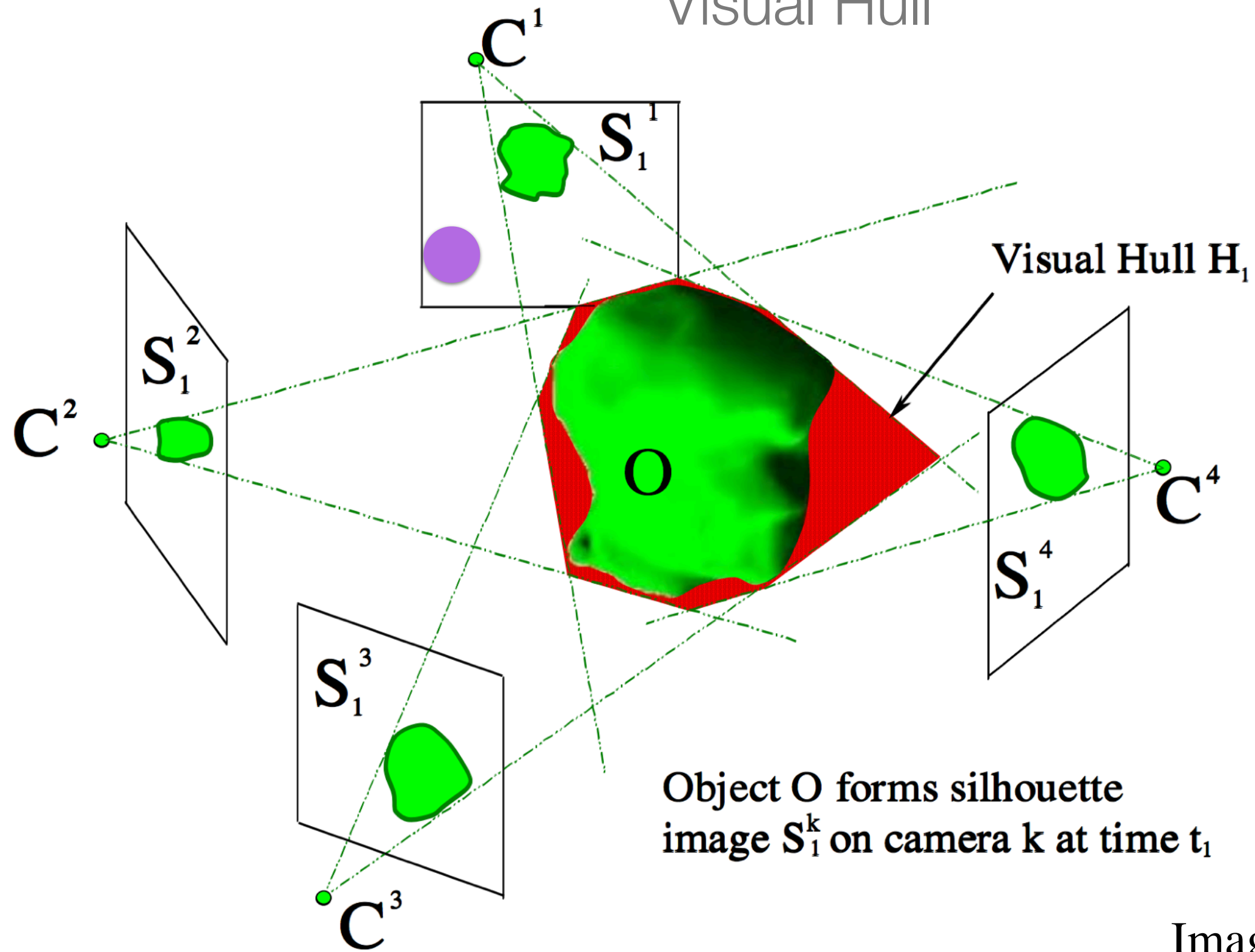
Measuring 3D Volume

Visual Hull



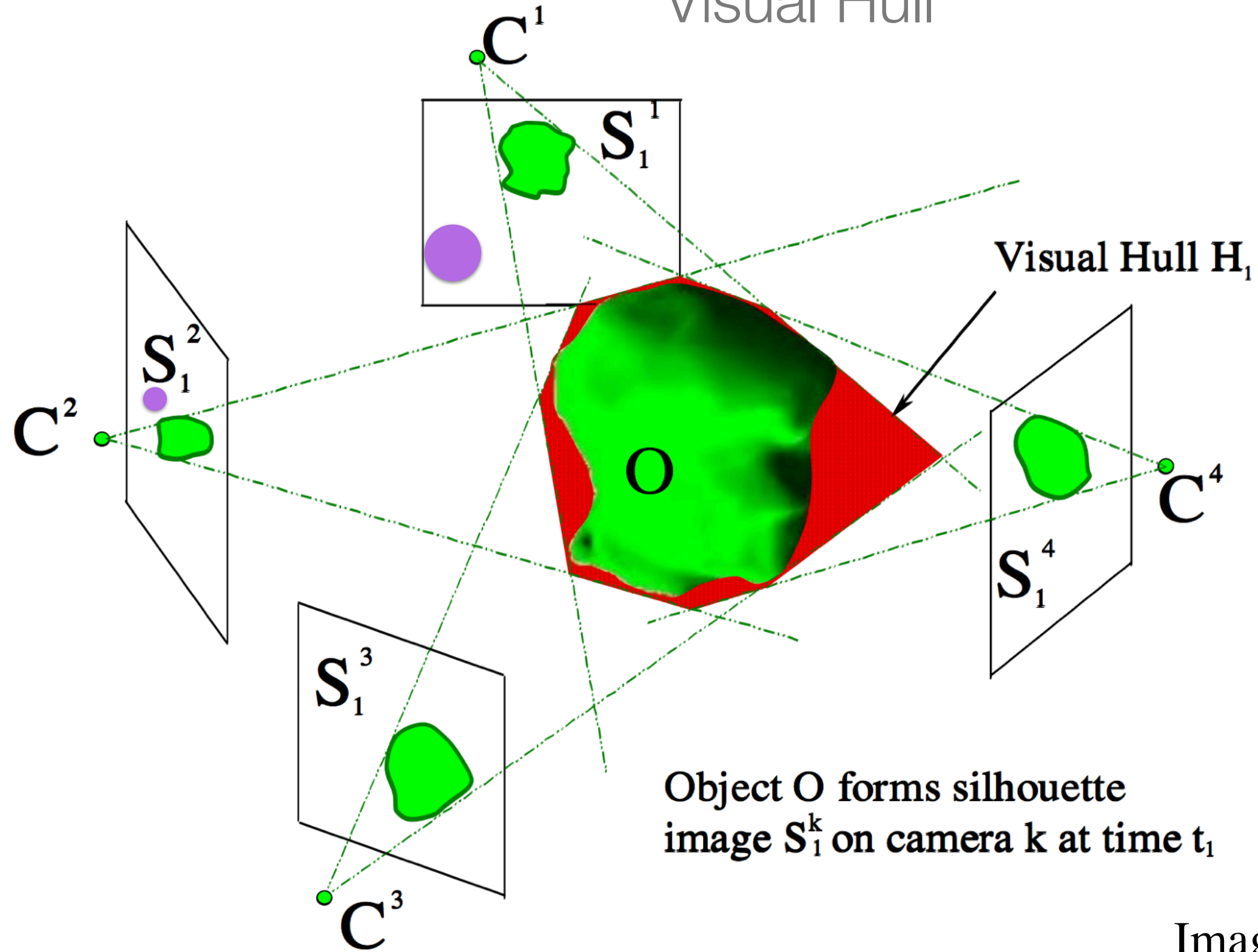
Measuring 3D Volume

Visual Hull

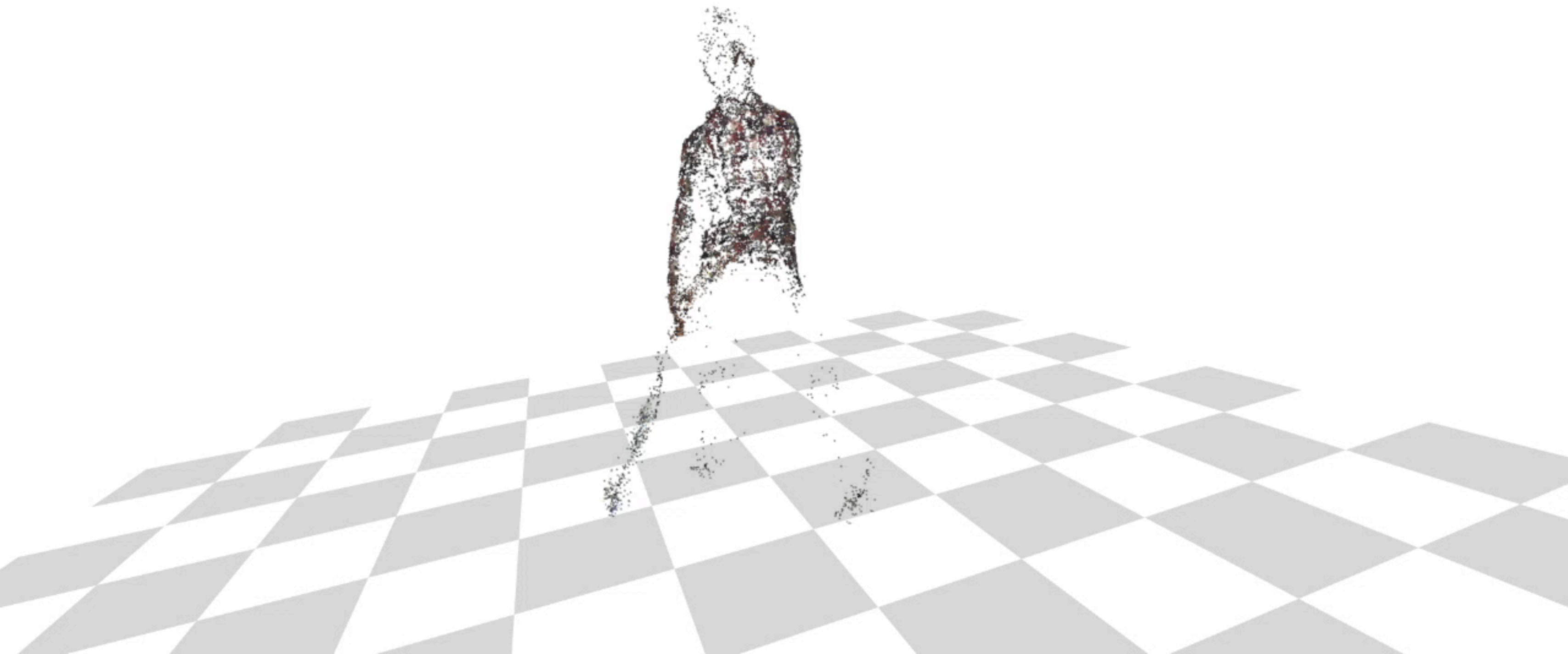


Measuring 3D Volume

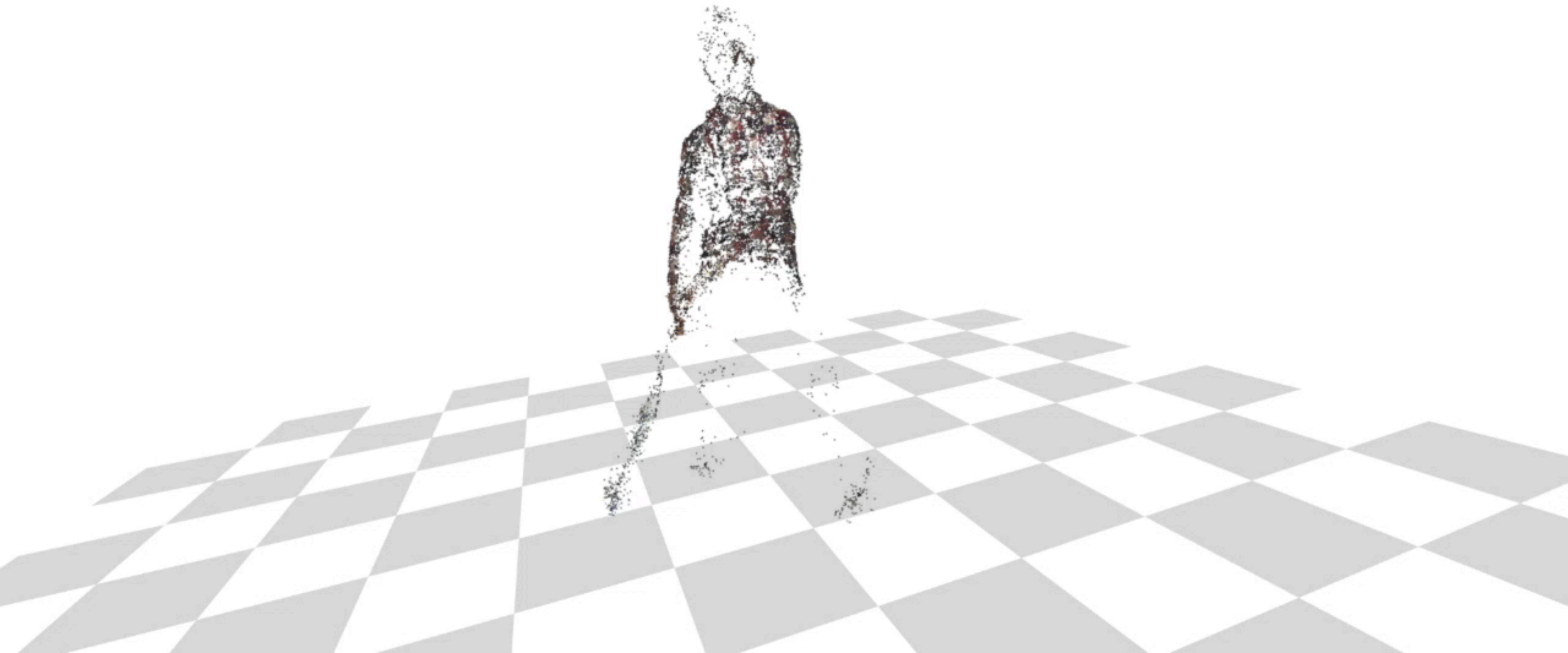
Visual Hull



Reconstructing 3D Point Cloud

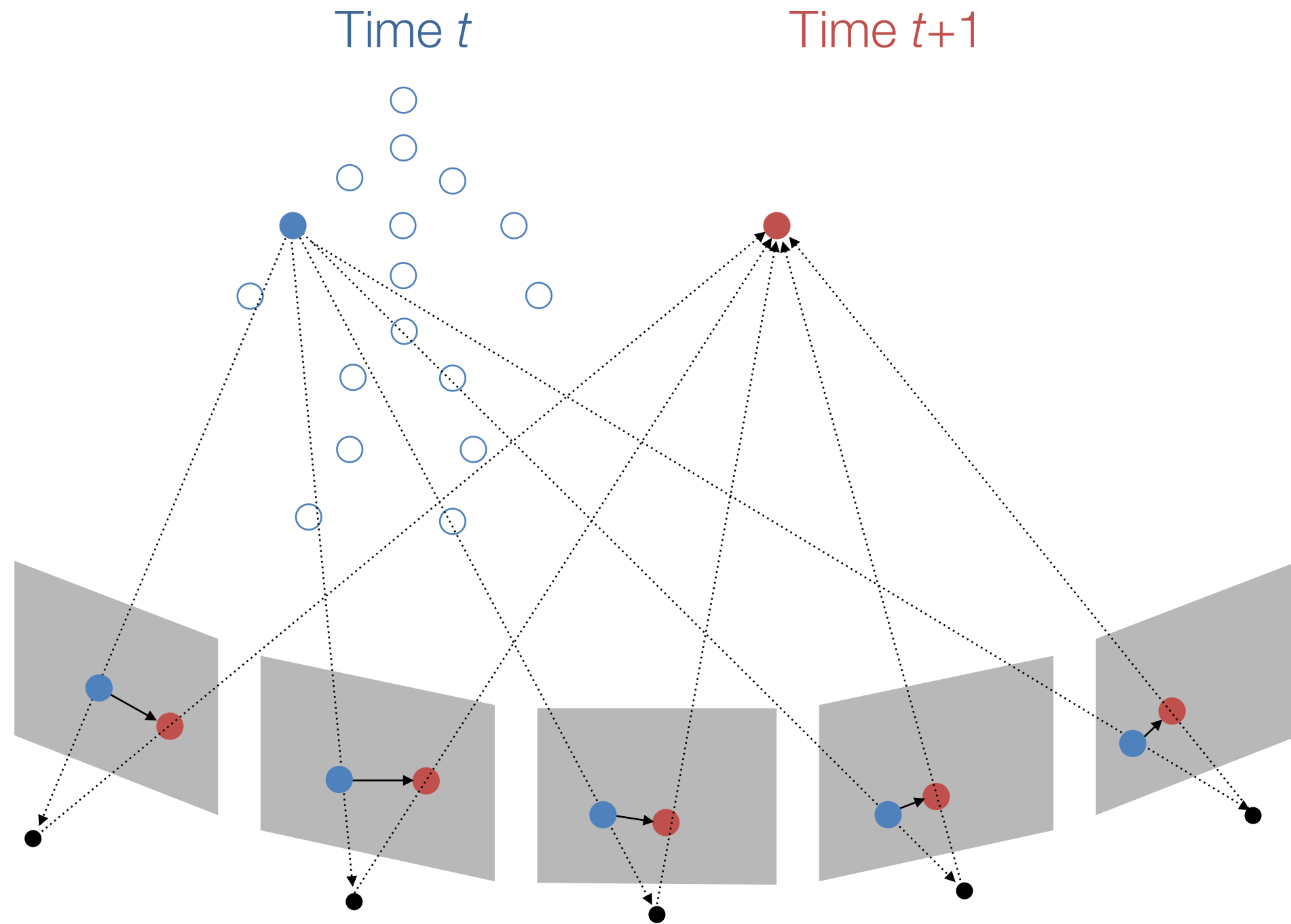


Reconstructing 3D **Trajectory Stream**



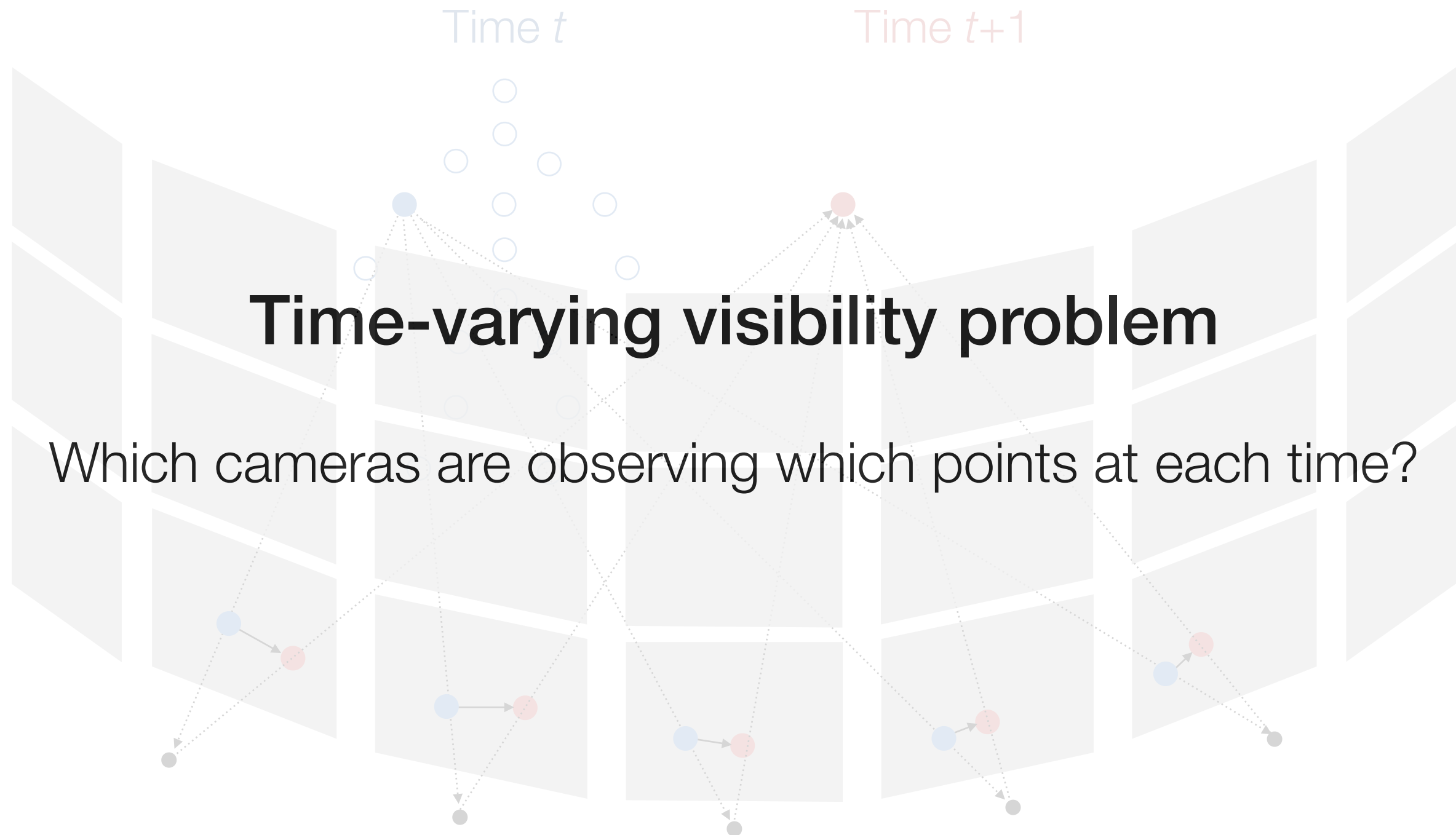
Reconstructing 3D Trajectory

2D Flow-based Method



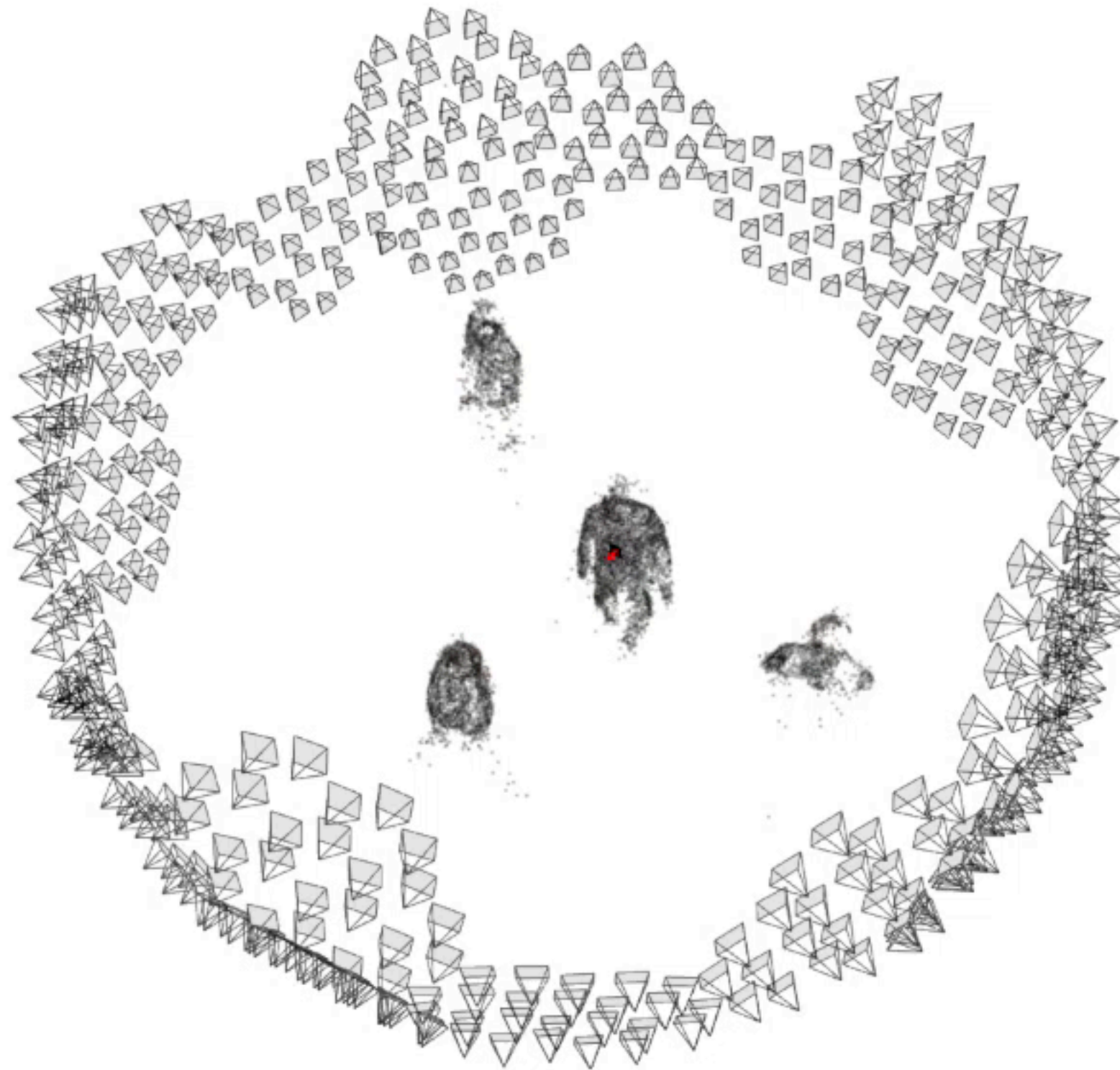
Reconstructing 3D Trajectory

Key Issue To Leverage a Large Number of Views



A Core Idea

Reasoning About Time Varying Visibility



Trajectory Stream Reconstruction

The Volleyball Sequence



Trajectory Stream Reconstruction

The Confetti Sequence



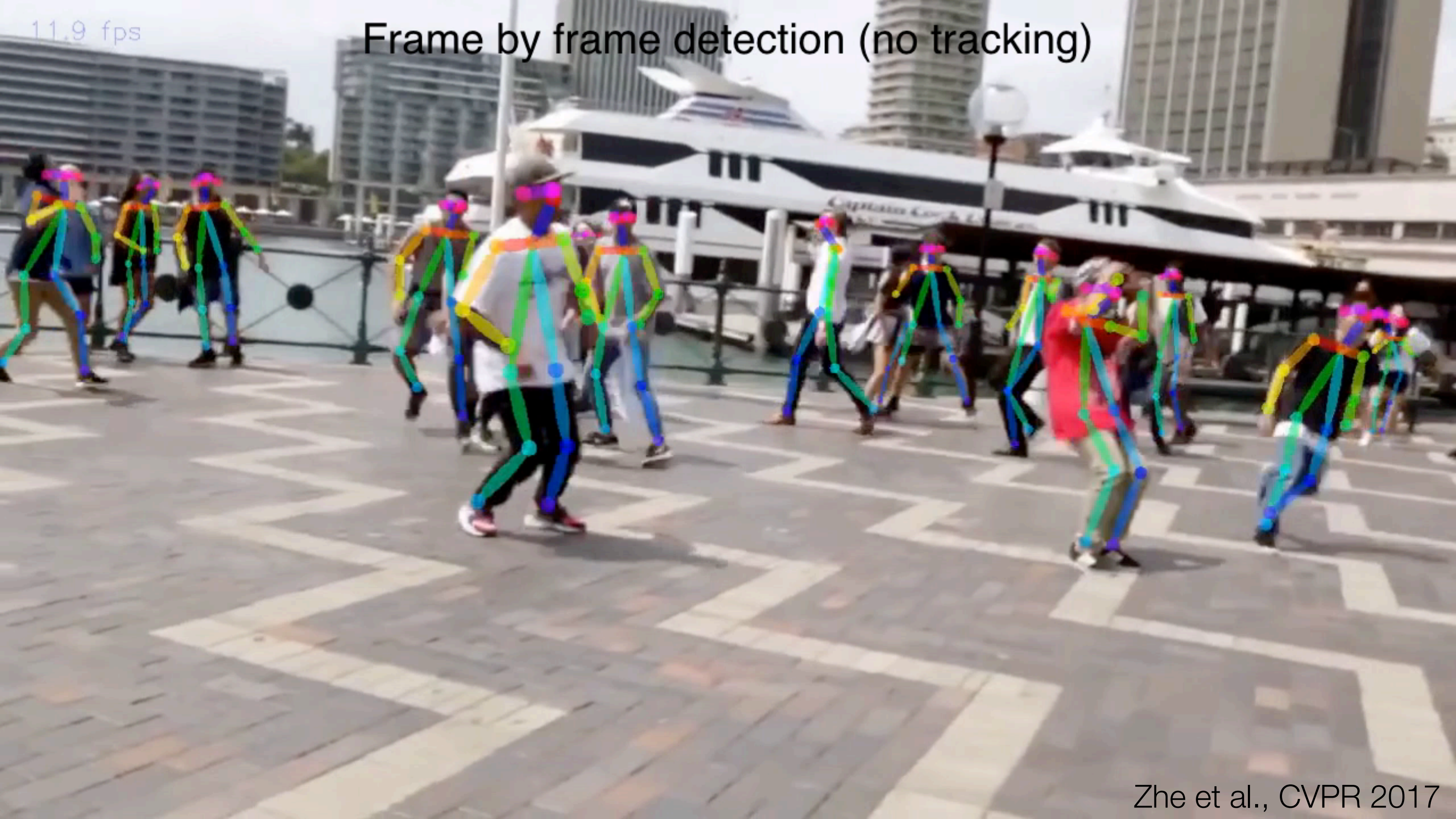
Trajectory Stream Reconstruction

The Fluid Motion Sequence

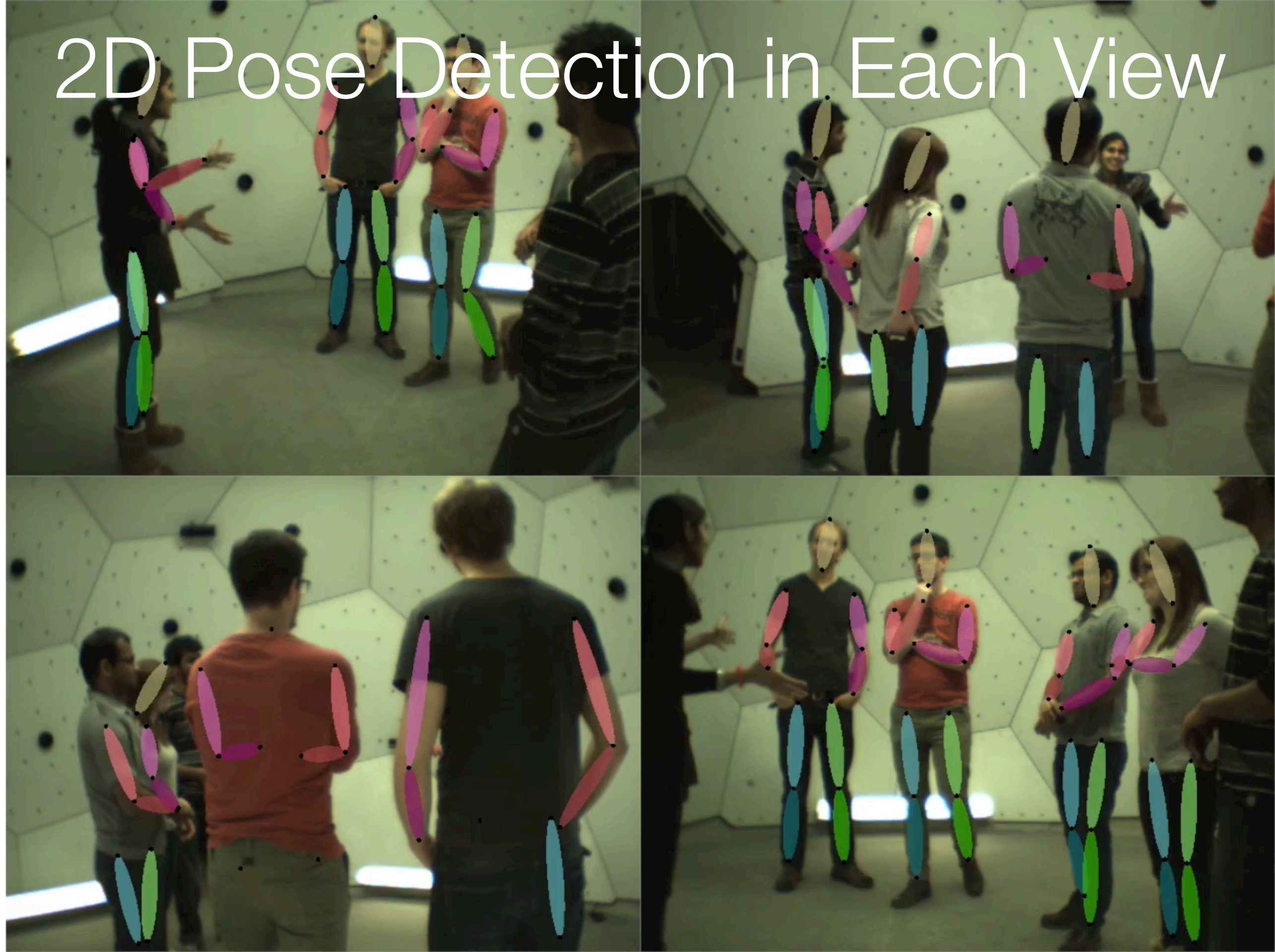


11.9 fps

Frame by frame detection (no tracking)

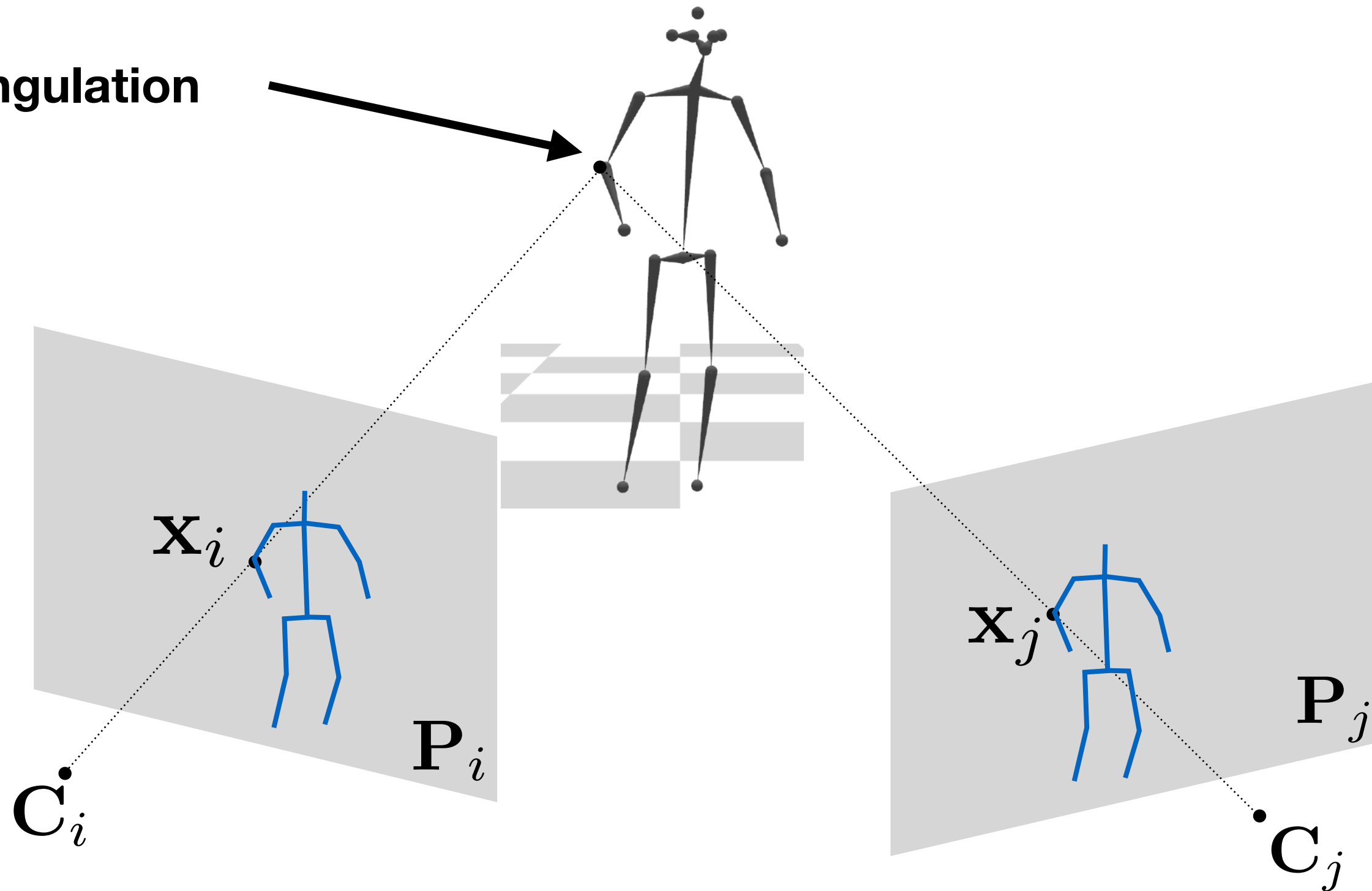


2D Pose Detection in Each View



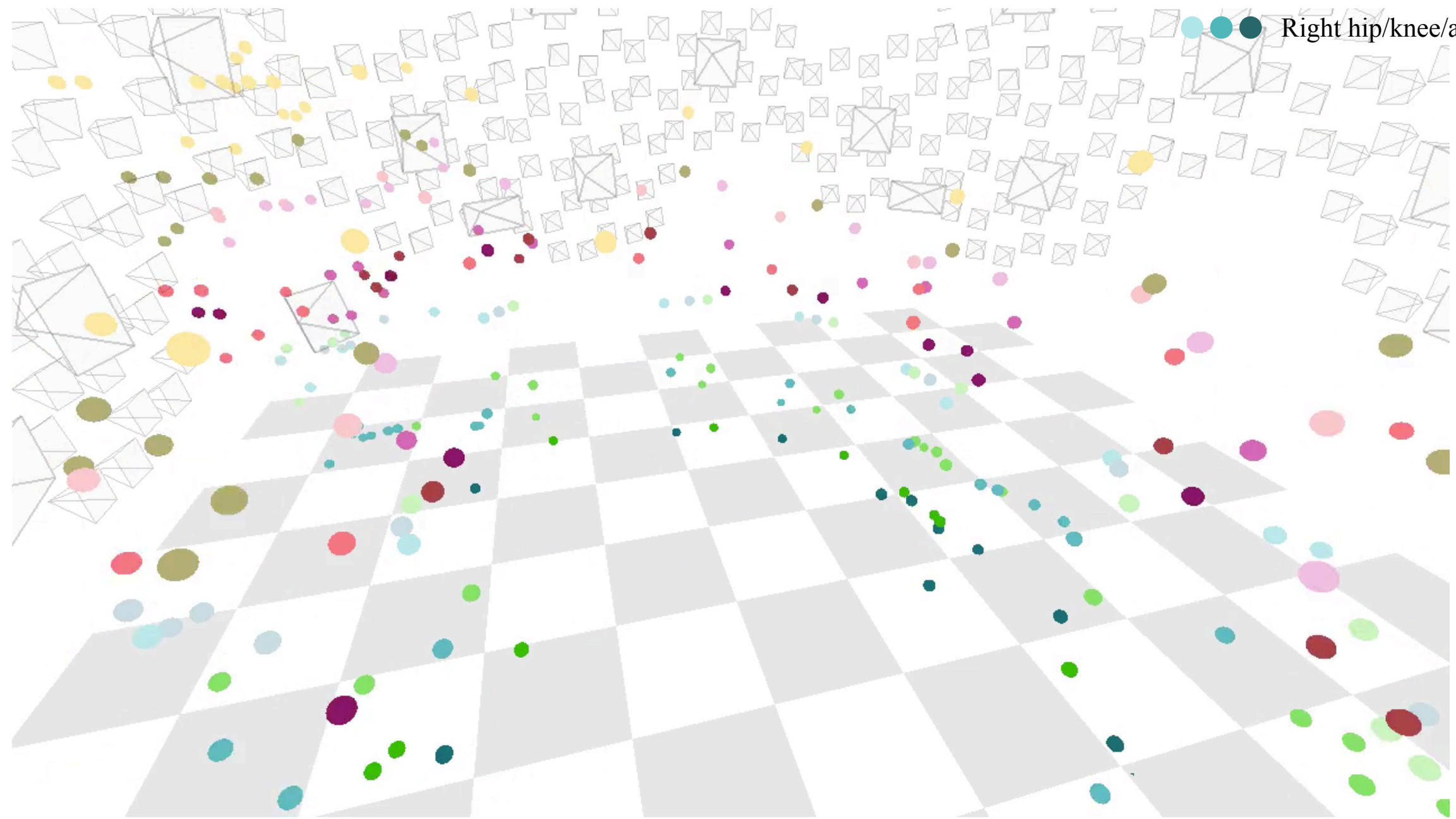
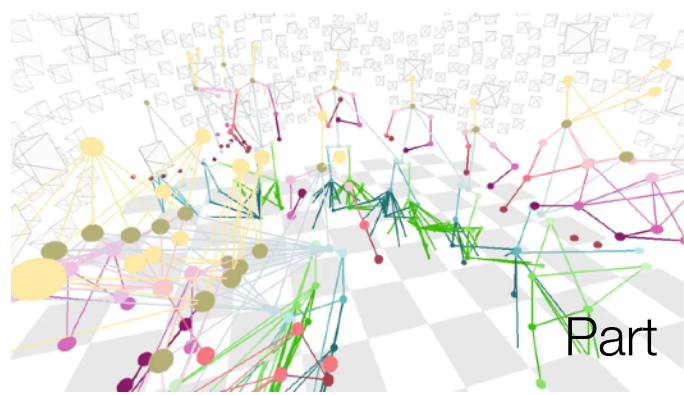
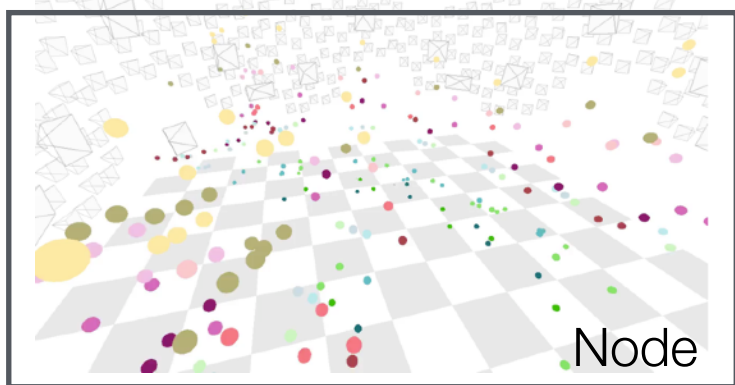
Triangulating 3D Skeletons

Triangulation



Fully Automatic Markerless Human Motion Capture

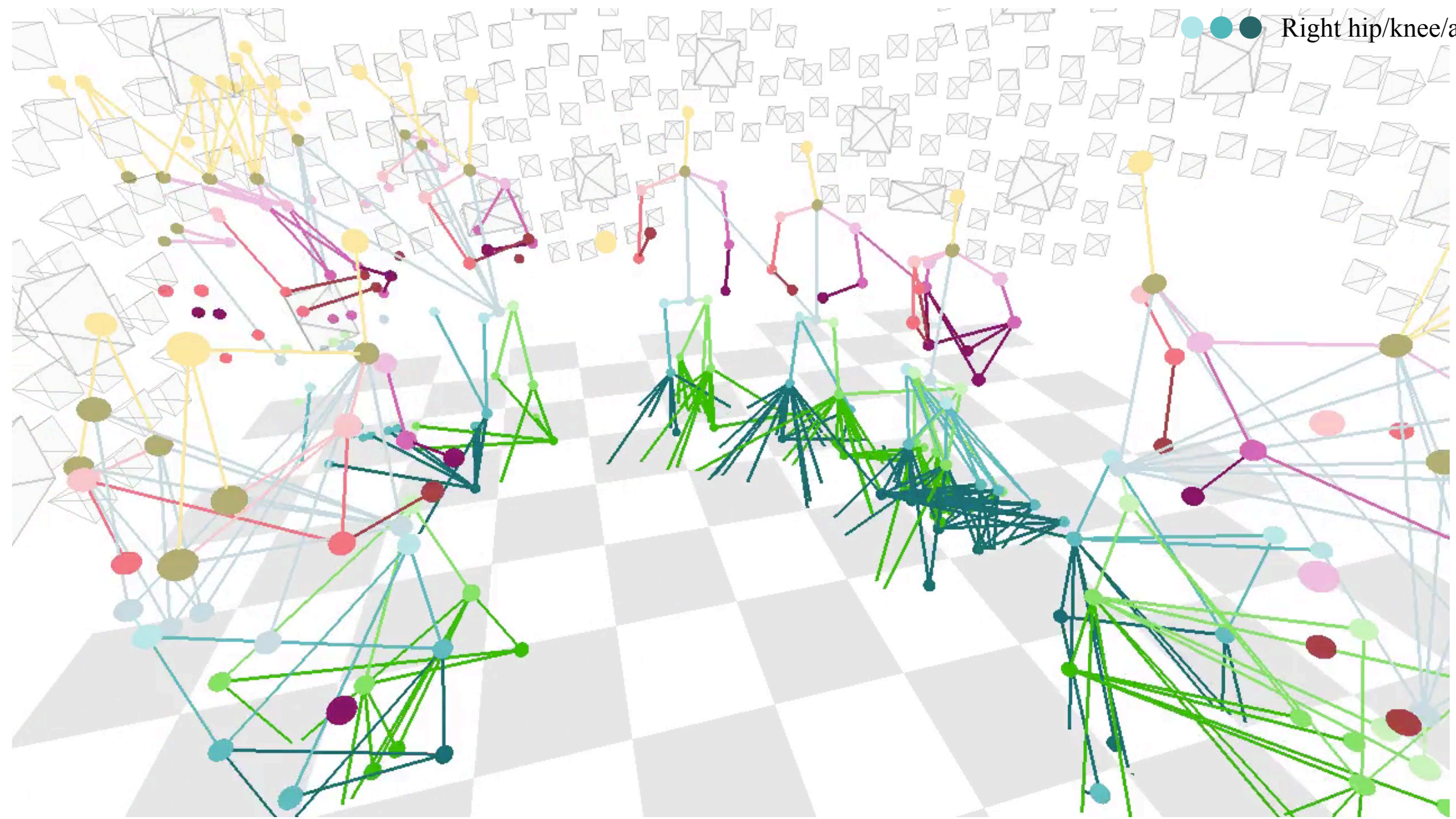
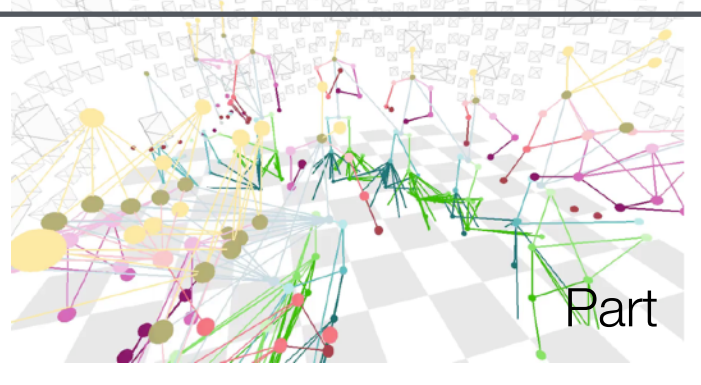
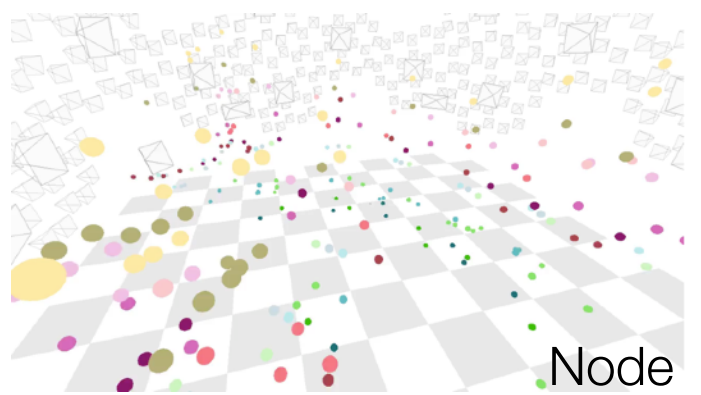
- ● ● HeadTop/neck/bodyCenter
- ● ● Left shoulder/elbow/wrist
- ● ● Right shoulder/elbow/wrist
- ● ● Left hip/knee/ankle
- ● ● Right hip/knee/ankle



Generating "Node" Proposals

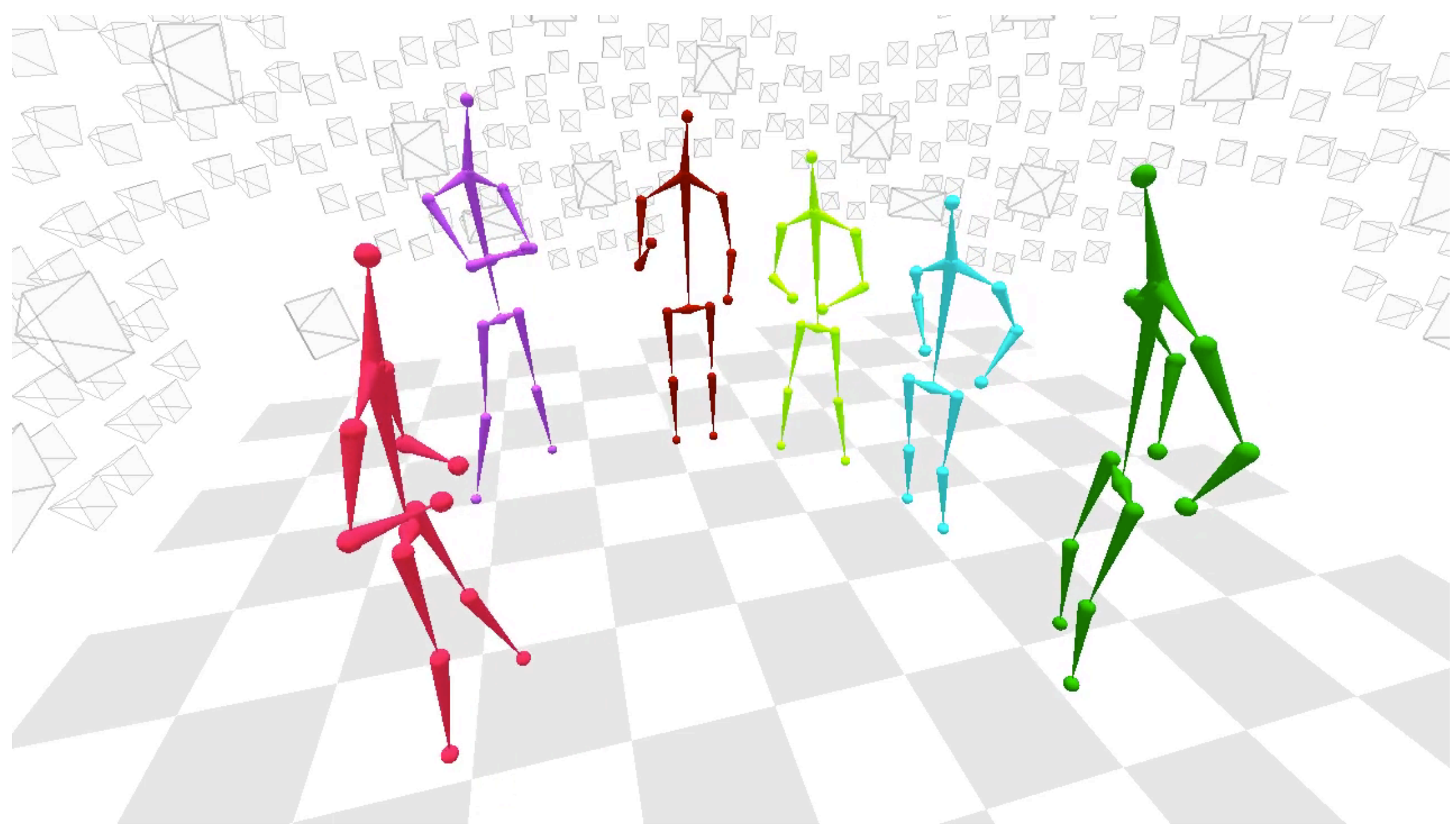
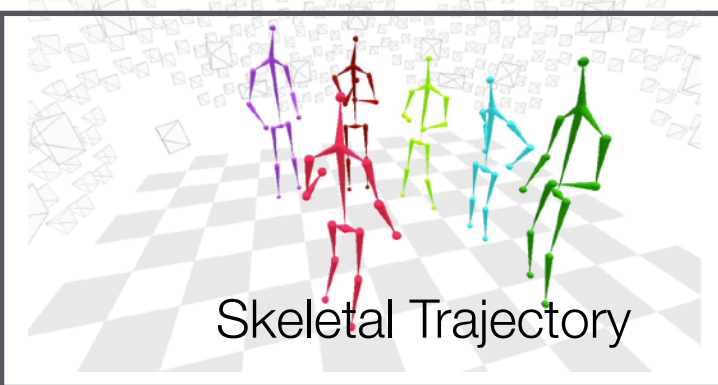
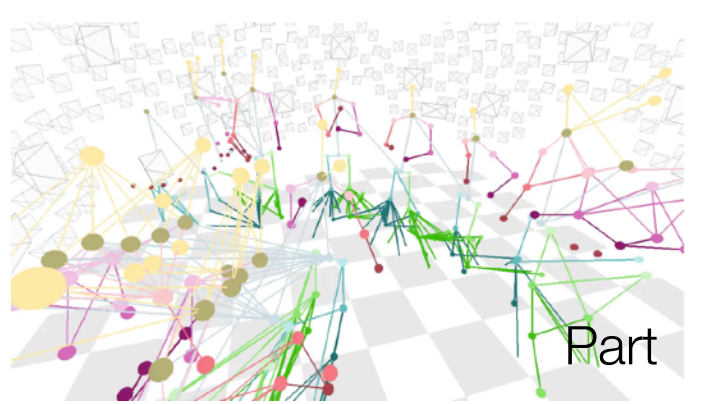
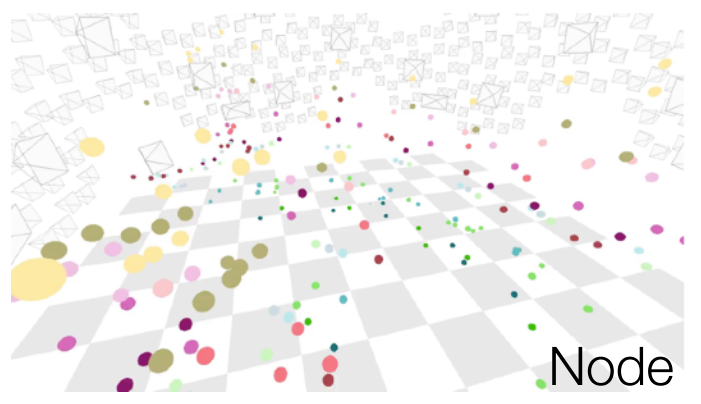
Fully Automatic Markerless Human Motion Capture

- ● ● HeadTop/neck/bodyCenter
- ● ● Left shoulder/elbow/wrist
- ● ● Right shoulder/elbow/wrist
- ● ● Left hip/knee/ankle
- ● ● Right hip/knee/ankle



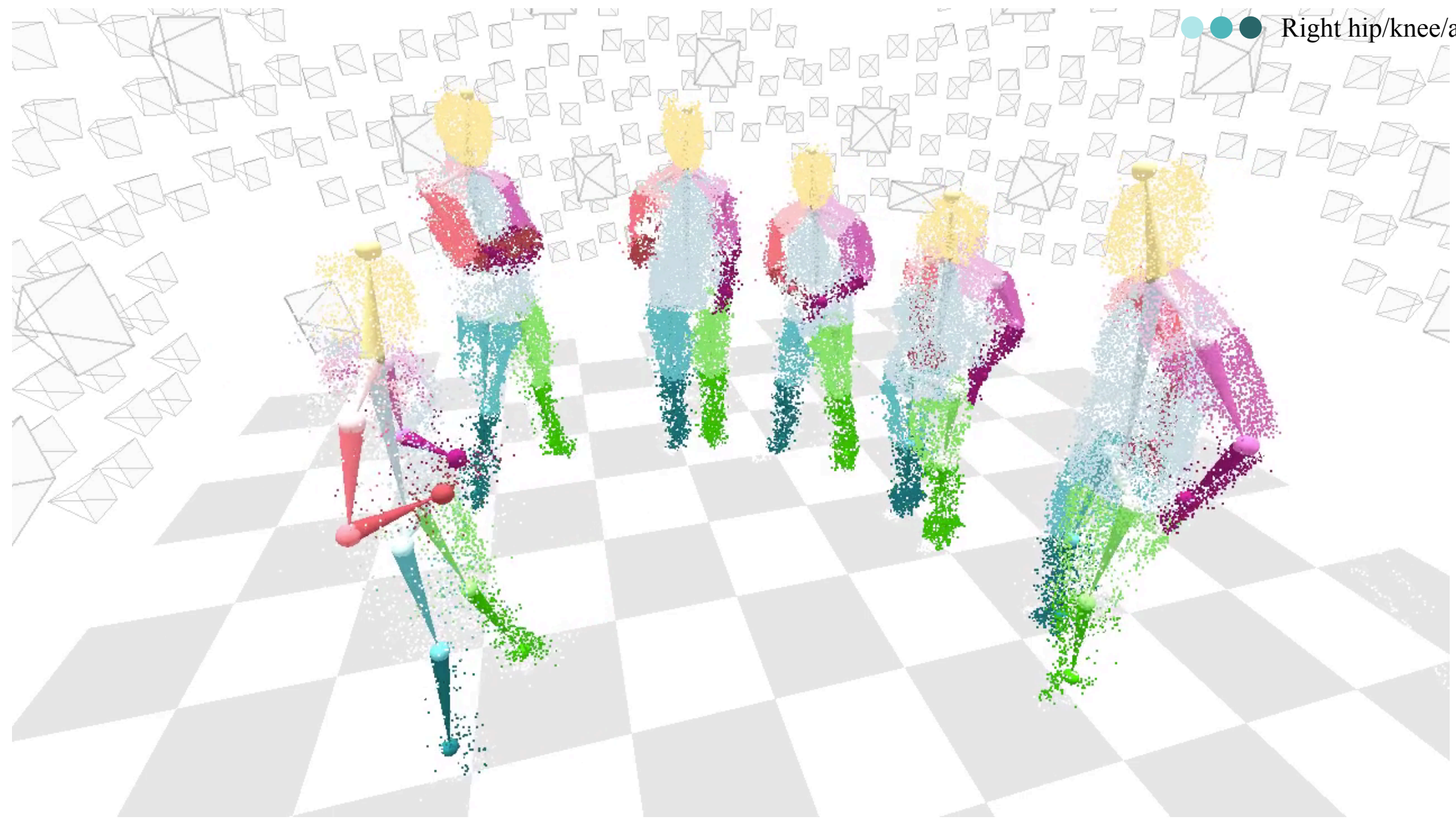
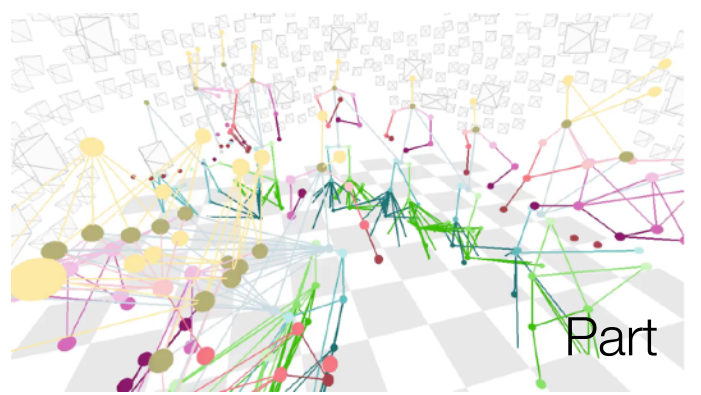
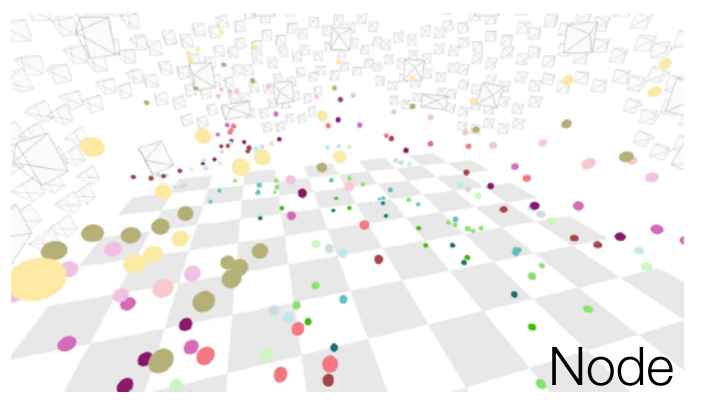
Generating "Part" Proposals

Fully Automatic Markerless Human Motion Capture



Fully Automatic Markerless Human Motion Capture

- ● ● HeadTop/neck/bodyCenter
- ● ● Left shoulder/elbow/wrist
- ● ● Right shoulder/elbow/wrist
- ● ● Left hip/knee/ankle
- ● ● Right hip/knee/ankle



Associating with Dense 3D Trajectories
Temporal Refinement

Fully Automatic Markerless Motion Capture

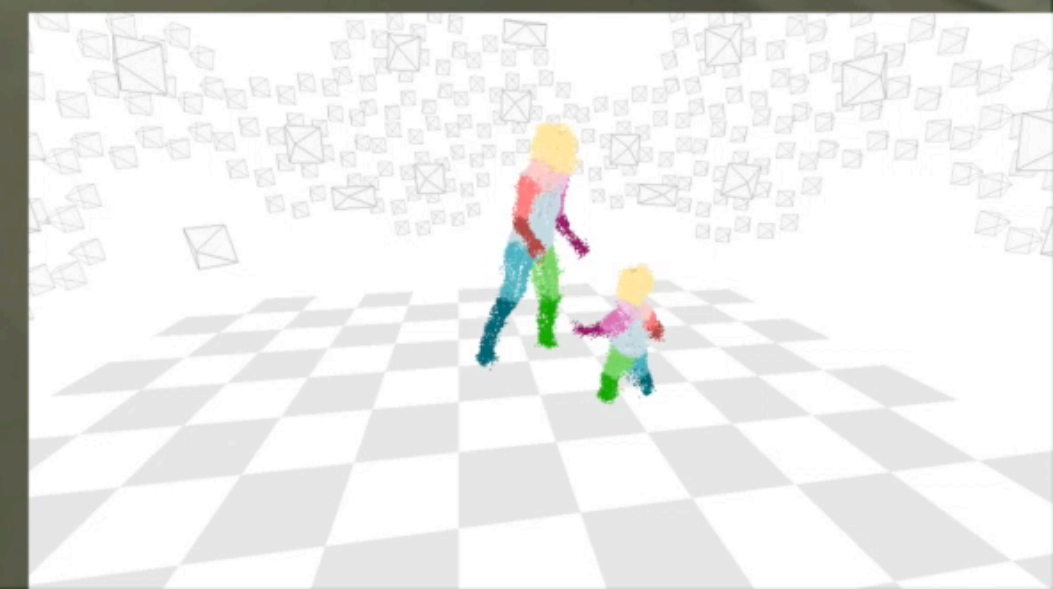
No assumption about the motion, appearance, and number of people



[Joo et al., ICCV 2015]

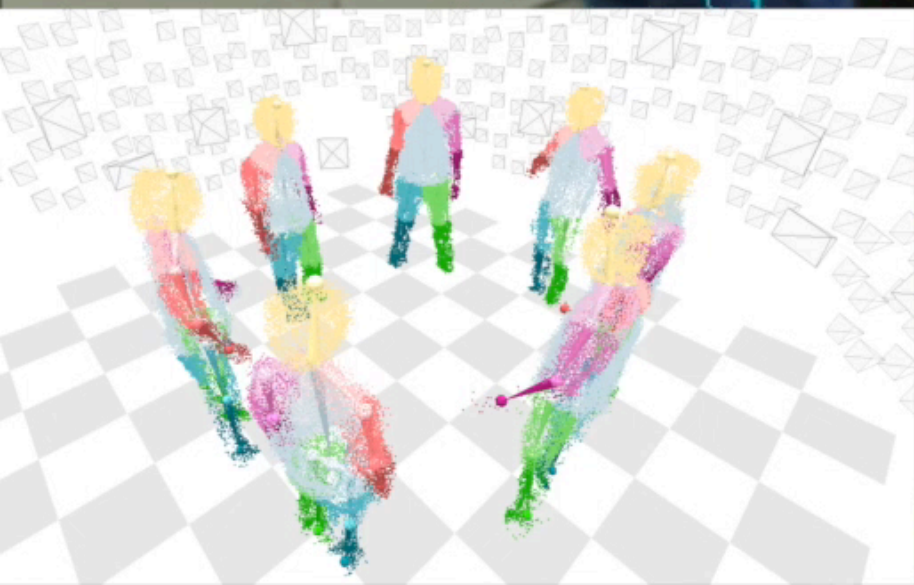
[Joo et al., TPAMI 2017]

Different Size of People



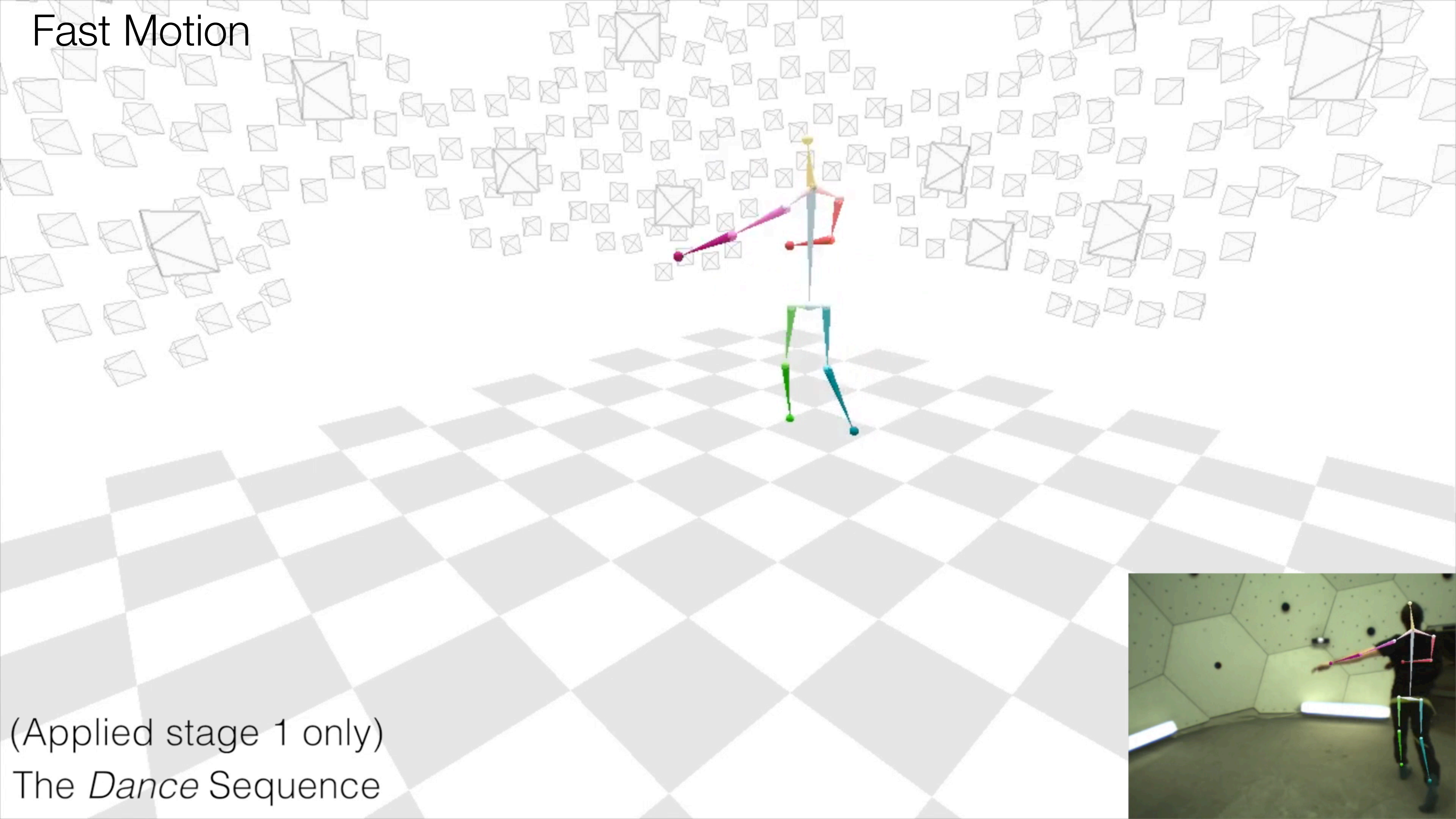
The *Ian* Sequence

Severe Occlusions

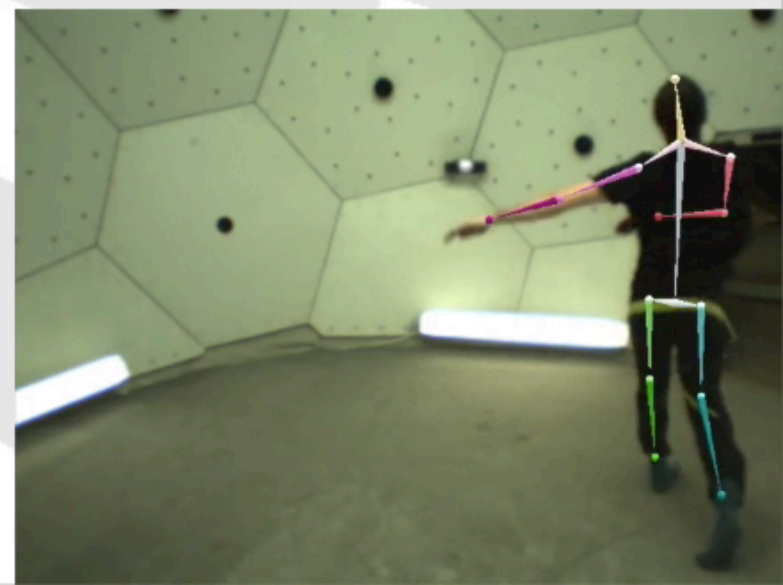


The 151125 *Bang* Sequence

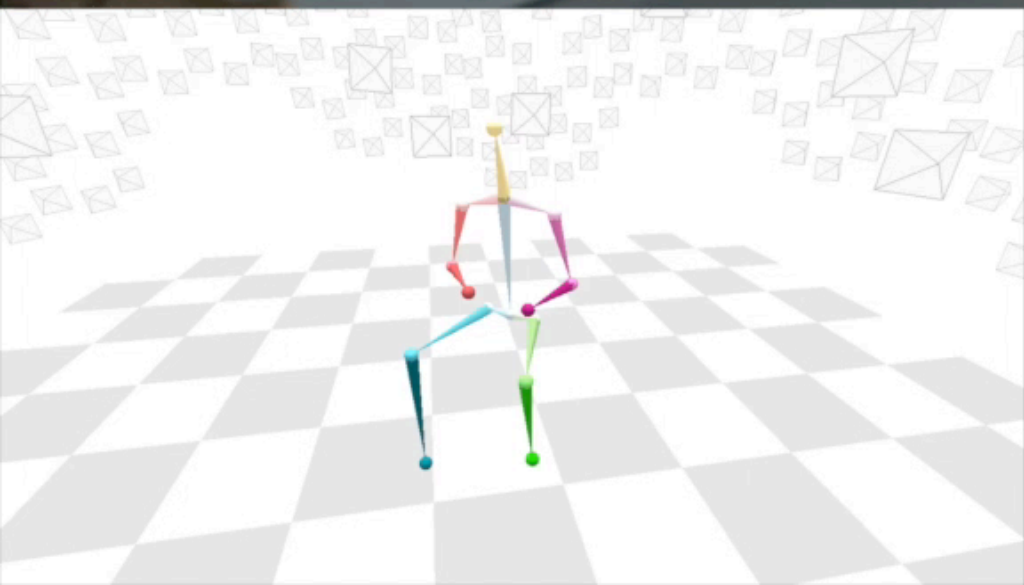
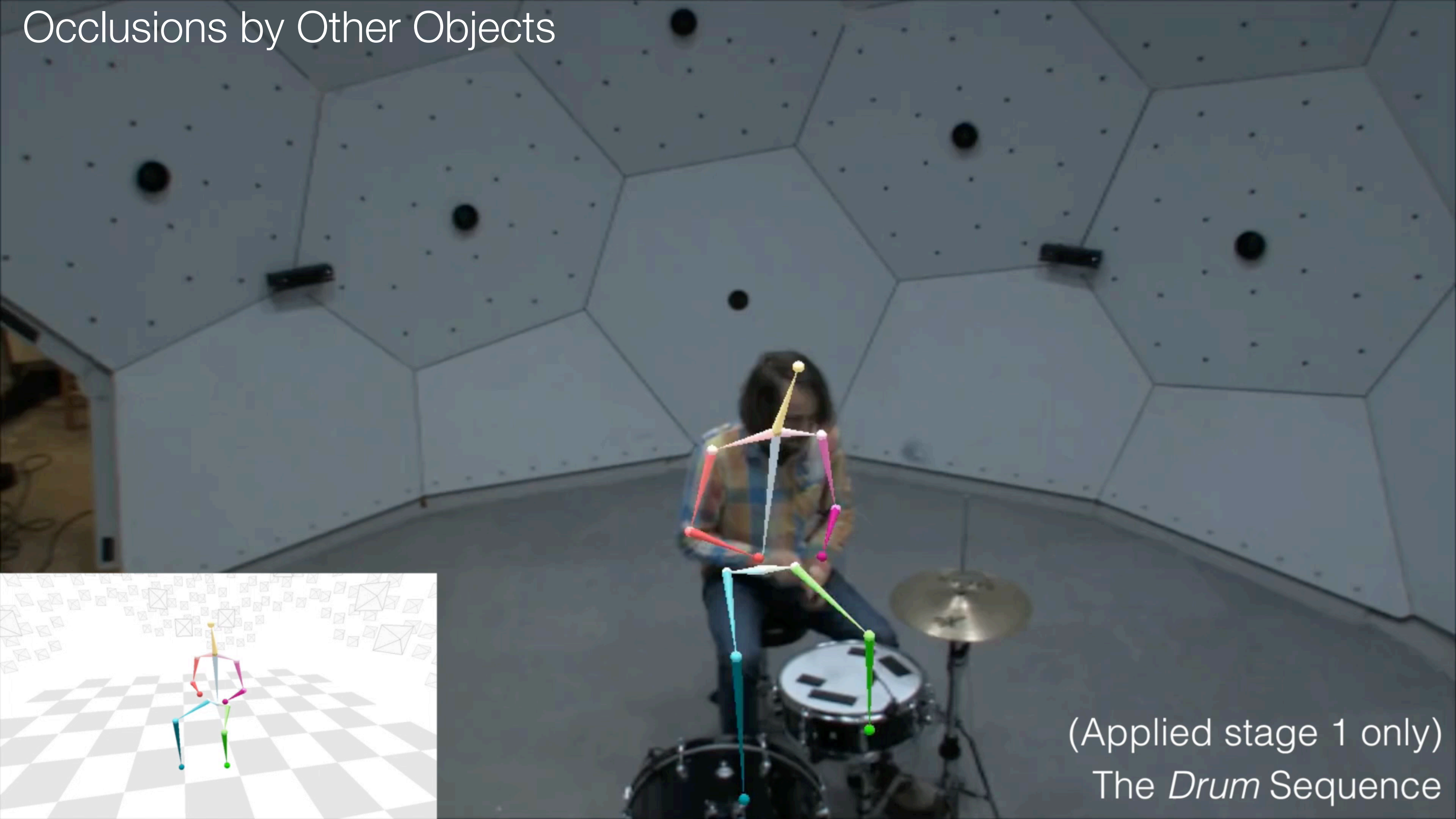
Fast Motion



(Applied stage 1 only)
The *Dance* Sequence



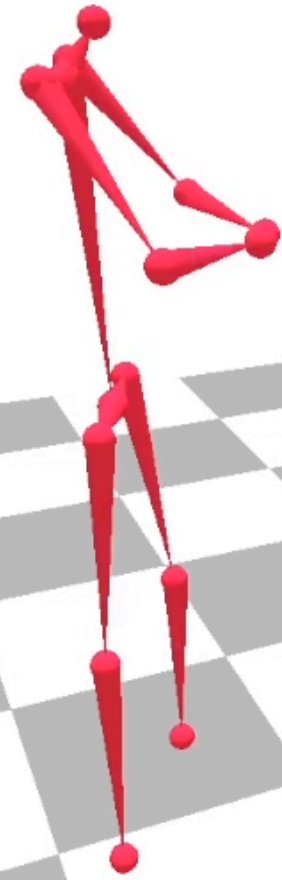
Occlusions by Other Objects



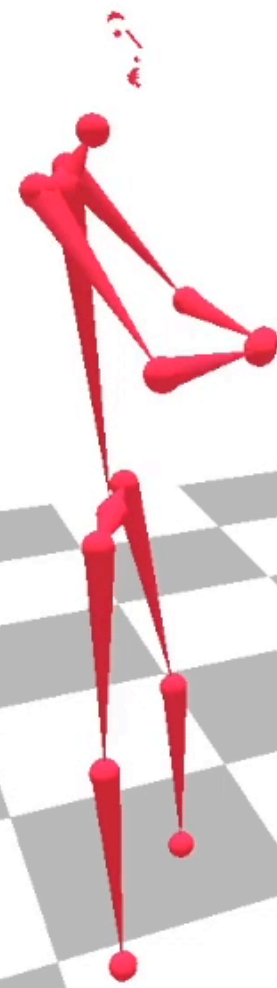
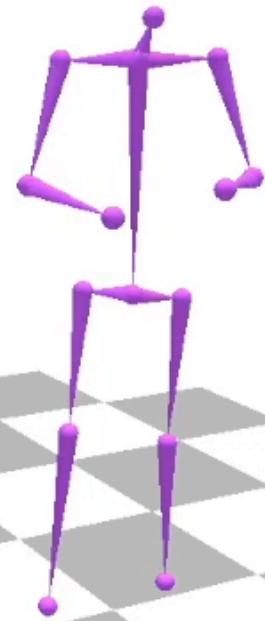
(Applied stage 1 only)
The *Drum* Sequence

Are Body and Face Enough?

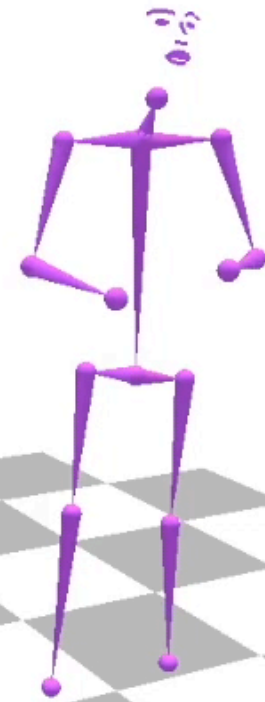
Important Nuances are Embedded in Hand Motion



Body Only



Face+Body



Face+Body+Hand

How To Make A Good 2D Hand Pose Detector

Leveraging Recently Advanced Deep Learning Framework



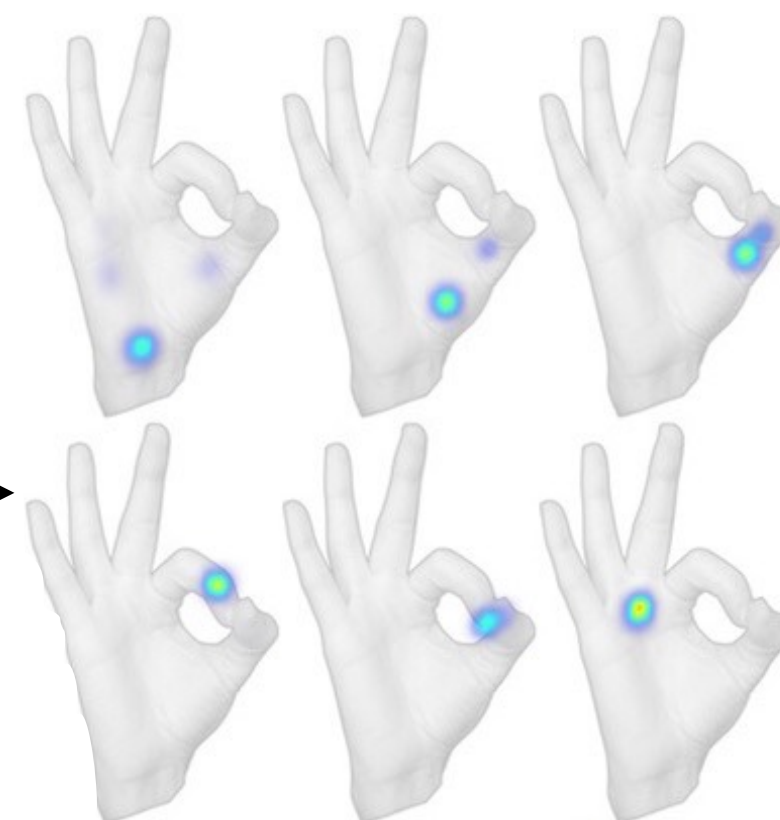
Image Input



Training Data



Deep Learning



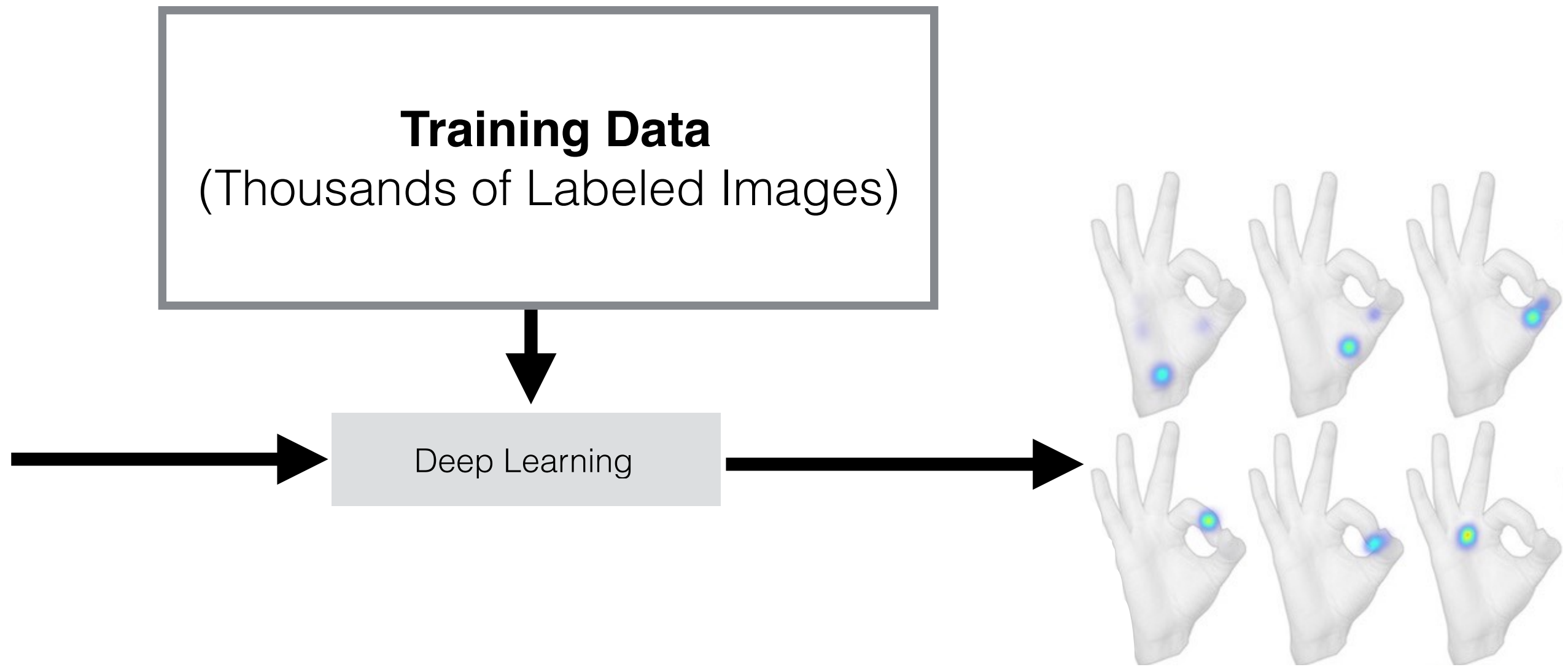
Part Confidence Maps

How To Make A Good 2D Hand Pose Detector

Depends on Availability of A Large Scale Dataset



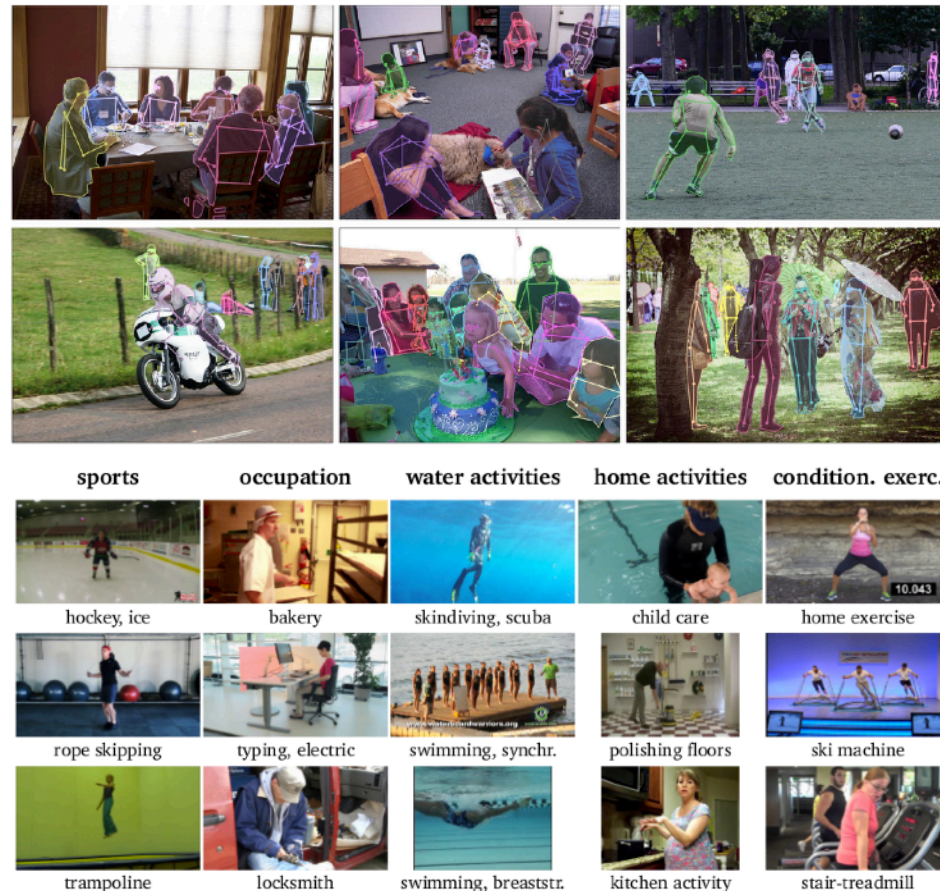
Image Input



Part Confidence Maps

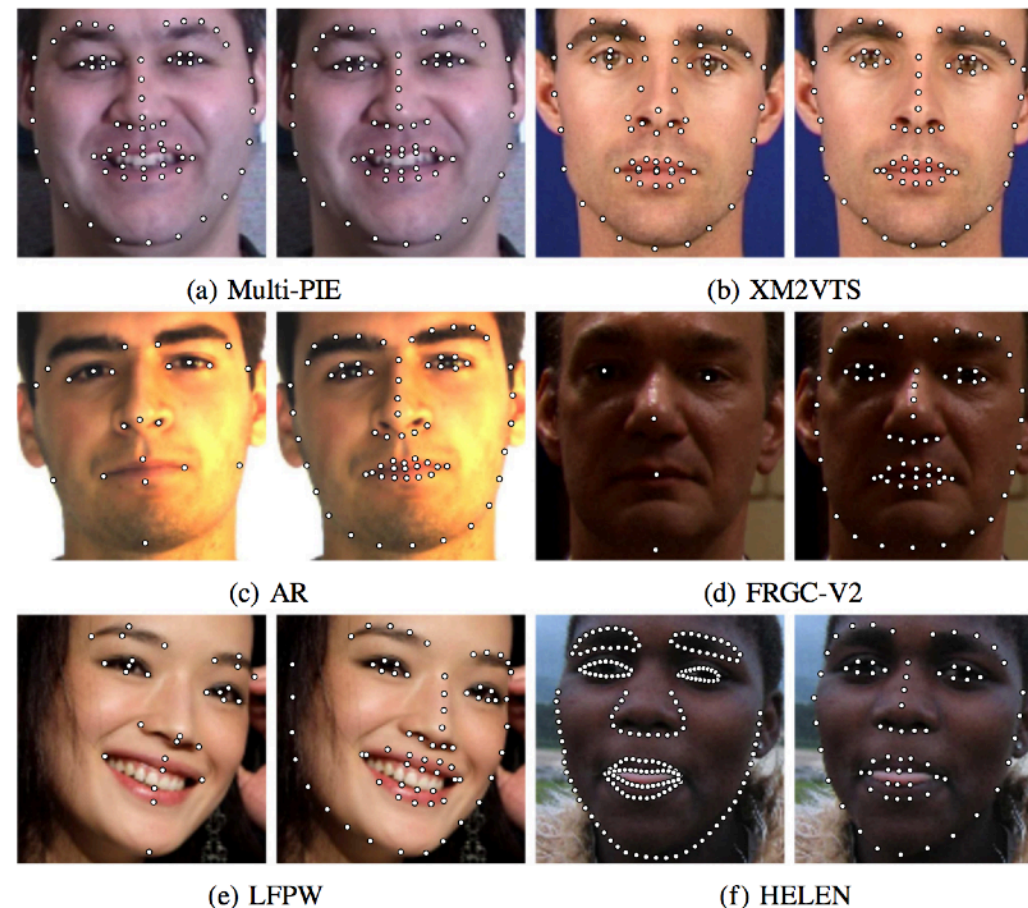
Human Keypoint Detectors from Single RGB image?

No Available Hand Keypoint Detector and Dataset



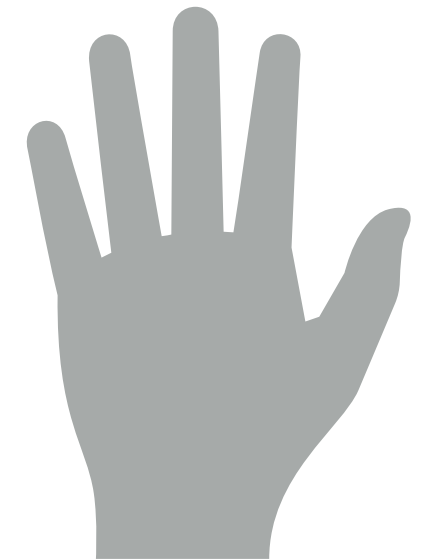
MPII with **40K** annotations
MS COCO with **250K** annotations

Wei et al., 2016, Newell et al., 2016
Cao et al., 2017, He et al., 2017



300-VW with **218K** annotations
ALFW with **26K** annotations
PUT with **10K** annotations
MUCT with **3K** annotations

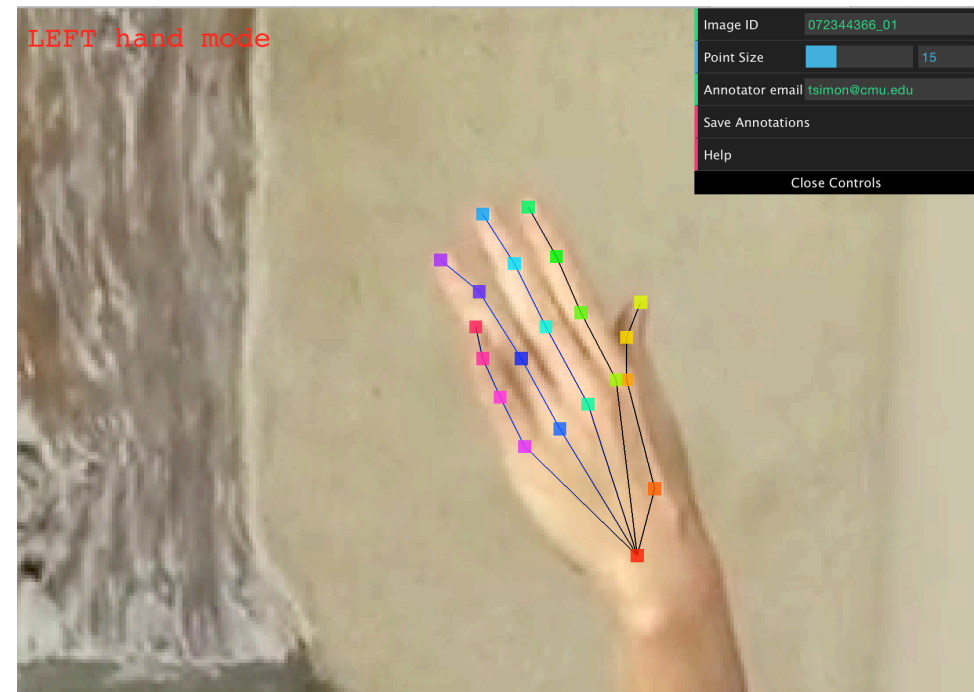
DeepFace 2014, FaceNet 2015
OpenFace 2016



No large scale dataset

Rehg and Kanade 1994, Lu et al., 2003
Stenger et al., 2006, Gorce et al., 2011

ANNOTATORS NEEDED TO LABEL IMAGES



We are looking for people to help annotate landmarks in images and video. The ideal candidate should be consistent, self-motivated, and have great attention to detail. The position will be paid hourly at \$12/hour, hours flexible.

- Work from home using any browser.
- ATTENTION TO DETAIL required.
- Proofreading and/or editing skills helpful
- Payment is up to \$12 per hour

Contact: Tomas Simon (tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

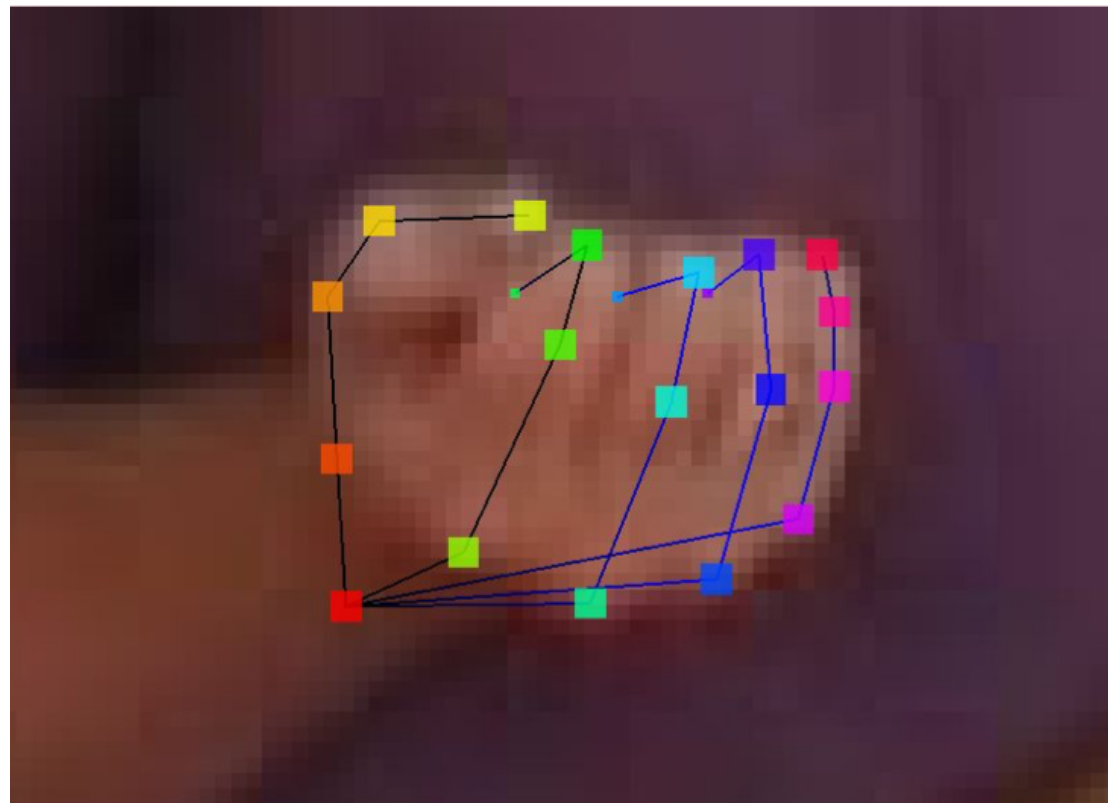
Tomas Simon
(tsimon@cs.cmu.edu)

Tomas Simon
(tsimon@cs.cmu.edu)

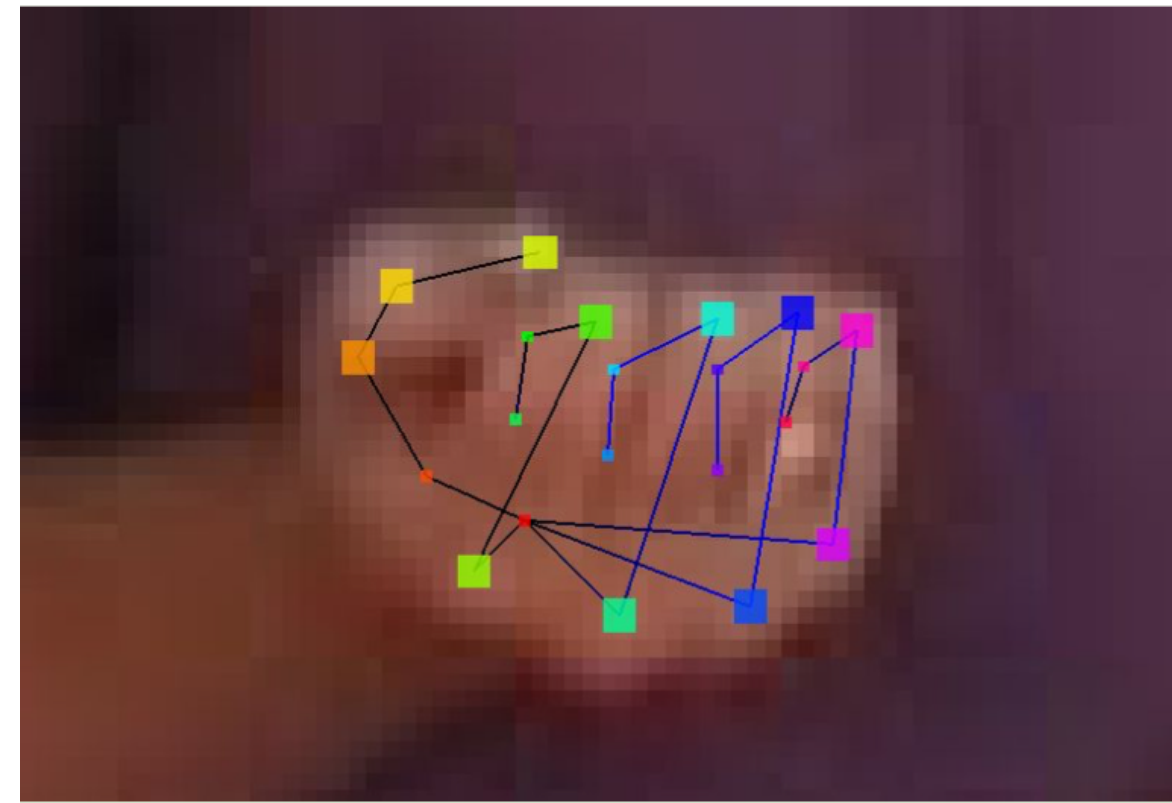
Tomas Simon
(tsimon@cs.cmu.edu)

How To Make A Good 2D Hand Pose Detector

Difficulties in Labeling Hand Joints



Annotator 1

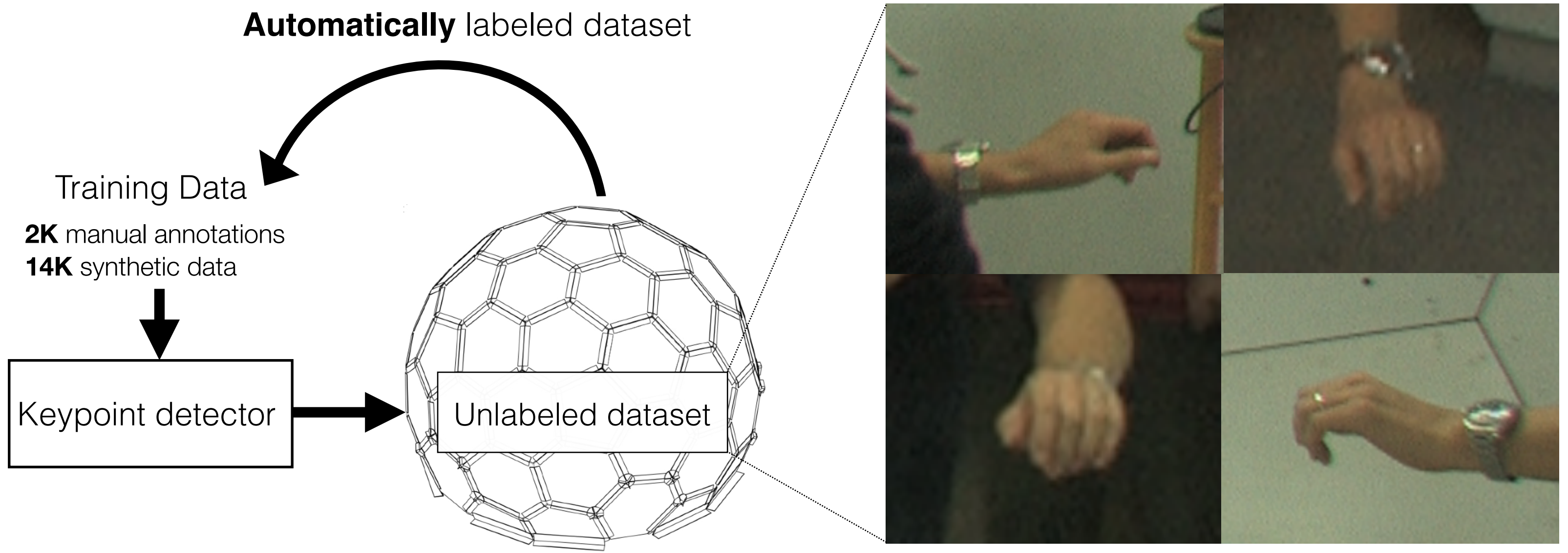


Annotator 2

Occluded joints should be guessed
We ended up generating **2K** images

Multiview Bootstrapping

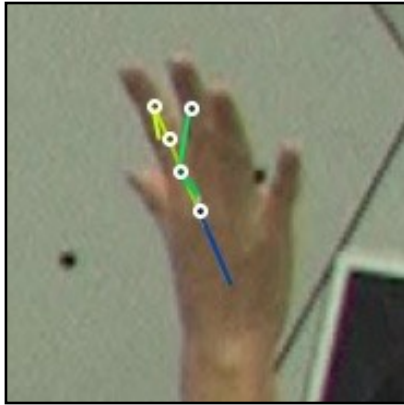
Geometric Cues as A Supervision



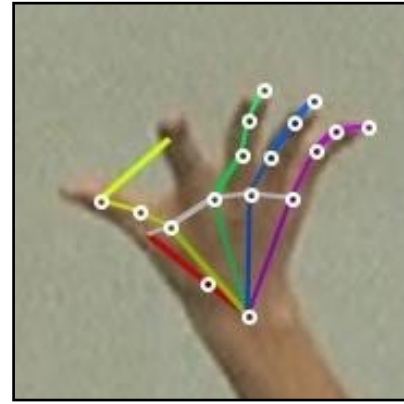
Multiview Bootstrapping

An Example

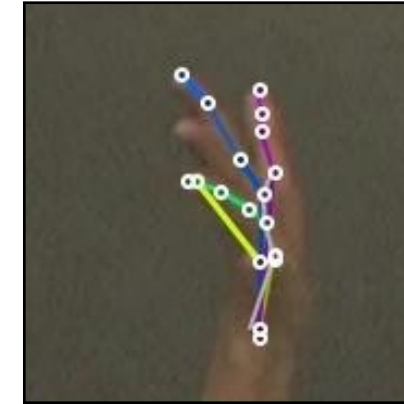
● View 7



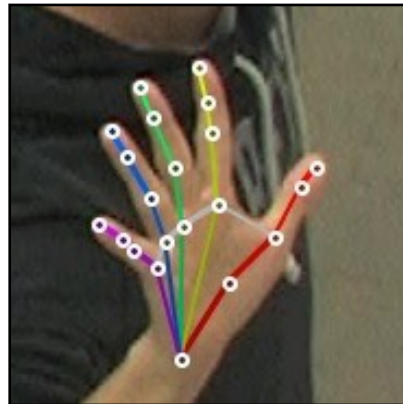
● View 6



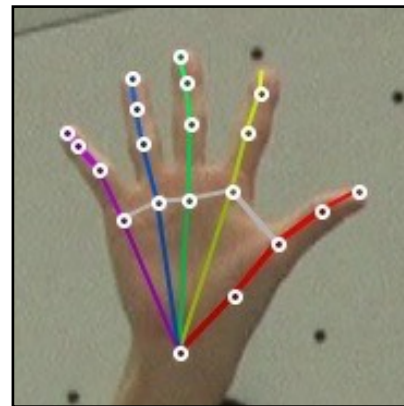
● View 5



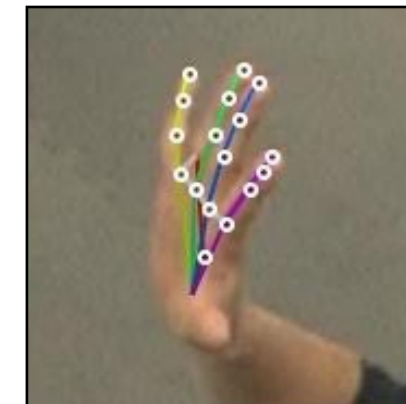
● View 1



● View 2



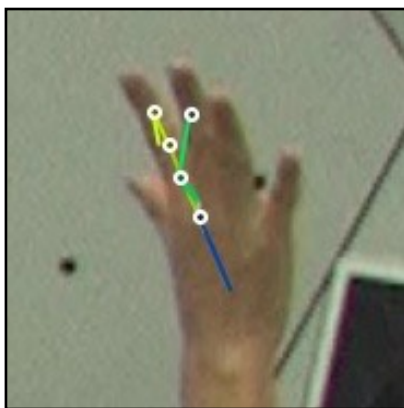
● View 3



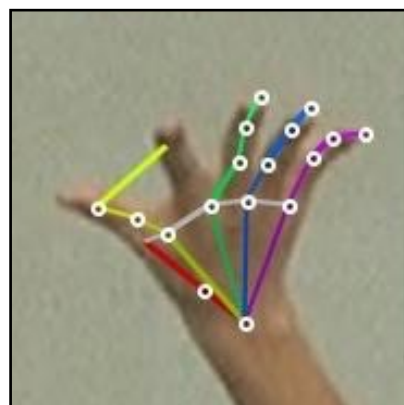
Multiview Bootstrapping

Checking Ray Intersection

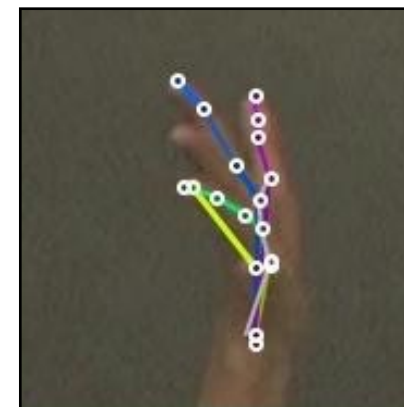
● View 7



● View 6

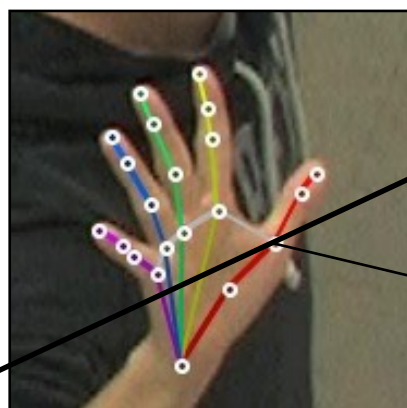


● View 5



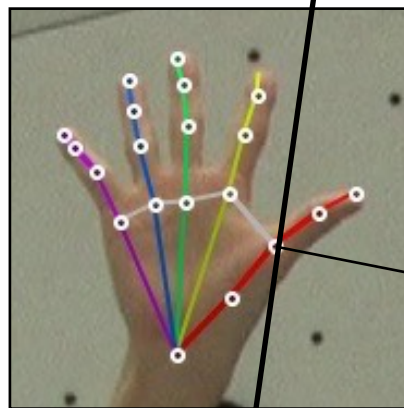
X

● View 1



X₁

● View 2



X₂

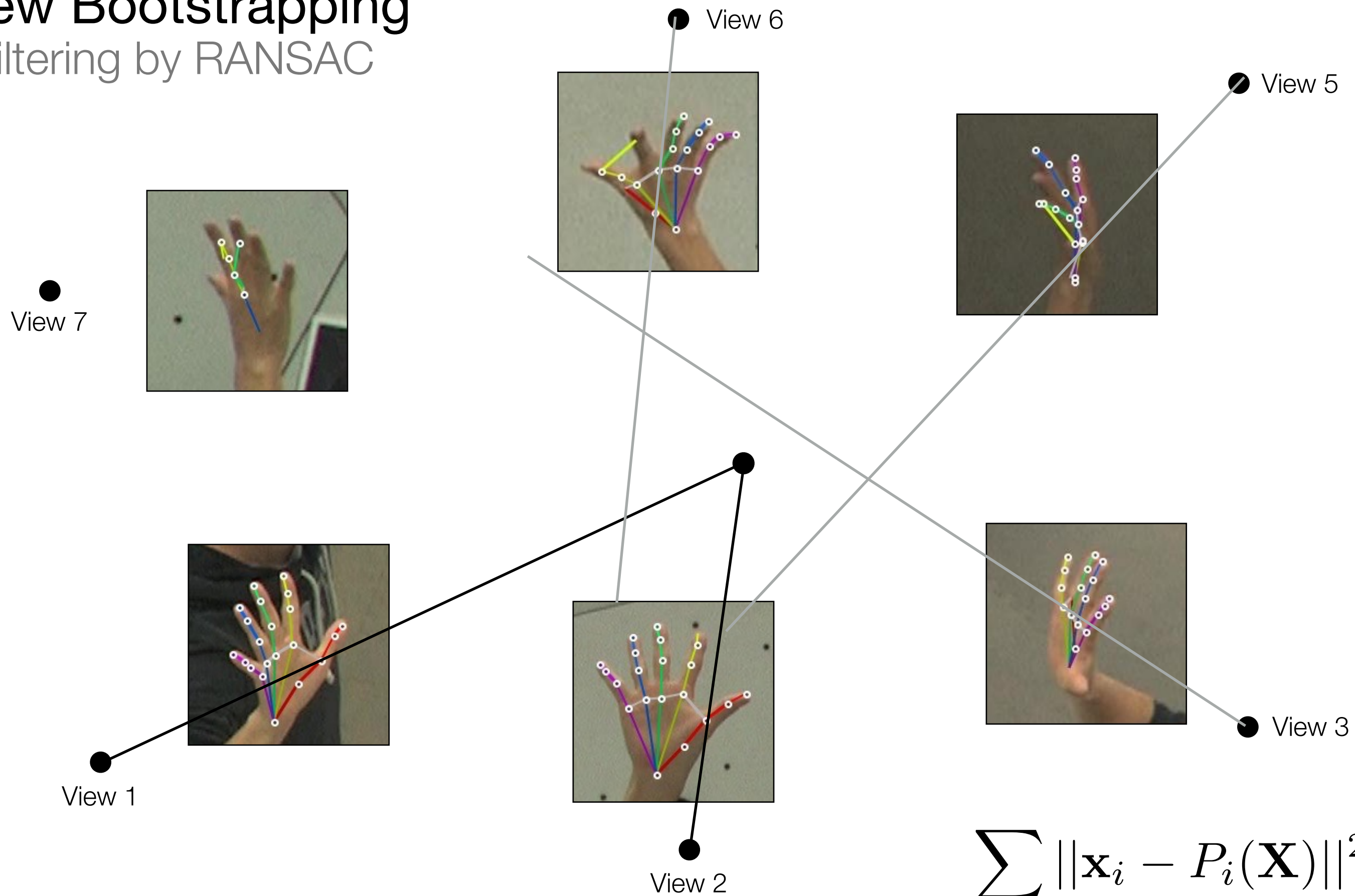
● View 3



$$\sum_i ||\mathbf{x}_i - P_i(\mathbf{X})||^2 < \tau$$

Multiview Bootstrapping

Outlier Filtering by RANSAC



$$\sum_i ||\mathbf{x}_i - P_i(\mathbf{X})||^2 < \tau$$

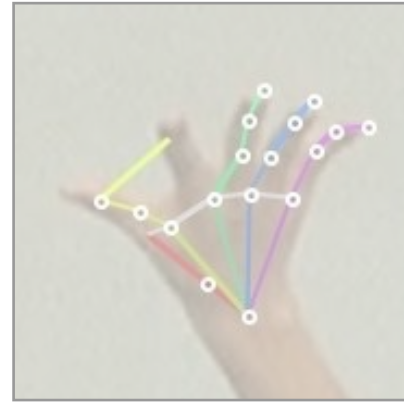
Multiview Bootstrapping

Outlier Filtering by RANSAC

● View 7



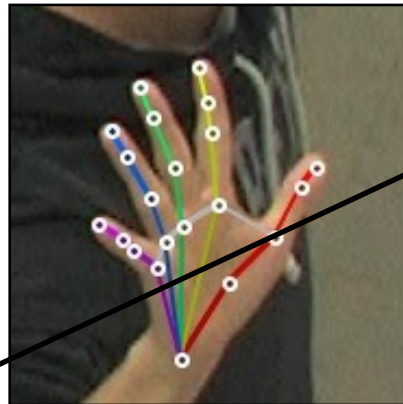
● View 6



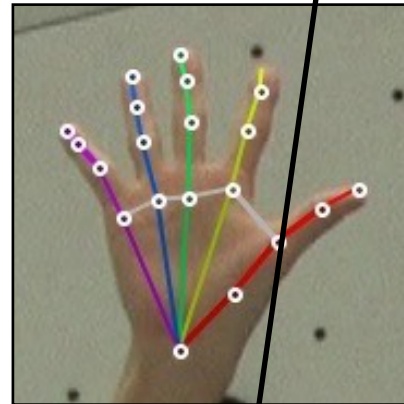
● View 5



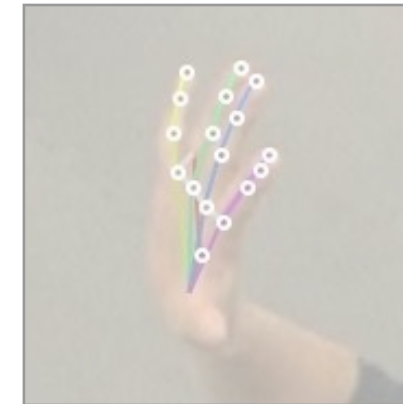
● View 1



● View 2



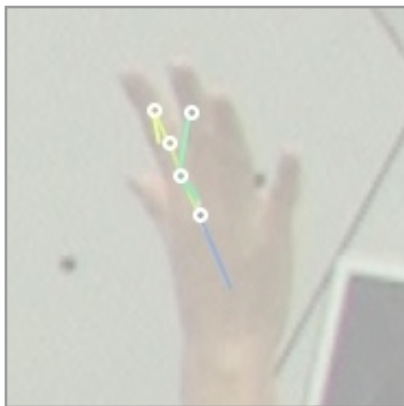
● View 3



Multiview Bootstrapping

3D Reconstruction

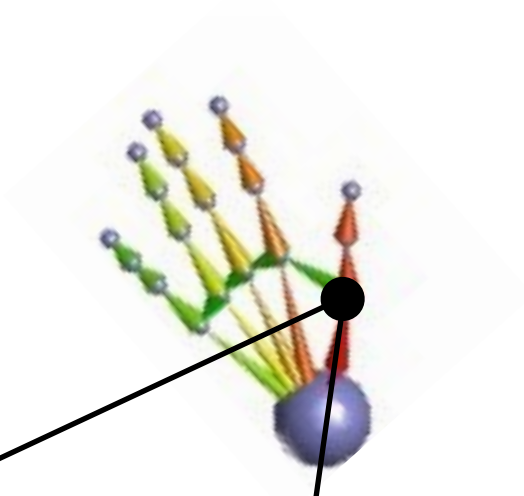
● View 7



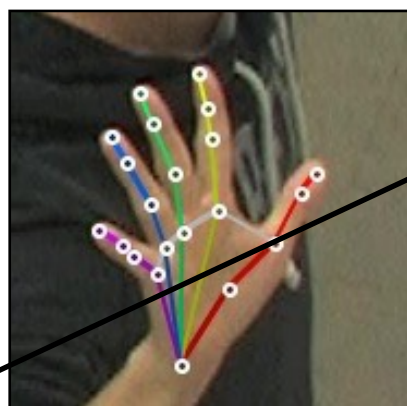
● View 6



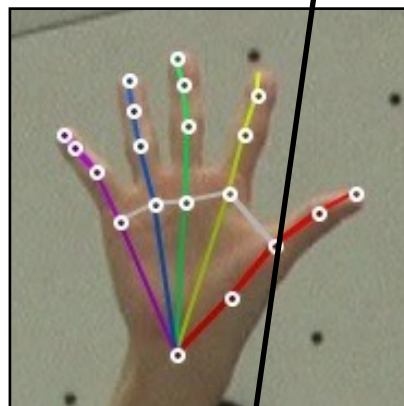
● View 5



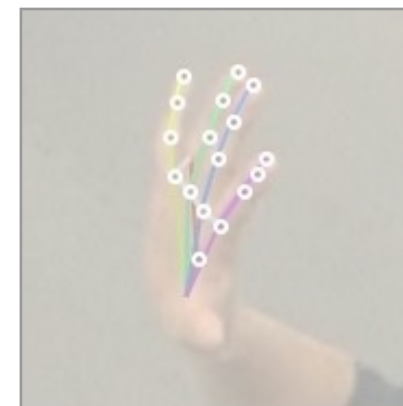
● View 1



● View 2



● View 3



Multiview Bootstrapping

Reprojection

● View 7



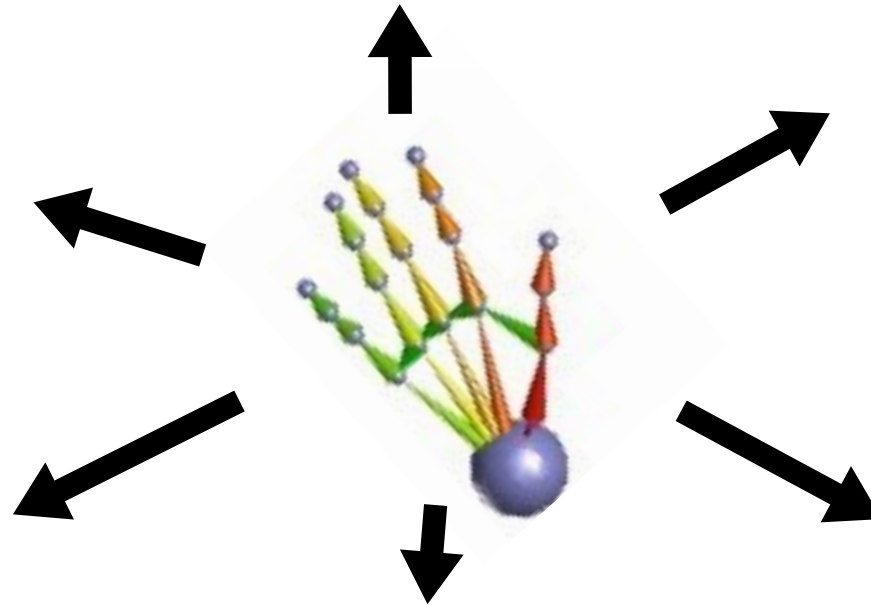
● View 1



● View 6



● View 5



● View 2

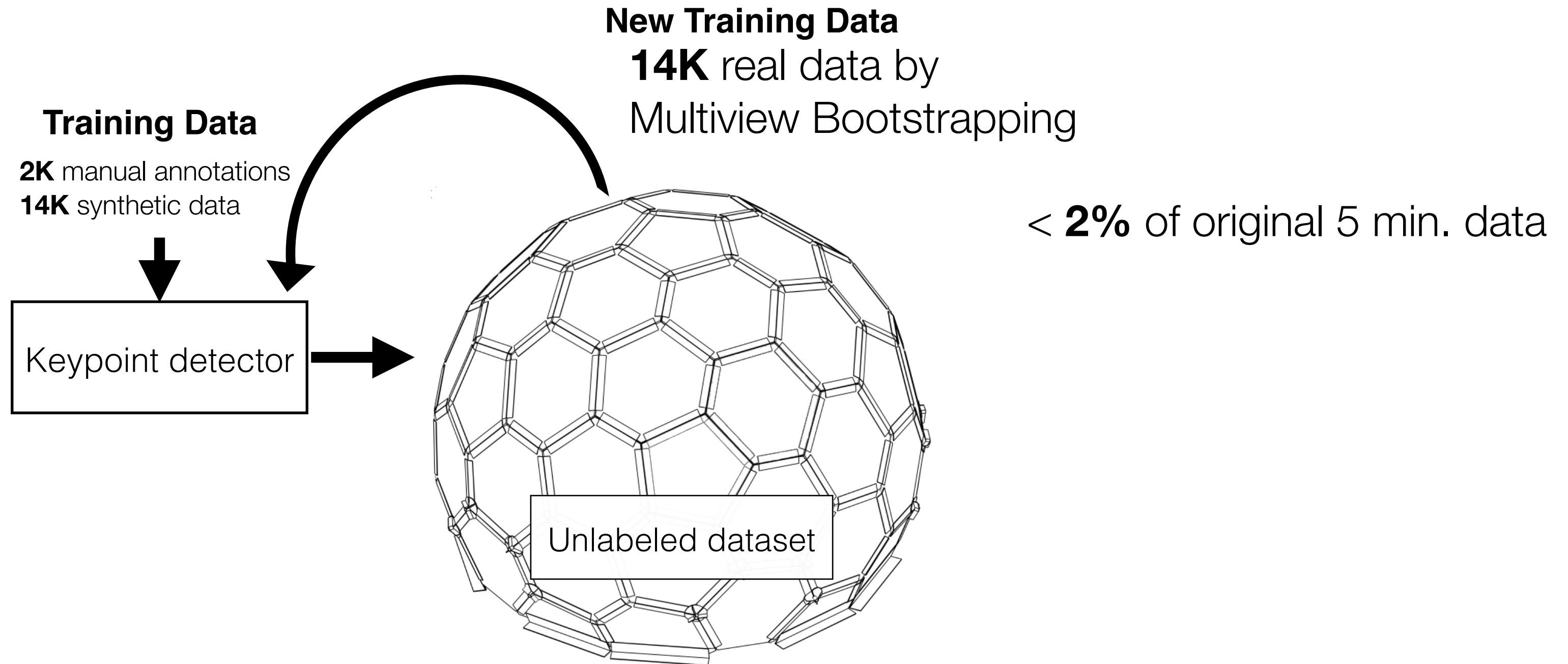


● View 3

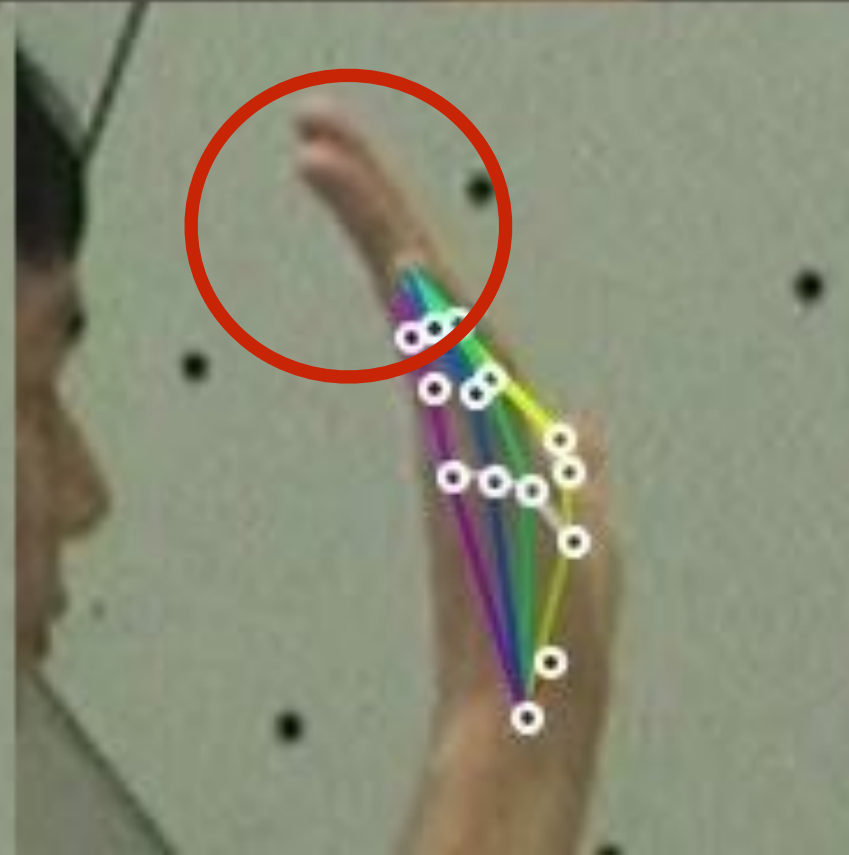
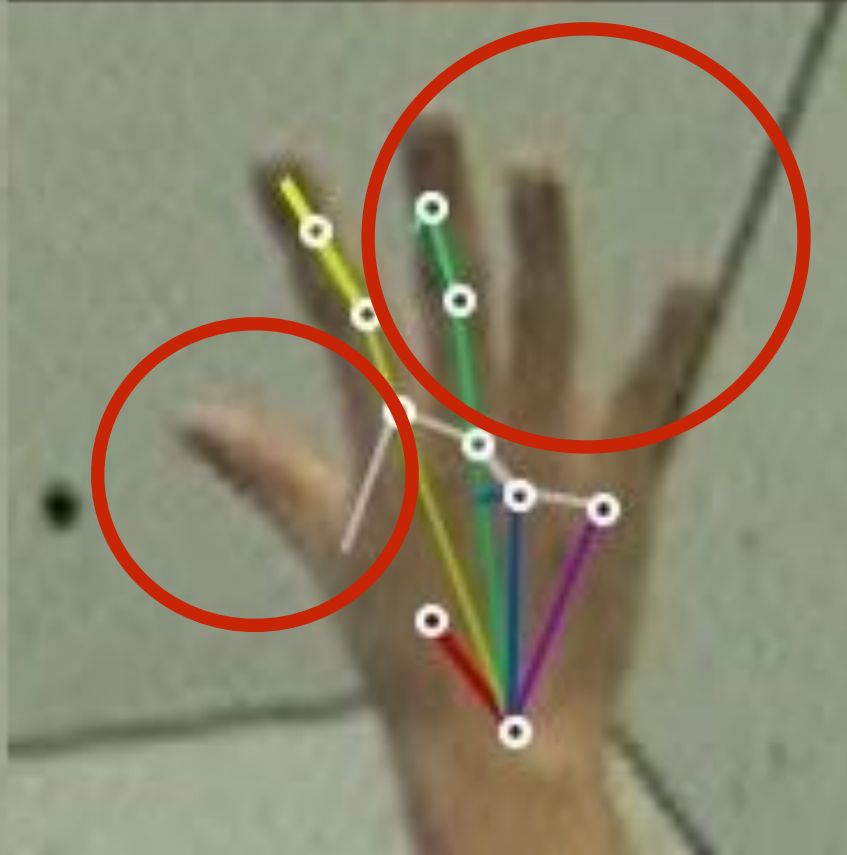
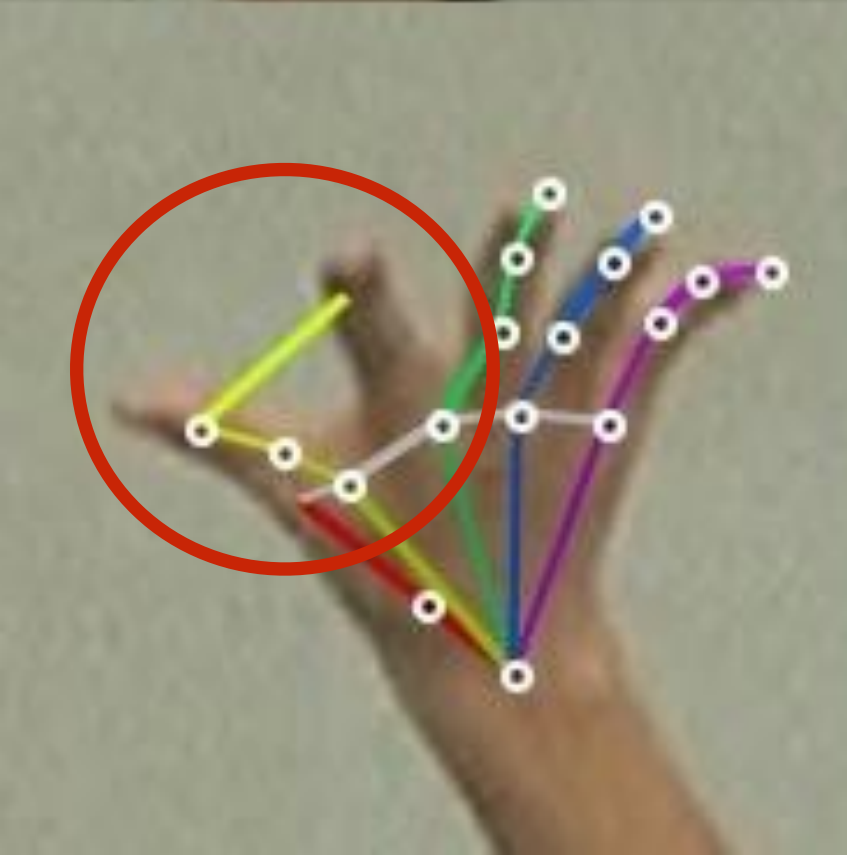
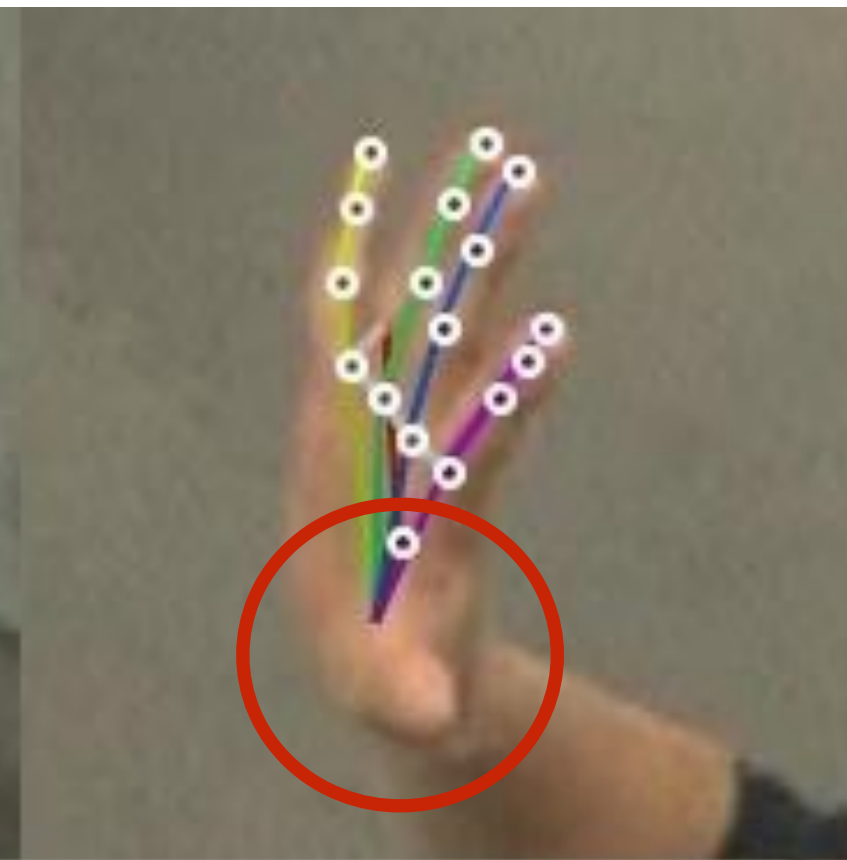
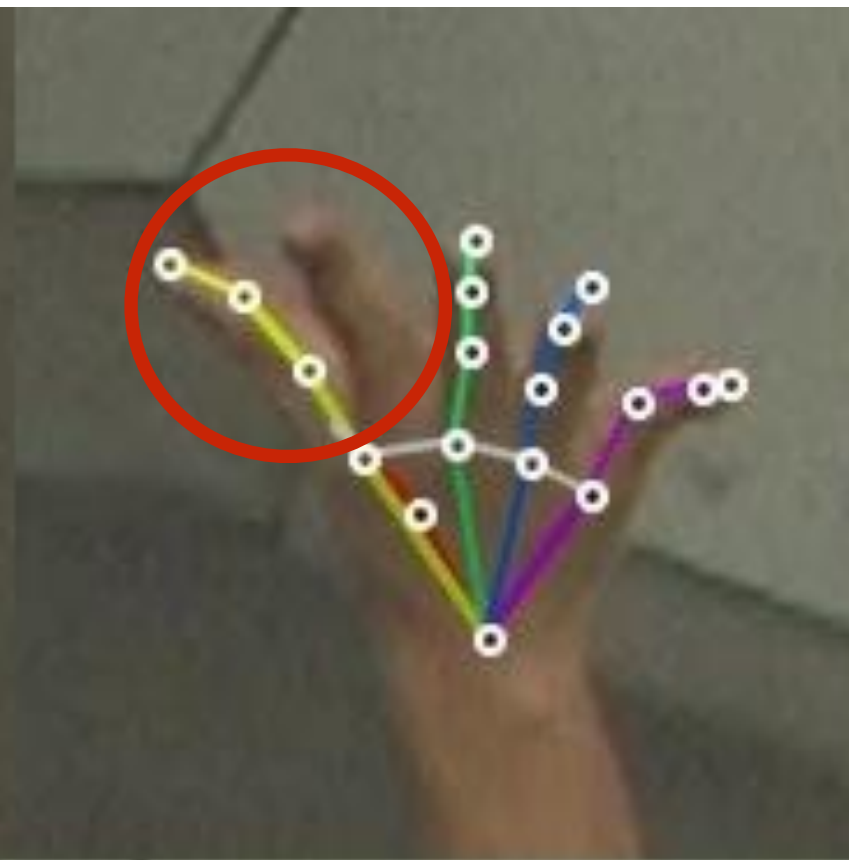
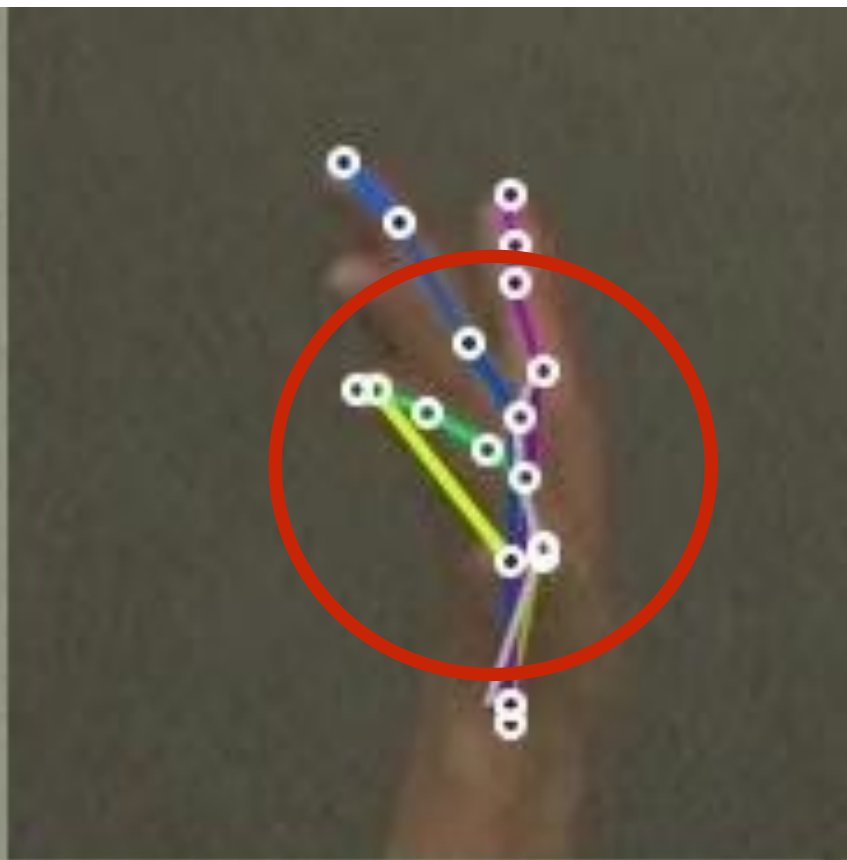
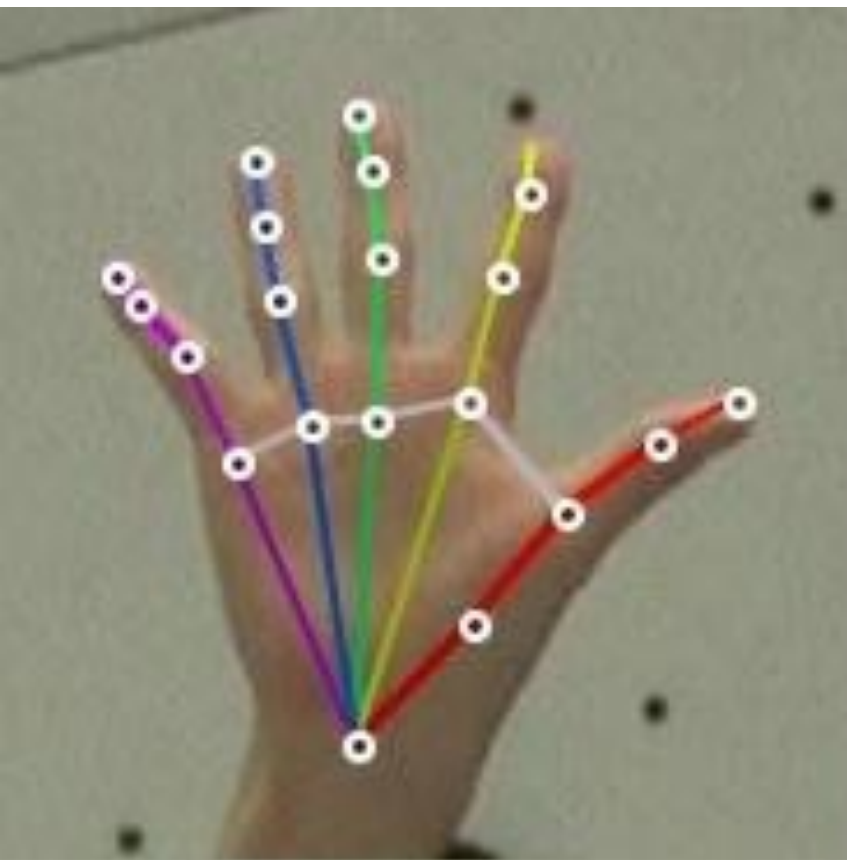


Multiview Bootstrapping

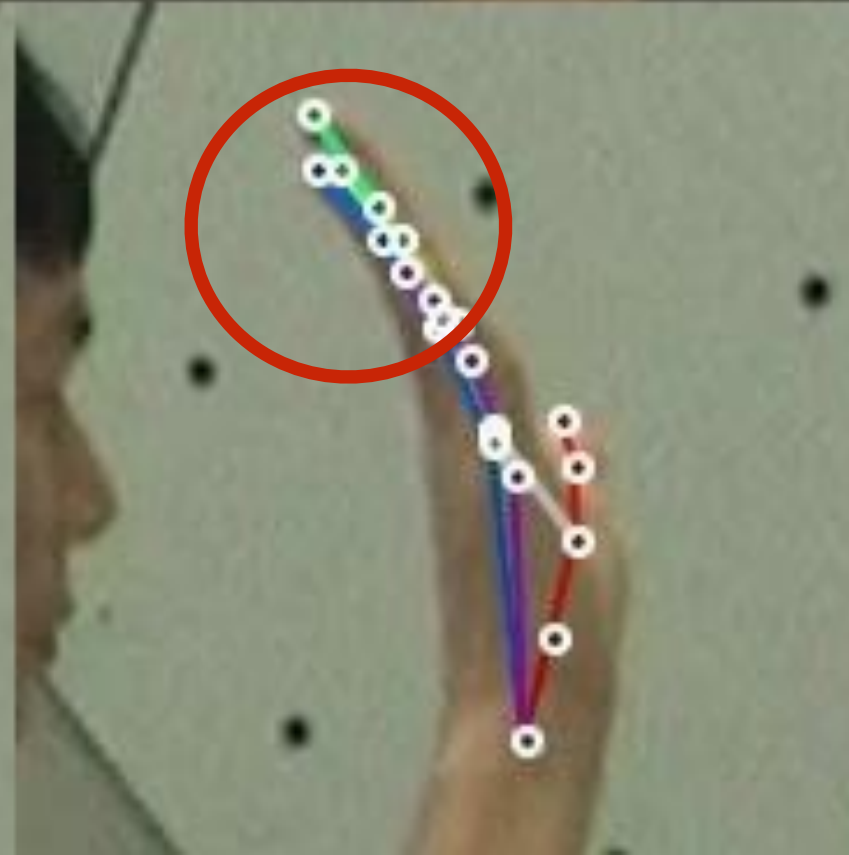
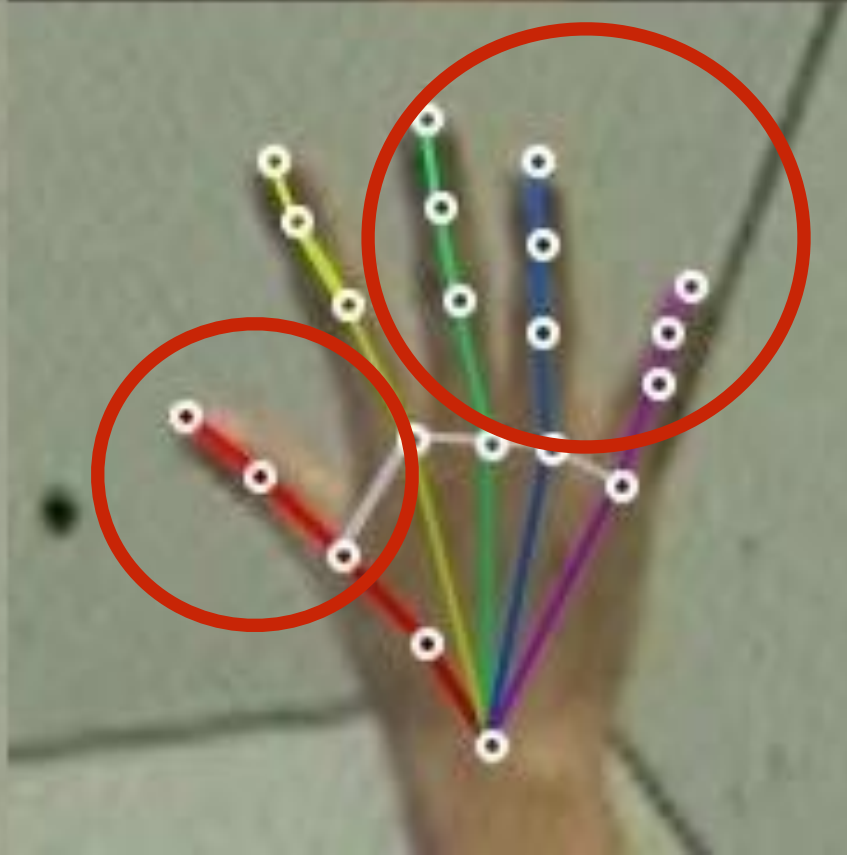
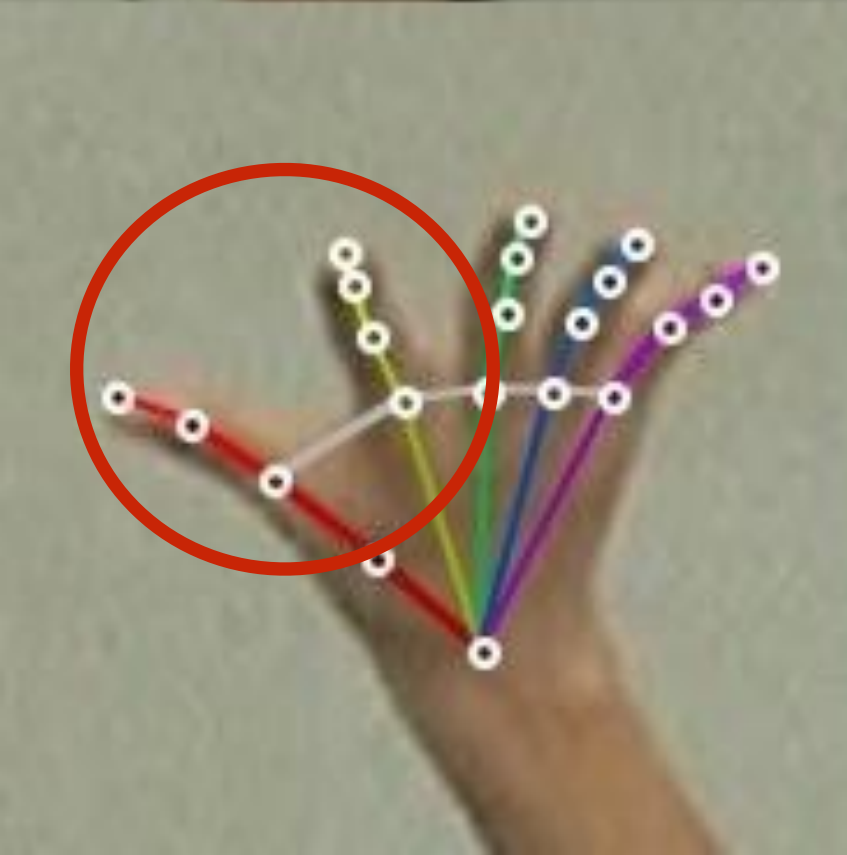
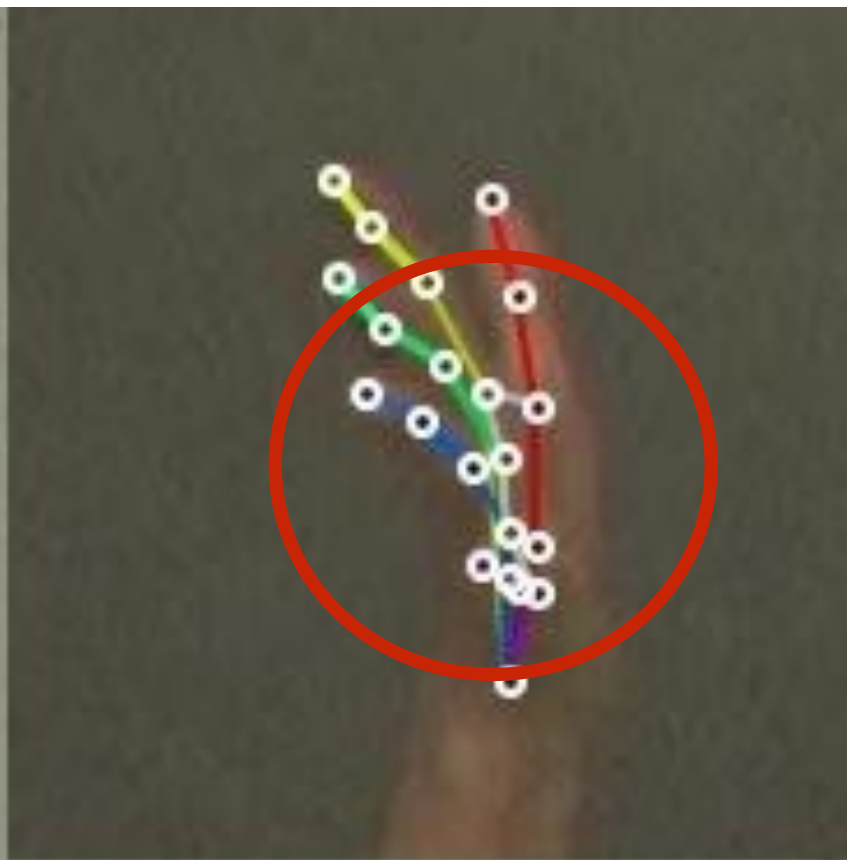
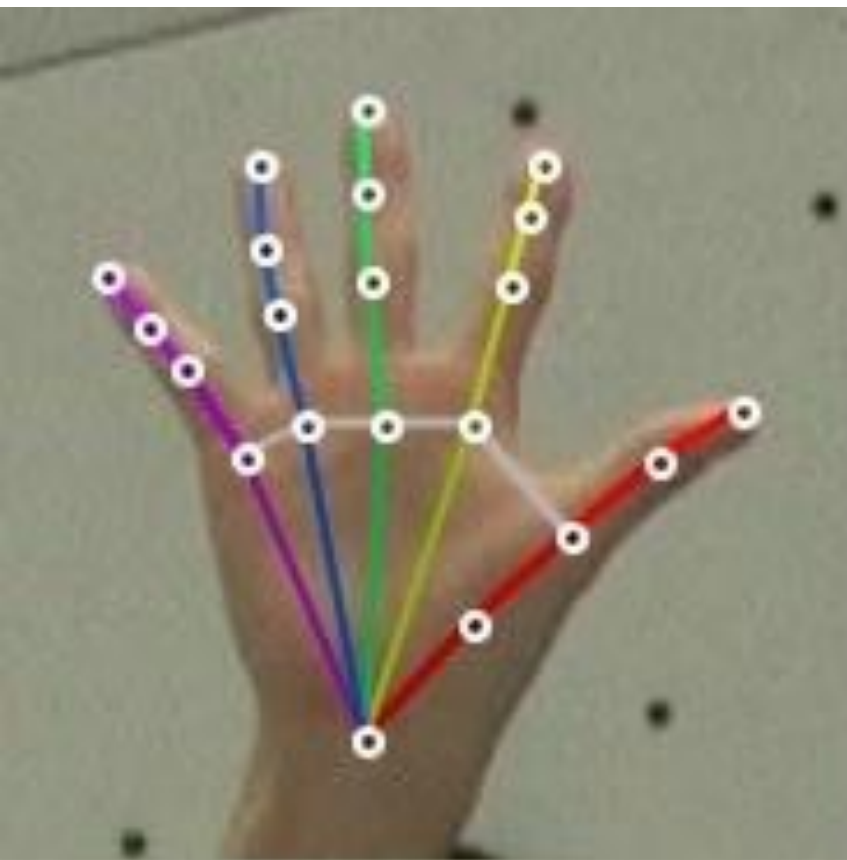
A Method to Automatically Generate Annotations



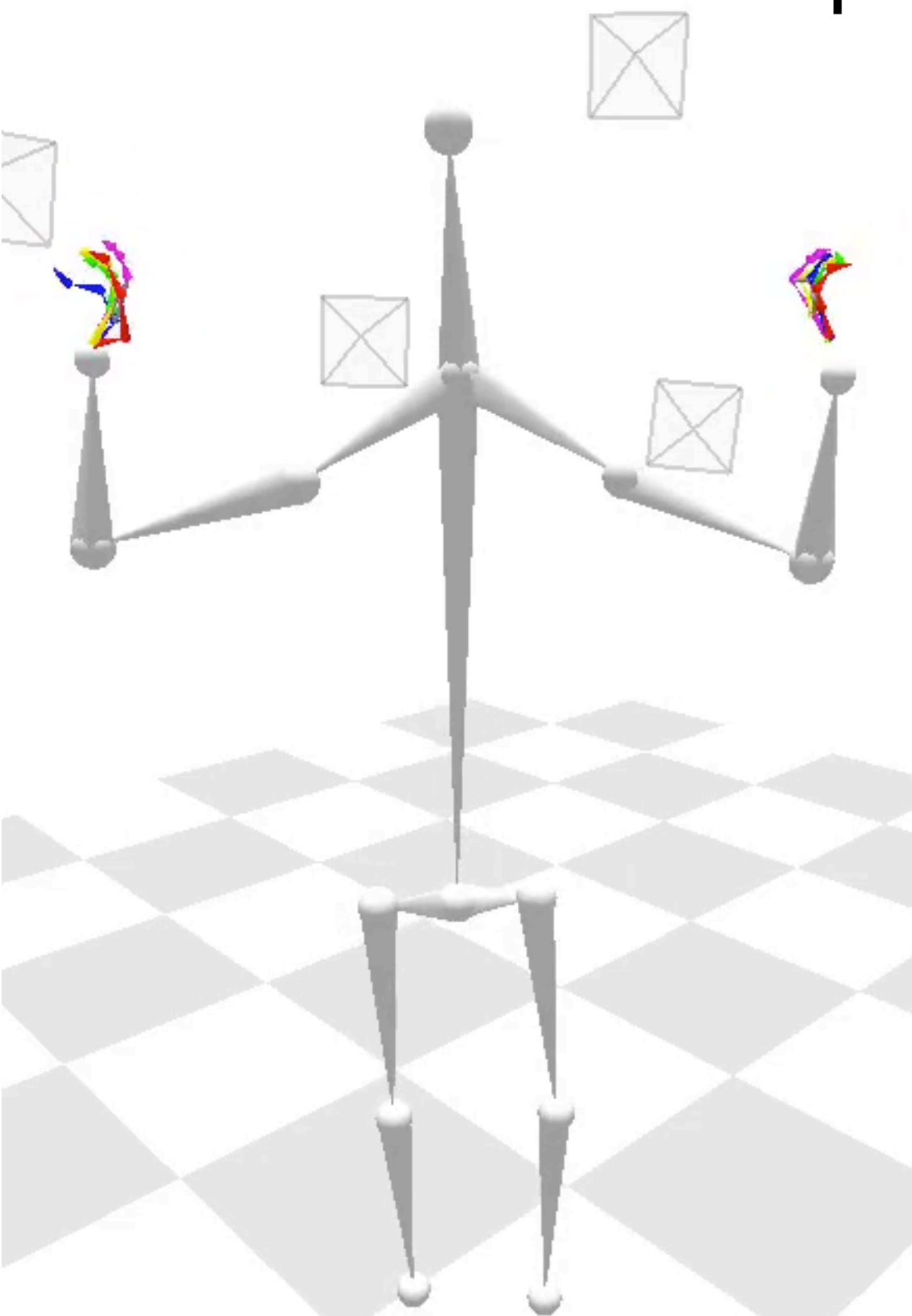
2D Detection (Initial)



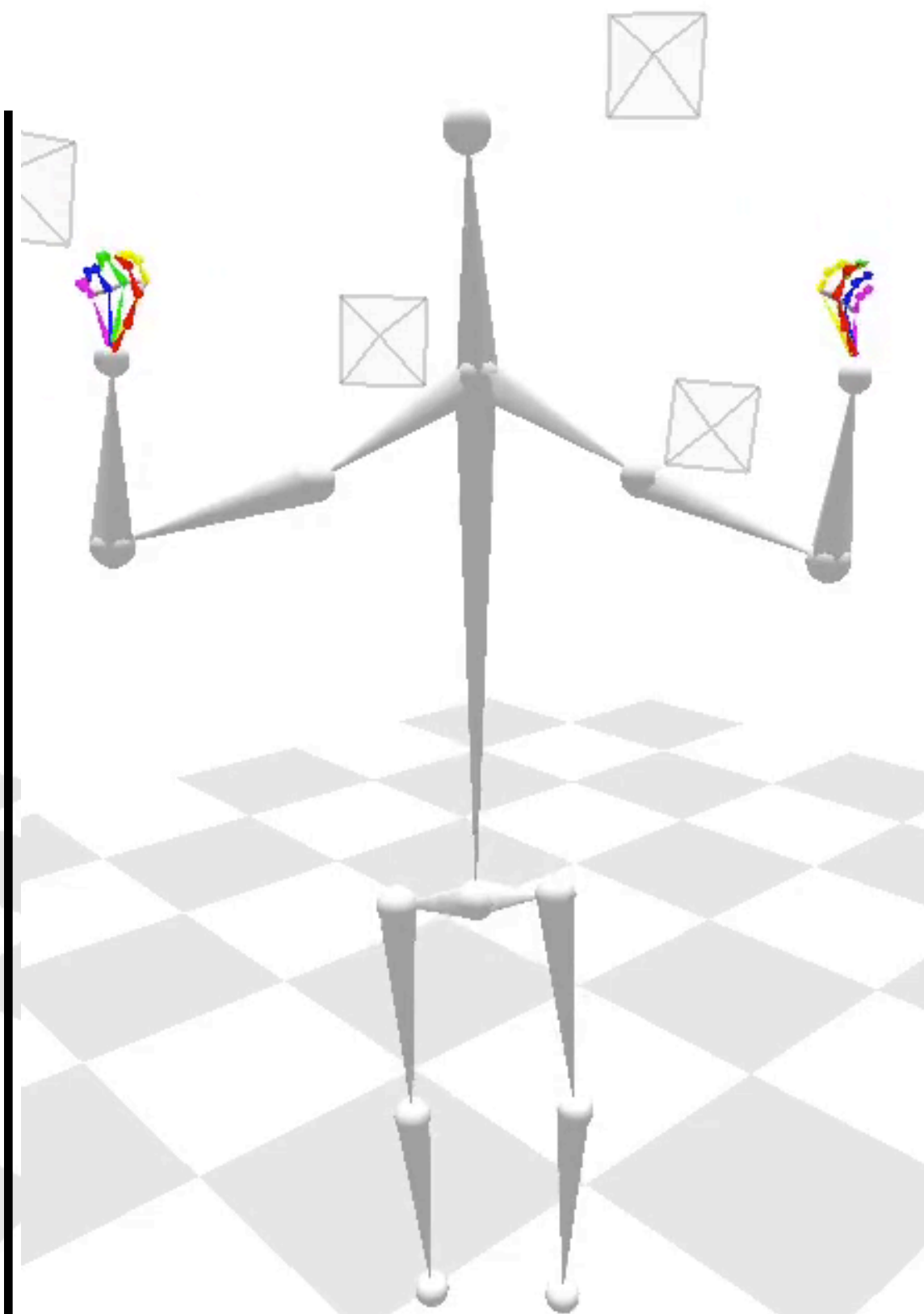
2D Detection (Iteration 1)



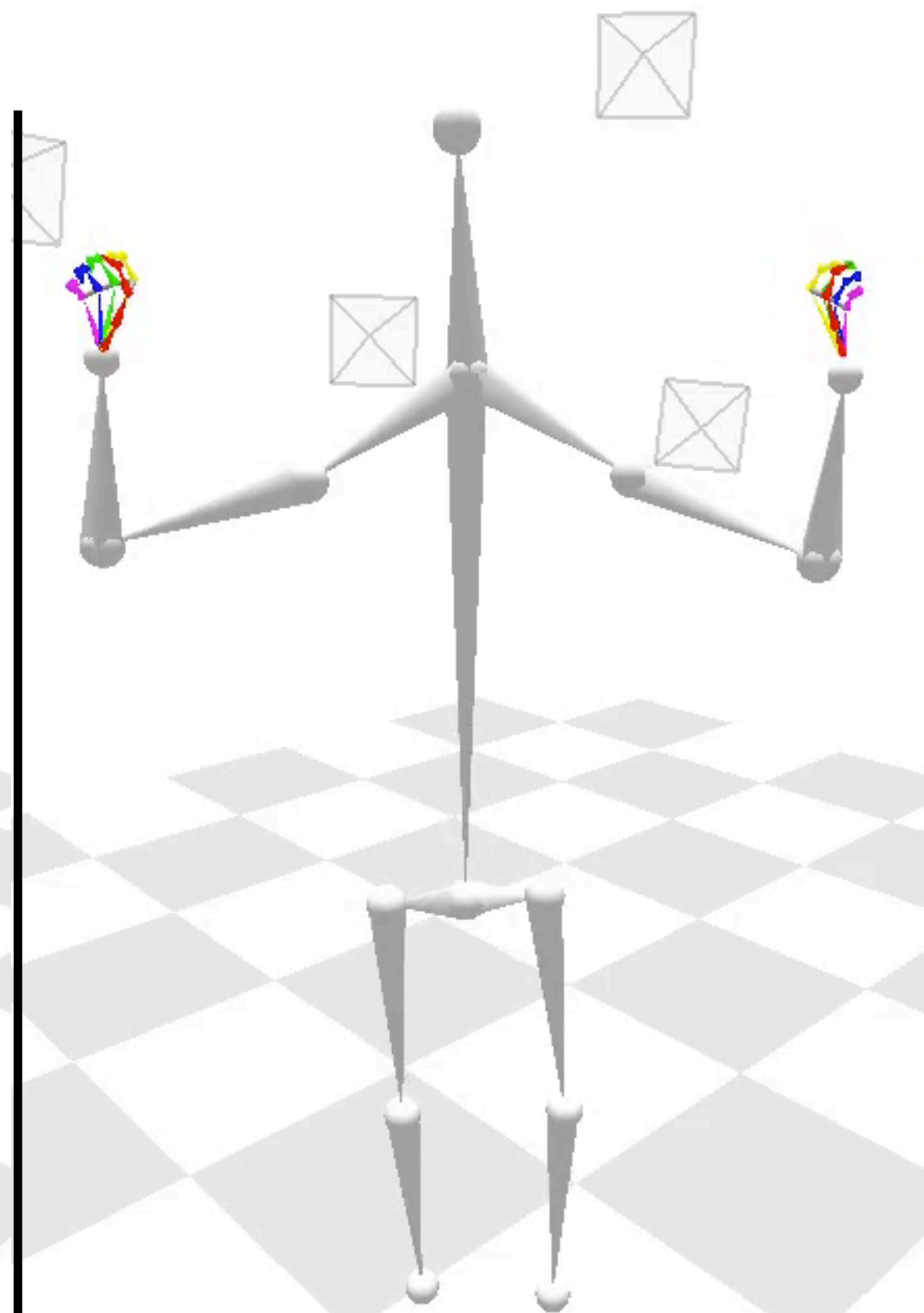
Comparison Between Iterations



Initial (Iteration 0)



Iteration 1



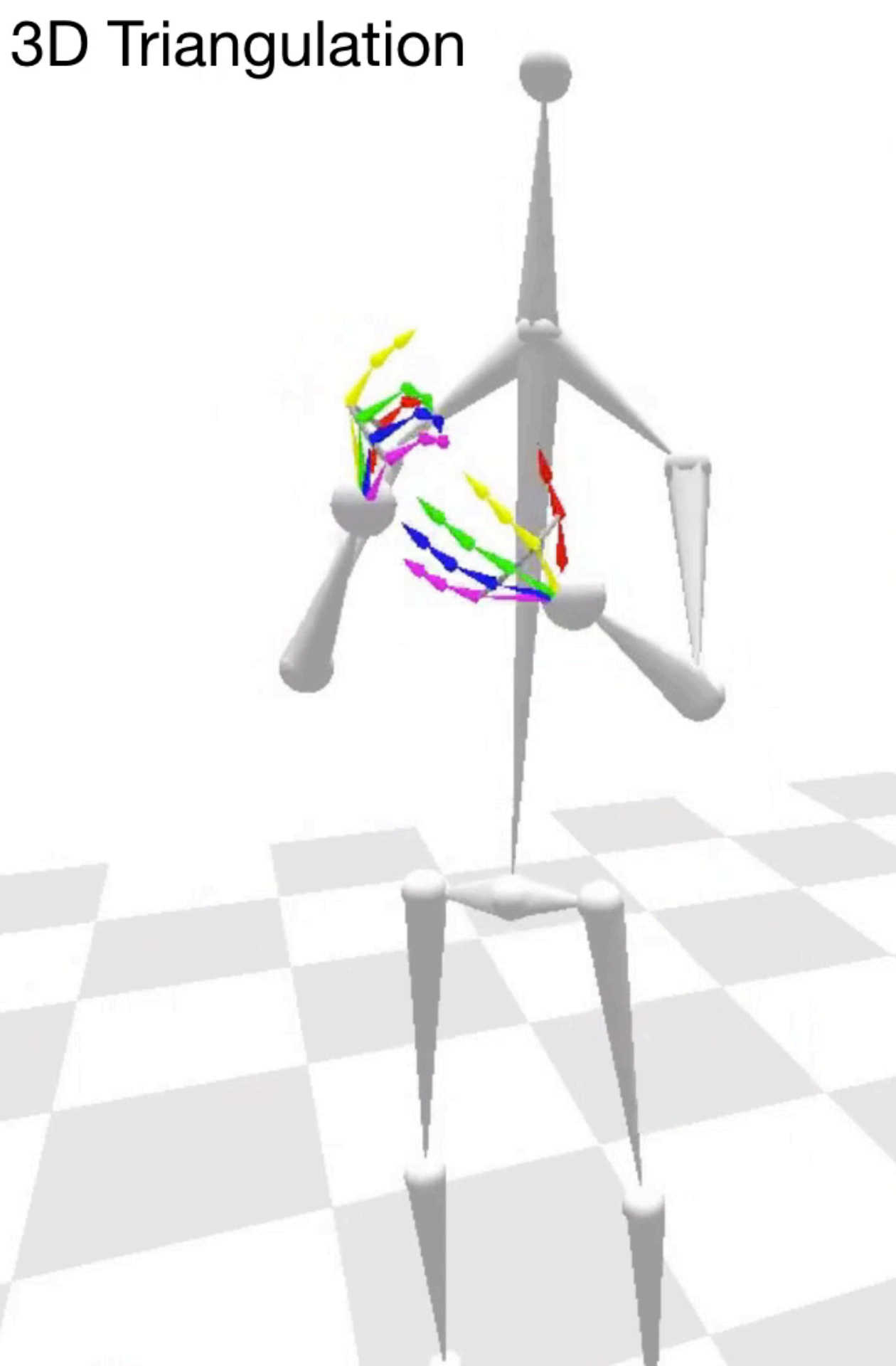
Iteration 2

Frame-by-Frame 2D Hand Pose Detection In-the-wild (No tracking)

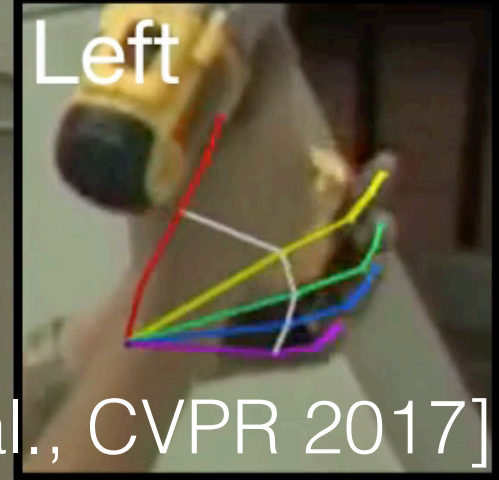
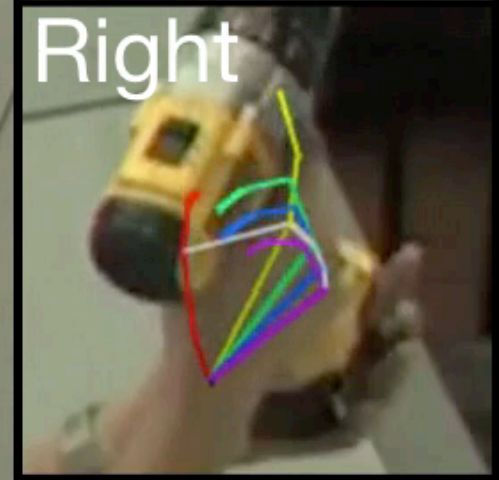
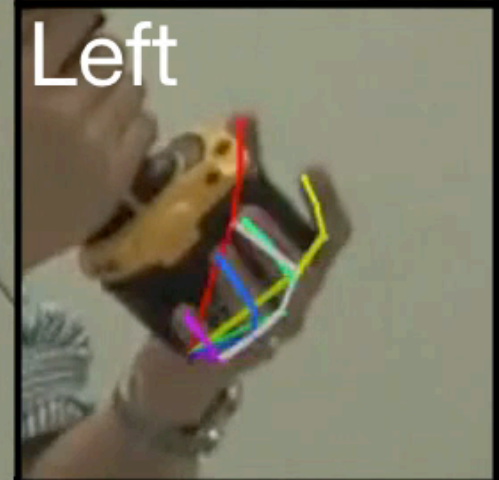
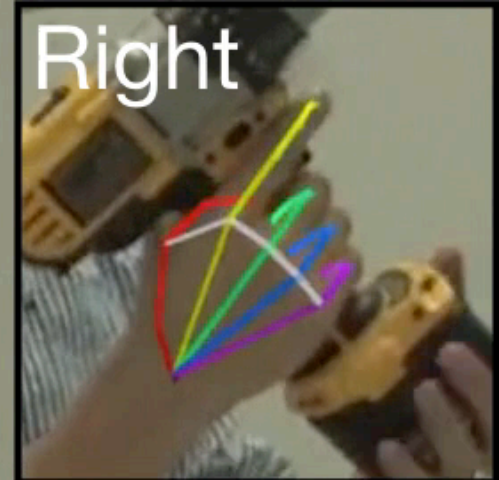


[Simon, **Joo**, et al., CVPR 2017]

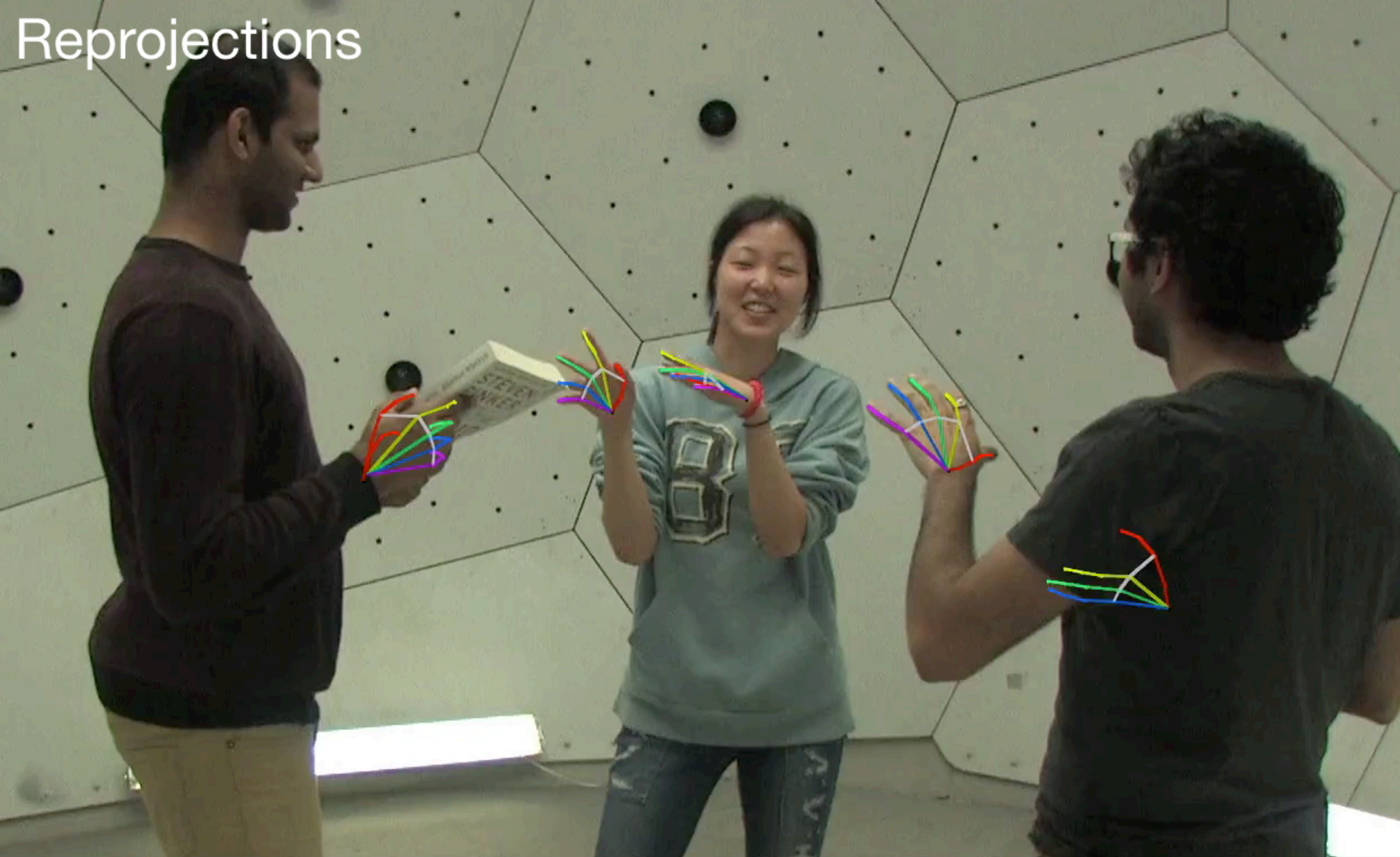
3D Triangulation



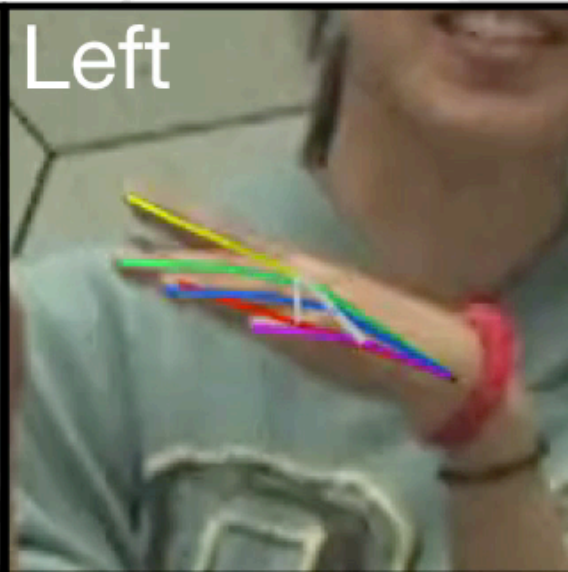
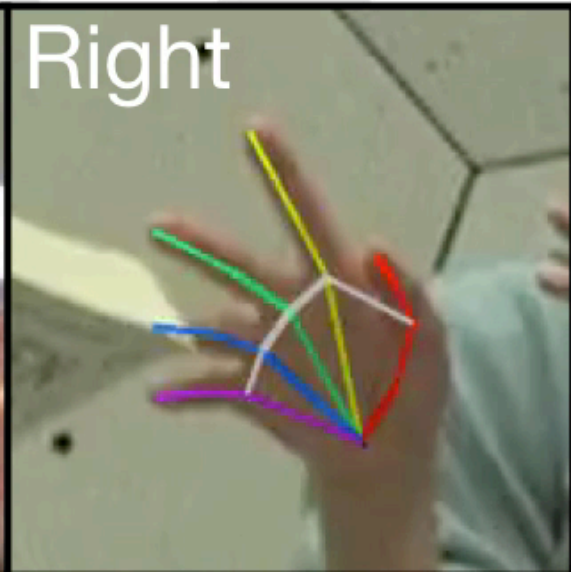
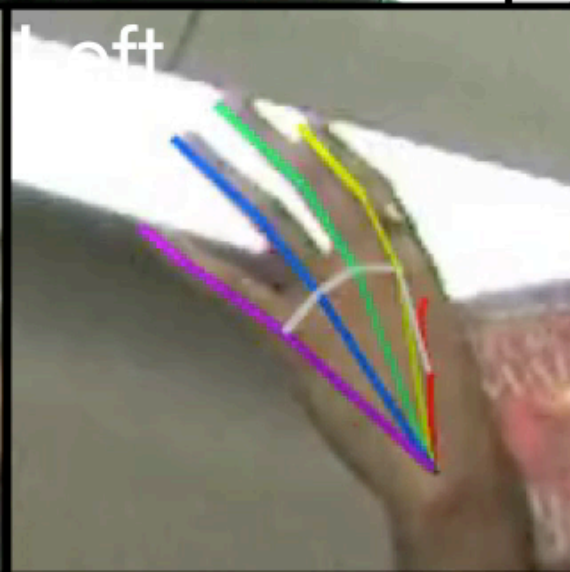
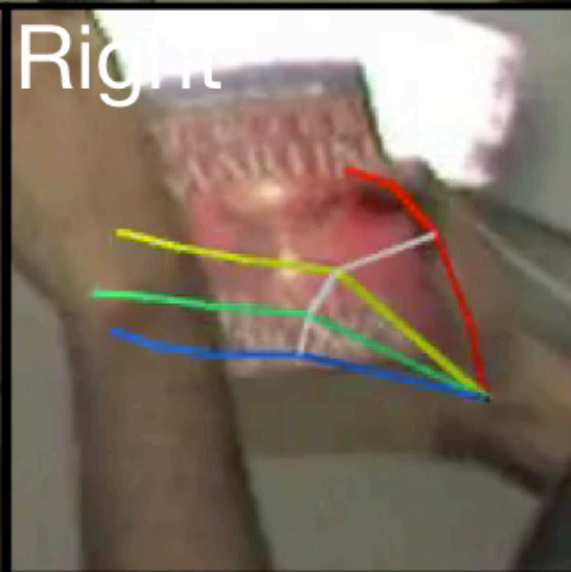
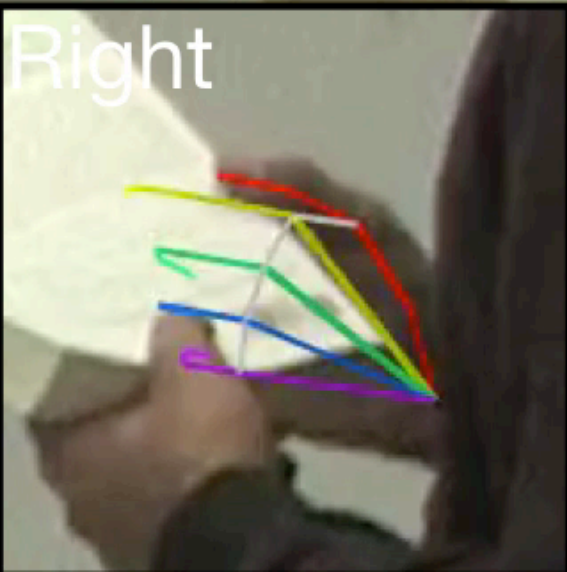
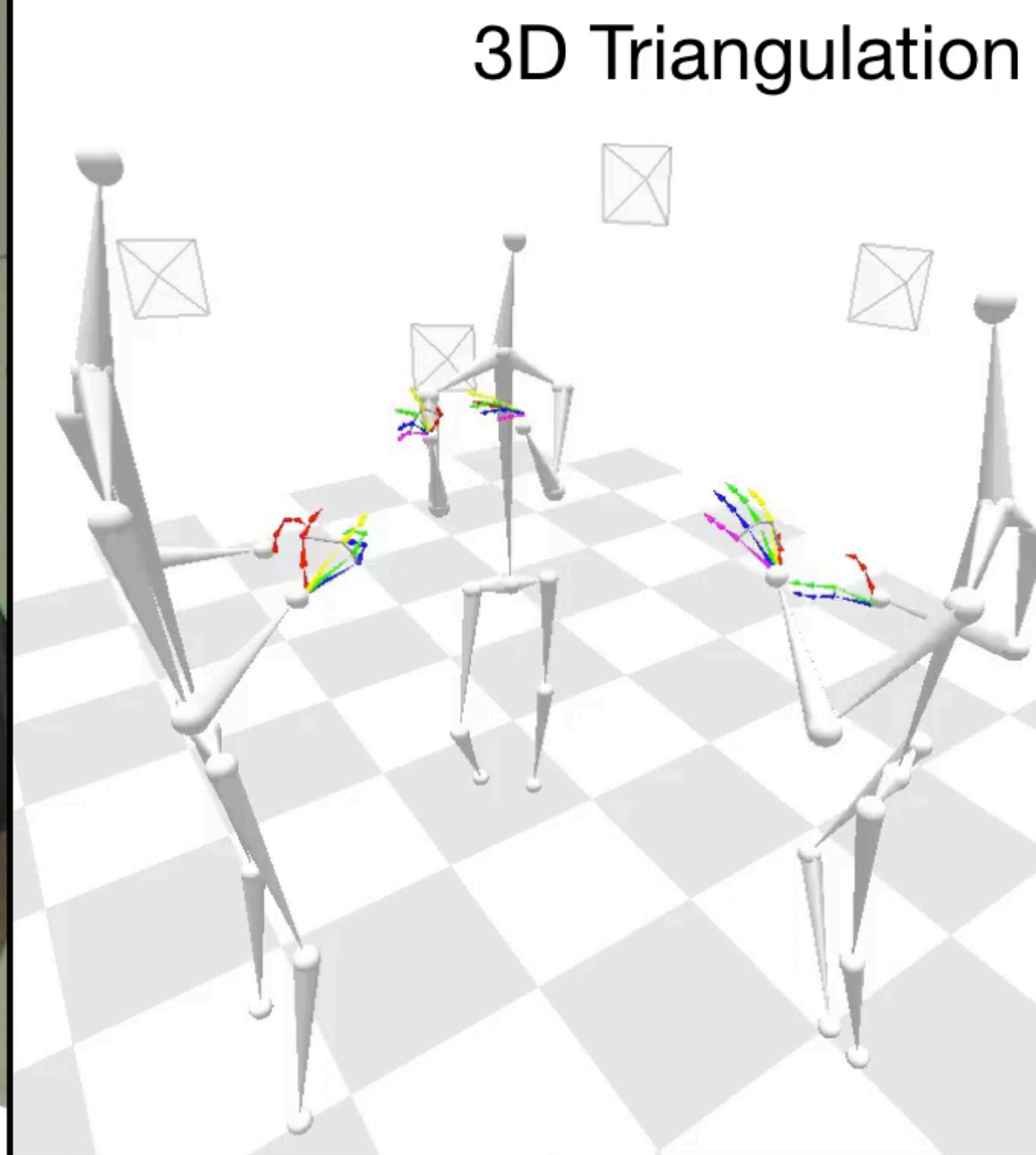
Reprojections

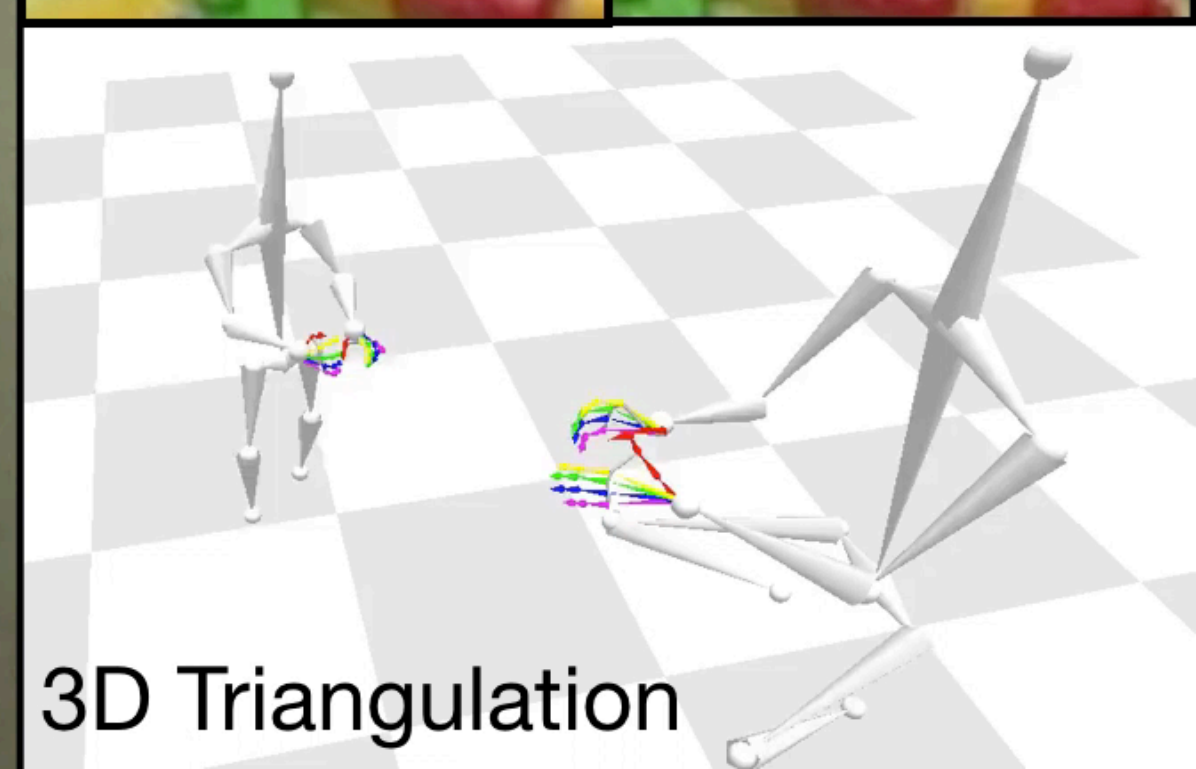
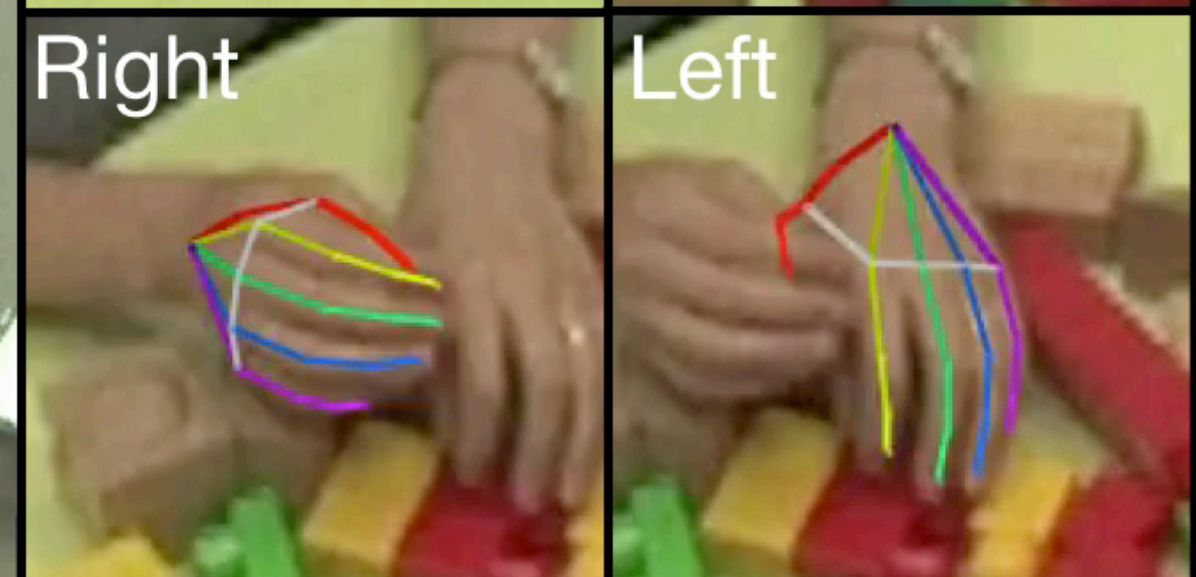
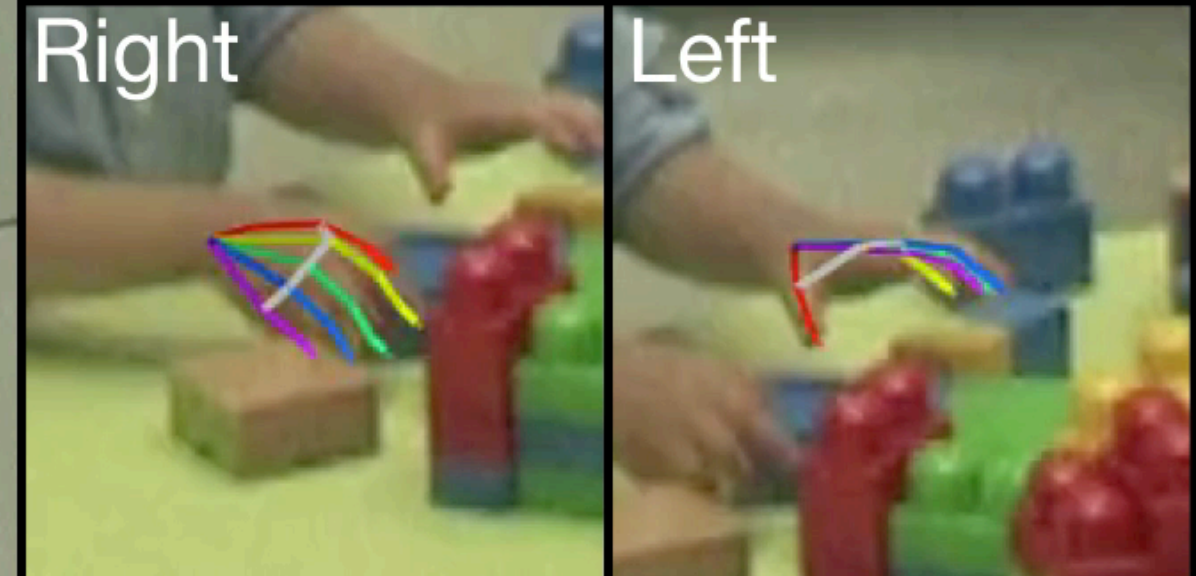


Reprojections



3D Triangulation





Reprojections

Multiview Bootstrapping

Applied For Face Keypoint Detector



Initial Detections (Iteration 0 --- Manual labels MultiPIE, Helen, AFW, ...)



Retrained Detections (Iteration 1)



OpenPose: Measuring Signals In the World

On GitHub: 6,632 ★ 1,654 Forks A Trending C++ repository (top 25) Included in OpenCV 3.4.1

Going Beyond Sparse Signals

Reconstructing Dense 3D Mesh Vertices Consistently Defined Across People



Adam Model with A Unified Parameterization

With Much Simpler Parameterization



Adam

$$= LBS \left(\underbrace{\text{Mean Shape}} + \underbrace{\sum_j \phi_j^U \text{Eigenvector 1}}_{\text{Total Body Shape Variations}} + \underbrace{\sum_k \phi_k^U \text{Eigenvector 1}}_{\text{Facial Expression}}, \underbrace{\theta^U}_{\text{Body+ Hand Pose}} \right)$$



An example view

Frame-by-frame Reconstruction

Measuring the Full Spectrum of Social Signals



An Example View



An Example View



The "Piano" Sequence

Frame-by-frame processing
Using 3D Keypoint Only



An example view

Assets > Scripts >

NOT_IN_U...	AnimData	BvhPaser
CameraCon...	CharacterA...	Controller
DataFrame...	InputContro...	Loader
SceneCont...	UDPRceiv...	UIController



Create Q All

- Main*
 - CameraCenter
 - Canvas
 - EventSystem
 - Adam_byTomas
 - Tara
 - ybot
 - Skeleton
 - Girl
 - TestHuman
 - boy_SKELMESH
 - GameManager
 - Test Environment
 - kitchen
 - Island02

Tag

- Position
- Rotation
- Scale
- Script
- Human M
- Size
- Eleme
- Eleme
- Eleme
- Eleme
- Eleme
- Scene M
- Size
- Eleme
- Eleme
- Eleme
- Cam Foc
- Script
- File Nam
- Frame Ti
- Speed M
- Script
- Port
- Frame D
- Script



The CMU Panoptic Studio: Introduction (short version)

Massively Multiview System

- > 480 VGA camera views
- > 30+ HD views
- > 10 RGB-D sensors
- > Hardware-based sync
- > Calibration

Interesting Scenes with Labels

- > Multiple people
- > Socially interacting groups
- > 3D body pose
- > 3D facial landmarks
- > Transcripts + speaker ID

* See the full length version of this video [here](#)

Dataset Size

Currently, 65 sequences (5.5 hours) and 1.5 millions of 3D skeletons are available.

What's New

- Dec. 2017 [Hand Keypoint Dataset Page](#) has been added. More data will be coming soon.
- Jun. 2017 We organize a tutorial in conjunction with CVPR 2017: "[DIY A Multiview Camera System: Panoptic Studio Teardown](#)"
- Jun. 2017 Hand keypoint detection and reconstruction paper will be presented in CVPR 2017: [Project Page](#).
- Dec. 2016 Panoptic Studio is featured on [The Verge](#). You can also see the video version [here](#).
- Dec. 2016 The social interaction capture paper (extended version of ICCV15) is available on [arXiv](#).
The CMU PanopticStudio Dataset is now publicly released.
- Sep. 2016 Currently, **480 VGA videos**, **31 HD videos**, **3D body pose**, and **calibration data** are available. **Dense point cloud** (from 10 Kinects) and **3D face reconstruction** will be available soon. Please contact [Hanbyul Joo](#) and [Tomas Simon](#) for any issue of our dataset.
- Sep. 2016 The [PanopticStudio Toolbox](#) is available on GitHub.
- Aug. 2016 Our dataset website is open. Dataset and tools will be available soon.

Dataset Examples

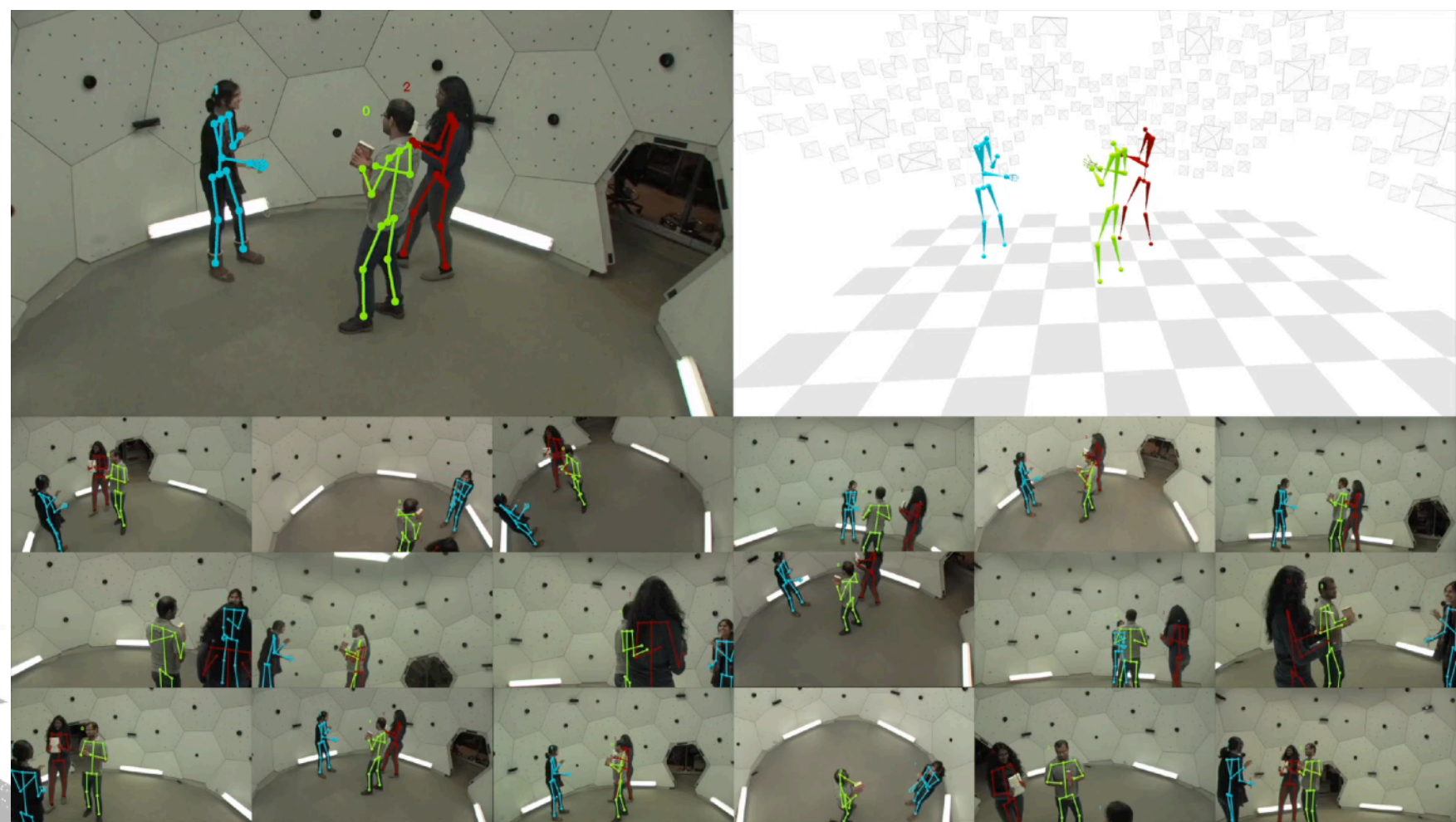
Panoptic Studio Dataset

<http://domedb.perception.cs.cmu.edu/>

- 30HDs + 10 RGBDs + Calibration + Point Clouds + 3D Keypoints (bodies + faces + hands)
- **6 hours** of diverse scenes (social games, range of motion, musical instruments, etc.)

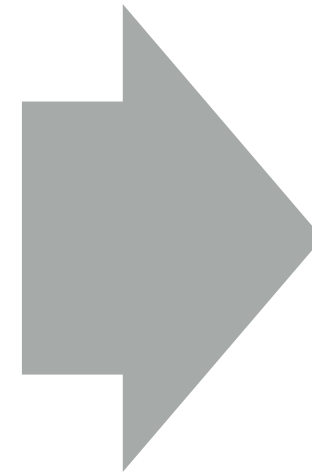
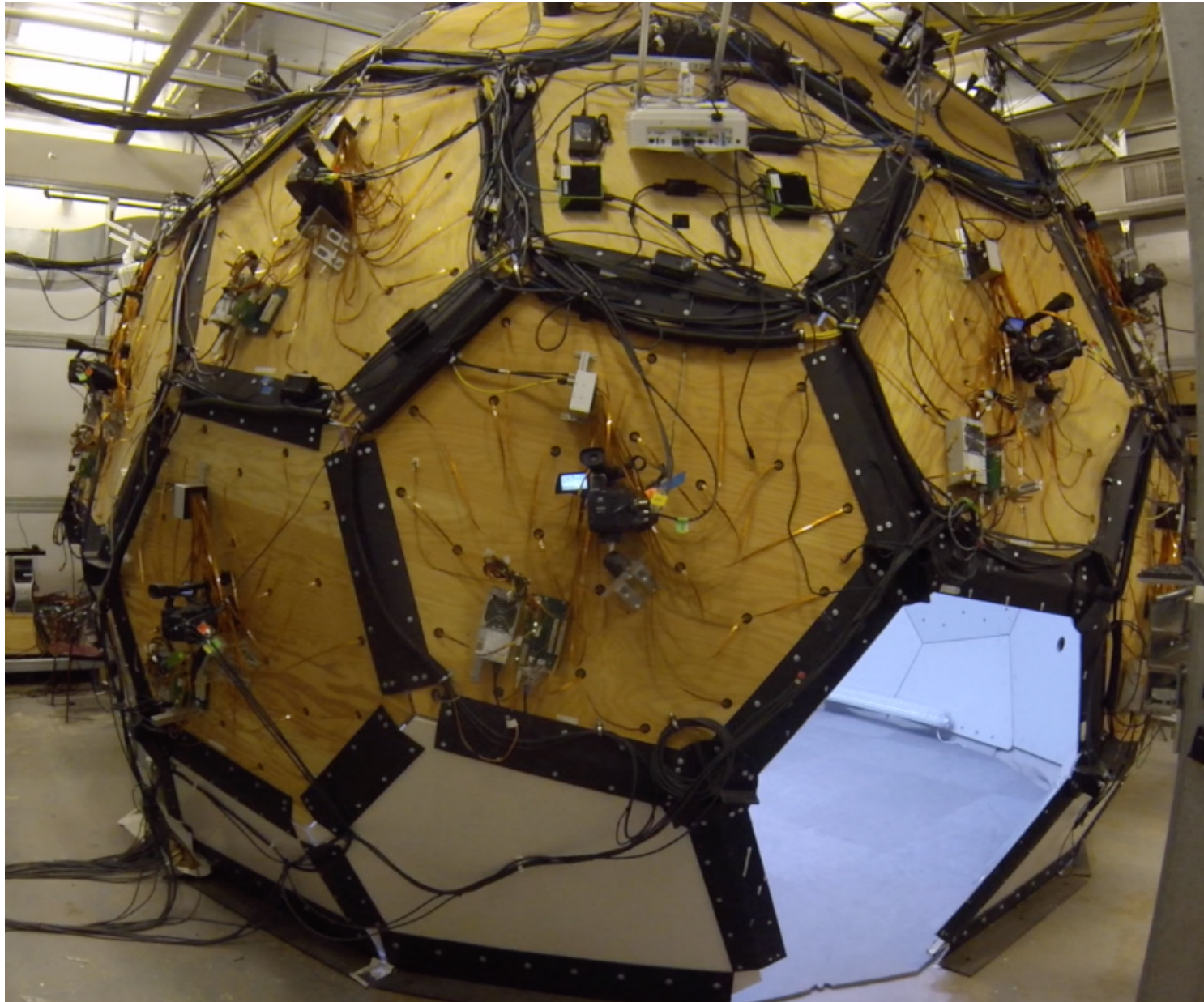


3D Point Clouds



3D Keypoints (Bodies+ Faces + Hands)

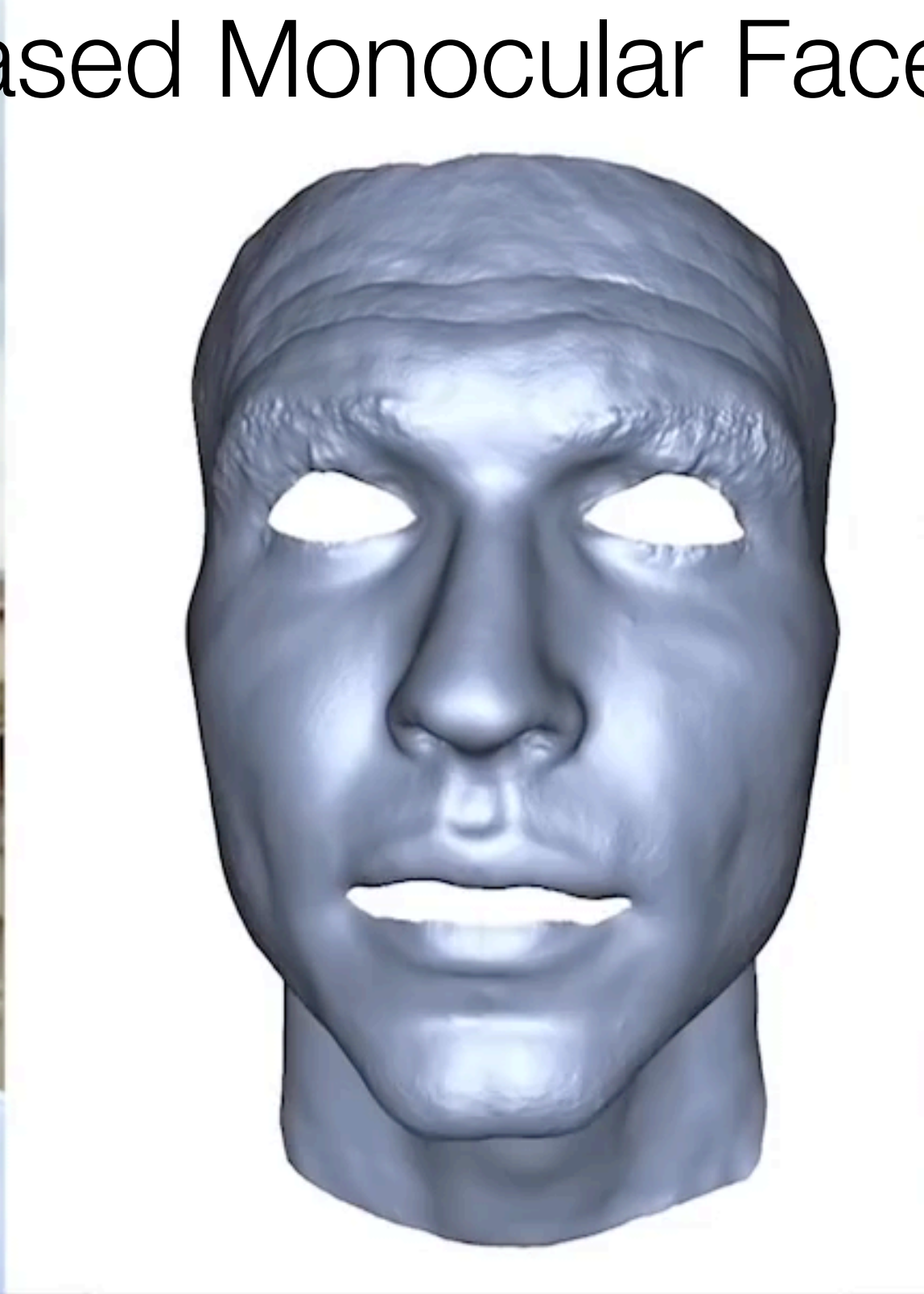
Do We Really Need 500 Cameras?



Model based Monocular Face Capture



iPhone Video



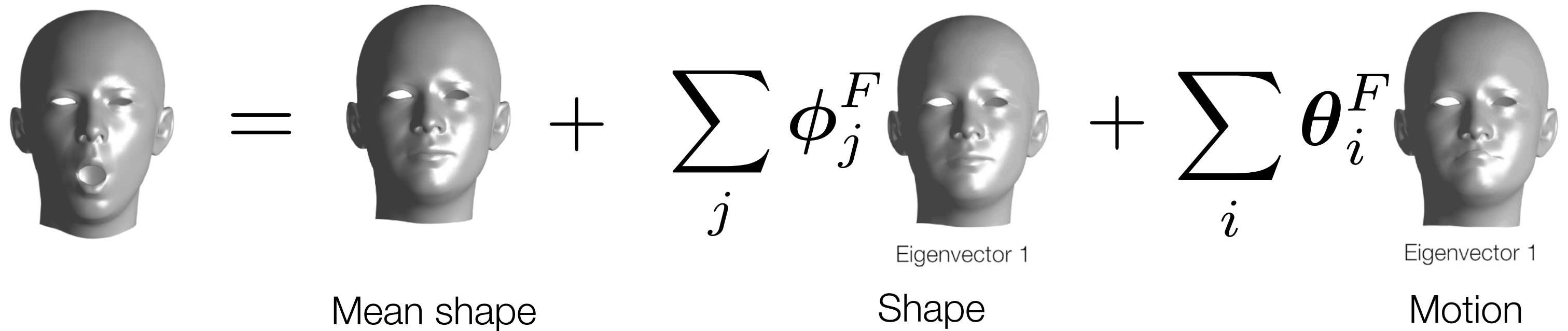
Our Result



Overlay

The Face Part Model

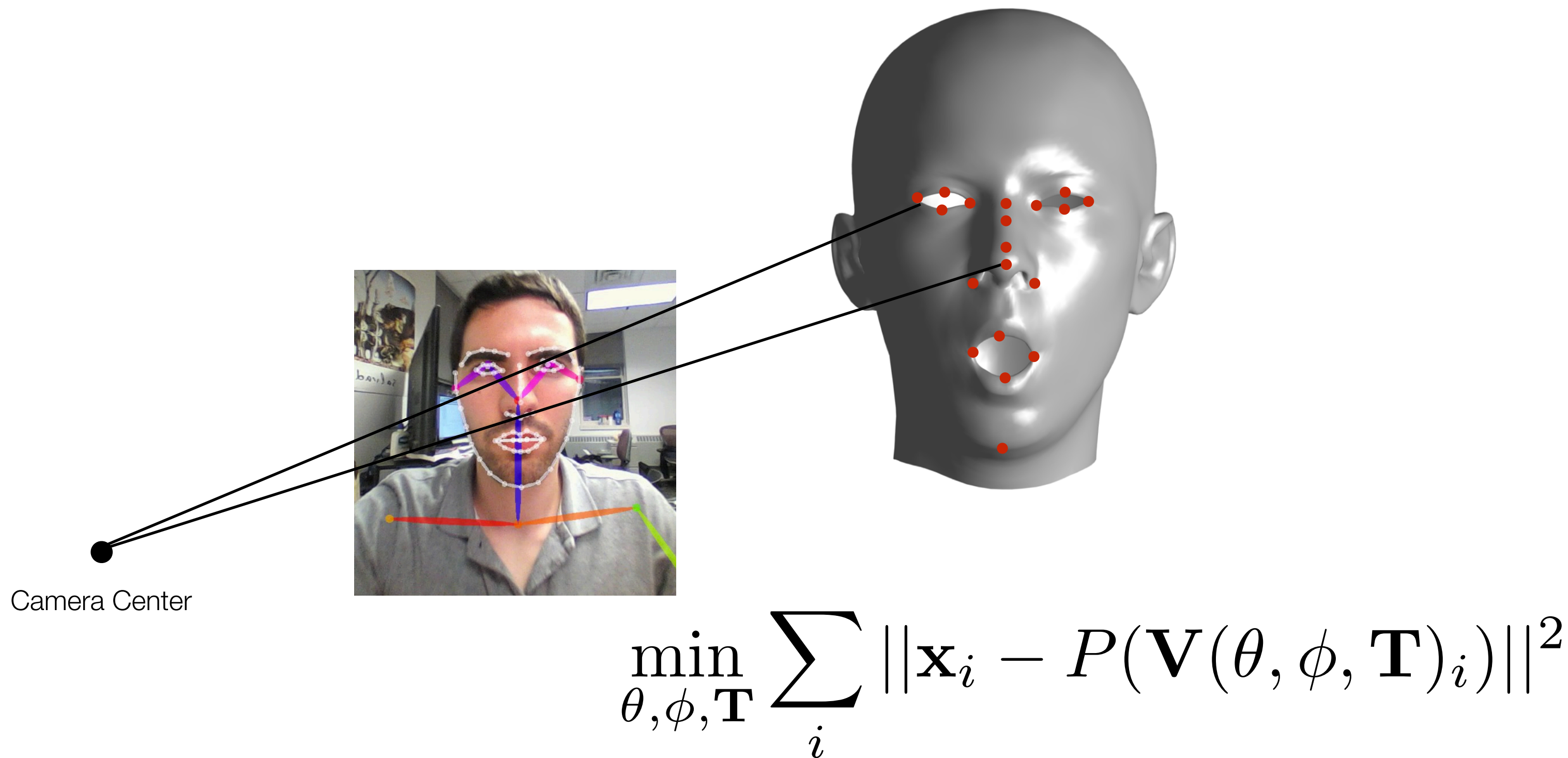
FaceWarehouse [Cao et al., TVCG 2014]



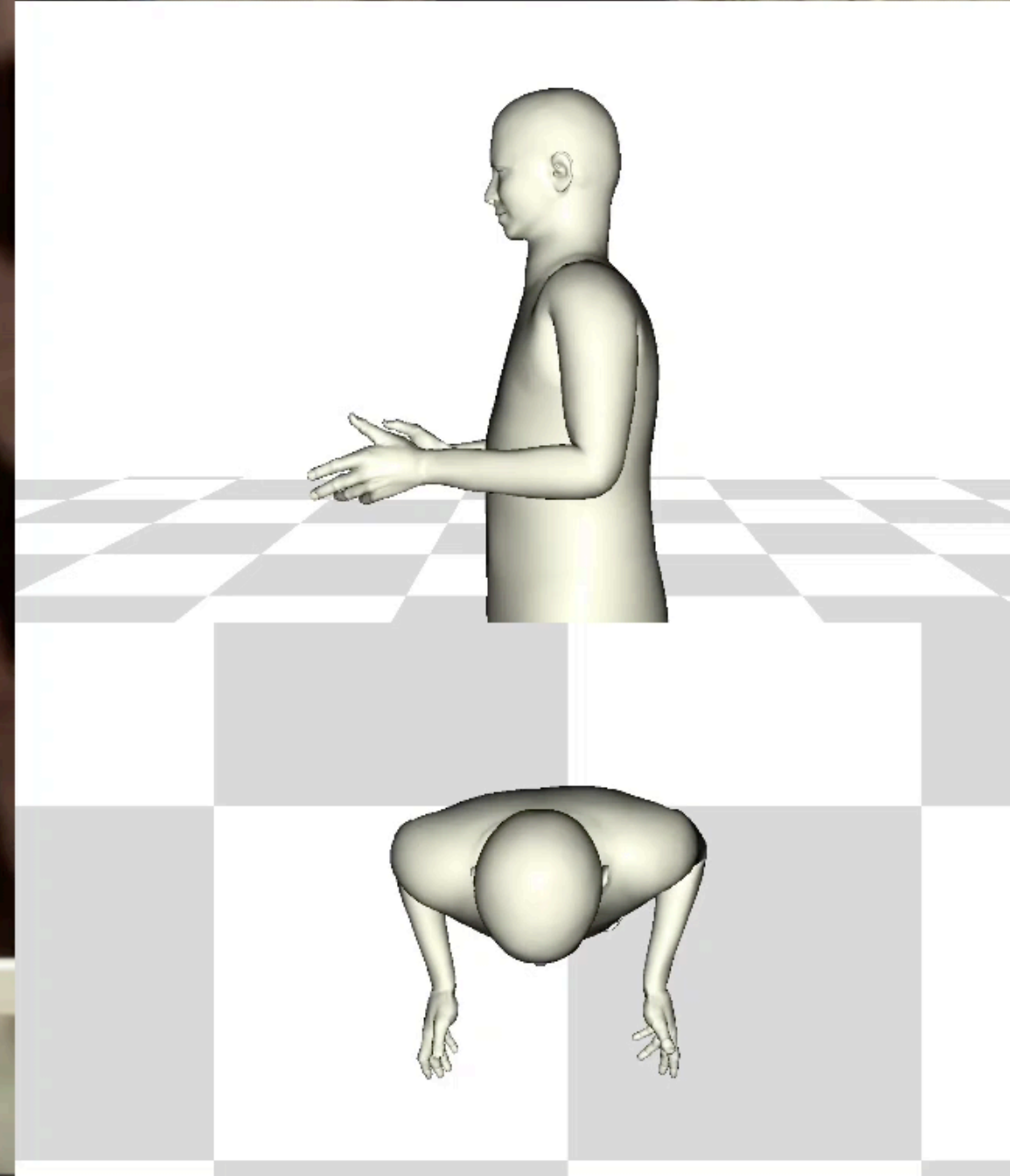
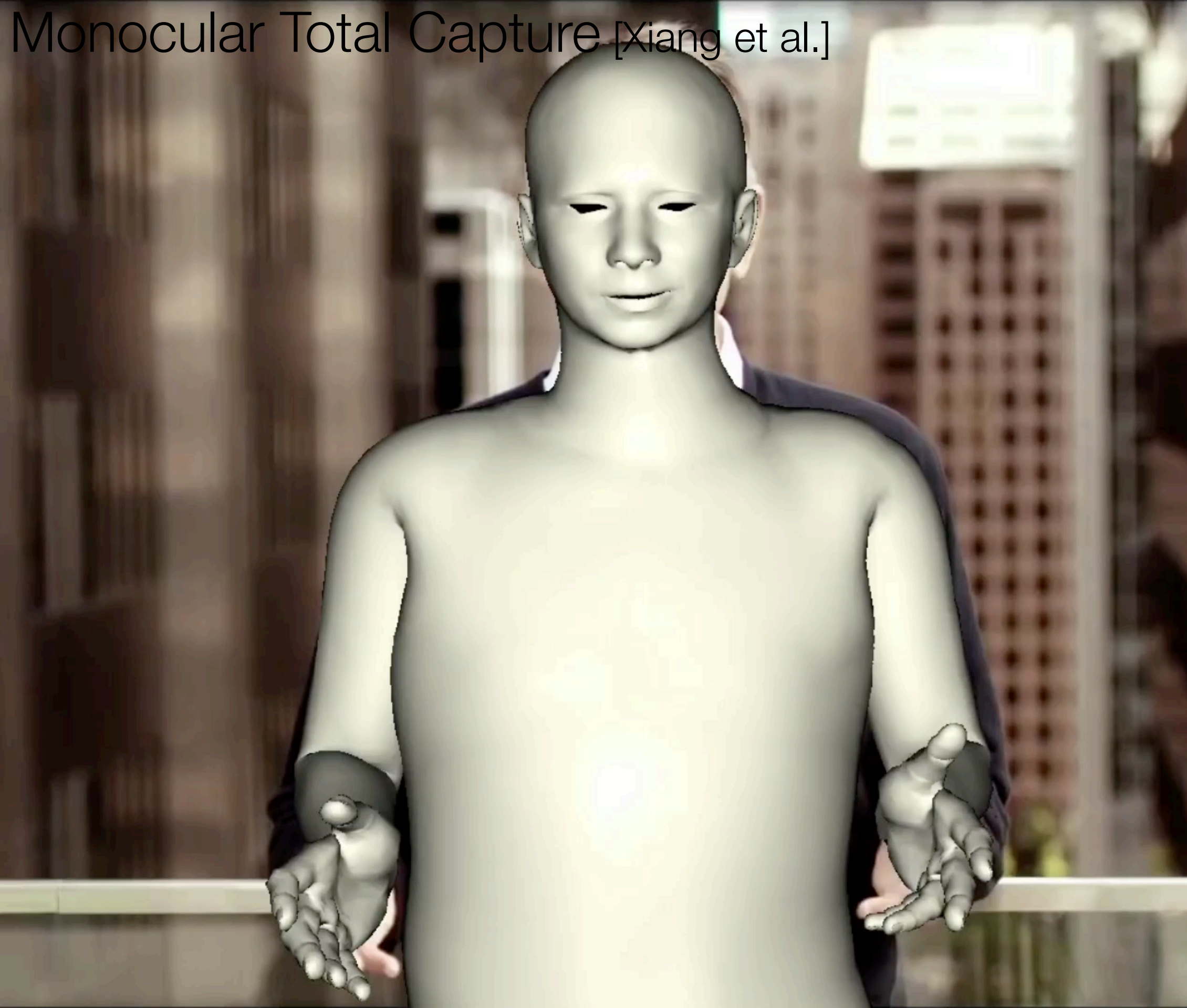
$$\text{3D vertices } \mathbf{V}^F = M^F(\boldsymbol{\phi}^F, \boldsymbol{\theta}^F, \mathbf{T}^F) \text{ Rigid transformation}$$

Model based Monocular Face Capture

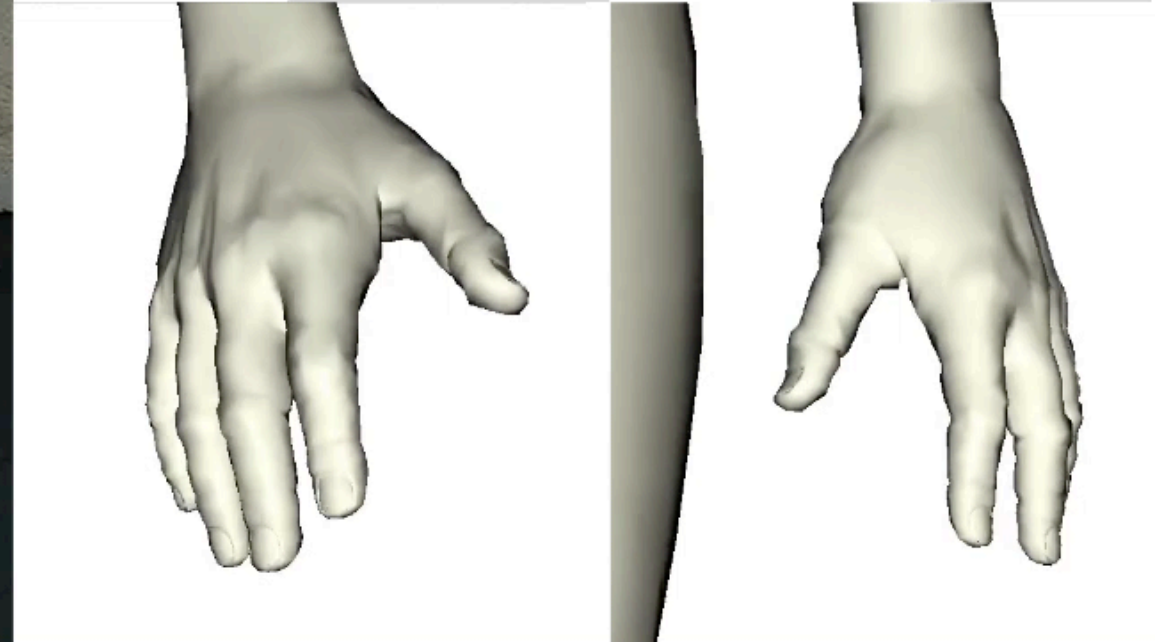
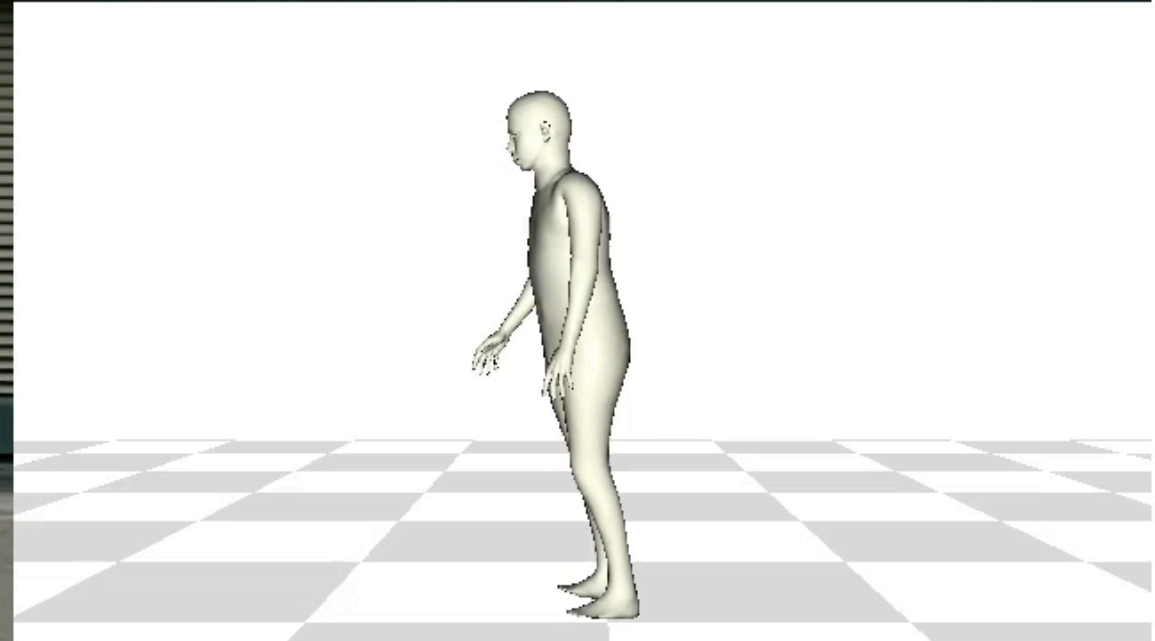
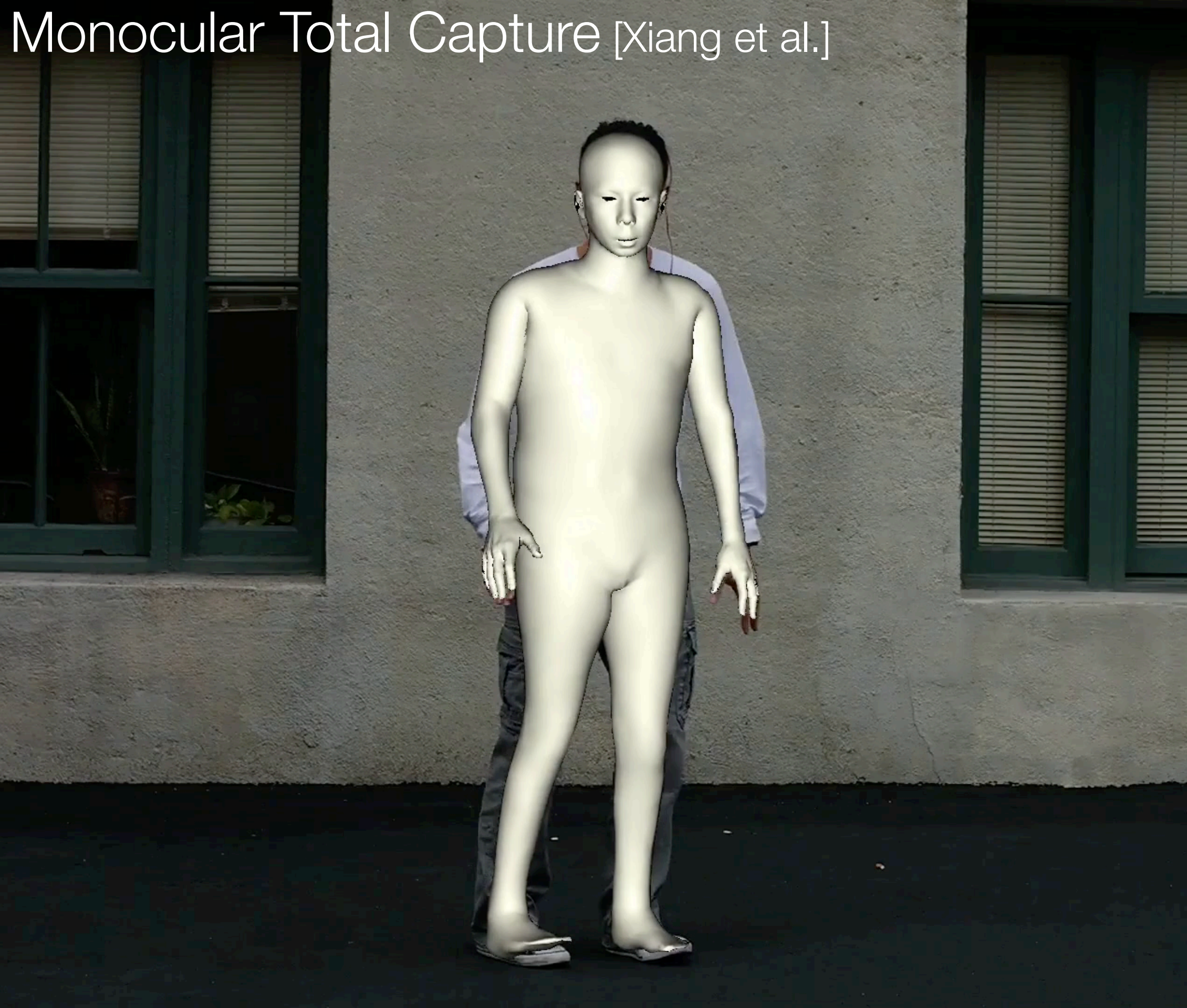
The Basic Idea



Monocular Total Capture [Xiang et al.]



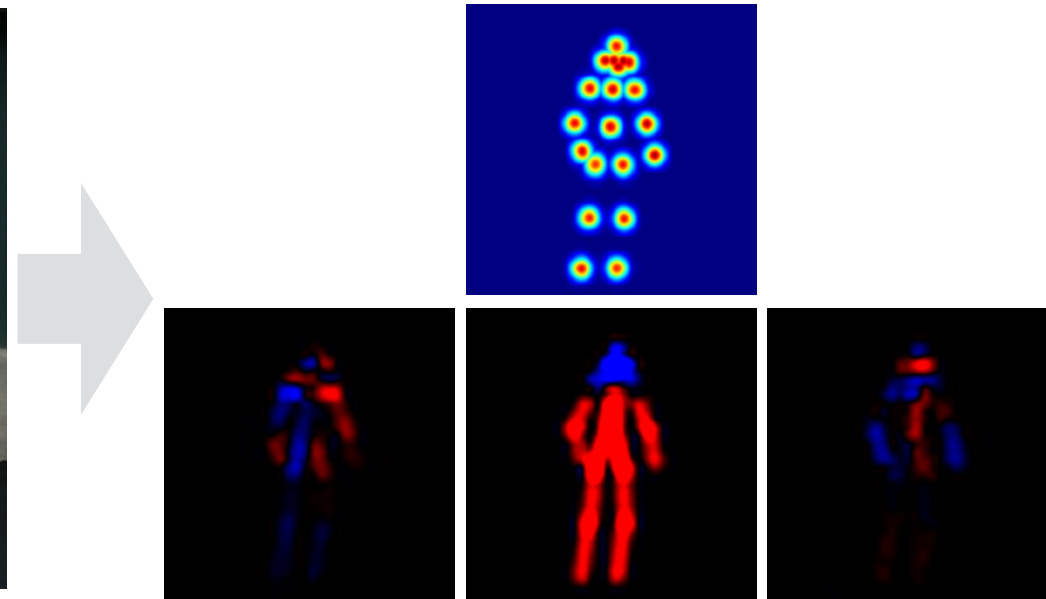
Monocular Total Capture [Xiang et al.]



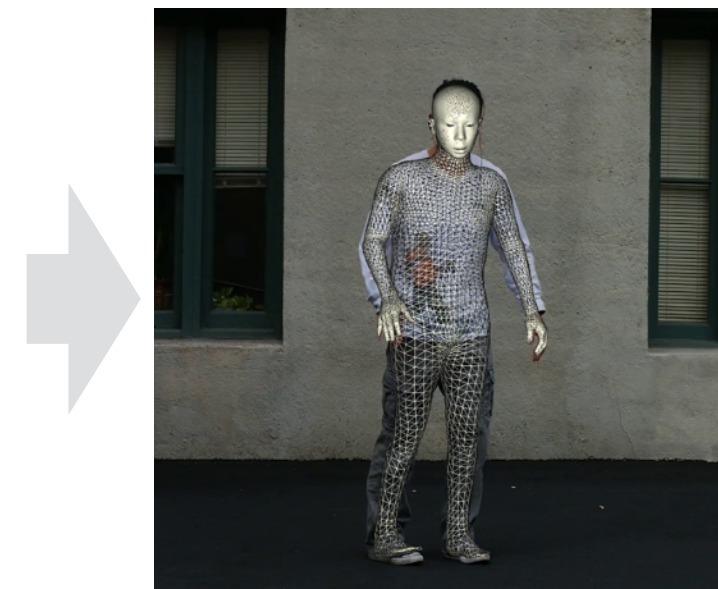
Monocular Total Capture: Posing Face, Body, and Hands



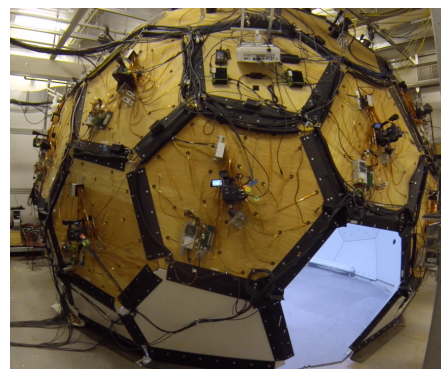
Input image



Predict 2D keypoint and
3D Part Orientation Field (POF)



Deformable Model
Fitting



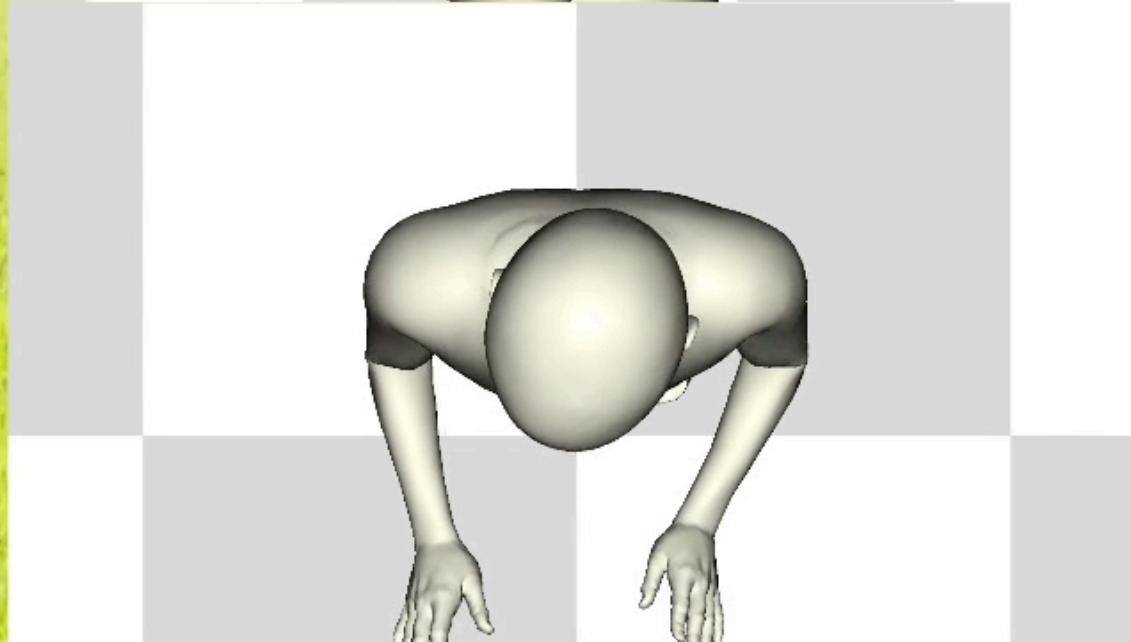
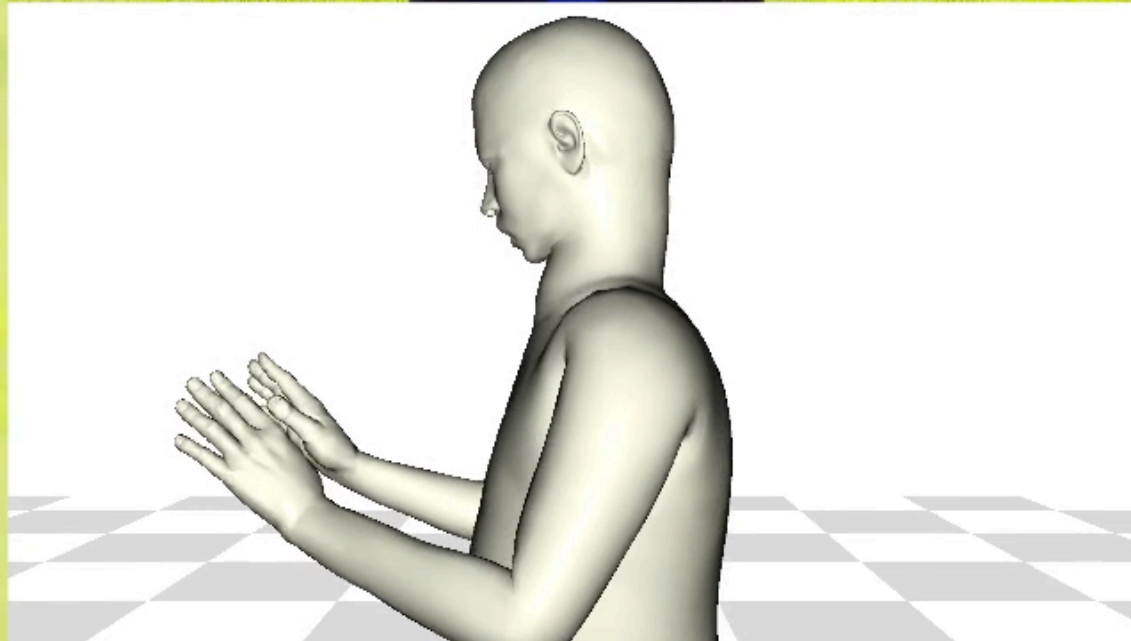
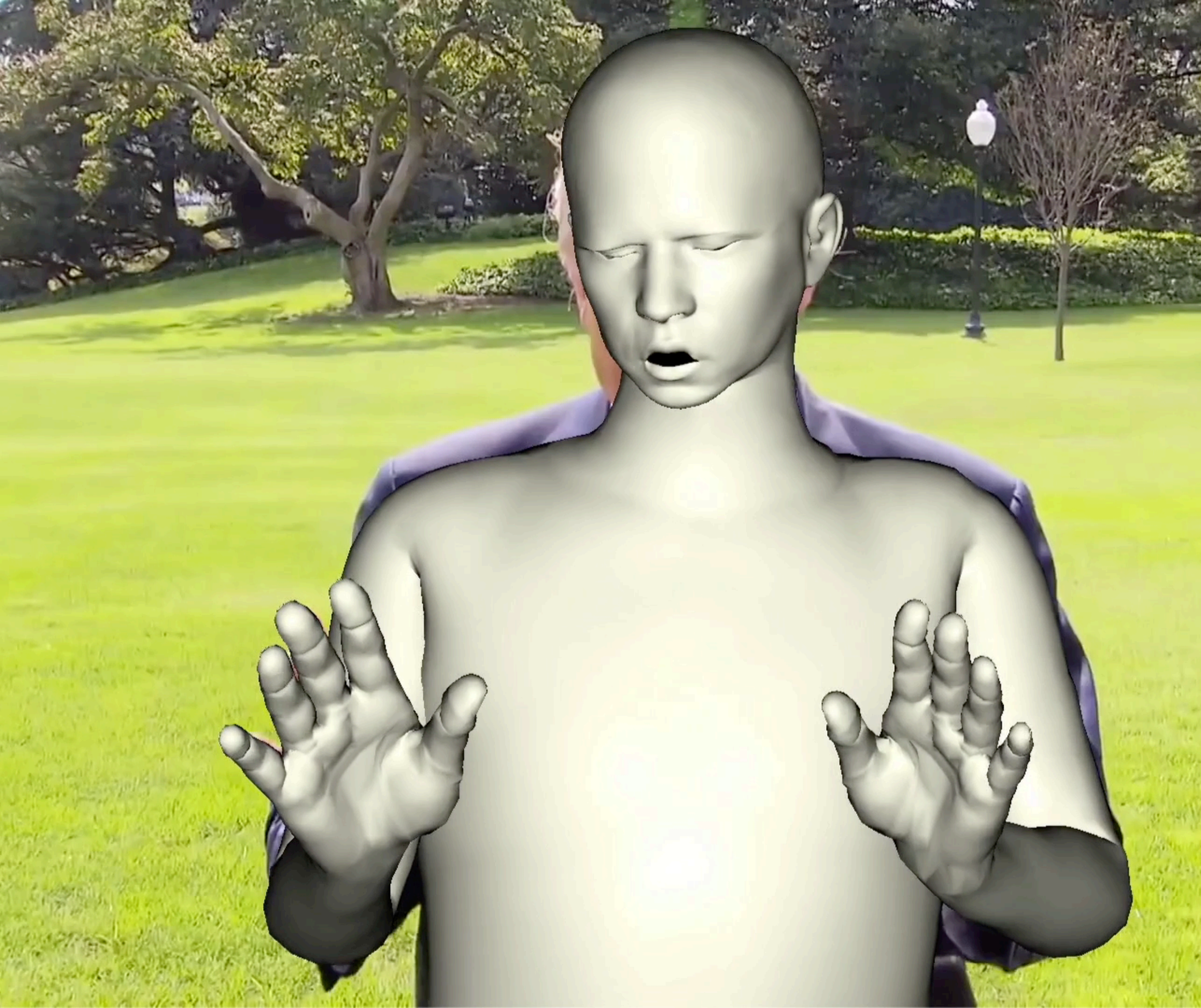
2D/3D Dataset
3D Human Model

Machine Learning

Parametric Space



work by Donglai Xiang



Question?

Hanbyul (Han) Joo (hanbyulj@cs.cmu.edu)