Panoptic Studio

A Massively Multiview System for **Social Motion Capture**

Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews
Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh

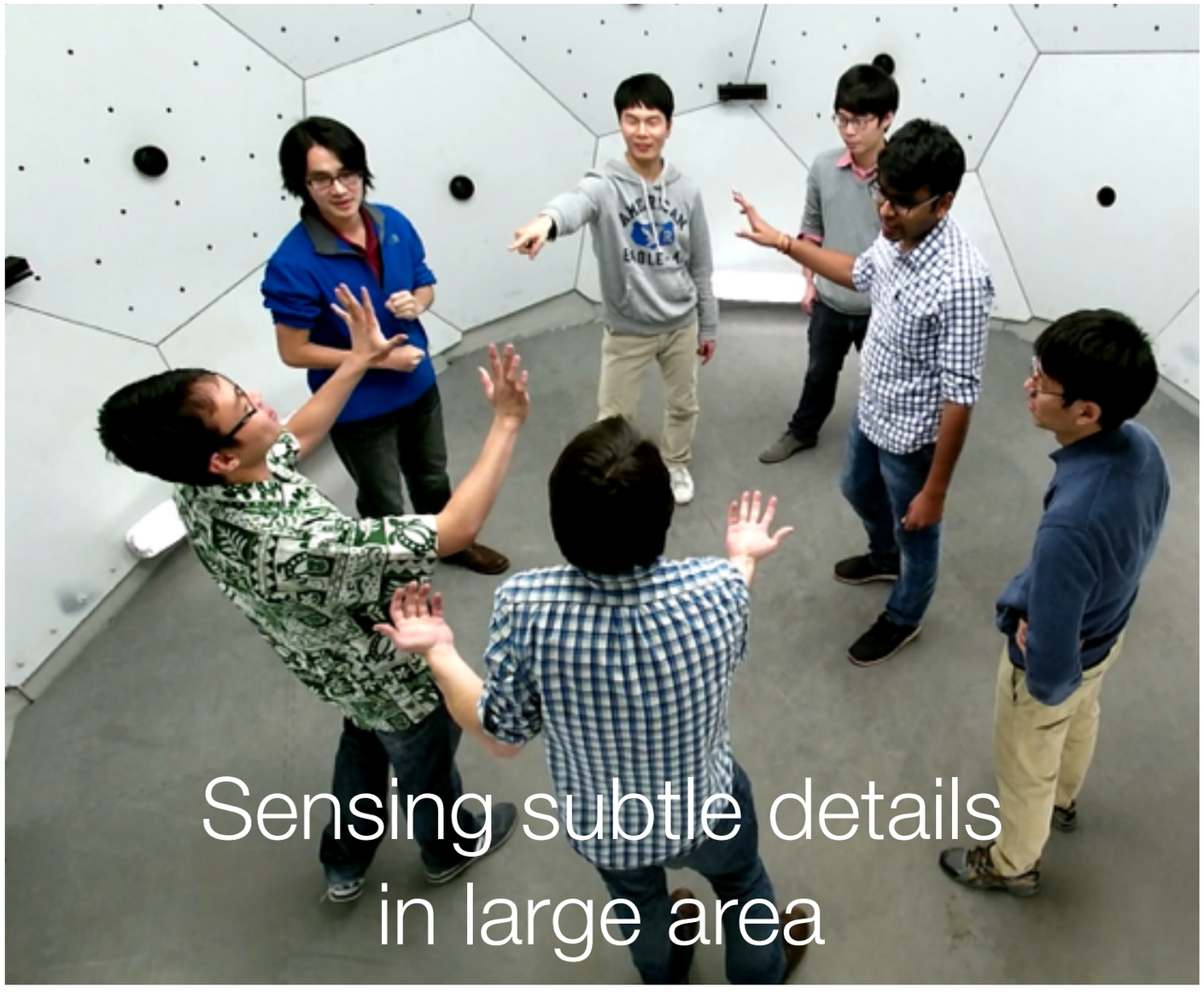
Robotics Institute
Carnegie Mellon University

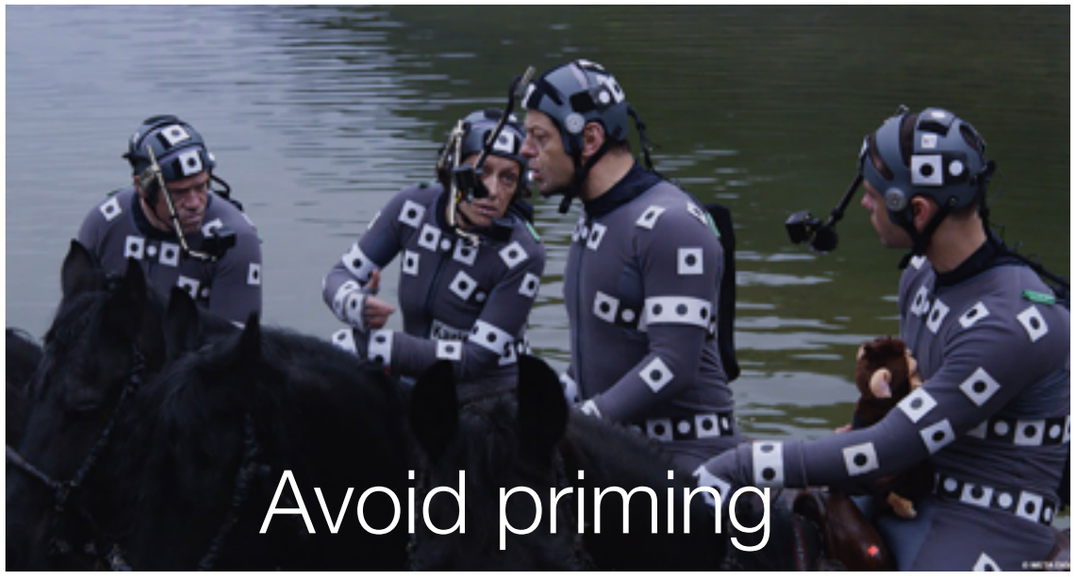
Social Motion Capture

Measuring Nonverbal Signals of Socially Interacting People

Technical Challenges

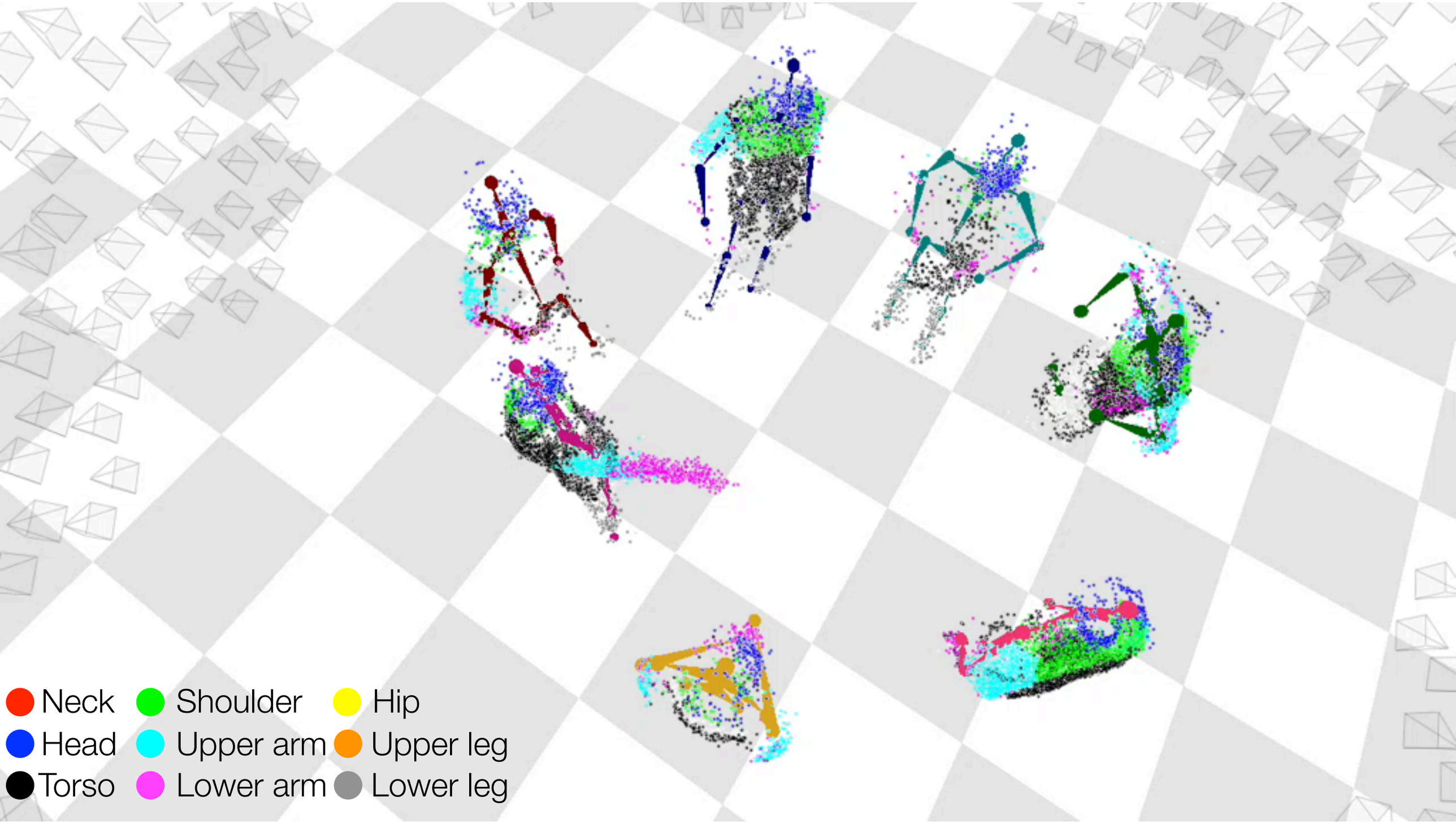
Sensing Challenges and Priming Issue





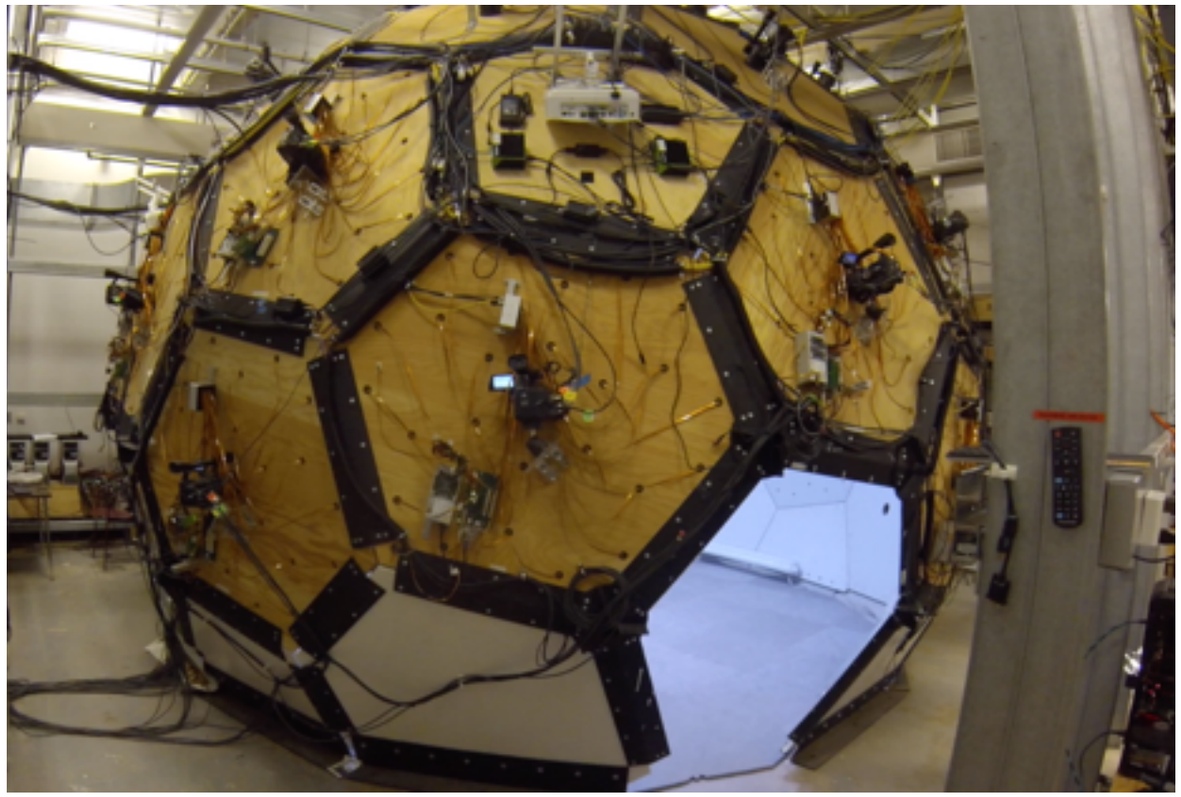
Output

Labelled Point Trajectories with Skeletal Structures



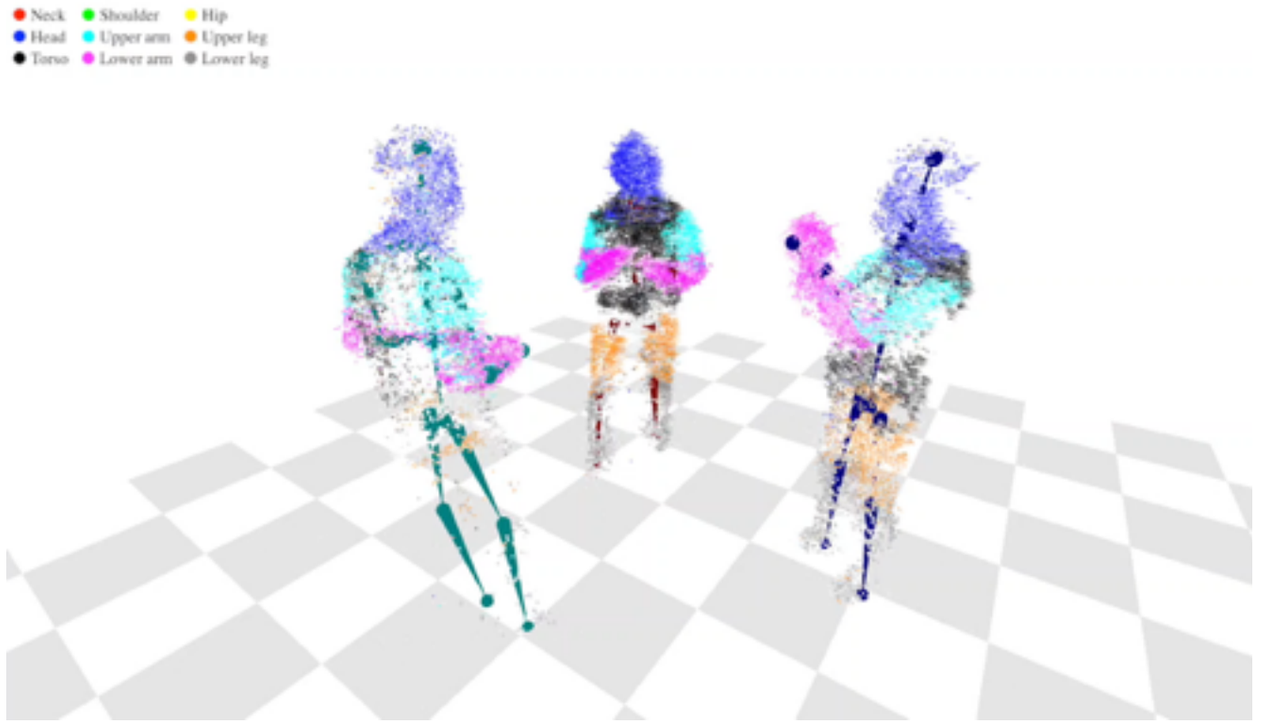
Technical Contribution

Novel Hardware and Software Systems for Social Motion Capture



The Panoptic Studio

- 480 Cameras
- Modularized design
- Synchronization

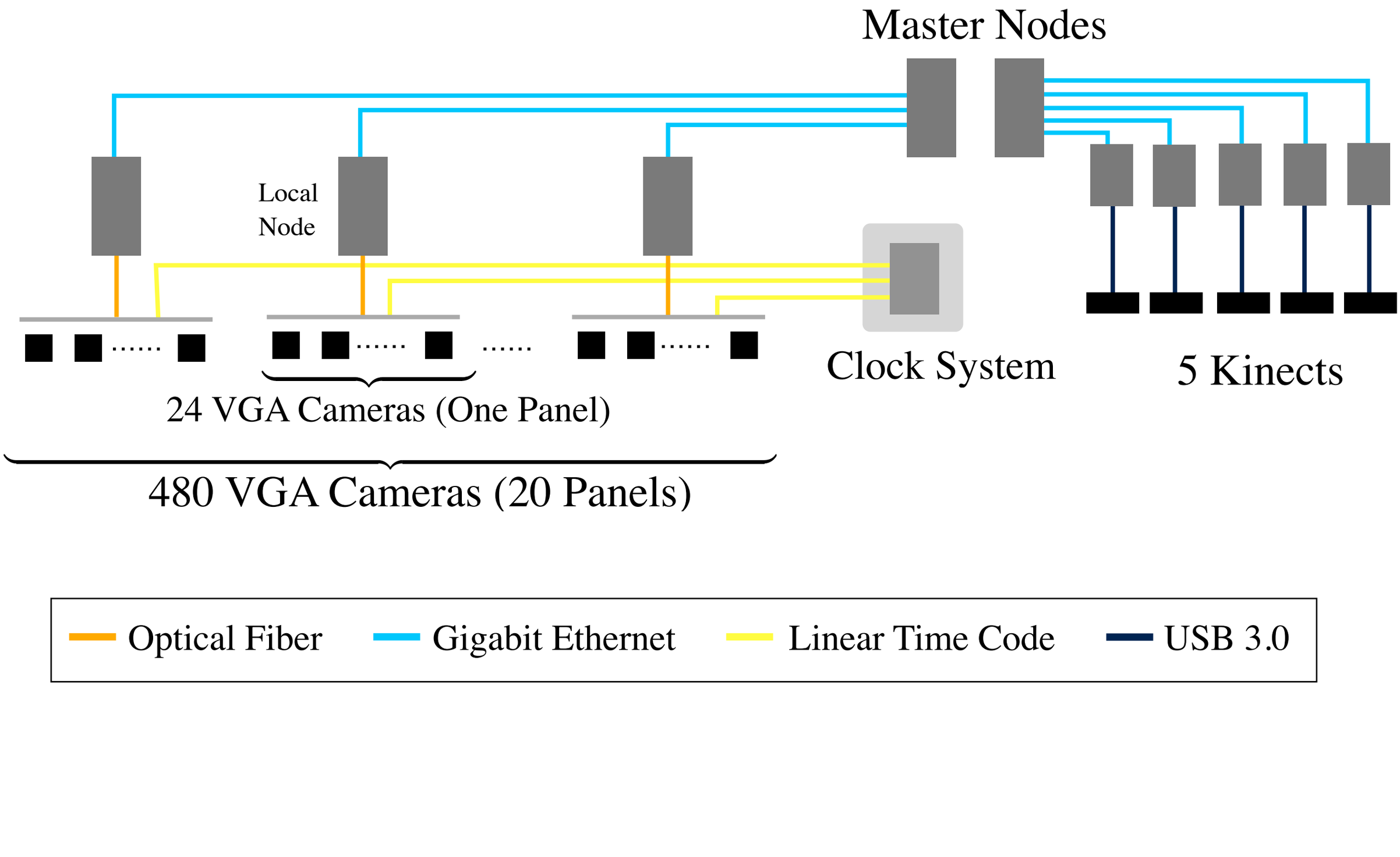


Social motion capture

- Boost large number of simple processes
- No subject-specific template is needed

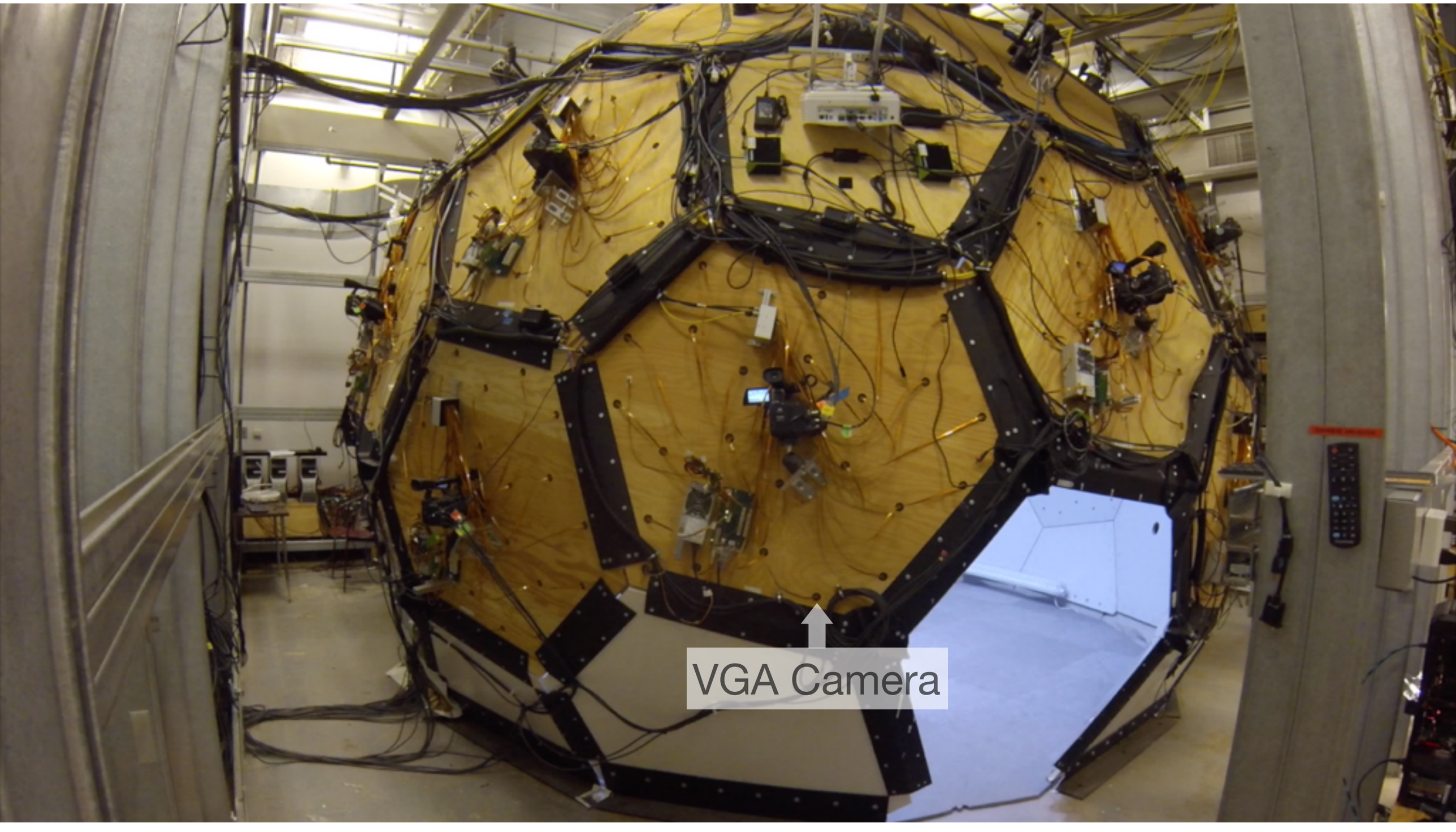
The Panoptic Studio

Hardware Architecture



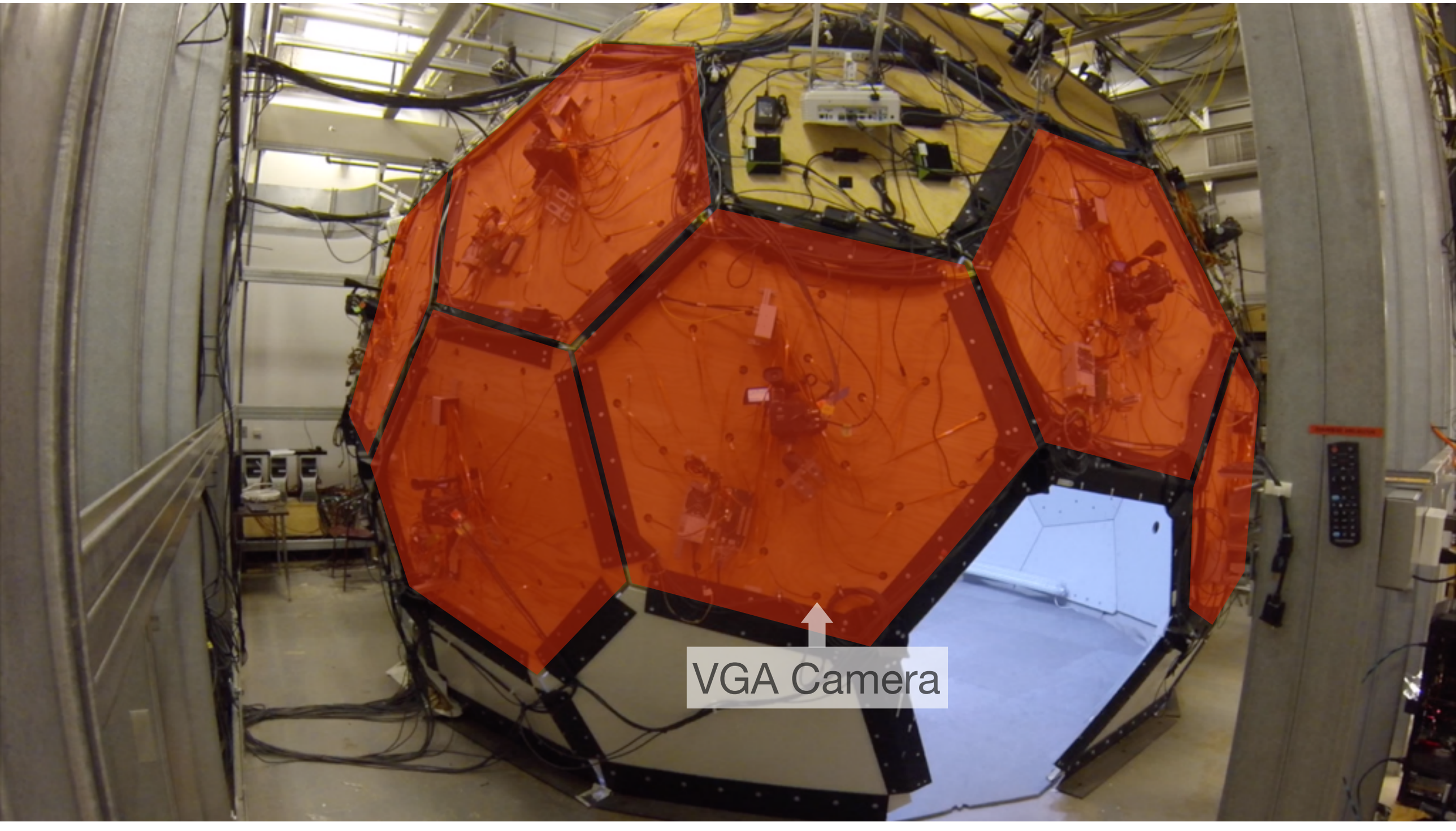
The Panoptic Studio

A Massively Multiview System with 480 Cameras



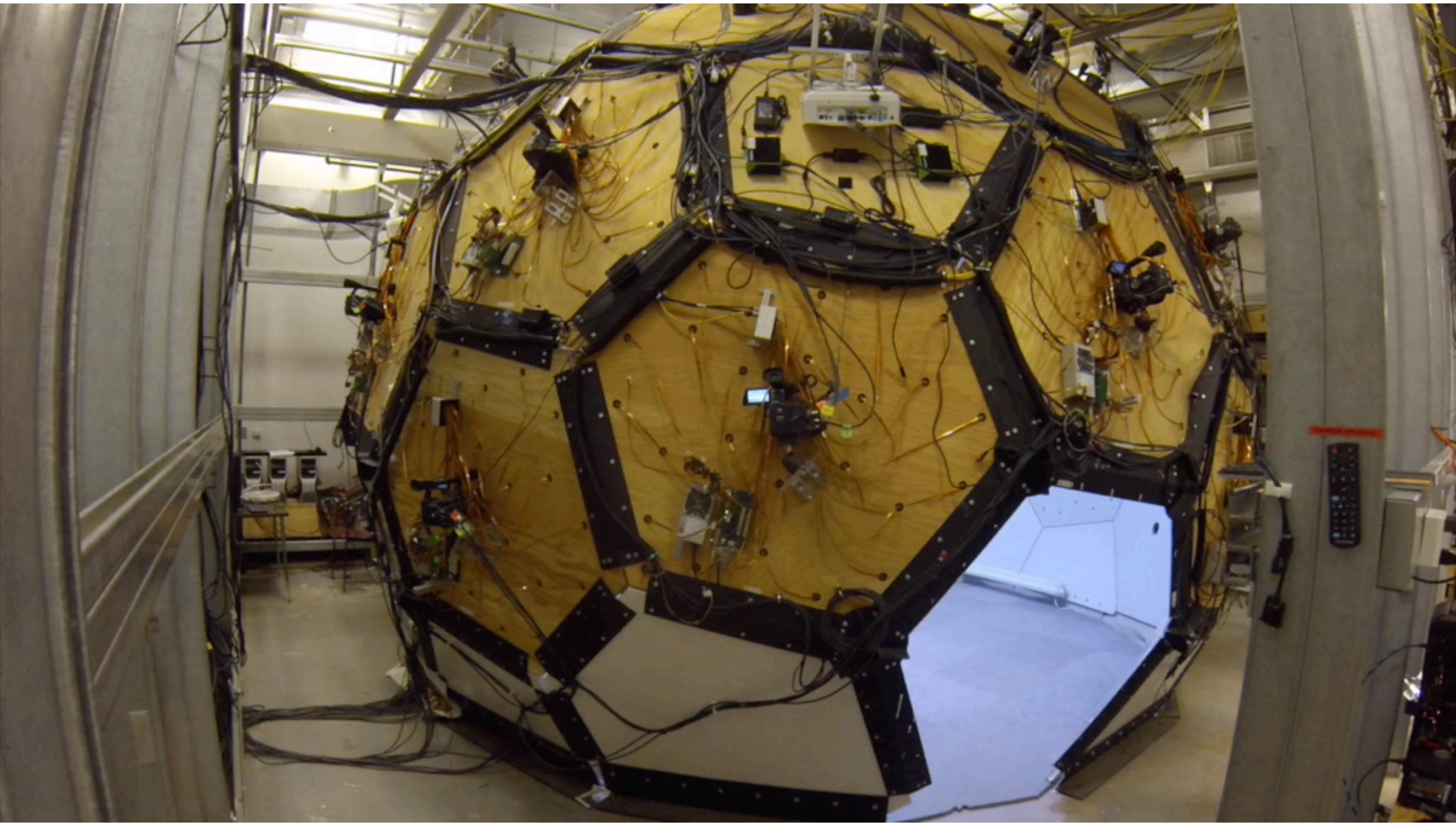
The Panoptic Studio

Modularized Design with 20 Panels



The Panoptic Studio

Modularized Design with 20 Panels



An Example Social Scene

An Example Video

An Example Social Scene

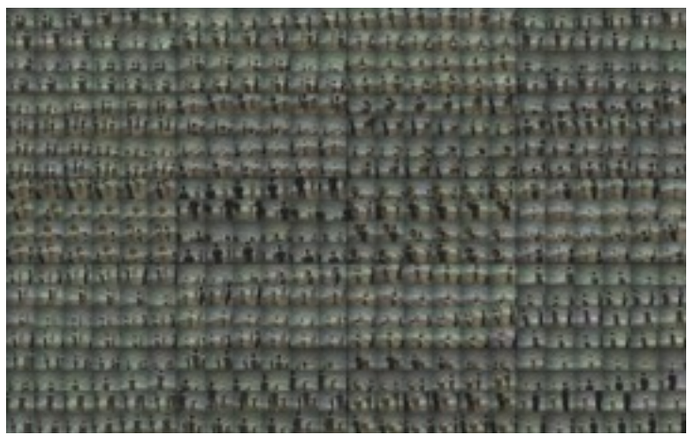
480 Unique Input Views

220GB/min

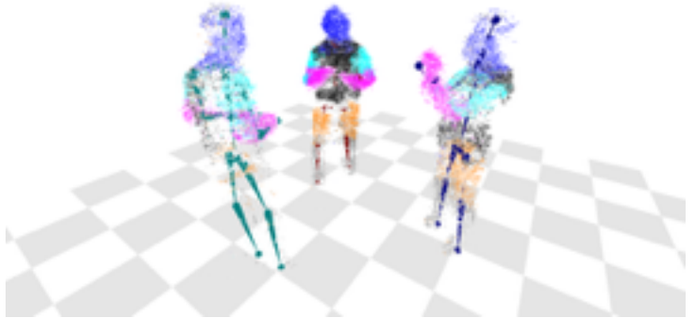
Social Motion Capture Algorithm

Boost A Large Number of “Weak” Perceptional Processes

Goal

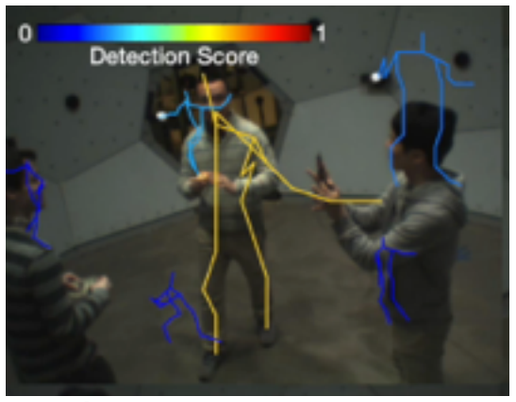


480 view



Automatically labelled trajectories

Low-level cues

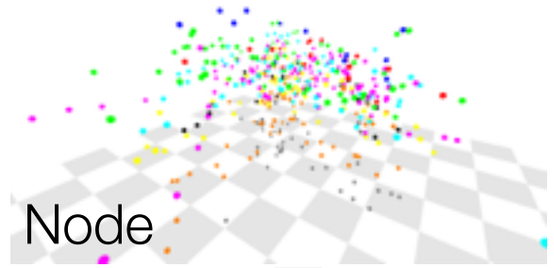


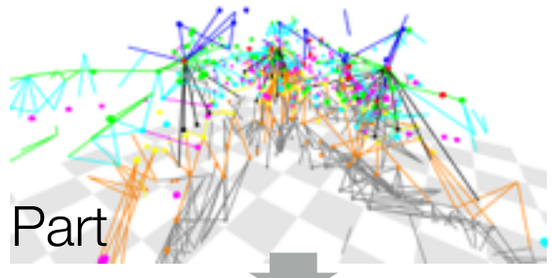
Appearance cue
(Pose detection)

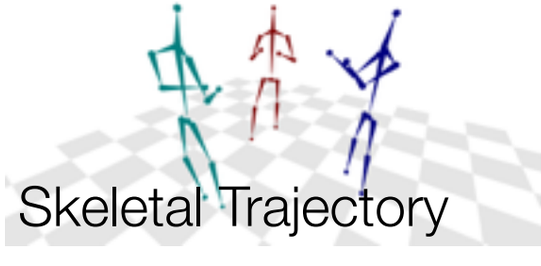
+

Motion cue
(Dense 3D trajectories)

Algorithm flow

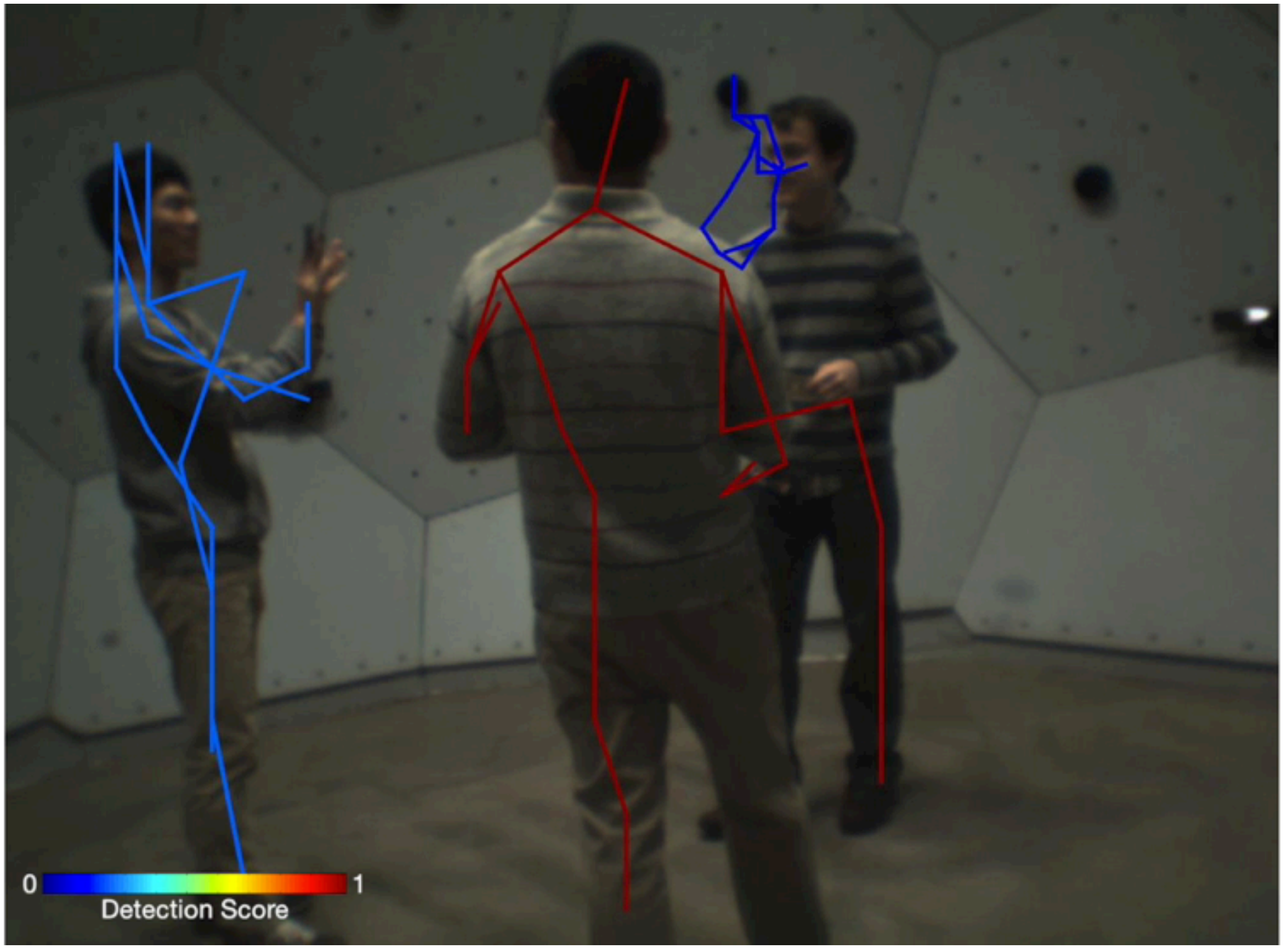






Human Pose Detection

An Example View



[Yang et al., 2012]

Generating 3D Node Score Maps

3D Voting from 2D Score Maps

Camera 1

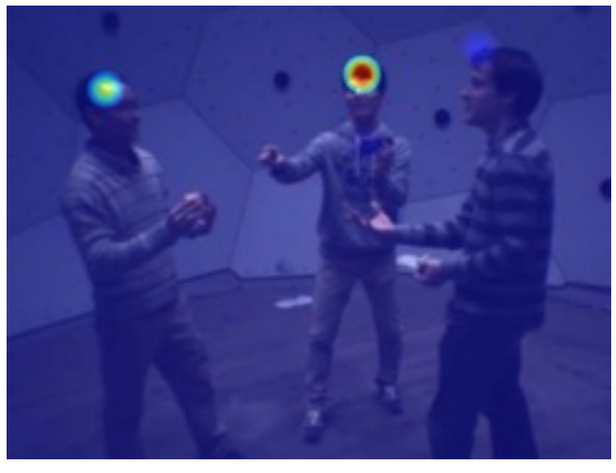
Camera 2

Camera 3

Generating 3D Node Score Maps

3D Voting from 2D Score Maps

Camera 1

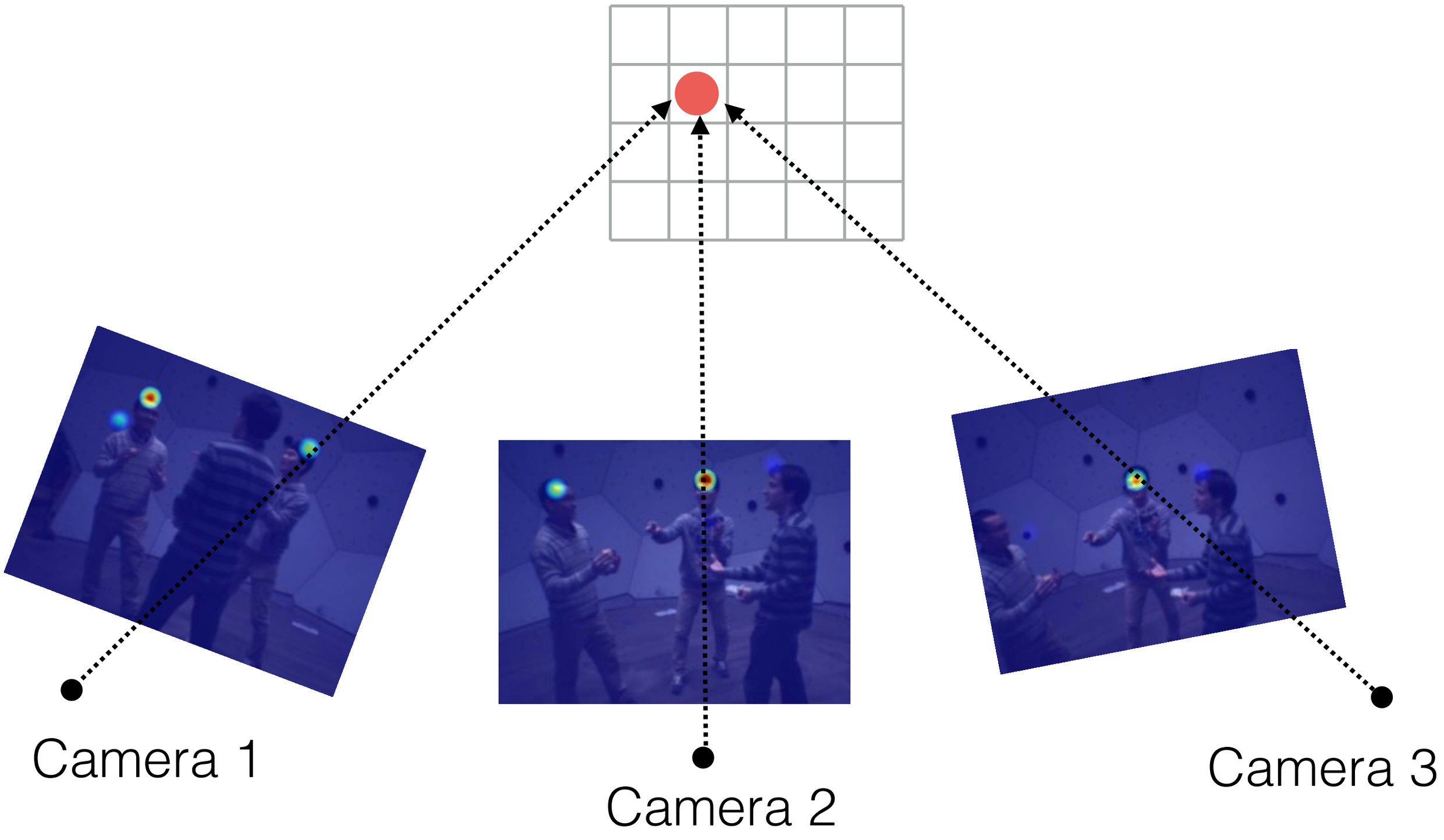


Camera 2

Camera 3

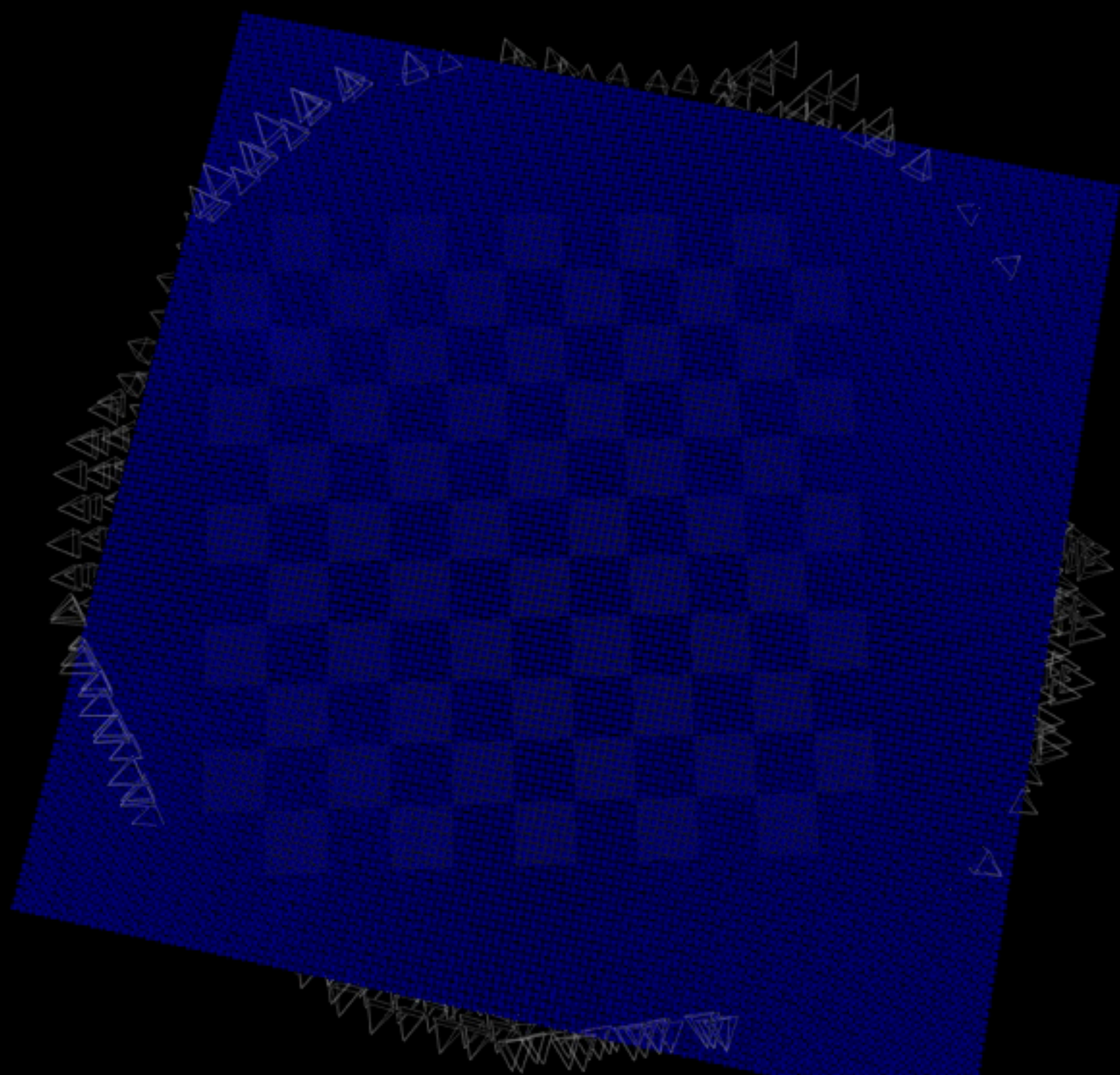
Generating 3D Node Score Maps

3D Voting from 2D Score Maps



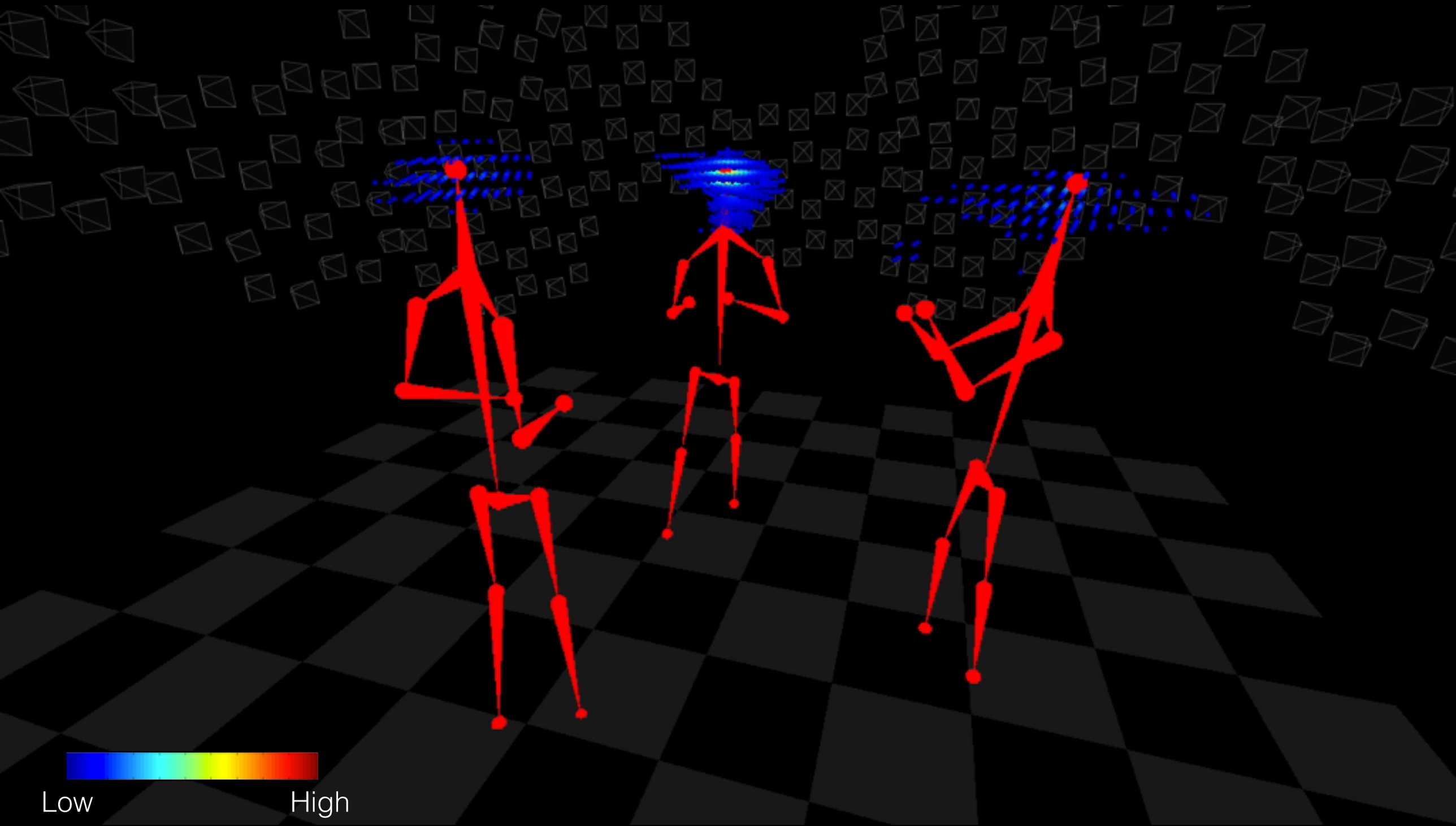
Generating 3D Node Score Maps

Example of Head-top Node



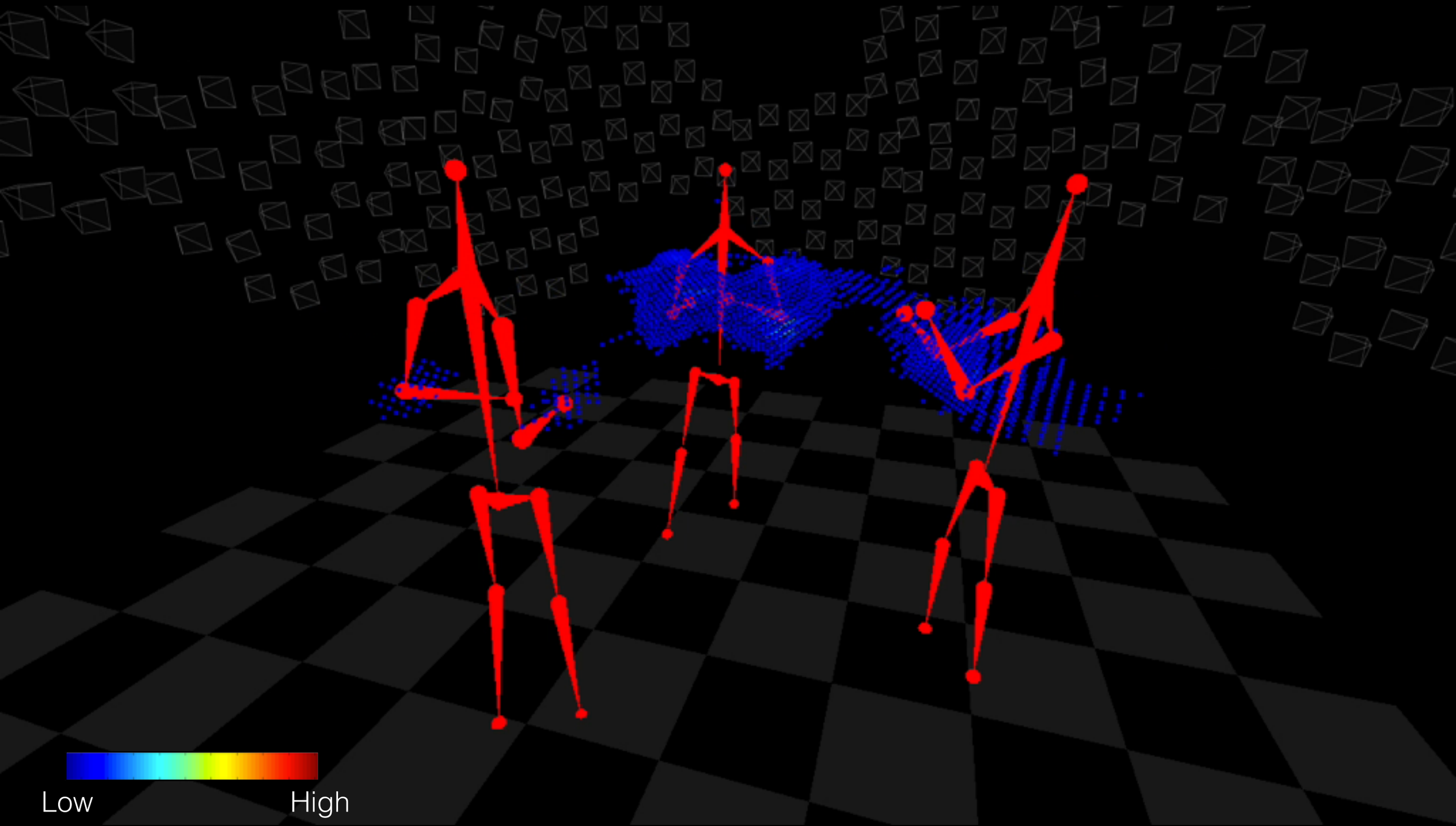
Generating 3D Node Score Maps

Example of Head-top Node



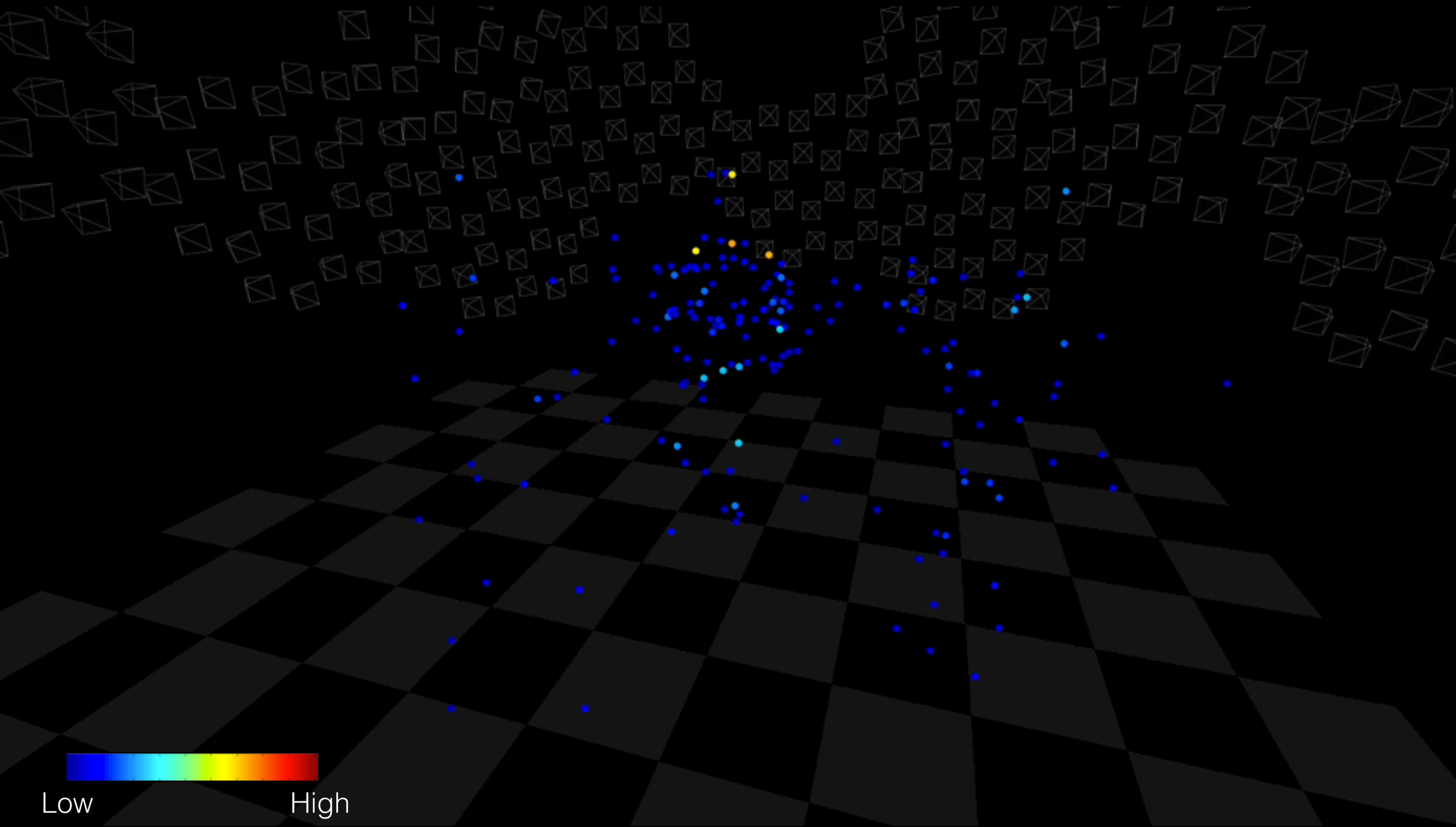
Generating 3D Node Score Maps

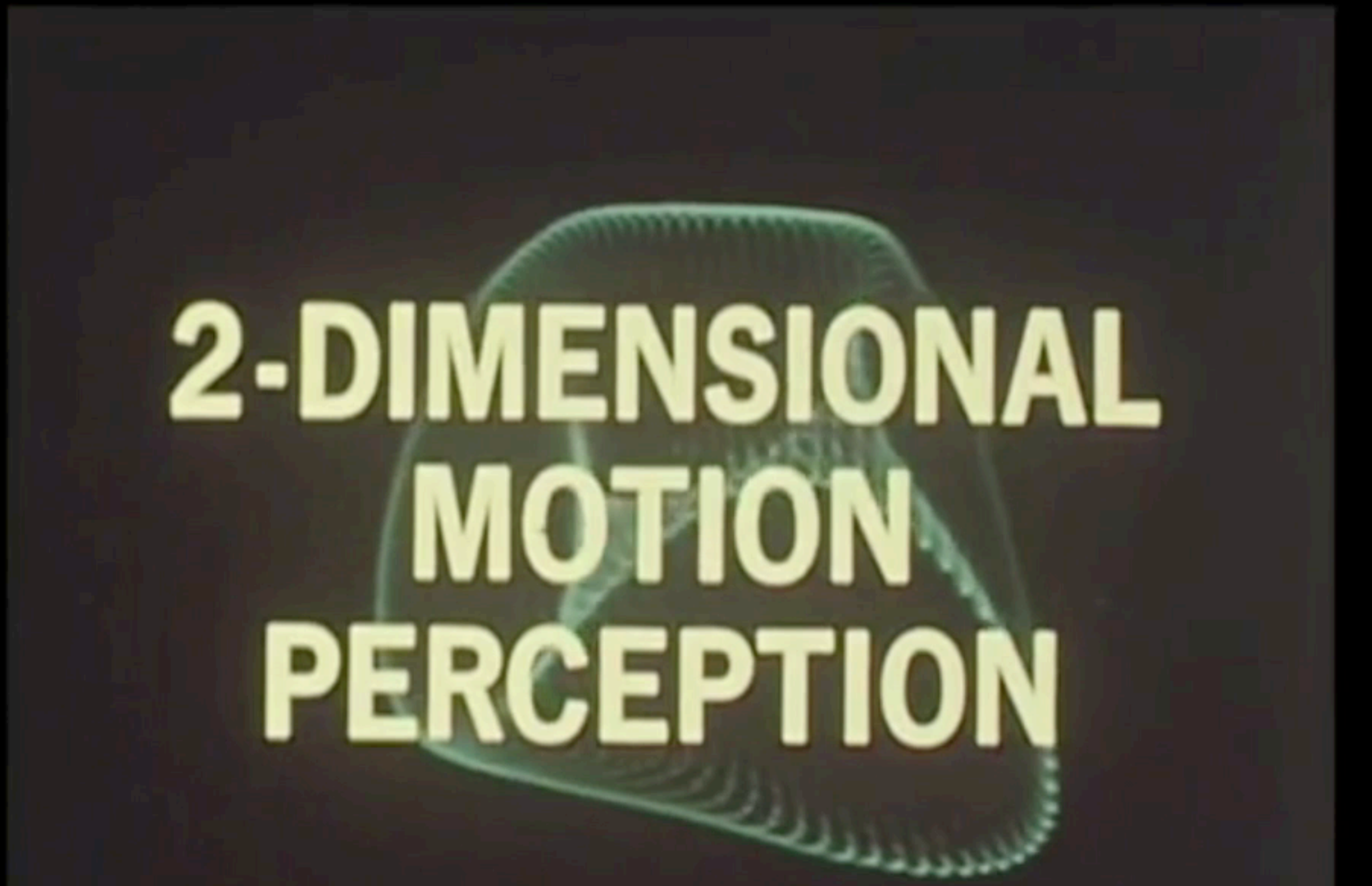
Example of Elbow Node



Generating “Node” Proposals

After Non-Maximum Suppression and Thresholding





G. Johansson (1973). "Visual perception of biological motion and a model for its analysis”

Perceiving rigidity

G. Johansson (1973). "Visual perception of biological motion and a model for its analysis”

Dense 3D Trajectory Stream

Leverage 2D Flows from Large Number of Views

[Joo et al. CVPR 2014]

Generating “Part Trajectory” Proposals

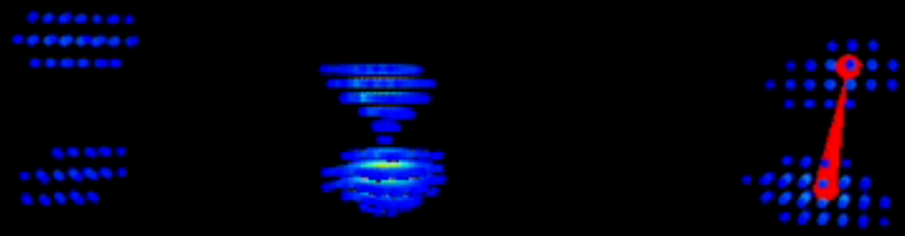
Associating Part with 3D Dense Trajectories

Generating “Part Trajectory” Proposals

Associating Part with 3D Dense Trajectories

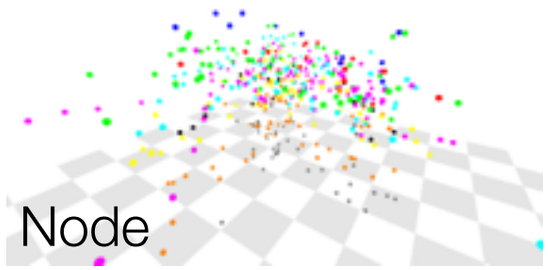
Generating “Part Trajectory” Proposals

Part Trajectory Scoring

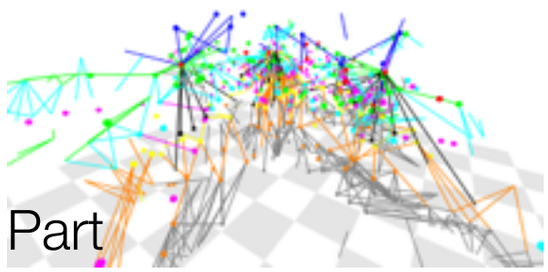


Possible to determine rigidly moving body part
<u>without a prior template</u>
(shape, texture, bone-length, number of subjects)

Algorithm Flow

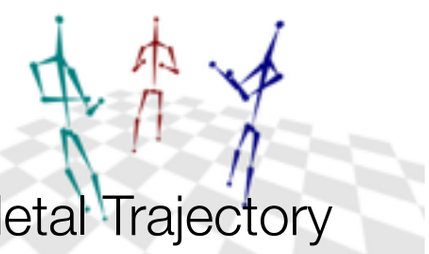


Node



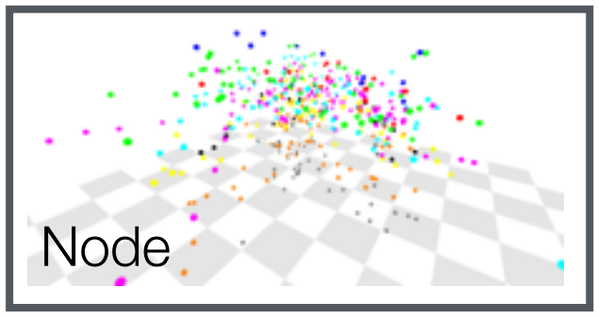
Part

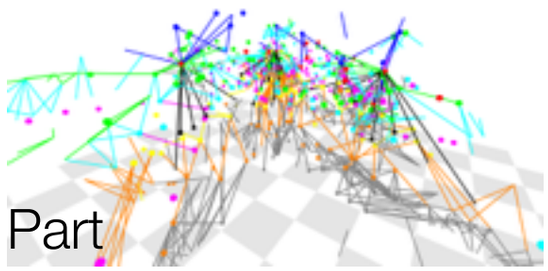
Part Trajectory

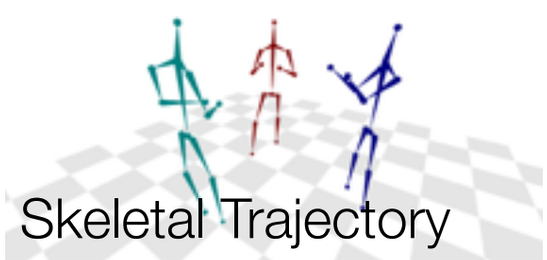


Skeletal Trajectory

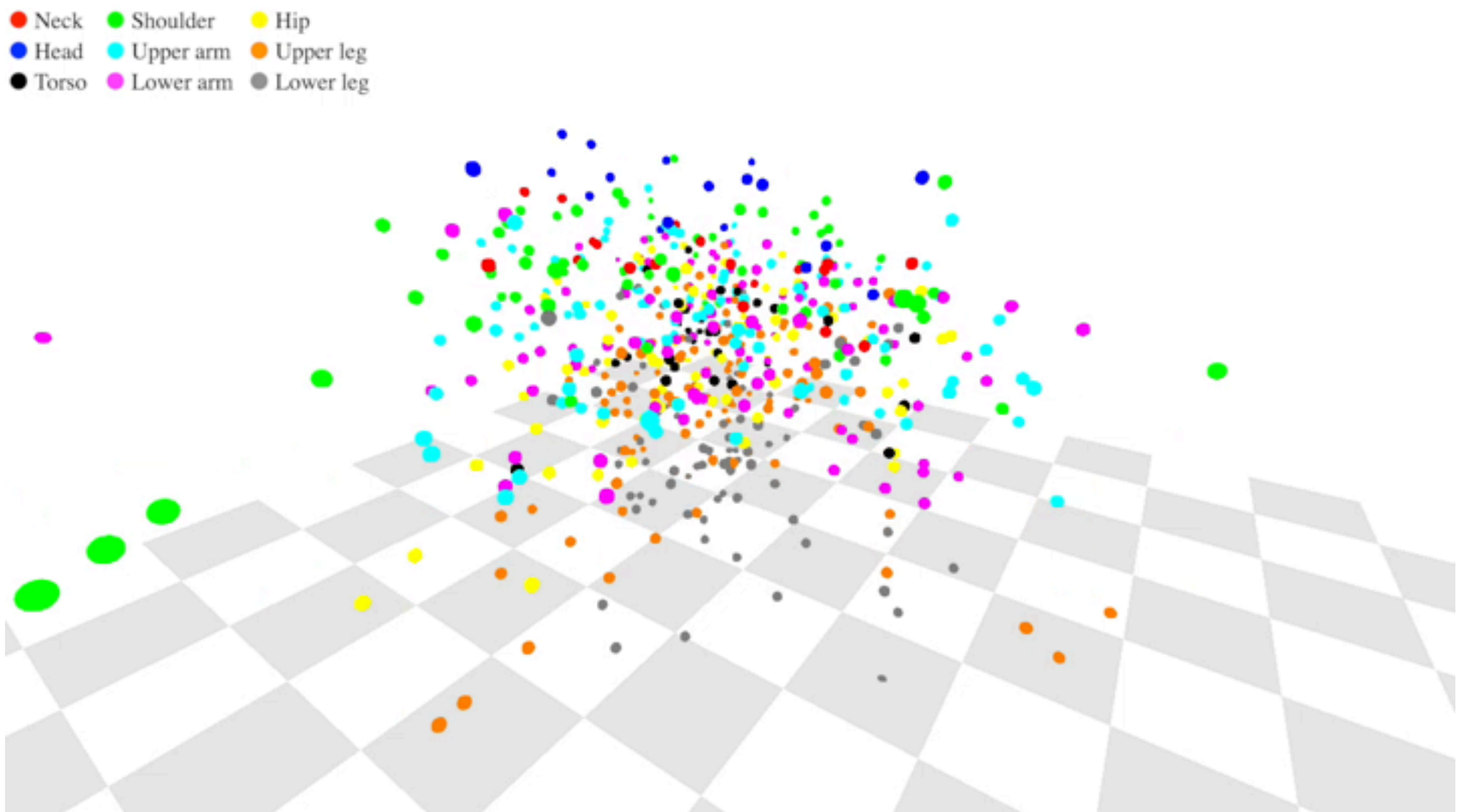
Algorithm Flow



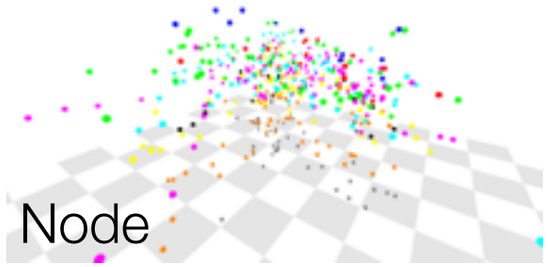


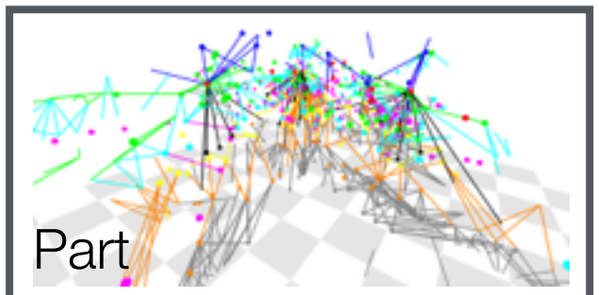


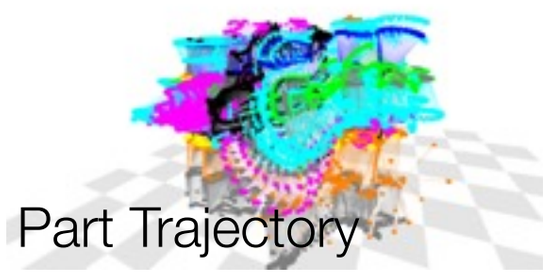
“Node” Proposals

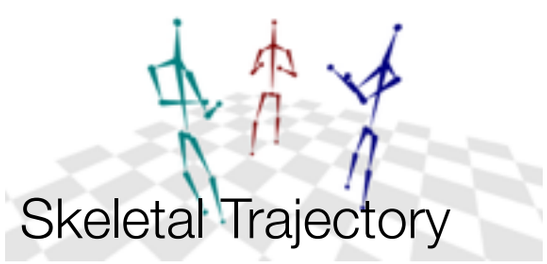


Algorithm Flow

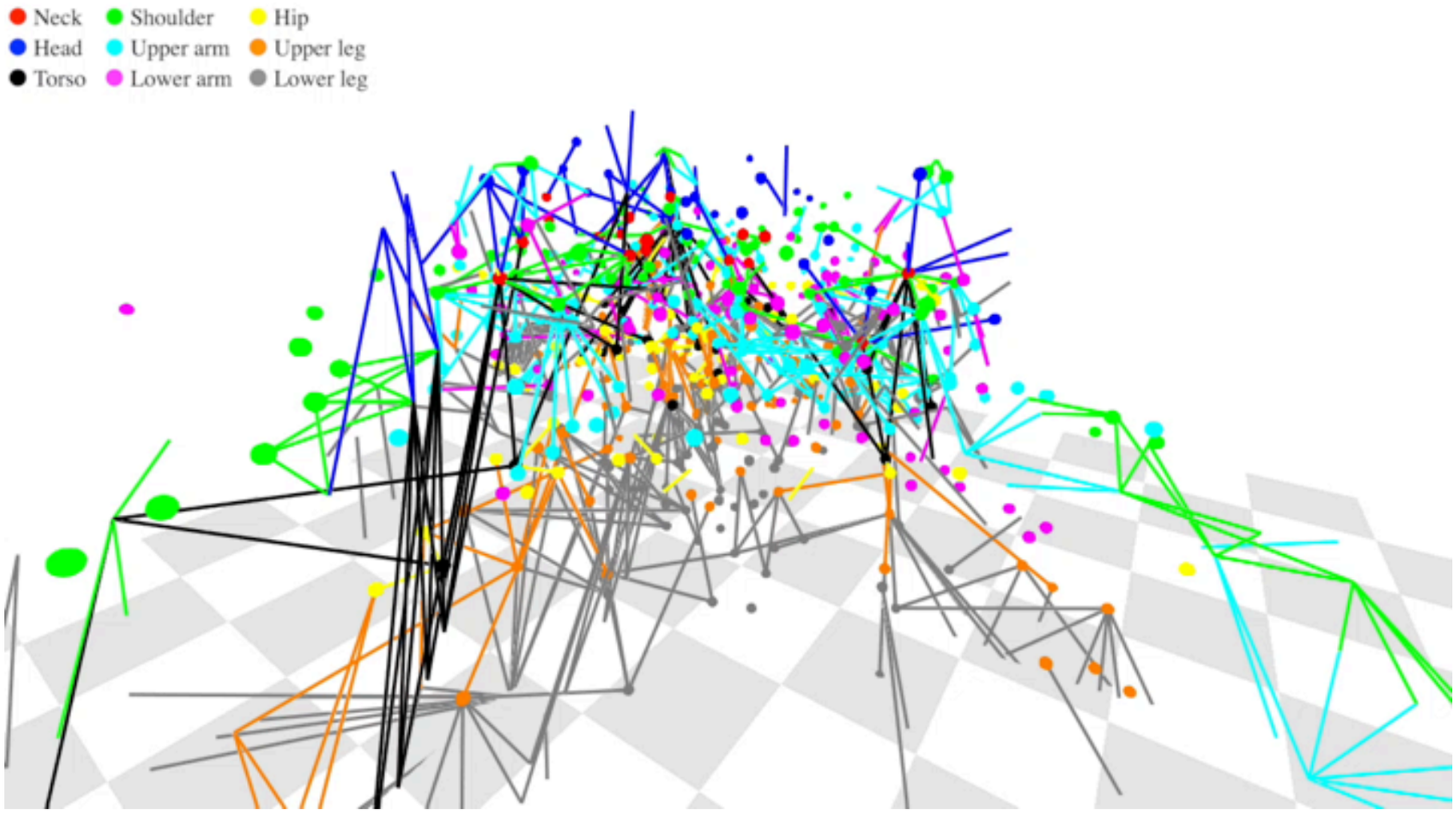




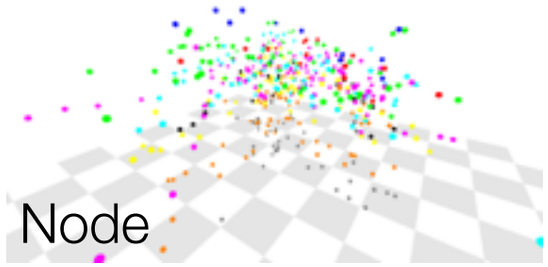


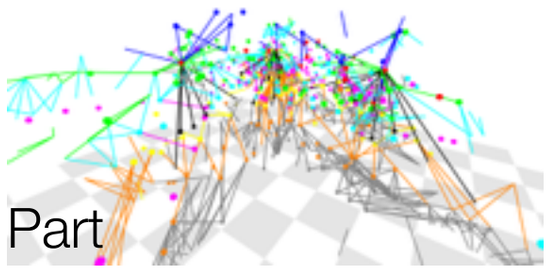


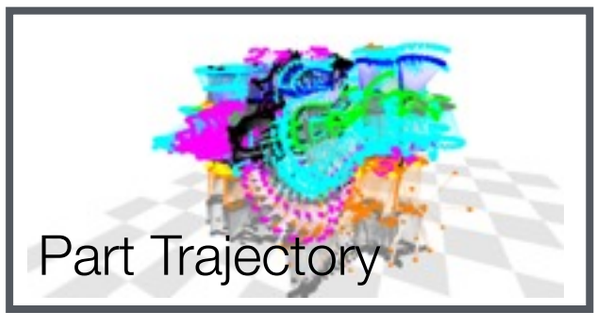
“Part” Proposals

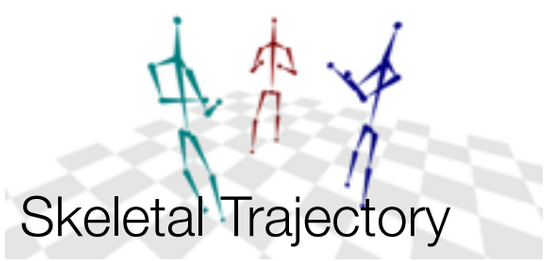


Algorithm Flow

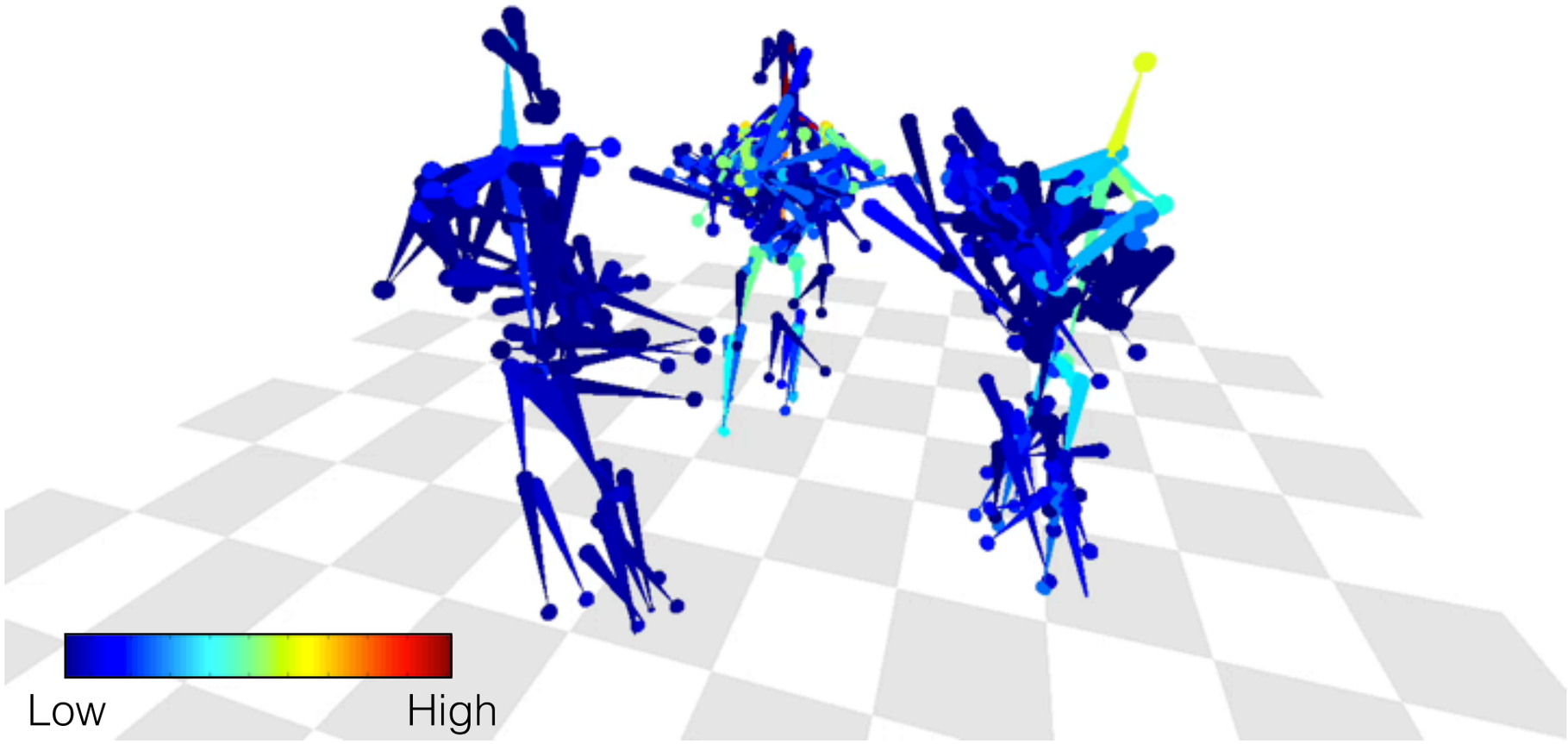




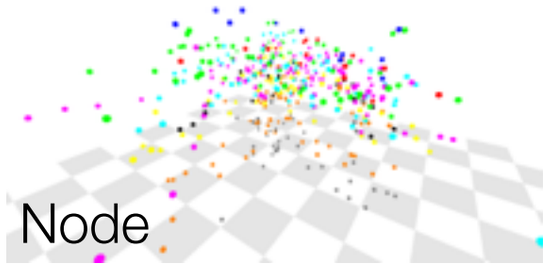




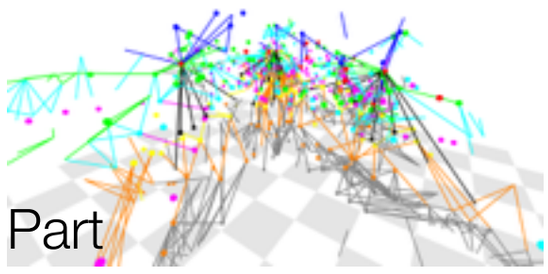
“Part Trajectory” Proposals



Algorithm Flow

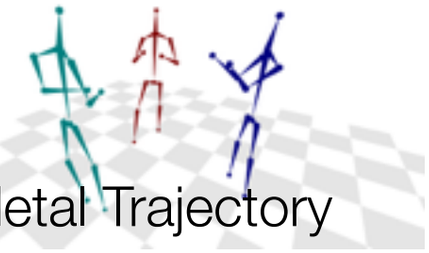


Node



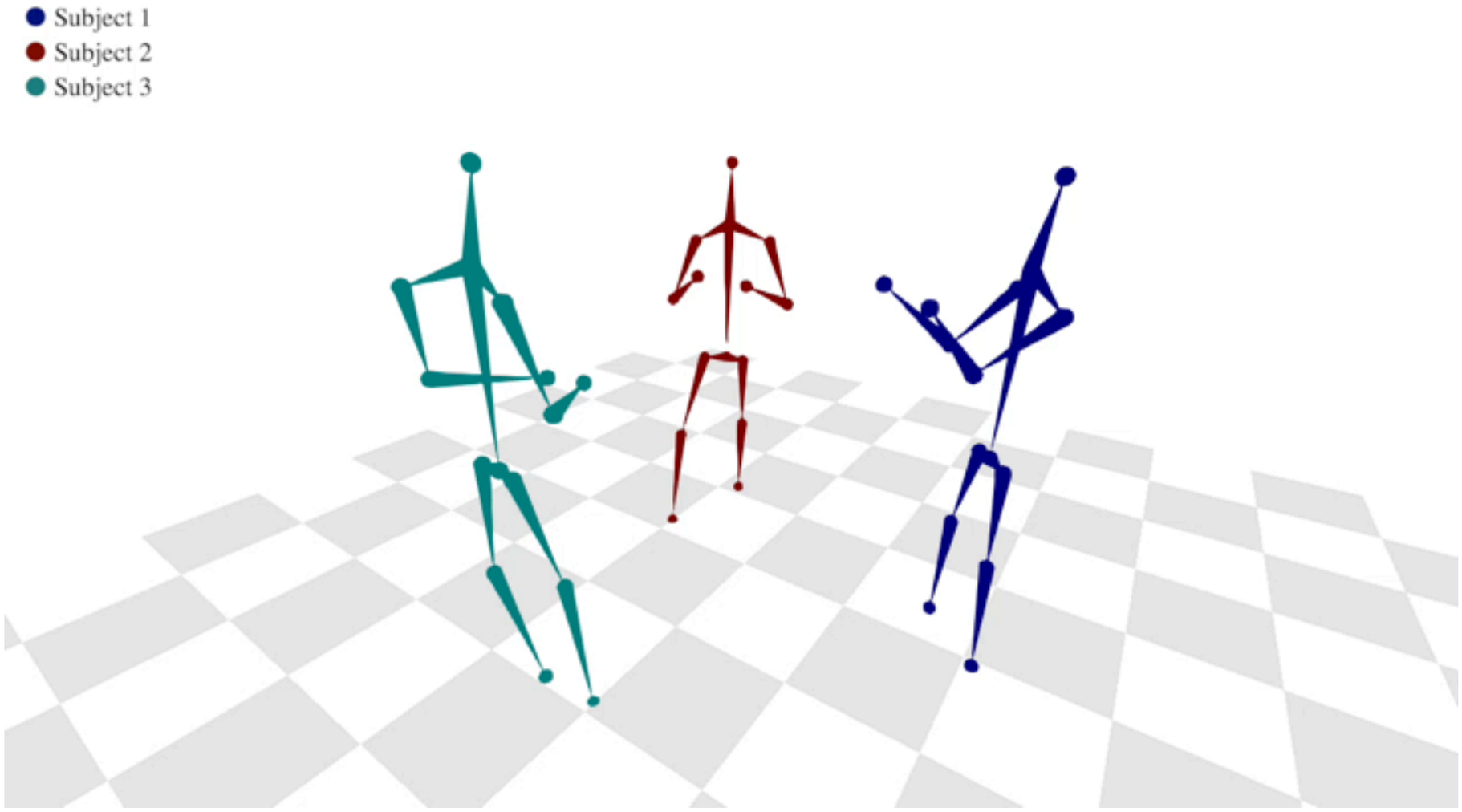
Part

Part Trajectory



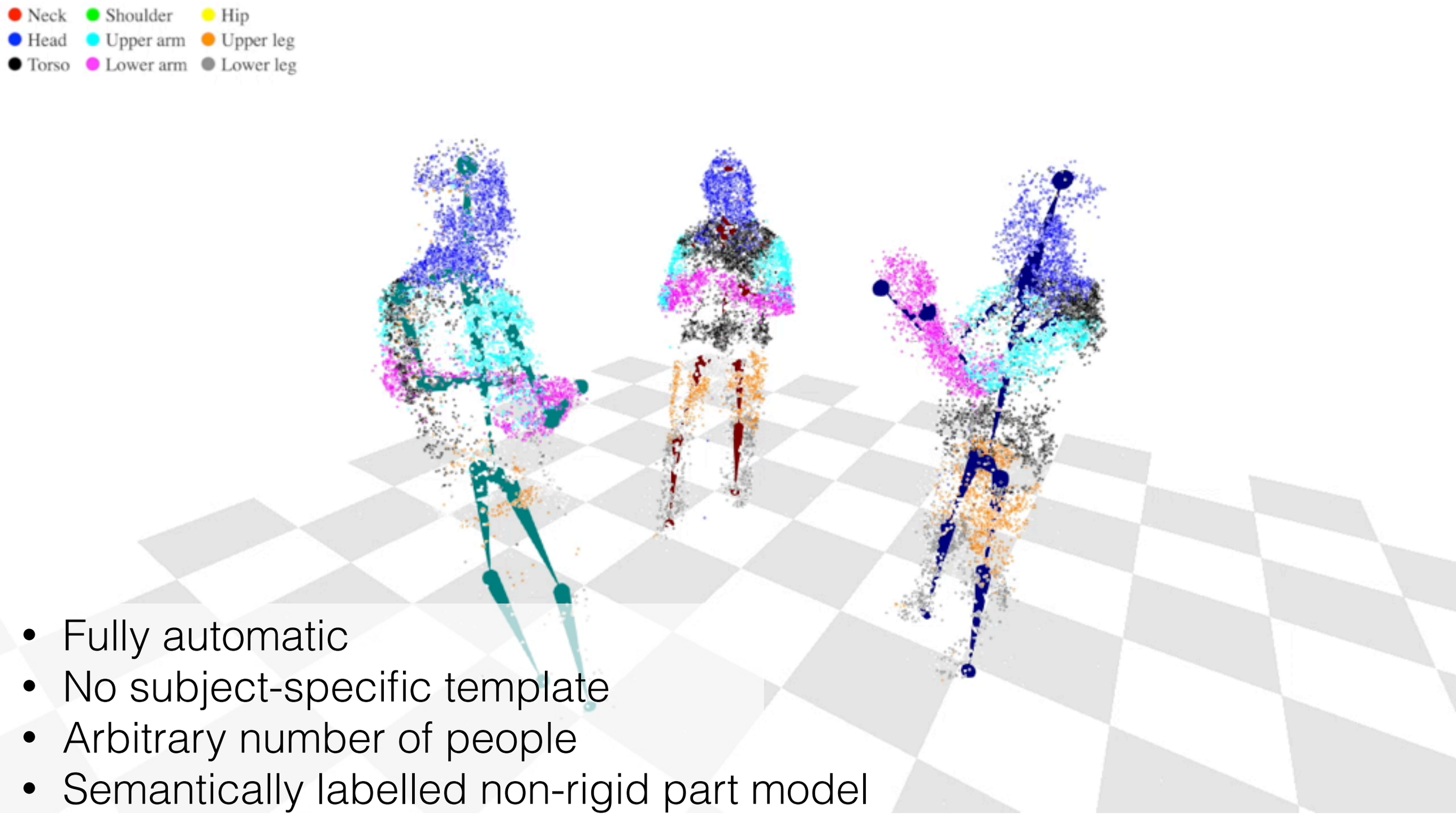
Skeletal Trajectory

“Skeletal Trajectory” Proposals



“Skeletal Trajectory” Proposals

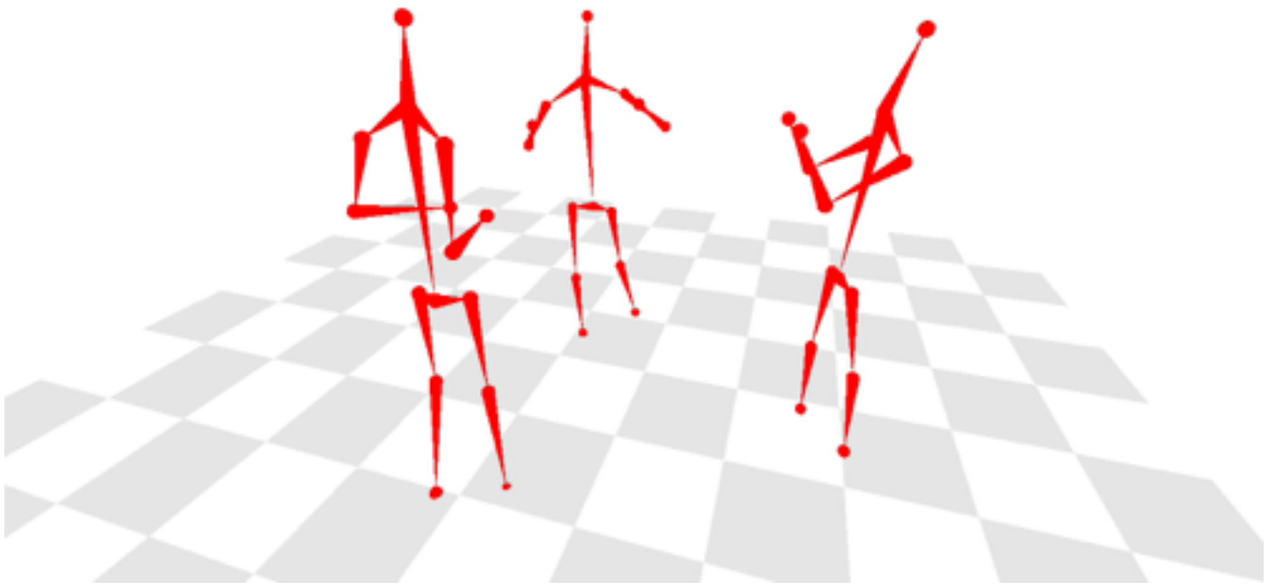
Labelled Non-Rigid Part Representation



- Fully automatic
- No subject-specific template
- Arbitrary number of people
- Semantically labelled non-rigid part model

Evaluation

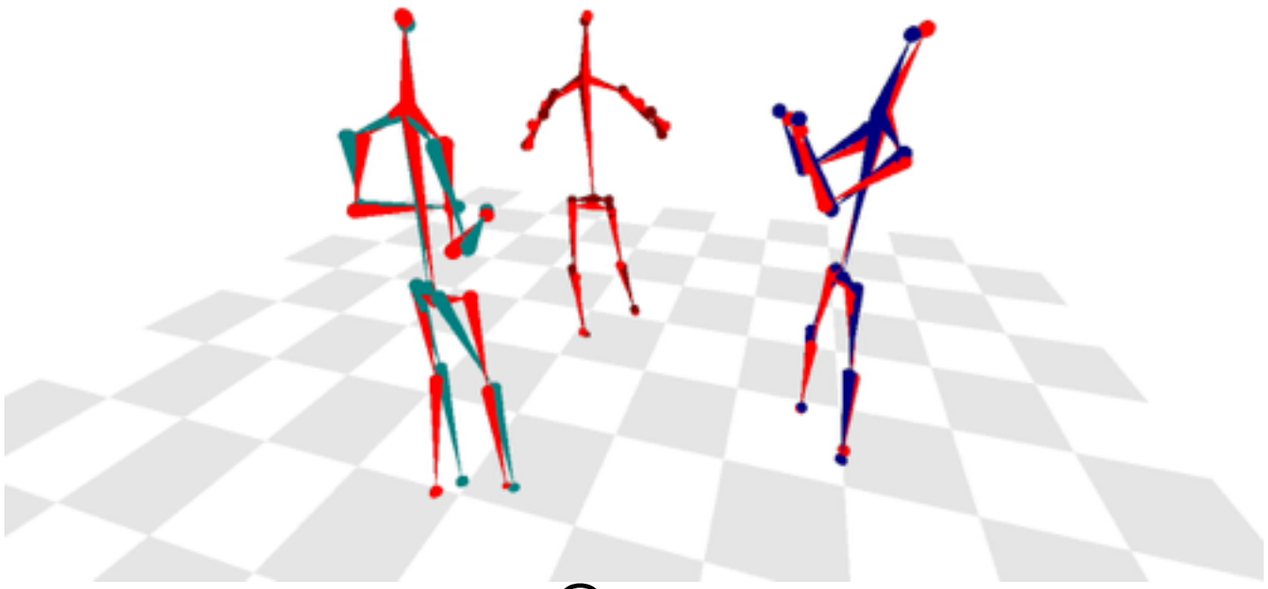
Two Baselines: 3DPS and Multiple Kinects



Ground truth
by manually annotating 2D locations
(every 4th frame)

Evaluation

Two Baselines: 3DPS and Multiple Kinects

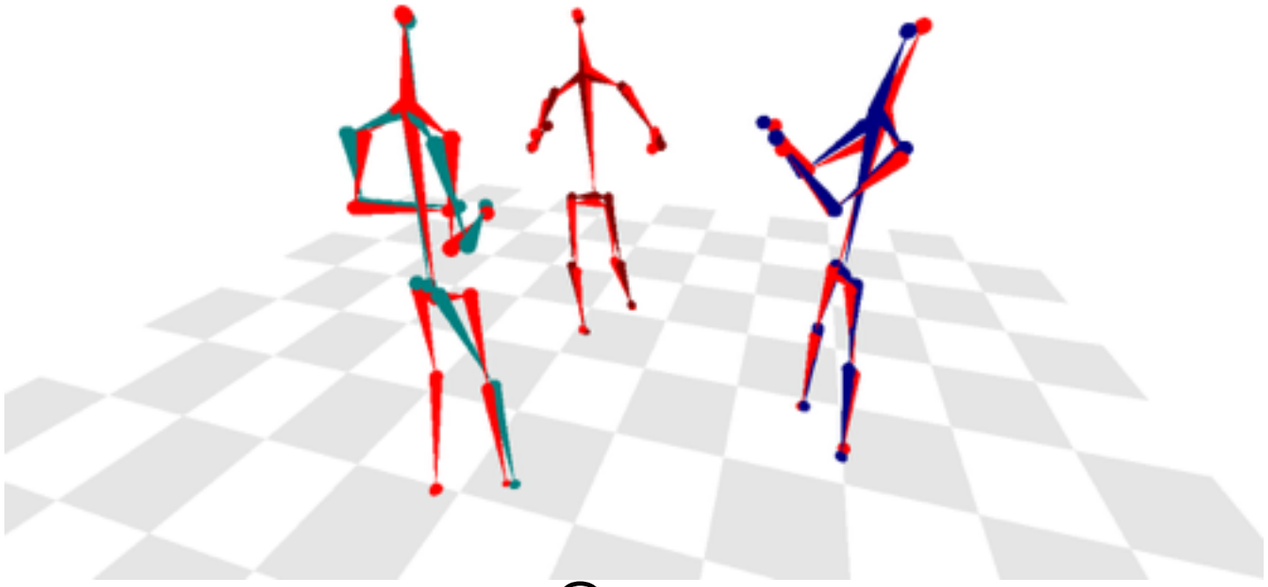


Ours

Average node errors: 4.88 cm

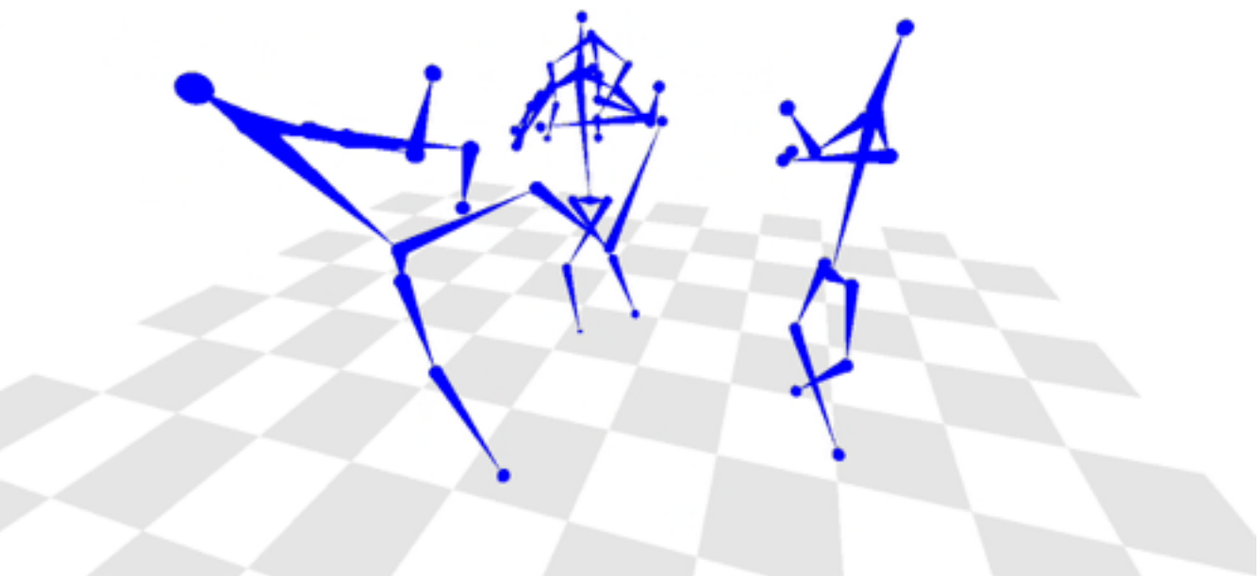
Evaluation

Two Baselines: 3DPS and Multiple Kinects



Ours

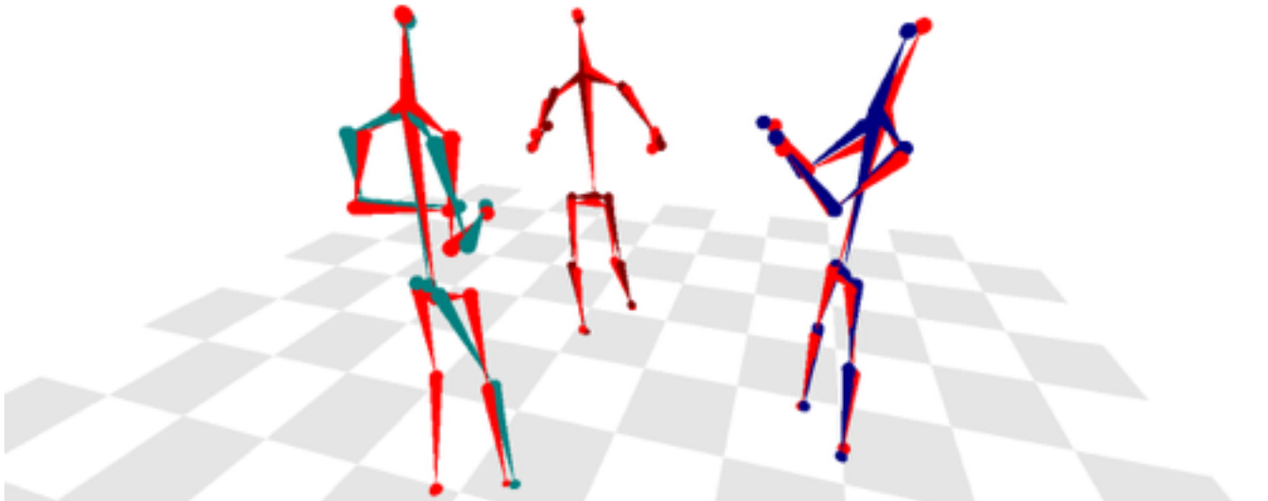
Average node errors: 4.88 cm



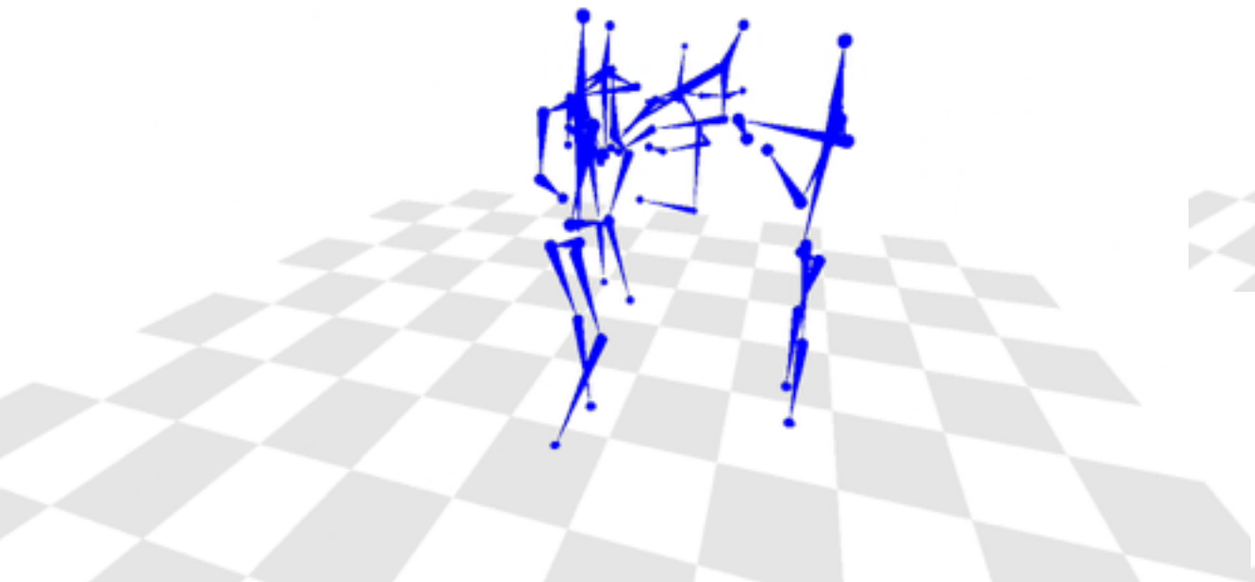
3D Pictorial Structure (All candidates)

Evaluation

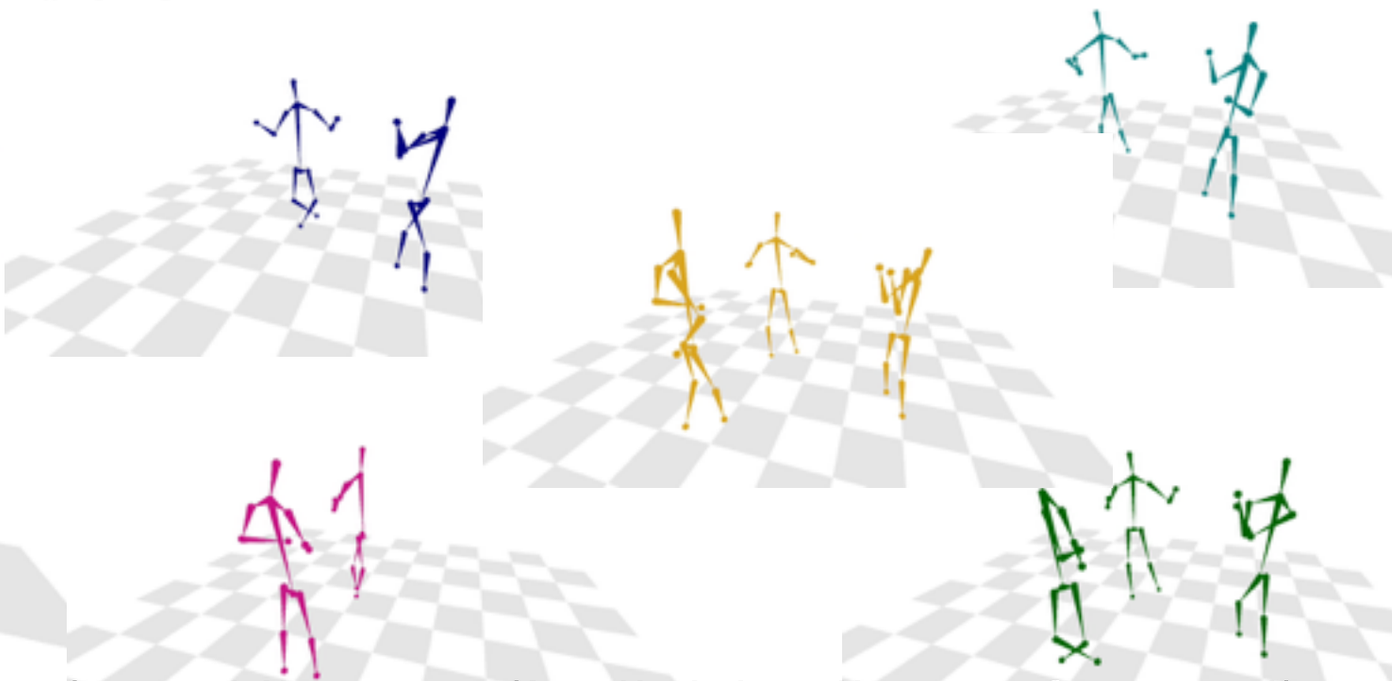
Two Baselines: 3DPS and Multiple Kinects



Ours



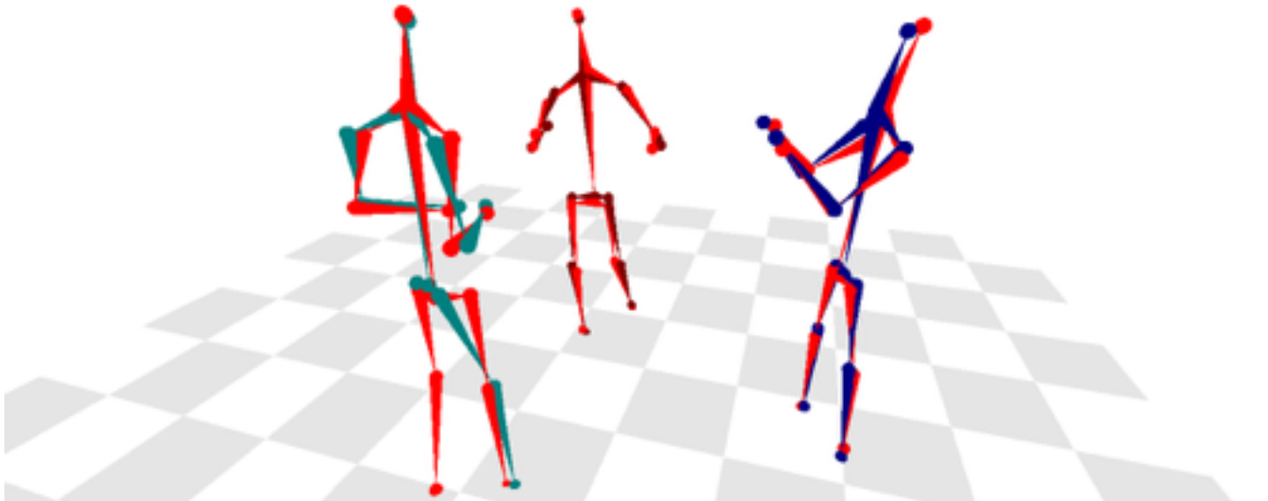
3D Pictorial Structure (All candidates)



5 Kinects (Individually performed)

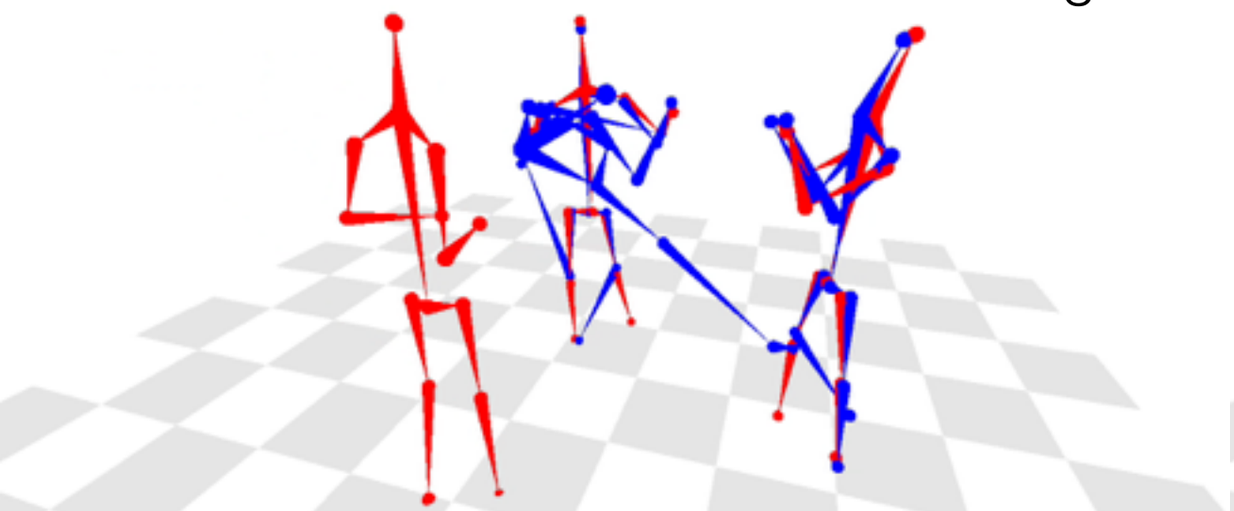
Evaluation

Two Baselines: 3DPS and Multiple Kinects



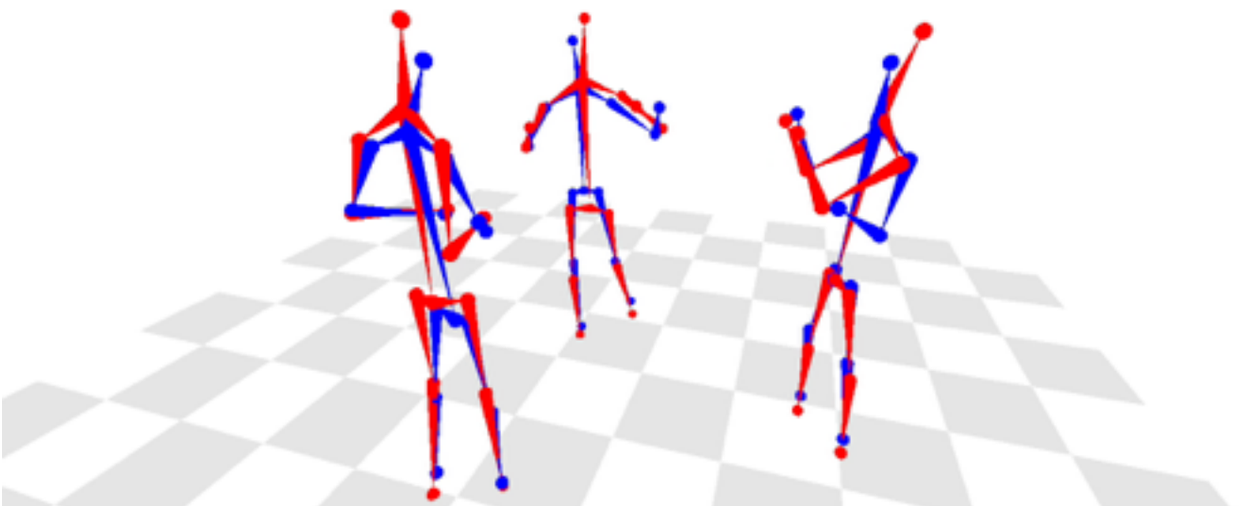
Ours

Average node errors: 4.88 cm



3D Pictorial Structure (Oracle)

Average node errors: 15.35 cm



5 Kinects (Oracle)

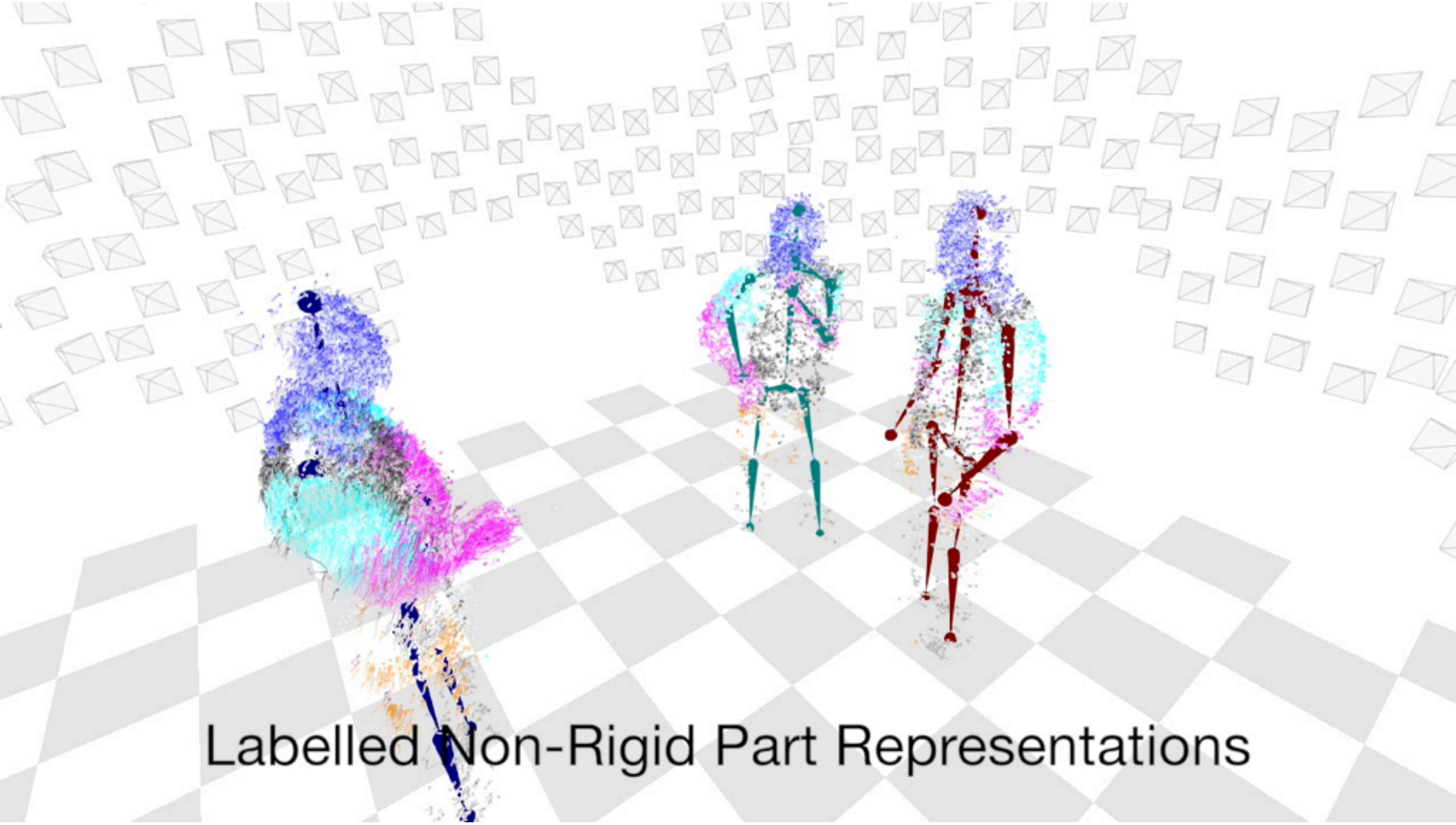
Average node errors: 5.55 cm

Social Motion Capture Result

The Ultimatum Sequence

Social Motion Capture Result

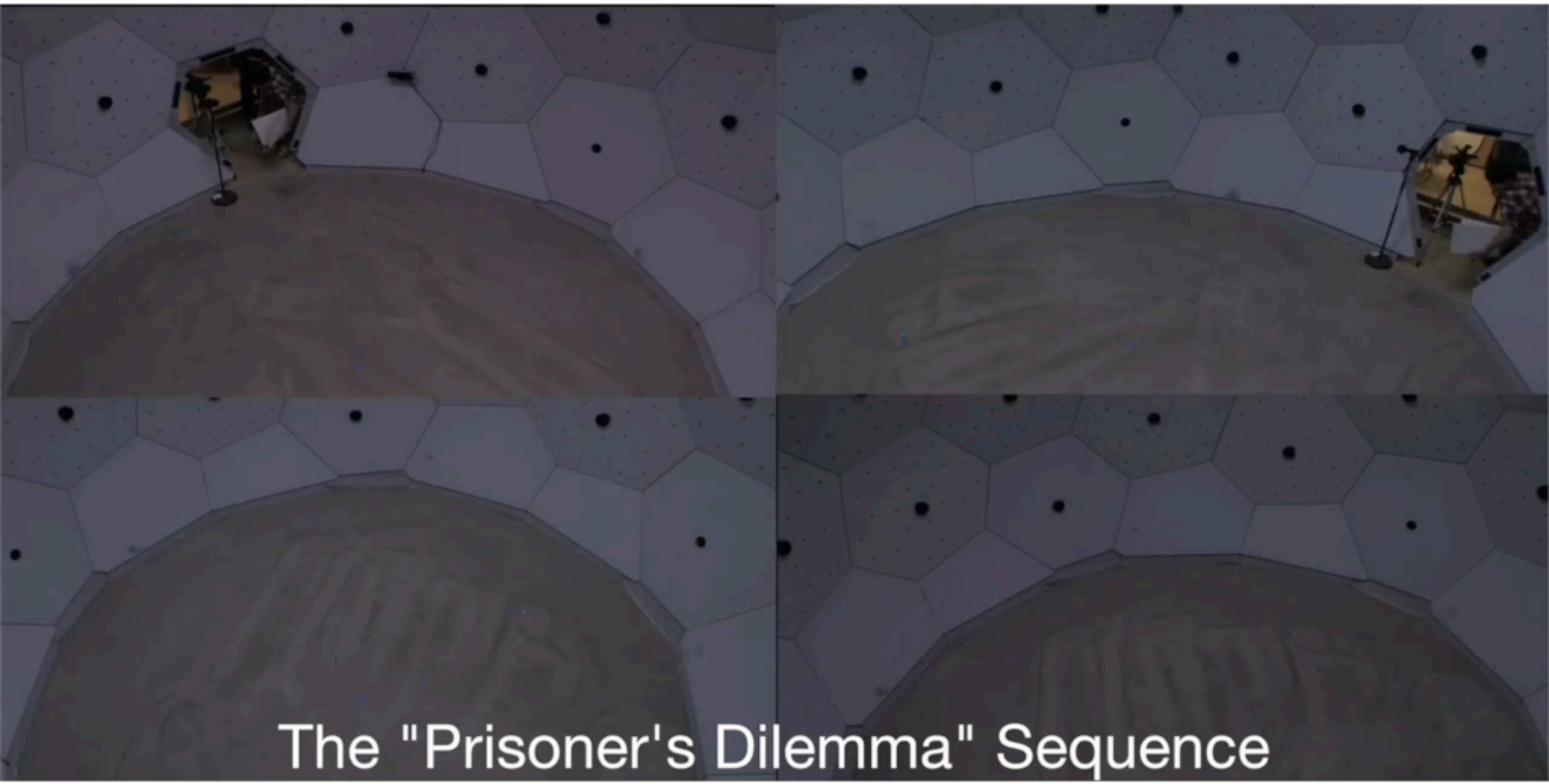
The Ultimatum Sequence



Labelled Non-Rigid Part Representations

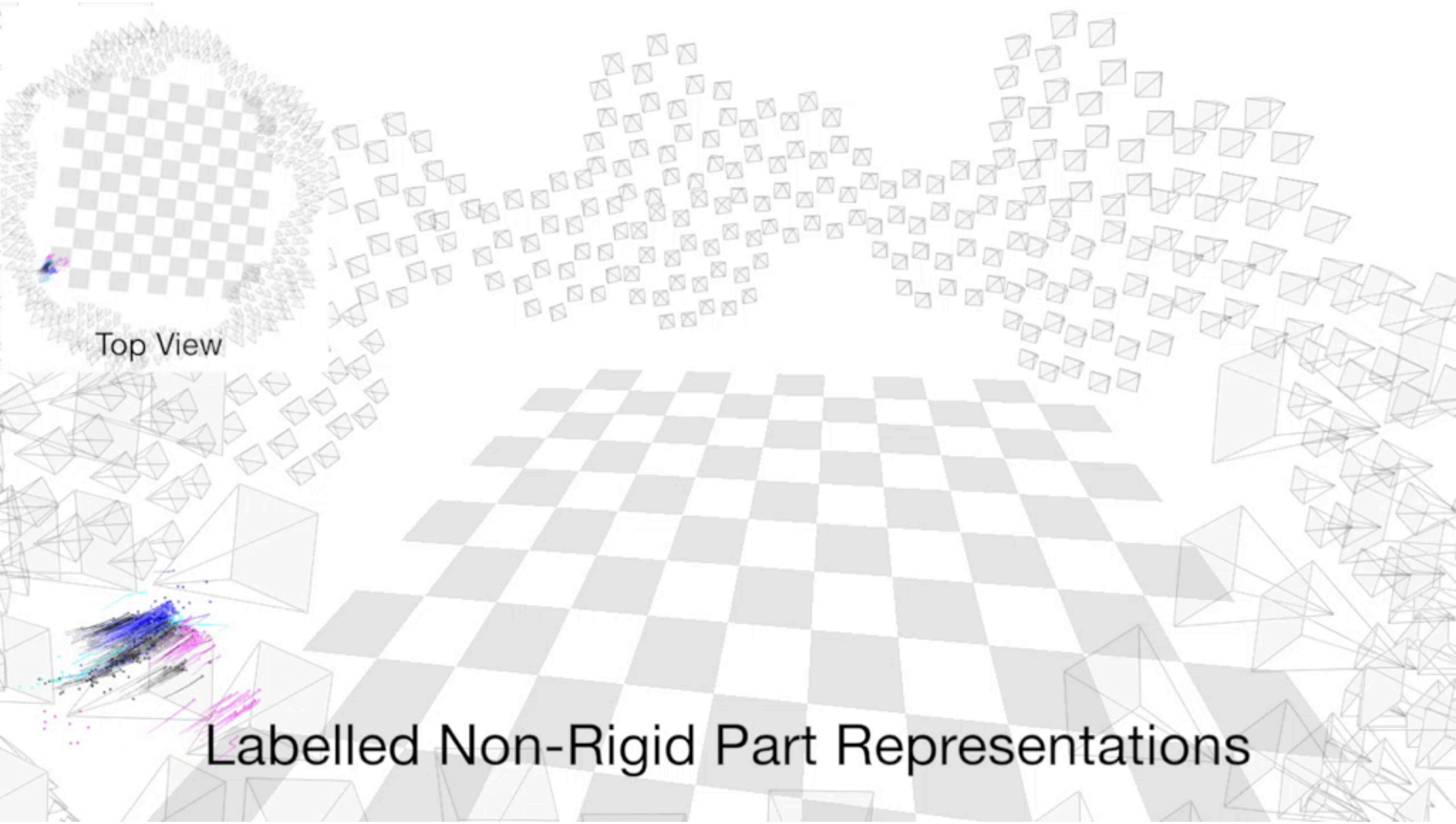
Social Motion Capture Result

The Prisoner’s Dilemma Sequence



Social Motion Capture Result

The Prisoner's Dilemma Sequence



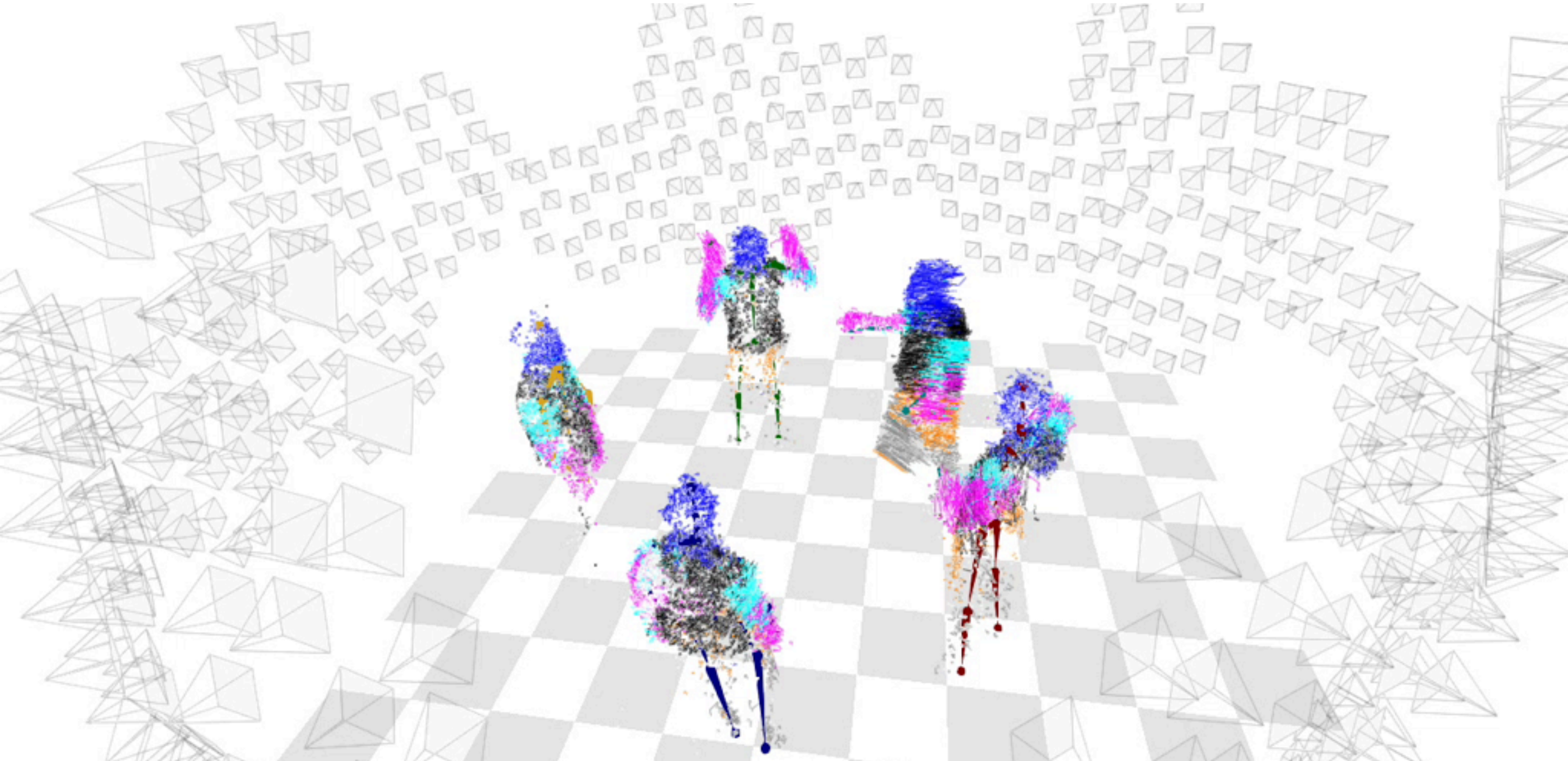
Social Motion Capture Result

The 007-Bang Sequence

The "007-Bang" Sequence

Social Motion Capture Result

The 007-Bang Sequence



Labelled Non-Rigid Part Representations

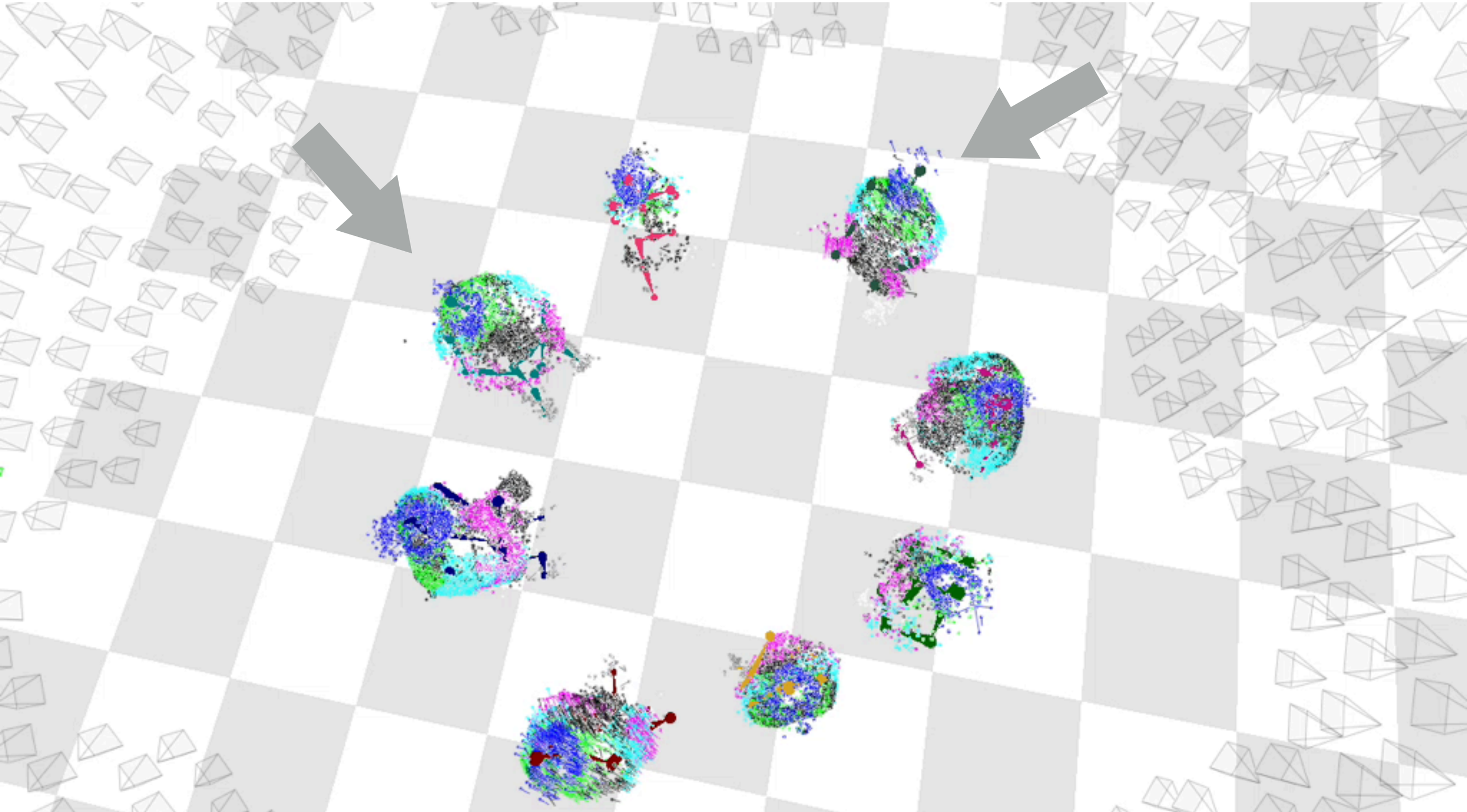
Social Motion Capture Result

The Mafia Sequence

The "Mafia" Sequence

Social Motion Capture Result

The Mafia Sequence



Future Work

Analyzing Nonverbal Signals of Interacting People

Thank you

Dataset will be available:

http://www.cs.cmu.edu/~panoptic-studio/