

人間挙動の常時センシングのための コンピュータビジョン技術

斎藤 英雄 (慶應義塾大学)

hs@keio.jp

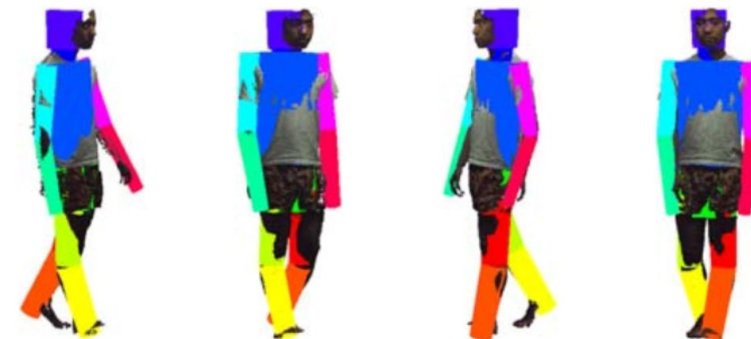
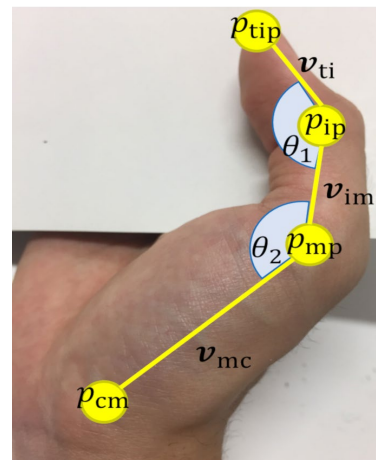
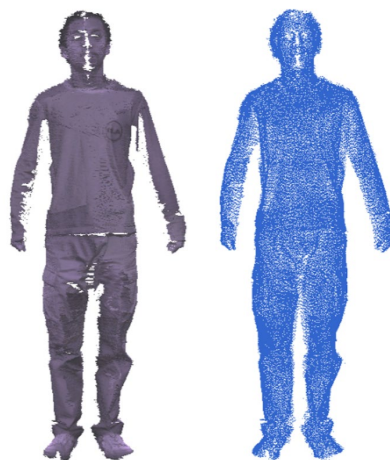
コンピュータビジョン

- 人間の視覚のコンピュータによる実現
 - 手段：カメラ＋コンピュータ．カメラは非接触なセンサ．
 - 要素となる技術
 - センシング（眼）：**物理量の計測**
 - 大きさ，形，色，材質．．．．
 - 理解・認識（脳）：**対象の意味・状態を判断**
 - 名前，意味，感情，意図
 - 人間挙動をセンシング・理解・認識
-

人間挙動：カメラでセンシングできそうなデータ

■ 幾何学的物理量

- 形状
- 関節角度
- 姿勢



■ 光学的物理量

- 色, 明るさ, 温度 (物体からカメラに到達する光のエネルギー)

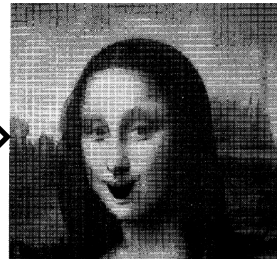
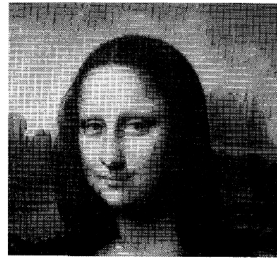
■ 上記から推測する内部状態

- 意図, 感情
- 生理的状态
- 機能

人間をカメラでセンシング

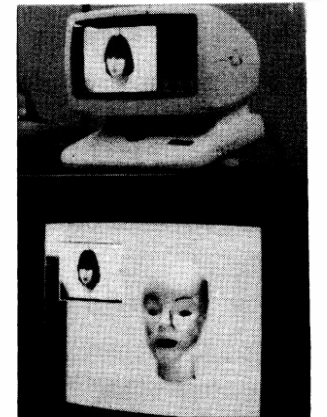
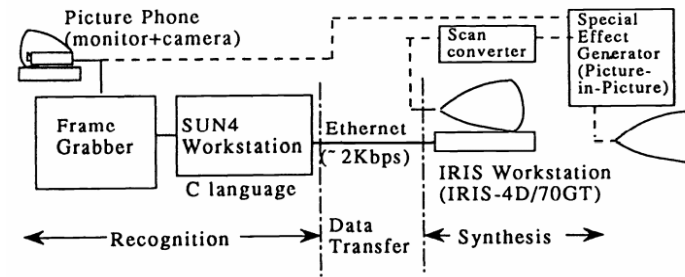
■ 昔からある.

- 知的符号化
- 動作で機械をコントロール
- ゲーム用インタフェース



[原島88]

[Mase90]



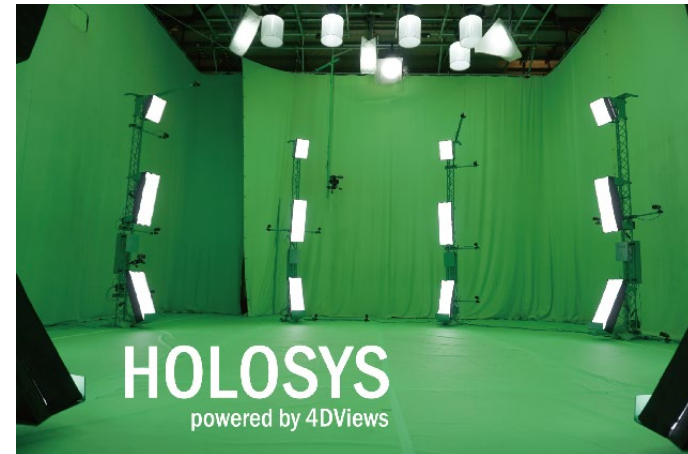
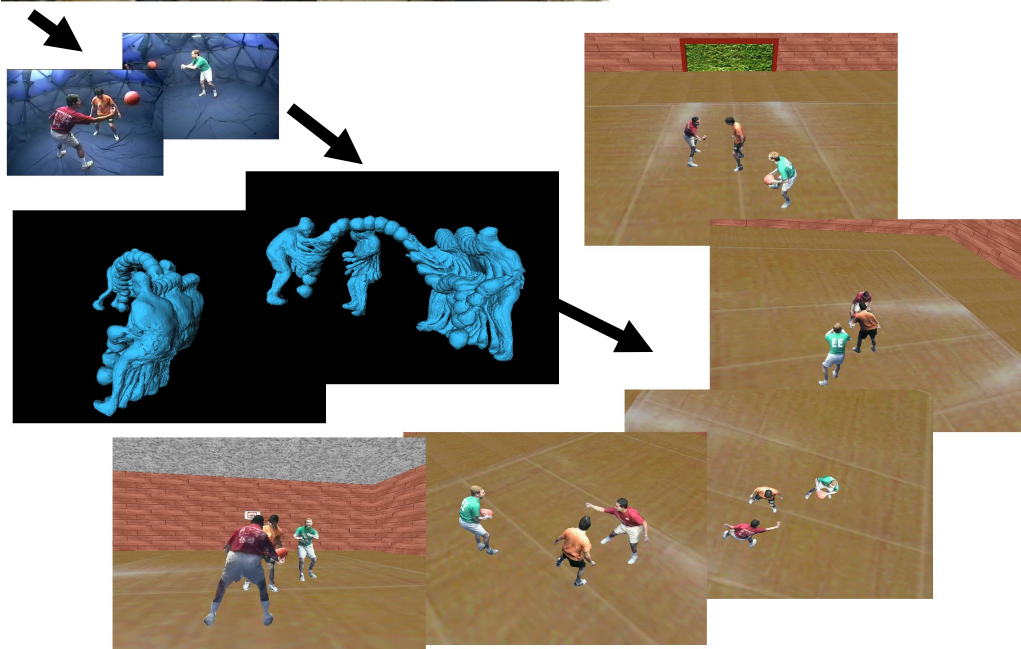
■ ブレークスルー

- カメラの数 > 人の数
- 撮影したデータ量 > 見ることが可能な量
- 機械学習の進化

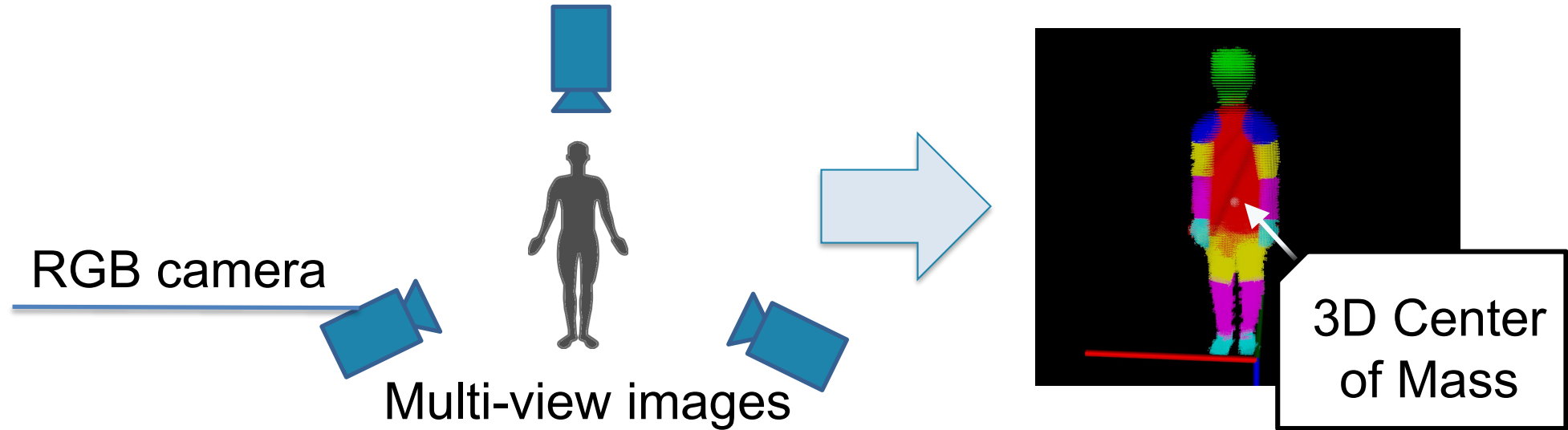


環境側にカメラを取り付ける

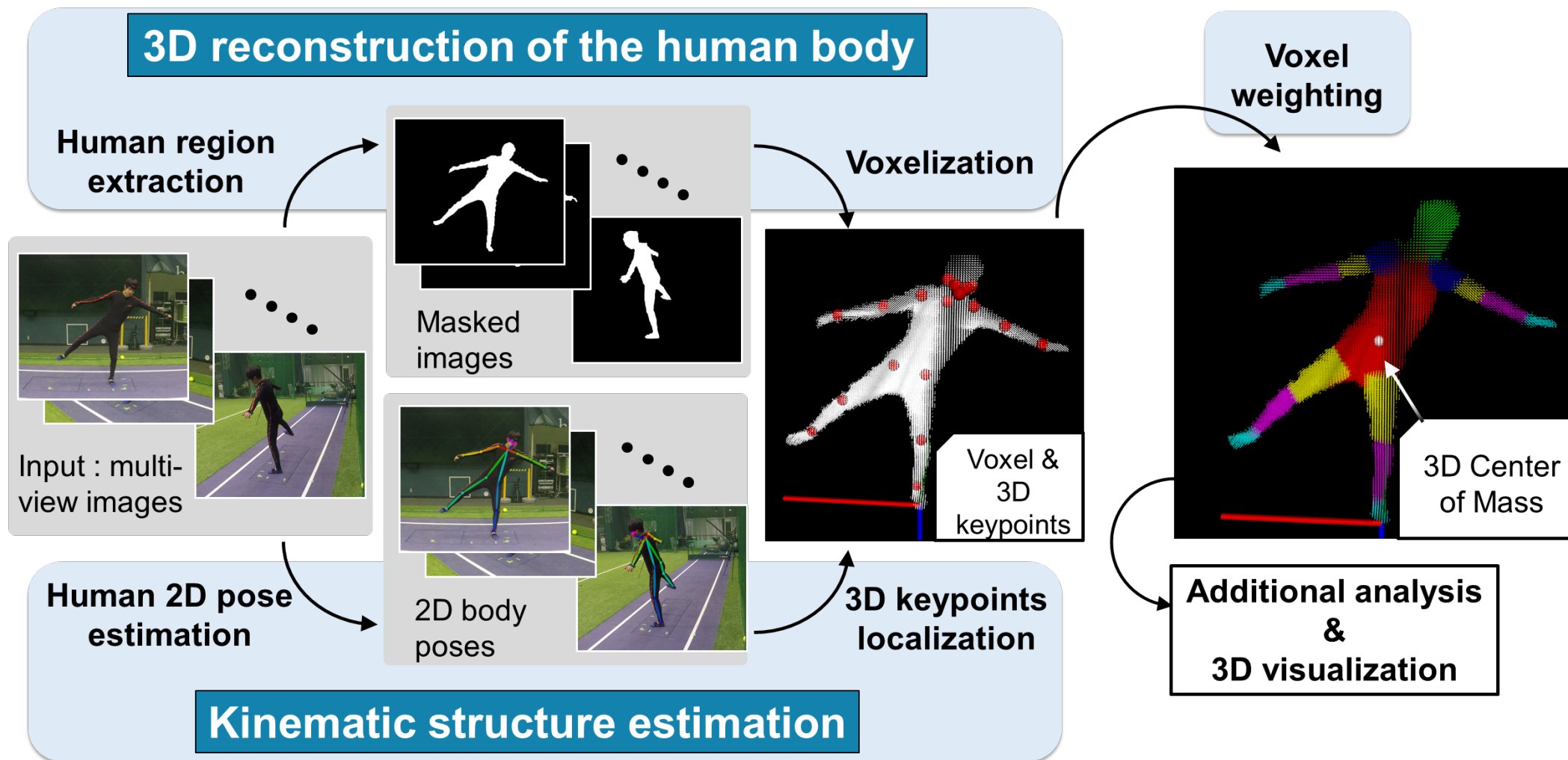
- Virtualized Reality[金出96]
- Vicon
- The CMU Panoptic Studio



複数固定カメラ映像による重心位置推定 [Kaichi CVSports18]



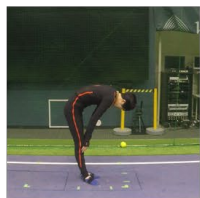
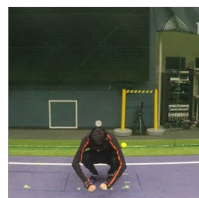
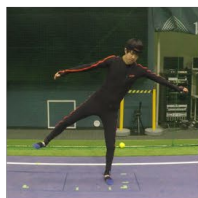
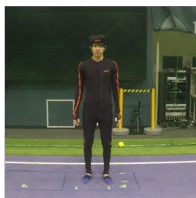
重心推定の流れ



CoPの推定精度評価

Setups

- Five cameras calibrated beforehand
- Three subjects (two male and one female) each stood on the force plate with four static posture
- Measured distance between the CoP and the point where the estimated CoM was projected vertically



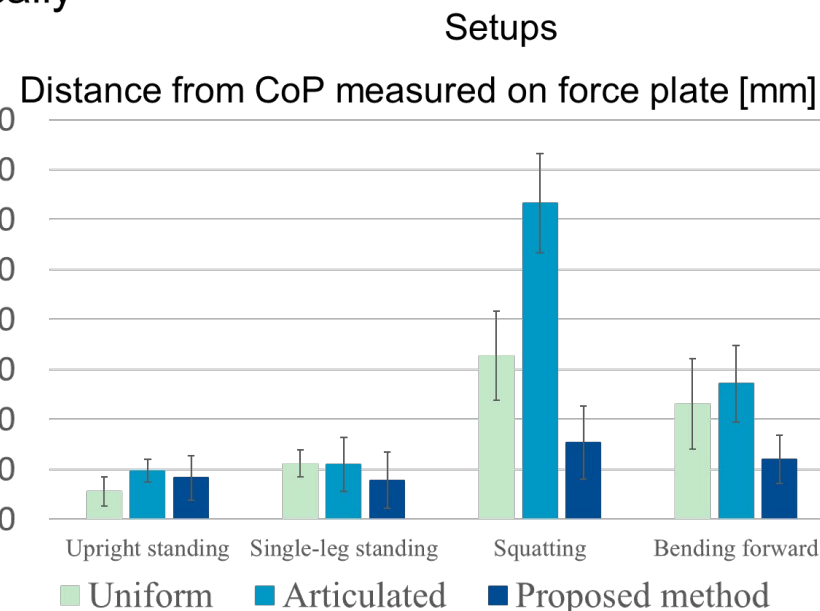
Baseline

Uniform: the CoM of the reconstructed 3D model which are assigned a uniform weight.

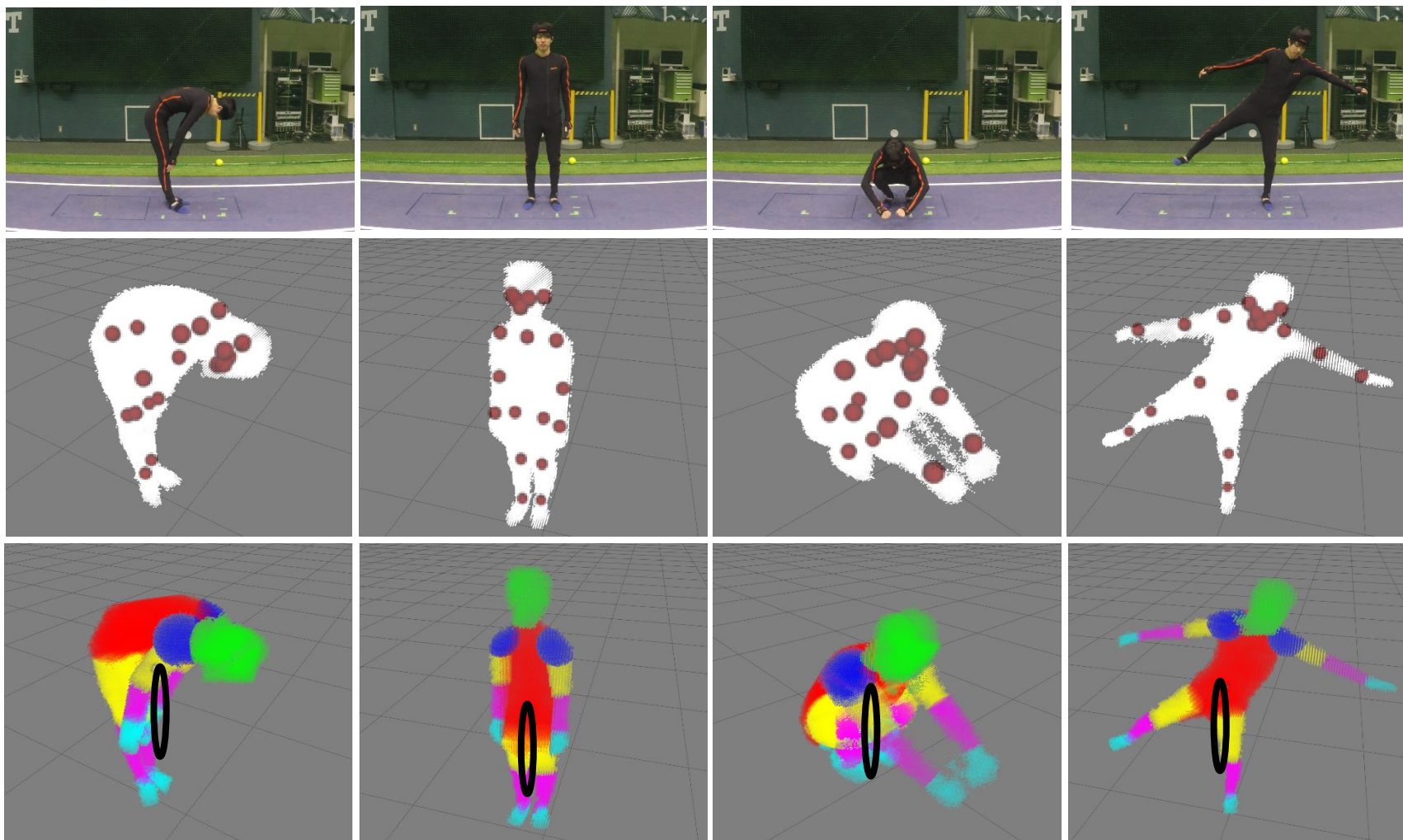
Articulated: the CoM as the center of the weighted articulated joint model.

Results

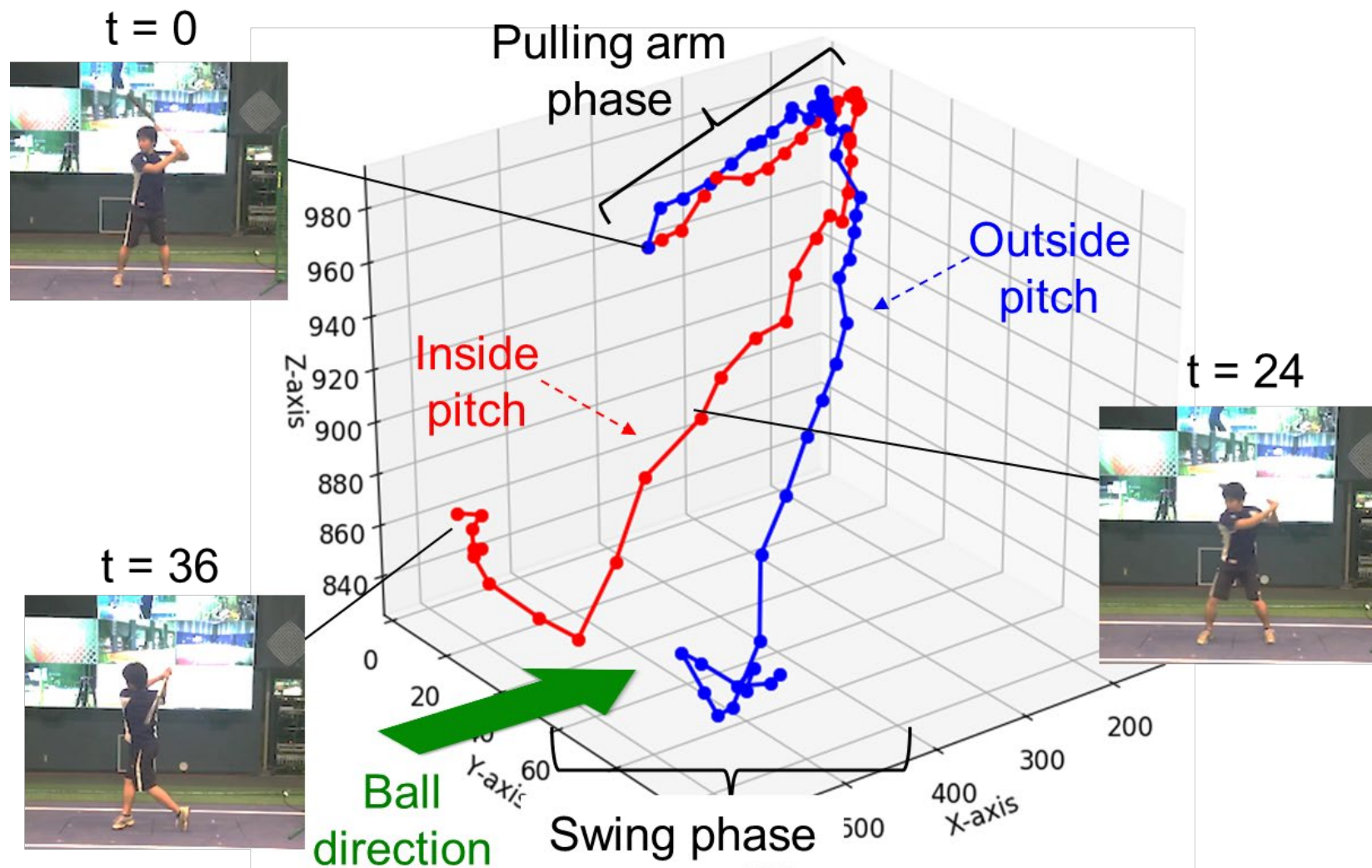
Our method robustly estimates the CoM with errors of around 10 mm in terms of CoP.

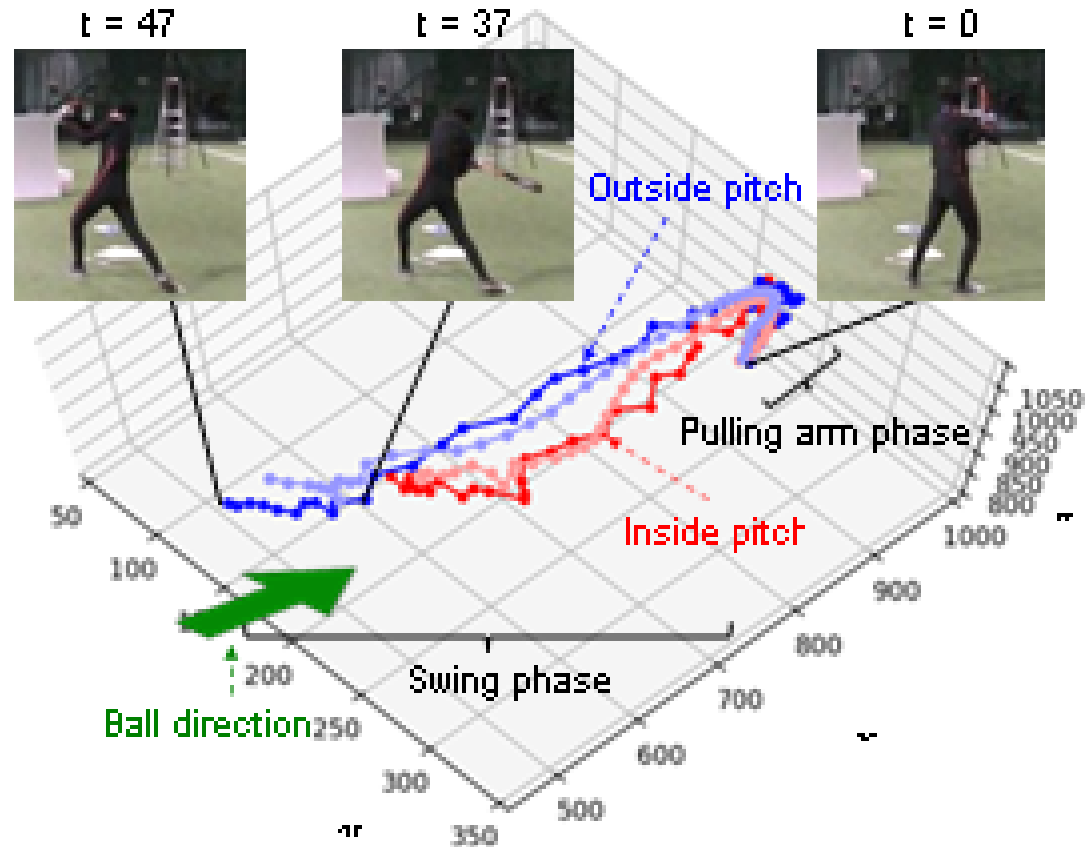
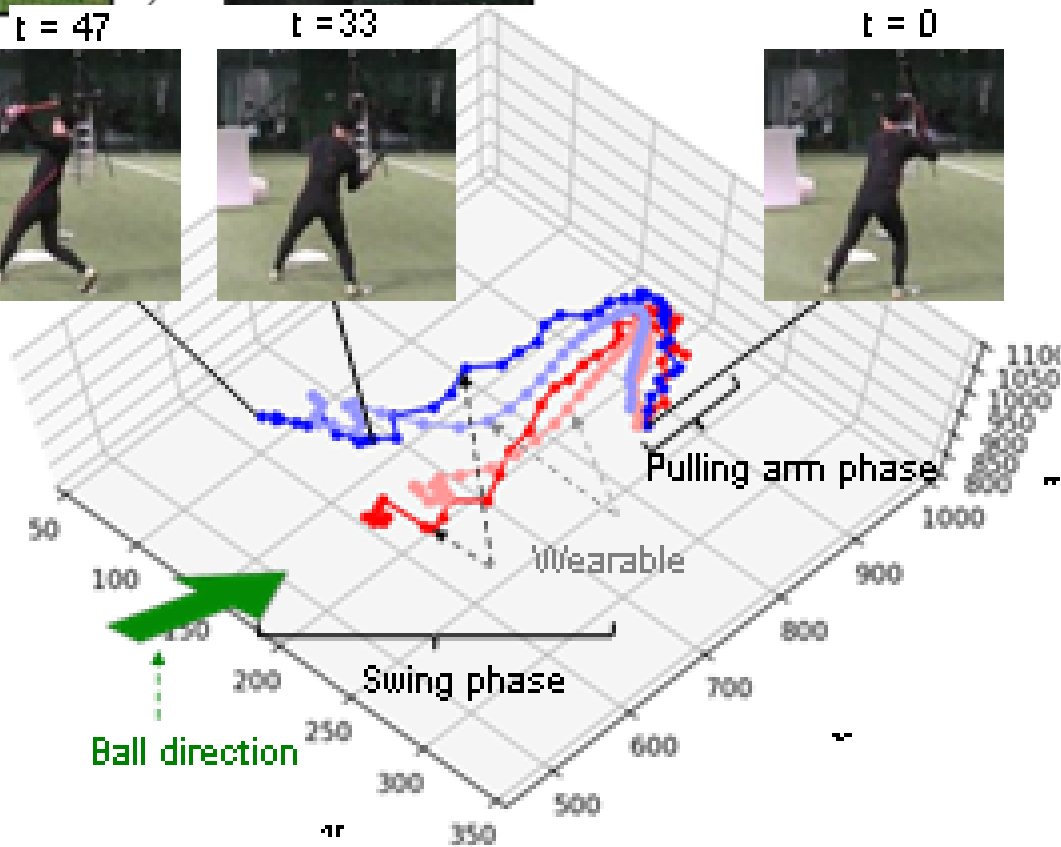
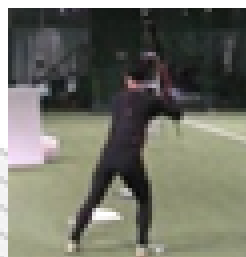
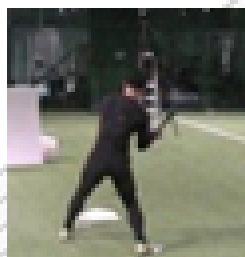
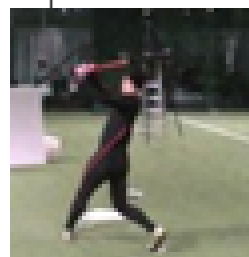
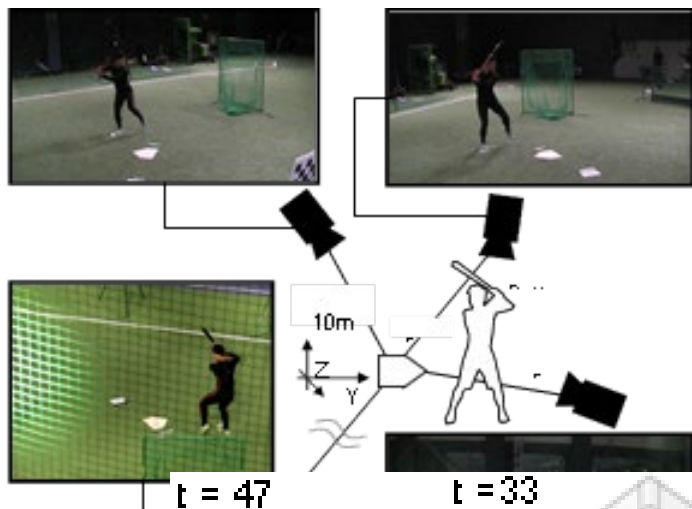


静止状態での重心位置推定結果



野球の打撃時の重心推移





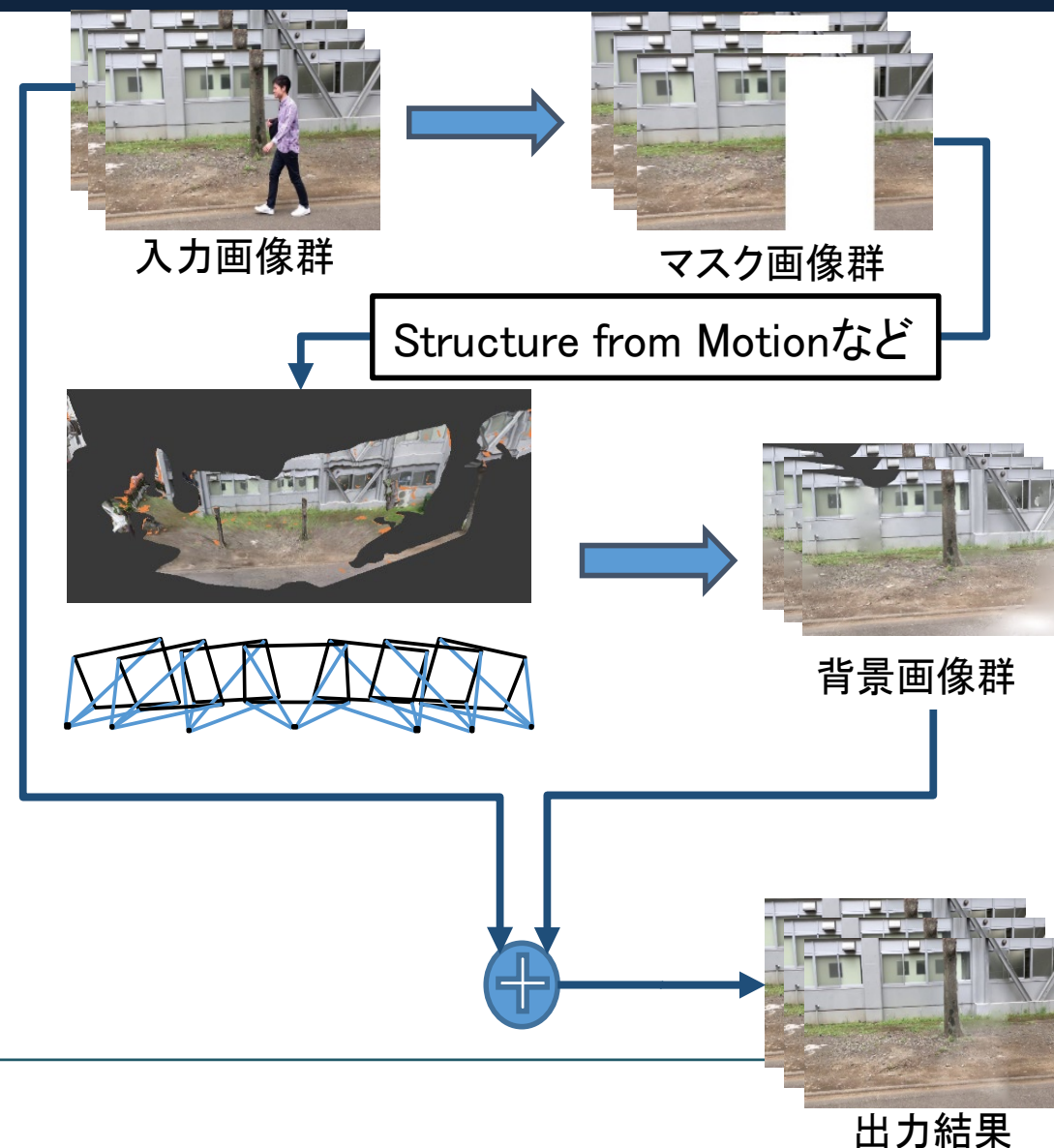
多視点カメラ：問題点と解決例

■ 台数の制約

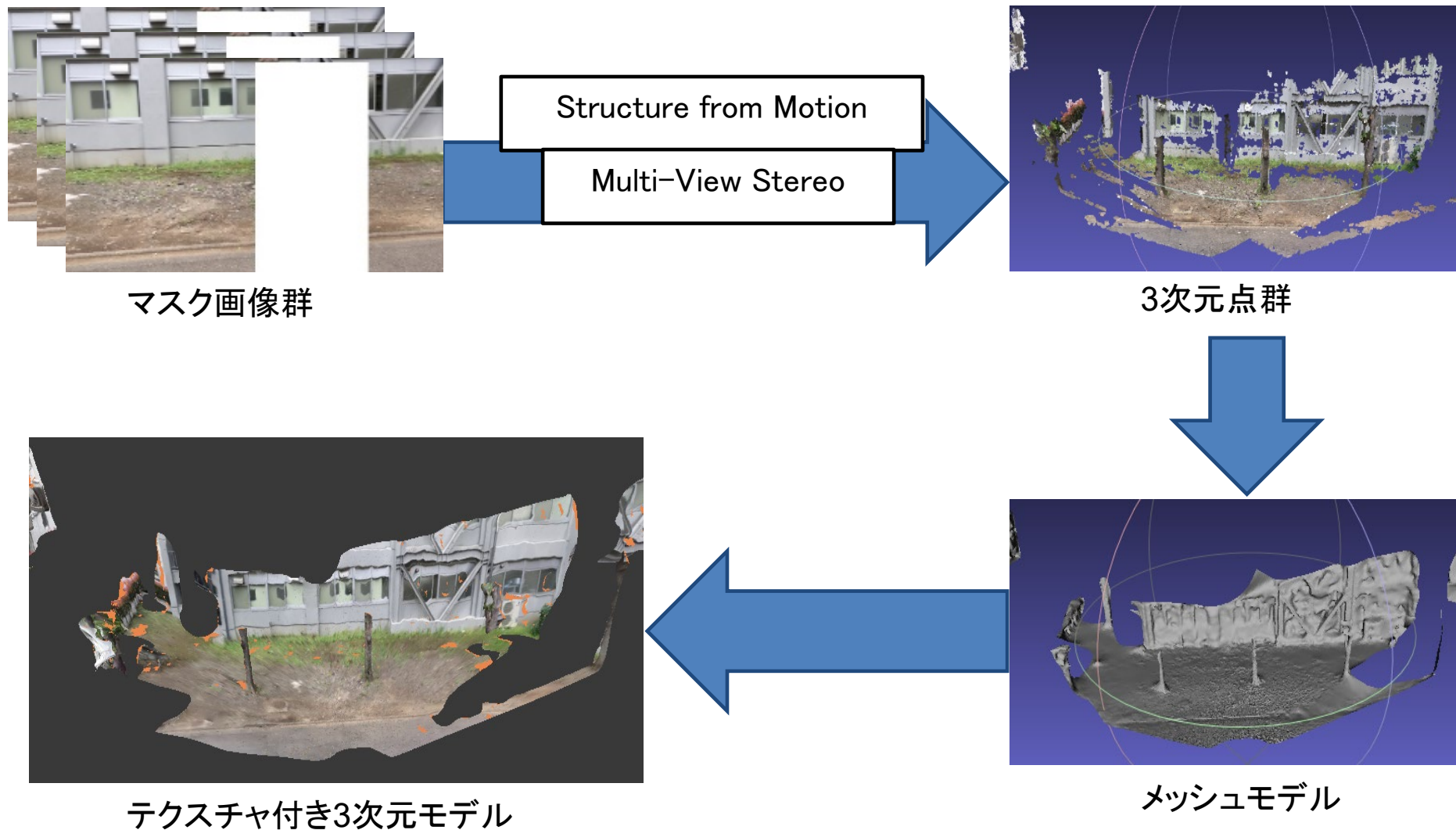
- 距離カメラの利用：Kinect
- [Open Pose](#), [Monocular Total Capture](#)

■ 移動範囲の制約：人は動き回る

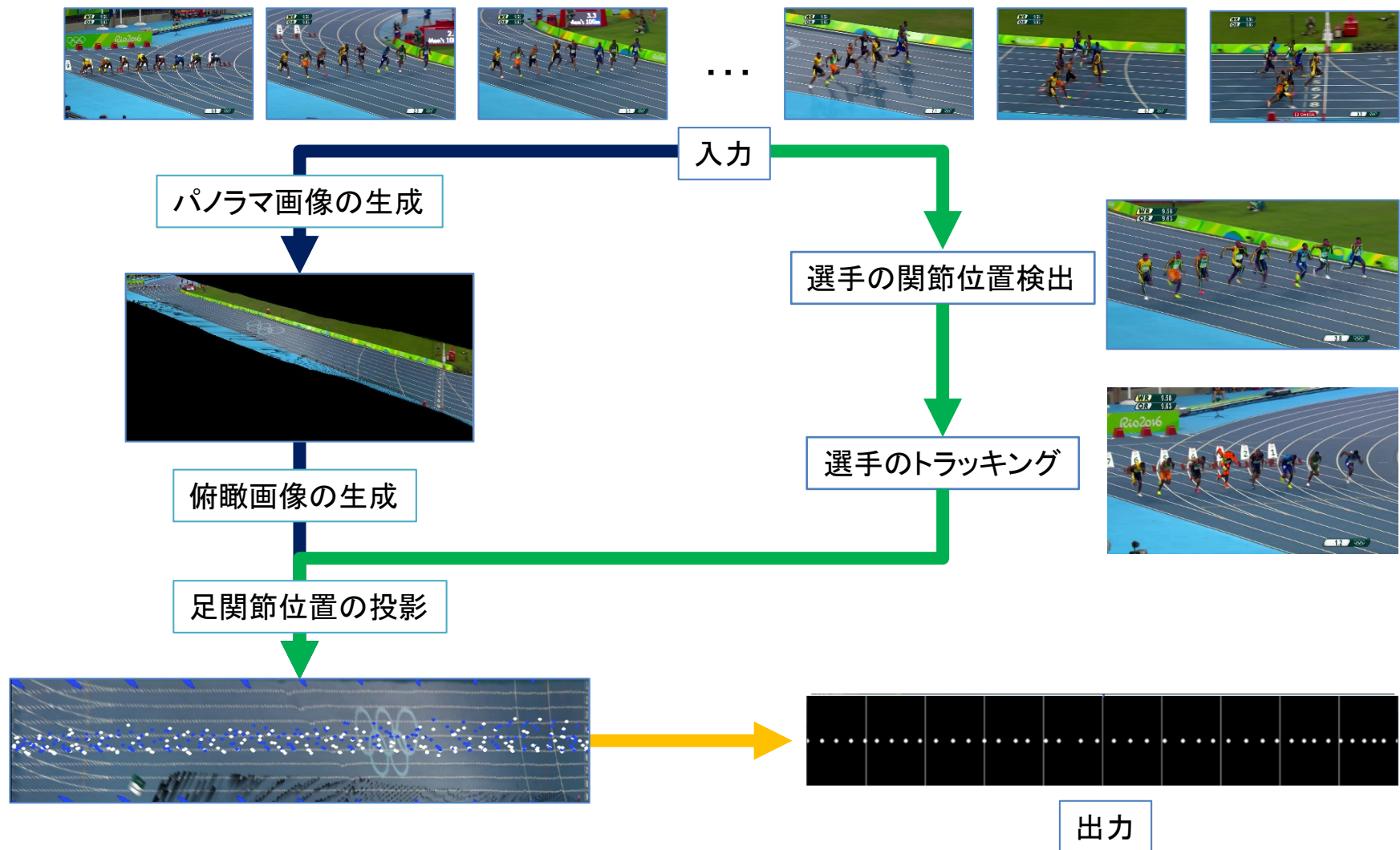
- 移動カメラで人の動きに合わせて撮影
→背景基準で空間位置を整合させる



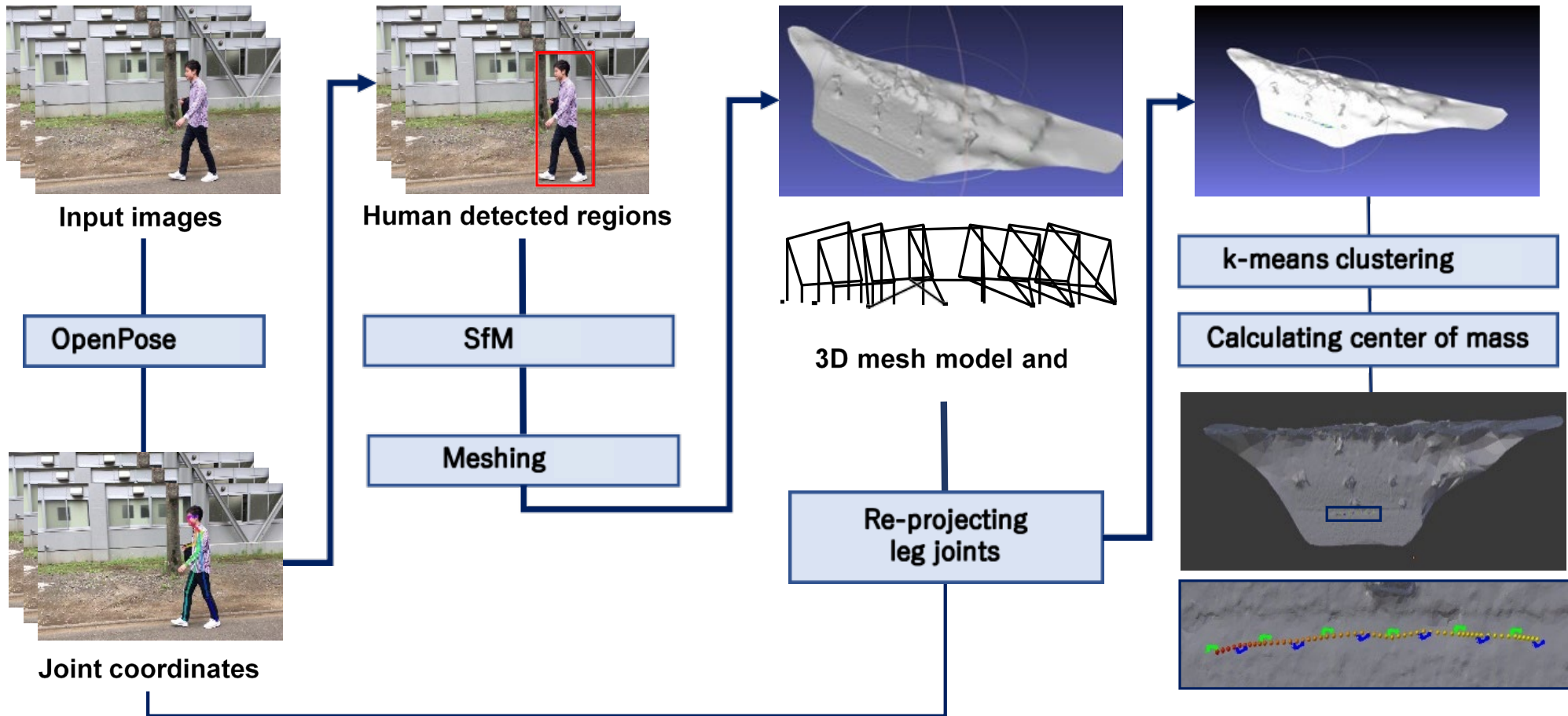
背景基準で空間位置を整合



カメラ映像を入力とした100m走選手の動作解析



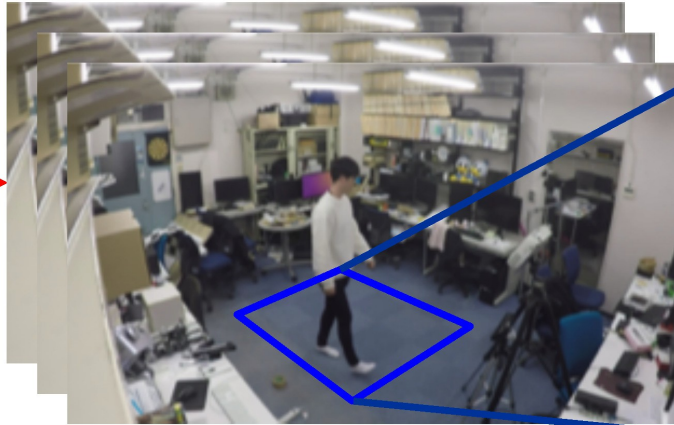
歩行時の接地位置計測手法



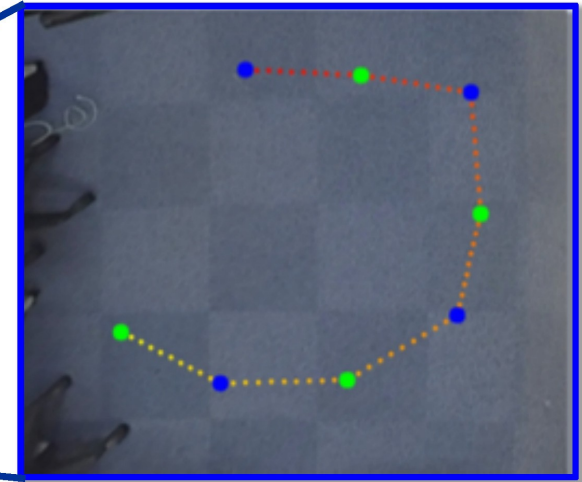
固定カメラによる歩行ステップ計測



Measurement environment



Input video



Result

- Left step
- Right step
- Trajectory

人が邪魔な場合もある



input



output

人が生活する環境のセンシングと認識も重要[Nakajima IROS18]

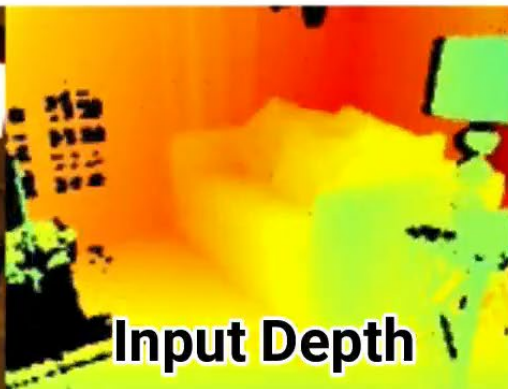
<http://www.youtube.com/watch?v=nYG8jqFBqWI>



Geometric Map



Semantic Map



環境側設置→人側に設置

- 万歩計（IMU1個）
- IMUを複数→詳細な動作センシング
- First Person View (FPV)



Pose Estimation from First Person View (FPV)

Indirect method

- 画角に人体が入らない

Direct method

- 広角レンズで人体を捉える

3D Ego-Pose Estimation via Imitation Learning

Paper ID 3017

- Hao Jiang et al. "Seeing invisible poses: Estimating 3D body pose from egocentric video." in CVPR. 2017.
- Ye Yuan et al. "3D ego-pose estimation via limitation learning." in ECCV. 2018.

Mo²Cap²: Real-time Mobile 3D Motion Capture with a Cap-mounted Fisheye Camera

Weipeng Xu¹ Avishek Chatterjee¹ Michael Zollhoefer^{1,2} Helge Rhodin³
Pascal Fua³ Hans-Peter Seidel¹ Christian Theobalt¹

¹Max Planck Institute for Informatics, Saarland Informatics Campus
²Stanford University ³EPFL



Stanford
University



- Weipeng Xu et al. "Mo2Cap2: Real-time Mobile 3D Motion Capture with a Cap-mounted Fisheye Camera." arXiv preprint arXiv:1803.05959 (2018).
- Helge Rhodin et al. "Egocap: egocentric marker-less motion capture with two fisheye cameras." ACM Trans. on Graphics (TOG) 35.6 162. 2016.

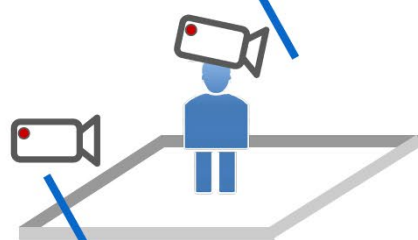
FPVとTPV(Third Person View)の連携

Actor and Observer: Joint Modeling of First and Third-Person Videos (INRIA)
github.com/gsig/actor-observer

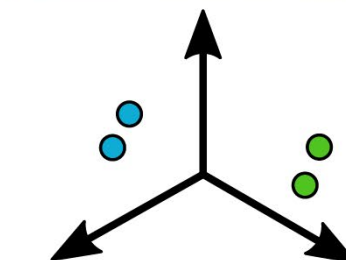
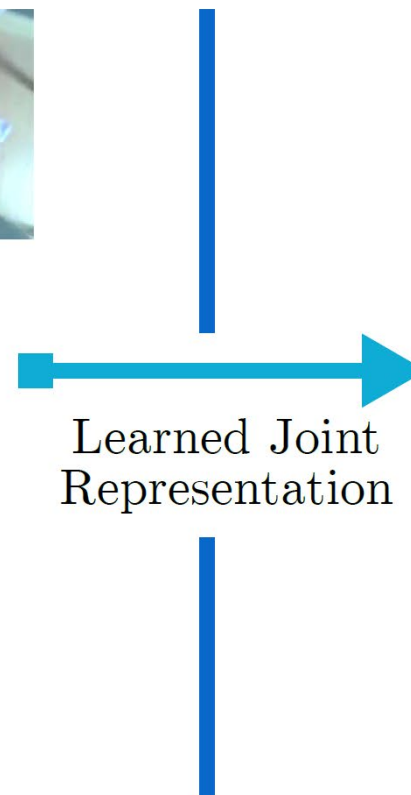
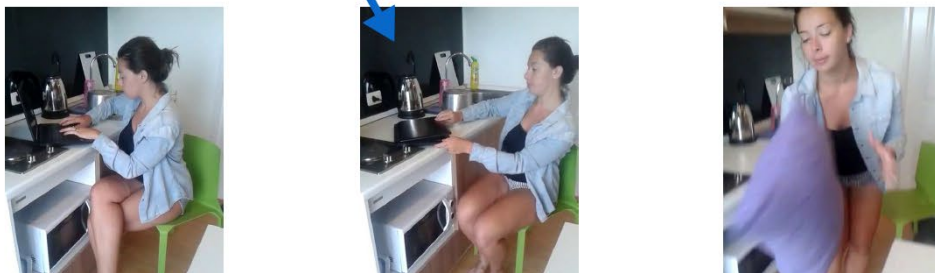
First
Person



*"Person is typing on a laptop.
Then they put down the laptop
and pick up a pillow."*



Third
Person



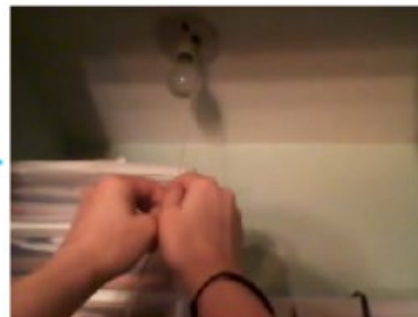
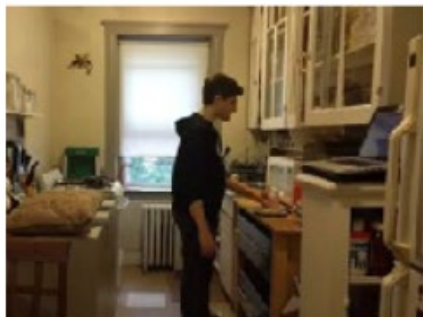
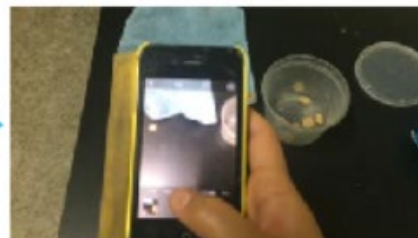
"Cleaning Dishes"



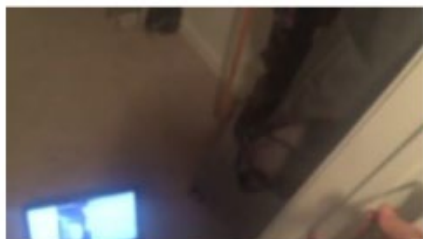
Third-Person Frame



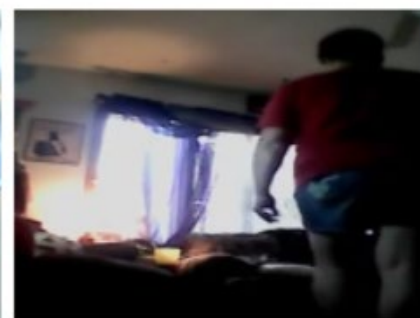
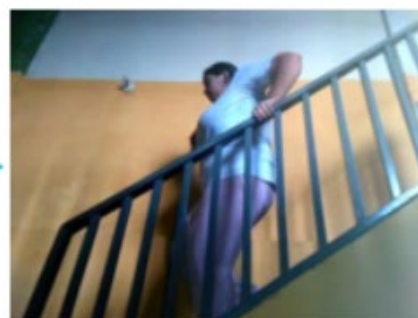
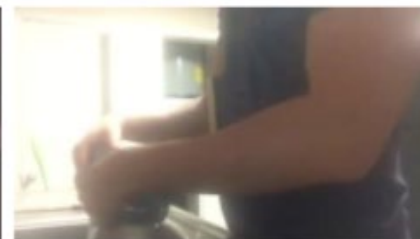
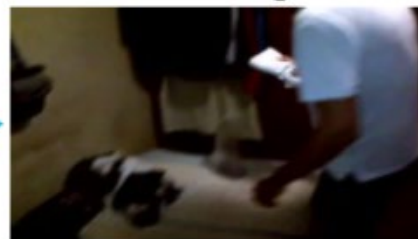
Nearest Neighbors in First-Person



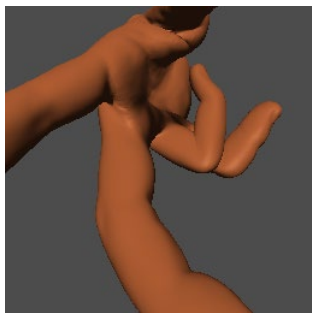
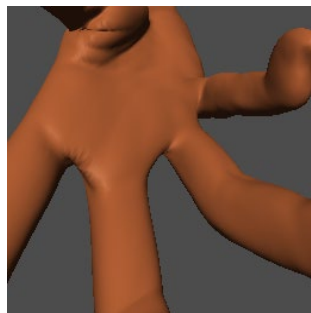
First-Person Frame



Nearest Neighbors in Third-Person



人が扱う物体に設置 [Ino 3D BodyTech18]

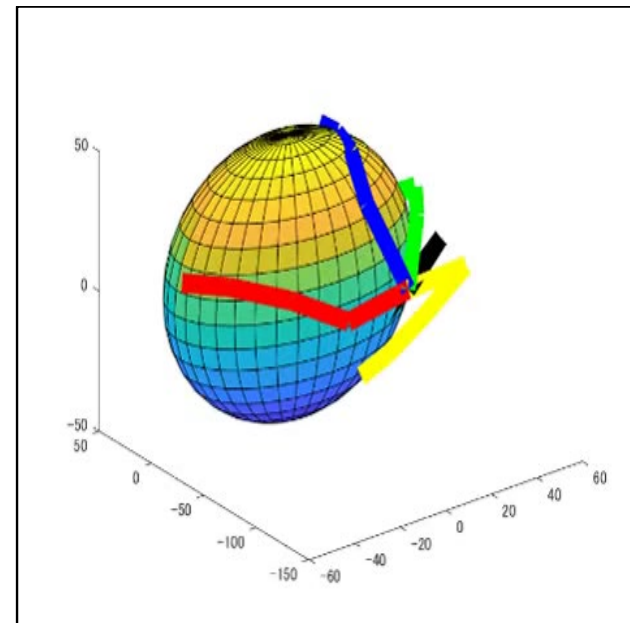
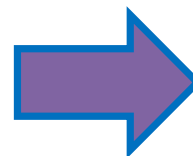


学習データ



入力

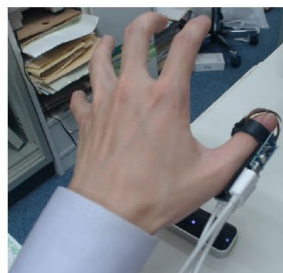
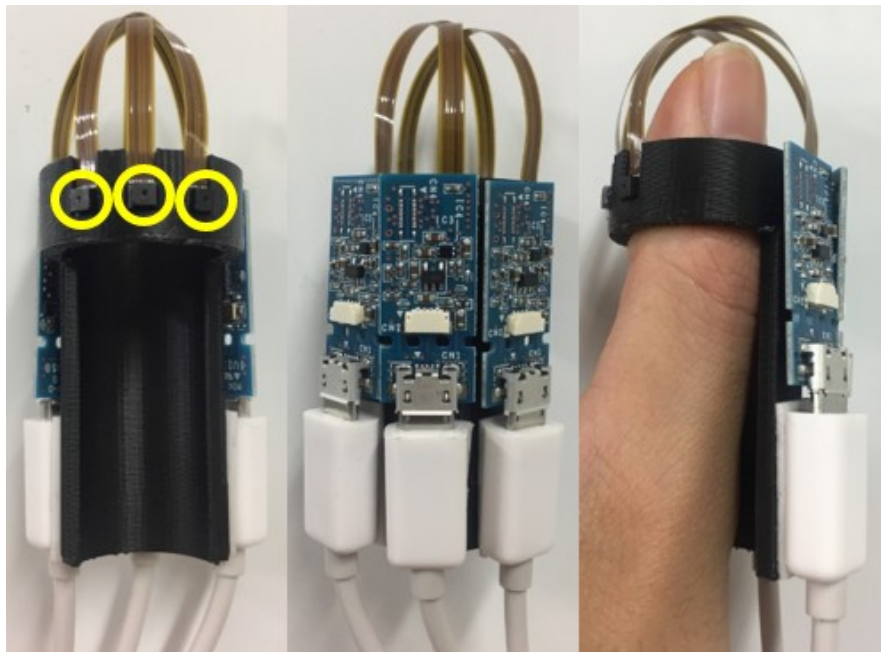
姿勢推定



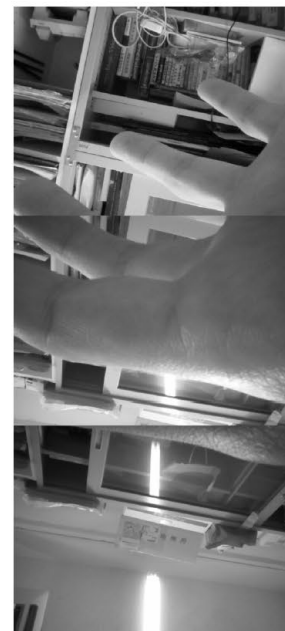
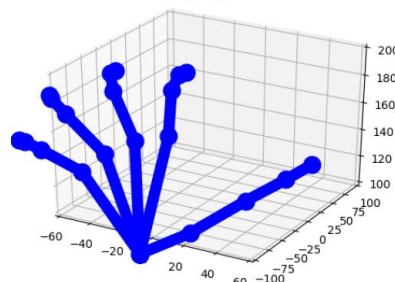
出力

- 物体を持った時の手の姿勢推定は困難
⇒ 握り方の制約が大きい
- 機械学習を用いることによって握り方の制約なく、姿勢推定が可能

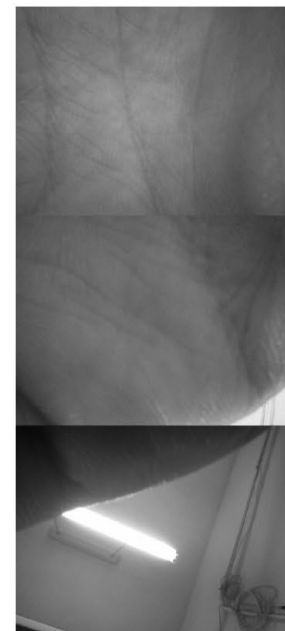
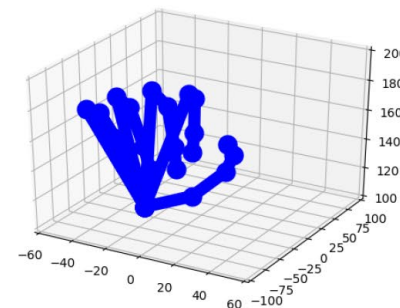
指にカメラを設置[Lenaga ACCV-WS18]



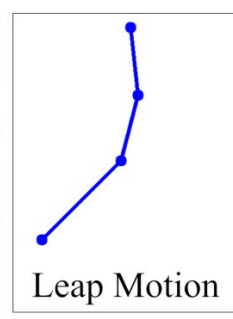
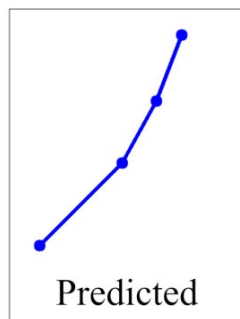
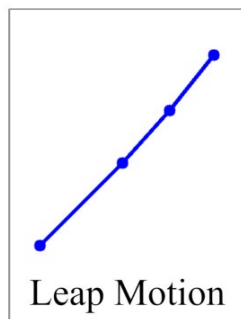
(a)



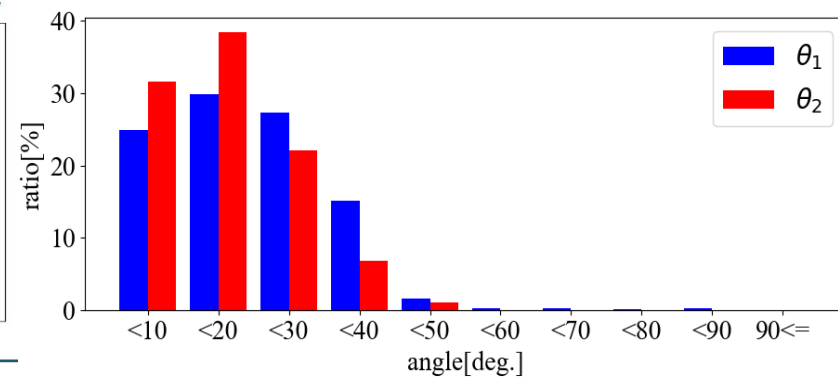
(a')



scene 1



scene 2



まとめと今後の展望

- カメラ映像に含まれる多くの潜在的情報から多様な人間挙動データを
得ることができる
 - カメラの小型化・大量化が「常時計測」を可能に
 - 大量の学習データの利用
 - 可視光以外の物理量を計測できるカメラもどんどん登場するはず
 - 人だけではなく、環境をセンシング・認識することも、人の挙動の認
識のために重要.
-
- 現状は可能性の実証レベル.
要求精度の達成のために、実証フィールドと密な連携が鍵
-