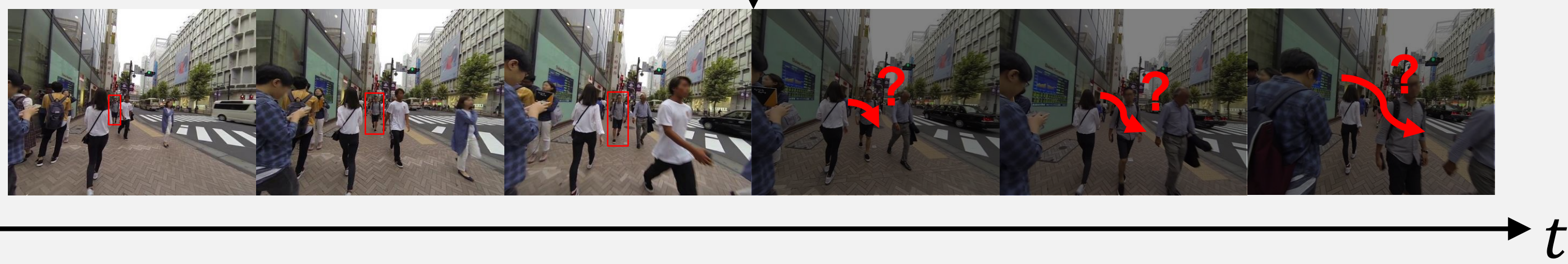


八木 拓真<sup>1</sup> マンガラム カーティケヤ<sup>2</sup> 米谷 竜<sup>1</sup> 佐藤 洋一<sup>1</sup><sup>1</sup>東京大学 <sup>2</sup>インド工科大学カンプール校

## 研究目的

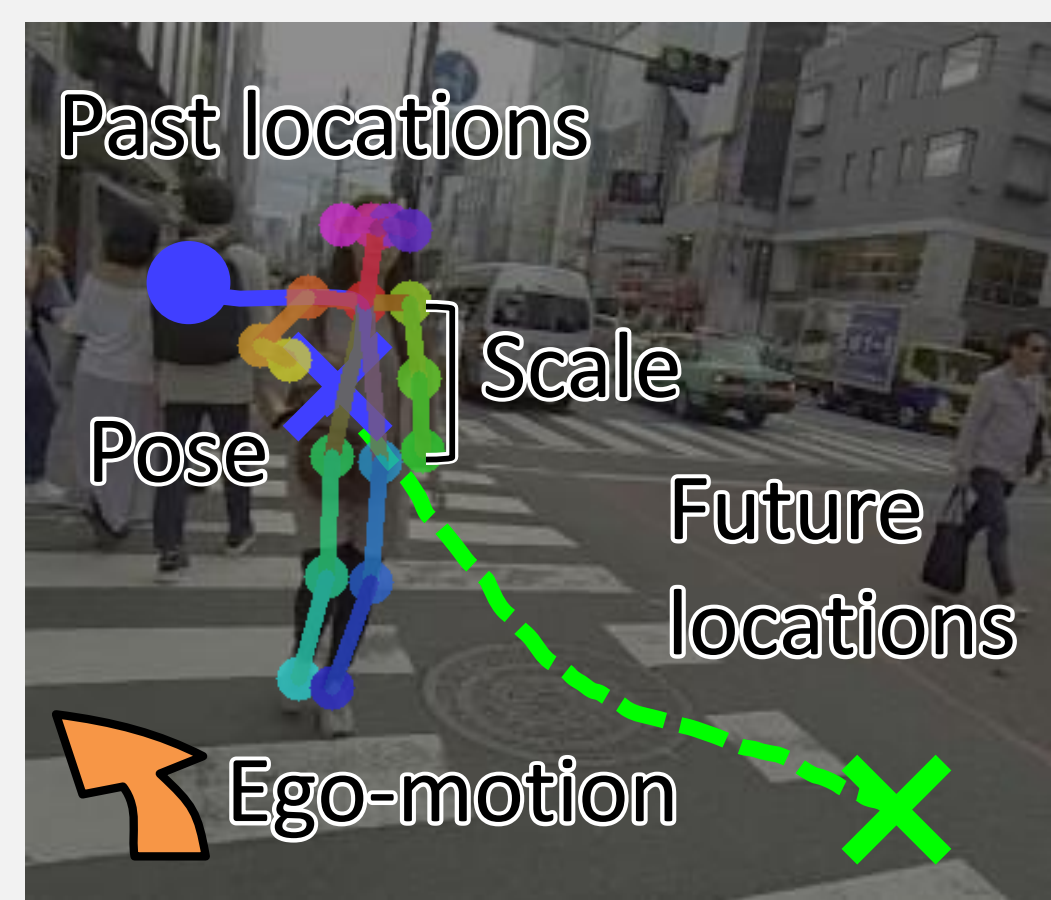
- 一人称視点映像中に映る人物の位置予測

入力                  現在                  出力



## 本研究のアプローチ

- 一人称視点映像に特有の手掛かりを位置予測に利用
- Tri-Stream CNNにより映像中の人物が近い将来出現する位置の系列を予測



## スケール

対象人物までの距離を近似

## 姿勢

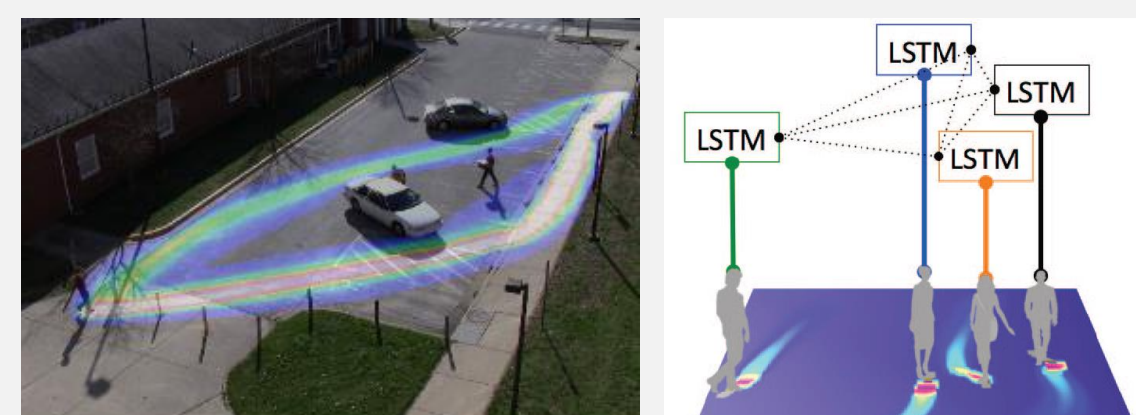
対象人物の頭部/身体方向を捉える

## 自己運動

装着者自身の運動をエンコード

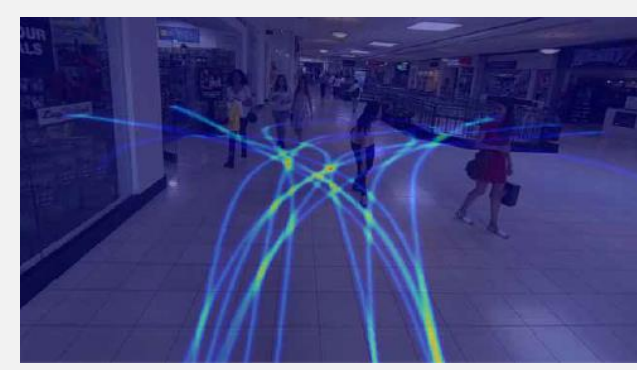
## 固定視点映像における行動予測

- Activity forecasting [Kitani+, ECCV'12]
- Social LSTM [Alahi+, CVPR'16]
  - 人物間の相互作用をモデル化



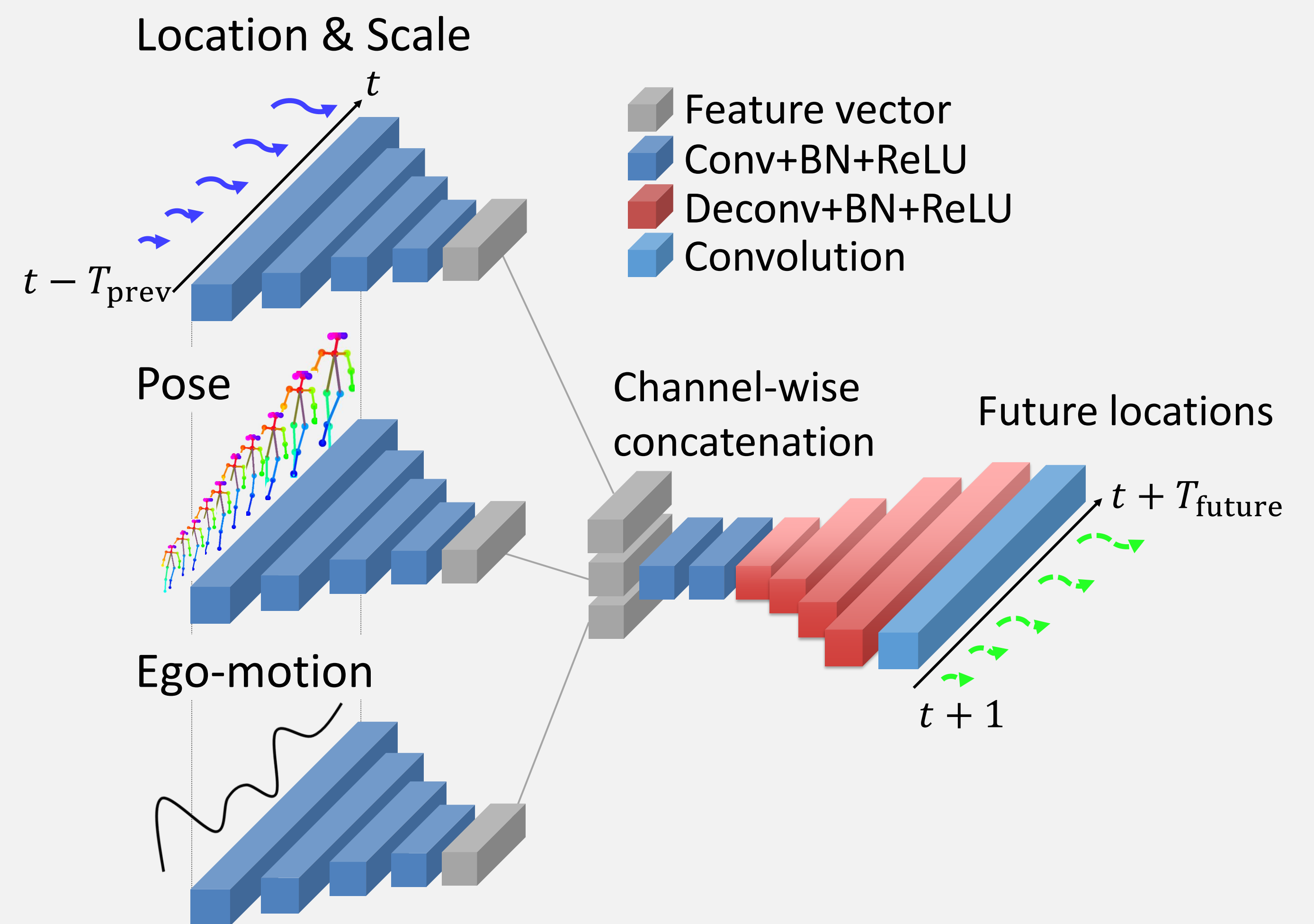
## 一人称視点映像における装着者の行動予測

- Egocentric future localization [Park+, CVPR'16]
  - カメラ装着者の将来位置予測
- Indoor activity forecasting [Rhinehart+, CVPR'17]
  - 屋内に特化した装着者の行動予測



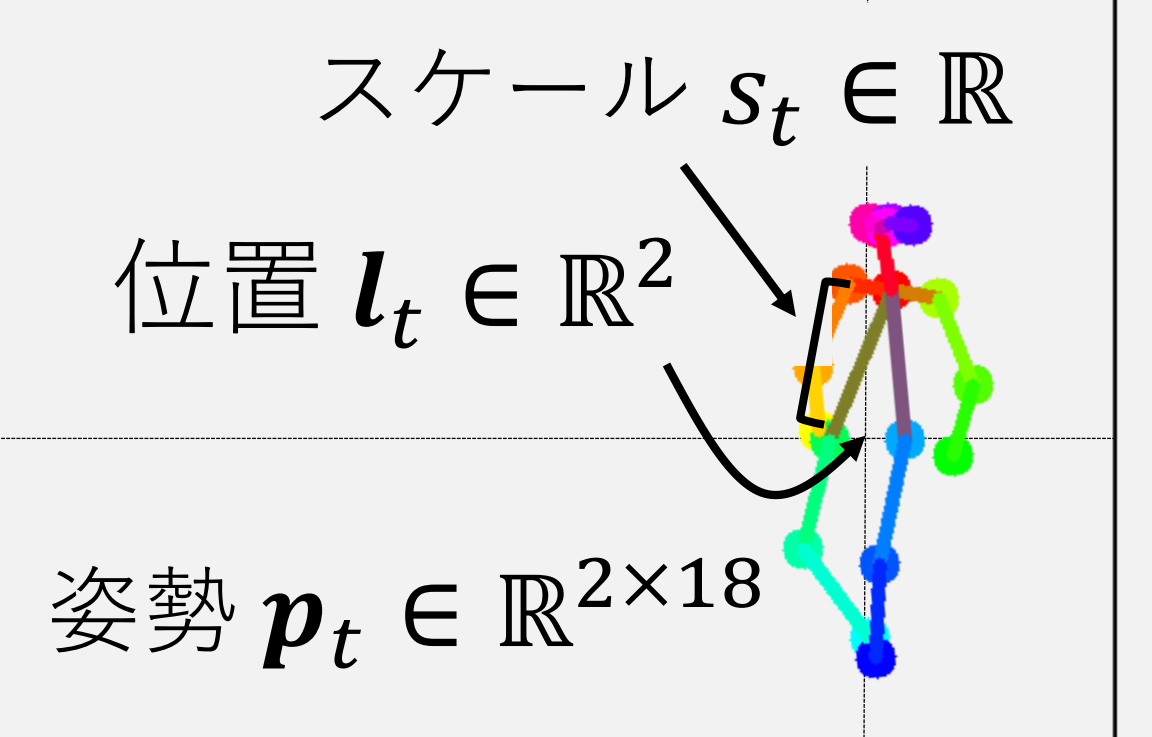
## 提案手法

- Tri-Stream 1D-CNN (時間方向への畳み込み)
- 各モダリティについて特徴量の時系列を入力
- 異なるモダリティ間の高次の相関を捉える



## スケール特徴

- 2次元位置 (腰の中心)
- 1次元スケール (首から腰の中心)



## 姿勢特徴

- 2次元身体部位位置 × 18点
  - OpenPose [Cao+, CVPR'17] を利用
  - 欠損値をk-NN法にて補完

## 自己運動特徴

- フレーム  $t$  から  $t + \Delta t$  にかけての6自由度大域運動
  - 3次元並進 + 3次元回転
  - SfMLearner [Zhou+, CVPR'17] を使用

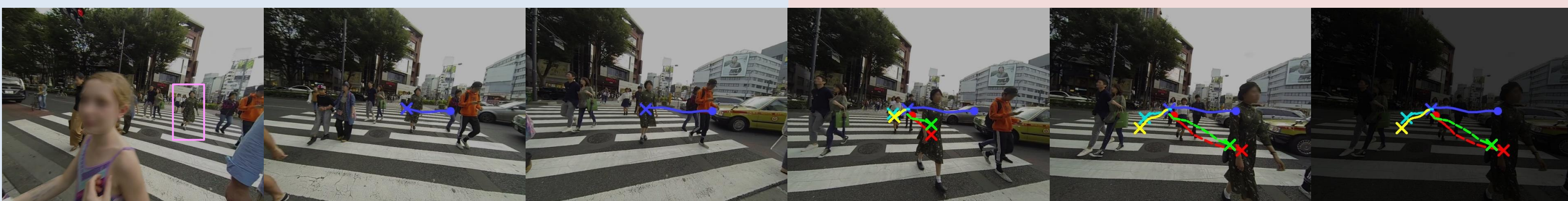
## 実験結果

## 予測例

— 入力    - - - 正解    - · - · - NNeighbor    - · - · - Social LSTM    - · - · - 提案手法

-0.9s      -0.5s      現在      +0.6s      +1.0s      (予測強調)

Toward: 対象人物が装着者と相対する



Away: 対象人物と装着者が同方向に移動する



Across: 対象人物が装着者の前を横切る

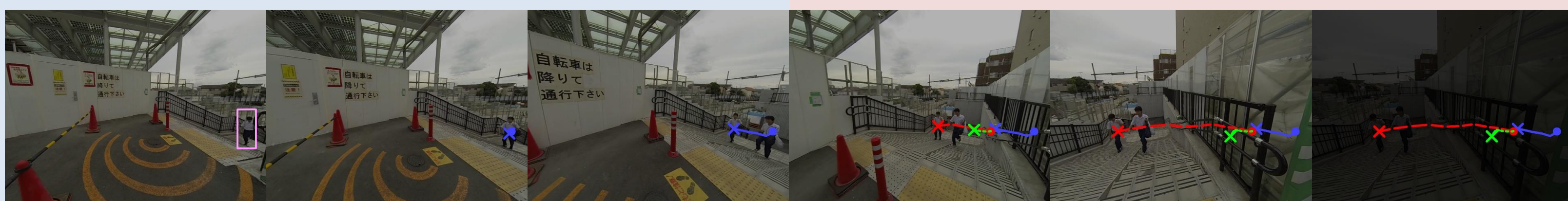


## 失敗例

シーンコンテキストの欠如 (e.g. 障害物、歩行不能箇所)



装着者及び歩行者の急な方向転換



## First-Person Locomotionデータセット

- 胸部に装着したGoProカメラを使用
- 4.5時間、計5,000人分の歩行映像を収集
- 人物追跡により2秒間のサンプルを50,000本生成



## 実験条件

- 入力、出力は各1秒間 (10フレーム)
- 評価: 最終予測誤差 (FDE, 予測-正解間のL2距離)

## 1秒後の予測誤差の比較

- 姿勢及び自己運動特徴が性能向上に相補的に寄与

手法	歩行方向			
	Toward	Away	Across	Average
(単位: 画像幅に対する%)				
位置のみ (L)	11.50	6.32	8.19	6.89
位置 + スケール (L+S)	9.89	6.18	8.05	6.40
L + S + 姿勢	8.85	6.14	7.84	6.29
L + S + 自己運動	9.54	5.99	7.76	6.18
Nearest Neighbor	12.95	7.02	9.67	7.69
Social LSTM [Alahi+, CVPR'16]	13.52	8.69	11.63	9.23
提案手法	<b>8.52</b>	<b>5.90</b>	<b>7.27</b>	<b>6.04</b>

## 論文・ソースコード

- <https://artilects.net/projects.html>