

# 歩行者の会話の活発さ認識の軽量化における骨格時空間グラフの有効性検証

近藤翼 井上路子 米田駿介 西山正志 (鳥取大学)



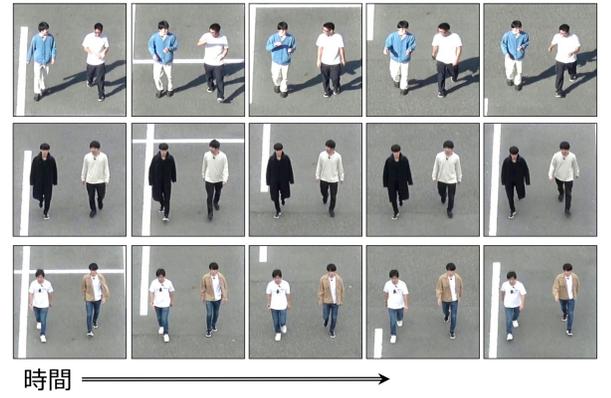
## 目的

動画を用いた会話の活発さ認識において  
計算時間やGPUメモリ量が限られた場面での  
使用を想定し骨格時空間グラフを用いる手法を検証

会話の活発さ認識の関連研究

- ・ジェスチャは会話分析に重要 [McNeill, 1994]
- ・歩行者の会話の活発さ認識にジェスチャを利用 [Ganaha+, ICPR'24]

- 活発**  
大きなジェスチャが  
高頻度で発生する
- 非活発**  
小さなジェスチャが  
たまに発生する
- 会話なし**  
ジェスチャが  
ほぼ発生しない

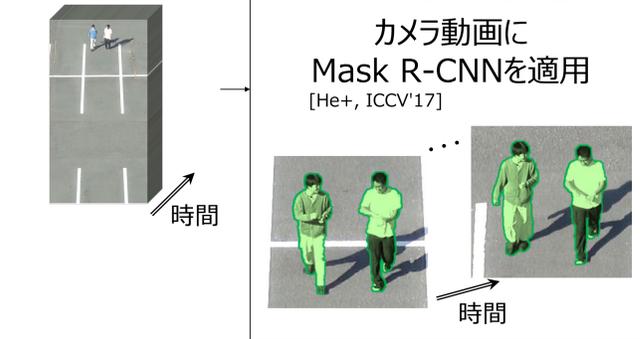


## 既存手法

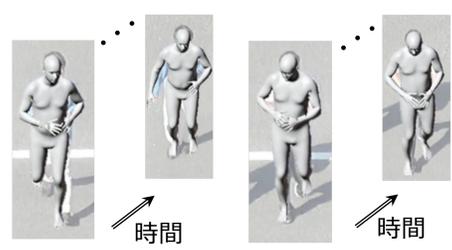
ジェスチャをインタラクション動画で  
表現し特徴量として利用 [Ganaha+, ICPR'24]

入力：カメラ動画

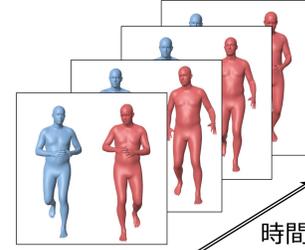
**P1: 人物領域推定**  
カメラ動画に  
Mask R-CNNを適用  
[He+, ICCV'17]



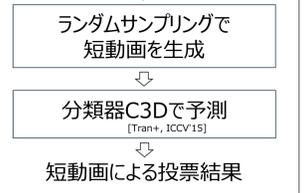
**P2: SMPL推定**  
人物領域の見え方に  
PHALPを適用  
[Rajasegaran+, CVPR'22]



**P3: インタラクション動画生成**  
SMPLモデルの  
3Dレンダリングを適用



**P4: クラス分類**  
インタラクション動画



出力：会話の活発さラベル



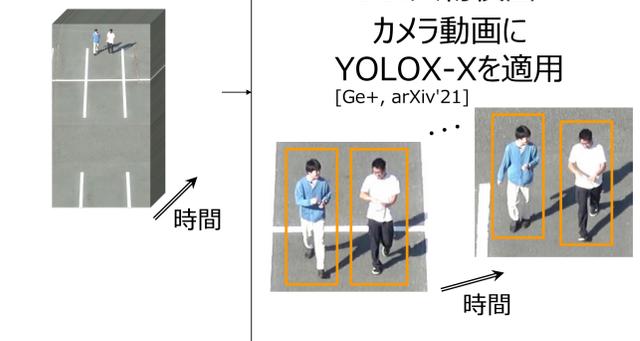
**X課題点** 計算時間とGPUメモリ量が増大

## 提案手法

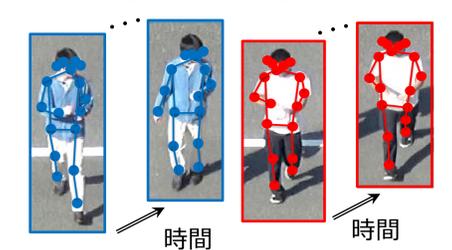
ジェスチャを骨格時空間グラフで  
表現し特徴量として利用

入力：カメラ動画

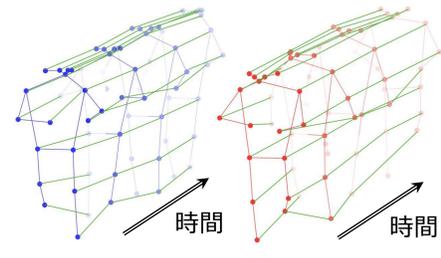
**P1: 人物検出**  
カメラ動画に  
YOLOX-Xを適用  
[Ge+, arXiv'21]



**P2: 骨格推定**  
人物の矩形領域に  
ViTposeを適用  
[Xu+, NeurIPS'22]



**P3: 時空間グラフ設定**  
各時刻の骨格を  
時間方向に連結



**P4: クラス分類**  
時空間グラフ



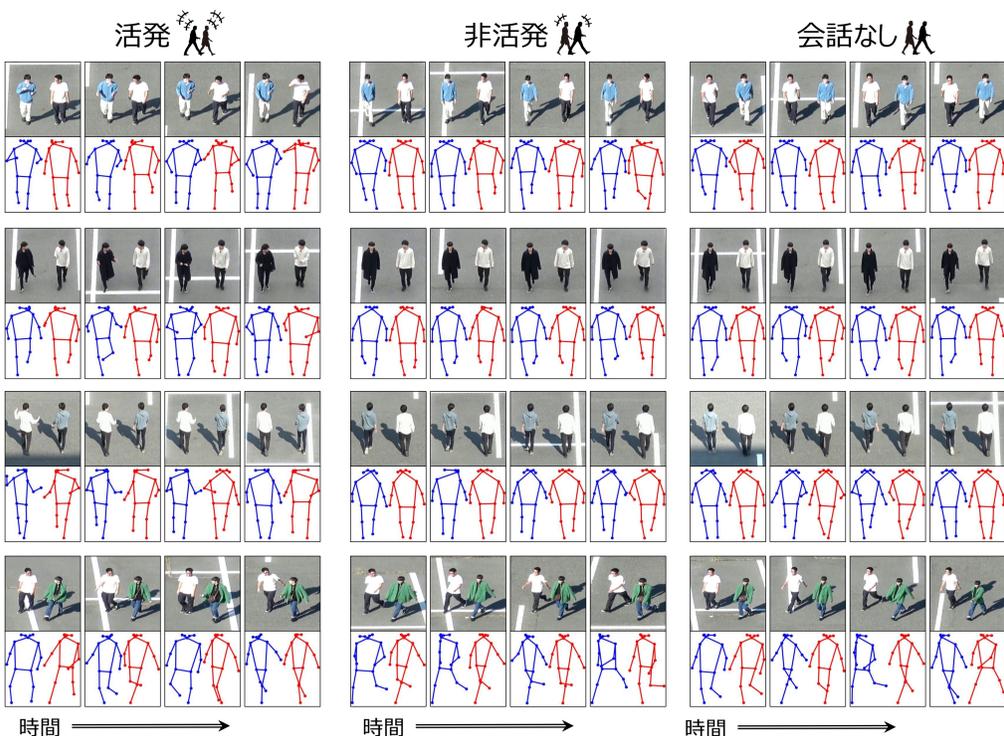
出力：会話の活発さラベル



**✓利点** 計算時間とGPUメモリ量を削減

## 実験

評価データセット ● グループ数：52組 ● 動画数：624個



検証内容 既存手法と提案手法の計算時間・GPUメモリ量・認識率を比較

既存手法 認識率：76.2±0.7%

手続き番号	P1	P2	P3	P4	合計
手続き名	人物領域推定	SMPL推定	インタラクション動画生成	クラス分類	
計算時間	0.249秒/枚	0.326秒/枚	0.542秒/枚	0.007秒/短動画	1.124秒
手続き番号	P1	P2	P3	P4	合計
手続き名	人物領域推定	SMPL推定	インタラクション動画生成	クラス分類	
GPUメモリ量	1020MiB	1220MiB	1227MiB	1690MiB	5157MiB

提案手法 認識率：80.3±0.8%

手続き番号	P1	P2	P3	P4	合計
手続き名	人物検出	骨格推定	時空間グラフ設定	クラス分類	
計算時間	0.085秒/枚	0.021秒/枚	0秒/枚	0.001秒/短グラフ	0.107秒
手続き番号	P1	P2	P3	P4	合計
手続き名	人物検出	骨格推定	時空間グラフ設定	クラス分類	
GPUメモリ量	978MiB	1002MiB	0MiB	1194MiB	3174MiB

認識精度を維持したまま  
計算時間10.5倍高速化しGPUメモリ量3.8割削減