

歩行者間の会話の盛衰認識における複数グループ対応の検討



尾崎匠 我那覇航 井上路子 西山正志 (鳥取大学)

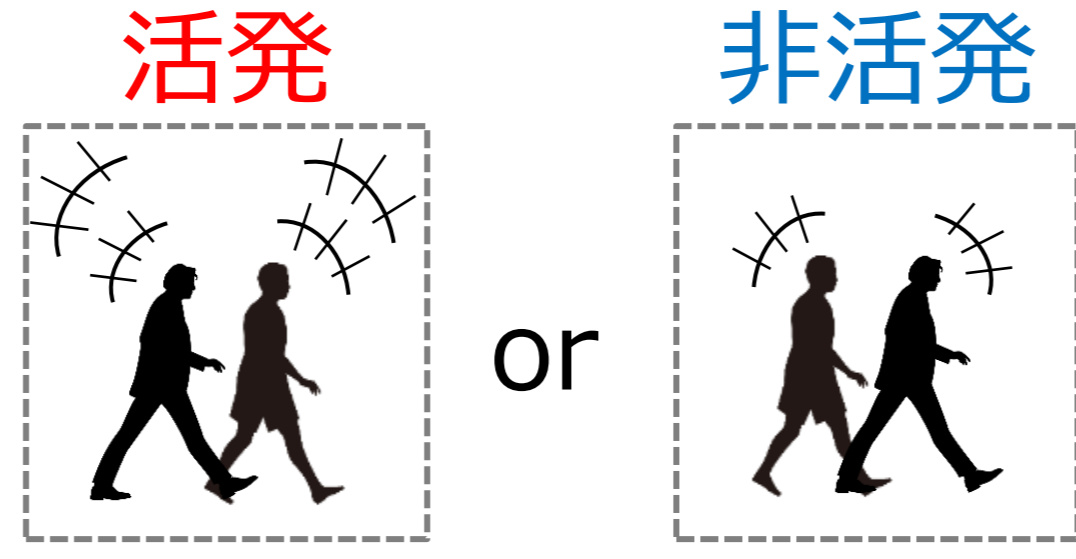
目的

屋外で歩行中の人物グループの動画のみから、認識精度が高く人間が目視でも確認できる**インタラクション動画**を特徴量として抽出し会話の盛衰を認識する手法を提案



会話の盛衰とは？

人物グループ内で発生している会話が活発か非活発か



インタラクション動画とは？

人物グループ内のインタラクションを抽出した3Dモデル動画

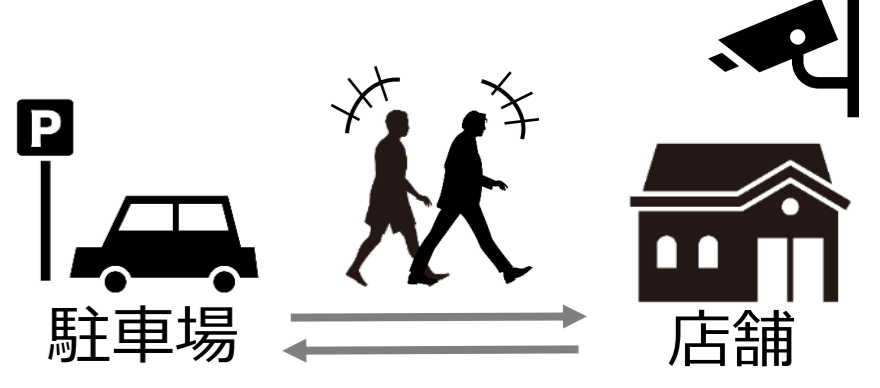
- ・身振り手振り：会話分析に利用 [McNeill, D., 1994]
- ・互いの体の向き・互いの立ち位置：ソーシャルグループ検出に利用 [Chamveha et al., 2013] [Zanlungo et al., 2014]

既存手法 [Han et al., CVPR2022]

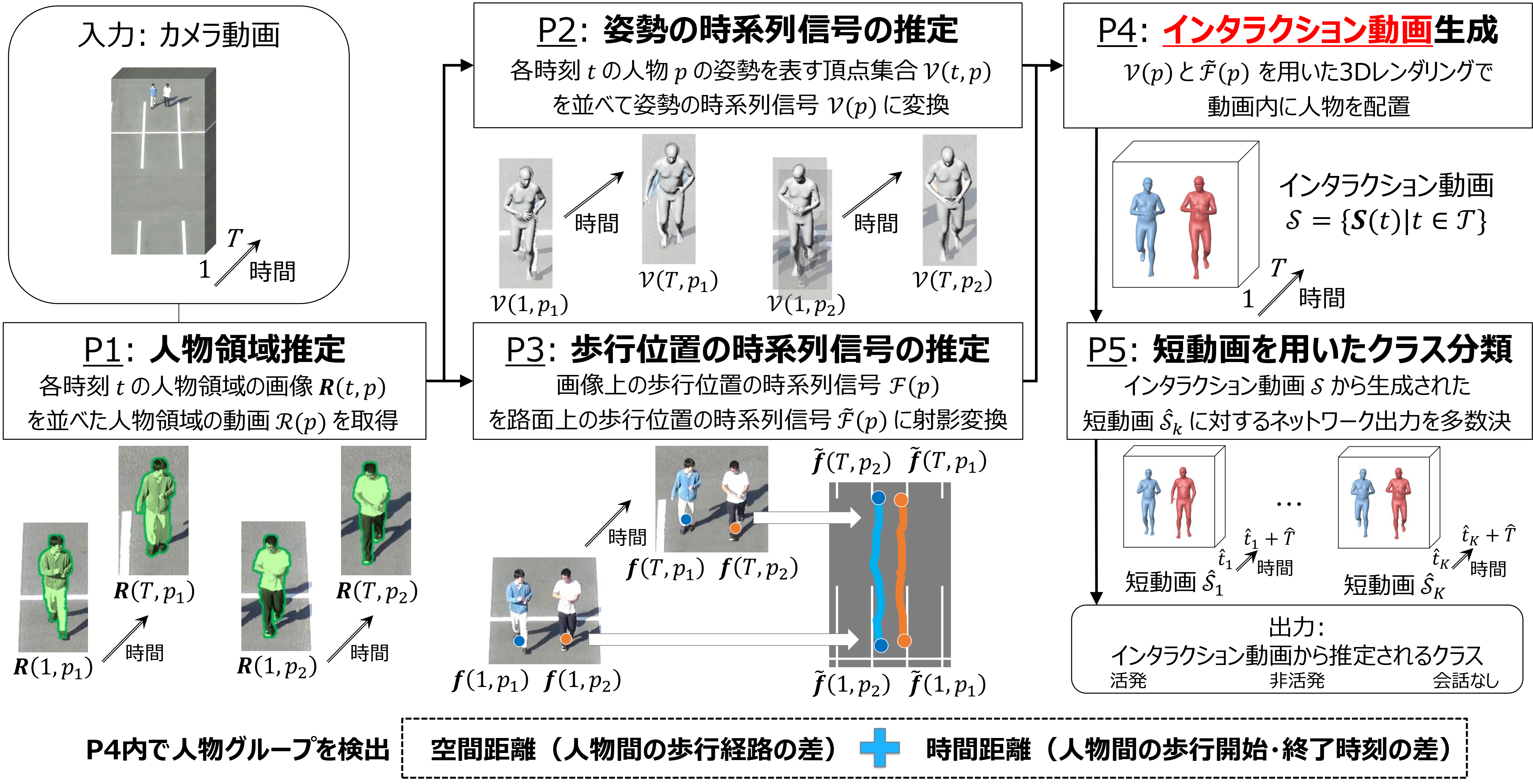
動画から歩行中の人物グループを検出し会話の有無を認識
 ✗ 会話の盛衰を認識する手法は存在しない

応用シーン (大型ショッピング店舗)

会話が活発なグループをカウントし行きと帰りで比較 → 来客者の満足度を把握

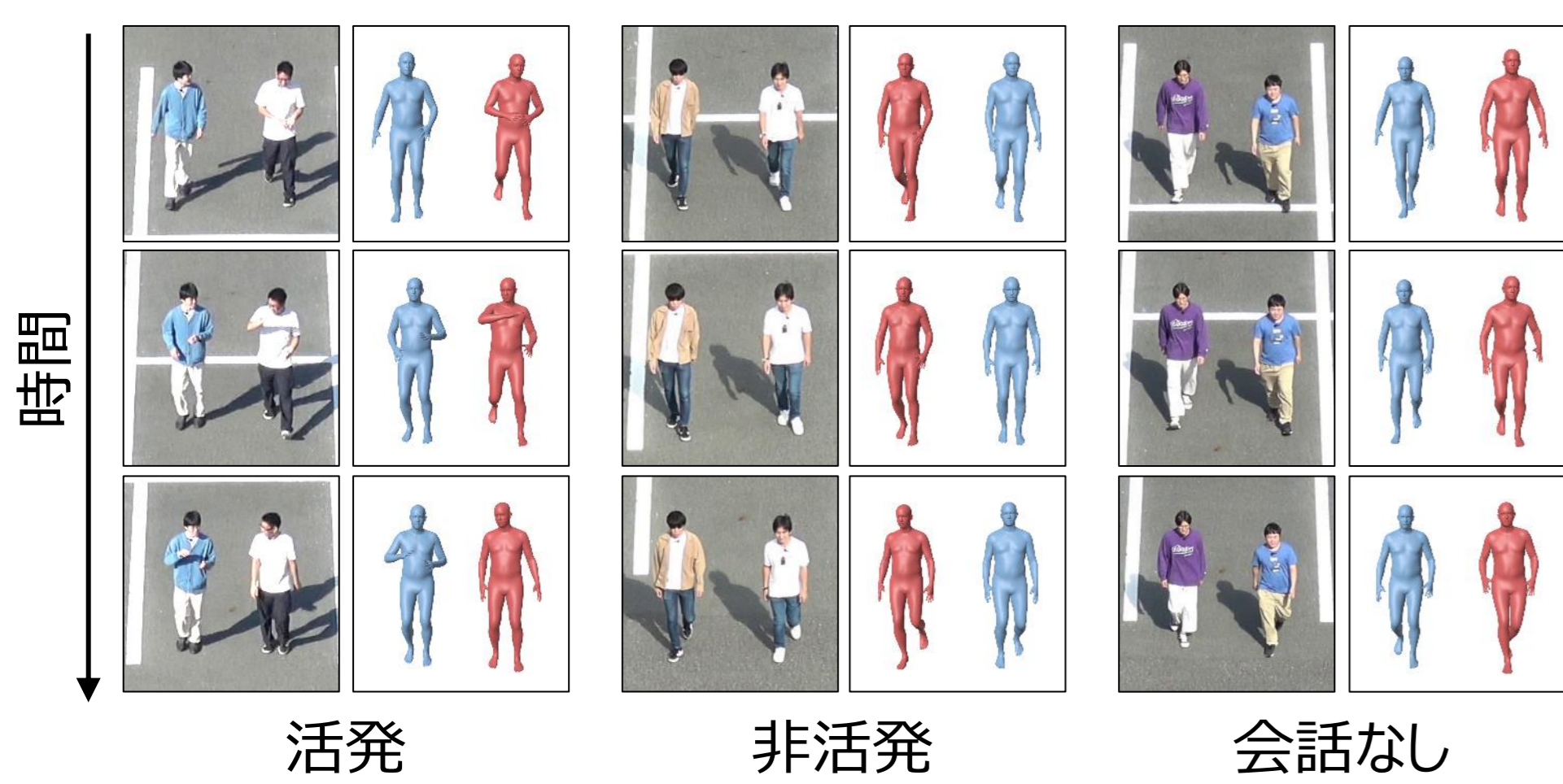


提案手法

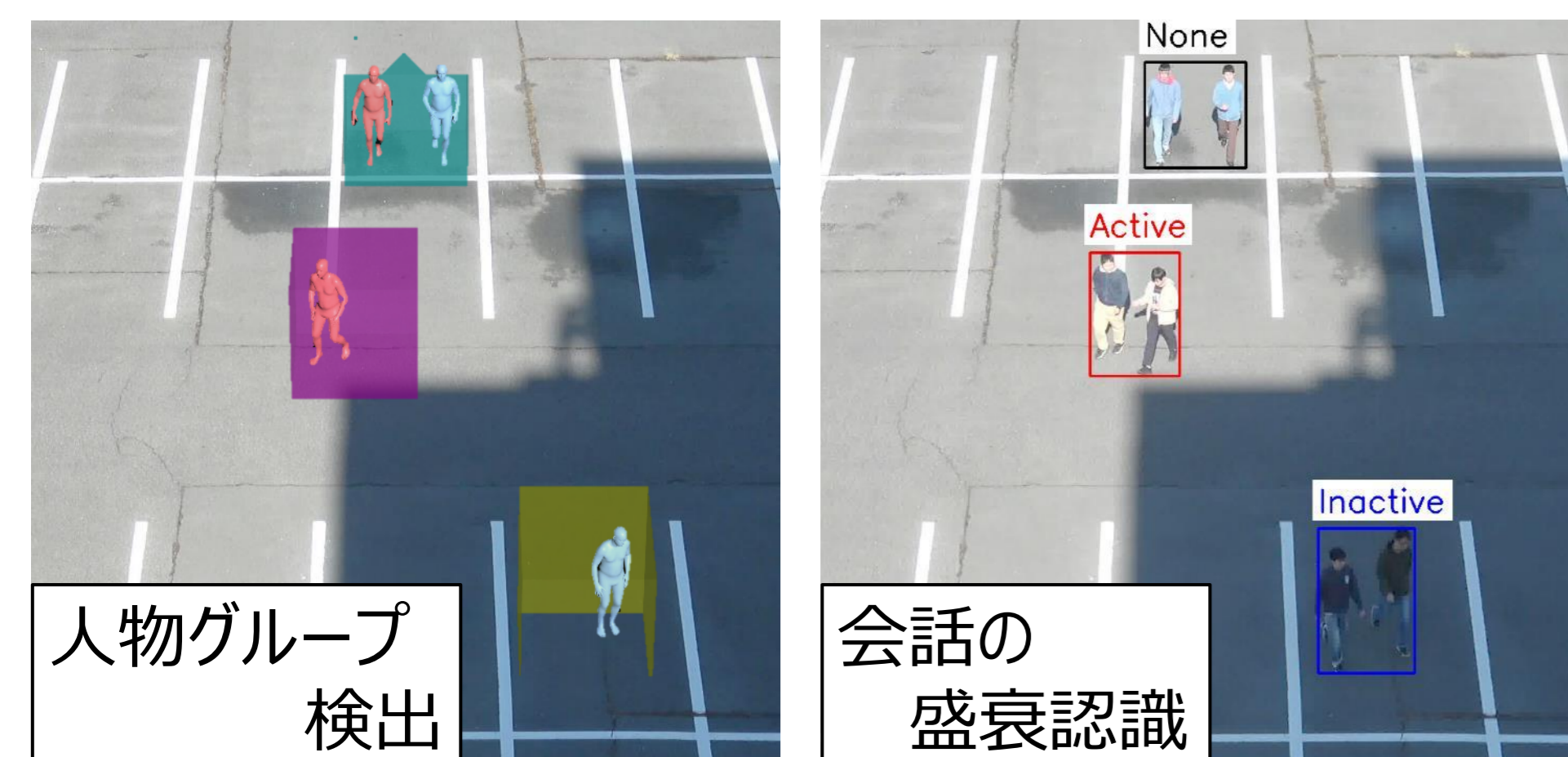


実験結果

評価データセットの一部



動作例



会話の盛衰認識の精度

インタラクション動画を特徴量として用いる提案手法の有効性を評価

- ・人物グループ数: 52組
- ・歩行経路数: 4方向
- ・平均歩行時間: 19.7秒
- ・フレームレート: 30 fps
- ・クラス分類ネットワーク: C3D [Tran et al., ICCV2015]
- ・姿勢推定アルゴリズム: PHALP [Rajasegaran et al., CVPR2022]
- ・評価方法: Leave-one-pair-out

特徴量	精度
インタラクション動画 (P1~P4)	76.2%
人物グループ動画 (P1)	57.3%
姿勢の時系列信号 (P1~P2)	72.9%
姿勢と歩行位置の時系列信号 (P1~P3)	74.1%

提案手法は人物領域の動画や姿勢と歩行位置の時系列信号と比べて会話の盛衰認識に有効

仮想カメラ視点の比較

人物グループ内インタラクションの見え方の違いが精度に与える影響を検証

仮想カメラ視点	精度
正面	76.2%
背後	74.8%
足元	72.0%
頭上	70.2%

人間が目視で確認しやすい正面が最も精度が高い