原著論文

行動の一部に見られる特徴に着目する歩行者グループ検出†

佐藤　僚太*1・波部　斉*2・満上　育久*3
佐竹　聡*4・鷲見　和彦*5・八木　康史*3

公共空間内を往来する歩行者の属性や行動目的などを推定する技術は，施設の利用状況を自動的に観測し，各人物に最適な情報提供を行う情報環境の構築に貢献できる．その際には，どの人物同士が共に行動しているのかを推定し，歩行者をグループとして扱う技術が求められる．歩行者のグループを検出するには，2人の歩行者間の歩行軌跡や注意の向きの関係を手がかりとする手法が多く用いられる．これらの手法は，観測データ全体から特徴量を抽出して識別しており，歩行者グループは常にグループらしい動きをすること前提としている．しかし，実空間での歩行者グループは各個人の興味の対象の違いや障害物を回避する道筋の違いなどから常にグループらしい行動をする訳ではない．これによって，これまでの手法ではグループ検出が困難となる事例が見られる．本研究では，歩行データの時系列分割とマルチプル・インスタンス・ラーニング（MIL）によって，行動中の短時間に存在するグループらしい振る舞いを検出する手法を提案する．提案手法では，歩行データを時系列分割し，各時間区間の特徴量をMILを用いて別々に識別する．MILは教師あり学習の一種であり，複数の要素データの集合であるバッグ内に，少なくとも1つの正の要素データがあるかどうかを識別する．上記の時系列分割によって，時間区間の一部にグループらしい特徴があればそれを検出することができる．実験では，グループ動作を模擬したデータと，実際のショッピングモールで取得したデータを利用し，提案手法の有効性を確認した．

キーワード：人物行動解析，グループ行動解析，レンジセンサ，機械学習，特徴抽出

1. はじめに

公共空間を往来する歩行者は複数人のグループを構成している場合があるが，その属性や状態・目的を知ることができれば，歩行者を個人として扱う場合よりも高価値の情報を得られると考えられる．図1のように，緑枠で囲われた子どもがグループＡもしくはグループＢのどちらに属しているかによって，それぞれのグループに最適な広告表示や情報提供は異なると考えられる［10］．例えば，家族連れには子ども向けの施設を案内することなどが挙げられる．このように，グループの構成や行動の意図や目的によって歩行者が求めている情報は大きく異なることが予想されるため，デジタルサイネージに表示する広告や案内の内容を周囲の状況に応じて動的に変化させたり，ターゲットを絞った情報提供を行えば，歩行者に有益な情報を提供でき，情報提供側の効率向上にもつながると考えられる．同様に，歩行者の属性情報などを蓄積し客層分析などを行う際にも，利用客がどのようなグループ構成で訪れているかを把握することが重要であると考えられる．また，将来的には道案内ロボットが自ら話しかける対象者を選択する際に用いることも可能である．歩行者はグループ単位で行動しているため，対象者の探索の効率化や人間との自然なコミュニケーションを図るには，ロボットが話しかける対象はある個人ではなくグループ全体である方が望ましいと考えられる．

歩行者グループの情報を得るための前段階の処理として，歩行者の集合からどの複数人が共に行動しているのかを見出す必要がある．これまで，歩行者グループ検出を目的とする多くの研究が行われてきた［3, 14］．これらの研究は複数人の位置や向きの関係から各人物がどのグループに属すかを推定するものであるが，歩

† Social Group Discovery Extracting Useful Features using Multiple Instance Learning
Ryota SATO, Hitoshi HABE, Ikuhisa MITSUGAMI, Satoru SATAKE, Kazuhiko SUMI and Yasushi YAGI

*1 近畿大学大学院総合理工学研究科
Graduate School of Science and Engineering, Kindai University
*2 近畿大学理工学部
Faculty of Science Engineering, Kindai University
*3 大阪大学産業科学研究所
ISIR, Osaka University
*4 国際電気通信基礎技術研究所知能ロボティクス研究所
ATR Intelligent Robotics and Communication Laboratories
*5 青山学院大学理工学部
College of Science and Engineering, Aoyama Gakuin University

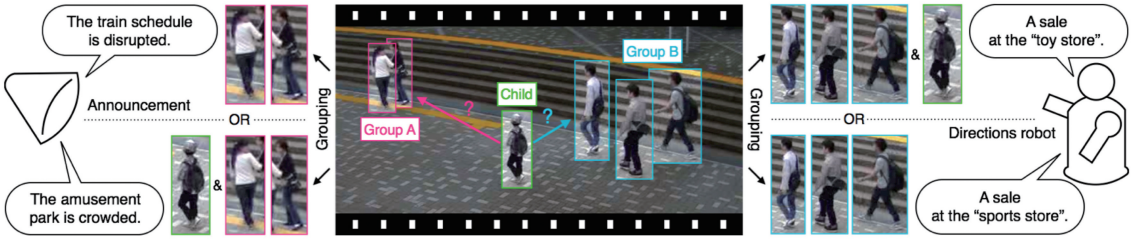


図 1 グループ情報の応用例．中央の子ども（緑枠）がグループ A（赤枠）かグループ B（青枠）のどちらに属するかで，各グループに最適な情報は異なる．

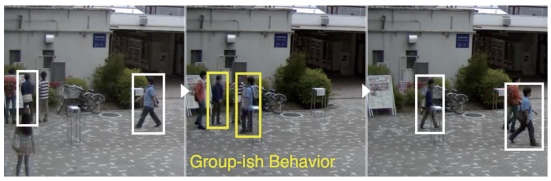


図 2 グループ行動の例．枠で囲まれた 2 名は行動の一部でのみ会話する振る舞いを見せる（中央）．それ以外の場面では人物間距離が大きく離れる（左右）．

行者が観測された時間全体のうちグループらしさを特徴付ける動きが多くの割合を占めていなければならないという制約がある．一方で，実際の公共空間内のグループは，個人の興味を持った対象の違いや障害物を回避する道筋の違いなどから常にグループらしい行動をとる訳ではないため，従来の手法では検出が困難な場合がある．

本研究では公共空間内を往来する歩行者集合からのグループ検出を行う．特に，あるグループが場面に現れてから立ち去るまでの時間全体に対して，グループらしく振る舞う時間帯が少ない場合（図 2）でもグループ検出可能とすることを目的とする．そこでは，時系列上の一部のみに存在するグループらしい振る舞いを観測する手段として，歩行データの時系列分割とマルチプル・インスタンス・ラーニング（MIL）を用いた歩行者グループ検出手法を提案し，実装と評価を行う．

ここでのグループらしい振る舞いとは，例えば，複数の人物が並行な歩行軌跡を描いたり，隣り合ったり体を向き合わせた状態で立ち止まったりする振る舞いのことを指す．このような振る舞いは，グループ内の人物同士がまとまって移動したり立ち話をした際に，歩行軌跡や体の向きに顕著に現れるグループらしさを強く示す特徴であると考えられる．

本論文では，まず 2 節で集団行動解析や歩行者グループ解析に関する関連研究について述べてから，3 節で我々の提案手法について説明し，4 節で実験手法と結果･評価を示したのち，最後に 5 節で結論を述べる．

2. 関連研究

空間内の歩行者の行動を自動で把握する手段として，カメラなどのセンサを用いて，人物を検出・追跡する技術が挙げられる［2］．これらの技術が成熟していくにつれ，さらに複雑な対象としてグループの行動にも着目されるようになっている．

先述のとおり，どの歩行者同士がグループを構成しているかを推定し，個人の人物属性推定技術と統合すれば，デジタルサイネージやロボットによるターゲットを絞った効果的な情報提供の実現が期待される．歩行者グループを対象とした研究では，例えば，グループ内で起こるインタラクションや集団行動の解析［4, 8, 11］や意思決定者の推定［6］，集団とその行動の検出［7］が行われている．このような歩行者グループの内部の状態を知るためには，その前段階の処理として，群衆からグループを抽出する必要がある．

グループ検出の研究には，例えば，各個人がスマートフォンなどの通信端末を持っていることを前提に，GPS や Wi-Fi から得られる位置情報や Bluetooth の電波強度をもとに複数人物間のグループ関係を推定するものがある［12, 13］．歩行者が通信端末を持つことを必要としない手法としては，歩行者のセンシングにカメラや赤外線センサを用いるものが提案されている［3, 14］．本研究では対象人物に制約を課さない後者のアプローチに着目する．

また，グループ検出の研究では，2 人の歩行者が同一のグループに属しているかどうかを識別する手法が広く採用されている［3, 14］．Chamveha らは歩行者の移動軌跡と注意の向きの関係から，ある歩行者ペア間にグループの関係性があるかどうかを判断する手法を提案している［3］．守口らの研究では，各フレームでの歩行者ペア間のグループ関係性を SVM により識別し，一定の閾値以上のフレーム数でグループだと判断されたペアだけを最終的にグループとみなすことで，歩行軌跡中に含まれる不規則な動きなどのノイズの影響を軽減する手法が示されている［14］．

Chamveha ら［3］のグループ検出手法では，歩行者間の位置や向きの関係性の流動的な変化によって生じる，行動中に存在するグループらしくない振る舞いを考慮していない．また，守口ら［14］はペアの歩行軌跡を複数の時間区間に分け，それらを別々に学習・識別することで時系列方向でのグループ内の人物の位置関係性の変化を考慮しているが，各区間に教師ラベルを付与する必要がある．本研究では，歩行データの時系列分割を行うことで行動中に含まれるグループらしくない振る舞いの存在を許容し，マルチプル・インスタンス・ラーニング（MIL）を導入することで分割後の各区間に付与する教師ラベルの正確性と作成効率の問題を解決する．

3. 行動の一部に見られる特徴に着目する歩行者グループ検出

本節では提案手法について詳しく述べる．

3.1 グループ検出手法の概要

本研究では，これまでのグループ検出研究で広く用いられている手法［3, 14］と同様に，グループ検出を歩行者集合から抜き出した 2 人の歩行者が同一のグループに属しているかを調べる問題に帰着させる．ここでは，歩行者ペア間の歩行軌跡と胸部向きの関係から，このペアが同一のグループに属しているかを機械学習によって識別する．ここでいう胸部とは，肩から腰にかけての上半身を指す．このとき検出・追跡に失敗した歩行者および取得した歩行データに大きな欠損が生じた歩行者は，以降のグループ検出の処理には用いない．

提案手法では，ペアの歩行データ（歩行軌跡と各フレームでの胸部向き）を L フレーム間隔で分割し，各小区間から Chamveha ら［3］と同様の特徴量を抽出して，このペアが同一のグループに属しているかどうかを Doran ら［5］の MIL を用いて識別する（図 3）．時系列分割された各小区間の少なくとも 1 つに，グループらしい振る舞いが高い比率で含まれるものがあるかどうかに注目することで，グループらしくない振る舞いの多いグループが検出できないという従来手法の問題を解決する．このように本論文では，人間の行動を解析する上で，ある一定の期間観測したデータの一部に含まれる情報を抽出することが有用であると考える．それが有効であることを示すために，ペア間でのグループらしさに着目する．提案手法で，あるペアについて同じグループに属するかどうかが判断できれば，それをベースに 3 人以上の集団の行動を知ることができると考えられる．その際には，Odashima ら［7］などのアプローチを採用することも有用であると考える．

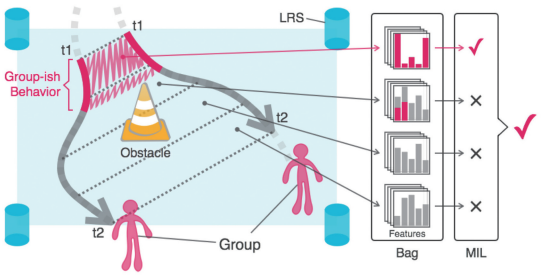


図 3 提案手法でのグループ検出の流れ

また，MIL を用いた学習を行うことで，教師データのラベル付けの正確性と効率の問題を解決する．本来同一のグループに属しているペアでも，検出された歩行データが数十秒に及ぶ場合には，グループを特徴付ける振る舞いから逸脱した振る舞いが含まれる可能性がある．このとき，歩行データの時系列全体を一つの単位として扱い場合には，各データに付与される教師ラベルが正確なものであるとはいいがたい．守口ら［14］は歩行データを時系列分割することでこのような振る舞いを許容しているが，分割後の歩行データの各区間に別々の教師ラベルを与えることは効率的でない．本研究では MIL を用いることで，教師ラベル付与の効率を維持したまま，分割後の歩行データの各区間に別々の適切なラベル付けを可能とする．

なお，歩行データの時系列分割自体は新しいアイディアではない．例えば文献［13］などでも行われているが，センサデータに含まれるノイズの影響を抑えるために分割した小区間で平滑化を行うことが主目的となっている．これに対して本論文での小区間はグループらしい振る舞いを捉えるための単位として導入したものであり，先行研究での時系列分割とは目的が大きくことなる．我々が知る限り，このような目的での時系列分割を行った事例は存在していない．

3.2 歩行データの時系列分割

本研究では，歩行データを複数の時間区間に細かく切り分け，各区間での振る舞いがグループらしいものかどうかを別々に識別することで，行動の一部のみに存在するグループらしい振る舞いを検出する．これまでの多くの研究で提案された手法では，ペアの歩行データ全体から一まとまりの特徴量群を抽出し，それがグループらしいものかどうかを識別する（図 4 (a)）．この手法では，グループらしくない振る舞いがペアの行動の多くを占めるとき，グループの検出が困難となる問題がある．これはペアの行動全体に比べて，グループらしい振る舞いが時系列全体の中で小さな割合しか

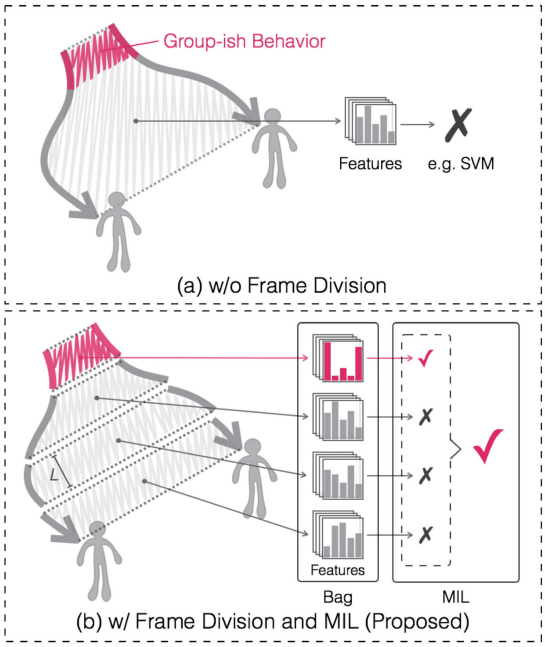


図 4　歩行データの時系列フレーム分割．(a) は従来手法．(b) は歩行データの時系列分割に MIL を適用する提案手法．

ないために，グループでないと判断されるためである．これに対し提案手法では，歩行データを L フレーム間隔で区切って複数の区間を作成し，各区間から別々に特徴量を抽出して，それぞれ識別を行う（図 4 (b)）．こうすることで，各区間に含まれるグループらしい特徴の割合を相対的に上げることができ，区間毎の識別精度を向上させることができると考えられる．

3.3　特徴量抽出

本節では，歩行データから抽出する特徴量の詳細と記述方法について述べる．

3.3.1　用いる特徴量

3.2 節で述べた手順で L フレーム間隔で分割された歩行データの各区間からそれぞれ特徴量を抽出する．各歩行者の特徴量は，その人物の歩行の様子や他の人物との位置や向きの関係から，フレームごとに算出される（図 5）．本研究では，Chamveha ら［3］と同様の 9 種類を用いる．Chamveha ら［3］は歩行者の注意の向きを示す特徴として頭部の向きを採用していたが，ここでは胸部の向きを使用する．詳細は 4.1 節で述べる．

図 5 は歩行者 j との組み合わせを考慮したときのフレーム t における歩行者 i の状態である．ベクトル $x_t^{(i)}$ と $x_t^{(j)}$ はフレーム t での人物 i と j の位置である．ベクトル $v_t^{(i)}$ と $c_t^{(i)}$ はそれぞれフレーム t での人物 i の歩行速度と胸部向きである．ベクトル $d_t^{(i,j)}$ はフレーム t での人物 i と j の間の距離である．$\psi_t^{(i)}$，$\theta_t^{(i)}$，$\phi_t^{(i)}$ はそれぞれベクトル $v_{t-1}^{(i)}$ と $v_t^{(i)}$，ベクトル $c_{t-1}^{(i)}$ と $c_t^{(i)}$，ベクトル $d_{t-1}^{(i,j)}$ と $d_t^{(i,j)}$ のなす角である．

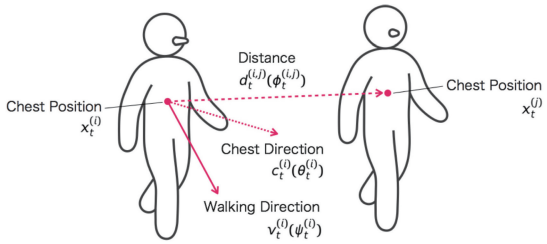


図 5　フレーム t における歩行者 i の状態

以下に，ここで用いる特徴量を列挙する．

人物間の距離と胸部向きの差　$a_{1t}^{(i,j)} = |\theta_t^{(i)} - \phi_t^{(i,j)}|$

胸部向きの差　$a_{2t}^{(i,j)} = |\theta_t^{(i)} - \theta_t^{(j)}|$

胸部向きと歩く方向　$a_{3t}^{(i)} = |\psi_t^{(i)} - \theta_t^{(i)}|$

歩行速度　$a_{4t}^{(i)} = ||v_t^{(i)}||$

人物間の距離　$p_{1t}^{(i,j)} = ||d_t^{(i,j)}||\ (d_t^{(i,j)} = x_t^{(j)} - x_t^{(i)})$

歩行速度の差　$p_{2t}^{(i,j)} = |\ ||v_t^{(j)}|| - ||v_t^{i}||\ |$

歩く方向の差　$p_{3t}^{(i,j)} = |\psi_t^{(i)} - \psi_t^{(j)}|$

歩く方向と距離　$p_{4t}^{(i,j)} = |\frac{\psi_t^{(i)} + \psi_t^{(j)}}{2} - \phi_t^{(i,j)}|$

観測された重複時間の長さの割合

$$p_5^{(i,j)} = \frac{|T^{(i)} \cap T^{(j)}|}{|T^{(i)} \cup T^{(j)}|}\ (T^{(i)} = \{t' | t' \leq t, s_{t'}^{(i)} \neq 0\})$$

なお，a_{3t} と a_{4t} についてはペアを構成する各人物について求まるので，ある時刻では合計で 11 個の特徴量が得られる．

3.3.2　特徴量の記述方法

3.3.1 節で述べた特徴量のうち p_5（ペアの人物が両方とも場面に現れている重複時間の割合）特徴量以外の 10 個の各特徴量からヒストグラムを作成する．ヒストグラムのビンは M 個設け（ここでは $M = 10$ とした），それぞれの頻度は 0.0〜1.0 に正規化する．最終的には，各区間内で生成されたすべてのヒストグラムと 0.0〜1.0 の実数で表現される p 5 特徴量を連結し，各区間での歩行者 i と j の特徴量とする．

3.4 マルチプル・インスタンス・ラーニング（MIL）によるグループ検出

本節では，MIL の説明と本研究への適用について詳細に述べる．

3.4.1 MIL の概要

MIL は弱教師あり学習の一種である．学習時のラベル付けでは，複数の要素データの集合をバッグとして，その中に 1 つでも正要素が含まれるバッグには正のラベルを与え，それ以外には負のラベルを付与する［1］．識別時にも同様に，バッグ内に少なくとも 1 つの正要素があるかどうかを基準に正負の 2 クラスに識別する．

3.4.2 MIL のグループ識別への応用

本研究においては，各人物ペアの歩行データを L フレーム間隔で複数区間に分割し，各区間を要素データとし，その全体を 1 つのバッグとみなす．これに対して MIL を適用すれば，グループらしい振る舞いを含む区間が少なくとも 1 つ存在するかどうかを判断できる．したがって，他の区間でのグループらしくない振る舞いの存在を許容することができる．

3.4.3 教師データのラベル付け

従来手法のようにペアの歩行データ全体を機械学習の教師データとして用いるときには，ペアの歩行データ全体に対して 1 つのラベルを付与する（図 6（a））．一方で，提案手法のようにペアの歩行データをいくつかの区間に分割したものを教師データとするには，一般的にはすべての区間に各々ラベルを付与する必要がある（図 6（b））．ここで問題となるのが，ペアによる具体的などの振る舞いがグループらしいか否かを人間が評価するのが困難であることである．分割時間幅が狭くなるほどその困難さは増し，ラベル付けの数も膨大なものとなる．これらの理由から，分割後の各区間にそれぞれ正負の教師ラベル付けを行うのは現実的でない．例えば，守口ら［14］はペアの歩行軌跡のすべてのフレームから別々に特徴量を抽出しているが，それらの特徴量に対して共通の教師ラベルを付与しているため，各フレームでのペアの行動の変化に対応した教師データとはなっていない．本研究では，分割後の歩行データの全体を 1 つのバッグとみなし，MIL を適用することで各時間区間についての明示的な教師ラベルを不要とする（図 6（c））．

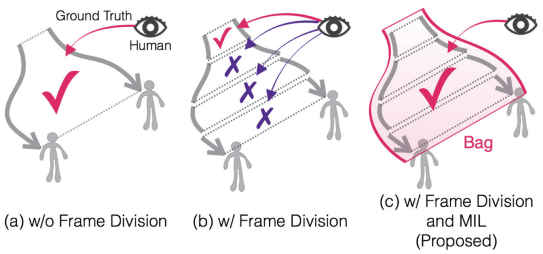


図 6 教師データのラベル付け方法．（a）は歩行データの分割を行わない場合．（b）はフレーム分割を行う場合．（c）はフレーム分割した歩行データに MIL を適用する提案手法．

4. 評価実験

本節では，歩行者グループ検出の実験手法と結果を述べる．

4.1 用いる歩行者データ

4.1.1 データ A：被験者実験によるデータ

まず，真値がはっきりとわかる状況での評価を行うために，被験者に集団行動を模擬した動きを依頼し，その動きをセンサで獲得した．被験者にはグループの構成と出発地点・経由地点・目標地点のみを指示し，行動中の具体的な振る舞いは各グループに委ねることで，自然な歩行を再現した．学習器に与える教師データの真値には，このとき実験協力者に指示されたグループ構成をそのまま流用する．被験者が動作を開始して終了するまでの一連の動きを一つの「シーン」と呼ぶことにし，22 シーンのデータを取得した．

歩行データは，人間の胸部周辺の高さに設置された 5 台のレーザレンジセンサ（図 7（b））を使用して 40 fps で取得されたものである．図 7（a）はある歩行者の各フレームでの胸部位置・向き推定の様子を，図 7（c）は歩行データ取得環境の一場面（図 7（f））を，それぞれ模式的に表したものである．歩行者の胸部の高さに設置された 5 台のレンジセンサで取得された検出物（図 7（d））から，各歩行者の胸部位置・向きを全てのフレームにおいて推定する（図 7（e））．図 7（d）は赤点部に設置されたレーザレンジセンサで検出された物体を白色で示したものである．図 7（e）では，人物の推定胸部位置・向きに黄色の楕円を当てはめている．図 7（g）はある人物の歩行軌跡であり，各フレームでの人物の推定胸部位置・向きをつなぎ合わせて作成する．これによって 22 シーンに出現する延べ 423 ペアの歩行データを獲得した．

Chamveha ら［3］の研究では，歩行者の注意の向きを表す特徴として頭部の向きを用いていたが，本研究においては胸部の向きを使用する．実環境下においては，各人物の歩行軌跡と各フレームの頭部向きを推定することが困難なためである．頭部向きを含む人物

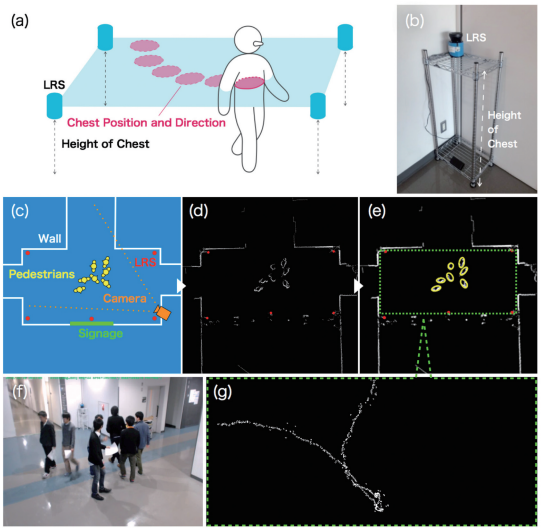


図 7 実験に用いる歩行者データ．(a) は歩行データ取得時の模式図．(b) は使用したレーザレンジセンサ．(c) はデータ取得を行ったビルの廊下の T 字路を俯瞰した模式図．(d) はレーザレンジセンサによって観測された検出物．(e) は検出物から推定された人物の胸部の位置と向き．(f) はデータ取得時の一場面．(g) はある人物の歩行軌跡．

の姿勢推定にはカメラで撮影された映像を用いることが多いが，これには各人物が鮮明に写る解像度が必要とされる．一方で，本研究が想定しているような公共空間内で撮影される群衆の映像では，各人物に高い解像度を割くことはできない．また，本研究のように，ある高さの水平面を観測するレーザレンジセンサを用いた場合であっても，人の頭部の大きさは胸部に比べて非常に小さいため，観測の漏れが生じ易い．さらに，レーザレンジセンサで観測される人の頭部と胸部の水平断面をおおよそ楕円形だと考えると，頭部の長辺と短辺の長さの差は胸部のものと比べて著しく小さく，観測されたデータからでは頭部の向きを推定することはできない．本研究では，映像を用いずに人の姿勢を観測できる手段として，レーザレンジセンサによる胸部位置・向き推定を採用し，グループ検出に応用する．

4.1.2 データ B：大規模商業施設でのデータ

次に，より実体に近い状況での結果を得るために，商業施設での歩行者データを利用して実験を行った．ここでは，Pedestrian Group Dataset [2, 9, 15] として公開されているデータを利用する．これは大阪南港にある ATC に設置されたセンサで行き交う通行者の動きを獲得したものであり，データ A と比べてより現実的なシナリオでの評価が可能である．ただし，このデータではグループへの分割はデータを見て人手で行っている．

このデータは，観測空間で人物を見下ろすような位置（天井など）に取り付けられた 3 D レンジセンサを用いて歩行する人物の 3 次元データを計測する．そこから人物領域の抽出を行った後，頭と肩に相当する領域を得る．前者からは人物の 3 次元位置を，後者からは身体の向きを推定する．これらの結果を時系列で追跡して通行する人物の移動軌跡と，その身体方向を得ている．詳細は文献 [2] を参照されたい．

今回の評価では公開データのうち 2013 年 1 月 9 日（水）の 10 時から 11 時までに取得されたものを用いる．データの中には 1090 名の人物が含まれ，そのうち 68 組が同じグループに属している．

4.2 実験の手順

グループ検出の評価を行う際には，あるシーンに出現した全歩行者から全ての組み合わせで 2 人の人物を抜き出し，そのペアが同一のグループに属しているかを識別する．歩行データを複数区間に分割して MIL を適用する提案手法と，歩行データの分割と MIL を適用を行わない Chamveha らの手法 [3] を比較することで，本手法の有効性を確認する．提案手法において各ペアの歩行データを L フレーム間隔で分割するとき，L をさまざまに変化させて性能を比較する．L を十分に大きくして歩行データの時系列分割を行わず，歩行データ全体を 1 つの要素データとするバッグとみなすことで，Chamveha ら [3] のものに相当する手法を再現する．L フレーム間隔で分割された歩行データを教師データとして用いるとき，識別する歩行データも同じく L フレーム間隔で分割されたものを用いる．学習・識別にはいずれの実験においても Doran ら [5] の MISVM を用いる．この識別器の学習時にはパラメータを選択する必要があるので，その中で最良なものを探索する．具体的には MISVM のパラメータ *cost* を 0.1 から 1000 の間で 10 倍ごとに変化させて，各 L において最良の識別結果となるものを採用する．性能評価の際には cross validation 法によって提案手法と比較対象手法の汎化性能を評価し，適合率と再現率を用いて比較する．

データ A は複数のシーン（一連の集団行動の開始から終了まで）が含まれてるため crossvalidation 法の中でも leave-one-scene-out によって評価する．すなわち 1 シーン分のすべての歩行者ペアをテスト用として抜き出し，残りの 21 シーン分のすべての歩行者ペアの歩行データを教師データとして用いる．22 シーンそれぞれの歩行者ペアをテスト用として用いるまで

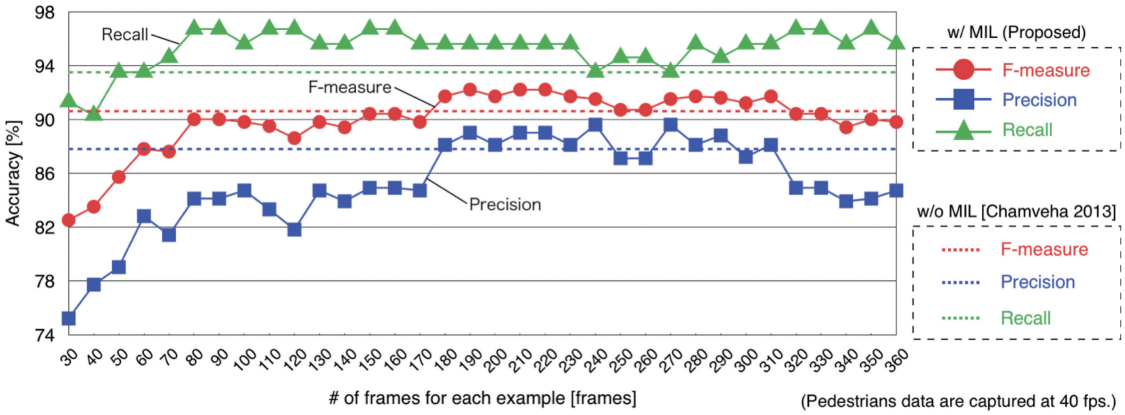


図 8　グループ識別結果（データ A：時系列分割幅と精度の関係）

同様の処理を繰り返し，すべての歩行者ペアのグループ識別結果を得る．これは同じシーンの中のデータを学習とテストに用いると適切に評価できない恐れがあるためである．

データ B はこのシーンのような明確な区切りがないため leave-one-pair-out 法とした．すなわち，グループに含まれるペアのうち一つをテスト用として抜き出し，残りのデータを学習用として用いた評価を繰り返す．ここで，データ B は著しく負例（グループに含まれないペア）が多いため，正例と負例を同数選んで学習した．よって，学習データの負例を選択する方法によって結果が大きく変わることが予想される．その影響を低減させるため，異なる負例データによる評価を 5 回繰り返し，その平均を求めた．

4.3　実験結果・評価

本節では，4.2 節で述べた実験手順に従って，グループ検出の性能を評価した実験結果を述べる．

4.3.1　データ A

まず，データ A についてのグループ検出評価結果を示す．

図 8 が時系列分割のフレーム幅 L を 30〜360 の間で 10 フレーム間隔で変化させたときの適合率，再現率，F 値の変化を示している．L を 190, 210, 220 としたときに最も高い性能を示した．表 1 にこのときの評価結果を詳細に示す．適合率と再現率が共に，Chamveha らの手法［3］に相当するものよりも高くなっていることがわかる．L を小さくしすぎると，ほとんどのペアが「グループでない」と識別され，L を大きくしすぎると，再現率と適合率が共に大きく悪化する傾向が見られた．以下，ここまでの実験で最良の結果のうち，L を 190 としたときの結果について，個々の事例別に詳細に述べる．

表 1　グループ識別結果（データ A：最良パラメータでの定量評価結果）

	Proposed (L = 190)	w/o MIL[3]
Precision	**89.0%**	87.8%
Recall	**95.6%**	93.5%
F-measure	**92.2%**	90.6%

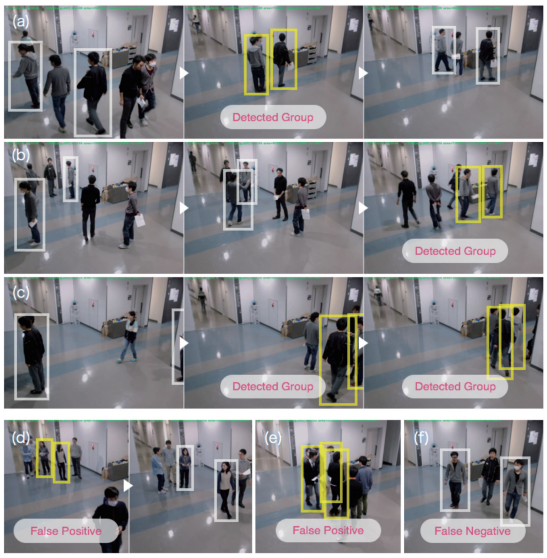


図 9　グループ検出の典型例（データ A）．(a)，(b)，(c) は人物間距離が変化する振る舞いを含むグループの検出に成功した例．(d) は一時的に隣り合ったペアをグループだと誤検出した例．(e) は一箇所に密集した 3 グループ全員を同一グループだと誤検出した例．(f) は検出漏れした例．

提案手法を適用することで，人物間の距離が大きく変化するように振る舞う歩行者グループが検出可能となった（図 9 (a)，(b)，(c)）．図 9 (a)，(b)，(c)

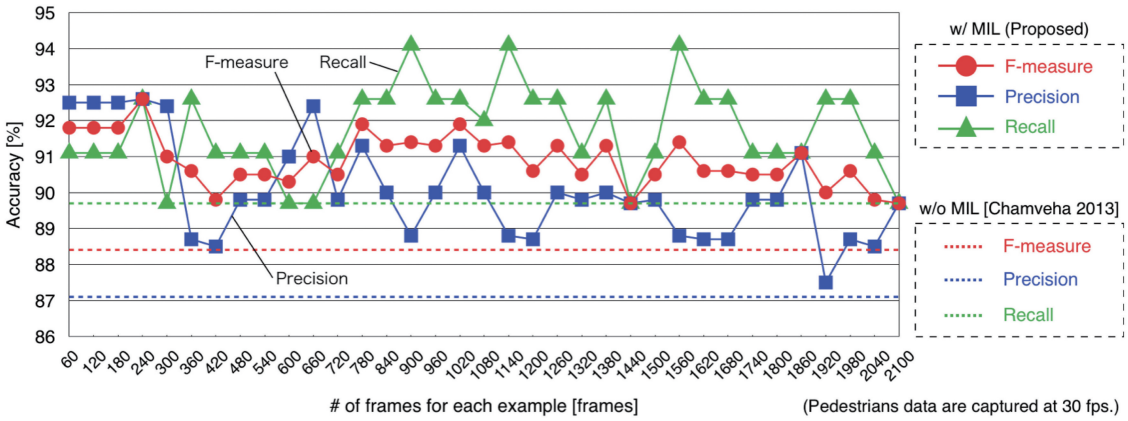


図 10　グループ識別結果（データ B：時系列分割幅と精度の関係）

は，いずれの場面でもペアは行動の多くの場面において背を向け合ったり互いの距離が大きく離れるなどのグループらしくない振る舞いを取るが，提案手法を用いることで，これらのペアが合流したり接近したりするなどの一時的なグループらしい振る舞いを捉えることで，これらのグループを検出することができた事例である．図 9（a）は，まとまって移動するグループ内の互いの距離が離れたペアが一時的に近距離になった場面を検出できた事例である．図 9（b），（c）は，離れた位置にいるペアが合流する振る舞いを検出できた事例である．これらの結果より，グループらしくない振る舞いの存在を許容しながら，一時的なグループらしい振る舞いを見出すことでグループを検出できており，提案手法の有効性が示されたといえる．

一方で，図 9（d），（e），（f）はグループの識別に失敗した事例である．図 9（d），（e）のシーンでは，複数のグループが一時的に密集した場面を，1 つのグループによる振る舞いによるものであると誤認識した．また，同一グループ内の歩行者が並行な歩行軌跡を描く場合でも，互いの距離が一度も密接しない場合には，ペアの間にグループ関係を見出すことはできなかった（図 9（f））．

4.3.2　データ B

大規模商業施設でのデータを用いた評価結果を次に示す．図 10 が時系列分割のフレーム幅 L を 60〜2100 の間で 60 フレーム間隔で変化させたときの適合率，再現率，F 値の変化を示している．L を 240 としたときに最も高い性能を示した．表 2 にこのときの評価結果を詳細に示す．適合率と再現率が共に，Chamveha らの手法［3］に相当するものよりも高くなっていることがわかる．これは 4.3.1 節に示したデータ A の場合の結果と全く同じ傾向である．

表 2　グループ識別結果（データ B：最良パラメータでの定量評価結果）

	Proposed (L = 240)	w/o MIL[3]
Precision	**92.6%**	87.1%
Recall	**92.6%**	89.7%
F-measure	**92.6%**	88.4%

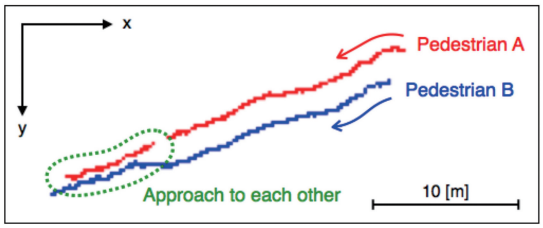


図 11　グループ識別の典型例（データ B：提案手法で検出できた例）

図 11 に提案手法による効果を現す典型例を示す．この図は環境を真上から俯瞰して，人物の移動軌跡をプロットしたものである．このペアは本来同一のグループに属すが，比較手法ではグループ関係を正しく検出できていなかった．それに比べて提案手法では正しくグループと判定されている．このペアは概ね平行な軌跡を描いて移動するが，二者間の距離は検出された歩行軌跡の大部分で密接ではない．先行手法では，時系列全体を一つの単位として扱うので，このような事例を誤判定していたと考えられる．しかしながら，このペアの行動中には，互いの距離が極めて密接になる場面が一部に含まれている．提案手法では，時系列分割によって行動中の一部の特徴的な振る舞いを検出できるようになったため，この誤判定が解消できたと考えられる．

本実験では，データ A とデータ B を用いた場合で，最良の識別結果となる最適な時系列フレーム分割幅 L

が異なった．この分割幅の差は，各データ間での歩行者の密度の差が関係していると考えられる．データAでは分割幅を5秒程度（40 fpsで$L=190, 210, 220$）とした周辺の結果が最良であったことに対して，データBでは8秒程度（30 fpsで$L=240$）が最良の結果となった．データAの環境に対してデータBの環境の方が歩行者の密度が高い．歩行者の密度が高い空間では，本来同一グループに属していないある二者間であっても，グループに良く見られる振る舞いが起こる頻度が高くなることが予想される．そのため，データBのような比較的混雑した環境下では，フレーム分割幅Lを大きく取ることで，本来グループでない二者間に一時的に見られるグループらしい振る舞いの影響を小さくできたと考えられる．

4.4 考察

本研究は人物の位置や向きの情報から読み取れるグループらしい振る舞いが行動の一部のみに存在する場合に対応したグループ検出を可能とするものである．実験では，提案手法を適用することによって，行動の一部で人物間の距離が近接するグループの検出に成功したことから，本手法の有効性が示されたといえる．

3.1節で述べたように，本論文で示した実験では大きな欠損が生じたデータを除外したが，提案手法は行動の一部のみに存在する振る舞いを抽出できるため，データが欠損している場合にも対応可能となることが期待される．これは実応用への適用を考えたときに大きな利点となる．

提案手法では，行動中のいずれの場面でも人物間の距離が密接しないグループは検出が困難である．図9(d)のような，並行な軌跡を描き，並んで歩いているグループであっても，互いの距離が離れるほど検出漏れが生じる恐れが高い．今回の実験では，同一グループ内のペア間に人間を1人挟んで歩ける程度の距離があれば，グループだと識別することはできなかった．

また，本手法は対象とする群衆の疎密度合いに大きな影響を受ける．本手法には，一時的にでもグループらしい振る舞いを示したグループの検出漏れが起きにくくなる利点がある．一方で，グループでないペアが偶然に似たような振る舞いを示すことによって，グループを誤検出し易くなる欠点がある．こうした誤検出は歩行者同士の距離が密接する状況で大幅に増加するため，極端に混雑した環境では正しい推定結果が得られないことが考えられる．データBで用いた商業施設でのデータでは通勤ラッシュや混雑したイベントといった混み合った状況ではなく，データAも同じような密度の流れを模擬しているため，本手法が有効に作用したと考えられる．この問題に対しては，例えば，人物間のインタラクションを考慮することで正しい推定が可能であると考えられる．このような別の手がかりを用いることが今後の課題である．

5. まとめ

本研究では，人物間の距離の開きや体の向きの差によって生じる，グループらしくない振る舞いの存在を許容できるグループ検出手法を提案した．本研究は，実空間内での歩行者グループが障害物や注意を引き付けられるオブジェクトなどの外的な要因によって集団の列やまとまりを崩すような一般的で自然な振る舞いを想定し，それによってグループの検出が困難となる事例に対応したものである．歩行データの分割とMILを用いる提案手法を適用することによって，行動中の多くの場面でグループらしくなく振る舞う歩行者グループの検出に成功し，本手法の有効性を示した．

今後の課題として，様々な環境での検証や本手法の制約条件の緩和と精度の向上および複雑なグループ関係への対応が挙げられる．まず，本論文では二つのデータセットを用いて評価を行ったが，対象となる人物の動きには多種多様なものがありシーンやコンテキストに依存して変化する．それらに適応できるかどうか検証を行っていき，結果の統計的有意性についても検討する必要がある．また，本手法は空間内が混雑した状況下では歩行者を複数のグループを切り分けることができず，シーン中のいずれの場面でも人物間の距離が密接にならないグループは検出できない．改善策には，空間の疎密度合いや複数人物間の位置や向きの関係に依存しない特徴量を用いることが必要である．代表的な例としては，ジェスチャーや会話などによるグループ内のインタラクションに着目することが有効であろう．さらに，本研究はペア情報のみに着目しグループ関係の推定を行っている．3人以上で構成される複雑なグループ関係の推定や行動認識には，提案手法に加えて，Odashimaら［7］の手法などを採用することが有効であると考える．

謝辞

本研究は，科学技術振興機構（JST）戦略的創造研究推進事業（CREST）およびJSPS科研費JP 25330215の支援のもとに推進された．また，有益な議論とプログラム作成にご協力いただいた飯坂瑞生氏，実験に用いたデータ取得にご協力いただいた酒井美沙紀氏と青山学院大学の皆様に感謝する．

参 考 文 献

[1] Amores, J., Multiple Instance Classification : Review, Taxonomy and Comparative Study, Artificial Intelligence, Vol. 201, pp. 81－105, 2013.

[2] Brscic, D., Kanda, T., Ikeda, T., and Miyashita, T., Person Tracking in Large Public Spaces using 3 D Range Sensors, IEEE Transaction on Human-Machine Systems, 2013.

[3] Chamveha, I., Sugano, Y., Sato, Y., and Sugimoto, A., Social Group Discovery from Surveillance Videos : A Data-Driven Approach with Attention-Based Cues, BMVC, 2013.

[4] Cristani, M., Bazzani, L., Paggetti, G., Fossati, A., Tosato, D., Bue, A., Menegaz, G., and Murino, V., Social Interaction Discovery by Statistical Analysis of F-formations, BMVC, 2011.

[5] Doran, G., and Ray, S., A Theoreticaland Empirical Analysisof Support Vector Machine Methodsfor Multiple-Instance Classification, Machine Learning, Vol. 97, pp. 79－102, 2014.

[6] Habe, H., Kajiwara, K., Mitsugami, I., and Yagi, Y., Group Leadership Estimation Based on Influence of Pointing Actions, ASVAI 2013, 2013.

[7] Odashima, S., Shimosaka, M., Kaneko, T., Fukui, R., and Sato, T., Collective Activity Localization by Spatiality Preservation Search, Advanced Robotics, Vol. 30, No. 11－12, pp. 784－794, 2016.

[8] Zanlungo, F., Ikeda, T., and Kanda, T., Potential-forthe Dynamics of Pedestriansina Socially Interacting Group, Physical Review E, Vol. 89, No. 1, 012811, 2014.

[9] Zanlungo, F., Brscic, D., and Kanda, T., Spatial-size Scaling of Pedestrian Groups under Growing Density Conditions, Physical Review E, Vol. 91, No. 6, 062810, 2015.

[10] 井上淳子，“購買行動における同伴者の影響：母娘ショッピングの観点から”，産研アカデミック・フォーラム，Vol. 13, pp. 29－40, 2005.

[11] 金子卓弘，下坂正倫，小田嶋成幸，福井類，佐藤知正，“多様なスケールでの関係を考慮した条件付き確率場による集団行動認識”，CVIM, 2013.

[12] 瀬古俊一，西野，正彬，青木政勝，山田智広，武藤伸洋，阿部匡伸，“誤差情報を考慮した同行判定手法”，UBI, 2008.

[13] 樋口雄大，山口弘純，東野輝夫，“デッドレコニングと Bluetooth の受信電波強度を用いたスマートフォンユーザ間の位置関係認識”，情報処理学会論文誌，Vol. 54, No. 8, pp. 2048－2060, 2013.

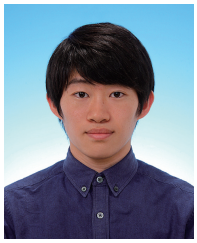
[14] 守口裕介，岡田亜沙美，浮田宗伯，萩田紀博，“軌跡間の時空間特徴量を用いた人のグループ検出”，信学論，Vol. J 96－D,No. 11, pp. 2776－2783, 2013.

[15] Pedestrian Group Dataset : http : //www.irc.atr.jp/sets/groups/

（2016 年 3 月 23 日　受付）
（2016 年 11 月 7 日　採録）

［問い合わせ先］
〒577-8502　大阪府東大阪市小若江 3-4-1
近畿大学理工学部情報学科
波部　斉
TEL：06-6721-2332
E-mail : habe@kindai.ac.jp

著 者 紹 介

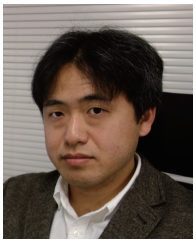


佐藤（さとう）　僚太（りょうた）［非会員］

2016 年近畿大学理工学部情報学科卒業．現在，同大学大学院総合理工学研究科エレクトロニクス系工学専攻博士前期課程に在学中．コンピュータビジョンに関する研究に従事．

波部（はべ）　斉（ひとし）［非会員］

1997 年京都大学工学部電気工学第二学科卒業．1999 年同大学大学院工学研究科電子通信工学専攻修士課程修了．三菱電機（株），京都大学助手，奈良先端科学技術大学院大学助手・助教，大阪大学産業科学研究所特任講師（常勤）を経て 2012 年より近畿大学理工学部情報学科講師．京都大学博士（情報学）．コンピュータビジョンに関する研究に従事．IEEE，ACM，電子情報通信学会，情報処理学会，人工知能学会，日本水産学会各会員．



満上（みつがみ） 育久（いくひさ）［非会員］

2001年京都大学工学部電気電子工学科卒業．2003年奈良先端科学技術大学院大学情報科学研究科博士前期課程終了．2007年同研究科博士後期課程了．同年京都大学学術情報メディアセンター研究員．2010年より大阪大学産業科学研究所助教．博士（工学）．ジオメトリを中心としたコンピュータビジョン，対象検出・追跡，歩行解析等に関する研究に従事．第14回画像の認識・理解シンポジウムMIRU優秀論文賞等受賞．IEEE，電子情報通信学会，情報処理学会，日本バーチャルリアリティ学会，日本ロボット学会各会員．

鷲見（すみ） 和彦（かずひこ）［非会員］

京都大学博士（工学）：1982年京都大学工学部電気工学科卒業，84年同大学院修士課程電気電子工学専攻修了，同年三菱電機株式会社生産技術研究所勤務，その後，産業システム研究所，先端技術研究所勤務を経て，2011年から青山学院大学理工学部情報テクノロジー学科教授（現職）．その間，1989年メリーランド州立大学客員研究員，2003年－2006年京都大学大学院情報学研究科研究員（COE）客員教授，神戸大学システム情報学研究科客員教授．専門分野は画像の認識理解およびセキュリティ．情報処理学会・電子情報通信学会（フェロー）・計測自動制御学会・ロボット学会会員．

佐竹（さたけ） 聡（さとる）［非会員］

2003年慶應義塾大学理工学部情報工学科卒業．2005年同大学開放環境科学専攻前期博士課程修了．博士（工学）．現在ATR知能ロボティクス研究所研究員．ヒューマンロボットインタラクション，ネットワークロボットシステム，ヒューマンインターフェイスに興味を持つ．情報処理学会，日本ロボット学会，IEEE各会員．

八木（やぎ） 康史（やすし）［非会員］

1983年大阪大学基礎工学部制御工学科卒業．1985年同大大学院修士課程修了．同年三菱電機（株）入社．同社産業システム研究所にてロボットビジョンの研究に従事．1990年大阪大学基礎工学部情報工学科助手．同学部システム工学科講師，同大学院助教授を経て，2003年～2016年同大学産業科学研究所教授．2008年～2009年同研究所所長補佐，2012年～2015年同研究所所長を経て，2015年より同大学理事・副学長．1995年～1996年英オックスフォード大学客員研究員，2002年仏ピカルディー大学招聘助教授，全方位視覚センシング，人画像理解，医用画像処理，知能ロボットに関する研究に従事．IEEE FG 1998（Financial Chair），OMINVIS 2003（Organizing chair），IEEE ROBIO 2006（Program co-chair），ACCV（2007 Program chair，2009 General chair，2010 Steering committee member），PSVIT 2009（Financial chair），ICRA 2009（Technical Visit Chair），IEEE ICRA the Editor of Conference Editorial Board（2007，2008，2009）．IPSJコンピュータビジョンとイメージメディア論文誌編集委員長，IPSJ Transactions on Computer Vision& Applications編集副委員長．1996年度電子情報通信学会論文賞，2003年ACM VRST 2003 Honorable Mention Award，2006 IEEE ROBIO Finalistfor T. J. Tarn Best Paperin Robotics，2008年IEEE ICRA 2008 Finalistfor Best Vision Paper，2008年画像センシングシンポジウム優秀論文賞等受賞，MIRU（2008年長尾賞，2009年デモセッション賞，2010年優秀論文賞），情報処理学会フェロー，IEEE，電子情報通信学会，日本ロボット学会各会員．博士（工学）．

Social Group Discovery Extracting Useful Features using Multiple Instance Learning

by

Ryota SATO, Hitoshi HABE, Ikuhisa MITSUGAMI, Satoru SATAKE, Kazuhiko SUMI and Yasushi YAGI

Abstract :

If we can detect groups of visitors in public spaces and commercial facilities, we can provide information depending on the attributes of the groups, and we can also provide statistics with regards to the usage of the facilities for the owners of the facilities. The features, such as person-to-person distance and gaze direction, is useful for group detection and have been used in a number of works. However, almost all of the works extract the features from the whole data. This causes miss-detection in some cases. Even when we are walking with a fiend or colleague, we do not interact with the others all of the time. This means that a meaningful information for group detection is embedded within a part of time-series data, not all of the data. We have to pick up the meaningful information and ignore the others. To this end, we divide whole of the time-series data into a set of data along the time axis. We apply the multiple instance learning (MIL) to find out the meaningful information among the data. The features computed for each time slot are treated as instances in MIL. MIL can extract one or more positive instance from all of the instances, including positive and negative ones. We conducted experiments using two types of data : simulated group actions and actual actions. Our method outperforms the existing method for both of the data.

Keywords : HumanAction Analysis, Group Action Analysis, Range Sensor, Machine Learning, Feature Extraction

Contact Address : **Hitoshi HABE**

Department of Informatics, Faculty of Science Engineering, Kindai University

3-4-1, Kowakae, Higashiosaka City, Osaka 577-8502, Japan

Phone : +81-6-6721-2332

E-mail : habe@kindai.ac.jp