

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 齊^{2,a)} 満上 育久³ 鷲見 和彦⁴

概要: 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、一部の時間区間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を利用し、提案手法の有効性を確認した。

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。例えば、家族連れには子供向けの店舗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。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 (図1)。

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][11]。これらの研究は複

数人の位置や向きから各人物がどのグループに属するかを推定するものであるが、歩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づける動きが多く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常にグループらしく振る舞っているわけではなく、興味を持った対象の違いや障害物などに起因してしばしば人物間の距離が大きく離れることがあるため、従来の手法では検出が困難である。

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。ここでは、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、歩行軌跡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

¹ 近畿大学大学院総合理工学研究科

² 近畿大学理工学部

³ 大阪大学産業科学研究所

⁴ 青山学院大学理工学部

a) habe@kindai.ac.jp

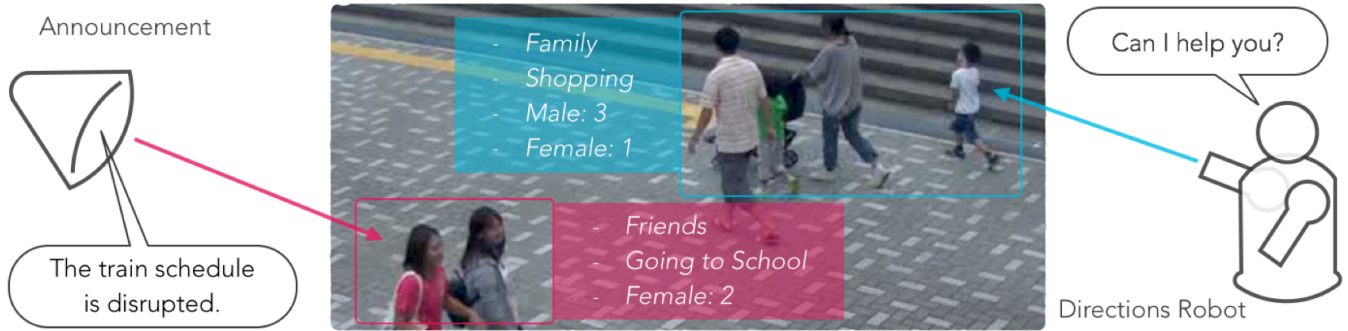


図 1 グループ情報の応用例



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)。

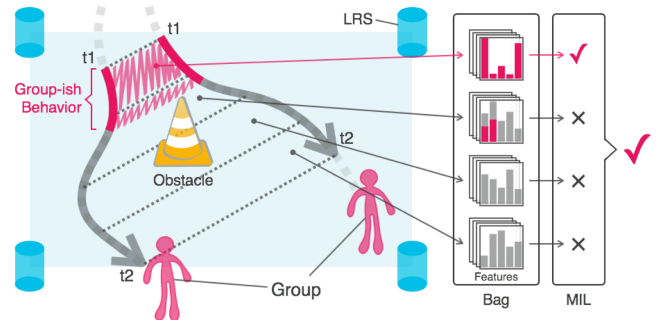


図 3 提案手法でのグループ検出の流れ

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く。

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][7][9] や意思決定者の推定 [6] 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][11]。Chamveha らは歩行者の移動軌跡と注意の向きとの関係から、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 [11]。

Chamveha ら [3] のグループ検出手法では、歩行者間の位置や向きと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 [11] 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、歩行軌跡の時系列分割を行って、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、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、分割後の各区間に正確な教師ラベルを付与しなくても学習と識別を可能にする。

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細に述べる。

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][11] と同様に、グループ検出をある一定時間に観測された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との関係から、このペアが同一のグループに属しているかを機械学

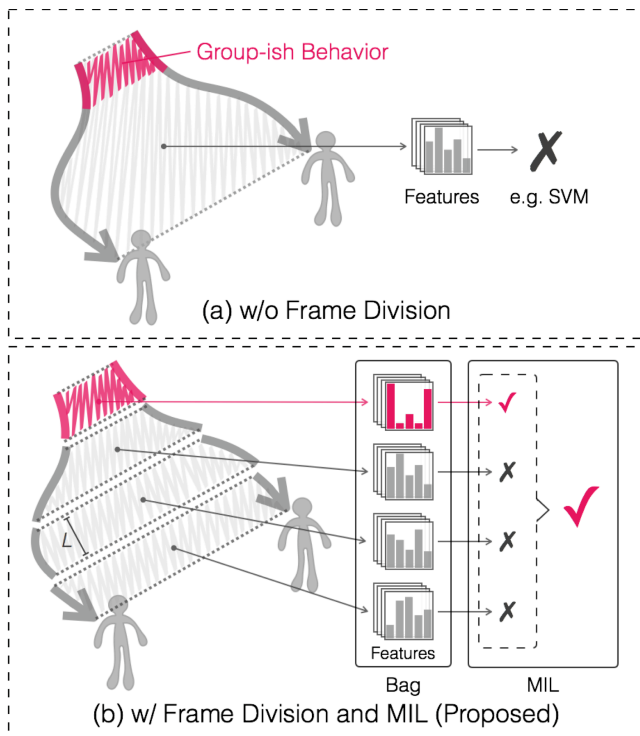


図 4 歩行軌跡の時系列フレーム分割. (a) は従来手法. (b) は歩行軌跡の時系列分割に MIL を適用する提案手法.

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.

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含まれているかどうか注目することで, 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.

このようにすることで, 他人同士が偶然に示した一時的なグループらしい振る舞いを誤検出する可能性があるが, ペア間の注意方向の関係を考慮することで, その影響を小さく留めることが期待できる.

3.2 歩行軌跡の時系列分割

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, 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, 歩行軌跡を 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.

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 $p_{2t}^{(i,j)} = \left| \|v_t^{(j)}\| - \|v_t^{(i)}\| \right|$

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

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

観測された重複時間の長さの割合

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

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, 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

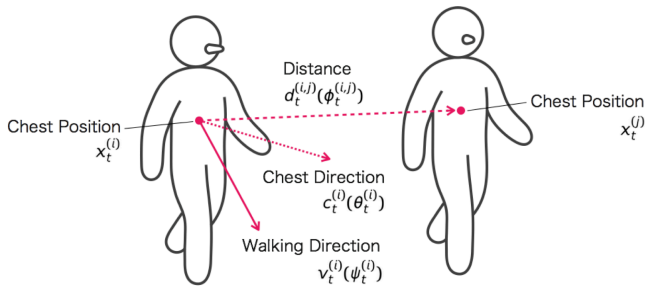


図5 フレーム t における歩行者 i の状態

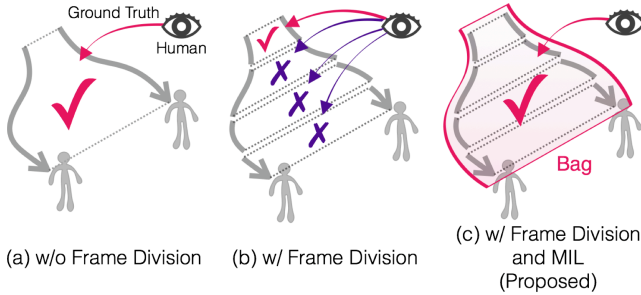


図6 教師データのラベル付け方法。(a)は歩行データの分割を行わない場合。(b)はフレーム分割を行う場合。(c)はフレーム分割した歩行データにMILを適用する提案手法。

ラムと $0.0 \sim 1.0$ の実数で表現される $p5$ 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軌跡全体を機械学習の教師データとして用いるときには、ペアの歩行軌跡全体に対して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軌跡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

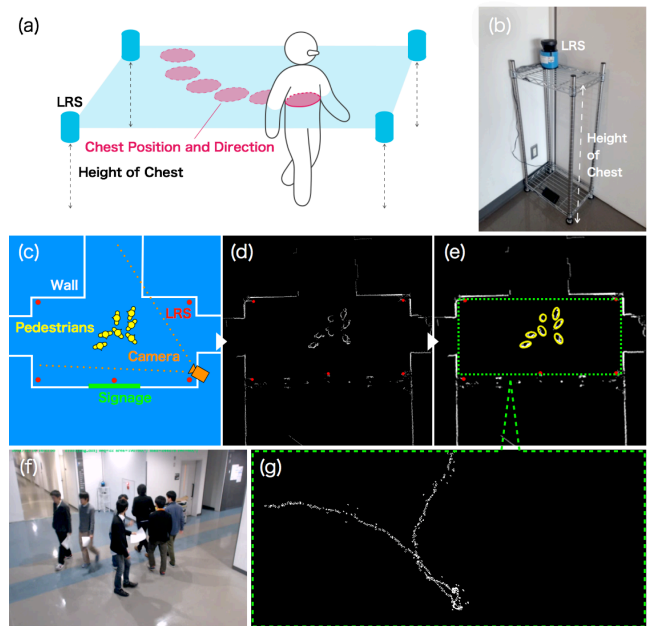


図7 実験に用いる歩行者データ。(a)は歩行データ取得時の模式図。(b)は使用したレーザレンジセンサ。(c)はデータ取得を行ったビルの廊下のT字路を俯瞰した模式図。(d)はレーザレンジセンサによって観測された検出物。(e)は検出物から推定された人物の胸部の位置と向き。(f)はデータ取得時の一場面。(g)はある人物の歩行軌跡。

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[11]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))。

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

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・経路地点・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40fps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に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。

4.2 実験の手順

実験は2つの段階に分けて行う。まず、任意の2人の歩行者が同一グループに属すかどうかを識別する第一の実験を行い、複数の手法での性能を評価し比較する。最後に、第一の実験結果を用いて、ペアごとの識別結果を組み合わせることで3人以上で構成されるグループを再現する第二の実験を行う。

第一の実験では、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[3]の手法を比較することで、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。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。

歩行データは複数のシーン（一連の集団行動の開始から終了まで）が含まれるためcross validation法の中でもleave-one-scene-out法によって評価する。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。この検証自体を各シーンについて N 回行い、 $\frac{N}{2}$ 回よりも多く正と識別されたペアをグループだとみなす。ここでは $N=10$ とする。

第二の実験では、第一の実験で得られた各ペアの識別結果を組み合わせることで、3人以上で構成されるグループを再現する。各ペアのグループ関係は、第一の実験において L をさまざまに変化させて行った検証の中で、最も良い性能が得られたときの結果を採用する。ここでは、あるペア間に直接のグループ関係がなくとも、他のいずれかの人物を介して間接的にでもグループ関係が存在するとき、そのペアは同一グループに属しているものとみなす。

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 第一の実験：ペアごとのグループ識別結果

まず、ペアごとのグループ関係のみを考慮したグループ検出評価結果を示す。

図8は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210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[3]の手法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

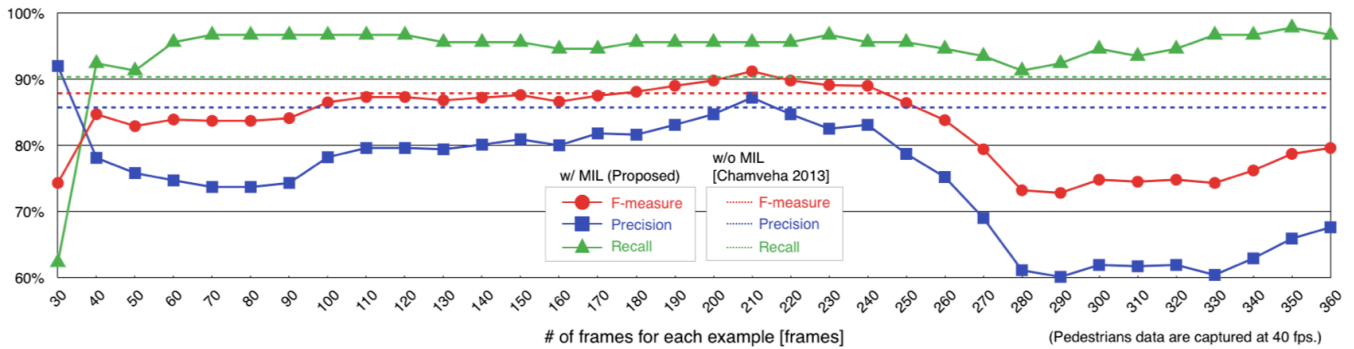


図 8 ペアごとのグループ識別結果 (時系列分割幅と精度の関係)

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である、 L を210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。

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)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、まとめ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、互いの距離が常に離れている場合には、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 第二の実験：ペアの識別結果の統合

本節では、4.3.1節の結果を用いて3人以上で構成されるグループを再現する。

表 1 ペアごとのグループ識別結果 (最良パラメータでの性能評価結果)

	Proposed ($L = 210$)	w/o MIL[3]
Precision	87.2%	85.7%
Recall	95.6%	90.3%
F-measure	91.2%	87.9%

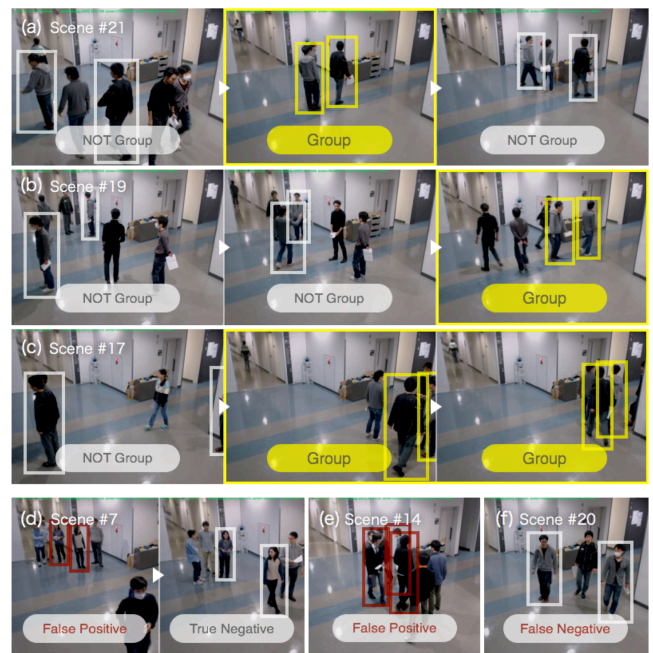


図 9 ペアごとのグループ検出の典型例。(a), (b), (c)は人物間距離が変化する振る舞いを含むグループの検出に成功した例。(d)は一時的に隣り合ったペアをグループだと誤検出した例。(e)は一箇所に密集した3グループ全員を同一グループだと誤検出した例。(f)は検出漏れした例。

このとき、あるペアが直接グループと識別されなくても、他の人物を介して間接的にでもグループ関係があれば、当該ペアは同一グループ内の歩行者であるとみなす。

ここでは、提案手法での実験結果で最良であった、 $L = 210$ とした場合と歩行データを複数区間に分割しない場合でのグループ検出精度を比較する。ペアごとの識別結果を組み合わせ、最終的に構築された任意の人数のグループを単位として、適合率と再現率によって性能を評価する。

提案手法を適用することで、再現率が1.5ポイント向上

表 2 ペア識別結果の統合の性能評価結果

	Proposed ($L = 210$)	w/o MIL[3]
Precision	73.4%	75.5%
Recall	83.7%	82.2%
F-measure	78.1%	78.6%

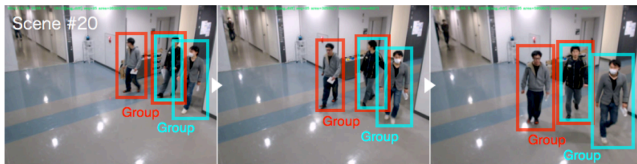


図 10 ペアごとの識別結果の統合によるグループ検出例

したものの、適合率が2.1ポイント悪化したことで、F値が0.5ポイント悪化した。これは、図9(d), (e)のように複数グループが一箇所が密集したことで、一部のペア間に誤検出が生じたことから、全体が1つのグループであると識別されるなどし、適合率が悪化したことに起因するものである。

個々の事例をみると、4.3.1節の実験においてペア間の直接のグループ関係を検出できなかった図9(f)のシーンでは、ペアごとの識別結果の統合により、両端の人物に関して、中央の人物を介した間接的なグループ関係を検出することに成功した(図10)。

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。提案手法を適用することによって、行動の一部でグループらしく振る舞う例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。

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。例えば、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。ここで用いた歩行者データは、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い密度の人物集合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。

また、対象とする環境やデータセットに応じて、分割フレーム幅 L を適切な値に調整する必要がある。例えば、 L を大きくしすぎた場合には行動中の一部のみに存在するグループらしさを見逃す可能性があり、小さくしすぎた場合には振る舞いの特徴を捉えるために十分な時間幅が確保できない可能性がある。実験での最良パラメータである $L=210$ は、これらの悪影響が最も小さくなり、グループらしい振る舞いの特徴を学習・識別するために適したものであったと考えられる。

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ペア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。4.3.1節の実験では、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。4.3.2節の実験では、各ペア間の識別結果を統合し、他の人物を介した間接的なペア関係の有無に着目することで、一部のグループについてはこの問題を解決できたが、全体としては過検出となり、識別精度の悪化を引き起こした。

これらの問題に対しては、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。橋本ら[10]の研究では、グループ内で発生する身振り手振りや顔向き変化などのジェスチャーによるインタラクションを特徴量として用いることで、グループ検出精度を改善可能であることが示されている。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。歩行者データの分割とMILを用いる提案手法を適用することで、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、本手法の制約条件の緩和と精度の向上が挙げられる。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。

謝辞

本研究の一部は、科学技術振興機構(JST)戦略的創造研究推進事業(CREST)「歩容意図行動モデルに基づいた人物行動解析と心を写す情報環境の構築」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*, 201, 81-105, 2013.

- [2] Brscic, D., Kanda, T., Ikeda, T., and Miyashita, T., Person Tracking in Large Public Spaces using 3D 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 Theoretical and Empirical Analysis of Support Vector Machine Methods for 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.
- [7] Zanlungo, F., Ikeda, T., and Kanda, T., Potential for the Dynamics of Pedestrians in a Socially Interacting Group, *Physical Review E*, Vol. 89, No. 1, 012811, 2014.
- [8] Zanlungo, F., Brscic, D., and Kanda, T., Pedestrian Group Spatial Size Scaling under Growing Density Conditions, *Physical Review E*, Vol. 91, No. 6, 062810, 2015.
- [9] 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", 情報処理学会研究報告 CVIM, 2013.
- [10] 橋本 知典, 波部 斉, 満上 育久, 鷺見 和彦, 八木 康史, "ジェスチャータイミングを用いたグループ検出精度の向上", 情報処理学会研究報告 CVIM, 2016.
- [11] 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", 電子情報通信学会論文誌, vol. J96-D, no. 11, pp. 2776-2783, 2013.