

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

橋本 知典¹ 波部 齊² 満上 育久¹ 鶴見 和彦³ 八木 康史¹

概要：公共空間や商業施設などを行き交う人々が構成するグループを検出できれば、そのグループに応じた情報提供が可能となり、さらには施設の利用状況を知る貴重な情報となる。グループの存在を知るにはお互いの距離や視線方向などが有用であるが、観測場所が混雑している場合や、グループが分かれて行動しているときなど、それらだけでは十分ではない場合がある。我々はこのような場合に有用な特徴として人物間のインタラクションを示すジェスチャーに着目し、ジェスチャーが発生したか否かを映像から判定し、それを加味してグループを検出する手法を提案する。ジェスチャーは身振り手振りや顔向き変化などによって観測できるが、それらによる画像上で見え方変化は多種多様となるため判定は簡単ではない。これに対して提案手法では、オブティカルフローの変化によってジェスチャーを検出する。さらに、そのジェスチャーの発生時刻を人物間のインタラクション発生を特徴づける情報を捉え、ジェスチャー情報を加味したグループ検出を行う。

キーワード：グループ、インタラクション、ジェスチャー

1. はじめに

近年では、公共空間での効果的な情報提供やロボットによるサービスを行う技術が注目されている。サービスの対象となる空間内的人物の関係性を理解し、発信する情報を選択することができれば、空間内の人物に対してよりよいサービスの提供、効率のよい宣伝などを行うことができる。そのためには、空間内のどの人物がグループを構成し

ているのか、グループとして何をしようとしているのか、グループの中でどのような役割分担が行われているのか、などが分かれば非常に有用な情報になると考えられる。例えば、図1に示すようにシーンの中で共通の目的を持ったグループを見つけることができれば、その人達にだけ情報を提示することで効果的な宣伝や情報伝達が可能となる。ここでグループとは、家族・同僚・友人などの社会的なつながりがあり、空間内において、一緒に何かを行おうとしている複数人の人物のことである。

防犯カメラなどの環境に設置されたセンサは、不特定多数が通行する公共空間でのグループ行動解析に適しており、これまでにも多くの事例が提案されている。例えば、比較的古くから行われている事例としては、それぞれの人物がどのグループに属しているかを推定するグループ検出があり、最近ではそのグループの属性推定やグループ内の役割推定も取り上げられるようになっている。特にグループ検出はそれらの第一ステップとなる処理であり、既に様々な研究が行われている [2], [3], [8], [9]。これらは対人距離を用いて、距離が近いと同じグループに属する可能性が高い、という仮定に基づいている。人流の密度が低い場合はこの仮定が成立するが、混雑した環境では対人距離では不十分になることが多い。

対人距離だけでは不可能なグループ検出を可能とするに

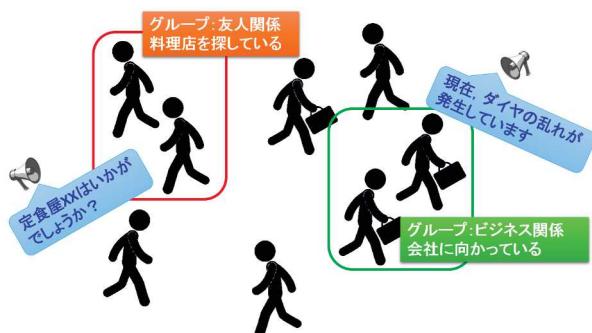


図1 グループの目的に応じた情報発信

¹ 大阪大学
Osaka University

² 近畿大学
Kinki University

³ 青山学院大学
Aoyama Gakuin University

は、観測する複数の人物間に発生するインタラクションに着目することが有用であると考えられる。実際、我々がシーンを観察する場合には、必ずしも対人距離からグループを推定しておらず、むしろ、特定の人物への顔向き方向の変化、何か意思を伝えるための身振り手振りなどのジェスチャーの発生、他人の呼びかけに反応する動作などのような、他人との関わりを含んだインタラクションに着目している。すなわち、これらの情報はグループについて考える上で重要な要素であり、グループ行動解析には必要な特徴であると言える。

そこで本研究では、身振り手振りなどのジェスチャーの発生をインタラクション発生を特徴づける情報と捉え、この情報を用いてグループ検出精度を向上させる手法について提案する。そこでは、まず、ジェスチャーの検出をオプティカルフローを用いて行う。ジェスチャーとして検出すべき動作は、身振り手振りや指差し動作をするなどの動作であり、これを検出するために身体部位の姿勢推定を用いることが考えられる。しかし、本研究で想定しているような複数人物が映るシーンにおいては通常、各人物に割かれる解像度は高くなく、そのため身振りや手振りなどは姿勢の推定が難しいなどの問題が生じる。そこでジェスチャーによる人物の見え方の変化を密なオプティカルフロー[7]として捉え、ジェスチャー時のフローの様子の違いからジェスチャー発生の確からしさを「ジェスチャー発生度合」として定量的に表現する。人物がジェスチャーや指さし動作を行ったとき、その人物の手の動きは他の人体部位の動きと比べて大きく、オプティカルフローも大きくなることが予測される。この性質を用いて、ジェスチャー発生度合を表す特徴量を計算する。

また、ジェスチャータイミングを考慮したグループ検出については、2人の人物に着目したとき、少なくとも片方の人物がジェスチャーを起こした場合、どちらもジェスチャーを起こしていない場合と比べて、2人はグループである可能性が高いという考え方のもと、上記のジェスチャー度合を表す特徴量を加味して、グループらしさを記述することを考える。具体的には、Chamveha ら[5] や佐藤ら[6] のグループ検出の特徴量のうち、対人距離を表す特徴量をジェスチャー度合で割る、すなわち、ジェスチャー発生時にはより近くに人物がいるとみなすこととしたものを新しい特徴量として加えて、機械学習によるグループ検出を行った。

2. 関連研究

グループ行動解析の目的を詳細に考えると、まず、グループに所属する人物を特定するグループ検出が挙げられる。その結果を踏まえてさらに、主導的な立場は誰かといったグループ内での人物の役割、家族連れ、若者の友人同士といったグループの属性や、何かを探している、目的

に向かっているといったグループの状態などを推定することが考えられる。さらに、これらの要素とグループの行動履歴を対応付けることで、グループの属性やグループの状態と、その行動の関連を知ることもできると考えられる。

先に挙げた様々な要素の中で、第一段階となるグループ検出が多くの研究者によって注目され、多くの研究が行われている[1]-[6]。その中では、最も基本的であり、かつ効果的な特徴量である対人距離を用いた例が多い。井上ら[2]は、大型ディスプレイ広告の前で立ち止まった複数人がどのような関係にあるのかを推定し、推定結果に基づいた広告を提示することで、より訴求効果のある広告提示ができる可能性があることを示している。また、岡本ら[3]は、歩行者は互いの関係に応じて、近づいたり離れたりして距離を調節しながら歩行を行っていることに着目し、歩行者間の人間関係の推定を行っている。しかしながら、空間内の人物の数が多く、人物関係と対人距離が対応しないことがある場合などは、対人距離だけでは不十分であると思われる。そのため、対人距離以外の特徴量が必要となる。

対人距離以外の特徴に着目した関連研究として、波部ら[4]による人間関係の推定や異常行動抽出を行う手法がある。波部らは歩行パターン変化発生の時間関係に着目しており、そこでは、立ち止まる、歩き始める、曲がるといった歩行者の歩行パターン変化を検出し、他の人間がその変化にどれだけ影響を与えたかを測る尺度を導入している。しかしながらこの研究では、人物の移動軌跡のみを用い、身振りや手振りといった、インタラクション理解に有用な情報は一切利用していない。

インタラクションを特徴量として用いた関連研究としては、人物の顔向きや胸部方向に着目した研究[5], [6]がある。これらは、同じグループに属する人間は頻繁に視線を合わせたり、同じ対象物を見ることが多い、などといった人間の注視方向に関する性質を用いてグループを推定する手法を提案している。二人の歩行者の顔向きや、その差といった特徴量を定義し、それらと人物の位置関係を用いて推定を行っている。この研究は、二人の間で顔を見合わせる、などグループ内的人物同士に発生するインタラクションに着目していることになる。しかしながら、人物間のインタラクションとして用いている特徴量は顔向きに関するものだけであり、身振り手振りや指さし動作といったジェスチャーを伴うインタラクションについては考慮されていない。

そこで本研究では、身振り手振りなどのジェスチャーによる人物の見え方の変化をオプティカルフローとして捉え、ジェスチャー発生度合を計算し、グループ検出に用いる特徴量を新しく記述することを考える。本研究では、上記[5], [6]の Chamveha らや佐藤らが用いた特徴量にジェスチャーに関する特徴量を追加した場合に、グループ検出精度にどのような影響を与えるかを評価することにより、

グループ検出にジェスチャーを用いることの有用性を示すことを目標とする。以降の章で、提案手法の詳細について述べていく。

3. オプティカルフローを用いたジェスチャー発生の検出

本研究では、グループ検出において従来より用いられてきた人物の移動軌跡や身体の向きといったセンサから取得できる特徴量に、画像から取得できる人物のジェスチャーの特徴量を加味してグループ検出を行う枠組みを想定している(図2)。すでに述べたように、そういった身振り手振りや指さし動作といったジェスチャーの検出を行うために姿勢推定を用いることは、本研究で想定しているような複数人物が映るシーンにおいては難しいと考えられる。そこで、ジェスチャーが発生したことを人物の見え方が変化したと捉え、これを密なオプティカルフローとして表現することを考える。

図3のようにジェスチャーを起こしている人物について、そのうちの2フレーム間におけるオプティカルフローを可視化したものを図4に示す。図4のオプティカルフローについては、映像から取れた2枚の画像間でフローを計算し、その後、人物領域に合わせて切り出したものを示している。(A)については手を振り上げる動作をしており、オプティカルフローも腕が色濃く出ていることが確認でき、腕部分のフローが大きく出ていることが予測される。また(B)については、人物が歩行中であることから体全体にフローが出ていることが確認できるが、その中でも指をさす動作によって手の部分でフローが色濃く出ていることが確認できる。この性質を用いてジェスチャーが発生した度合を表す指標を計算し、ジェスチャー検出を行う詳細について、以下に述べる。

3.1 ジェスチャー発生度合の計算

先ほど、身振り手振りや指さし動作発生時は、特定の体の部位にてその他の部位よりフローの大きさが大きくなると

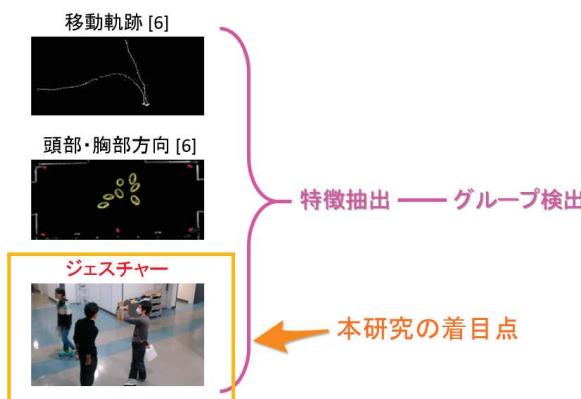


図2 本研究の枠組み

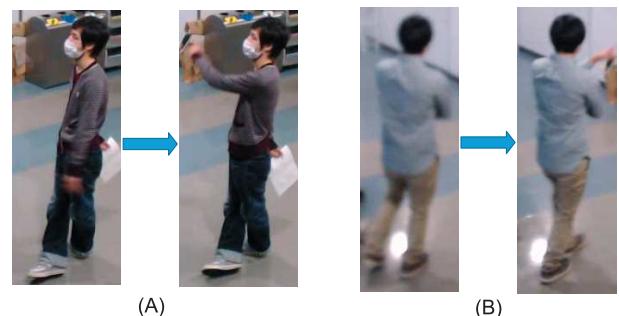


図3 ジェスチャーの例

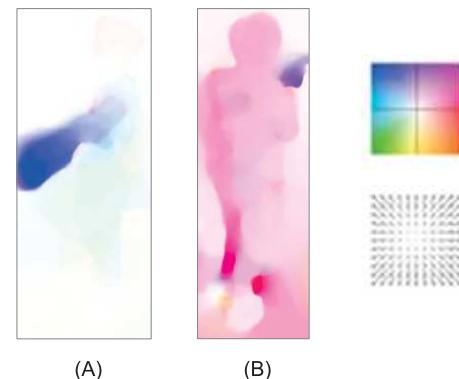


図4 オプティカルフローによるジェスチャー表現。色合いはフロー方向を表しており、色の濃淡でフローの大きさを表している。

いう傾向があることを確認したが、人体全体に着目すると、歩行により足においてもフローが大きくなることが考えられる。そこで、この影響を受けないためにジェスチャー検出には上半身のみを用いることにする。ジェスチャーは手や腕を用いて行うものが大半で、そのためジェスチャーは上半身部分に現れやすいと考えられるので、着目する領域を上半身に限定することはジェスチャー発生検出に影響を与えないと考えられる。

ジェスチャーが発生しているとき、その部分のフローは他の部分のフローより大きくなることから、人物領域内においてフローの値が最大であることが予測される。そのため、人物領域内のフローの最大値が他のフローの値より大きければ、ジェスチャーが発生している可能性があると予測される。このことを用いて、ある1フレーム t におけるジェスチャー発生度合 $G(t)$ を以下のように定義した。

$$G(t) = \frac{V_{max}(t)}{V_{mean}(t)} \quad (1)$$

ここで $V_{max}(t)$ は上半身のフローの大きさの最大値、 $V_{mean}(t)$ は前景における上半身のフローの大きさの平均値を表す。ここでの前景とは、ある画素 p のフローの大きさ $v(p)$ がある閾値 T 以上であるとき、その画素を前景 P_{fg} と呼ぶ。すなわち、

$$P_{fg} = \{p | v(p) > T\} \quad (2)$$

$$V_{mean} = \frac{1}{N} \sum_{p \in P_{fg}} v(p) \quad (3)$$

N は P_{fg} の要素数である。なお、人物が止まっているなどでフローの値が小さく、人物領域内のすべての画素で前景とみなされない場合があり、その場合は $G(t) = 0$ とする。まとめると、

$$G(t) = \begin{cases} 0 & (P_{fg} = \phi) \\ \frac{V_{max}(t)}{V_{mean}(t)} & (otherwise) \end{cases} \quad (4)$$

これに基づき、各フレームでジェスチャー発生度合を計算し、実際にジェスチャーが発生しているフレームでこの特徴量が示す値が高くなれば、ジェスチャー発生を捉えることができていると言える。その結果は 5.1 章にて示す。

4. ジェスチャータイミングを用いたグループ検出

本研究におけるグループ検出は、複数人物がいる中で 2 人の人物に着目し、その 2 人がグループであるか否かの判定を機械学習によって行う。ここでグループ検出に用いる特徴量は Chamveha ら [5] や佐藤ら [6] と同じものに、ジェスチャー情報を含んだ特徴量を追加したものを用いる。また特徴量の記述の方法、及び学習方法は佐藤らの手法に倣うこととする。

4.1 相互関係特徴を用いたグループ検出

Chamveha や佐藤らがグループ検出に用いた特徴量は、移動軌跡から計算できる対人距離や 2 人の人物の歩行速度や歩行方向の差といった特徴量と、頭部や胸部の向きから計算できる注視方向の一致度などの人物の注視方向に関する特徴量である。

ここで、Chamveha らは 2 人の人物について人物ペアが出現してから去って行くまでの時系列データ全体に対してこれらの特徴量を計算し、ヒストグラムを作てグループ検出を行っていた。それに対し佐藤らは、グループの行動でも時系列データ全体でグループらしい行動をするわけではない点に対応すべく、時系列データ全体を一定フレーム幅で時系列分割し、分割後のそれぞれの区間に上記の特徴量を計算し、ヒストグラムを生成した（インスタンスと呼ぶ）。このようにインスタンスに分割すれば、グループらしい情報を含むものとそうでないものが存在することになる。この考え方と親和性の高い Multiple Instance Learning を導入することにより、Chamveha らでは検出できなかったようなグループを検出することに成功している。以下ではインスタンスを得るための分割フレーム幅を L として考えていく^{*1}。

^{*1} 以降の実験では $L = 100$ フレームとしている

4.2 ジェスチャー情報を含む特徴量の記述

本研究では、注目している 2 人の人物うちどちらか一方がジェスチャーを起こした場合、どちらもジェスチャーを起こしていない場合と比べて、2 人はグループである可能性が高いという考えに基づいて、新しくジェスチャー情報を含む特徴量を記述することを考えている。そこで、ある 2 人について、2 人の距離が離れていてもジェスチャーが起こればグループである可能性を高めて、逆に、2 人の距離が近くても 2 人の間にインタラクションが皆無であればグループである可能性を低くすることを考える。具体的には、Chamveha らや佐藤らの用いている特徴量の内、人物間距離を示す特徴量にジェスチャー度合の逆数に比例する重みをかけることにする。ジェスチャー度合が高いほど、この値は小さくなり、グループである可能性が高くなることを表現できると期待される。

2 人の人物 i, j について、2 人のジェスチャー度合を $G_i(t), G_j(t)$ と表す。2 人の内、どちらかがジェスチャーを起こしてもグループの可能性を高めることを考えているので、 $G_i(t), G_j(t)$ のうち大きい値を選んでいったものを 2 人の間のジェスチャー度合 $G_{i,j}(t)$ として表現する。すなわち、

$$G_{i,j}(t) = \max\{G_i(t), G_j(t)\} \quad (5)$$

佐藤らの用いているインスタンスは、 L フレーム分の 9 つの特徴量の正規化済みヒストグラムを連結させたものとなっており、その内、対人距離を表す特徴量は b ビンのヒストグラムで表されている。この対人距離について平均距離となる値を算出し、2 人の間のジェスチャー度合の逆数に比例する重みをかける。まず、あるインスタンス I におけるビン i の対人距離を表す特徴量を $p(i)$ で表したとき、平均距離となる値 p^I を、

$$p^I = \sum_{i=1}^b ip(i) \quad (6)$$

で計算する。この値はシーン中の連続する L フレーム分のものであり、ジェスチャー度合もそれに相当する時間幅において適切に選ぶ必要がある。ここで佐藤らの実験によると L フレームは数秒に相当する幅に設定されているので、数秒において目立ったジェスチャーが存在するのであれば、それを代表として選んでも構わないと考えられる。よって、インスタンス I に新しく追加する特徴量 $F_{gesture}^I$ を以下のように定義する。

$$F_{gesture}^I = \frac{p^I}{\max\{G_{i,j}^I(t)\}} C \quad (7)$$

$G_{i,j}^I(t)$ はインスタンス I に相当する一定フレーム幅における $G_{i,j}(t)$ を表す。また、 C は定数であり、これはその他と特徴量と比べてスケールが異なる場合にジェスチャー情報がグループ推定に加味されないことを避けるために用意

している。5.2 章の実験では、 C を変化させたときに性能がどのように変化するかを確認する。

そして、この計算した特徴量 $F_{gesture}^I$ を元の 9 つの特徴量を含むヒストグラムに追加して、これを新しい特徴量として Multiple Instance Learning によって機械学習をさせ、グループ検出精度が向上するかどうかを評価する。

5. 実験

本章では、3 章のジェスチャー検出と 4 章のグループ検出についての評価を行う。

5.1 ジェスチャー検出の評価

5.1.1 評価方法

ジェスチャーが発生したかどうかを判定するには、3 章で述べた式(4)を用いる。評価方法としては、実際のシーンを見てジェスチャーが起こっているフレームを主観で決定してこれをジェスチャー発生区間とし、各フレーム t について $G(t)$ を求め、あらかじめ決定したフレームと比較する。ジェスチャーは数フレームに渡って行われるため、上記のジェスチャー発生区間は時間的な幅を持つように設定している。この数フレーム内で $G(t)$ の値が高くなっていることが確認できれば、式(4)の妥当性を示すことができると言える。

5.1.2 実験結果

ジェスチャーを上手く $G(t)$ によって捉えられている例を図 5 に示す。図 5 では、15~20 フレームにおいて $G(t)$ の値が高くなっていることが確認でき、実際にジェスチャーを起こしているフレームが 14~28 フレームの間であることから、ジェスチャー発生を検出できていると考えられる。図 6 では、14~28 フレームでの実際のジェスチャーシーンを示している。注目している人物は左手で指をさすというジェスチャーを行っており、正しくジェスチャーの発生を捉えているとわかる。

また別の実験結果を図 7 に示している。この例では、



図 6 図 5 における実際のジェスチャーシーン

1~17, 36~54 フレームにおいて実際にジェスチャーが発生しており、そのときの $G(t)$ の値も高い値を示しており、ジェスチャーを捉えることに成功していると考えられる。しかし、実際にはジェスチャーが発生していないにも関わらず、 $G(t)$ が高い値を示していることが 70 フレーム付近で確認することができる。70 フレームあたりの実際のシーンを図 8 に示している。注目している人物はマスクを直し手を下しているだけで、ジェスチャーとは考えられない動作をしている。しかし、手を下すという動作に対してフローが大きく出てしまうことで、 $G(t)$ が高い値を示してしまった例になっている。データには、こういったジェスチャーでないような手の振りなどに反応てしまい、高い値を示してしまう例が多い。これは、提案手法のみでは原理的に識別できないため、ジェスチャー種別の詳細な推定やインタラクション相手の推定など別の情報と組み合わせる必要がある。また、注目している人物領域の上半身部分に、他の人の足が写ってしまい、その人の足に反応してしまう、という例も少なくなかった。実環境においては、人物同士が重なるシーンが多く含まれるため、解決していくべき問題である。

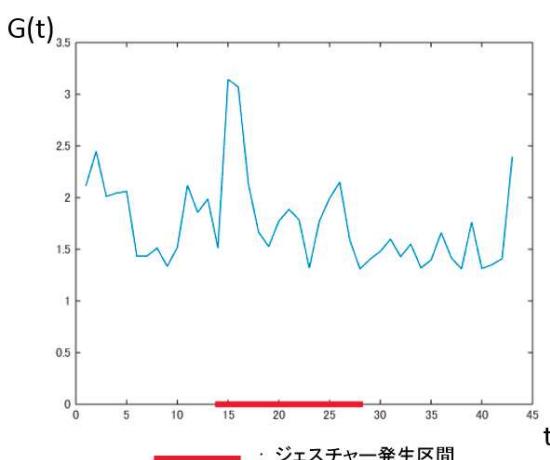


図 5 実験結果 1

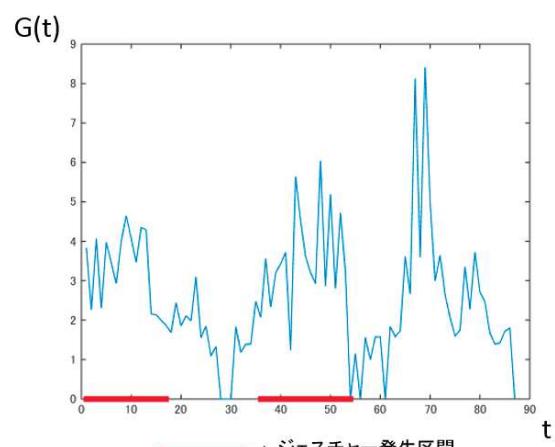


図 7 実験結果 2



図 8 $G(t)$ が誤って高く反応してしまう例

また、実際にはジェスチャーを起こしているが、高い値を示さなかった例もあり、これは、ジェスチャー動作自体が小さくフローとして鮮明に表現されなかつた、あるいは、歩行中の人物などで人体部分のフローが全体的に大きく、ジェスチャーのフローと大きく差が出なかつたなどの理由があると考えられる。

本実験では、オプティカルフローを用いたジェスチャー発生度合の計算の基本的な有効性は確認できたが、ジェスチャー発生度合の記述の方法を改善することで、よりジェスチャー発生を捉えることができる可能性があると考えられる。

5.2 グループ検出の評価

5.2.1 評価方法

グループ検出の評価は4章で述べた通り、シーン中の2人に着目し、その2人がグループであるかどうかを機械学習により判定する。用いる特徴量は佐藤らが用いた特徴量に新しい特徴量(式(7))を加えたもので、学習方法は佐藤らに倣う。ここで、[6]ではMultiple Instance Learningを用いて、最低一つでもグループらしいインスタンスがあればグループとみなすようにしていたが、本研究では、推定の感度を調整した場合の精度をみるために、グループらしいインスタンスが最低 A 個ある場合に全体をグループと判定するようにして、このパラメータ A と式(7)における定数 C を変化させたときのグループ検出精度について評価を行った。

用いたデータは大学内でグループ行動を模擬する動きを撮影したものであり、186組のペアのうち、グループであるペアとそうでないペアが93組ずつ含まれる。

5.2.2 実験結果

グループ検出精度の結果を図9、図10の表に表しており、ジェスチャー情報なしが佐藤らの手法にあたる。それぞれの表は陽性率とF値の値を計算したものとなつてゐる。また、図11は図10の表をグラフ化したものである。

これらを見ると、まず、閾値 A が小さいときは、ジェ

A	陽性率	ジェスチャー情報		ジェスチャー情報あり						
		なし		C=1 C=2 C=3 C=4 C=5 C=6 C=7						
		真	偽	94.6	94.6	94.6	93.5	91.3	90.3	89.2
1	真	94.6	5.4	94.6	6.5	6.5	7.6	19.4	26.9	28.0
	偽	5.4	94.6	6.5	6.5	7.6	19.4	26.9	28.0	30.2
2	真	88.1	1.1	87.0	1.1	2.2	6.5	14.0	19.4	20.5
	偽	1.1	88.1	1.1	2.2	6.5	14.0	19.4	20.5	22.6
3	真	73.1	1.1	73.1	75.2	82.7	87.0	86.0	82.7	82.7
	偽	1.1	73.1	75.2	82.7	87.0	86.0	82.7	82.7	82.7
4	真	69.8	0.0	69.8	1.1	2.2	3.3	11.9	16.2	17.3
	偽	0.0	69.8	1.1	2.2	3.3	11.9	16.2	17.3	17.3
5	真	67.7	0.0	67.7	72.0	79.5	86.0	84.9	82.7	82.7
	偽	0.0	67.7	72.0	79.5	86.0	84.9	82.7	82.7	82.7
6	真	65.5	0.0	65.5	66.6	70.9	77.4	83.8	84.9	82.7
	偽	0.0	65.5	66.6	70.9	77.4	83.8	84.9	82.7	82.7
7	真	64.5	0.0	64.5	66.6	70.9	76.3	83.8	84.9	82.7
	偽	0.0	64.5	66.6	70.9	76.3	83.8	84.9	82.7	82.7
8	真	64.5	0.0	64.5	66.6	70.9	75.2	82.7	84.9	82.7
	偽	0.0	64.5	66.6	70.9	75.2	82.7	84.9	82.7	82.7
9	真	63.4	0.0	63.4	65.5	69.8	75.2	82.7	83.8	81.7
	偽	0.0	63.4	65.5	69.8	75.2	82.7	83.8	81.7	81.7
10	真	63.4	0.0	63.4	65.5	68.8	75.2	82.7	83.8	81.7
	偽	0.0	63.4	65.5	68.8	75.2	82.7	83.8	81.7	81.7

図9 グループ検出精度表。真陽性率はグループを正しく検出した割合、偽陽性率は他人をグループと誤認識した割合を表す。

A	ジェスチャー情報	ジェスチャー情報あり						
		なし	C=1	C=2	C=3	C=4	C=5	C=6
1		0.946	0.941	0.941	0.930	0.872	0.832	0.822
2		0.932	0.926	0.914	0.881	0.866	0.838	0.827
3		0.840	0.840	0.854	0.880	0.876	0.851	0.828
4		0.823	0.818	0.847	0.884	0.870	0.845	0.828
5		0.808	0.808	0.832	0.871	0.870	0.845	0.828
6		0.792	0.800	0.830	0.857	0.857	0.845	0.828
7		0.784	0.800	0.830	0.850	0.857	0.845	0.828
8		0.784	0.800	0.830	0.843	0.851	0.845	0.828
9		0.776	0.792	0.823	0.843	0.851	0.839	0.822
10		0.776	0.792	0.815	0.843	0.851	0.839	0.822

図10 グループ検出精度表。各値はF値を表す。

スチャー情報を加えても検出精度が向上しなかつた。これは、もともと94%という高い精度のため、これ以上精度向上させる余地がなかつたものと考えられる。

一方、閾値 A が大きいときは、ジェスチャー情報を加えると検出精度が向上している。具体的に見ると、真陽性率、偽陽性率がともに上昇していることが分かる。このことから、ジェスチャー情報を加えるとグループらしいインスタンスの数を増加させていることが考えられる。

また、 C を変化させたときについて観察すると、値を上昇させるほど、すなわちジェスチャー情報の重みを強くすると、偽陽性率が増加していることが分かる。今回の定義したジェスチャー発生度合では、人物の重なり、歩行などの手の振りによるジェスチャー誤認識も多くあったため、他人をグループと識別してしまう傾向が見られたと考えられる。

図12はジェスチャー情報を用いたことでグループ検出できた例である。赤枠で囲った2人はグループであり、最初にジェスチャーを伴った意思疎通で二手に分かれる。そして、最後に再びジェスチャーによる意思疎通で同方向に進んでいく事例となっている。この例では、2人が同時に行動しているフレームが少なかつたこともあり、佐藤らの

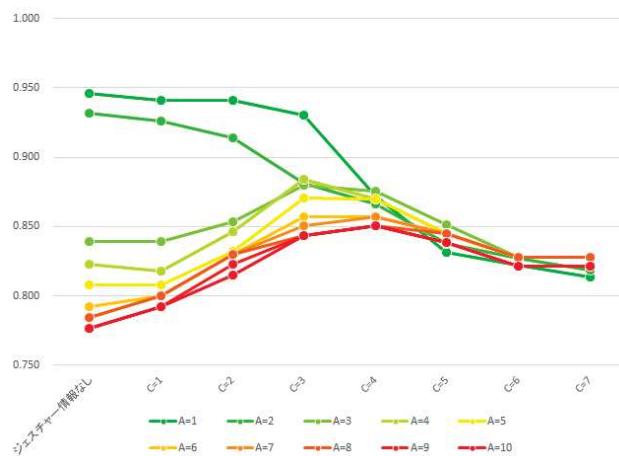


図 11 グループ検出精度グラフ。各閾値についてジェスチャー情報の有無や C の値を変更させたときの F 値の変化を表す。

手法ではグループとは捉えきれなかったが、本手法ではグループと認識することに成功している。実際に、この 2 人の特徴量について 3 つのインスタンスが存在していたが、佐藤らの手法ではすべてのインスタンスにグループらしいとラベルが付けられていた。しかし本手法では、1 つ目と 3 つ目のインスタンスにグループらしいとラベルが付けられており、最初と最後にジェスチャーを行ったという事実と一致していることを確認することができた。

6. おわりに

本研究では、グループ検出を行う際の対人距離以外の特徴量として、人物間に発生するインタラクションに着目した。そこでは、インタラクションの発生を特徴づける情報として、人物による身振り手振りなどのジェスチャーが発生したタイミングを用いることを考えた。具体的には、まず、ジェスチャーによる人物の見え方の変化をオプティカルフローによって捉え、ジェスチャー発生時はフローが大きく出る性質を利用して、ジェスチャー発生度合を表す特徴量を計算した。そして、2 人の人物において、少なくとも

片方の人物がジェスチャーを起こした場合、どちらもジェスチャーを起こしていない場合と比べて、2 人はグループである可能性が高いという考え方のもと、上記のジェスチャー度合を表す特徴量を用いて 2 人の間のジェスチャー情報を表す特徴量を記述し、既存の特徴量群に追加して機械学習によるグループ検出を行った。

評価実験では、ジェスチャー検出については、実際にジェスチャーが発生しているシーンにて定めた特徴量が高い値を示しており、うまくジェスチャー発生タイミングを捉えることができている例を確認できた。しかし、人物の重なりや、その他手の動きなどの影響でジェスチャーと誤認識をしてしまう例や、ジェスチャーが発生しているにも関わらずそれを捉えられない例があった。このことからジェスチャーを捉える特徴量の記述を改善することが今後の課題として挙げられる。

グループ検出については、既存の手法ではグループと捉えきれなかった事例でも、ジェスチャー情報を用いることにより検出に成功している例があることを確認することができた。また、ジェスチャー情報を用いても従来手法を上回らない結果も存在したが、上記のジェスチャー発生検出の改善を行うことで、提案手法の精度も改善し、従来手法を上回る可能性があると予測される。今後の課題としては、上記の通りジェスチャー発生検出の改善を行うことであるが、その他にも、ジェスチャー情報がグループ検出に与える影響度を確認することなど、結果に対する考察を進めることが挙げられる。

謝辞 本研究の一部は、科学技術振興機構「JST」の戦略的創造研究推進事業「CREST」における研究領域「共生社会に向けた人間調和型情報技術の構築」の研究課題「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」の支援により行った。また、本研究を行うにあたりデータ作成やプログラム開発に協力いただいた近畿大学の佐藤僚太氏に感謝する。

参考文献

- [1] 神田崇行, 石黒浩：“対話型ヒューマノイドロボットからの日常生活の中の友達関係の推定”，情報処理学会論文誌 vol.45, no8, pp.2098-2104, 2004.
- [2] 井上智雄, 瓶子和幸：“グループに適応する公共空間向け広告システム GAS”，情報処理学会論文誌 vol.49, no6, pp.1962-1971, 2008.
- [3] 岡本宏美, 西尾修一, 馬場口登, 森井藤樹, 萩田紀博：“移動軌跡を用いた歩行者間の人間関係の推定”，情報処理学会研究報告, CVIM, 第 2009 卷, pp.299-304, 2009.
- [4] 波部齊, 本田和久, 木戸正出繼：“歩行パターン変化に着目した人物間インタラクションの解析”，「画像の認識・理解シンポジウム (MIRU2009)」, 2009.
- [5] Isarun Chamveha, Yusuke Sugano, Yoichi Sato and Akihiro Sugimoto : “Social Group Discovery from Surveillance Videos: A Data-Driven Approach with Attention-Based Cues”, in Proc. British Machine Vision Conference (BMVC 2013), September 2013.

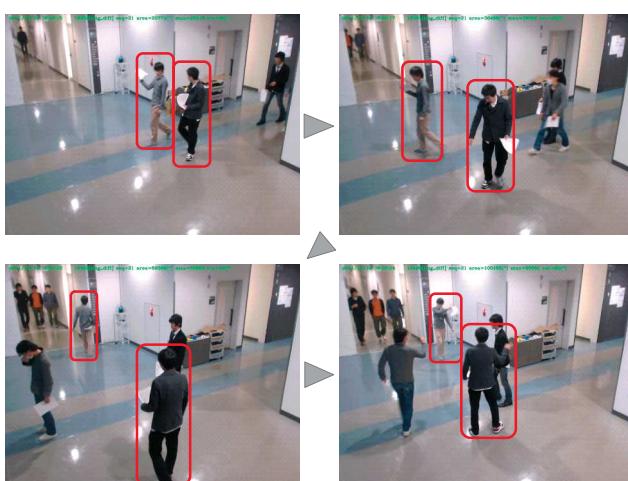


図 12 ジェスチャーによるグループ検出成功例

- [6] 佐藤 僚太, 酒井 美沙紀, 飯坂 瑞生, 波部 齊, 鶴見 和彦, 満上 育久, 八木 康史：“時系列上に局在する相互関係特徴を用いた歩行者グループの発見”, 「2015 年度 情報処理学会関西支部 支部大会」, 2015.
- [7] C. Liu. Beyond pixels: exploring new representations and applications for motion analysis. Doctoral Thesis. Massachusetts Institute of Technology. 2009.
- [8] Zanotto M., Bazzani L., Cristani M., Murino V. Online Bayesian Non-parametrics for Social Group Detection. BMVC 2012.
- [9] Solera F., Calderara S., Cucchiara R. Structured learning for detection of social groups in crowd. AVSS 2013.