ショートノート

人物のジェスチャーを加味した歩行者グループ検出†

波部　斉*1・橋本　知典*2・満上　育久*3・鷲見　和彦*4・八木　康史*3

環境に設置されたセンサで獲得した情報から，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，混雑していたり，グループが分かれて行動しているときなどでは十分ではない．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た歩行者グループ検出手法を提案する．提案手法では，防犯カメラで捉えた画像のオプティカルフローの変化でジェスチャーを検知し，人物間距離や視線方向にこの検知結果を加味してグループを検出する．実環境で収集したデータを用いた評価実験を行い，ジェスチャー情報を利用するとグループの見落としが低減できることを確認した．

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，ジェスチャー，機械学習，特徴抽出

1. はじめに

環境に設置されたセンサで獲得した情報から，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（図 1）．これまでに，歩行者グループ検出手法として提案されているもの [1–3] の多くは人物間距離を求めて距離が近いと同じグループに属すると判定する．しかし，混雑していてグループでない人でも距離が近づいてしまう場合や，グループが分かれて行動している場合では正しい判定ができない．

友人や知人同士で歩いているときお互いの意思疎通のためインタラクションが観察される．身振り手振りや，他人の呼びかけに反応する動作などが典型的な例である．これら人物間インタラクションを捉えられれば，人物間距離だけではグループの判定が難しい場合でも対処可能であると期待される．

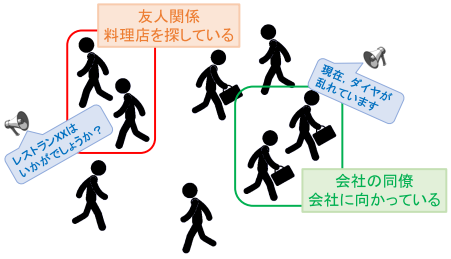


図 1　グループの目的に応じた情報発信

インタラクションに着目した例として，注意方向を利用した歩行者グループ検出手法が提案されている [4,5] が．本研究では，インタラクションをより積極的に表現するものとして，身振り手振りなどの相手に働きかける動作（以下，ジェスチャー）を考慮してグループ検出精度を向上させる手法を提案する（図 2）．そこではまず，映像から得られるオプティカルフローを用いてジェスチャーを検出する．環境に設置されたカメラで撮影した映像では人物は様々な向きをしているので，姿勢推定 [6] による身振りや手振りの検出が難しい場合がある．そこで，密なオプティカルフロー [7] の大きさ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」として定量的に表現する．

「ジェスチャー発生度」が大きい，つまり，ある人物が意思疎通のためのジェスチャーをしている可能性が高い場合は，シーンの中に意思疎通の対象となる仲間がいる可能性が高い．これを表現した特徴量を，移動軌跡と注意方向に基づく特徴量 [4,5] に加えて機械学習によるグループ検出を行った．

提案手法では，4 節で述べるようにレーザレンジセンサを用いて人物移動軌跡と注意方向（胸部方向で代用する）を獲得し，カメラを用いて人物の動きを捉える．提案手法ではこれらのデータはすべて正しく獲得できていると仮定する．

2. ジェスチャー発生度

図 3 はジェスチャーをしている人物とそこから算出したオプティカルフローの例である．オプティカルフローは色合いで方向を，濃淡でフローの大きさを示している．

手を振り上げる動作をしている (a) ではオプティカルフロー

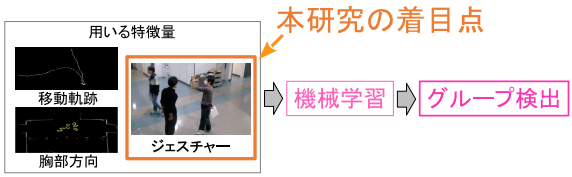


図 2　提案手法の概要

† Pedestrian Group Detection with Gesture Features
Hitoshi HABE, Tomonori HASHIMOTO, Ikuhisa MITSUGAMI, Kazuhiko SUMI, and Yasushi YAGI

*1 近畿大学理工学部
Faculty of Science Engineering, Kindai University
*2 大阪大学大学院情報科学研究科
Graduate School of Information Science, Osaka University
*3 大阪大学産業科学研究所
ISIR, Osaka University
*4 青山学院大学理工学部
College of Science and Engineering, Aoyama Gakuin University

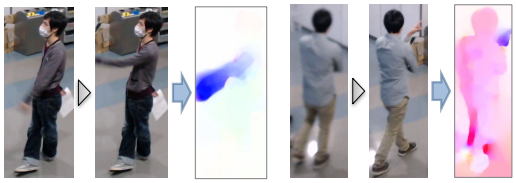


図３　ジェスチャーとオプティカルフロー算出結果の例

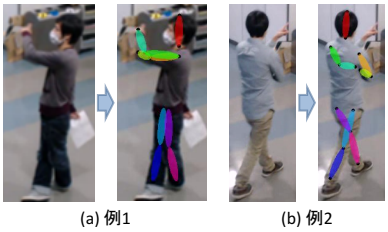


図４　Convolutional Pose Machines [6] による姿勢推定結果

でも腕部分で大きなフローが検出されている．(b) は人物が歩行しているため下半身で歩行方向のフローが検出されているが，上半身だけみると指さし動作の部分のフローの大きさが他と大きく異なっている．このように，ジェスチャーに対応するフローは他の部分と比べて大きくなる性質が見て取れる．

一方，このような画像に対して姿勢推定を行うことは簡単ではない，State-of-the-art として知られる Convolutional Pose Machines [6] を適用した結果を図 4 に示す．(a) は正しく推定できているが，(b) はカメラに対して背中を向けているため腕の姿勢推定に失敗している．このように，環境に設置されたカメラに対して人物は必ずしも正対していないため，姿勢推定が成功しない場合も多い．そこで本研究では，オプティカルフローを用いてジェスチャー発生の「確からしさ」を測る．そのための指標を以下で定義する．

ジェスチャーは手や腕を用いて行うため，ここでは上半身のフローのみに着目する．前節で示したように，ジェスチャー発生時は一部分で極端に大きなフローが得られる．そこで，人物領域内のフローの大きさの最大値 $V_{max}(t)$ と平均値 $V_{mean}(t)$ に着目し，その比をジェスチャー発生度合いを表す指標とする．

図 3 のように人物領域には背景が含まれるため，$V_{mean}(t)$ の計算時には前景のフローのみを考慮することにする．前景 P_{fg} は画素 p のうちフローの大きさ $v_p(t)$ がある閾値 T 以上であるものの集合とする．すなわち，以下のように定義する，

$$P_{fg} = \left\{ p | v_p(t) \geq T \right\}, \tag{1}$$

$$V_{mean}(t) = \frac{1}{\left|P_{fg}\right|} \sum_{p \in P_{fg}} v_p(t). \tag{2}$$

ここで $\left|P_{fg}\right|$ は P_{fg} の要素数である．なお，人物が止まっているときなどで前景画素が存在しない場合はジェスチャーは発生していないとする．これらをまとめ，最終的にジェスチャー発生度 $G(t)$ を次式で定義する．

$$G(t) = \begin{cases} 0, & \left(\left|P_{fg}\right| = 0\right) \\ \dfrac{V_{max}(t)}{V_{mean}(t)}. & (\text{otherwise}) \end{cases} \tag{3}$$

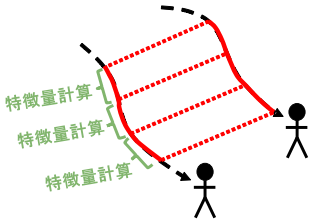


図５　時系列分割を用いた特徴量の計算

3. ジェスチャー発生度を用いたグループ検出

ジェスチャー発生度を用いたグループ検出について述べる．なお，本研究では先行研究 [4,5] に倣い，グループ検出問題をシーン中のある人物ペアが同じグループに属するかどうかを判定する問題に帰着させる．

3.1　グループ検出のための特徴量

[4,5] でグループ検出に用いた特徴量は，人物の移動軌跡と注視方向から計算される．前者からは人物間距離や歩行速度やその差が計算され，後者からは注視方向の一致度などが計算される．後者はインタラクションに着目しているといえる．[4] ではある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から特徴量を計算し，そのヒストグラムをグループ検出に利用している．しかし，実際のグループ行動ではグループらしい振る舞いは一部のみで見られ，時系列全体で特徴計算するとそれらを抽出できないことがある．[5] では時系列データを分割してそれぞれの区間において特徴量を計算する（図 5）．区間のうち 1 つ以上がグループらしいと判断されれば，全体をグループであるとみなすことでこの問題を解決している．

本研究では [5] の枠組みをベースとし，そこに 2 節で定義したジェスチャー発生度を加味して，ジェスチャー情報利用の有効性を確かめる．具体的には，[5] で用いられた人物移動軌跡から得られる特徴量 F_t^I と注視方向から得られる特徴量 F_i^I に加えて，ジェスチャー検出に基づく特徴量 F_g^I を導入する．学習や識別を行う際には，これら 3 つの特徴量を結合して，

$$F^I = \begin{pmatrix} F_t^I & F_i^I & F_g^I \end{pmatrix} \tag{4}$$

を小区間 I における特徴量とする．以下で F_g^I について述べる．

3.2　ジェスチャーに基づくグループ検出特徴量

2 節で求めたジェスチャー発生度はある瞬間でのジェスチャーを捉えるものである．これから図 5 で分割した小区間 I（以下では分割幅を L とする）における人物 i と人物 j の間の「グループらしさ」を示す特徴量 F_g^I を以下の手順で求める．

まず，次式によって小区間内でのジェスチャー発生度の最大値を求める（図 6）．

$$M^I = \max\{G_i(t), G_j(t)\}. \tag{5}$$

ここで，$G_i(t)$ は人物 i の時刻 t におけるジェスチャー発生度を指す．M^I は小区間 I において人物ペアがジェスチャーによるインタラクションをしている確からしさを示している．着目している人物ペアのうち一方がジェスチャーをしていれば，両者

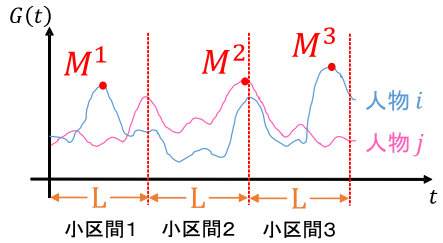


図 6　2 人の人物のジェスチャー発生度と M^I

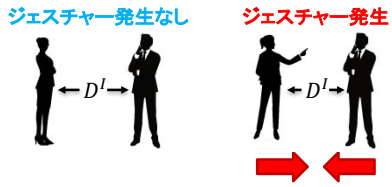


図 7　ジェスチャー補正距離

がグループである可能性が高くなるため，両者のジェスチャー発生度の最大値を求めている．

M^I だけでは本来グループでない場合でも大きくなることがあり得る．そこで，図 7 のように，物理的な人物間距離を M^I で補正して，ジェスチャーが発生している場合はより近い距離にいるとする仮想的な距離（ジェスチャー補正距離と呼ぶ）を次式のように導入する．

$$D_g^I = D^I - \lambda M^I. \tag{6}$$

ここで D^I は，小区間 I における 2 人の人物間距離の平均を表している．λ はこの 2 つの量のスケールを調整するための係数である[1]．このジェスチャー補正距離をグループ検出の特徴量として，すなわち，

$$F_g^I = \mu D_g^I, \tag{7}$$

として，式 (4) によって特徴量を求めてグループ検出を行う．なお，μ は他の特徴量と統合するときに，スケールの違いを調整するためのパラメータである．

4. 実験

2 節で導入したジェスチャー発生度の算出結果を示したあと，グループ検出（3 節）についての評価結果を示す．

4.1 データセット

評価のために被験者がグループ行動をする様子をセンサで記録したデータを用意した．被験者はあらかじめグループに分け，グループ以外の人物とは面識がないものとして行動する．行動の開始・終了の場所を指定して行動を開始するように指示し，終了までの一連の行動を一つのシーンとして記録する．環境中にはデジタルサイネージを設置しており，シーンによってはグループ構成員それぞれで異なる動作開始点を指定して，グループ内で様々なインタラクションが発生しやすいようにしている．ジェスチャーについては明示的な指示を与えておらず，

1　以下の実験では $\lambda = 1$ としている．

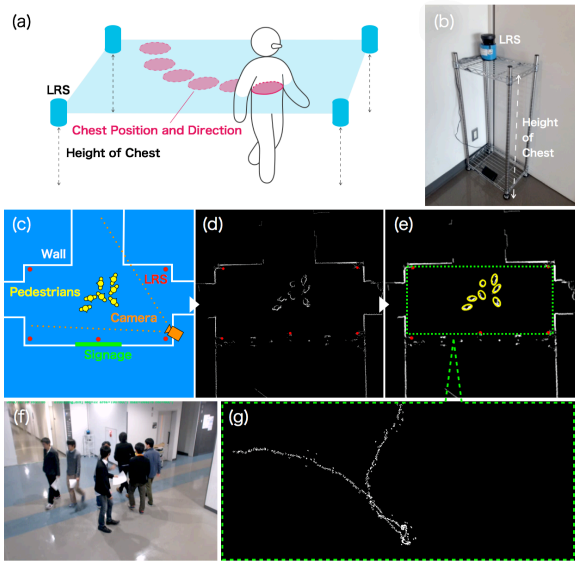


図 8　実験に用いる歩行者データ．(a) は歩行データ取得時の模式図．(b) はレーザレンジセンサ．(c) はデータ取得を行ったビルの廊下の T 字路を俯瞰した模式図．(d) はレーザレンジセンサの計測結果．(e) は (d) から推定された人物の胸部の位置と向き．(f) はデータ取得時の一場面．(g) はある人物の歩行軌跡．

シーン中のジェスチャーはすべて被験者が自発的に行ったものである．これによって，集団行動において自然に発生したジェスチャーの利用が集団検出の精度向上に寄与することを確認できると考えられる．

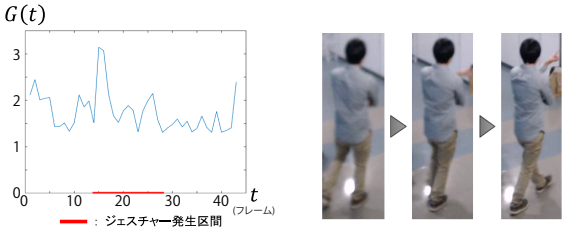
シーンの長さは行動によって異なるが概ね数分になり，あわせて 22 シーンを記録した．これらのシーンにはのべ 232 名の人物が出現している（重複を除いた実人数は 30 名である）．このうち，グループ行動をしているペアは 93 ペア含まれている．グループではないペアは無数の組み合わせが考えられるが，グループ行動するペアと同数の 93 ペアをランダムに抽出し，グループ検出の学習やテストに利用している．

歩行データは [5] と同様に，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を使用して 40 fps で取得されたものである．レーザレンジセンサで検出された胸部に楕円をあてはめ，人物位置および胸部方向を計測した（図 8）．これと環境中に設置された 1 台のカメラで撮影した映像（図 8(f)）から求まったジェスチャー発生度を合わせてグループ検出を行う．ジェスチャー発生度を求めるために必要な人物領域の切り出しは人の手で行った．

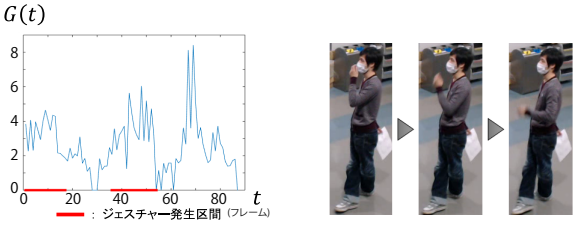
4.2 ジェスチャー発生度の算出結果

図 9 がジェスチャー発生度 $G(t)$（式 (3)）を算出した結果である．グラフの赤線はジェスチャーが実際に発生していた区間である．ジェスチャーの発生は 1 節の定義に従い人の目で確認した．図 9(a) では左手の指さし動作が発生したときに $G(t)$ が大きくなっている．一方，図 9(b) ではジェスチャー発生区間だけでなく，それ以外の区間でも $G(t)$ が大きくなっている．これは手を下ろす動作（図右）に反応しているものである．

このように，実シーンでは様々な見え方変化が発生するため

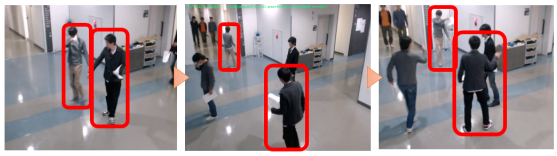


(a) 算出結果 1



(a) 算出結果 2

図 9　ジェスチャー発生度算出結果



(a) 見落としを低減できた例（グループと判定）

(b) 誤検出を低減できた例（グループでないと判定）

(c) 提案手法で誤検出した例（グループと判定）

図 10　検出結果の典型例

表 1　フレーム幅 $L = 100$ におけるグループ検出精度

	ジェスチャー情報なし	ジェスチャー情報あり			
		$\mu = 0.07$	$\mu = 0.08$	$\mu = 0.09$	$\mu = 0.1$
適合率	96.7	94.6	94.6	95.6	92.4
再現率	93.7	94.6	95.6	95.6	95.5
F 値	95.2	94.6	95.1	95.6	93.9

$G(t)$ が大きいときが常にジェスチャー発生に対応しているわけではない．しかし，防犯カメラ映像でのジェスチャー検出はそれ自体困難な問題であり，今回の目的はあくまでグループ検出の一つの手がかりとして利用することであるので，この段階での正確性はそれほど必要ではないと考えられる．

4.3　グループ検出精度の評価

分割フレーム幅 $L = 100$（2.5 秒）としてグループ検出精度を求めた．文献 [5] に倣って leave-one-sence-out によって精度を評価した．1 つのシーンに含まれる歩行者ペアから得られた特徴量をテスト用として抜き出し，残りのシーンの歩行者ペアの特徴量を学習データとして用いる．全てのペアのデータをテスト用として用いるまでこの手順を繰り返し検出精度を得た．その結果を表 1 に示す．この表はグループ検出結果の適合率，再現率と F 値の値を示したものとなっている．「ジェスチャー情報なし」の列がジェスチャーに基づく特徴量 F_g^I を用いない場合，「ジェスチャー情報あり」が F_g^I を用い，式 (6) における μ を変化させた場合の精度を示している．

これをみると，全体的に，ジェスチャー情報をみることで再現率が向上している．すなわち，従来の軌跡情報などでは捉えられなかったインタラクションを捉えることで，グループの見落としが少なくなっていることが確認できた．反面，適合率をみるとジェスチャー情報を加味することで精度が悪化している．これは，本来検出すべきでないものを検出していることになる．しかし，適合率と再現率を総合した F 値では，$\mu = 0.09$ のときにジェスチャー情報を用いた方がグループ検出の精度が良いことがわかる．このように，μ の選択が重要ではあるが，ジェスチャー情報の利用によってグループ検出精度が向上することが確認できた．

検出結果の典型例を図 10 に示す．ここでは，枠で囲んだ 2 名がグループであるかどうかを判定している．枠の色が正解を示し，赤はグループ，黄はグループでないと判定するのが正しい結果である．図 10(a) はジェスチャー情報なしでは見落としていたものをジェスチャー情報を加味して検出可能となった例である．逆に，図 10(b) は誤検出を低減できた例である．左端の画像の瞬間にたまたま近くにいたペアを，ジェスチャー情報なしでは誤検出していたが，ジェスチャー情報を加味することで相対的に人物間距離に対する判定条件が厳しくなり，正しい判定ができるようになった．最後に，図 10(c) は提案手法で誤検出が増加した例である．中央の画像で手前の人物がジェスチャーを行っており，そのときたまたま近くにいた人物との間のジェスチャー補正距離が近くなったためである．

5. おわりに

本研究では人物のジェスチャーに着目することでグループ検出の精度向上を図る手法を提案した．ジェスチャーには，身振り手振りなどで相手に働きかける動作が含まれており，人物間のインタラクションの発生を特徴づける情報として有用と考えられる．提案手法では，ジェスチャーによる画像上での人物の見え方の変化を捉え，ジェスチャー発生度が高い場合に人物間の距離が近くなるように補正した特徴量を導入した．

評価実験では，人物間距離や注意方向などのみの場合に比べ，ジェスチャー情報を用いることで一定の精度向上効果があることを確認した．今後の課題としては，ジェスチャーをより詳細に認識する手法の開発や，ジェスチャー発生タイミングとその前後の動作の時間的前後関係などを考慮することで，図 10(c) のような誤検出を低減させていくことが考えられる．

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 JP17K00256 の支援のもとに推進された．また，有益なご助言をいただいた奈良先端科学技術大学院大学 木戸出正繼名誉教授，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] 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，pp.299-304，2009.
[2] Zanotto, M., Bazzani, L., Cristani, M., Murino, V., “Online Bayesian Non-parametrics for Social Group Detection”, BMVC 2012, 2012.
[3] Solera. F., Calderara. S., Cucchiara. R., “Structured learning for detection of social groups in crowd”, AVSS 2013, 2013.
[4] Chamveha, I., Sugano, Y., Sato, Y., Sugimoto, A, “Social Group Discovery from Surveillance Videos: A Data-Driven Approach with Attention-Based Cues", BMVC 2013, 2013.
[5] 佐藤僚太，波部斉，満上育久，佐竹聡，鷲見和彦，八木康史，“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”，日本知能情報ファジイ学会誌，Vol. 28，No. 6，pp. 920-931，2016.
[6] Wei, S., Ramakrishna, V., Kanade, T., Sheikh, Y., “Convolutional Pose Machines", CVPR 2016, 2016.
[7] Liu, C., “Beyond pixels: exploring new representations and applications for motion analysis", Doctoral Thesis, Massachusetts Institute of Technology, 2009.

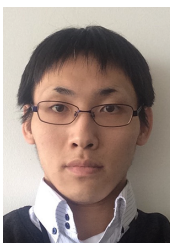
（2017 年 3 月 22 日　受付）
（2017 年 4 月 6 日　採録）

［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
E-mail: habe@kindai.ac.jp

著 者 紹 介

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．



橋本（はしもと）　知典（とものり）［非会員］

2014 年大阪大学基礎工学部卒業，2016 年大阪大学大学院情報科学研究科博士前期課程修了．在学中にコンピュータビジョンに関する研究に従事．

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，1984 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，2003-2006 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．

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，2003～2016 年同大学産業科学研究所教授．2008～2009 年同研究所所長補佐，2012～2015 年同研究所所長を経て，2015 年より同大学理事・副学長．1995～1996 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 FG1998（Financial Chair），OMINVIS2003（Organizing Chair），IEEE ROBIO2006（Program Co-chair），ACCV（2007 Program Chair, 2009 General Chair, 2010 Steering Committee Member），PSVIT2009（Financial Chair），ICRA2009（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 VRST2003 Honorable Mention Award，2006 IEEE ROBIO Finalist for T.J.Tarn Best Paper in Robotics，2008 年 IEEE ICRA2008 Finalist for 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）．

Pedestrian Group Detection with Gesture Features

by

Hitoshi HABE, Tomonori HASHIMOTO, Ikuhisa MITSUGAMI, Kazuhiko SUMI, and Yasushi YAGI

Abstract:

If groups of visitors in public spaces and commercial facilities can be detected, information depending on the attributes of the groups can be provided, 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, are useful for group detection and have been used in a number of works. However, if the scene is crowded or people in a group act separately, the features don't seem to work well. In this work, we focus on gestures, which indicate the interaction of people, and propose a group detection method using the information of gestures. Experimental results using dataset collected in an actual scene demonstrate gesture information improve the accuracy of group detection, especially the recall rate.

Keywords: human action analysis, group action analysis, range sensor, gesture, machine learning, feature extraction

Contact Address: **Hitoshi HABE**
Department of Informatics, Faculty of Science Engineering, Kindai University
3-4-1, Kowakae, Higashi-Osaka-Shi, Osaka 577-8502, Japan
E-mail: habe@kindai.ac.jp