

集団検出のためのジェスチャコミュニケーションの認識

山田花穂[†] 吉田武史[†] 鷲見和彦[†] 波部齊[‡] 満上育久[§]
 青山学院大学[†] 近畿大学[‡] 大阪大学[§]

1 はじめに

近年、駅や街角などの公共空間において有効な情報提供をするための技術の実現が期待されている。この技術には、空間内の人物の属性を理解し、発信する情報を選択する必要がある、そのためにはまず高精度な集団検出の実現が要求される。集団検出は従来から盛んに研究されており、例えば Chamveha らは、歩行者ペアの相対的な位置・頭の向きなどを特徴量として集団を検出する手法を提案している [1]。しかし、この方式は人物間距離の影響が支配的で、集団検出が不正確となる場合が多く存在する。そこで、インタラクションを特徴付けるジェスチャ検出と組み合わせることで集団検出の精度向上が期待される。ジェスチャ検出手法には、特徴点軌跡を用いた手法が多く提案されている [2]。これは入力動画の全域と時系列全体で特徴点を追跡していた。しかし実環境においてジェスチャは、歩行などの動作と同時に発生する場合や、途中で似たような動きが存在することが多い。

そこで本研究では、身体のパーツ検出手法を用いて局所的な領域の設定を行い、それを一定のフレーム間隔で時系列分割した後、各区間から特徴量を作成することでジェスチャを検出する手法を提案する。これにより、ジェスチャごとに必要なパーツ領域のみから特徴点の追跡ができ、またある区間に似ている動きが含まれていても他の区間でそれを補うことができる。

2 提案手法

本研究で提案する手法は図 1 に示す”領域探索”，”特徴点検出・追跡”，”時系列分割・特徴量抽出”，”動作識別”の 4 つの処理から成る。次項より各処理を詳しく解説する。

2.1 領域探索

領域探索処理には、Brandon らの身体のパーツ検出手法を用いる [3]。図 2(a) はその結果で、これにより頭・胴体・足のパーツの領域を設定する(図 2(b))。また、腕

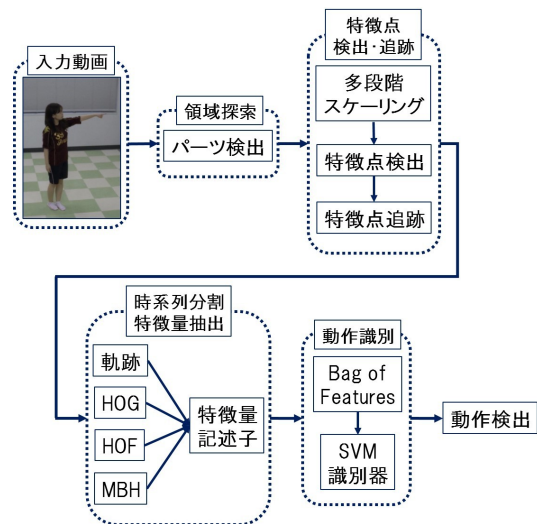


図 1 提案手法の概要

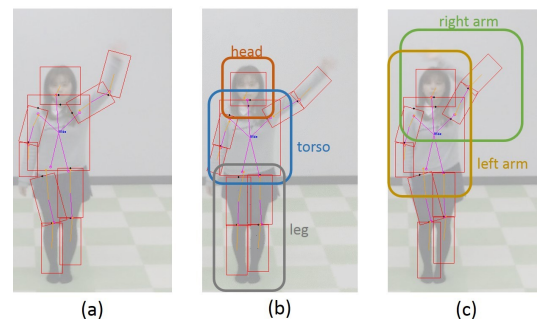


図 2 領域探索処理

は検出率が低いため、頭と腕の領域を合わせた広めの領域を設定する(図 2(c))。

2.2 特徴点検出・追跡

特徴点検出・追跡処理では Wang らの手法 [2] に倣って、多段階スケールリングした画像それぞれに DT (Dense Trajectory) という密な特徴点追跡を行い、各特徴点の移動軌跡を生成する。

2.3 時系列分割・特徴量抽出

時系列分割・特徴量抽出処理では DT で得られた軌跡データを、図 3 のように一定のフレーム間隔で時系列分割した後、Wang ら [2] に倣って、30 次元の Trajectory

Communicating Gesture Recognition For Group Detection
[†]Kaho Yamada [†]Takeshi Yoshida [†]Kazuhiro Sumi
[‡]Hitoshi Habe [§]Ikuhisa Mitsugami
[†]Aoyama Gakuin University [‡]Kinki University [§]Osaka University

と、その周辺領域から 96 次元の HOG (Histograms of Oriented Gradients), 108 次元の HOF (Histograms of Optical Flow), 192 次元の MBH (Motion Boundary Histograms) を特徴量として記述する. 最終的には追跡開始から終了までの 1 軌跡につき, これらの特徴量をすべて組み合わせた 426 次元の特徴量を作成する.

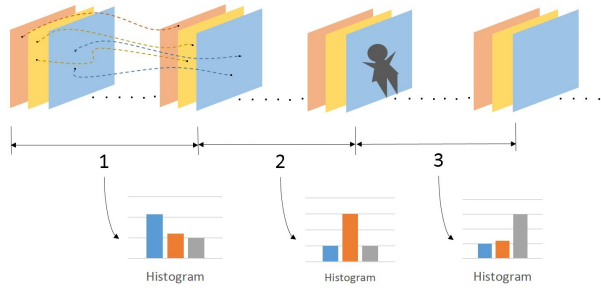


図 3 時系列分割処理

2.4 動作識別

動作識別処理では得られた特徴量を元に Bag-of-Features (BoF) 法と Support Vector Machine (SVM) により人物動作を識別する.

3 実験

本手法の有効性を評価するため, 従来手法 [2] に加えて従来手法に領域探索処理のみを加えた手法, 時系列分割処理のみを加えた手法で精度を比較した.

3.1 実験条件

検出するジェスチャには集団で行動している人物が頻繁に行うと考えられる「指をさす」, 「うなづく」, 「手をふる」の 3 種類とした. 用いるデータは, 監視カメラを想定した角度・距離から, GoPRO によって 1 人につき 3 種類のジェスチャを撮影したものであり, 各ジェスチャにつき学習用データに延べ 80 人, 評価用データに延べ 20 人の動きの情報を使用した. 実験に際し, 各ジェスチャの規定時間長として「指をさす」は「手を挙げ始めてから指をさすまでの間」, 「うなづく」は「頭を上下に 2 回振る間」, 「手をふる」は「手を左右に 2 回振る間」に設定した. また時系列分割数は 3 分割とした. この値は対象ジェスチャの特徴点軌跡に最も違いがでる最良の分割数と考え, 定めたものである. BoF 法のクラスタ数 k は 1,000 に設定した.

3.2 実験結果

実験結果を表 1 に示す. 表 1 より, 提案手法における正解率の大幅な向上が見られた. 「うなづく」は小さな動きであるため, 従来手法では頭以外の他の領域から動

きが見られると他のジェスチャと誤検出されやすかったが, 領域探索処理を加えることによって正解率が向上した. また「指をさす」は, 似ている動きが含まれる「手をふる」との誤検出が多かったが, 時系列分割処理を加えることによって 3 分割目の動きに違いが現れ, 正解率が向上した. よって本手法が有効であったといえる. しかし, 全体的に「指をさす」の正解率が他のジェスチャに比べて低く, その原因としては人によって指をさしている時間が異なることにある. これによって「指をさす」と「手をふる」の動きが大きく違う 3 分割目で, 指をさすまでの動きが終わり止まっている人物と, 指をさす途中でまだ動いている人物が現れ, 後者は「手をふる」と誤検出されてしまった. このような動きの個人差をいかに改善していくかが今後の課題となる.

表 1 実験結果

行動	従来	領域探索のみ	時系列分割のみ	提案
指をさす	45	50	65	65
うなづく	50	85	65	90
手をふる	65	55	80	70
全体	53.3	63.3	70	76.7

正解率 (%)

4 おわりに

本研究では, 集団検出を最終目標として, 身体のパーツ検出手法を用いて局所的な領域の設定を行い, それを一定のフレーム間隔で時系列分割した後, 各区間から特徴量を作成することでジェスチャを検出する手法を提案した. 評価実験の結果, 従来手法より正解率を向上させることができた. 今後の課題として, 動きの個人差の改良法の考案, ジェスチャ検出を特徴量として加えた集団検出が挙げられる.

本研究の一部は JST 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」の補助によって行われた.

参考文献

- [1] I.Chamveha, et al. Social group discovery from surveillance videos: A data-driven approach with attention-based cues. 2013.
- [2] H.Wang, et al. Dense trajectories and motion boundary descriptors for action recognition. *International journal of computer vision*, Vol. 103, No. 1, pp. 60–79, 2013.
- [3] B.Rothrock, et al. Integrating grammar and segmentation for human pose estimation. In *CVPR*, pp. 3214–3221. IEEE, 2013.