

報告番号	※甲	第	号
------	----	---	---

主 論 文 の 要 旨

論文題目 ハフ変換に基づく行動検出の頑健化に関する研究

氏 名 原 健 翔

論 文 内 容 の 要 旨

カメラを目としてコンピュータが実世界を理解するための技術が急速に進展している。実世界における重要な情報の一つとして、人の行動がある。コンピュータが人を支援したり、人とコミュニケーションを取ったりするためには、人が何をしているかを理解することは必要不可欠である。加えて、安価なカメラやスマートホンなどの普及もあり、映像というメディアは世界に溢れているものとなった。このような大量の映像を、内容を理解した上で整理したり特定のシーンの容易な検索を可能にしたりするためには、映像において主な被写体となる人の振る舞いを理解することが重要となる。

本研究では、コンピュータによる映像中の人物行動の理解を目標とする。これを実現するために、まず映像中の人物行動を検出するための技術に注目する。これは、映像中のいつ、どこで、何が起きたかを特定するための技術である。ここで、対象とした行動としては、「ボールを蹴る」、「走る」のように、「足を上げる」、「腕を振る」といった基本的な要素の組み合わせで表現可能な単純な行動である。

映像中の人物行動の検出手法には様々なものが存在する。その中でも、本研究ではハフ変換に基づく手法に焦点を当てている。ハフ変換に基づく手法は、映像から多数の局所特徴を抽出し、各局所特徴に基づいて行動のクラス、位置、スケールを推定し投票する。ここで、行動位置は行動の時空間を代表する点であり、一般的に空間的・時間的な中心で定義される。そして、投票がどのくらい集まっているかを表す投票スコアがしきい値以上となる点を行動として検出する。投票処理はそれぞれの局所特徴に対して独立なため、ハフ変換に基づく手法は、部分的な観測のみから行動を検出可能である。したがって、人物領域の一部が別の物体に隠れているオクルージョンに頑健ということや、行動の前半のみから早期検出が可能といった利点がある。

映像中の行動を高精度に検出するのは、様々な要因により困難な課題となる。例えば、オクルージョン、人物向きの変化、異なるクラス間での類似動作、行動の時間変動、行動の仕方の変動などがある。様々な環境下で安定して行動を検出するためには、このような要因に頑健な手法が望ましい。ハフ変換に基づく手法は元来オクルージョンに頑健という利点がある。そのため、本研究ではそれに加えて、人物向きの変化、異なるクラス間での類似動作、行動の時間変動、行動の仕方の変動の、四つの頑健性を実現するための手法を研究し、各手法の有効性を確認した。本論文の構成を以下に示す。

第一章では、まず映像中の行動検出に関する研究の背景と課題について述べる。そして、本論文の目的と位置付けを示すとともに、各章の位置付けについて説明する。

第二章では、関連研究を紹介する。行動の特徴表現手法や、行動検出におけるその他のアプローチについて述べる。

第三章では、ハフ変換に基づく行動検出の基本的なアルゴリズムを示し、本研究におけるベースラインとなる手法の実装について述べる。

第四章では、人物向きの変化への頑健性を実現することを目的とする。カメラに対する人物向きの変化に伴い行動の見え方は大きく変化する。これにより、学習時と検出時の対象の向きが大きく異なると、ハフ変換に基づく手法では行動の検出が困難になる。ここで、複数のカメラで同期撮影された多視点映像を利用することで、向きの違いを小さく保ち頑健な検出を実現できることが知られている。本研究では、ハフ変換に基づく手法を多視点映像へ適用する。これには、各視点で異なる座標系の変換による投票の統合が必要となる。本研究では、人は足元が地面に接することに着目し、行動の空間的な位置を人の足元に設定する。そして、各視点の地面の座標間をホモグラフィ変換することで、各視点の地面上にされた投票を統合する手法を提案する。評価実験から、提案手法は投票の統合により多視点の情報を有効に活用し、人物向きの変化に頑健な行動検出を実現したことが確認できた。

第五章では、異なるクラス間での類似動作に対して頑健な検出手法を提案する。例えば、「座る」と「ものを拾い上げる」には、体を屈める共通した類似動作が存在する。このような類似動作から抽出される局所特徴に基づく投票は、適切な行動のクラスや位置の推定が困難であり誤りやすいため、誤検出を引き起こす。本研究では、クラス間の関係により、このような誤投票は実際にされている行動のクラスごとに傾向が異なることに着目する。この誤投票の傾向を捉えるために、提案手法では投票分布という各クラスに対する投票スコアの分布を導入する。ここで、投票スコアとは投票がどの程度されているかを表す値である。誤投票のクラスごとの傾向を捉えた投票分布を学習したモデルを構築し、このモデルに基づくクラス確率を計算することで、誤投票の影響を抑制する。評価実験では、従来のハフ変換に

基づく手法と比較して、提案手法は類似動作により発生する誤検出を抑制することで検出精度を向上させることが確認できた。

第六章では、行動の時間変動の問題に注目する。同じクラスの行動であっても、例えば「蹴る」という行動で足を蹴り上げてから下ろすまでの時間が違うことがあるように、行動に時間的な変動が生じることがある。このような違いが学習時と検出時の行動の間に存在する場合、ハフ変換に基づく投票は時間方向にばらつき正しい位置に集中しないため、検出に失敗しやすい。本研究では、ばらついた投票を時間伸縮させることで一点に集中させる手法を提案する。提案手法では、時間変動による投票のばらつきがあっても、正解位置では投票スコアは極大点となることを仮定し、ばらついた投票を極大点に向けて時間伸縮する。これにより、時間変動による投票のばらつきの影響を抑制し、時間変動への頑健性を実現する。評価実験では、従来のハフ変換に基づく手法と比較して、投票の時間伸縮による集中による検出精度の向上が確認できた。

第七章では、各行動から抽出される局所特徴の数の変化に頑健な手法を提案する。オクルージョン、人物向きの変化、時間変動、行動の仕方の変動といった様々な要因により、行動を記述する特徴ベクトルだけでなく、行動から抽出される局所特徴の数も変化する。従来のハフ変換に基づく手法は、局所特徴数の変動がある場合に精度が低下するという問題がある。これは、投票スコアが局所特徴数に依存して変化することで、投票スコアに対する適切なしきい値の決定が困難になるためである。本研究では、二つの点からハフ変換に基づく手法を改良する。一つ目は、局所特徴の抽出処理である。これは、対象を探索するスケールによっても局所特徴数は変化するため、その影響を抑制するための手法である。提案手法では、スケールに応じて局所特徴をサンプリングすることで、局所特徴数を合わせる。二つ目は、検出におけるしきい値処理である。本研究では局所特徴数に応じて適切なしきい値を計算する手法を提案する。局所特徴数と投票スコアの関係をあらかじめ学習しておくことで、適切なしきい値の計算を可能とする。評価実験の結果から、従来のハフ変換に基づく手法と比較して、提案手法により局所特徴数の変化の影響を抑制し、行動検出の精度が向上することを確認できた。すなわち、上述した様々な要因への頑健性を、これまでの章とは異なる観点から実現したといえる。

第八章では、本論文を総括する。本研究で提案した手法を組み合わせることで、様々な要因に対する頑健性の実現が期待されることを述べ、今後の研究方向を展望する。