

研究終了報告書

「能動的分散協調視覚による群衆の3次元行動理解」

研究期間：2018年10月～2022年3月

研究者：延原 章平

1. 研究のねらい

ひとは雑踏のように多くの周辺人物が自由に行動する空間の中であっても自己位置を見失わず、また周辺人物の行動意図を理解しながら、衝突のないスムーズな移動を行うことができる。本研究のねらいは、群衆の中で人々と同じよう行動することができる、つまり実環境に参加できる自律エージェントを実現することであり、その応用によって、高齢者や視覚障がい者の移動支援、また防犯や警備など、労働人口の減少に伴う人手不足の解消への貢献を図る。

この目標を達成するうえで解くべき課題は、人物が密集した状況で生じる観測の不完全性への対処である。すなわち周辺人物が互いに遮蔽しあうことから、各個人の視点で撮影された映像から、周囲の人物の様子を完全に観測することは期待できない。本研究ではこのような観測が不完全になるという問題に対して、各個人視点の映像や、俯瞰支店の映像を統合してシステムが群衆の行動を予測し、不完全な情報を補うべく能動的に観測視点群を制御することによって、より行動理解に適した観測を行う能動的分散協調視覚の実現を目指す。

2. 研究成果

(1)概要

本研究では、自由に移動する人物にマウントされたカメラで撮影される個人視点映像と、環境に設置された固定カメラで撮影される俯瞰視点映像を入力として、雑踏のように群衆が入り乱れて移動するシーンにおける各人物の運動理解技術として、下記の成果を得た。

- 成果 1: 個人視点映像における周辺人物の見掛け上の移動量から、自身と周囲の相対移動軌跡を推定する技術。
- 成果 2: 移動する人物を複数の俯瞰視点で撮影することで、特別な校正物体を使用することなく俯瞰視点間の位置姿勢を自動的にキャリブレーションする技術。
- 成果 3: 机や椅子など一般物体を撮影した1枚の画像から、その3次元形状と撮影カメラの姿勢を推定する技術。
- 成果 4: 低解像度の映像に写された人物の視線方向を推定する技術。
- 成果 5: 同じ動作を繰り返し撮影した映像から、その動作の3次元運動を推定する技術。

これらのうち、成果1と2は個人視点・俯瞰視点の相対的な位置姿勢を、特別なキャリブレーション物体を用意することなく自動的に理解する技術であり、多視点カメラにより計測技術を導入する際に障壁となるキャリブレーション作業を不要とすることで、実社会への実装の助けとなることが期待できる。

また成果3によって、各視点の映像中で発見された机や椅子といった一般物体の3次元形状を推定すると同時に、その物体に対するカメラ位置姿勢を理解することができることで、移動経路上の障害物の有無や、移動可能範囲の推定を行うことが可能となる。

成果4と5は視界に入った人物の視線やジェスチャー認識を行う技術であり、成果3を移動環境の構造推定として、成果4と5を人物の行動意図推定として用いて、成果1を組み合わせることで、周辺人物の行動予測をより精緻なものとするのが期待できる。

(2) 詳細

本研究では、自由に移動する人物にマウントされたカメラで撮影される個人視点映像と、環境に設置された固定カメラで撮影される俯瞰視点映像を入力として、雑踏のように群衆が入り乱れて移動するシーンにおける各人物の運動理解技術として、下記の成果を得た(図1)。

- 成果 1: 個人視点映像における周辺人物の見掛け上の移動量から、自身と周囲の相対移動軌跡を推定する技術。
- 成果 2: 移動する人物を複数の俯瞰視点で撮影することで、特別な校正物体を使用することなく俯瞰視点間の位置姿勢を自動的にキャリブレーションする技術。
- 成果 3: 机や椅子など一般物体を撮影した1枚の画像から、その3次元形状と撮影カメラの姿勢を推定する技術。
- 成果 4: 低解像度の映像に写された人物の視線方向を推定する技術。
- 成果 5: 同じ動作を繰り返し撮影した映像から、その動作の3次元運動を推定する技術。

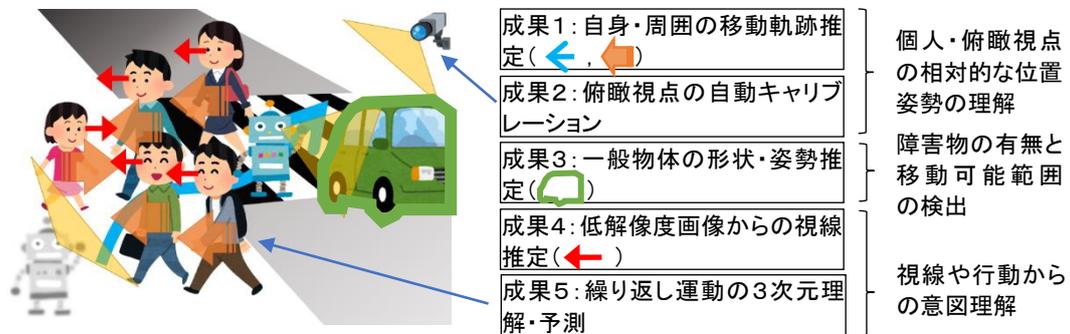


図1: 個人視点および俯瞰視点映像を用いた群衆の行動理解によって、群衆の中で人々と同じように行動することができる自律エージェントを実現する技術。

1 個人視点映像からの自身および周囲の相対移動軌跡推定



図2: 混合ガウス分布による3次元形状推定(左:入力画像, 右:各コンポーネント数 K における推定結果)

群衆の中を移動するためには、個人視点で撮影された映像をもとにして、自分自身の移動軌跡と周囲の人物の移動軌跡を理解・予測することができる必要がある。一般に、人やロボット、自動運転車両などのような移動体に取り付けられたカメラで得られる映像を入力として、自己位置を推定する問題は SLAM(simultaneous localization and mapping)として知られているが、従来の SLAM は個人視点映像中に映り込んだ静止物体を追跡することを前提としており、周囲を群衆に囲まれて静止物体を観測することができない状態では自己位置推定を行うことができなかった。

これに対して本研究では、図2に示すように、個人視点で撮影された映像を入力として、自己と周囲の移動軌跡の推定を同時に行い、あたかも上から見下ろしたような状況理解を実現している。この技術の要点は、自分自身も、周囲の人物も、ともに一定の行動モデルに沿って移動している群衆の一部であると仮定して、この行動モデルのもとでの次時刻における尤もらしい移動先の推定を逐次的に行うとともに、各時刻の観測によってその推定を修正する点にある。

2 人物観測による俯瞰視点の自動キャリブレーション

カメラを用いて計測を行うには、カメラを物理的に設置することに加えて、設置環境におけるカメラの位置姿勢を推定し、実世界との対応付けを行うカメラキャリブレーション作業が不可欠である。しかしその実施には特別な参照物体を用いた熟練者による作業が必要とされ、カメラシステム導入の妨げとなっていた。

これに対して本研究では、常に撮影され続ける被写体、つまり人体そのものや椅子などの一般物体を参照物体とする全自動カメラキャリブレーションを実現した。具体的には深層学習による視線方向・骨格姿勢推定、一般物体の位置姿勢推定を用いて、視点間の位置姿勢を推定すると同時に、逆に推定された多視点カメラ群の相対位置姿勢を用いて3次元人物骨格構造を三角測量によって計測し、これを新たな学習データとすることで深層学習モデルのドメイン適応も実現している。

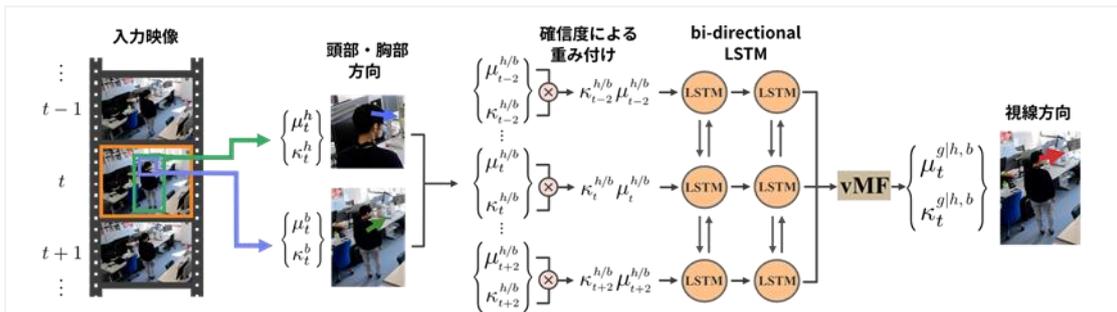


図3: 低解像度映像を入力とした視線方向推定 (左から入力フレーム, vMF 分布としての頭部・胸部方向推定, 視線方向推定)

3 一般物体の形状・姿勢推定

ひとが環境中を移動するときには、無意識に机や椅子など、その環境に存在する様々な物体の存在を検出するとともに、その形状を理解して衝突を避けるように適切な移動軌跡を設定している。このとき、ひとは必ずしも精緻な3次元形状理解を必要としておらず各物体のおおよその形状と姿勢を把握しているに過ぎない。

本研究ではこれと同様に、撮影された画像から被写体のおおよその形状を混合ガウス分布として推定すると同時に、その物体に対するカメラの姿勢推定を実現する(図3)。この技術によって、ステレオ法や LiDAR などによる詳細な3次元形状計測を経ることなく、撮影画像そのものから3次元形状を得ることができると同時に、混合ガウス分布という形状表現によって、コンポーネント数に応じた自然な LoD (Level-of-Detail) 表現を実現すると同時に、その物体からの距離の計算が容易となるため、前述の経路計画・衝突回避の実現を簡単化することができると期待される。

4 低解像度映像からの人物視線方向推定

群衆中を移動するひとは、視界に入った周囲の人物の視線方向を無意識に理解してその行動を予測し、衝突を回避できる適切な移動方向を決定している。またこのような移動方向予測に限らず、一般に視線はひとの行動予測や意図理解に有用な手がかりとされており、これまでも各種の視線計測装置を用いた行動解析がマーケティングなどの分野で実用化されている。しかし従来の視線計測は眼球の映像を高精細に撮影することができることを前提としており、通常の個人視点あるいは俯瞰視点から撮影された映像のように、眼球を判別することが困難であるような低解像度の映像を用いた視線推定は実現されていなかった。

本研究ではこのような低解像度の映像を入力とした、人物の視線方向推定を実現した。この技術の要点は、視線方向と人物頭部・胸部方向の間には時間遅れを伴う相関関係を利用する点にあり、この相関関係を深層学習によって自動的にモデル化することができる手法を考案している(図3)。本技術により、周辺人物の移動方向推定の予測に事前知識を与えることができるようになるだけでなく、複数の周辺人物の視線方向から三角測量することによって、環境中で視線が集中している領域を同定し、事故などの異常検出につなげることが期待できる。

5 繰り返し撮影した映像を入力とした3次元運動推定

一般に、ひとは歩行する、ものを投げるといった動作を一目見るだけでその3次元運動を理解し、かつケガや疲労などに起因するその微細な違いも認識することができる。これに対して現在の計算機では、モーションキャプチャのように三角測量による計測を行うか、あるいは深層学習によって映像からおおよその運動を推定することしかできない。

本研究では、ひとはその動作の理解を過去に繰り返し様々な方向から見た経験と照らし合わせることで実現しているのではないかという着想から、繰り返し行われた動作を時間的・空間的に異なる方向から観測することで、疑似的な三角測量を実現し、その3次元運動を推定することを可能にした。この技術により、単に深層学習によって3次元運動を推論するのではなく、三角測量としての裏付けを伴った推定が可能となり、3次元運動の微細な差の検出も可能となると考えられる。

3. 今後の展開

1 宅内環境における見守り支援

高齢者が在宅で自立した生活を続けることができる健康寿命を延伸することを目的として、宅内に設置したカメラによって日常動作に潜むわずかな異常を検出し、軽度認知障害の早期発見を実現する。このような技術の実現と社会実装には、①高齢者やその家族といった非熟練者による観測カメラの設置、②日常動作に潜む微細な行動の変化の検出、などが必要となるが、①は主に本研究の成果2で、また②は主に成果5によって実現が見込まれる。

2 自動運転支援

現在の自動運転技術は欧米のように歩車分離が比較的進んでいる環境が想定されており、日本のように時には雑踏の中を進行するような状況は想定されていない。また高速走行する車両が、前方の歩行者が道路の横断を考えているか否かを予測するためには、遠方を撮影した低解像度の画像から瞬時にその意図を推定しなくてはならない。これに対して本研究で得られた周辺人物の移動軌跡予測(成果1)、周辺物体形状の概形認識(成果3)、低解像度画像からの視線方向推定(成果4)は、いずれも自動運転車両から見た歩行者の行動理解・予測の手がかりとして活用することができる。

3 オンサイトマーカーレスモーションキャプチャ

本研究で得られた成果2による全自動カメラキャリブレーションは、被写体である人間を校正物体として利用していたが、この技術は人間に限らず、骨格構造を有して多関節剛体運動をする哺乳類や昆虫、爬虫類など、数多くの生物に対しても応用することができる。また成果4による体の部位間の連鎖運動のモデル化も、同様に様々な生物に対しても適用できる。つまり実験室だけでなく野外環境におけるオンサイトでのマーカーレスモーションキャプチャや動作解析の自動化が実現可能であり、人間であればリハビリやスポーツ解析、実験動物であれば様々な条件下での動作変化の解析、また野生動物であれば実世界での行動観察・解析などに応用することができる。

4. 自己評価

本研究では群衆の3次元行動理解を目的として、前述のように個人視点映像と俯瞰視点

映像を組み合わせた主に5つの成果を得ることができた。これらはいずれも群衆の移動予測など研究目標の実現につながる重要な要素技術であるといえる。しかし新型コロナウイルスの感染拡大によって、群衆を対象とした大規模な実験は実施することができなかった。このため、当初予定していたモバイル型の同期撮影装置の整備を想定していた研究費は、シミュレーション実験に用いる計算機環境の整備に充てた。

また本研究で得られた成果は、前述の項目3で挙げたように、高齢化社会における課題解決や、日本での自動車両運転の実現に向けた基盤技術としてだけでなく、医療やスポーツあるいは生物学のような他の学術分野の進展に寄与することが期待できる。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 3件

1. Soma Nonaka, Shohei Nobuhara, and Ko Nishino, "Dynamic 3D Gaze from Afar: Deep Gaze Estimation from Temporal Eye-Head-Body Coordination," in In Proc. of Conference on Computer Vision and Pattern Recognition, 2022.

自由に行動する人物を写した監視カメラ映像を入力として、その人物の3次元視線方向を推定する手法を提案した。具体的には頭部および胸部運動と視線方向の時間的依存関係を学習することで、映像から頭部・胸部方向を推定することができれば、眼球を直接計測できない場合であっても、視線方向推定が可能であることを実証した。

2. Keisuke Shibata, Sangeun Lee, Shohei Nobuhara, and Ko Nishino, "Consistent 3D Human Shape from Repeatable Action," in In Proc. of The 3rd International Workshop on Dynamic Scene Reconstruction (DynaVis), 2021. **Best Paper Award.**

ひとは投球動作などある運動を1方向から観測するだけでその3次元動的な動きを理解することができる。本研究はこれを過去に別方向から観測した似た運動との間でマッチングを行い、仮想的に多視点撮影を行うことで実現するものであり、別方向から撮影された映像同士の時間的・空間的キャリブレーションと、被写体の3次元形状復元を行う新たなアルゴリズムを提案している。

3. Kohei Yamashita, Shohei Nobuhara, and Ko Nishino, "3D-GMNet: Single-View 3D Shape Recovery as A Gaussian Mixture," in Proc. of British Machine Vision Conference, 2020.

ひとは1枚の画像から被写体の全周囲3次元形状をおおよそ推定することができる。本研究では同様に、1枚の画像から被写体のおおよその形状を混合ガウス分布として推定する手法を考案した。またこのような形状表現の応用例として、異なる視点から復元された同じ物体の混合ガウス分布を比較することで、視点間の相対姿勢を推定できることを実証した。

(2) 特許出願

研究期間全出願件数: 0 件(特許公開前のものも含む)

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- 山下浩平, 延原章平, 西野恒, “混合ガウス分布推定ネットワークを用いた単一画像からの 3 次元物体形状復元,” 情報処理学会コンピュータビジョンとイメージメディア研究会(CVIM), 2019-CVIM-217(19), May 2019. **情報処理学会 CVIM 研究会卒業論文セッション最優秀賞受賞**