

博士論文

行動特徴に基づく人物追跡

Exploiting Motion Characteristics to Track
Humans in Static and Dynamic Scenes



東京大学大学院
情報理工学系研究科
電子情報学専攻

48-077406 杉村大輔

指導教員 佐藤洋一 准教授

概要

近年の凶悪犯罪の増加に伴い、安心・安全な社会環境の実現のために監視カメラの普及が進んでいる。監視カメラの設置による犯罪の抑止効果は証明されつつあるが、その一方で、アルコール摂取に伴う事件などの衝動的犯罪には、その抑止効果は小さいことが報告されている。このような犯罪の防止、または事件の早期解決の実現には、観測視野内において何が起きているのかをシステムが自動的に理解することが重要となる。人間の行動をシステムが理解するためには、どこに人物がいてどのように動いているのかを認識する人物行動計測技術が必要である。そのため、人物行動計測技術の確立への期待は大きい。

このような社会背景のもと、本研究では、人間の行動計測において最も重要な役割を果たす人物追跡技術について研究を行う。

実環境において頑健に対象の人物を追跡するためには、複雑な背景、追跡対象の見えの変化、障害物による遮蔽など、行動計測の性能低下を引き起こす様々な問題へ対処しなければならない。本研究では、人物の行動特徴に基づく人物追跡手法を実現することにより、その問題へ対処する。

人間の行動特徴とは、行動に関する個人が持つ特性のことであり、これは個々の人物を追跡するための特徴的な指標と見なすことができる。本研究では、環境に基づく行動特徴、人物の個人性に基づく行動特徴の二つの行動特徴を考える。

環境に基づく行動特徴とは、環境により特徴付けられる人物の行動の偏りを意味する。対象環境における机や商品棚、ディスプレイなどの配置により、人物の行動しやすい領域に偏りが生じると考えられる。このような環境に即した行動の性質を利用することにより、人物追跡の安定化を実現することが可能となると考えられる。

これに対し個人性に基づく行動特徴とは、人物が生来持つ行動の癖に基づくもの意味する。一般に、他の人物には持ちえない個人特有の性質であるため、追跡対象の人物を他の人物と区別するための指標に利用できると考えられる。

本研究では、このような二つの人間の行動特徴を用いることにより、(1) 屋内環境における行動履歴に基づく人物追跡の安定化、(2) 歩容特徴に基づく混雑環境下人物追跡、(3) 一貫した人物追跡のための歩容特徴に基づく人物動線の対応付け、の三つ

の技術を開発する。

コンビニエンスストアやスーパーマーケットといった屋内環境を想定した場合、決まった通路の頻繁な通行、立ち読みのための本棚付近での滞留といった人物の行動は、対象空間の特定の領域で頻繁に観測される。このような人物の行動を長時間観測することにより、行動履歴に基づいた人物の存在確率分布を得ることができる。この人物存在確率（本研究では環境属性と呼ぶ）を、追跡安定化のための情報源（importance function）として時系列フィルタの一つであるパーティクルフィルタの枠組みに組み込む。これにより、安定な人物追跡、特に高速な追跡初期化機能を実現する。また、環境属性は、追跡器により毎フレーム推定される追跡結果を用いて逐次的に更新される。このような人物追跡と環境属性の構築の同時実行により、環境に適応した追跡安定化を実現できる。

行動履歴に基づく人物存在確率の利用により、環境に基づく行動特徴が顕著に表れる屋内環境において安定した人物追跡が可能となる。その一方で、一人一人の人物を頑健に追跡するためには、追跡対象と観測視野内に存在する他の人物とを区別しなければならない。これに対し本研究では、人物の個人性に基づく行動特徴を利用することによりこの問題の解決を図る。

観測視野内に多数の人物が存在するような混雑環境（朝のラッシュ時における駅の構内、イベント会場など）において、頻繁に発生する遮蔽や、複数の人物が非常に近接していることにより、個々の人物を正しく追跡することが難しい。そこで本研究では、人物の個人性にあたる歩容特徴と局所的な見えの時間変動の一貫性という二つの指標を、特徴点軌跡のクラスタリングに基づく追跡の枠組みへ組み入れた人物追跡手法を開発する。

周波数空間における歩容特徴は、生体認証の分野において頻繁に利用されている指標であり、個人を識別するための重要な手掛かりであることが知られている。また、局所領域における見えの時間的な変化は、人物の動きが周りと類似する傾向のある混雑環境下において個々の人物を区別するための効果的な指標となる。このような動きと見えの異なる種類の指標を利用することにより、混雑環境下において頑健な人物追跡を実現する。

次に本研究では、歩容特徴という個人性に基づいた行動特徴を利用することにより、人物追跡における大きな課題の一つである遮蔽の問題について対処する。

障害物や他の人物により追跡対象が遮蔽されるとき、対象の追跡は失敗する。これに伴い、人物の行動の軌跡（動線）はいくつかの動線片に分断されることが考えられる。観測視野内において一貫した人物追跡を実現するために、分断された行動

の動線片を対応付ける枠組みが必要となる。

これに対し本研究では，人物固有の特徴である歩容特徴に基づく人物動線の対応付けを実現する．歩容特徴は個々の人物特有の性質であるため，動線の対応付けにおいて効果的な指標になることが期待される．本手法では，移動する人物より得られる特徴点の動きの軌跡の集合から歩容特徴を抽出する．各特徴点の動きから得られる歩容特徴の周波数特性に基づき，Pyramid Match Kernel により対応付けるべき人物動線間の類似度を求める．このようにして得られる類似度の集合から最適化な組み合わせを探索することにより，人物動線の対応付けを実現する．

以上のように，本研究では，人間の行動特徴に基づく人物追跡技術を開発した．これにより，静的な屋内環境，動的に変動する混雑環境における頑健な人物追跡が可能となった．また人物の動線の対応付け手法を開発した．これにより，観測視野内における一貫した人物追跡が可能となる．

本研究により得られた成果は，非定常行動検出や行動認識といった，人物の行動理解へ向けた研究への活用が期待される．そのため本研究の貢献は大きい．

目次

第1章	序論	1
1.1	研究背景	1
1.2	本研究の目的	2
1.2.1	環境に基づく行動特徴の利用：行動履歴に基づく人物存在確率の利用による人物追跡の安定化	4
1.2.2	人物の個人性に基づく行動特徴の利用：歩容特徴と局所的な見えに基づく混雑環境下人物追跡	5
1.2.3	人物の個人性に基づく行動特徴の利用：歩容特徴を用いた人物動線の対応付け	6
1.3	本論文の構成	6
第2章	行動履歴に基づく人物存在確率の利用による人物追跡の安定化	7
2.1	はじめに	7
2.2	関連研究	9
2.2.1	観測モデルの高精度化	9
2.2.2	複数情報の利用による追跡初期化	11
2.2.3	本研究の位置付け	12
2.3	環境属性の人物頭部追跡への統合	14
2.3.1	パーティクルフィルタによる人物追跡	14
2.3.2	環境属性の統合	16
2.3.3	観測による人物頭部らしさの評価	18
2.4	環境属性の更新	20
2.4.1	環境属性の確率表現	20
2.4.2	追跡結果の動的反映による環境属性の更新	21
2.4.3	人物の行動状態推定に伴う行動履歴の適応的反映	23
2.5	実験・考察	26
2.5.1	環境属性の獲得	27

2.5.2	環境属性による追跡初期化の高速性	28
2.5.3	環境属性と一様分布を利用した場合の追跡精度の比較	31
2.5.4	遮蔽への対処	34
2.5.5	環境属性の更新に要する計算コスト	34
2.5.6	環境属性の利用の課題	36
2.6	本手法のまとめ	37
第3章	歩容特徴と局所的な見えに基づく混雑環境下人物追跡	39
3.1	はじめに	39
3.2	関連研究	41
3.2.1	人物の形状のモデル化	41
3.2.2	見えの情報の利用	42
3.2.3	群集の動きの流れの学習に基づく大多数人物追跡	43
3.2.4	特徴点軌跡のクラスタリング	44
3.2.5	本研究の位置付け	44
3.3	特徴点軌跡のクラスタリングに基づく人物追跡	45
3.3.1	特徴点軌跡の生成	45
3.3.2	特徴点軌跡群のグラフのクラスタリング	48
3.4	軌跡間の相違度の計算	50
3.4.1	周波数領域における歩容特徴	50
3.4.2	局所領域における見えの時間変動	53
3.4.3	軌跡間の空間関係	54
3.4.4	動きの一貫性	55
3.5	実験・考察	56
3.5.1	CG 実験	56
3.5.2	実環境下追跡実験	60
3.5.3	追跡失敗に関する考察	67
3.6	本手法のまとめ	69
第4章	歩容特徴を用いた人物動線の対応付け	71
4.1	はじめに	71
4.2	関連研究	73
4.2.1	人物動線の対応付け	73

4.2.2	視野を共有しない複数視野における人物の対応付け	75
4.2.3	歩容特徴の利用による個人認証	77
4.2.4	本手法の位置付け	78
4.3	マルコフ連鎖に基づく人物動線の対応付け	79
4.4	特徴点軌跡群を用いた人物動線片間の連結確率の計算	82
4.4.1	Pyramid Match Kernel	82
4.4.2	動線片間の連結確率の計算	87
4.5	実験・考察	92
4.5.1	実験結果	93
4.5.2	対応付け失敗に関する考察	100
4.6	本手法のまとめ	100
第5章	結論	101
5.1	本研究のまとめと貢献	101
5.1.1	行動履歴に基づく人物存在確率の利用による人物追跡の安定化	101
5.1.2	歩容特徴と局所的な見えに基づく混雑環境下人物追跡	103
5.1.3	歩容特徴を用いた人物動線の対応付け	105
5.2	将来展望	107
	謝辞	111
	参考文献	113
	発表文献	125

目次

1.1	環境に基づく行動特徴	4
1.2	人物の個人性に基づく行動特徴	5
2.1	人物追跡の難しさ	10
2.2	人物存在可能性の偏り	13
2.3	システム概要	15
2.4	確率密度分布の離散表現	16
2.5	パーティクルフィルタ	17
2.6	滞留行動による環境属性の局所化	24
2.7	行動状態遷移モデル	25
2.8	対象空間のレイアウト	27
2.9	頻繁に観測された行動	28
2.10	環境属性の構築結果（行動状態推定無し）	29
2.11	環境属性の構築結果（行動状態推定有り）	29
2.12	追跡初期化検証実験の評価例	30
2.13	人物頭部追跡結果	32
2.14	2名の人物頭部追跡の様子（一様分布の利用）	33
2.15	2名の人物頭部追跡の様子	33
2.16	完全遮蔽からの追跡再開	35
3.1	混雑環境の例	40
3.2	特徴点軌跡の生成	48
3.3	特徴点軌跡のクラスタリングによる人物追跡	49
3.4	ドローネー三角形分割によるグラフ構造	51
3.5	周波数領域における歩容特徴計測のための前処理	52
3.6	特徴点軌跡により張られる三角パッチ	54
3.7	特徴点軌跡間の空間関係と動きの一貫性	55
3.8	歩容特徴の利用の効果	58

3.9	平均 HS 色ヒストグラムの比較	59
3.10	ヒストグラム距離 $d_{hist}(\mathbf{h}(t), \bar{\mathbf{h}})$ の時間変動結果	59
3.11	シーケンス (A) における追跡結果比較	61
3.12	シーケンス (B) における追跡結果比較	62
3.13	シーケンス (C) における追跡結果比較	63
3.14	UCSD データ [74] における追跡結果比較	64
3.15	False positive と False negative	65
3.16	追跡結果 (再現率, 適合率)	67
3.17	追跡失敗例	68
4.1	遮蔽に伴う一貫した追跡の失敗	72
4.2	PMK による類似度計算 [29]	84
4.3	PMK [29] vs. VG-PMK [30]	85
4.4	VG ヒストグラムピラミッドの構築	86
4.5	主成分分析による周期運動の抽出	89
4.6	得られた周期成分の例	89
4.7	動きに基づく類似度計算	91
4.8	動線片の生成	93
4.9	シーケンス (1) の一部	94
4.10	歩容特徴を用いない場合の対応付け失敗例	94
4.11	シーケンス (1) における動線の対応付け結果 (歩容特徴の利用)	95
4.12	シーケンス (1) における動線の対応付け結果 (歩容特徴無し)	95
4.13	シーケンス (2)	97
4.14	シーケンス (2) における動線の対応付け結果 (歩容特徴の利用)	98
4.15	シーケンス (2) における動線の対応付け結果 (歩容特徴無し)	98

表目次

2.1	追跡初期化の速さの比較結果	31
2.2	追跡誤差	34
2.3	人物存在確率が低い領域における追跡初期化の速さの比較結果	37
3.1	実環境下追跡実験で用いた映像のパラメータ	60
3.2	実環境下追跡実験に用いたパラメータ	65
3.3	追跡結果の定量評価	66
3.4	追跡失敗率の詳細	68
4.1	シーケンス(1)における動線片対応付け結果	94
4.2	歩容特徴に基づく動線片間の類似度	96
4.3	シーケンス(2)における動線片対応付け結果	99

第1章 序論

1.1 研究背景

近年の凶悪犯罪の増加に伴い、安心・安全な社会環境の実現のために社会的な防犯意識が高まってきている。このような社会背景のもと、店舗や駅、街頭において、防犯を目的とした監視カメラの普及が進んでいる。監視カメラ大国として知られるイギリスでは、監視カメラの総設置台数は数百万台を超え、西ヨーロッパ諸国全体では、監視カメラの合計台数は一千万台にも及ぶ。アジア諸国においても三百万台以上の監視カメラの設置が確認されており、その普及率は目覚ましい。

監視カメラの設置の効果について広く調査研究がされている [28, 54, 102]。これらの調査によると、駐車場や病院を含む混合用途地区での防犯効果は大きく、特に車窃盗については顕著な犯罪件数の減少が確認されている。また防犯目的以外にも、凶悪犯罪の早期解決のための手掛かりとして監視カメラ映像は頻繁に利用されている。例えば、イギリスで2005年7月7日に起きた爆弾テロにおける犯人の迅速な検挙は、監視カメラの存在の寄与が大きいと報告されている。

このように、防犯や犯罪捜査など様々な場面において監視カメラは利用されており、その効果が実証されてきている。安心・安全な社会環境の実現のために、監視カメラの設置は必要不可欠になりつつある。

監視カメラを設置するだけでは、全ての種類の犯罪について防犯効果があるとは言い難い。監視カメラの設置についての調査結果をまとめている文献 [54] によると、車窃盗などの計画的犯罪に比べ、衝動的犯罪（アルコール摂取に伴う事件など）に関しては、犯罪の抑止効果が小さいことが報告されている。このような衝動的犯罪の防止、または事件の早期解決を実現するためには、監視カメラが観測している領域において何が起きているのかをシステムが自動的に理解することが重要となる。

しかしながら、監視カメラにより撮影される映像は、管理する人間の目により後日解析されるといった事後参照の用途が一般的である。そのため、衝動的犯罪のような現行犯の検知が重要になる場合に迅速な対応ができない。監視カメラの映像を常に人間が監視するようにすれば現行犯の検知も容易であると考えられるが、人件

費などの管理コストの面から現実的ではない。

以上のような様々な問題から、監視カメラの映像を用いた人間の行動の自動計測、理解の技術の確立が望まれている。ネットワークで接続された監視カメラシステムから得られる映像を、コンピュータによりリアルタイムで解析し、人間の行動を自動計測することが可能となれば、不審者の検出、警報の発令が可能となる。これは、対人暴行やテロなどの凶悪犯罪の防止に繋がることが考えられる。また凶悪犯罪の防止のための用途だけでなく、コンビニエンスストアやスーパーマーケットといった店舗内における万引き防止、スリの検出といった、セキュリティ分野全般への有効活用が期待される。

人間の行動計測技術は、セキュリティ分野以外にも様々な分野への応用が考えられる。一つに、駅のホームにおける事故の未然防止が考えられる。線路内への人物の落下の検出、ドア開閉時の安全確認など、その用途は多岐にわたる。これ以外にも、駅構内において道に迷っている利用客の検出、案内表示、混雑状況解消のためのラッシュ時の利用客の人数計測、利用客へのその情報提示などが考えられる。

マーケティングへの応用についても期待されている。店舗内において人物の行動を計測できれば、POSデータからでは知り得ない顧客の購買行動を知ることができる。また顧客の動きの履歴情報を収集することが可能になるため、顧客の目に留まりやすいような商品の配置方法への指針として履歴情報を利用することができる。これらの応用以外にも、人物の行動の履歴に基づいた照明器具や空調の適応的制御などへの利用も提案されている [20]。

このように、カメラ映像を用いた人物の行動計測技術は様々な産業分野への応用が期待されており、その社会的意義は大きい。

1.2 本研究の目的

観測視野内における人間の行動をシステムが理解するためには、どこに人物がいてどのように動いているのかを認識する人物行動計測技術が必要となる。そこで本研究では、人間の行動計測において最も重要な役割を果たす人物追跡技術について研究を行う。

実環境において頑健に対象の人物を追跡するためには、複雑な背景、追跡対象の見えの変化、机や棚などの障害物、他の人物による遮蔽など、行動計測の性能低下を引き起こす様々な問題へ対処しなければならない。これに対し本研究では、人間の行動特徴に基づく人物追跡を実現することにより、その問題へ対処する。

人間の行動特徴とは、行動に関する個人が持つ特性のことであり、これは個々の人物を追跡するための特徴的な指標と見なすことができる。本研究では、環境に基づく行動特徴、人物の個人性に基づく行動特徴の二つの行動特徴を用いる。これらを追跡の枠組みに利用することにより、頑健な人物追跡を実現する。

環境に基づく行動特徴

環境に基づく行動特徴とは、環境により特徴付けられる人物の行動の偏りを意味する。

対象環境における机や商品棚、ディスプレイなどの配置に伴い、人物の行動しやすい領域に偏りが生じると考えられる。例えば、弁当や飲み物などを購入しようとしている人物は、商品棚の周りに存在しやすいと考えられる。また雑誌の立ち読みを目的とする人物は、雑誌が陳列されている棚付近に長時間滞留しやすい(図 1.1)。

このような対象環境に即した人物の行動特性を利用することにより、人物追跡の安定化を実現することが可能となる。

人物の個人性に基づく行動特徴

環境に基づいた行動特徴がある一方で、個人性に基づく行動特徴が考えられる。

個人性とは、一般に指紋や顔、血管などその人物の身体的な特徴に基づくものが多いが、行動に関する特徴についてもいくつか報告されている。これは、人物が生来持つ行動の癖に基づくものを意味する(図 1.2)。文献 [97] にあるように、個人性が表れる行動特徴は様々なものが考えられるが、このうち、特に歩き方(歩容)などは人物追跡に密接に関わっていると考えられる。これらは一般に他の人物には持ちえないその個人特有の性質であるため、追跡対象の人物を他の人物から区別するための指標に利用できると考えられる。

本研究では、このような人物の行動特徴を用いることにより、(1) 屋内環境における行動履歴に基づく人物追跡の安定化、(2) 歩容特徴に基づく混雑環境下人物追跡、(3) 歩容特徴に基づく人物動線の対応付け、の三つの技術を開発する。



図 1.1: 環境に基づく行動特徴 .

1.2.1 環境に基づく行動特徴の利用：行動履歴に基づく人物存在確率 の利用による人物追跡の安定化

コンビニエンスストアやスーパーマーケットといった屋内環境を想定した場合、決まった通路の頻繁な通行、立ち読みのための本棚付近での滞留といった人物の行動は、対象空間の特定の領域で頻繁に観測される。このような人物の行動を長時間観測することにより、対象環境における行動履歴に基づいた人物の存在確率分布を得ることができる。この人物存在確率（本手法では環境属性と呼ぶ）を、追跡安定化のための情報源（importance function）として時系列フィルタの一つであるパーティクルフィルタの枠組みに組み込む。

本システムは人物追跡処理部と環境属性更新処理部の二つの処理機構が互いに作用しあうことにより実現される。人物追跡は、視野を共有する複数台のカメラを用い、観測視野内で移動する複数の人物の頭部を三次元的に追跡する。また環境属性は、追跡器により毎フレーム推定される人物三次元位置を用いて更新される。具体的には、環境属性を混合正規分布により表現し、オンライン EM アルゴリズムと呼ばれる最尤推定手法により逐次的にパラメータを更新することで実現する。これにより人物追跡の安定化、特に高速な追跡の再初期化を可能とする。

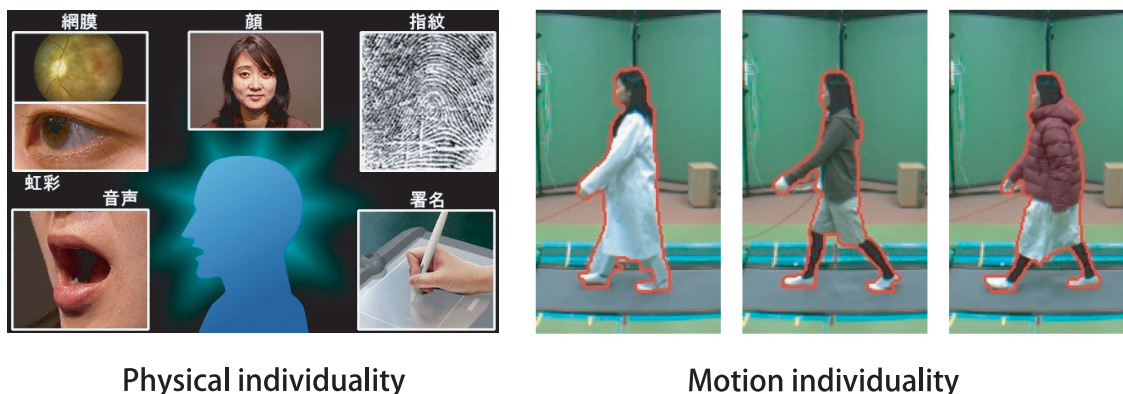


図 1.2: 人物の個人性に基づく行動特徴．指紋や顔などの身体的な特徴に基づくもの（左図 [40]）が多いが，歩き方（歩容）などの行動に関する癖にも個人性は表れる（右図 [35]）．

1.2.2 人物の個人性に基づく行動特徴の利用：歩容特徴と局所的な見えに基づく混雑環境下人物追跡

観測視野内に多数の人物が存在するような混雑環境（朝のラッシュ時における駅の構内，イベント会場など）において，頻繁に発生する遮蔽や，複数の人物が非常に近接していることにより，個々の人物を正しく追跡することが難しい．そこで本手法では，人物の個人性にあたる歩容特徴と局所的な見えの時間変動の一貫性という二つの指標を，特徴点軌跡のクラスタリングに基づく追跡の枠組みへ組み入れた人物追跡手法を開発する．

周波数空間における歩容特徴は，生体認証の分野において頻繁に利用されている指標であり，個人を識別するための重要な手掛かりであることが知られている．また，局所領域における見えの時間的な変化は，人物の動きが周りと類似する傾向のある混雑環境下において個々の人物を区別するための効果的な指標となる．

本手法は，同じ人物に属する特徴点の動きの軌跡の類似性に着目する．特徴点の動きの軌跡が類似した性質を持つならば，それは同じ人物に属している仮定する．このような仮定のもと，特徴点軌跡の性質に基づき特徴点軌跡をクラスタリングすることで，人物追跡を実現することが可能になる．特徴点軌跡のクラスタリングのための評価指標として，本研究の特長である歩容特徴の周波数特性，三本の特徴点軌跡により張られる局所三角パッチ内の見えの時間変動を利用する．これにより多

数の人物が存在するような混雑環境下においても頑健な人物追跡を実現することが可能となる。

1.2.3 人物の個人性に基づく行動特徴の利用：歩容特徴を用いた人物動線の対応付け

障害物や他の人物により追跡対象が遮蔽されるとき，対象の追跡は失敗する．これに伴い，人物の行動の軌跡（動線）はいくつかの動線片に分断されることが考えられる．観測視野内において一貫した人物追跡を実現するために，分断された動線片を対応付ける枠組みが必要となる．

これに対し本手法では，人物固有の性質である歩容特徴に基づく人物動線の対応付けを実現する．歩容特徴は個々の人物特有の性質であるため，人物の動線の対応付けにおいて効果的な指標になることが期待される．

具体的には，追跡対象である人物に属する特徴点の動きの軌跡群より得られる手掛かりを，動線の対応付けの指標として利用する．文献 [88] にあるように，特徴点軌跡群から人物の歩容に相当する周波数特徴を求める．そして特徴点軌跡群がそれぞれ持つ歩容特徴を多次元データのサンプル集合とみなし，特徴量集合間の類似度を測る Pyramid Match Kernel [29, 30] により動線間の類似度を計算する．このようにして計測される動線片間の類似度をもとに，二部グラフの最大重みマッチング手法の一つであるハンガリアン法を適用することで，最適な動線の対応付けを実現する．

1.3 本論文の構成

本論文は以下のような構成となっている．

第二章では，環境に基づく人物の行動特徴を利用した人物追跡手法について詳しく述べる．観測視野内を移動する人物の行動の履歴に基づいた人物存在確率を利用することにより，人物追跡の安定化を実現する．続く第三章では，人物の個人性に基づく行動特徴である，歩容特徴を利用した追跡手法について説明する．先に述べたように，本手法は大多数の人物が存在する混雑シーンについて対処する．そして第四章では，個人性に基づく行動特徴を用いた人物動線の対応付け手法について詳しく述べる．最後に，第五章において本研究の全体を総括し，残された課題，今後の展望について議論し，まとめる．

第2章 行動履歴に基づく人物存在確率 の利用による人物追跡の安定化

コンビニエンスストアやスーパーマーケットといった屋内環境を想定した場合、決まった通路の頻繁な通行、立ち読みのための本棚付近での滞留といった人物の行動は、対象空間の特定の領域で頻繁に観測される。このような人物の行動を長時間観測することにより、行動履歴に基づいた人物の存在確率分布を得ることができる。本研究では、この人物存在確率を人物追跡の枠組みに組み込むことにより、安定な人物追跡、特に高速な追跡初期化を実現する。

2.1 はじめに

コンピュータの高速化やカメラの小型化、低価格化に伴い、カメラ映像を用いた応用技術が注目されている。なかでも人物追跡技術は、防災・防犯を目的とするセキュリティ分野、効果的な店内レイアウトの設計に向けたマーケティング分野など、様々な分野から大きな期待が寄せられている。しかしながら、実環境においては、障害物や人物相互による遮蔽、複雑背景、見えの変化などの様々な問題があり、いかなる状況においても安定に人物追跡を実現することは難しい。

カメラ映像を用いた人物追跡技術は、これまでに多くの手法が提案されているが、なかでも時系列フィルタの一つであるパーティクルフィルタの有用性が数多く報告されている（[41,42]など）。パーティクルフィルタによる人物追跡は、追跡対象を表す確率密度分布を状態量と尤度を持つ多数の仮説群を用いて離散的に近似し、事前の状態と状態遷移モデルに基づく“予測”，観測画像を用いて人物らしさを評価する“観測”の処理を逐次的に処理することで実現される。動きの変動や観測のノイズ、特に観測値が非ガウス型になるような場合においても頑健な追跡を実現できることから、実環境下における追跡手法として広く利用されている。

しかしながら、上述した問題から実環境において追跡の失敗を防ぐことは難しい。

例えば、追跡している人物が障害物によって完全に遮蔽され、そして遮蔽された場所とは異なる位置から再度現れる場合、システムは追跡していた人物を見失ってしまうため、正しい追跡を直ちに再開することは困難である。

このような問題へ対処するために、追跡が不安定になる、もしくは失敗した際に、修正、再初期化を速やかに行うことで正しい追跡を再開する手法がいくつか提案されている（[42,46,70] など）。これらの手法は、観測画像から得られる指標のうち、追跡に利用する指標とは異なるものを追跡の初期化用として利用することで、追跡失敗の問題に対処している。

一方、鈴木らは、画像から得られる指標に加え、環境モデルと呼ばれる対象空間内の人物の存在確率分布を導入することで追跡の安定化を図る手法を提案している [91]。彼らは、対象空間内の机や椅子などの物理的障害物からの距離や配置に基づき環境モデルを定義している。これにより、壁、床の下など人物が存在しえない領域に仮説が発生することを抑制することができるため、安定な追跡が実現される。しかしながら、この手法は予め環境モデルが獲得されていない場合には使用できないという問題がある。また、環境の物理的な制約だけでは対象空間内の人物の存在確率分布を必ずしも表現できているとは言えない。

例えば、対象空間内に二つの通路があり、そのうち一つの通路のみが頻繁に使われるシーンを考える。この場合、物理的な制約のみに基づいて人物存在確率を表現すると、両者はほぼ等しい確率となってしまうが、実際には人物が頻繁に通行する通路の方が人物の存在確率は高いと考えられる。これは、人物の存在確率は物理的制約よりも対象空間中の人物の行動履歴に強く依存することを示唆している。

このような人物の行動の履歴に着目し、それを利用することで、対象空間に特化した人物存在確率分布を表現することが可能となる。そしてこれを人物追跡の初期化のための情報源として利用することで、対象環境に適応した高速、高精度な追跡の再初期化を実現することができると考えられる。

このような知見に基づき、本手法では、人物の行動履歴に基づいて対象空間の人物存在確率分布を獲得し、追跡の枠組みに組み込むことで、人物追跡の安定化を実現する。以後本手法では、このような人物存在確率分布を“環境属性”と呼ぶ。

本手法は具体的には次のように実現される。ICONDENSATION [42] の考え方に基づき、環境属性を importance function としてパーティクルフィルタの追跡の枠組みに組み込む。環境属性は混合正規分布でモデル化し、オンライン EM アルゴリズム [79] を用いることで、追跡器により毎フレーム推定される人物頭部位置から逐次的に学習させる。これより、対象シーンに特化した適応的な環境属性の獲得と追跡

の安定化を同時に実現することが可能となる。

本章の構成は以下の通りである。まず、2.2 節において関連研究についてまとめると共に、本手法の位置付けを確認する。続く 2.3 節では、環境属性の人物追跡への統合方法について述べる。そして 2.4 節において環境属性の構築法について詳しく述べる。2.5 節では本手法の有効性、適用範囲について実験を通して考察し、最後に 2.6 節において本手法についてまとめる。

2.2 関連研究

観測シーンの影響（図 2.1）に対して頑健な人物追跡を実現するために、パーティクルフィルタを利用した追跡手法が数多く提案されている。本節では、これらの先行研究について概観すると共に、本手法の位置付けについて確認する。

2.2.1 観測モデルの高精度化

観測画像から人物らしさを評価する時系列フィルタの観測モデルを高精度化することにより、頑健な人物追跡を実現することが可能となる。

複数情報の統合

観測画像から得られる特徴量のうち、人物らしさを評価する特徴量として用いられるものは、色情報、エッジ情報、背景差分による前景領域などが挙げられる。頑健な人物追跡を実現するために、一般にこれらの特徴量を複数組み合わせるアプローチがとられている（[8, 43, 61, 69, 82, 87, 91, 96] など）。

Yang らは、複数の画像特徴量を階層的に評価することにより高速な人物追跡を実現している [8]。Coarse-to-Fine カスケードの考え方により、各特徴量を段階的に評価する。明らかに評価値が小さくなるものは、初期のステージにおける単純かつ高速な評価器により棄却され、尤もらしいものは後段のより高精度な評価器によって評価される。このような段階的な処理により、計算コストのかかる後段の高精度な仮説の評価を必要最低限に抑えることができるため、高速な処理が実現される。

複数の画像特徴量を利用する一方で、複数視点の観測を組み合わせた手法も提案されている（[64, 65] など）。対象環境を複数台のカメラで観測することにより、カメラ一台では観測できない領域において人物追跡を行うことができる。これにより、



図 2.1: 人物追跡の難しさ .

遮蔽による追跡失敗を抑制することができる。また複数台のカメラを用いたシステムでは、計算量の面から一台の PC での処理に限界があるため、統合処理のためのサーバ PC と、カメラ一台に対して一台のクライアント PC を割り当てた分散システムを構築する方法が一般にとられる。

識別器の尤度評価への利用

色情報やエッジ情報などの比較的単純な画像特徴を利用した追跡手法は、実環境において必ずしも十分ではない。そこで小林らは、カスケード型 Adaboost を用いた顔識別器 [99] を仮説の人物頭部らしさの評価に利用した追跡手法を提案している [51,53]。人物頭部を楕円でモデル化し、観測される画像中の対象領域に顔識別器を適用する。そして、この識別器のカスケードを通過した段数を人物らしさ評価値とする。これにより頑健な追跡が実現できる。更に、人物頭部の向き毎に学習を行った顔識別器をそれぞれ用意し、顔の向きに伴い識別器を選択的に利用することにより、人物頭部の向きの不定性による見えの変化に対処している。

同様に、Chateau らもまた識別器をパーティクルフィルタの枠組みに組み入れた手法を提案している [11]。彼らは、識別器による評価値を Sigmoid 関数を用いて射影

することにより，人物らしさを表す尤度として追跡に利用している．

環境の物理的な制約の利用

追跡対象の確率密度分布が正確に得られない場合，パーティクルフィルタの性質上，サンプリングした仮説が壁の中や床の下など，人物が存在しえない場所に発生することがある．このような領域に発生する仮説は，人物追跡において意味のない情報に他ならず，出来る限りこのような位置に仮説が発生することを防ぐ必要がある．

そこで鈴木らは，画像情報以外に，レンジセンサによって得られた室内の三次元形状から空間的な人物の存在可能性を定義し，それを追跡に利用することで追跡の安定化を図っている [91]．彼らは，椅子，棚，机などの障害物の位置関係から対象空間をいくつかの領域に分割し，それぞれの領域に人物の存在可能性に基づく重みを設定している．この環境による重みを，観測画像から得られる手がかりに統合することで，仮説の無駄打ちを抑制した安定な追跡を実現している．

2.2.2 複数情報の利用による追跡初期化

環境の影響に起因した追跡の不安定化，失敗は一般に防ぐことが難しい．このような問題へ対処するために，追跡とは異なる情報源を利用し，追跡の安定化・初期化を実現する手法が提案されている．

よく知られている手法に ICONDENSATION と呼ばれるものがある [42]．パーティクルフィルタの追跡の枠組みに，Importance Sampling を組み合わせることで，不安定な追跡の修正，もしくは追跡の初期化を実現する．

Importance Sampling とは，事前分布が複雑な形状を持つ場合や，正確に得られていない場合に，事後分布について重要な情報を含んでいると考えられる別の情報源 (importance function と呼ばれる) から仮説をサンプリングすることで，事後確率の分布の推定の不安定化を防ぐ手法である．これをパーティクルフィルタの枠組みに適用することで，追跡性能を高めることが可能となる．具体的には，ICONDENSATION は，(1) Importance Sampling による追跡初期化，(2) Importane Sampling による追跡修正，(3) パーティクルフィルタによる追跡，の三つの機能を持つ．これらの処理を同時に利用することにより，追跡失敗の回復 (新規対象のための) 追跡初期化，追跡の安定化を実現することが可能となる．

Importance function は、任意の形状の確率密度分布を表現できるが、仮説をサンプリングしやすい分布である必要がある。主な理由は、追跡の処理速度を保つためである。過去の事後分布を元に事前分布を予測し、仮説をサンプリングする時系列フィルタリングとは異なり、Importance Sampling は観測された情報から仮説をサンプリングする。そのため、分布の形状によってはサンプリングに時間がかかることが懸念される。このような理由から、多くの手法では正規分布、もしくは混合正規分布が用いられることが多い。

Importance Sampling の利用による追跡の安定化について、近年様々な手法が提案されている。Jin らは、パーティクルフィルタによる追跡と同時に、背景差分、エッジ検出を組み合わせた頭部検出器を用い、視野内に観測される人物頭部の検出処理を行っている [46]。パーティクルフィルタの事前分布からのサンプリングと検出位置からのサンプリングを併用することで、検出結果を追跡に反映した頑健な追跡を実現している。また Wang らは、三次元人物追跡を行うと同時に、複数台のカメラそれぞれにおいて人物検出を行い、検出結果を追跡初期化のための情報源として利用している [101]。この追跡初期化処理において、検出器による仮説のサンプリングは各カメラの観測にのみ依存する。人物が観測されたカメラ画像の検出結果から仮説がサンプリングされる。

このような Importance Sampling の考え方を基に、物体検出の技術を複数対象追跡の枠組みに組み入れた Boosted Particle Filter と呼ばれる手法がよく知られている [70]。彼らは、カスケード型 Adaboost 識別器 [99] に基づく人物検出を確率的に取り扱い、importance function としてパーティクルフィルタの枠組みに組み入れている。Mixture Particle Filter [98] の考え方にに基づき、事前確率と検出器による初期化用の確率分布の混合分布を利用することで、追跡初期化、追跡の頑健化を同時に実現している。また、これにより観測視野内における対象の数に依らない柔軟な複数対象追跡が実現される。

2.2.3 本研究の位置付け

複雑な背景、追跡対象の見えの変化、照明状況の変化、遮蔽といった観測環境における様々な問題へ対処するために、数多くの研究が行われている。本節では、観測モデルの高精度化 (2.2.1 節)、複数情報の利用による追跡初期化 (2.2.2 節) の二つのアプローチに分類し、それぞれについて紹介した。

2.2.1 節において概観したように、複数特徴量の利用、複数視点からの観測や、識

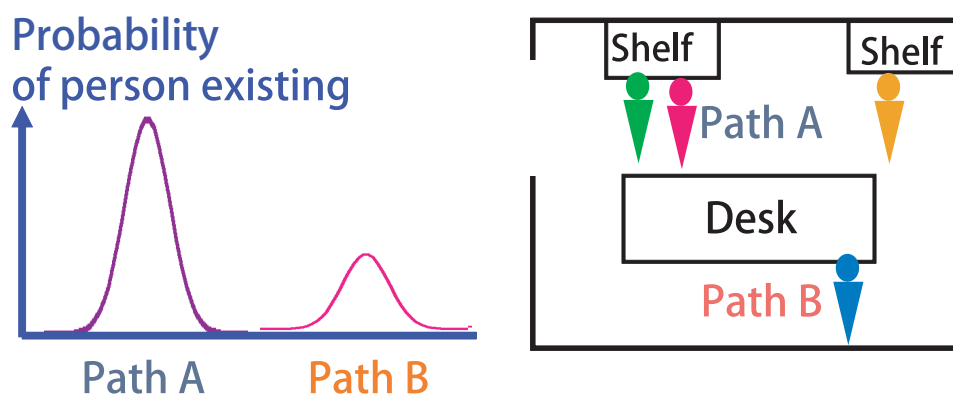


図 2.2: 人物存在可能性の偏り．コンビニなどの店舗内を対象空間とした場合，立ち読みをするために本棚周辺において数人の人が滞留することがある．このような人物の行動に伴い人物の存在可能性は大きく偏ると見なすことができる．

別器を用いることにより，頑健な人物追跡が実現されつつある．また，遮蔽などなんらかの要因により追跡が失敗した場合においても，2.2.2 節で述べた Importance Sampling の併用に基ついた追跡手法により，追跡失敗からの速やかな復旧が可能となる．

しかしながら，これらの追跡初期化は，あくまで観測画像に基づき実現しているため，三次元人物追跡を考えると必ずしも十分ではない．

一方，鈴木らは，画像から得られる指標に加え，環境モデルと呼ばれる対象空間中の人物の存在確率分布を導入することで追跡の安定化を図っている．このような観測画像に依らない三次元的な事前知識を追跡に組み込むことにより，頑健な三次元追跡が実現できる．

しかしながら，この手法はあらかじめ環境モデルが獲得されていない場合には使用できないという問題がある．また対象環境の物理的な制約だけでは対象空間中の人物の存在確率分布を表現できていないとは必ずしもいえない．例えば，対象空間内に二つの通路があり，そのうち一つの通路のみが頻繁に使われるシーン（図 2.2）を考えると，明らかに人物が頻繁に通行する通路の方が人物の存在確率は高いと考えられる．このような人物の行動履歴に基づいた人物存在確率を定義することにより，対象空間に特化した人物存在確率を得ることができると考えられる．

このような知見に基づき，本手法では人物の行動履歴に基づく対象空間の人物存在確率分布（環境属性）を用いた人物三次元追跡の安定化を実現する．ICONDEN-

SATION の考え方に基づき，環境属性を importance function としてパーティクルフィルタの枠組みに組み込むことにより，追跡の安定化，特に柔軟な追跡初期化を実現する．

2.3 環境属性の人物頭部追跡への統合

図 2.3 に本システムの全体図を示す．本手法は“人物頭部の三次元追跡”と“環境属性の獲得”の二つの処理単位が相互に作用することで実現される．これにより人物追跡の安定化と環境属性の獲得を同時に実現することが可能となる．以下，本節では，“人物頭部の三次元追跡”の処理単位について述べる．

2.3.1 パーティクルフィルタによる人物追跡

時刻 t における対象の状態量を \mathbf{X}_t ，時刻 t までに得られた画像による観測列を $\mathcal{Z}_t = \{z_1, \dots, z_t\}$ とする．時系列フィルタによる人物追跡は，追跡対象を表す密度分布 $p(\mathbf{X}_t | \mathcal{Z}_t)$ を逐次推定することであり，それは，追跡対象の事前の状態に基づく“予測”と，観測画像を用いて人物らしさを評価する“観測”のステップを逐次実行することで実現される．具体的には，ベイズの法則により定式化される．

$$p(\mathbf{X}_t | \mathcal{Z}_t) \propto p(\mathbf{Z}_t | \mathbf{X}_t) p(\mathbf{X}_t | \mathcal{Z}_{t-1}). \quad (2.3.1)$$

ここで $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ は時刻 t における事前分布を表す．また， $p(\mathbf{Z}_t | \mathbf{X}_t)$ は尤度を表し，観測画像から推定される．

推定対象がマルコフ過程に従うと仮定したとき，事前分布 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ は Chapman-Kolmogorov 方程式，

$$p(\mathbf{X}_t | \mathcal{Z}_{t-1}) = \int p(\mathbf{X}_t | \mathbf{X}_{t-1}) p(\mathbf{X}_{t-1} | \mathcal{Z}_{t-1}) d\mathbf{X}_{t-1}, \quad (2.3.2)$$

により計算される．これにより逐次的な状態量推定が可能となり，対象の追跡が実現される．

パーティクルフィルタは，図 2.4 に示すように，密度分布 $p(\mathbf{X}_t | \mathcal{Z}_t)$ を重み $\pi_t^{(n)}$ を持つ仮説 $s_t^{(n)}$ の集合 $\{(s_t^{(n)}, \pi_t^{(n)}) \mid n = 1, 2, \dots, N\}$ を用いて離散的に近似する．これにより，任意の形状の分布を表現することが可能となるため，ノイズや環境変動に対して頑健な追跡を実現できる．具体的には次の処理を繰り返すことで実現される．

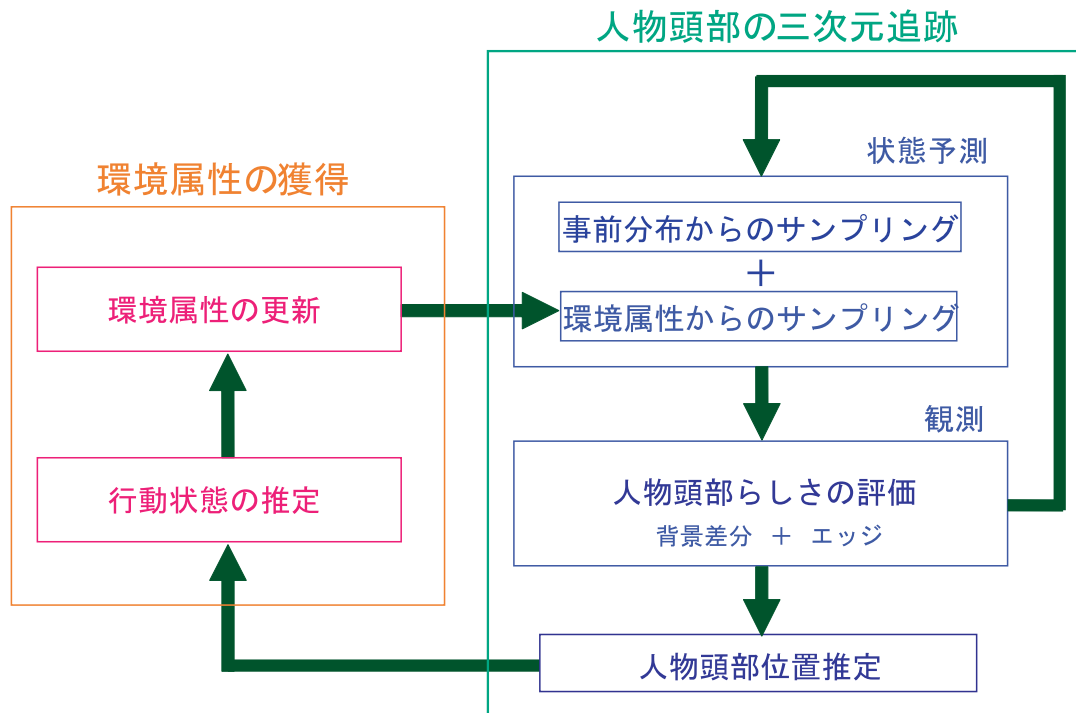


図 2.3: システム概要 .

(i) 仮説の選択

時刻 $t - 1$ における事後分布 $p(\mathbf{X}_{t-1} | \mathcal{Z}_{t-1})$ を表す仮説群 $\{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$ が持つ重み $\{\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}\}$ の比に従い, 仮説群 $\{s'_{t-1}^{(1)}, \dots, s'_{t-1}^{(N)}\}$ を選択する .

(ii) 状態遷移モデルに基づく仮説の伝播

選択された仮説群 $\{s'_{t-1}^{(1)}, \dots, s'_{t-1}^{(N)}\}$ を, 予め定義した状態遷移モデル $p(\mathbf{X}_t | \mathbf{X}_{t-1})$ に従いそれぞれ伝播させ, 時刻 t における事前分布 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ に従う仮説群 $\{s_t^{(1)}, \dots, s_t^{(N)}\}$ を生成する .

(iii) 観測による重み $\pi_t^{(n)}$ の決定

仮説 $s_t^{(n)}$ の重み $\pi_t^{(n)}$ を, 時刻 t の観測画像から得られる観測モデル $p(\mathcal{Z}_t | \mathbf{X}_t)$ を用いて評価することにより決定する .

図 2.5 に一連の処理の概要を示す .

本手法の人物追跡器は, 視野を共有する複数台のカメラを用いて人物頭部を三次元的に追跡する . 室内空間において, 床面を XY 平面に一致させ, 高さ方向に Z 軸をとる三次元世界座標系 $O - XYZ$ を定義する . 追跡対象である人物頭部を楕円体

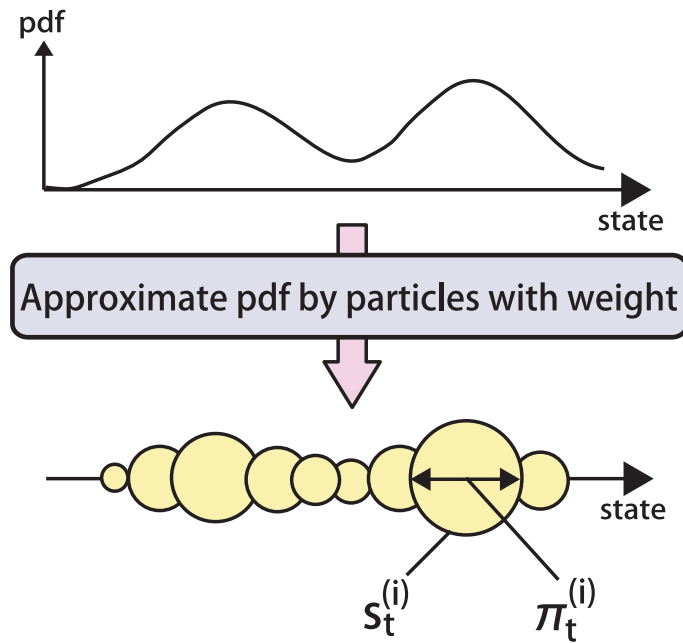


図 2.4: 確率密度分布の離散表現 .

でモデル化し，その中心座標を (x, y, z) とする．また時刻 t における n 番目の仮説は， $\mathbf{s}_t^{(n)} = [x_t^{(n)}, y_t^{(n)}, z_t^{(n)}]^T$ という三次元ベクトルの状態量を持つものとする．

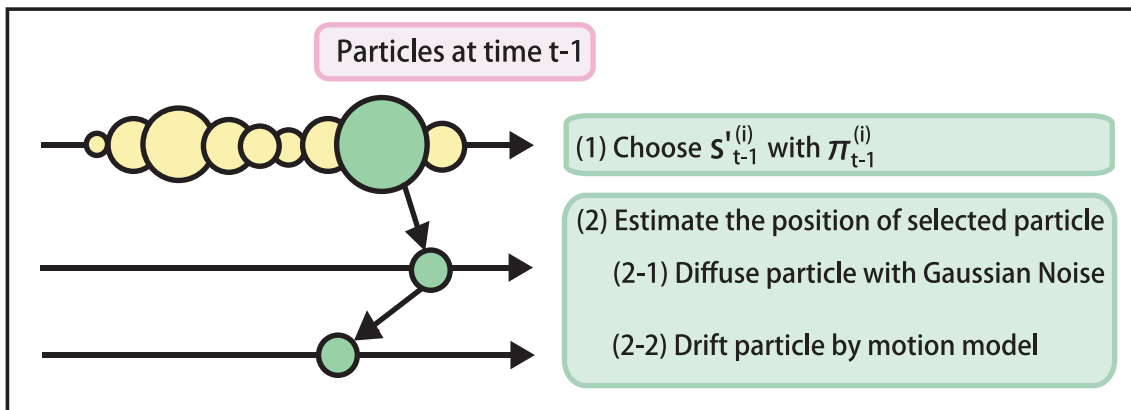
2.3.2 環境属性の統合

環境属性を importance function としてパーティクルフィルタの追跡の枠組みに導入する．すなわち，仮説をサンプリングするとき，事前分布 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ と環境属性 $g_t(\mathbf{X}_t)$ をサンプリング源として同時に利用する．具体的には，ICONDENSATION [42] の考え方にに基づき，以下の3つのサンプリング手法を併用する．またこれらのサンプリング手法はある一定の割合で選択する．

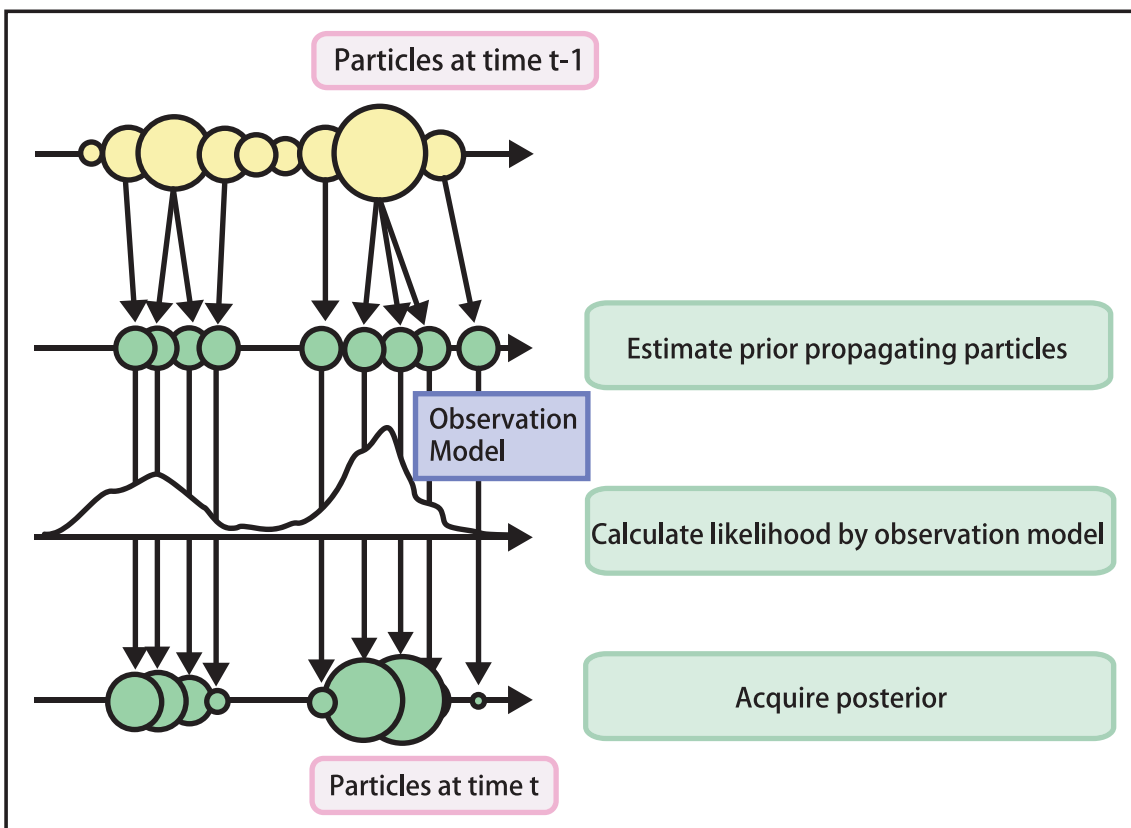
(a) 追跡

パーティクルフィルタの枠組みにおける仮説のサンプリング，すなわち事前分布 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ からサンプリングする．状態遷移モデル $p(\mathbf{X}_t | \mathbf{X}_{t-1})$ は，本手法では次式のように定義する．

$$\mathbf{s}_t^{(n)} = \mathbf{s}_{t-1}^{(n)} + \omega , \quad (2.3.3)$$



(a) Propagating particles for acquiring prior



(b) Particle filtering

図 2.5: パーティクルフィルタ. (a) 仮説の伝播による事前分布 $p(\mathbf{X}_t | \mathcal{Z}_{t-1})$ の推定, (b) パーティクルフィルタ全体の処理手順.

ここで ω は平均 0, 分散 σ^2 を持つガウス雑音を意味する.

(b) 追跡初期化

追跡初期化のための importance sampling . 環境属性 $g_t(\mathbf{X}_t)$ からサンプリングする .

(c) 追跡修正

追跡修正のための importance sampling . 環境属性 $g_t(\mathbf{X}_t)$ からサンプリングする . 仮説の重みを評価するときに環境属性により修正する .

これにより環境に適応した追跡の修正, 初期化を行うことが可能となるため, より安定な追跡を実現することができると考えられる . 環境属性の詳細については次節 2.4 節において述べる .

2.3.3 観測による人物頭部らしさの評価

各カメラで観測される画像の特徴量を用い, 仮説 $s_t^{(n)}$ の人物頭部らしさを評価する . 具体的には, 仮説をカメラに投影したときの画像において, 背景差分をすることで得られる前景領域と, Sobel Filter をかけることにより得られるエッジ勾配を用いて評価する . そして, 複数台のカメラによる評価を統合することで, 仮説の最終的な評価値とする .

仮説の画像平面への投影

時刻 t における n 番目の仮説を表す楕円体を $\Gamma_t^{(n)}$, i 番目のカメラ画像平面へ投影したときに得られる楕円領域を $\Omega_{i,t}^{(n)}$ とする . 世界座標系と i 番目のカメラ画像座標系との対応関係を, 投影関数 $\mathcal{F}_i(\cdot)$ で表現したとき, 仮説の投影は次式のように表現される . また, 投影された人物頭部の中心位置を $\mathbf{p}_{i,t}^{(n)}$ とする .

$$\Omega_{i,t}^{(n)} = \mathcal{F}_i(\Gamma_t^{(n)}) . \quad (2.3.4)$$

三次元空間中の人物頭部の大きさは, 簡単のために, 個人差による変動を考えず, ある決まった値を用いる . これにより, 楕円体の中心座標が決まることで頭部の端点の位置も決定される . 得られた端点の座標を同様に画像平面へ投影することで, 画像平面に投影された人物頭部の横幅 $w_{i,t}^{(n)}$ と縦幅 $h_{i,t}^{(n)}$ を得ることができる .

背景差分による評価

背景差分により得られる i 番目のカメラの前景画像において，投影された仮説の位置を中心とした楕円 $\Omega_{i,t}^{(n)}$ 内部の全画素数を $|\Omega_{i,t}^{(n)}|$ ，楕円 $\Omega_{i,t}^{(n)}$ 内部の前景領域に含まれる画素数を $f_{i,t}^{(n)}$ としたとき，背景差分による人物頭部らしさ $\pi_{i,t}^{fore,(n)}$ は次式により与えられる．

$$\pi_{i,t}^{fore,(n)} = \frac{f_{i,t}^{(n)}}{|\Omega_{i,t}^{(n)}|}. \quad (2.3.5)$$

また，計算高速化のために，前景領域の画素数の計算には Integral Image [99] を用いる．

エッジの類似度による評価

i 番目のカメラから得られる観測画像 $I_{i,t}(x, y)$ に Sobel Filter をかけることにより，水平方向のエッジ画像 $D_{x\ i,t}(x, y) \equiv \partial I_{i,t}(x, y)/\partial x$ と，垂直方向のエッジ画像 $D_{y\ i,t}(x, y) \equiv \partial I_{i,t}(x, y)/\partial y$ を得る．これらの情報を用い，仮説 $s_t^{(n)}$ のエッジ情報に基づく人物頭部らしさ $\pi_{i,t}^{edge,(n)}$ を求める．

本手法では，楕円の輪郭位置におけるエッジ勾配 $\mathbf{D} \equiv (D_{x\ i,t}(x, y), D_{y\ i,t}(x, y))^T$ と法線ベクトル $\mathbf{n} \equiv (n_x(x, y), n_y(x, y))^T$ の内積を評価する手法を用いる [64, 65]． i 番目のカメラの画像平面に投影された n 番目の仮説の位置 $\mathbf{p}_{i,t}^{(n)}$ と，人物頭部の横幅 $w_{i,t}^{(n)}$ ，縦幅 $h_{i,t}^{(n)}$ から，楕円 $\Omega_{i,t}^{(n)}$ の輪郭位置は次式で与えられる．

$$x_{c\ i,t}^{(n)}(\theta) = \frac{w_{i,t}^{(n)}}{2} \cos \theta, \quad (2.3.6)$$

$$y_{c\ i,t}^{(n)}(\theta) = \frac{h_{i,t}^{(n)}}{2} \sin \theta. \quad (2.3.7)$$

ここで θ は画像座標系 (u, v) における u 方向を 0 [deg.] とし，半時計回りの方向に増加する値をとる角度パラメータを意味する．また，人物頭部の輪郭位置を， $\mathbf{c}(\theta) \equiv (x_{c\ i,t}^{(n)}(\theta), y_{c\ i,t}^{(n)}(\theta))^T$ とする．

人物頭部の輪郭位置において，正規化されたエッジ勾配と法線ベクトルの内積の和を計算する．本手法では，頭部領域における輪郭位置，すなわち角度 $-30[\text{deg.}] \leq \theta \leq 210[\text{deg.}]$ の範囲のみを考慮する．これより，仮説 $s_t^{(n)}$ のエッジ情報による人物頭部らしさ $\pi_{i,t}^{edge,(n)}$ は次式で与えられる．

$$\pi_{i,t}^{edge,(n)} = \sum_{-30 \leq \theta \leq 210} \frac{\mathbf{D}|_{\mathbf{c}(\theta)} \cdot \mathbf{n}|_{\mathbf{c}(\theta)}}{\|\mathbf{D}|_{\mathbf{c}(\theta)}\| \|\mathbf{n}|_{\mathbf{c}(\theta)}\|}. \quad (2.3.8)$$

評価の統合

背景差分による評価値 $\pi_{i,t}^{fore,(n)}$ と、エッジ情報による評価値 $\pi_{i,t}^{edge,(n)}$ の積を、 i 番目のカメラによる n 番目の仮説の人物頭部らしさ $\pi_{i,t}^{(n)}$ とする。そして、各カメラ毎に得られる人物頭部らしさ $\pi_{i,t}^{(n)}$ を視野を共有する全てのカメラについて統合することにより、時刻 t の n 番目の仮説の最終的な人物頭部らしさ $\pi_t^{(n)}$ を求める。視野を共有するカメラの台数を M としたとき、次式で表現される。

$$\pi_t^{(n)} = \prod_{i=1}^M \pi_{i,t}^{(n)}. \quad (2.3.9)$$

ただし、“追跡修正”のために環境属性 $g_t(\mathbf{X}_t)$ からサンプリングされた仮説の場合 (2.3.2 節, (c) のサンプリング), 仮説の重みは次式で与えられる。

$$\pi_t^{(n)} = \frac{p(\mathbf{X}_t = \mathbf{s}_t^{(n)} | \mathcal{Z}_{t-1})}{g_t(\mathbf{s}_t^{(n)})} \prod_{i=1}^M \pi_{i,t}^{(n)}, \quad (2.3.10)$$

ここで $g(\mathbf{X}_t)$ は importance function であり、環境属性を意味する。

人物頭部位置推定

全ての仮説に対して重みを計算した後、事後分布の期待値をとることにより、時刻 t における人物頭部の三次元位置 \mathbf{s}_t^{head} を推定する。具体的には次式により与えられる。

$$\mathbf{s}_t^{head} = \sum_{n=1}^N \pi_t^{(n)} \mathbf{s}_t^{(n)}. \quad (2.3.11)$$

2.4 環境属性の更新

前節において述べた人物追跡の枠組みにより、人物頭部の三次元位置 \mathbf{s}_t^{head} が逐次推定される。このようにして推定された三次元頭部位置を用いることにより、環境属性の更新を実現する。本節では、環境属性の更新方法について詳しく説明する。

2.4.1 環境属性の確率表現

広域の環境を対象とする場合、人物の存在しやすい領域は複数あると考えられる。このような環境において人物存在確率を表現するために、環境属性は多峰性の分布に

対応する必要がある．また，前述したように，環境属性を importance function $g_t(\mathbf{X}_t)$ として追跡の枠組みに組み込むため，仮説のサンプリングが容易なモデルである必要がある．そこで本手法では，環境属性を混合正規分布を用いて表現する．すなわち，

$$g_t(\mathbf{X}_t) = \sum_{i=1}^K \omega_{i,t} \mathcal{N}(\mathbf{X}_t \mid \boldsymbol{\mu}_{i,t}, \boldsymbol{\Sigma}_{i,t}). \quad (2.4.1)$$

ここで $\mathcal{N}(\cdot)$ は平均 $\boldsymbol{\mu}_{i,t}$ ，共分散 $\boldsymbol{\Sigma}_{i,t}$ を持つ正規分布を表し， $\omega_{i,t}$ は重み，また K は正規分布の数を表す．

このようなパラメトリックモデルで表現した環境属性に行動履歴を反映させるということは，行動履歴を用いてモデルが持つパラメータを更新することと同義である．

2.4.2 追跡結果の動的反映による環境属性の更新

追跡器により毎フレーム得られる人物頭部の三次元位置 \mathbf{s}_t^{head} を環境属性に逐次反映させることにより，行動履歴に基づく環境属性を獲得する．具体的には，時刻 $t-1$ において各正規分布が持つパラメータ， $\omega_{i,t-1}$ ， $\boldsymbol{\mu}_{i,t-1}$ ， $\boldsymbol{\Sigma}_{i,t-1}$ と人物頭部の三次元位置 \mathbf{s}_t^{head} を用いて，EM アルゴリズムをオンライン利用に拡張したオンライン EM アルゴリズムを環境属性に逐次適用することで実現する．

EM アルゴリズムとは，最尤推定手法の一つであり，観測できない，もしくは未観測の事象が存在する場合においても用いることができる汎用的な最尤推定手法である [21, 23]．

推定すべきパラメータを θ ，観測されるデータ集合を $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ とし，観測できない値をとる確率変数を \mathbf{Y} とする．このとき，EM アルゴリズムでは以下の2つのステップを解が収束するまで繰り返し計算する．

E ステップ：

次の対数尤度の条件付期待値を計算する．

$$\begin{aligned} Q(\theta \mid \theta^{(t)}) &= E [\ln p(\mathcal{X}, \mathbf{Y} \mid \theta) \mid \mathcal{X}, \theta^{(t)}] \\ &= \sum_j \ln p(\mathcal{X}, \mathbf{y}_j \mid \theta) p(\mathbf{y}_j \mid \mathcal{X}, \theta^{(t)}). \end{aligned} \quad (2.4.2)$$

M ステップ：

$Q(\theta \mid \theta^{(t)})$ を θ について最大化し，その θ を $\theta^{(t+1)}$ とする，すなわち，

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta \mid \theta^{(t)}). \quad (2.4.3)$$

EM アルゴリズムを混合正規分布のパラメータ推定に適用する場合，以下の計算を全ての i についてパラメータが収束するまで行う．

$$\boldsymbol{\mu}_i^{(t)} = \frac{\langle \mathbf{x} \rangle_i(m)}{\langle 1 \rangle_i(m)}, \quad (2.4.4)$$

$$\boldsymbol{\Sigma}_i^{(t)} = \frac{\langle \mathbf{x} \mathbf{x}^T \rangle_i(m)}{\langle 1 \rangle_i(m)} - \boldsymbol{\mu}_i^{(t)} \left(\boldsymbol{\mu}_i^{(t)} \right)^T, \quad (2.4.5)$$

$$\omega_i^{(t)} = \langle 1 \rangle_i(m). \quad (2.4.6)$$

ここで $\langle \cdot \rangle_i(m)$ は i 番目の正規分布が選択される事後分布 $p(i | \mathbf{x}_l, \theta^{(t)})$ の重み付き平均，

$$\langle f(\mathbf{x}) \rangle_i(m) \equiv \frac{1}{m} \sum_{l=1}^m f(\mathbf{x}_l) p(i | \mathbf{x}_l, \theta^{(t)}), \quad (2.4.7)$$

を意味する．また，事後分布 $p(i | \mathbf{x}_l, \theta^{(t)})$ は，ベイズの定理により次のように表される．

$$\begin{aligned} p(i | \mathbf{x}_l, \theta^{(t)}) &= \frac{p(i | \theta^{(t)}) p(\mathbf{x}_l | i, \theta^{(t)})}{p(\mathbf{x}_l | \theta^{(t)})} \\ &= \frac{\omega_i^{(t)} \mathcal{N}(\mathbf{x}_l | \boldsymbol{\mu}_i^{(t)}, \boldsymbol{\Sigma}_i^{(t)})}{\sum_{j=1}^K \omega_j^{(t)} \mathcal{N}(\mathbf{x}_l | \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)})}. \end{aligned} \quad (2.4.8)$$

このように，EM アルゴリズムは，全観測データを用いたバッチ処理によって実現される．このため，多数の観測データを用いる場合，多くの計算量を必要とする．

一方で，EM アルゴリズムにおける計算量の問題を克服し，オンライン処理に拡張したオンライン EM アルゴリズムが提案されている [79]．そこでは，時刻 t において得られた観測データ $\mathbf{x}(t)$ と時刻 $t-1$ で推定されたパラメータを用い，時刻 t におけるパラメータを推定する．各混合正規分布のパラメータ $\omega_{i,t}$ ， $\boldsymbol{\mu}_{i,t}$ ， $\boldsymbol{\Sigma}_{i,t}$ は次のようにそれぞれ計算される．

$$\boldsymbol{\mu}_i^{(t)} = \frac{\langle \langle \mathbf{x} \rangle \rangle_i(t)}{\langle \langle 1 \rangle \rangle_i(t)}, \quad (2.4.9)$$

$$\boldsymbol{\Sigma}_i^{(t)} = \frac{\langle \langle \mathbf{x} \mathbf{x}^T \rangle \rangle_i(t)}{\langle \langle 1 \rangle \rangle_i(t)} - \boldsymbol{\mu}_i^{(t)} \left(\boldsymbol{\mu}_i^{(t)} \right)^T, \quad (2.4.10)$$

$$\omega_i^{(t)} = \langle \langle 1 \rangle \rangle_i(t). \quad (2.4.11)$$

ここで， $\langle \langle \cdot \rangle \rangle_i(t)$ は，学習の重みを α として，

$$\langle \langle f(\mathbf{x}) \rangle \rangle_i(t) = (1 - \alpha) \langle \langle f(\mathbf{x}) \rangle \rangle_i(t-1) + \alpha f(\mathbf{x}) p(i | \mathbf{x}(t), \theta^{(t-1)}). \quad (2.4.12)$$

として定義される．また $p(i | \mathbf{x}(t), \theta^{(t-1)})$ は， i 番目の正規分布が選択される事後分布を意味する．

このように、過去の推定結果を学習率 α で伝播させることにより、高速にパラメータ推定をすることを可能にする。そのため、オンライン EM アルゴリズムを利用することは、人物頭部追跡と環境属性の獲得を同時に実現する本手法に適している。

2.4.3 人物の行動状態推定に伴う行動履歴の適応的反映

人物の行動により環境属性の本来の役割である追跡の安定化、初期化の機能が失われる場合がある。この問題を引き起こす人物の行動例の一つとして、滞留行動が考えられる。人物が滞留している間、人物頭部の位置はほとんど変化しない。長時間このような状態が続いたとき、行動履歴の環境属性への反映により、滞留している周辺位置にのみ環境属性の確率密度値が集中する状態が起こる（図 2.6）。これにより、他の領域は人物が存在しえない領域であると判断され、人物追跡の不安定化、新規追跡対象の検出失敗などといった様々な問題を引き起こす可能性がある。

このような問題に対処するために、追跡対象である人物の行動状態を推定し、状態に応じて環境属性の更新の重み α を適応的に変化させる必要がある。そこで本手法では、図 2.7 に示す行動状態遷移モデルを導入することで、追跡対象の行動状態に伴う環境属性の適応的更新を実現する。

視野外 / 視野内の判定

追跡対象である人物が視野内に存在しているかどうかの判定条件を定義する。本手法では、この状態判定に追跡中における尤度の変化を利用する。もし人物が視野内に存在するならば、人物が存在する領域周辺で尤度が相対的に高くなることが考えられる。逆に人物が視野内に存在しないとき、いずれの領域においても人物らしさが高くなることはないため、尤度は一様に小さい値をとる。このような人物の有無に伴う尤度分布の変化を次のように利用する。

視野外 / 視野内 判定条件

- $\sum_{i=1}^N \pi_t^{(i)} < T_{exist}$: 視野外と判定 (1)
- $\sum_{i=1}^N \pi_t^{(i)} \geq T_{exist}$: 視野内と判定 (2)

ここで $\pi_t^{(i)}$ はパーティクルフィルタにおける i 番目の仮説が持つ尤度を意味し、 N は全仮説数を意味する。また T_{exist} は判定のための閾値である。

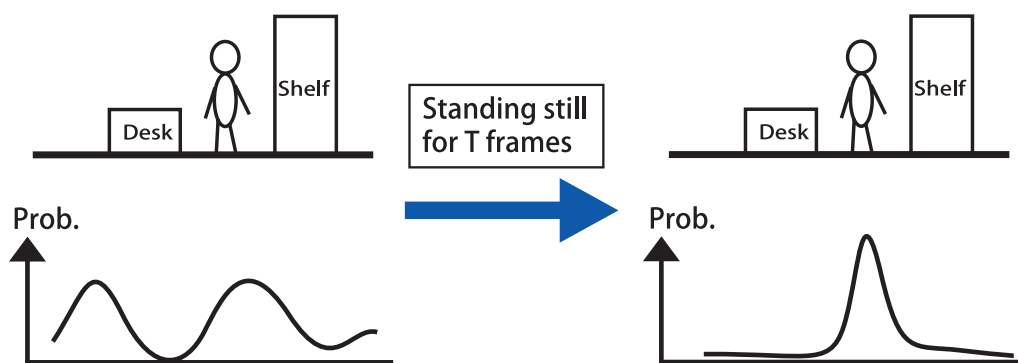


図 2.6: 滞留行動による環境属性の局所化．人物頭部位置を行動状態を考慮せずに環境属性へ反映させ続けると，長時間の滞留に伴い環境属性が一箇所にピークを持つ分布形状になる．これにより他の領域において追跡の初期化，再初期化の失敗を引き起こす．

行動状態遷移モデル

図 2.7 に示すように，行動状態遷移モデルは視野外，出現，活動，滞留の4つの状態を持つ．また人物はこのいずれかの状態に属すると仮定する．以下それぞれの状態，遷移条件について詳しく述べる．

視野外状態

人物がカメラにより観測されていない状態を視野外状態とする．このとき人物は視野内に存在しないため，環境属性の更新の重みは $\alpha = 0$ とする．視野外状態からの遷移条件は，前節の視野外 / 視野内判定の方法を用いて次のように定義する．

- 条件 (1) を満たす：遷移しない (視野外状態)
- 条件 (2) を満たす：出現状態へ遷移

出現状態

視野外から人物が新たに現れた状態を出現状態とする．この状態は追跡初期化に重要であると考えられるため，環境属性の更新の重みを $\alpha = k\alpha_{act}$ とする．ここで

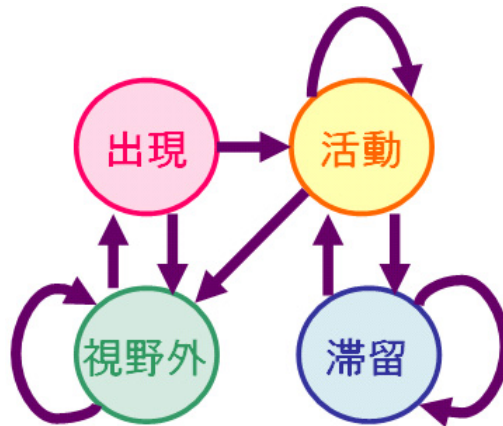


図 2.7: 行動状態遷移モデル .

$k > 1$ は任意の定数 , また α_{act} は活動状態における更新の重みを表す . 出現状態からの遷移条件を次のように定義する .

- 条件 (2) を満たす : 活動状態へ遷移
- 条件 (1) を満たす : 視野外状態へ遷移

滞留状態

ある特定の場所から人物が長時間ほぼ動かない状態を滞留状態とする . 環境属性の集中に伴う初期化の機能喪失を防ぐために , 滞留時間 t_{stay} の増加に伴い滞留状態における環境属性の更新の重みを減少させる . 具体的には $\alpha = \alpha_{act}/t_{stay}$ とする . 滞留状態からの状態遷移について以下の二つの条件を用いる .

- 人物の移動速度 v_t が閾値 T_{vel} [cm/frame] 以下であることを満たす . (3)
- 条件 (3) を閾値 T_s [frame] 以上満たす . (4)

これらの条件を利用し , 滞留状態からの遷移条件を次のように定義する .

- 条件 (2) , (3) を満たす : 遷移しない (滞留状態)
- 条件 (2) を満たす : 活動状態へ遷移

活動状態

視野内で人物が何らかの行動をしている状態を活動状態とする．多くの場合人物はこの状態に属する．環境属性の更新の重みは $\alpha = \alpha_{act}$ とし，活動状態からの遷移条件は次のように定義する．

- 条件 (2) を満たす：遷移しない (活動状態)
- 条件 (2), (4) を満たす：滞留状態へ遷移
- 条件 (1) を満たす：視野外状態へ遷移

2.5 実験・考察

本手法の有効性について確認するために，図 2.8 に示すような対象空間において，環境属性の獲得実験と人物頭部追跡実験を行った．

室内天井に校正済みの IEEE1394 カメラ (Point Grey Research 社製 Flea) を 2 台設置し，解像度 640 x 480 ピクセル，毎秒 30 フレームの速度で得られるカメラ画像を用いた．カメラ 1 台毎に 1 台の汎用 PC (CPU: Intel Pentium4 2.8GHz, RAM: 2GByte, OS: Windows XP) を割り当て，それぞれをクライアント PC とした．またクライアント PC の情報を統合するサーバの役割を果たす汎用 PC (CPU: Intel Pentium4 3.2GHz, RAM: 1GByte, OS: Windows XP) を用意し，これらを通信速度 1Gbps の Gigabit Ethernet で接続してサーバ・クライアントモデルを構築した [65,91]．具体的な追跡の処理手順は次のようになる．

- (1) サーバ : 仮説群を全てのクライアントに送信
- (2) クライアント : 仮説群受信
- (3) クライアント : 観測画像に基づく人物頭部らしさの評価
- (4) クライアント : 評価値をサーバに送信
- (5) サーバ : 各クライアントの評価値を統合し，最終的な評価値計算
- (6) サーバ : 仮説サンプリング．その後 (1) へ戻る

データ通信には DirectX9.0c に含まれる API の一つである DirectPlay [71] を利用し，通信プロトコルは TCP/IP プロトコルを用いた．これにより Winsock [105] を利用するよりも容易にネットワークプログラムを構築できる．

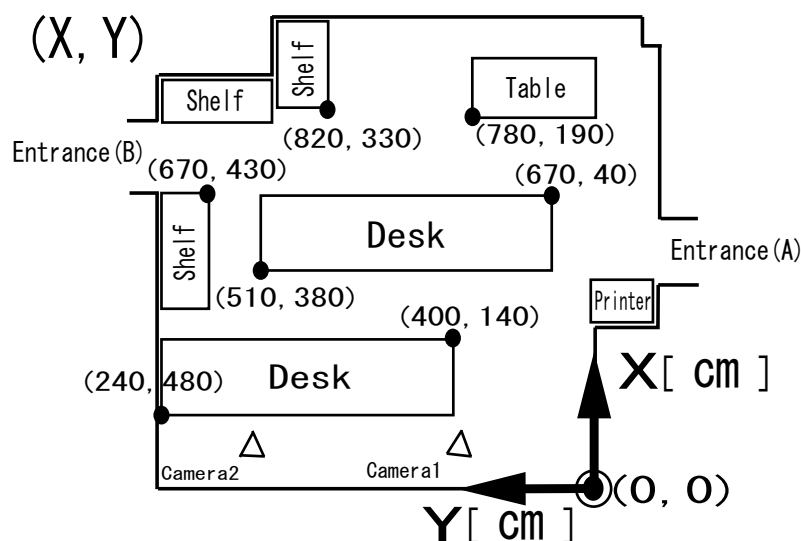


図 2.8: 対象空間のレイアウト．図中の数値はそれぞれの位置における座標値を意味する．

実験で用いたパラメータは次のように設定した．環境属性の正規分布の個数は $K = 7$ とし，行動状態推定に用いたパラメータはそれぞれ $\alpha_{act} = 0.0005$ ， $k = 2$ ， $T_{exist} = 2.5$ ， $T_{vel} = 1[\text{cm/frame}]$ ， $T_s = 200[\text{frame}]$ とした．また，Mixture Particle Filter [98] の考え方にに基づき，複数人の追跡の枠組みに拡張して実験を行った．人物 1 人当たりの追跡に利用した仮説数は $N = 150$ とした．

2.5.1 環境属性の獲得

対象空間をおよそ 30 分間観測することにより，環境属性の獲得実験を行った．本実験では，行動状態推定の効果を確認するために，行動状態推定を伴う場合とそうでない場合のアルゴリズムを予め撮影した映像に対してそれぞれ適用し，得られた分布を比較した．ここでは，片方の通路の頻繁な通行，特定個所における雑談による滞留が主にみられた部分の映像を利用した．図 2.9 にその様子を示す．

それぞれの推定結果を図 2.10，2.11 に示す．動線の反映のみにより得られた環境属性(図 2.10) は，ある領域 ($650 < X < 750$ ， $0 < Y < 100$) における確率密度値がほかの部分に比べ極端に高くなっていることがわかる．これは人物の滞留行動による環境属性の確率密度値の集中によるものと考えられ，実験に利用した映像においても二人の人物が長時間雑談している様子が観測された(図 2.9-(ii))．このこと



(i) 特定の通路の歩行

(ii) 雑談による滞留

図 2.9: 頻繁に観測された行動 .

から、実際に観測された滞留行動が正しく反映されていることが確認できる。しかしながら、その一方で、滞留行動以外に頻繁に観測された片方の通路の歩行による反映の効果がみられない。これは (2.4.12) 式における学習率を一定の値として行動履歴を反映し続けたことに起因すると考えられる。

行動状態推定を利用することにより得られた環境属性 (図 2.11) では、前者の環境属性においても同様の領域でみられた分布のピーク、そして奥側の通路領域 ($650 < X < 750$, $0 < Y < 450$) に沿ったなだらかな分布の峰がみられる。これは、片側の通路の頻繁な通行 (図 2.9-(i)) によるものと考えられる。このことから行動履歴が正しく反映されているとみなすことができる。これは、人物の行動状態に伴い行動履歴が適応的に環境属性へ反映されたことによる効果であると考えられる。

以上のことから、人物の行動履歴が正しく環境属性に反映されていることが確認することができた。また行動状態推定は行動履歴を正しく反映させるために必要であることを確認することができた。

2.5.2 環境属性による追跡初期化の高速性

前節で得られた行動状態推定を伴う環境属性 (図 2.11) を用いて、追跡初期化に関する環境属性の有効性を確認した。

本実験では、人物が視野内に現れたとき、もしくは遮蔽などなんらかの要因によって追跡に失敗した状態から正しい追跡が再開されるまでの速さを評価した。追跡対象が実際に観測されるフレームを f_{gt} 、正しく追跡が初期化されたときのフレームを

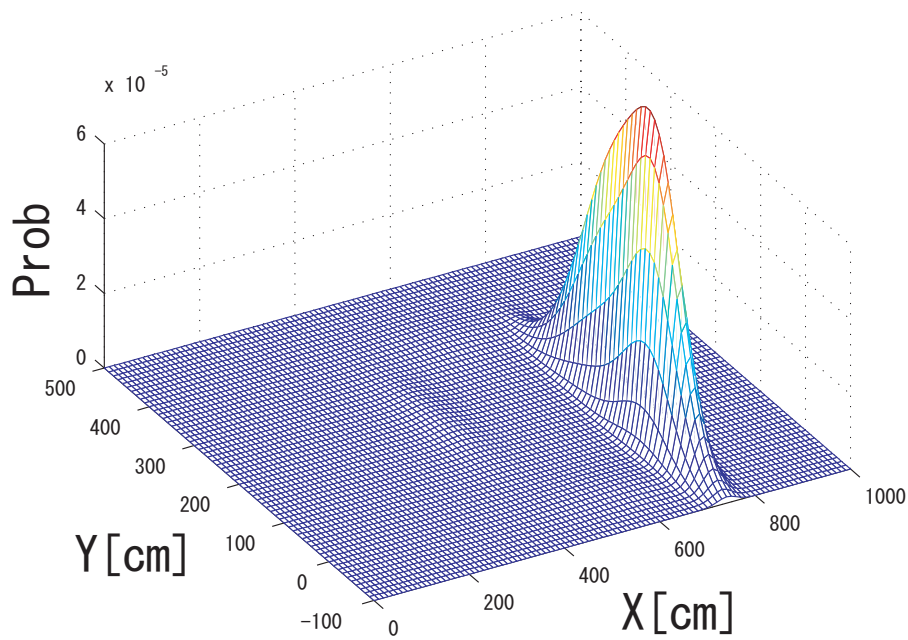


図 2.10: 環境属性の構築結果 (行動状態推定無し)

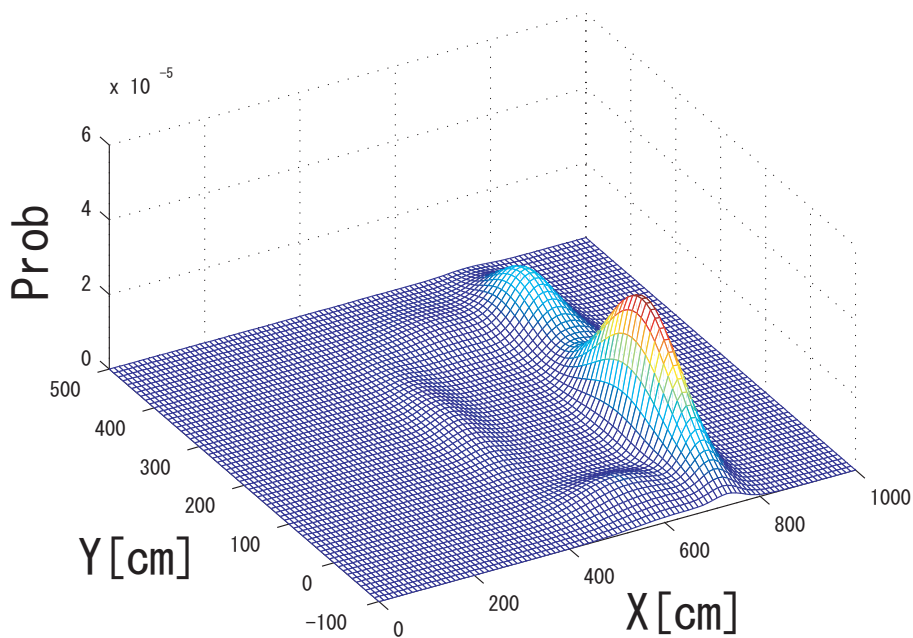


図 2.11: 環境属性の構築結果 (行動状態推定有り)



(a) frame No: 1038. 新規追跡対象出現

(b) frame No: 1054. 追跡開始

図 2.12: 追跡初期化検証実験の評価例．frame No: 1038 (左図(a)) では人物が新たに現れたことが観測できるが，追跡は開始されていない．そして frame No: 1054 (右図(b)) で追跡が開始されている．この例では，frame No: 1038 を追跡が開始されるべき真値 f_{gt} ，frame No: 1054 を実際に追跡開始されたフレーム f_{init} とする．そしてこれらの差分 $f_{diff} = |f_{gt} - f_{init}|$ を計算する．

f_{init} とする．このとき両者のフレーム差， $f_{diff} \equiv |f_{gt} - f_{init}|$ を追跡初期化が発生する様々な場合について計算し，統計的に評価した．これは， f_{diff} が 0 に近いほど高速に追跡初期化が行われていることを意味する．図 2.12 に例を示す．

本手法の有効性を比較検証するために，行動履歴を反映することにより得られた環境属性を用いた場合と一様分布を環境属性として利用した場合の両者を用いて，単一シーンにおいて比較実験を行った．本実験では，人物は得られた環境属性に従うような行動，すなわち人物存在確率が高い領域において頻繁に行動する映像を用いた．また本映像では，人物同士のすれ違いによる部分遮蔽，人物が机下に入り込むことによる完全遮蔽などが頻繁に観測された．試行回数は 5 回であり，追跡の初期化，再初期化は全試行合わせて 40 回行われた．

初期化の速さの比較結果を表 2.1 に示す．示されている値はそれぞれ f_{diff} の平均，標準偏差，すなわち追跡が初期化されるまでに必要な平均フレーム数とそのばらつき具合を意味する．環境属性と一様分布それぞれの平均を比較すると，環境属性によるものは一様分布のそれに比しておよそ 40% 程度のフレーム数で追跡初期化が行われている．この環境属性による平均遅れは時間にして 1 秒弱の遅れであり，高速な追跡初期化が実現されているといえる．また標準偏差の比較結果について着目す

表 2.1: 追跡初期化の速さの比較結果．人物が実際に観測されたフレームと正しい追跡初期化が働いたフレームとの差， f_{diff} の平均と標準偏差を意味する．

	平均 [frame]	標準偏差 [frame]
一様分布	48.31	37.92
環境属性	21.62	16.45

ると，環境属性によるものは一様分布のそれに比しておよそ 50% 程度に抑えられており，安定な初期化が実現できていることがわかる．

2.5.3 環境属性と一様分布を利用した場合の追跡精度の比較

人物二名（人物 A，人物 B）の頭部の追跡実験を単一シーンにおいて行った．環境属性の効果を比較検証するために，環境属性，一様分布をそれぞれ利用して実験を行った．人物 A は出入口（A）から視野内に現れ，別の出入口（B）に向かって歩いた．それと同時に，人物 B は出入口（B）から視野内に現れ，人物 A が歩いた通路と同じ通路を歩き，部屋の奥の方へ歩いた．また追跡精度を定量的に調べるために，画像中の人物頭部位置を手作業で求め，2 枚の画像から逆投影することで求めた三次元座標を真の位置と見なした．試行回数は 5 回であり，それぞれのシーケンスについて追跡結果と真値の比較を行った．

全試行について統計的に評価した追跡誤差結果を表 2.2 に示す．それぞれ Z 軸方向及び XY 平面上での平均誤差と標準偏差を表す．環境属性を利用した場合，一様分布を利用した場合の評価結果を比較すると，両者ほぼ同等の追跡精度が得られていることがわかる．XY 平面上の平均誤差が 15cm 程度であるが，観測領域が広く，画面上での 1 ピクセルが実世界での 5cm 程度に対応することを考慮すると，安定な追跡を実現できているといえる．

あるシーケンスにおける人物追跡結果を図 2.13 に，追跡の様子を図 2.14, 2.15 にそれぞれ示す．これらの様子からも両者共に安定な追跡を実現できていることが見て取れる．

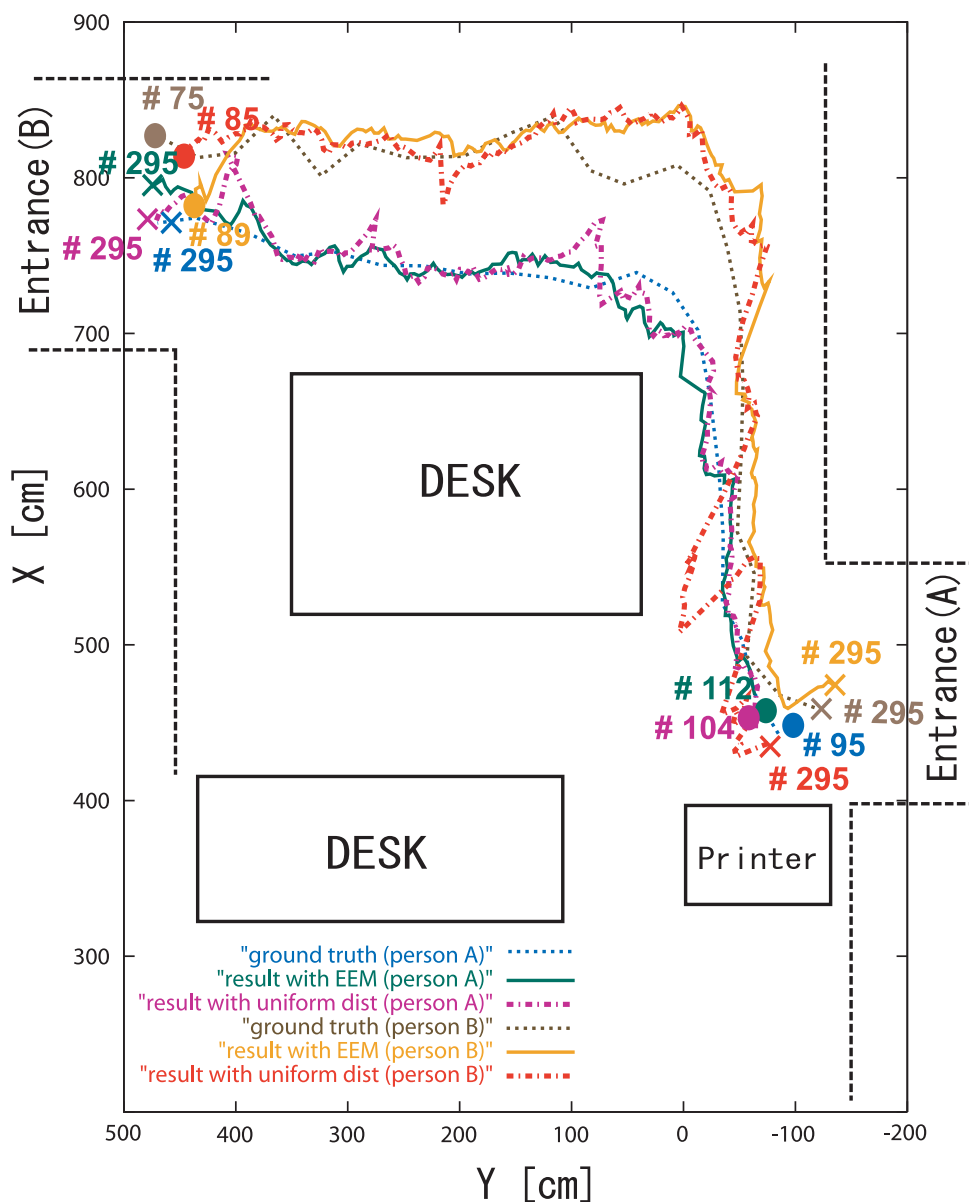
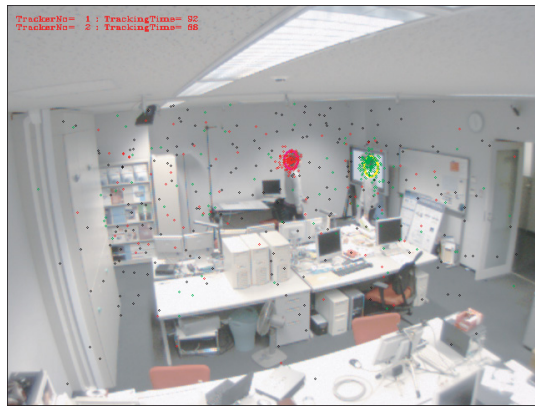
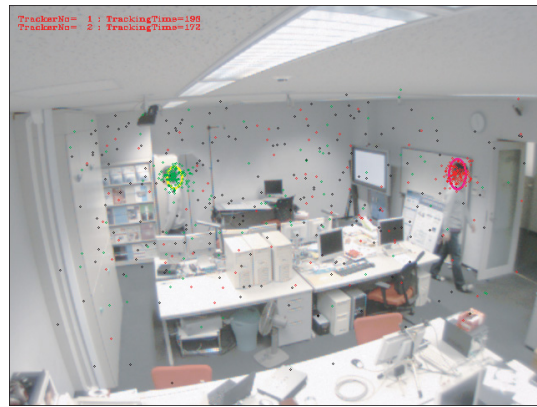


図 2.13: 人物頭部追跡結果．図中の線は2名の人物頭部の追跡結果をそれぞれ表している．凡例の“ ground truth ”は真値，“ result with EEM ”は環境属性を用いた追跡結果，そして“ result with uniform dist ”は一様分布を用いた場合の追跡結果を意味する．また推定された動線の端点に記されている数値はそれぞれ追跡開始・終了時刻を表す． \bullet で記された点が初期化位置，時刻であり， \times で記された点が追跡終了位置，時刻である．

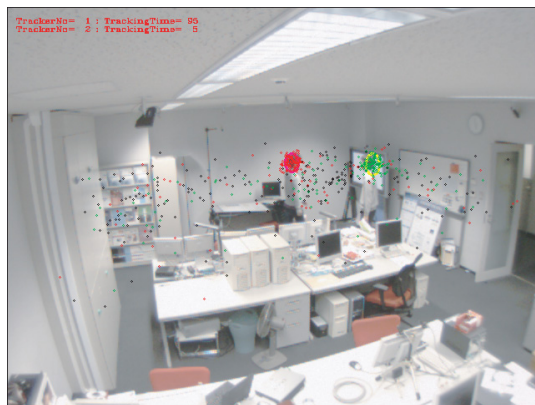


frame No: 169

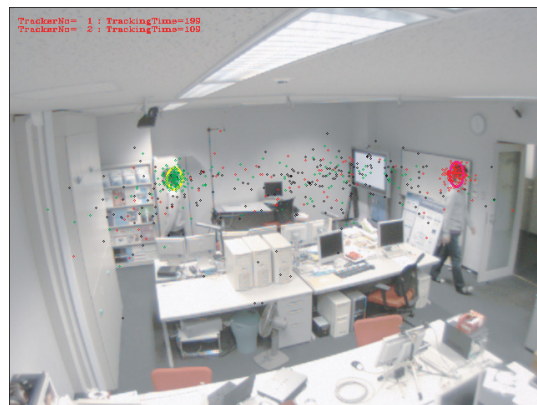


frame No: 273

図 2.14: 2名の人物頭部追跡の様子（一様分布の利用）．図中に示される点群はパーティクルフィルタの仮説を表している．また，人物以外の領域に分布している仮説群は環境属性からサンプリングされた仮説である．本実験では一様分布を利用しているため，観測環境全体から一様にサンプリングされていることが見て取れる．



frame No: 169



frame No: 273

図 2.15: 2名の人物頭部追跡の様子（環境属性の利用）．人物が頻繁に通行する領域（奥の通路）から集中的に仮説のサンプリングが行われていることが見て取れる．

表 2.2: 追跡誤差 .

		平均誤差 [cm]		標準偏差 [cm]	
		一様分布	環境属性	一様分布	環境属性
人物 A	XY 平面	15.75	14.54	10.06	9.10
	Z 軸	4.98	4.55	5.53	5.23
人物 B	XY 平面	15.18	13.52	8.23	7.32
	Z 軸	5.22	3.79	6.18	4.08

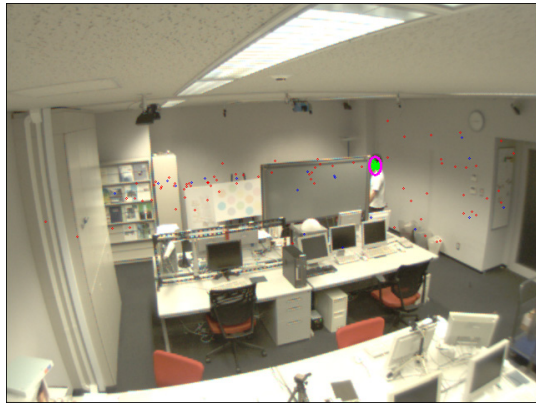
2.5.4 遮蔽への対処

図 2.16 は、人物が奥の通路を歩き (frame No: 345)、途中設置されたボードにより完全に遮蔽された後 (frame No: 360)、ボードの逆側から現われ (frame No: 720)、奥の部屋へ向かうシーケンスに対する追跡結果の一部である。この結果にあるように、ボードの完全な遮蔽により追跡が失敗していることが確認できる (frame No: 360 から frame No: 719)。しかしながら、その人物がボードの逆側から現れた際 (frame No: 720)、追跡器は速やかに追跡を再開していることがわかる (frame No: 723)。このように、環境属性を利用することで完全な遮蔽により追跡対象を見失った場合においても正しい追跡を再開できることがわかる。

以上の検証実験の結果から、環境属性は追跡の安定化、特に追跡対象の初期化、再初期化に関して効果的であるといえる。

2.5.5 環境属性の更新に要する計算コスト

環境属性の更新は、環境属性を利用しない場合 (一様分布を利用する場合) に比べていくらかの計算コストを要する。一方、パーティクルフィルタを用いた追跡において、仮説数の増加に伴い追跡精度が向上することが知られている。これらのことから、環境属性の更新に用いる計算コストをパーティクルフィルタの仮説数の増加に用いることで、追跡の安定化を図ることができるとも考えられる。このような計算コストの面から本手法の有用性を検証するために、環境属性の更新に必要な計算コストを実験的に求めた。



frame No: 345



frame No: 360



frame No: 720



frame No: 723

図 2.16: 追跡結果．完全な遮蔽からの追跡再開．

2.5.2 節の実験で利用した映像に対して本手法を適用し，環境属性の更新に要した処理時間を各フレーム毎に計測した．そしてその処理時間の平均値を環境属性の更新に必要な処理時間として採用した．本節冒頭で述べた実験環境で計測した結果，1回の環境属性の更新に必要な処理時間はおおよそ0.0025[ms]であった．

次に，環境属性の更新を行わない場合（一様分布の利用）において，1仮説を処理するために必要な処理時間を同様の方法で計測した．計測した結果，1仮説を処理するために必要な処理時間はおおよそ0.0211[ms]であった．これらの計測結果を用いて，追加可能な仮説の数 N' を算出すると，仮説の数はおおよそ0.0946[個]となる．これはすなわち，環境属性の更新に必要な計算コストを利用しても仮説の追加を行うことはできないことを意味する．このことから，環境属性の更新はパーティクルフィルタの仮説を増加させることに比べて低い計算コストで行うことができること

を確認できた。また2.5.2節の実験結果と併せて考えると、環境属性を利用することで仮説の数を増加させることなく追跡の安定化を図ることができると言える。

2.5.6 環境属性の利用の課題

観測履歴と異なり、環境属性の人物存在確率が低い領域において人物が頻繁に行動する場合、環境属性を利用することにより逆に追跡が不安定化することが考えられる。環境属性の適用範囲を知るために、本実験では環境属性が不得手とする場合について検証を行った。具体的には、2.5.1節にて得られた環境属性において人物存在確率が低い領域 ($400 < X < 500$, $0 < Y < 400$) を頻繁に人物が行動する状況を考えた。またこれまでの実験結果からわかるように、追跡安定化に関する環境属性の寄与は主に追跡初期化にあることから、2.5.2節と同様の追跡初期化の速さに関する評価を行った。実験に用いた映像は、上述したように人物存在確率が低い領域において人物が頻繁に行動するものとした。その映像では人物同士のすれ違いによる部分遮蔽、柱による遮蔽、人物が机下に入り込むことによる遮蔽などが頻繁に観測された。試行回数は5回であり、追跡の初期化、再初期化は全試行合わせて35回行われた。

初期化の速さの比較結果を表2.3に示す。両者の結果を比較すると、一様分布を用いた方が高速かつ安定に追跡初期化が行われている。これは、追跡失敗が人物存在確率が低い領域のみで頻繁に起こったことにより、いずれの追跡初期化においても多くの時間を要したためであると考えられる。また2.5.2節の実験結果(表2.1)と比較すると、一様分布を用いた場合は平均、標準偏差共に結果に大きな差はないといえる。それに対し環境属性を用いた結果では、人物が環境属性に従う行動をとった場合(表2.1)に比べ平均、標準偏差両面において明らかに不安定になっていることを確認できる。

これらの比較結果より、人物が観測履歴と大きく異なる動きをとる場合、環境属性は逆に追跡を不安定化させる可能性があるといえる。このような問題を防ぐために、より長時間の観測を対象とし、人物存在確率の極端な低下を防ぐ必要がある。また、環境属性だけでなく、他の異なる枠組みと併用する必要があると考えられる。

表 2.3: 人物存在確率が低い領域における追跡初期化の速さの比較結果 .

	平均 [frame]	標準偏差 [frame]
一様分布	36.22	38.78
環境属性	51.75	40.25

2.6 本手法のまとめ

人物は、特定の通路の頻繁な通行、特定の位置における滞留といった、対象環境に即した行動の特性を持つ。対象空間中の人物の存在確率は、このような行動特性に大きく依存すると考えられる。本手法ではこの点について着目し、人物の行動履歴を逐次利用することにより、対象シーンに特化した人物の存在確率を獲得できると考えた。そしてこれを環境属性として定義し、その構築法、追跡の枠組みへの統合を開発した。また実データを用いた実験により、本手法の有効性を確認した。

第3章 歩容特徴と局所的な見えに基づく混雑環境下人物追跡

混雑環境（朝のラッシュ時における駅の構内，イベント会場など）における人物追跡について研究する．頻繁に発生する遮蔽や，複数の人物が互いに近接していることにより，個々の人物を正しく追跡することが難しい．そこで本研究では，人物の個人性にあたる歩容特徴と局所的な見えの時間変動の一貫性という二つの指標を，特徴点軌跡のクラスタリングに基づく追跡の枠組みへ組み入れた人物追跡手法を開発する．

3.1 はじめに

近年，朝夕のラッシュ時の駅構内やイベント会場などの混雑環境下（図 3.1）における人物追跡への関心が高まっている．事故や事件の未然防止といったセキュリティ分野をはじめ，視野内の人数の自動計測，混雑解消のための情報源としての利用など，様々な分野への応用が期待される．

しかしながら，実際の混雑環境において頑健な追跡を実現するためには，頻繁に発生する遮蔽，非常に近接している複数の人物個々の識別といった様々な問題に対処する必要がある．そのため，背景差分に基づく人物の形状モデルを利用した検出手法や見えを手掛かりにした時系列フィルタリング手法は，混雑環境における追跡に不向きであると考えられる．

近年，このような形状や見えを用いた追跡手法とは異なった，画像中の特徴点の動きを利用した手法が提案されている [7, 58, 74]．これらの手法は，同じ人物に属している特徴点は類似した動きをするという仮定に基づき，特徴点軌跡群を動きの類似性と位置関係に基づきクラスタリングすることにより人物追跡を実現する．これらの手法は，対象の一部が観測されていれば追跡可能であるため，一般に部分遮蔽に頑健であると言える．



図 3.1: 混雑環境の例 .

しかしながら，混雑した状況下では，群集全体により形成される大きな動きの流れにより人物個々の動きは制限されやすい．これに伴い，隣接する人物と動きが類似する傾向にあるため，軌跡の動きの類似性と位置関係に基づいた手法では個々の人物を区別することは難しい．

このような問題に対処するために，本手法では，(1) 周波数領域における歩容特徴，(2) 局所領域における見えの時間変動という二つの異なる種類の指標を導入する．これらの特徴点軌跡のクラスタリング手法の考え方をを用いて追跡の枠組みへ組み入れることにより，混雑シーンにおける頑健な人物追跡を実現することが可能となる．

周波数領域における歩容特徴の利用は，群集における個々の人物の識別に非常に有効であると考えられる．互いに近接した複数の人物が同じ方向に同じ速度で歩いている場合においても，各人物の歩幅や歩行周期は異なるため，その周波数特徴には各人物固有の性質が現われる．このような個人特性は，その人物に属する特徴点の動きにも現われると考えられることから，本手法では特徴点軌跡の動きの周波数特性を個々の人物の区別に利用する．歩容特徴は，個人認証のための重要な手掛かりとしてバイオメトリクス分野で広く利用されており [63]，群集中の人物追跡においても歩容特徴は有用な指標になることを示唆している．

局所領域における見えの時間的な変化を利用することは，動きの周波数特徴さえも類似する場合において効果的に働くことが期待される．本手法では，三つの特徴点の軌跡によって張られる三角パッチを考え，その見えの時間変動に着目する．同

じ人物に属する特徴点軌跡により張られる三角パッチ内の見えは、時間的にほぼ変化しないことが期待されるが、その一方で他の人物との間に張られる三角パッチは、背景や他の人物の領域を含むためその時間変動は大きくなると考えられる。また局所領域のみに注目しているため、追跡対象全体の見えを必要としない。そのため部分的な遮蔽について頑健である。このように、動きと見えの異なる種類の指標を利用することにより、混雑環境においても頑健な人物追跡を実現することが可能になる。

本章の構成は以下のとおりである。まず、3.2節において本手法に関連する先行研究についてまとめ、本手法の位置づけおよび新規性について整理する。続く3.3節では特徴点軌跡のクラスタリングに基づく本手法の枠組みについて概観する。そして3.4節において歩容特徴、局所領域における見えの変化の二つの指標をどのように追跡の枠組みへ組み入れるのか詳しく述べる。3.5節では人工データ、実際の群集の映像を用いた実験結果について報告し、本手法の追跡性能、適用範囲について考察する。最後に3.6節において本手法についてまとめる。

3.2 関連研究

混雑環境下での人物追跡における大きな問題は、頻繁に発生する(部分)遮蔽、または個々の人物が互いに近接していることに伴う追跡の不安定化、失敗、の大きく二つが考えられる。このような問題へ対処するために、近年様々な研究が行われている。本節では、これらの先行研究について概観する。

3.2.1 人物の形状のモデル化

一つに、人物の形状をモデル化することにより群集の人物を追跡する方法が提案されている。Zhaoらは、人物を、頭部、胴体、両脚の4つのパーツに分割しそれぞれ楕円体でモデル化することにより、部分的な遮蔽の問題へ対処している[107]。具体的には、背景差分により抽出された前景領域に対し、モデルを当てはめることにより、各パーツの尤度を求める。そして最大事後確率推定(MAP推定)を行うことにより、混雑環境下における人物検出を実現している。

Wuらは、人物を、頭部、上半身、下半身、全身の四つの指標に基づき観測し、これらの領域それぞれについて検出器を学習することで部分遮蔽の問題へ対処している[103]。背景差分により抽出された領域に対してエッジ検出を行い、得られるエッジをいくつかの方向に基づき量子化した後、これらのヒストグラムを作成する。こ

のように得られるエッジ情報に基づくヒストグラムをサンプルとし、識別器の学習を行う。そして各部位の識別器による識別結果をベイズの定理に基づき組み合わせることにより、遮蔽に頑健な人物の検出、追跡を実現している。

人物をいくつかの部位に分割したモデルを用いる手法に対し、人物全体の形状を当てはめる手法もいくつか提案されている。Dongらは、予め定義した人物の形状モデルを用いることで、重なり合って観測される群集により作られる一つの大きな前景領域から個々の人物の検出を実現している [22]。また Khanらは、ホモグラフィを手掛かりに予め定義した人物モデルを前景領域にあてはめることで人物検出を実現している [48]。

街頭や歩道など人物が自由に行動できる環境を想定した場合、固定されたカメラに対して人物は異なるポーズで観測されることが考えられる。そのため、観測されるポーズによってはその検出性能が低下することが懸念される。これに対し Rodriguezらは、人物の様々なポーズの画像から形状情報を学習することにより、この問題へ対処している [77]。

3.2.2 見えの情報の利用

観測される人物の見えを手掛かりにした追跡手法もまた数多く提案されている。

Leibeらは、人物の局所的な領域における見えの情報と人物全体の見えを組み合わせた人物検出手法を提案している [57]。最初に、局所領域の情報を用いた人物検出のために事前に学習処理を行う。具体的には、SIFT 特徴点 [60] を検出し、その周りの局所領域のパッチを考慮する。これらの局所パッチを用い、コードブックを構築する。実際の局所領域の情報に基づく人物検出では、事前に学習により得られたコードブックを用い、カーネル密度推定の枠組みにより人物の検出を行う。そしてこれを人物検出の候補とし、人物全体の見えを手掛かりとした Chamfer Matching [3] により最終的な検出結果を得る。

また Tuらは、既存の人物検出アルゴリズムにより得られる人物の候補領域をグリッド状の局所パッチに分割し、そのパッチ内の見えを利用することで混雑状況において人物検出を行っている [95]。色情報に基づくパッチ間の類似度を計測し、EM アルゴリズムに基づく最適化を行うことにより頑健な人物検出を実現している。

見えを手掛かりにした、パーティクルフィルタに代表される時系列フィルタを用いた追跡手法もまた数多く提案されている。背景と人物の両方をモデル化し、追跡人数を状態量に含めることでシーン全体を追跡する手法 [43]、マルコフ確率場の利

用による複数対象追跡 [49]，追跡対象に特化した識別器のオンライン学習による人物追跡の高度化 [86]，Adaboost 検出器と追跡器の同時利用 [70] などが挙げられる．また部分遮蔽への対処のために，追跡対象の状態推定についてオプティカルフローによる動きの推定を用いた手法 [10, 85] もいくつか提案されている．

これ以外にも，時空間的な見えの相関を考慮した時空間マルコフ確率場を用いた人物追跡手法がよく知られている．文献 [39] では，時空間マルコフ確率場に基づいた追跡手法を拡張することにより，頻繁に発生する人物相互遮蔽の問題について対処している．

3.2.3 群集の動きの流れの学習に基づく大多数人物追跡

多くの研究者が取り扱う混雑環境は，駅構内や街頭といった観測視野内におよそ数十人の人物が存在する場合がほとんどである．その一方で，数百人から千人規模のより大多数の人物が視野内に存在するシーン（マラソン大会におけるスタート直後の選手の動き，メッカの巡礼，etc.）における人物追跡に関する研究が近年報告されている．

Ali らは，Floor Field と呼ばれるパニック時における人間の行動をモデリングしたセルオートマトンを追跡に利用することで，数百人規模の大多数人物追跡を実現している [2]．Floor Field は，観測環境に基づく静的なもの（Static Floor Field，SFF）と，人物を取り囲む他の複数人物の動きに基づく動的なもの（Dynamic Floor Field，DFF）の二種類が考えられる．

SFF とは「パニックが起きたとき人物は出入口へ向かって最短の経路を動こうとする」という動きの傾向をモデル化したものである．人物は自分の位置と出入口までに存在する障害物の位置などを考慮し，最適な経路を選択する．これに対し DFF とは「人物は周りを取り囲む多数の人物により形成される大きな動きの流れに従いやすい」という動きの傾向をモデル化したものである．

以上のような Floor Field に基づく手掛かりに加え，粉体力学の考え方により群集全体の大きな流れを検出する手法 [1] を利用する．これにより人物の状態遷移確率を求め，次フレームにおける位置推定を行うことができる．

このように，非常に混雑した状況特有の人物の動きの傾向を用いた研究もなされている．

3.2.4 特徴点軌跡のクラスタリング

人物の形状モデルや見えの情報をを用いた追跡手法は、人物がどのように観測されるのかを仮定することにより追跡を行っている。このような追跡手法とは対照的に、映像中の特徴点の動きのみを手掛かりに人物を検出する手法が提案されている。そのため、これらの手法は見えや形状など人物に関する事前知識を必要としない。

Brostow らは、画像中の局所特徴点の動きの軌跡をベイズの枠組みに基づきクラスタリングすることにより、混雑環境下において人物検出を実現している [7]。具体的には、二つの特徴点軌跡のペアの空間関係から事前確率を計算し、そしてその特徴点軌跡のペアの動きの一貫性を計測することにより事後確率が計算される。得られる事後確率を用い、最小記述長 (MDL) の考え方により再帰的に特徴点軌跡のペアをクラスタリングする。これにより人物検出が実現される。彼らは混雑した地下鉄の駅において頑健な人物検出を実現している。

Rabaud らは動きの一貫性の仮定に加え、同じ人物に属する特徴点は同じアフィン運動をするという仮定を導入することにより、屋外の混雑シーンにおいて個々の人物の追跡を実現している [74]。Li らは Brostow らと類似した追跡の枠組みに、特徴点軌跡のペアの動きの一貫性を測る尤度関数を事前に学習する処理を組み入れている [58]。これにより観測シーンに即した追跡を行うことができる。

3.2.5 本研究の位置付け

本節では、混雑環境下における人物追跡に関する先行研究について概観した。混雑環境における大きな問題である (部分) 遮蔽による追跡性能の低下、人物が互いに近接していることによる追跡の失敗に対処するために、人物の形状モデル、見えを利用したアプローチが一般にとられている。

先行研究において報告されているように、形状モデルの利用は一般に部分遮蔽について頑健である。しかしながら、視野内に数十人が存在するようなシーン (e.g., 図 3.1) において、背景差分により得られる前景領域から個々の人物に該当する領域を正しく検出することは難しい。

また見えに基づく手法は、頻繁に発生する遮蔽に伴い追跡対象が安定に観測されづらいことから、追跡性能の低下が懸念される。このような遮蔽の影響を抑えるために、近年提案された局所領域における見えを考慮した手法 [95] などは、混雑シーンにおいて有効であると考えられる。

一方で、画像中の特徴点の動きの軌跡群をクラスタリングすることにより追跡を実現するアプローチは、局所領域における情報を利用しているため、部分遮蔽について頑健であると言える。これらの手法は、特徴点の軌跡の動きの類似性を利用しているため、人物の見えや形状に関する知識を必要としない。そのため、観測が不安定な状況においても頑健な検出、追跡を実現することが可能となる。

しかしながら、混雑環境において個々の人物は周りに存在する多数の人物により動きを制限されるため、周りとは類似した動き方を示す傾向にある。そのため、特徴点軌跡の動きの類似度（動きの一貫性）を考慮するだけでは、個々の人物を区別するためには不十分である。

混雑環境特有の群集としての人物の動きの傾向を利用する手法は、個々の人物が自由に動くことはできないような、大多数の人物（数百人～千人程度）が非常に密接した状況を仮定している。本研究が対象とする観測環境は、駅構内や街頭といったおよそ数十人程度の人物が存在する混雑シーンを仮定しており、個々の人物は比較的自由に行動することができる。そのため、文献 [1,2] で提案されているような群集の動きの流れに基づく仮定を本研究に適用することは難しい。

以上のような先行研究に対し、本手法では個人性に基づく行動特徴である歩容特徴を人物追跡に利用する。歩容特徴は、個々の人物が持つ固有の性質であるため、個々の人物を区別するための有効な指標になると考えられる。また局所領域における見えの時間変動（一貫性）は、動きが類似する傾向にある混雑状況において個々の人物を区別するための大きな手掛かりとなる。これらの動きと見えに基づく指標を、部分遮蔽に対して頑健である特徴点軌跡のクラスタリングの枠組みで用いることにより、頑健な人物追跡を実現することが可能となる。

3.3 特徴点軌跡のクラスタリングに基づく人物追跡

本手法は、映像から得られる特徴点の動きの軌跡群を用いてグラフを形成し、それをクラスタリングすることにより人物追跡を実現する。以下本節では、本手法が用いる特徴点軌跡のクラスタリングの枠組みについて詳しく述べる。

3.3.1 特徴点軌跡の生成

最初に、観測される映像から特徴点の動きの軌跡群を生成する。本手法では、連続するフレーム間の特徴点の動きの推定に KLT (Kanade-Lucas-Tomasi) 特徴点追跡

法 [84, 93] を利用する．KLT による特徴点追跡を逐次行うことにより，特徴点の動きの軌跡を取得することができる．また遮蔽に伴い特徴点軌跡の数が減少することが懸念されるため，毎フレーム新たな特徴点の検出を行う．これと同時に，不定な特徴点軌跡のフィルタリング処理を行うことにより，信頼できる特徴点軌跡群のみを獲得することが可能となる．

KLT 特徴点追跡法

KLT 特徴点追跡法とは，時刻 t の画像 I における位置 \mathbf{x} にある特徴点が次のフレームにおいてどの位置に移動するか推定するアルゴリズムである．2次元画像平面上における特徴点の動きの推定は，言い換えると変形を表現するアフィンパラメータと2次元の移動ベクトルを求めることである．

時刻 t における位置 \mathbf{x} にある特徴点 $I(\mathbf{x})$ の時刻 $t+1$ における位置は，アフィン変形と画像平面上での移動を考慮することにより，次式で表現される．

$$J(A\mathbf{x} + \mathbf{d}) = I(\mathbf{x}), \quad (3.3.1)$$

ここで $J(\cdot)$ は時刻 $t+1$ における画像を意味し， A はアフィン行列， \mathbf{d} は移動量をそれぞれ意味する．

画像のノイズの影響，またはアフィンモデルが動きを正しく表現できない場合，(3.3.1) 式が必ずしも成り立つとは限らない．そこで，特徴点を表現するパッチ \mathcal{W} 内でその二乗誤差を最小化する枠組みを考えることにより，正しく特徴点の位置を推定することができる．パッチ \mathcal{W} 内における特徴点位置の二乗誤差は次式で表現される．

$$\epsilon = \int \int_{\mathcal{W}} [J(A\mathbf{x} + \mathbf{d}) - I(\mathbf{x})]^2 w(\mathbf{x}) d\mathbf{x}. \quad (3.3.2)$$

ここで $w(\mathbf{x})$ は重み関数を意味し，一般に定数，もしくはガウス分布が用いられる．

連続するフレーム間ではその移動量は微小であると仮定すると (3.3.2) 式は以下の線形方程式で表現される．

$$Z\mathbf{d} = \mathbf{e}. \quad (3.3.3)$$

ここで,

$$Z = \int \int_{\mathcal{W}} \mathbf{g}(\mathbf{x}) \mathbf{g}^T(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}, \quad (3.3.4)$$

$$e = \int \int_{\mathcal{W}} [I(\mathbf{x}) - J(\mathbf{x})] \mathbf{g}(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}, \quad (3.3.5)$$

$$\mathbf{g}(\mathbf{x}) = \left[\frac{\partial J}{\partial x}, \frac{\partial J}{\partial y} \right]^T, \quad (3.3.6)$$

である (3.3.3) 式を解くことにより, 次フレームにおける特徴点位置が推定される.

このような処理を逐次行うことにより, 特徴点の動きの軌跡を獲得することができる.

特徴点軌跡のフィルタリング

頻繁に発生する遮蔽, 複雑背景に基づく影響に伴い, 信頼しうる特徴点軌跡が減少することが懸念される. そこで本手法では, [74] において行われていたように, 毎フレーム新たな特徴点検出を行い, 特徴点軌跡数の減少を防ぐ. 更に得られる特徴点軌跡のうち信頼度が高いものを獲得するために, 以下の3つに該当する特徴点軌跡を除去する.

1. ほとんど動かない特徴点軌跡

本手法では, 観測視野内の人物はいずれも移動していることを仮定しているため, その特徴点は背景に属している可能性が考えられる. ある一定時間特徴点の移動量を計測し, それが微小であるなら背景に属した特徴点と見なし, その軌跡を除去する.

2. 連続するフレーム間での移動量が大きすぎる特徴点軌跡

観測環境の影響により, KLT 追跡器は誤った特徴点位置を推定することが考えられる. そこで, 次フレームにおける特徴点の推定位置が明らかに離れているものを除外する.

3. 生存時間が短い特徴点軌跡

時間方向の軌跡長が明らかに短い軌跡は信頼度が低いと見なす. 詳しくは後述するが, クラスタリングに利用する時間窓のサイズよりも短い軌跡を除外する.

以上のフィルタリング処理により得られる特徴点軌跡群の例を図 3.2 に示す. 生成された特徴点軌跡群は, 移動する人物にのみ属していることが見て取れる.

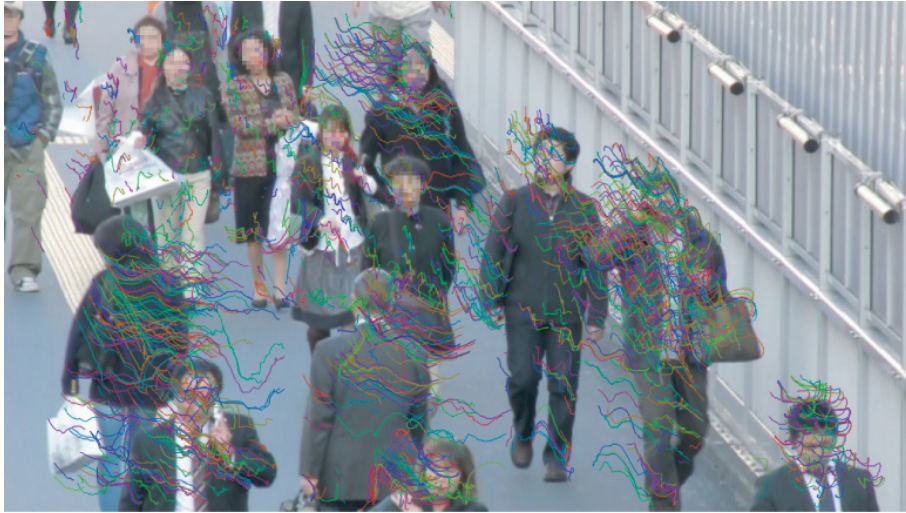


図 3.2: 特徴点軌跡の生成 .

3.3.2 特徴点軌跡群のグラフのクラスタリング

本手法における特徴点軌跡クラスタリングは、各特徴点軌跡をノードと見なした時空間グラフのクラスタリングにより実現される。

初期グラフの構築

KLT 追跡器により得られる特徴点軌跡群を用い、初期グラフ $G^0 = (V, E^0)$ を構築する。ここで V はノードの集合、すなわち特徴点軌跡群を表す。具体的には、時間窓 τ 内に存在する特徴点軌跡群を利用する。また E^0 は、初期グラフにおけるノード間に張られるエッジの集合を現す。

グラフ構築のアルゴリズムには、ドロネー三角形分割アルゴリズム [14] を利用する。得られる特徴点軌跡の集合についてドロネー三角形分割を適用することにより、初期グラフ G^0 を得る。図 3.3-(a) に概要を示す。また実際の映像における初期グラフの例を図 3.4 に示す。これらの図にあるように、初期グラフは特徴点軌跡を頂点に持つ三角パッチの集合により構成される。

軌跡間相違度に基づくクラスタリング

二つのノード（特徴点軌跡） $p \in V, q \in V$ を繋ぐエッジは、軌跡間の相違度を表す重みを持つ。軌跡間相違度は、先行研究 [7, 58, 74] において利用されていた空間関

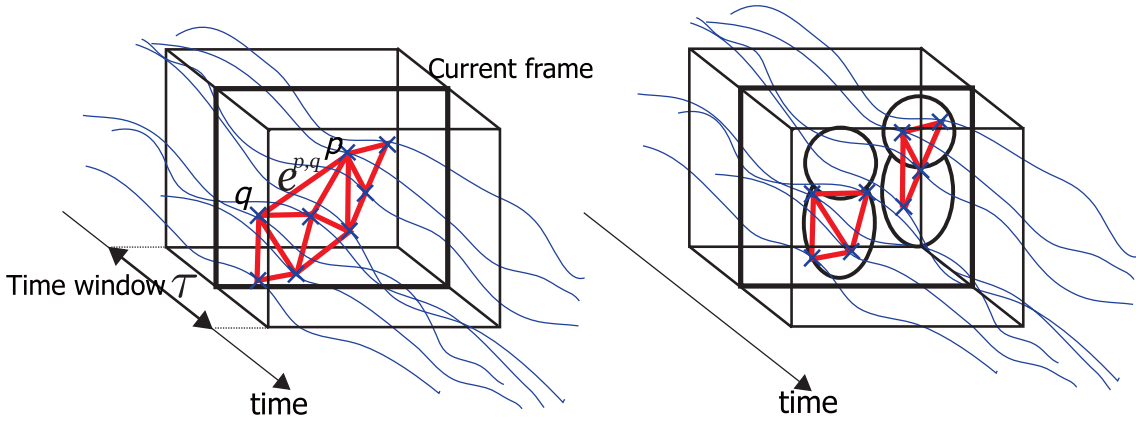


図 3.3: 提案手法概要 . (a) グラフの構築 . 時間窓 τ 内における特徴点軌跡群をノードとして用いることにより構成される . (b) クラスタリング結果 . 得られる特徴点軌跡間相違度に基づきエッジを除去することにより実現される .

係, 動きの一貫性に基づく尺度に加え, (1) 歩容特徴, (2) 局所領域の見えの時間変化の二つの新たな指標を用いることで計算される . それぞれの指標に基づき計算された重み (相違度) をもとに, 最終的な二つの軌跡 p, q の間の相違度 $d^{p,q}$ は以下のように計算される .

$$d^{p,q} = d_{freq}^{p,q} \cdot d_{app}^{p,q} \cdot d_{prox}^{p,q} \cdot d_{coh}^{p,q}, \quad (3.3.7)$$

ここで (3.3.7) 式の右辺はそれぞれ, 歩容特徴に基づく重み $d_{freq}^{p,q}$, 局所領域の見えの変化に基づく重み $d_{app}^{p,q}$, 空間関係に基づく重み $d_{prox}^{p,q}$, そして動きの一貫性に基づく重み $d_{coh}^{p,q}$ を意味する . これらの指標の詳細については次節で詳しく述べる .

ある指標に基づく重みが 0 になった場合, 他の指標に基づく重みの大きさに拘らず最終的な相違度は常に 0 になってしまう問題が考えられる . このような状況を避けるために, いずれかの重みが 0 になった場合, 事前に定めた定数 $d_{min} > 0$ で置き換える .

計算される軌跡間相違度に基づき, 初期グラフ G^0 をクラスタリングする (図 3.3-(b)) . クラスタリングにより得られた部分グラフ $G = (V, E) = \{G^{(1)}, \dots, G^{(n)}\}$ がそれぞれの追跡対象に相当する . 本手法におけるクラスタリング処理は, 相違度 $d^{p,q}$ がある閾値 th_p 以上のエッジを除去することにより実現される . すなわちクラスタリングにより得られる部分グラフ G を構成するエッジの集合 E は次式で与えられる .

$$E = \{e^{p,q} \mid p, q \in V, p \neq q, d^{p,q} < th_p\}. \quad (3.3.8)$$

このような特徴点軌跡のクラスタリングにより、人物検出が実現される。

クラスタリング結果に基づく人物追跡

特徴点軌跡のクラスタリングは、本手法では $\tau/4$ フレーム毎に行われる。そのため、人物追跡を実現するためには、毎クラスタリング処理で得られるクラスタ集合間の対応付けを行う必要がある。

クラスタの対応付けは、時刻 t におけるクラスタ集合 $\{G^{(i,t)}\}$ と、 $\tau/4$ フレーム前のクラスタリング結果、すなわち過去直近のクラスタリングにより得られる集合 $\{G^{(j,t-\tau/4)}\}$ 間において行われる。具体的には、各集合におけるクラスタ $G^{(i,t)}$ 、 $G^{(j,t-\tau/4)}$ が共有する特徴点軌跡数に基づき行われる。特徴点軌跡の消失などを考慮し、以下の評価式を満たすクラスタを対応付ける。

$$\text{Share}(G^{(i,t)}, G^{(j,t-\tau/4)}) > 0.8n^s, \quad (3.3.9)$$

ここで $\text{Share}(A, B)$ は、クラスタ A, B 間で共有する特徴点軌跡の個数を与える関数であり、 n^s は比較する二クラスタが持つ特徴点軌跡の個数のうち小さい方をとるものとする。この対応付け処理は全てのクラスタ間で行われる。

更に、不安定なクラスタリングの影響を抑えるために、隣接する両時刻において対応付けがなされているもの、すなわち $t - \tau/4, t, t + \tau/4$ それぞれにおいて対応付いているもののみを追跡結果と見なす。

3.4 軌跡間の相違度の計算

二つの特徴点軌跡（ノード）間のエッジの重みは、歩容特徴、局所的な見えの変化、そして先行研究でも利用されている動きの一貫性、空間関係の4つの指標に基づき計算される。これにより、これらの特徴点軌跡が同じ人物に属しているかどうか判断する。本節ではそれぞれの指標と類似度の計算について詳しく説明する。

3.4.1 周波数領域における歩容特徴

周波数領域における歩容特徴は、人物個々が持つ固有の特性であり、混雑環境下において各人物を区別するための有効な指標である。本節では、特徴点軌跡からの歩容特徴の抽出、相違度計算について詳しく述べる。

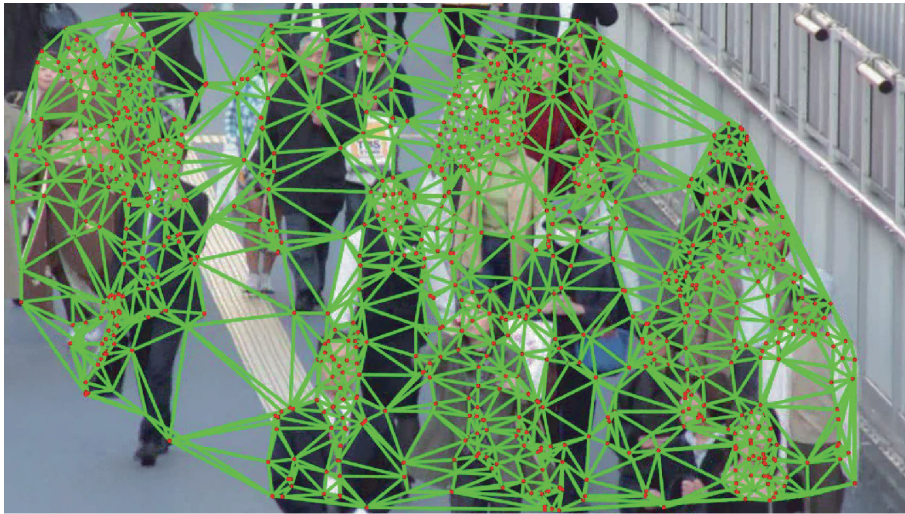


図 3.4: ドロネー三角形分割によるグラフ構造 .

歩容特徴の抽出

人物が歩くときに観測される歩容特徴は，人間の身体的な構造から垂直方向の周期的な動きに現われると仮定する．この仮定に基づき，特徴点軌跡の垂直方向の時間変位 $y(t)$ について着目する．

観測視野内において人物は移動していることを想定しているため，クラスタリングに利用する短時間窓 τ 内における特徴点の垂直方向の時間変化 $y(t)$ は，歩行による周期運動を伴う直線とみなすことができる．この仮定に基づき，最小二乗法による直線当てはめを $y(t)$ に対して行い，垂直方向の人物の並進移動量 $\bar{y}(t)$ を得る．歩容特徴に相当する周期成分 $y_p(t)$ は，垂直方向の時間変化 $y(t)$ から並進移動量 $\bar{y}(t)$ を差し引くことにより得られる，すなわち，

$$y_p(t) = y(t) - \bar{y}(t) . \quad (3.4.1)$$

図 3.5 に本処理の概略を示す．

歩容特徴に基づく相違度

歩容特徴は，歩行周期とその位相に特徴的な性質が表れると考えられる．そこで，特徴点軌跡より得られる周期成分 $y_p(t)$ について離散フーリエ変換を行う．周波数領

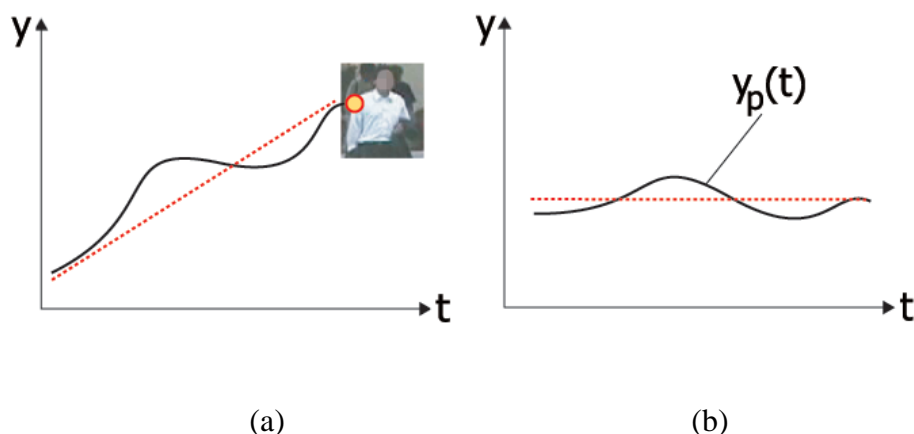


図 3.5: 周波数領域における歩容特徴計測のための前処理。(a) 歩容特徴に相当する垂直方向の周期的な動きの抽出。得られる特徴点の垂直方向の動き $y(t)$ について直線当てはめを行い、並進移動量 $\bar{y}(t)$ を得る。垂直方向の動き $y(t)$ から差し引くことにより、周期成分 $y_p(t)$ が得られる。(b) 抽出された垂直方向の動きの周期成分 $y_p(t)$ 。

域における歩容特徴は次式で与えられる。

$$y_k = \sum_{t=t_s}^{t_s+\tau} y_p(t) \exp(-j\omega_0 kt), \quad (3.4.2)$$

ここで $\omega_0 = 2\pi/\tau$ は基本角周波数であり、 $j = \sqrt{-1}$ は虚数単位である。周波数帯 k における振幅スペクトル a_k 、位相成分 ϕ_k は以下のように与えられる。

$$a_k = \sqrt{(\text{Re}[y_k])^2 + (\text{Im}[y_k])^2}, \quad (3.4.3)$$

$$\phi_k = \arctan\left(\frac{\text{Im}[y_k]}{\text{Re}[y_k]}\right). \quad (3.4.4)$$

$\text{Re}[\cdot]$ 、 $\text{Im}[\cdot]$ はそれぞれ実数成分、虚数成分を取り出す作用素を意味する。

歩容特徴に基づく二つの特徴点軌跡の相違度 $d_{freq}^{p,q}$ は、振幅スペクトル、振幅成分が最も含まれている周波数帯における位相成分の二つの尺度を用いることにより計算される。振幅スペクトルに基づく相違度は、それぞれの軌跡の振幅スペクトル間のユークリッド距離を計算することにより求められる。この評価尺度は、歩容認証に関する研究分野において一般に使われている [63]。また、位相に基づく相違度については単純に差の絶対値を利用する。これより、歩容特徴に基づく軌跡間の相違

度 $d_{freq}^{p,q}$ は次のように計算される．

$$d_{freq}^{p,q} = \sqrt{\sum_{k=0}^{\tau/2} [a_k^p - a_k^q]^2 |\phi^p - \phi^q|}, \quad (3.4.5)$$

ここで a_k^p と ϕ^p は p 番目の特徴点軌跡の k 番目の周波数帯の振幅スペクトル，振幅成分が最も含まれる周波数帯における位相成分をそれぞれ意味する．また動きの周期成分 $y_p(t)$ は実数値信号なので，半分の周波数帯のみ考慮すればよい．

3.4.2 局所領域における見えの時間変動

歩容特徴は，混雑状況において個々の人物を特徴づける有効な指標であるが，混雑環境下では周囲の人物の歩容特徴も類似したものになる傾向にある．このような問題へ対処するために，特徴点軌跡の動きに基づいたものとは異なる指標が必要となる．そこで本手法では，局所領域における見えの時間的な変動を手掛かりとして利用する．

初期グラフにおいて連結された3つの特徴点軌跡によって張られる三角パッチを考え，そのパッチ内の見えの時間変動を利用する．図3.6に概略図を示す．同じ人物に属する特徴点軌跡により張られる三角パッチ内の見えは，時間的にほぼ変化しないことが期待される．それに比して，他の人物との間に張られる三角パッチは，背景や他の人物を含むためその時間変動は大きくなると考えられる．この指標を利用することにより，動きが類似する傾向にある混雑環境においても個々の人物を区別することが可能となる．

見えに関する尺度は，色ヒストグラムをはじめ，テクスチャ情報や HoG (Histogram of Gradient) [19] など様々なものが考えられるが，本手法では簡単のため HS (Hue-Saturation) 色ヒストグラムの時間変動を考える．具体的には，時間窓 τ 内の各時刻の色ヒストグラム $\mathbf{h}(t)$ と，その平均色ヒストグラム $\bar{\mathbf{h}}$ を計算し，各時刻において Bhattacharyya 距離を求める．得られる Bhattacharyya 距離の RMS (Root Mean Square) を計算し，これを見えに基づく相違度 $d_{app}^{p,q}$ とする，すなわち，

$$d_{app}^{p,q} = \sqrt{\frac{1}{\tau} \sum_{t=t_s}^{t_s+\tau} d_{hist}^2(\mathbf{h}(t), \bar{\mathbf{h}})}. \quad (3.4.6)$$

ここで t_s は時間窓の最初のフレームを意味し， $d_{hist}(\mathbf{h}(t), \bar{\mathbf{h}})$ は Bhattacharyya 距離，

$$d_{hist}(\mathbf{h}(t), \bar{\mathbf{h}}) = \sum_{j=1}^m \sqrt{h_j(t) \bar{h}_j}, \quad (3.4.7)$$

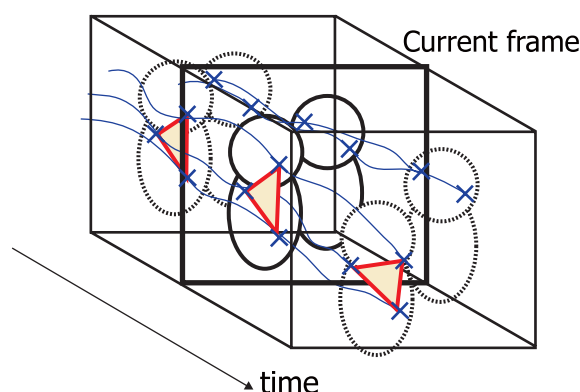


図 3.6: 特徴点軌跡により張られる三角パッチの利用．三角パッチは連結されている特徴点軌跡により張られ，パッチ内の見えの時間変動が利用される．

で与えられる．ここで m はヒストグラムのビンの数， $h_j(t)$ はヒストグラム $\mathbf{h}(t)$ の j 番目のビンの度数を意味する．

得られる相違度 $d_{app}^{p,q}$ は，三角パッチを構成するすべてのエッジに割り当てられる．またエッジは隣り合う二つの三角パッチにより共有される可能性があるが，この場合は二つの三角パッチのうち相違度が大きい方を割り当てるものとする．

3.4.3 軌跡間の空間関係

先行研究 [7, 58, 74] において利用されているように，特徴点軌跡間の空間関係，動きの一貫性（次節）に基づいた相違度を評価する．

二つの特徴点軌跡が同じ人物に属しているならば，それらは空間的に近い距離にあると考えられる．そこで，本手法では空間関係に基づく相違度 $d_{prox}^{p,q}$ に時間窓内における二つの軌跡間で最大となる距離を利用する．

特徴点軌跡 p, q の時刻 t における位置をそれぞれ $(x^p(t), y^p(t))$, $(x^q(t), y^q(t))$ としたとき，相違度 $d_{prox}^{p,q}$ は以下のように計算される．

$$d_{prox}^{p,q} = (\Delta x)_{\max} \cdot (\Delta y)_{\max} , \quad (3.4.8)$$

ここで， $(\Delta x)_{\max}$, $(\Delta y)_{\max}$ はそれぞれ正規化された水平方向，垂直方向の最大距離

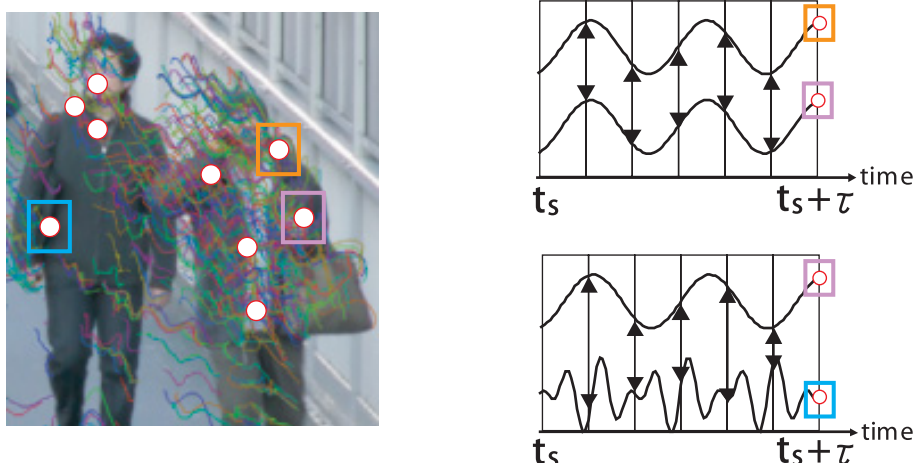


図 3.7: 特徴点軌跡間の空間関係と動きの一貫性．同じ人物に属する特徴点軌跡は近接しており，類似した動き方をする．

を表し，

$$(\Delta x)_{\max} = \frac{\max_{t_s \leq t \leq t_s + \tau} |x^p(t) - x^q(t)|}{\sum_{t=t_s}^{t_s + \tau} |x^p(t) - x^q(t)|}, \quad (3.4.9)$$

$$(\Delta y)_{\max} = \frac{\max_{t_s \leq t \leq t_s + \tau} |y^p(t) - y^q(t)|}{\sum_{t=t_s}^{t_s + \tau} |y^p(t) - y^q(t)|}, \quad (3.4.10)$$

である．

3.4.4 動きの一貫性

空間関係に基づく相違度に加え，動きの一貫性に基づく相違度を評価する．二つの特徴点軌跡が同じ人物に属しているならば，時間窓内において同じ距離関係を保ちやすいと考えられる（図 3.7）．そこで，時間窓内における二つの軌跡 p, q 間の距離関係の標準偏差を，動きの一貫性に基づく相違度 $d_{coh}^{p,q}$ として利用する．具体的には以下のように計算される．

$$d_{coh}^{p,q} = \sqrt{\frac{1}{\tau} \sum_{t=t_s}^{t_s + \tau} (l^{p,q}(t))^2 - \left(\frac{1}{\tau} \sum_{t=t_s}^{t_s + \tau} l^{p,q}(t) \right)^2}, \quad (3.4.11)$$

ここで $l^{p,q}(t)$ は、時刻 t における特徴点軌跡 p, q 間のユークリッド距離、

$$l^{p,q}(t) = \sqrt{(x^p(t) - x^q(t))^2 + (y^p(t) - y^q(t))^2}, \quad (3.4.12)$$

である。

3.5 実験・考察

本手法の有用性を確認するために、CG データと実際の群集の映像を用いてそれぞれ実験を行った。実験に用いた環境は、OS: Windows XP, CPU: Intel Core2 Quad 2.66GHz, RAM: 3.00GB である。歩容特徴と局所的見えの効果を検証するために、歩容特徴と局所的見えの変化に基づく指標を除いたものを基準のシステムとして用いた。これは、軌跡間相違度を空間関係と動きの一貫性のみ、すなわち ($d^{p,q} = d_{prox}^{p,q} \cdot d_{coh}^{p,q}$) で計算することを意味する。この基準システムは特に明記されない限り以下の全ての実験において同じものが利用される。

3.5.1 CG 実験

本手法で提案する歩容特徴、局所的見えの時間変化の二つの指標の有効性を確認するために、近接した二つのターゲットが同じ方向に同じ速度で動いている人工データについて実験を行った。また歩容特徴は人物の歩行における垂直方向の周期的な動きに表れるため、ターゲットの動きに垂直方向の周期的な動きを付加した。具体的には、各ターゲットの動き $x(t), y(t)$ は以下の式により与えた。

$$x(t) = at, \quad (3.5.1)$$

$$y(t) = x(t) + b + c \sin(dt + e). \quad (3.5.2)$$

ここで a, b, c, d, e はそれぞれ各ターゲットの動きを特徴づけるパラメータである。

歩容特徴の利用の効果

本実験ではまず歩容特徴の効果を確認した。隣接した二つのターゲット A, B は同じ方向、同じ速度で動くが、垂直方向の周期的動き（歩容特徴にあたるもの）の周波数、位相は互いに異なるものとした。具体的には (3.5.2) 式における各パラメー

夕を，ターゲット A : $a = 1, b = -70, c = 5, d = 0.3142, e = 3.14$ ，ターゲット B : $a = 1, b = -130, c = 5, d = 0.1653, e = 1.57$ とした．また歩容特徴の性能向上への寄与を確認するために，本実験では局所領域における見えの変化の指標は利用しないものとした．

図 3.8 に追跡の比較結果の一部を示す．基準としたシステム（図 3.8-(b)）では，近接する二つのターゲットを一つの大きなターゲットとみなして誤追跡してしまっていることがわかる．それに対し歩容特徴を組み入れた追跡（図 3.8-(a)）では，正しく二つのターゲットを追跡できていることがわかる．この結果より，歩容特徴の利用が追跡性能の向上に寄与していることがわかる．

局所領域における見えの一貫性の利用の効果

混雑環境における個々の人物の区別において最も難しいケース，すなわち動きの周波数特徴についても周りの人物と同調してしまう場合を想定し，二つのターゲットの動きが全く同じケースについて実験した．このようなケースにおいて実験をすることにより，局所領域における見えの変化に基づく指標の効果を確認することができる．そのため，本実験において基準とするシステムでは，空間関係，動きの一貫性そして歩容特徴の指標に基づき軌跡間相違度 ($d^{p,q} = d_{freq}^{p,q} \cdot d_{prox}^{p,q} \cdot d_{coh}^{p,q}$) を計算した．各ターゲットの動きのパラメータは，ターゲット A : $a = 1, b = -70, c = 5, d = 0.3142, e = 3.14$ ，ターゲット B : $a = 1, b = -130, c = 5, d = 0.3142, e = 3.14$ とした．

実験の結果，基準としたシステムでは，前節の実験と同様に二つのターゲットを一つの大きなターゲットとして誤追跡してしまった．これに対し見えの時間変化に基づく指標を組み入れた追跡手法（本手法）は，正しく二つのターゲットを追跡することができた．

図 3.9，図 3.10 に，ある時間窓における一つの三角パッチ内の平均ヒストグラム \bar{h} ，ヒストグラム距離 $d_{hist}(h(t), \bar{h})$ のばらつきをそれぞれ示す．図 3.9 の左図は背景を含む三角パッチ内平均ヒストグラムを，右図は一つのターゲット内に張られる三角パッチ内平均ヒストグラムをそれぞれ意味する．また図 3.10 において，赤色でプロットされた点群は一つのターゲットに属した三角パッチのヒストグラム距離の結果，緑色の点群は二つのターゲット間に張られたもの，すなわち背景を含むパッチの結果を意味する．この結果からわかるように，背景を含む三角パッチに比して一つのターゲットに属した三角パッチ内の色ヒストグラムはほぼ一貫していることがわかる．



frame No: 66



frame No: 123



frame No: 279

(a)

(b)

図 3.8: 歩容特徴の利用の効果 (a) 歩容特徴を組み入れた追跡結果 (b) 歩容特徴無しの追跡結果 .

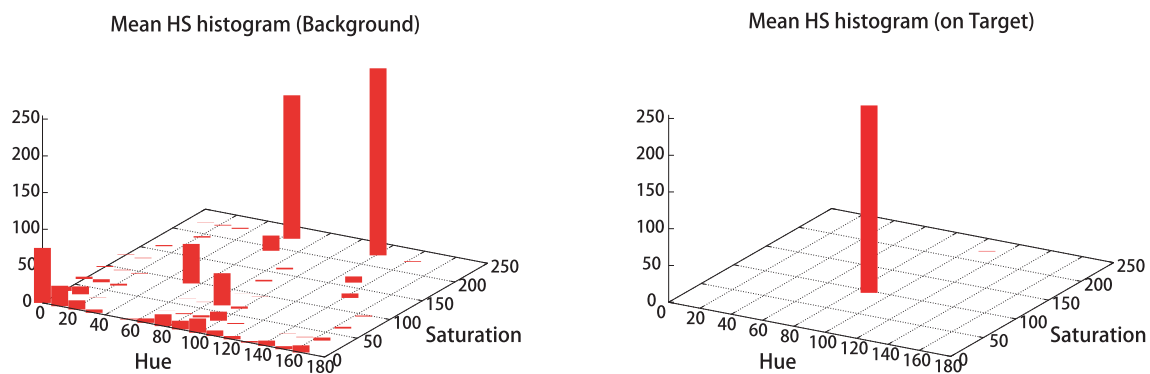


図 3.9: ある時間窓内における一つの三角パッチ内平均 HS 色ヒストグラム \bar{h} の比較 . 左図 : 背景を含む三角パッチの HS 平均ヒストグラム , 右図 : ターゲット内三角パッチの HS 平均ヒストグラム .

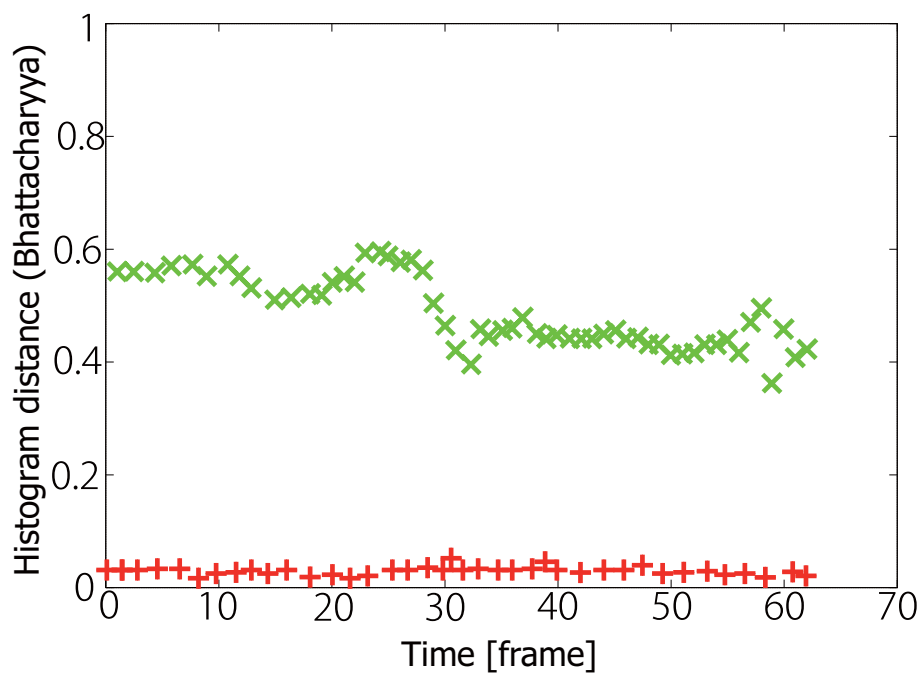


図 3.10: ヒストグラム距離 $d_{hist}(h(t), \bar{h})$ の時間変動結果 . 赤色の点は一つのターゲット内に張られた三角パッチによるものを意味し , 緑色の点は二つのターゲット間に張られた三角パッチによるものを表す .

表 3.1: 実環境下追跡実験で用いた映像のパラメータ

	解像度	fps	総フレーム数
シーケンス (A)	1280x720	60	700
シーケンス (B)	1280x720	60	700
シーケンス (C)	800x600	30	520
UCSD data [74]	720x480	30	1540

これより見えの時間変動に基づく指標は二つのターゲットを区別するための効果的な指標であると言える。

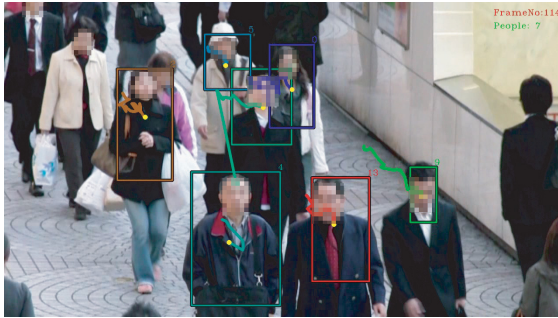
3.5.2 実環境下追跡実験

実際の群集の映像を用いて本手法の有用性を確認した。本実験では、それぞれ異なる場所から撮影した3つのシーケンス(A), (B), (C)と、UCSDのデータセット[74]を利用した。表3.1に用いた映像の詳細を示す。また本実験で利用したパラメータを表3.2に示す。エッジの枝打ち(クラスタリング)のための閾値 th_p は、予め獲得した軌跡間相違度 $d^{p,q}$ の集合からメディアン値をとることにより決定した。また時間窓 τ は、一般的な人物の歩行周期(およそ1秒)を確保するように、データのフレームレートに応じて決定した。

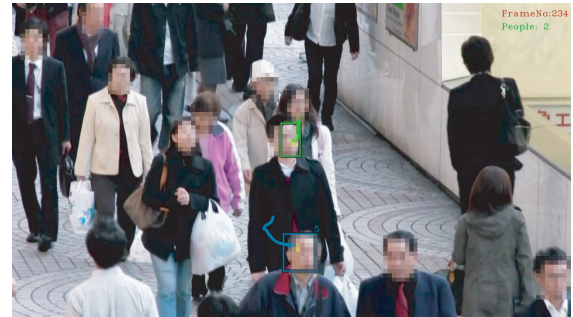
図3.11, 3.12, 3.13, 3.14にそれぞれのシーケンスにおける追跡比較結果を示す。いずれのシーンにおいても、基準システム($d^{p,q} = d_{prox}^{p,q} \cdot d_{coh}^{p,q}$)に対して提案手法の方がより多くの人物を正しく追跡できていることが確認できる。

追跡性能を定量的に評価するために、再現率と適合率を用いた。また本実験ではfalse positiveを、(1)一人の人物に複数のクラスタが生成される場合、(2)複数の人物が一人の人物として追跡される場合、の二種類とした。図3.15に例を示す。

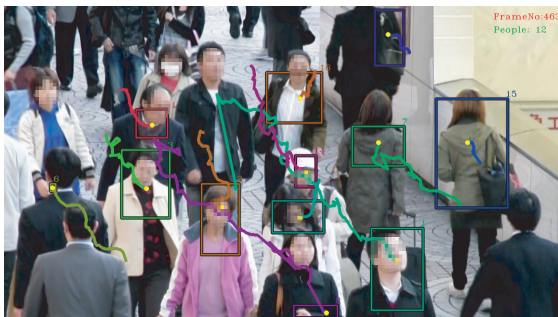
表3.3, 図3.16に各シーケンスにおける追跡結果を示す。いずれのシーケンスにおいても提案手法の方が追跡性能が向上していることがわかる。主に再現率についてその差が顕著であり、基準としたシステムと本手法にはおよそ2倍の性能差が見られる。このことから、歩容特徴と局所領域における見えの時間変動に基づいた指標が追跡性能に大きく寄与していると言える。



frame No: 114

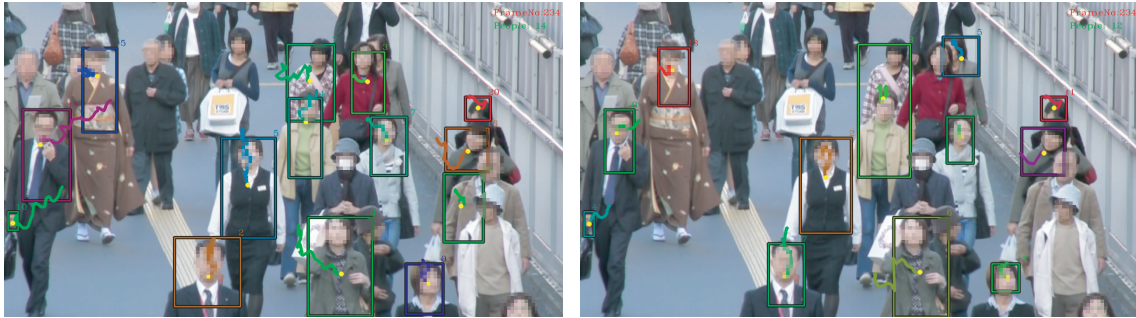


frame No: 234



frame No: 463

図 3.11: シーケンス (A) における追跡結果比較 (左図: 提案手法, 右図: 基準システム)



frameNo: 234



frameNo: 353



frameNo: 455

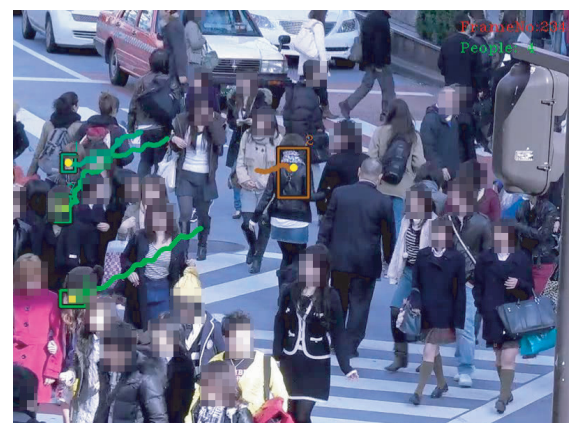
図 3.12: シーケンス (B) における追跡結果比較 (左図: 提案手法, 右図: 基準システム)



frameNo: 147

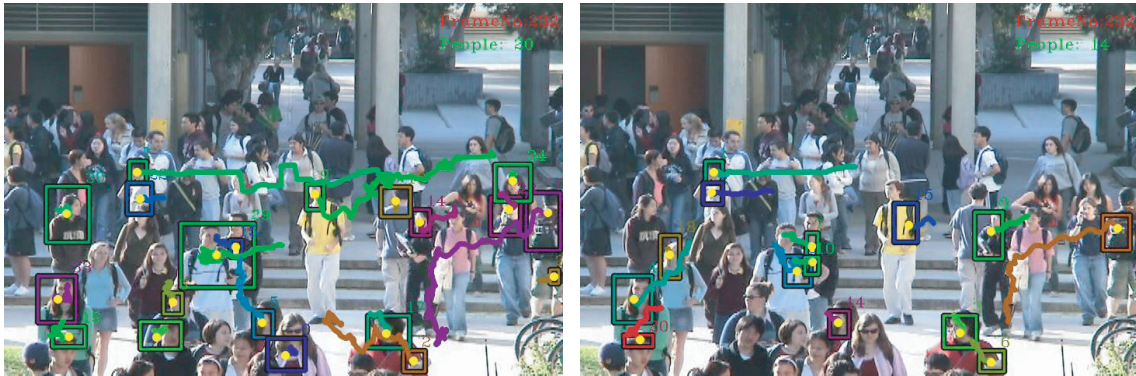


frameNo: 181

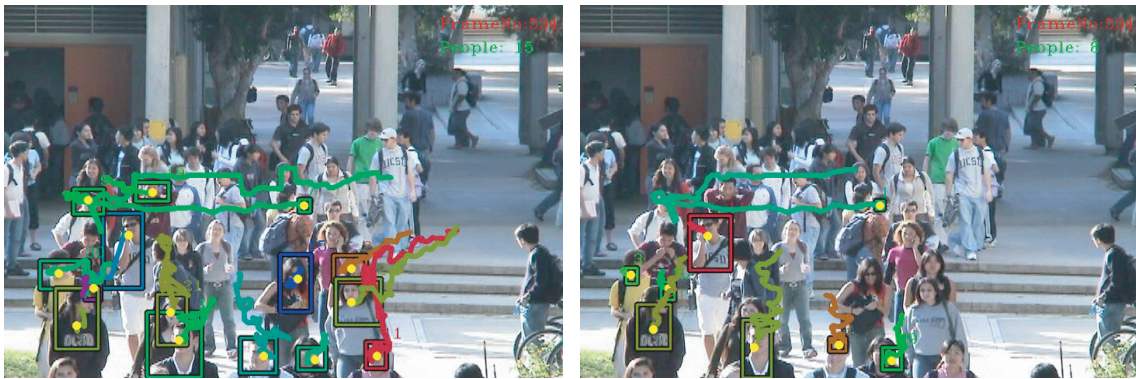


frameNo: 234

図 3.13: シーケンス (C) における追跡結果比較 (左図: 提案手法, 右図: 基準システム)



frameNo: 232



frameNo: 534



frameNo: 871

図 3.14: UDSD データ [74] における追跡結果比較 (左図: 提案手法, 右図: 基準システム)

表 3.2: 実環境下追跡実験に利用したパラメータ．閾値 th_p は軌跡間相違度 $d^{p,q}$ の集合のメディアン値とした．相違度の集合の獲得のために利用したシーケンスは実験に含まれない．

	τ	提案手法		基準システム	
		e_{\min}	th_p	e_{\min}	th_p
シーケンス (A)	64	0.1	0.27	0.1	64.8
シーケンス (B)	64	0.1	1.38	0.1	560.0
シーケンス (C)	40	0.1	0.68	0.1	66.6
UCSD Data [74]	40	0.1	1.11	0.1	207.7

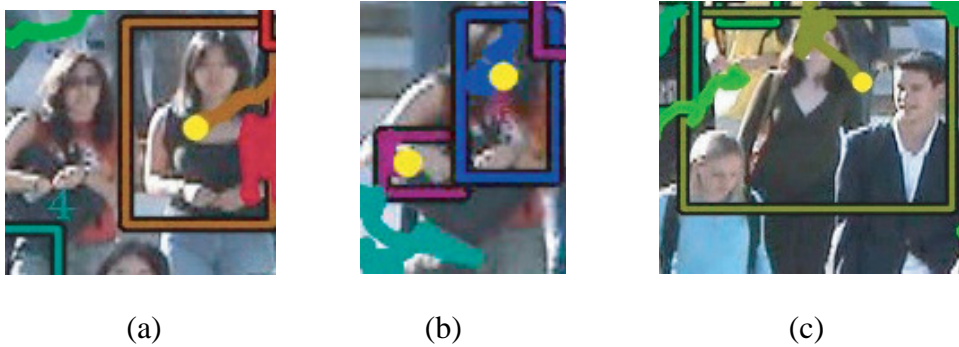


図 3.15: False positive と False negative の例．(a) 人物の未検出 (False negative)．(b) 一人の人物に複数のクラスターが生成される場合 (False positive-(1))．(c) 複数の人物が一人の人物として追跡される場合 (False positive-(2))．

表 3.3: 追跡結果： N は視野内における平均人数， μ ， σ はそれぞれ平均，標準偏差を表す．また t_c はクラスタリング処理に要する平均計算時間を表す．

	N	提案手法						基準システム					
		再現率		適合率		t_c [sec]	再現率		適合率		t_c [sec]		
		μ	σ	μ	σ		μ	σ	μ	σ			
シーケンス (A)	14.0	0.581	0.133	0.953	0.069	30.45	0.269	0.162	0.898	0.149	4.16		
シーケンス (B)	22.0	0.562	0.089	0.984	0.031	28.75	0.435	0.139	0.960	0.071	4.33		
シーケンス (C)	29.4	0.505	0.133	0.911	0.090	10.09	0.286	0.135	0.880	0.100	2.23		
UCSD data [74]	21.3	0.591	0.114	0.883	0.064	6.75	0.353	0.140	0.864	0.101	3.33		

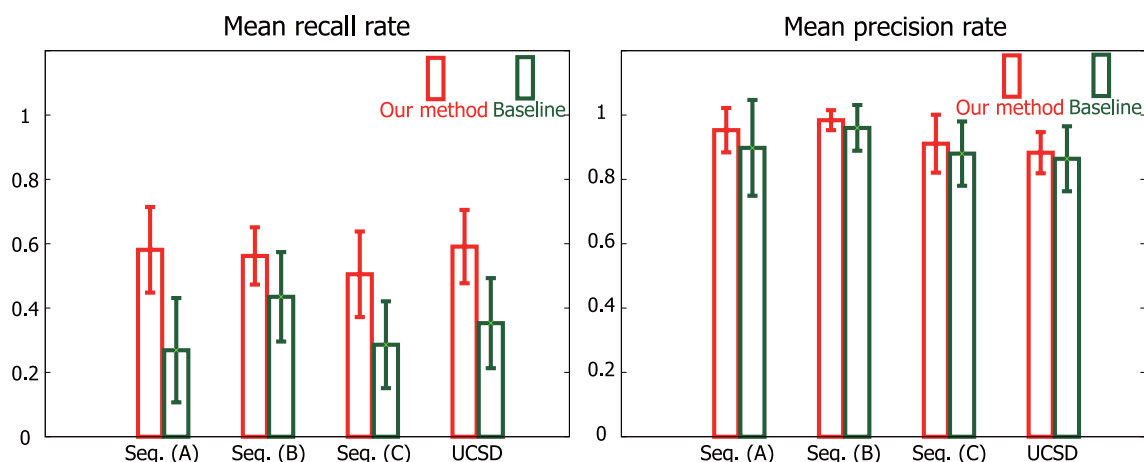


図 3.16: 追跡結果．左のグラフはそれぞれのシーケンスにおける再現率（平均，標準偏差）を表し，右のグラフは各シーケンスにおける適合率を表す．

3.5.3 追跡失敗に関する考察

表 3.4 に，それぞれのシーケンスにおける追跡失敗率の詳細（false negative rate，2 種類の false positive rate）を示す．この表からわかるように，それぞれのシーケンスにおいて false negative rate の方が false positive rate よりも高いことが見てとれる．主な原因は，クラスタリングすべき特徴点軌跡がほとんど発生していないことによると考えられる．図 3.17-(a) に示す実際の失敗例にあるように，人物がテクスチャがほとんどないような服装をしている場合，もしくはカメラに対して後ろ向きである場合，明らかに特徴点軌跡の数が不足していることが見て取れる．

このような問題に対処するために，追跡すべき特徴点を SIFT や SURF といったより安定した局所特徴量を利用することが考えられる．これに加え，特徴点追跡のアルゴリズムの改善も考慮すべき課題である．例を挙げると，SIFT 特徴量を追跡すべき特徴点とし，Mean Shift アルゴリズム [16] と組み合わせた特徴点追跡手法 [94, 108] などの利用が考えられる．

False positive rate について着目すると，(2) の“複数の人物が 1 人の人物として追跡される場合”に比べて (1) の“1 人の人物に複数のクラスタが生成される場合”の方が頻繁に発生していることがわかる．これは，同じ人物に属しているにも拘らず特徴点の動きが明らかに異なっていることに起因すると考えられる．実際，人物の動きには，腕を振る動作や後方を振り向く動作など，移動に関する動作に加え様々なものが観測される．このような同一人物内における動きのばらつきの問題につい

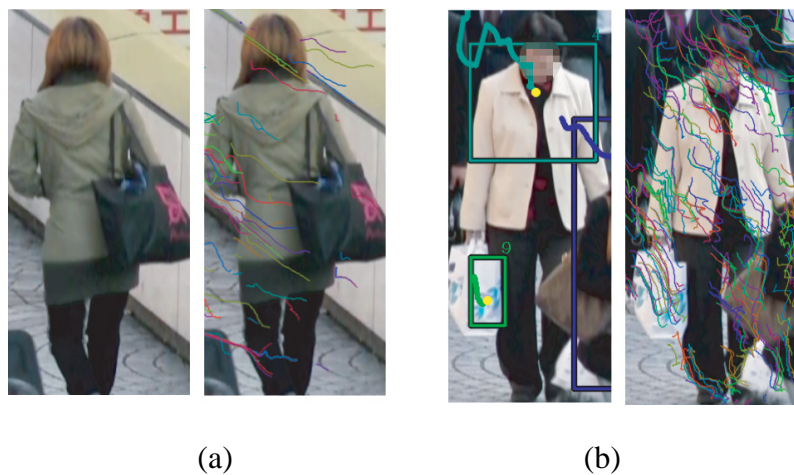


図 3.17: 追跡失敗例．(a) 特徴点軌跡の不足に起因する人物の未検出，(b) 周波数特徴の違いに起因した過検出．

表 3.4: 追跡失敗率の詳細．FN は false negative のサンプル数，FP は false positive のサンプル数の平均をそれぞれ意味する．また (1)，(2) はそれぞれ，(1) 1 人の人物に複数のクラスターが生成される場合，(2) 複数の人物が 1 人の人物として追跡される場合，の false positive を意味する．

	提案手法			基準システム		
	FN	FP		FN	FP	
		(1)	(2)		(1)	(2)
シーケンス (A)	5.70	0.32	0.09	9.80	0.50	0.05
シーケンス (B)	9.50	0.22	0.00	12.95	0.10	0.30
シーケンス (C)	12.90	1.05	0.22	19.25	1.03	0.08
UCSD data [74]	7.33	1.35	0.33	12.94	1.04	0.19

て対処する必要がある。

動きのばらつきに関する問題に加え、人物の体、手、人物が持つバッグなどの領域において複数のクラスタが発生することが観測された。これは、図 3.17-(b) の右図（特徴点軌跡の発生の様子）からもわかるように、人物の体の動きの周期性（周波数特徴）とバッグの動きのそれが明らかに異なっていることに起因すると考えられる。本手法は、特徴点軌跡の動き周波数特徴に基づきクラスタリングを行っているため、このような場合、システムは二人の人物が存在しているとみなしてしまっていると考えられる。

より頑健な人物追跡を実現するために、本節で議論した問題へ対処する必要がある。

3.6 本手法のまとめ

本章では、人物の歩容特徴と局所領域における見えの時間変動の一貫性を利用した混雑環境下における人物追跡手法を提案した。特徴点軌跡のクラスタリングに基づく考え方にに基づき、これらの指標を追跡の枠組みへ組み入れることにより、頑健な人物追跡が可能となる。人工データ、実際の群集の映像を用いた実験を通じて、歩容特徴は混雑環境において個々の人物の識別に有効であることを確認した。更に、局所領域の見えの時間変動を計測することにより、動きが類似する傾向にある混雑状況において有効な指標となることを確認した。

第4章 歩容特徴を用いた人物動線の 対応付け

障害物、他の人物による遮蔽により、人物追跡は失敗する。これに伴い、人物の動線がいくつかの動線片に分断されてしまうことが考えられる。観測視野内における一貫した人物追跡を実現するために、分断化された動線片を対応付ける枠組みが必要となる。本手法では、人物固有の性質である歩容特徴を利用することにより、人物動線の対応付けを実現する。

4.1 はじめに

安心・安全な社会環境の実現に向け、カメラ映像を利用した人物追跡について広く研究開発が行われている。環境の影響に対して頑健な人物追跡を実現するために、様々な問題に対処する必要があるが、なかでも障害物や他の人物により引き起こされる追跡対象の遮蔽は、人物追跡において不可避の重要な課題である。人物が、部分的、もしくは完全に遮蔽されることにより、追跡システムは人物を安定に観測できないため、追跡の不安定化、失敗を引き起こす。

遮蔽により引き起こされる追跡失敗の問題に対処するために、近年様々な研究が行われている。その多くは、部分遮蔽に対して頑健な追跡手法の開発である。遮蔽により人物全体が安定に観測されづらいことから、人物を、頭部、胴体、両足などの複数のパーツの集合としてモデル化するアプローチが一般にとられている [57, 77, 103, 104, 107]。各パーツ毎に事前に学習した検出器を用意することで、遮蔽により人物の一部の観測が不安定化する場合においても追跡を継続することが可能となる。

障害物による完全遮蔽により追跡が失敗した場合、遮蔽から回復した際に速やかに追跡を再開することが重要となる。これに対し、追跡初期化のための検出器の併用 [42, 46, 70, 101]、他センサの利用 [53, 67]、人物の行動履歴に基づく人物存在確率

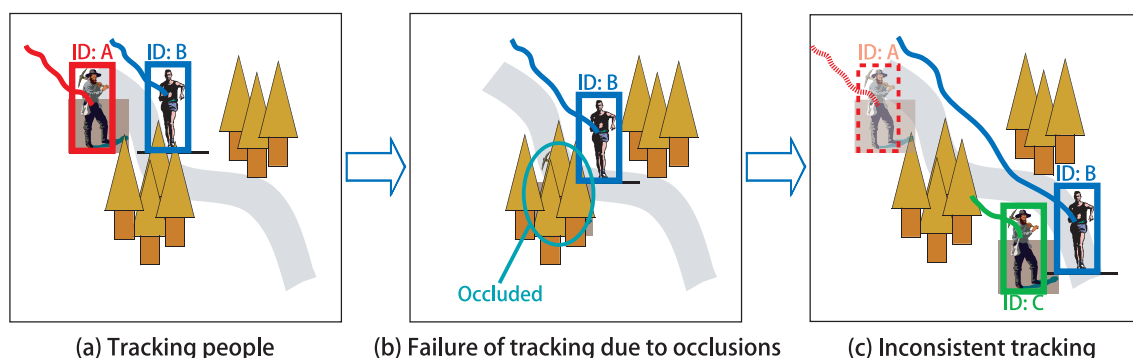


図 4.1: 遮蔽に伴う一貫した追跡の失敗．同じ追跡対象であるにも拘らず，遮蔽による追跡失敗前後において異なる人物として追跡する問題が考えられる（右図（c））．一貫した追跡を実現するために，遮蔽前後の動線の対応付けを行う必要がある．

の利用 [89] など，追跡初期化に関する手法が数多く提案されている．

このように，部分遮蔽に対する追跡の頑健化，追跡失敗からの高速な追跡初期化といった様々な方法により，遮蔽の問題への対処がなされている．しかしながら，一般に遮蔽による追跡失敗は不可避であることから，計測される人物の動線が観測視野内において一貫した人物の動きの軌跡であるとは限らない．言い換えると，観測視野内を移動する人物の動線は，遮蔽に伴いいくつかの動線片に分断されている可能性がある（図 4.1）．このような問題に対処するために，観測視野内において追跡器により推定される人物の動線に対応付ける枠組みについて考慮する必要がある．

これに対し本手法では，人物固有の性質である歩容特徴に基づいた人物動線の対応付けを実現する．具体的には，追跡対象である人物に属する特徴点の動きの軌跡群より得られる手掛かりを，動線の対応付けの指標として利用する．文献 [88] にあるように，特徴点軌跡群から人物の歩容に相当する周波数特徴を求める．そして特徴点軌跡群がそれぞれ持つ歩容特徴を多次元データのサンプル集合とみなし，特徴量集合間の類似度を測る Pyramid Match Kernel [29, 30] により動線間の類似度を計算する．このように計測される動線片間の類似度をもとに，二部グラフの最大重みマッチング手法の一つであるハンガリアン法を適用することで，最適な動線の対応付けを実現する．

人物の歩幅や歩行周期に基づく歩容特徴は，近年個人認証のための手掛かりとして注目されている性質であり，その有用性が報告されている [13, 35, 55, 63]．歩容特徴は人物固有の性質であるため，動線の対応付けにおいても効果的な指標になるこ

とが期待される。歩容特徴は、生態認証の分野で主に利用されているものであり、人物動線の対応付けに利用する研究は類を見ない。

Pyramid Match Kernel (PMK) とは、階層的ヒストグラムを用いたデータ集合間の類似度を測る手法である。ヒストグラム交差に基づく計算方法であることから、データ集合間においてデータのサンプル数が異なっている場合においても類似度を求めることができる。本手法では、個数が一定しない多数の特徴点の動きの軌跡群を用いるため、PMK による類似度計算手法は本手法に適していると言える。

このような考え方のもと、観測視野内全体における最適な人物動線の対応付けを実現する。

本章は以下のように構成される。次節 4.2 節では関連する先行研究について概観し、本手法の位置づけを確認する。そして 4.3 節では動線の対応付けを確率的な枠組みで取り扱う。この枠組みのもと、4.4 節において人物に属する特徴点軌跡群を用いた動線片間の類似度計算の方法について説明する。4.5 節では動線の対応付けの実験結果について報告すると共に、本手法の問題点について考察する。最後に 4.6 節においてまとめる。

4.2 関連研究

本節では、人物の動線の対応付けに関する先行研究について概観する。また同一視野内における動線の対応付けに関する研究だけでなく、視野を共有しない複数視野間での人物の対応付け、認識に関連する研究についても併せて概観する。そして紹介する既存研究をまとめることにより、本手法の位置づけについて確認する。

4.2.1 人物動線の対応付け

遮蔽に起因する追跡失敗に伴い、観測視野内を移動する人物動線は複数の動線片に分断される。本節で紹介する関連研究は、このような追跡失敗に起因して生成される動線片 (Tracklet) の対応付けを行うことにより、一人の人物の一貫した追跡を実現している。

局所的な動線片の対応付け

動線の対応付けによる人物追跡に関してよく知られている手法に、Kaucicらの手法が挙げられる [47]。彼らは、実世界の複雑な環境において同時に複数の人物追跡を行うことは困難であるとし、視野内の人物を一人ずつ追跡する。そして個々の追跡により得られる動線をそれぞれ対応付けることにより、複数の人物追跡を実現している。最初に、得られる映像から観測視野内のシーン認識を行うことで、遮蔽マップという遮蔽が発生しやすい領域の推定を行う。得られる遮蔽マップを事前知識として利用し、追跡対象の見え、移動軌跡の位置、速度に基づき軌跡の類似度を計算することにより、MAP推定の枠組みで最適な対応を求めている。

Nilliusらは、複数人追跡の結果に基づいた track graph と呼ばれる時空間グラフをベイジアンネットワークで表現し、動線片の対応付けを行っている [68]。Track graph とは、複数の追跡対象のインタラクション（複数の人物が合流する、分かれる、交差する、etc.）をリンクとした個々の人物の行動状態を表す時空間グラフを意味する。

この手法の主なアイデアは、複数人が近接したときに問題となる個々の追跡対象の区別について、追跡処理レベルにおいて対処しない点にある。複数人が近接する前後では個々の人物を正しく追跡できているとし、これらの人物動線の状態に基づき各追跡対象がどのように行動したのかをベイジアンネットワーク (Track graph) を用いて推定する。これにより、追跡処理では対処が難しい複数人が近接した場合においても個々の対象の正しい動きを推定することができる。

観測視野全体の最適な人物動線の対応付け

前述した [47] や [68] の手法では、局所的な（人物一人ずつの）動線の対応付けのみを行っている。それに対し文献 [56] では、人物の動きの軌跡の推定と人物検出を最適化問題の枠組みで同時に考慮することにより、観測シーン全体の最適な動線の対応付けを行っている。具体的には、動線の推定は検出結果に基づく見えの情報を手掛かりにした拡張カルマンフィルタにより推定する。得られる検出結果と推定された動線片から、EM アルゴリズムのような逐次最適化に基づき観測視野内全体の最適な動線の対応付けを行う。

これ以外にも、観測視野全体の動線の対応付けの最適化手法は、近年数多く提案されてきている。その多くは、MAP推定として動線の対応付けを捉え、ハンガリアン法や最小コストフローに基づき最適な対応付けを実現している [38, 59, 72, 104, 106]。

Huang らは、動線の対応付けを MAP 推定の枠組みで捉え、階層的な対応付処理を行うことにより視野内における一貫した人物追跡を実現している [38]。最初に、人物検出器による検出結果を対応付けることで、短時間であるものの信頼しうる動線片を生成する。得られる動線片を用い、中間レベルでは MAP 推定に基づいた対応付けを行う。具体的にはハンガリアン法を用いた Greedy な対応付けを行う。そして最後のステージでは、シーン認識（出入口、遮蔽領域の認識）を行うことで、中間レベルで推定された動線片の対応付けの改善を行う。これにより視野内で一貫した追跡を可能としている。

また Li らは、動線片の対応付けは、尤もらしさに関する順序付けと識別の 2 段階から成ると見なし、Rankboost [25]、Adaboost [80] の 2 種類の識別器を組み合わせた HybridBoosting により対応付けを行っている [59]。[38] と同様の MAP 推定の枠組みで動線片の対応付けの最適化を行うことにより、混雑環境下における頑健な人物追跡を実現している。

Xing らは、全体の対応付けの最適化の前段階である動線片の生成を高精度に行う手法を提案している [104]。具体的には、彼らはできる限り長時間の信頼できる動線を生成するために、時系列フィルタの一つであるパーティクルフィルタ [41] に、尤度関数を選択的に利用できる枠組みを組み入れている。尤度関数は人物のパーツについて事前に学習した識別器を利用し、安定に観測されるパーツについて尤度評価を選択的に行うことにより、混雑環境下で頻繁に発生する部分遮蔽に対処している。

4.2.2 視野を共有しない複数視野における人物の対応付け

人物の行動の軌跡の対応付けにおいて、その軌跡間の類似度の計測は、対応付け候補の人物の見えを手掛かりに行うことが一般的である。見えに基づく類似度の計測については、視野を共有しない複数観測領域における人物の対応付けや、人物認識に関する研究において広く議論されている。そこで本節では、複数の観測領域間での人物の対応付けに関する研究について概観する。

色情報の利用

人物の対応付けの問題について様々な研究がなされているが、色情報を手掛かりに対応付けを実現する手法が多く提案されている。

主な先行研究では色ヒストグラムを利用した手法が多く提案されている [12,17,36] . Chengらは主色ヒストグラムにより人物を表現することにより,照明環境の変化,姿勢の変動の問題に対処している [12] . 主色ヒストグラムとは,対応付けに必要な主要な色のみを用いた色ヒストグラムであり,これに基づき類似度を測ることにより,対象の対応付けを行う .

主色ヒストグラム表現の拡張もなされている . 人物は上半身,下半身で色ヒストグラムが明らかに異なることが多いことから,人物を上半身,下半身,全身の3種類の指標で表現し,それぞれの領域において主色ヒストグラムを計算している [9,36] . 主色ヒストグラムによる人物の対応付けをオンライン学習の枠組みにより実現する手法 [36] なども近年では提案されている .

主色ヒストグラムを利用したアプローチ以外にも,観測領域間の人物の時空間的な動きの制約を組み入れた手法 [44],色ヒストグラムに空間的な情報を組み入れた Spatiogram [6] を利用した手法 [17] や,色ヒストグラムとテクスチャ情報のアンサンブル学習による人物の認識・対応付け手法 [31] などがある . また,観測するカメラ間の色の違いの対応を表現する写像 (Color Transfer Function) を獲得し,それを対応付けに利用する研究 [45],カメラ間の明るさの対応を記述する写像 (Brightness Transfer Function) を利用した手法 [73] なども提案されている .

形状情報の利用

照明環境の変動に頑健なエッジ情報に基づいた形状情報を手掛かりに対応付けを行う研究がなされている .

Caiらは,対応づく人物はエッジ領域近傍は類似するという仮定に基づき,エッジ領域と幾何的な制約を考慮したコスト関数を最小化することにより,人物の対応付けを実現している [9] . Shanらは,自動車の識別の問題に対し,エッジ情報を利用した事例に基づいた対応付けを実現している [81] . また Maddenらは,観測画像より得られる人物のシルエットから人物の身長を推定することで,人物の識別のための手掛かりとしている [62] . これ以外にも,HoG (Histogram of Gradient) 特徴 [19] に基づいた形状と見えのモデリングを行うことで人物の識別に利用する手法 [100] が提案されている . これらの多くは,色情報や他の手掛かりと共に併用されることが多い .

局所特徴量の利用

SIFT [60] や SURF [4] などの局所特徴量の利用による物体認識，行動認識の分野の発展がよく知られているように，局所特徴量を用いた人物の対応付けを実現する研究も数多く報告されている．

Gheissari らは，Hessian Affine invariant interest operator [66] により局所特徴点を検出し，特徴点まわり半径 r の円内の領域における色ヒストグラム，エッジ特徴を利用している [26]．これらの特徴に基づくヒストグラムの距離を計算することにより，人物の対応付けを実現している．

物体認識の分野で広く使われる Bag of Words モデル [18] を適用した手法も提案されている．Teixeira は人物の観測画像から抽出した SIFT 特徴量から，Bag of Words モデルで人物を表現し，SVM (Support Vector Machine) により人物の識別を実現している [92]．彼らは，観測される人物は姿勢や服装が変化する可能性を考慮し，識別器を逐次更新するアプローチをとっている．

Hamdoun らは，人物の対応付けのために SURF 局所特徴量を用いて観測される人物を表現している [33, 34]．人物を記述する特徴点集合を KD-tree [5] に格納し，投票により識別処理が行われる．これにより高速な人物の対応付けを実現している．

4.2.3 歩容特徴の利用による個人認証

画像を用いた生体認証の研究分野において，歩容特徴を用いた人物認識に関する研究がなされている．歩容特徴とは，歩幅，歩行周期などにより特徴付けられる人物固有の性質 [97] であり，人物識別において効果的であることが報告されている．本節では，このような歩容特徴を用いた人物認識に関する研究について概観する．

歩容特徴は，生体認証の研究分野では一般に背景差分により抽出された人物のシルエットの変化に基づき計測される．そのため，対象人物の姿勢の変化に応じたシルエット変化がその認識性能に大きな影響を与えることが知られている．

このような問題に対し，Makihara らは様々な方向から撮影した人物の歩行の様子から学習した適応的な観測方向変換モデルを用いることで対処している [63]．具体的には，まず背景差分により抽出された人物のシルエット画像の時系列データから，正規化自己相関係数を計算することにより歩容周期を求め，画像列に対して離散フーリエ変換を施すことにより歩容特徴の周波数特性を求める．このようにして得られる歩容の周波数特性を， K 方向， M 人の人物のデータから予め取得することにより，

$K \times M$ の方向と人物の種類により構成される行列を得ることができる．これを特異値分解することにより，ある特定の方向の適切な変換が求められる．

Chengらは，歩容認証に問題となるシルエット画像のばらつきを吸収するために，Gaussian Process [76] を利用することにより適切な学習を実現している [13]．Gaussian Process は少ない学習サンプルから最適な推定を行う強力な手法であり，このようなシルエットのばらつきがあるような問題について適した枠組みである．

また Leeらは，Shape Variance-Based (SVB) Frieze Pattern と呼ばれる表現を利用することにシルエットのばらつきの問題へ対処している [55]．SVB Frieze Pattern は，歩容画像列においてキーポーズとなる画像を抽出し，そのフレーム間差分に基づくものであり，歩行動作中にシルエット画像が不安定に観測される問題を回避することができる．

服装の違いによるシルエットの差異の問題について主に取り組んだ研究として Hossainらの手法がある [35]．彼らは人物を，頭部，胴体，下半身，脚など全部で8つの種類のモデルで表現し，各パーツにおける対応付けを行うことで識別性能を改善している．

4.2.4 本手法の位置付け

本節では，本手法に最も関連する人物動線の対応付けに関する先行研究について概観した．これに加え，複数の観測領域間での人物の対応付け，そして生体認証の研究分野である歩容特徴を用いた人物の認識に関連する研究についても概観した．

動線の対応付けに関連する先行研究の多くは，MAP 推定などの確率的な枠組みで対応付けを取り扱い，ハンガリアン法や最小コストフローといった最適化手法を適用することにより最適な対応付けを見つけている．文献 [59] でも言われているように，観測視野内の最適な動線の対応付けの探索において，二つの動線片間の類似度の計算が最も重要な役割を果たす．

人物（動線）を対応付けるための類似度について，複数観測領域間における人物の対応付けに関する研究についての調査からわかるように，提案されている類似度の多くは見えや動きに基づいている．

その一方で，生体認証の分野において，人物の固有の性質である歩容特徴を用いた個人認証の研究が盛んに行われている．これは，同一視野内における動線の類似度の計算において効果的に働くことを示唆している．

歩容特徴の利用における本手法と歩容認証の研究との違いは，歩容認証の研究の

多くは歩容特徴は人物のシルエットの動きのパターンから獲得していることに対し、本手法では特徴点の動きの軌跡から歩容特徴を抽出している点である。人物のシルエットを用いる歩容認証では、安定にシルエットが検出されている必要があるが、本手法では人物に属する特徴点の動きの軌跡が獲得されればよい。そのため、部分遮蔽に対して頑健であるといえる。

また、本手法における人物に属する特徴点の動きの軌跡の性質を利用する考え方は、人物追跡に関する研究では個々の人物を区別するための指標として利用されることが多いものの [7, 58, 74, 88]、人物の動線の対応付けのための類似度計算には利用されていない。更に、データ集合間の類似度を測る Pyramid Match Kernel は、物体認識、行動認識の研究分野においては頻繁に用いられているが、人物追跡、特に動線の対応付けに関する研究についての利用は見られない。

このような研究背景のもと、本手法では歩容特徴を用いた人物動線の対応付けを実現する。

4.3 マルコフ連鎖に基づく人物動線の対応付け

追跡器により得られる人物の動線片の集合から、観測領域における全ての人物の動きを表現するような最適な動線の組み合わせを見つける。具体的には、動線片の組み合わせをマルコフ連鎖で表現し、その対応付けの確率を最大にする組み合わせを最適解とする。大域最適解の探索には、莫大な計算コストを要するため、本手法では二部グラフの最大重みマッチング手法の一つであるハンガリアン法に基づいた Greedy な方法 [38] により対応付けを実現する。

以下本節では、人物動線の対応付けを実現するための確率的な枠組みについて、具体的な定式化を行う。

最初に、観測領域内を移動する人物の動線片を取得する。人物の動線片の生成には、特徴点軌跡のクラスタリングに基づく人物追跡手法を利用する。追跡手法に関する詳細については、文献 [88] を参照されたい。本追跡器により得られる、各人物に属する特徴点群の動きの性質は、動線片の対応付けの確率計算において利用される。これらの詳しい計算方法については次節で説明する。

得られる動線片の集合を $\mathcal{T}^{in} = \{T_i^{in}\}$ とする。この動線片集合について最適な対応付けを行うことにより、観測視野内における各人物の一貫した動きを表現するような動線の集合 $\mathcal{T} = \{T_j\}$ が得られるものとする。ここで $T_j = \{T_{i_0}^{in}, T_{i_1}^{in}, \dots, T_{i_j}^{in}\}$

である．これは，対応付けられた j 番目の動線 T_j は， l_j 個の動線片により構成されていることを意味する．

以上の定義に基づき，対応付けられた最適な動線の集合 \mathcal{T}^* は，対応付けられた動線集合が \mathcal{T} となる確率 $P(\mathcal{T})$ を最大化することにより得られる，すなわち，

$$\mathcal{T}^* = \operatorname{argmax}_{\mathcal{T}} P(\mathcal{T}) . \quad (4.3.1)$$

追跡対象一人に対して同時に複数の動線は生成されないため，動線 $T_j \in \mathcal{T}$ は他の動線と同時に存在しない．これはすなわち，

$$T_j \cap T_k = \emptyset, \forall j \neq k, \quad (4.3.2)$$

が成り立つことを意味する．

対応付けられた動線 $T_j \in \mathcal{T}$ は互いに独立であると仮定すると (4.3.1) 式を次のように分解することができる．

$$\begin{aligned} \mathcal{T}^* &= \operatorname{argmax}_{\mathcal{T}} \prod_{T_j \in \mathcal{T}} P(T_j), \quad (4.3.3) \\ \text{s.t.}, & T_j \cap T_k = \emptyset, \forall j \neq k. \end{aligned}$$

(4.3.3) 式において，対応付けられた動線が T_j となる確率 $P(T_j)$ は，次式のようにマルコフ連鎖によりモデル化することができる．

$$\begin{aligned} P(T_j) &= P\left(\{T_{i_0}^{in}, T_{i_1}^{in}, \dots, T_{i_j}^{in}\}\right) \\ &= P_{init}(T_{i_0}^{in}) P_{link}(T_{i_1}^{in} | T_{i_0}^{in}) \cdots P_{link}(T_{i_j}^{in} | T_{i_{j-1}}^{in}) P_{term}(T_{i_j}^{in}), \quad (4.3.4) \end{aligned}$$

ここで $P_{init}(T_{i_0}^{in})$ と $P_{term}(T_{i_j}^{in})$ は，初期状態確率，終了状態確率をそれぞれ意味し， $P_{link}(T_{i_{n+1}}^{in} | T_{i_n}^{in})$ は，動線片 $T_{i_n}^{in}$ の終端が $T_{i_{n+1}}^{in}$ の先端へ連結される確率を意味する．

(4.3.4) 式のように定式化された目的関数について，ハンガリアン法を適用することにより最適な対応付けを見つける (4.3.3) 式に (4.3.4) 式を代入し，対数をとることで次式を得る．

$$\begin{aligned} \mathcal{T}^* &= \operatorname{argmax}_{\mathcal{T}} \sum_{T_j \in \mathcal{T}} \left[\ln P_{init}(T_{i_0}^{in}) + \ln P_{link}(T_{i_1}^{in} | T_{i_0}^{in}) + \cdots \right. \\ &\quad \left. + \ln P_{link}(T_{i_j}^{in} | T_{i_{j-1}}^{in}) + \ln P_{term}(T_{i_j}^{in}) \right]. \quad (4.3.5) \end{aligned}$$

(4.3.5) 式右辺より，次の推移確率行列 $C = \{c_{ij}\}_{2n \times 2n}$ が得られる．

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} & c_{1(n+1)} & -\infty & \cdots & -\infty \\ c_{21} & c_{22} & \cdots & c_{2n} & -\infty & c_{2(n+2)} & \cdots & -\infty \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} & -\infty & -\infty & \cdots & c_{n(2n)} \\ \hline c_{(n+1)1} & -\infty & \cdots & -\infty & 0 & 0 & \cdots & 0 \\ -\infty & c_{(n+2)2} & \cdots & -\infty & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -\infty & -\infty & \cdots & c_{(2n)n} & 0 & 0 & \cdots & 0 \end{bmatrix}_{2n \times 2n} . \quad (4.3.6)$$

ここで，推移確率行列の各要素 c_{ij} は以下のように定義される．自身の動線片への対応付けの防止を考慮すると，

$$c_{ij} = \begin{cases} -\infty, & \text{if } i = j \leq n \\ \ln P_{link}(T_j^{in} | T_i^{in}), & \text{if } i, j \leq n, i \neq j \\ \ln P_{init}(T_j^{in}), & \text{if } i = j + n \\ \ln P_{term}(T_i^{in}), & \text{if } i + n = j \\ 0, & \text{if } i > n, j > n \\ -\infty, & \text{otherwise} \end{cases} , \quad (4.3.7)$$

となる．

更に，推移確率行列を全要素について規格化する．規格化された行列 $C' = \{c'_{ij}\}$ は次式により計算される．

$$c'_{ij} = \begin{cases} \frac{c_{ij}}{\sum_{(k,l,c_{kl} \neq -\infty)} c_{kl}}, & \text{if } c_{ij} \neq -\infty \\ -\infty, & \text{otherwise} \end{cases} . \quad (4.3.8)$$

このように規格化された推移確率行列 C' についてハンガリアン法を適用する．一般に，ハンガリアン法はコストを最小化する問題であるため (4.3.8) 式をコスト行列 $\tilde{C} = \{\tilde{c}_{ij}\}$ に変換した後に適用する．コスト行列への変換は次式により行われる．

$$\tilde{c}_{ij} = \max(C') - c'_{ij} , \quad (4.3.9)$$

ここで $\max(C')$ は，規格化された推移確率行列 C' の要素のうち最大の値をとることを意味する．本手法では，各要素は確率の対数をとった値であることから， $\max(C') = 0$ である．

ハンガリアン法により対応付けられた動線片の対応関係が格納される行列 $\Gamma^* = \{\gamma\}_{2n \times 2n}$ を参照することにより，以下のようにして動線片を対応付ける．

1. $i, j \leq n, i \neq j$ のとき, T_i^{in} の終端を T_j^{in} の先端へ対応付ける
2. $i = j + n$ のとき, T_j^{in} は連結された動線における最初の動線片とする
3. $i + n = j$ のとき, T_i^{in} は連結された動線における最後の動線片とする

これにより最適な対応付けを実現する.

安定した動線片の対応付けの実現のためには, 二つの動線片間の連結確率 $P_{link}(T_{i_{n+1}}^{in} | T_{i_n}^{in})$ の計算が重要な役割を果たす. 本手法では, この二つの動線片間の連結確率の計算に人物の歩容特徴を利用する.

4.4 特徴点軌跡群を用いた人物動線片間の連結確率の計算

本節では, 動線片間の連結確率 $P_{link}(T_j^{in} | T_i^{in})$ の計算方法について詳しく述べる. 本手法では, 人物に属する特徴点軌跡群により得られる性質に基づき, Pyramid Match Kernel アルゴリズムによりその類似度を計算する.

最初に, Pyramid Match Kernel アルゴリズムについて概観した後, PMK を用いた具体的な類似度計算方法について述べる.

4.4.1 Pyramid Match Kernel

Pyramid Match Kernel (PMK)

画像間のマッチング, 一般物体認識, 行動認識などの研究分野において, データ集合間の類似度の計算は重要な課題である. これまで, データサンプル集合間の類似度計算について様々な手法が提案されているが, 用いる特徴点の個数が膨大になる場合計算コストの面で実用的ではなかった

このような問題に対し, Pyramid match kernel は, ヒストグラム交差に基づく方法によりデータ集合間の類似度の高速な計算を実現している. 具体的には, 特徴点の集合を複数の分解能を持つ (ビン幅が異なる) 階層的なヒストグラムへ写像し, ヒストグラム交差を計算する. そしてそれぞれのレベルにおけるヒストグラム交差の結果を足し合わせることで, 高速な類似度計算を実現することができる. 代表的なデータ集合の類似度計算は, データサンプルの次元を d , サンプル数を m としたとき, 計算コストのオーダーがおおよそ $O(dm^2)$ から $O(dm^3)$ 程度になる. その一方で, PMK では $O(dmL)$ (L はヒストグラムの階層数) の計算コストで実現できる [29].

d 次元の特徴量ベクトルを $\mathbf{x} \in X$ としたとき, L レベルの分解能を持つ階層的ヒストグラムへの写像を次式で定義する.

$$\Psi(\mathbf{x}) = [H_{-1}(\mathbf{x}), H_0(\mathbf{x}), \dots, H_L(\mathbf{x})]. \quad (4.4.1)$$

ここで $H_i(\mathbf{x})$ ($i > 0$) は, レベル i における d 次元のヒストグラムベクトルである. また $H_{-1}(\mathbf{x})$ は, それぞれの特徴量ベクトル唯一がいずれかのビンに射影されるような, 十分小さいビン幅を持つヒストグラムベクトルを意味する.

二つのデータ集合間の類似度 \tilde{K}_Δ は, 比較すべき二つのデータ集合 $\mathbf{y} \in Y, \mathbf{z} \in Z$ としたとき, 次式により計算される.

$$K_\Delta(\Psi(\mathbf{y}), \Psi(\mathbf{z})) = \sum_{i=0}^L w_i N_i. \quad (4.4.2)$$

ここで, w_i はレベル i における重みを表し, $w_i = 2^{-i}$ で与えられる. また N_i は, レベル i において新たに適合したデータの組の数を意味する. 具体的には次式により与えられる.

$$N_i = \mathcal{I}(H_i(\mathbf{y}), H_i(\mathbf{z})) - \mathcal{I}(H_{i-1}(\mathbf{y}), H_{i-1}(\mathbf{z})). \quad (4.4.3)$$

ここで $\mathcal{I}(H_i(\mathbf{y}), H_i(\mathbf{z}))$ はヒストグラム交差関数であり,

$$\mathcal{I}(H_i(\mathbf{y}), H_i(\mathbf{z})) = \sum_{j=1}^r \min(H_i^{(j)}(\mathbf{y}), H_i^{(j)}(\mathbf{z})), \quad (4.4.4)$$

により与えられる. r はヒストグラムのビン数, $H_i^{(j)}(\cdot)$ は j 番目のビンにおける頻度を意味する. 以上の一連の計算方法について, 例を図 4.2 に示す.

PMK は, 高速にデータ集合間の類似度を計算できることだけでなく, 類似度を測る二つの集合の大きさが異なっている場合においても利用することができる. これは, 属する特徴点軌跡の数が不定である本手法に適した類似度計測手法であると言える.

Pyramid Match Kernel の問題点

PMK では, ビン幅が異なる複数のヒストグラムを階層的に定義し, それぞれのレベルにおいてヒストグラムマッチングを行うことにより, データ集合間の類似度を計算する. 各レベルにおけるヒストグラムのビン幅は, データ集合の分布により

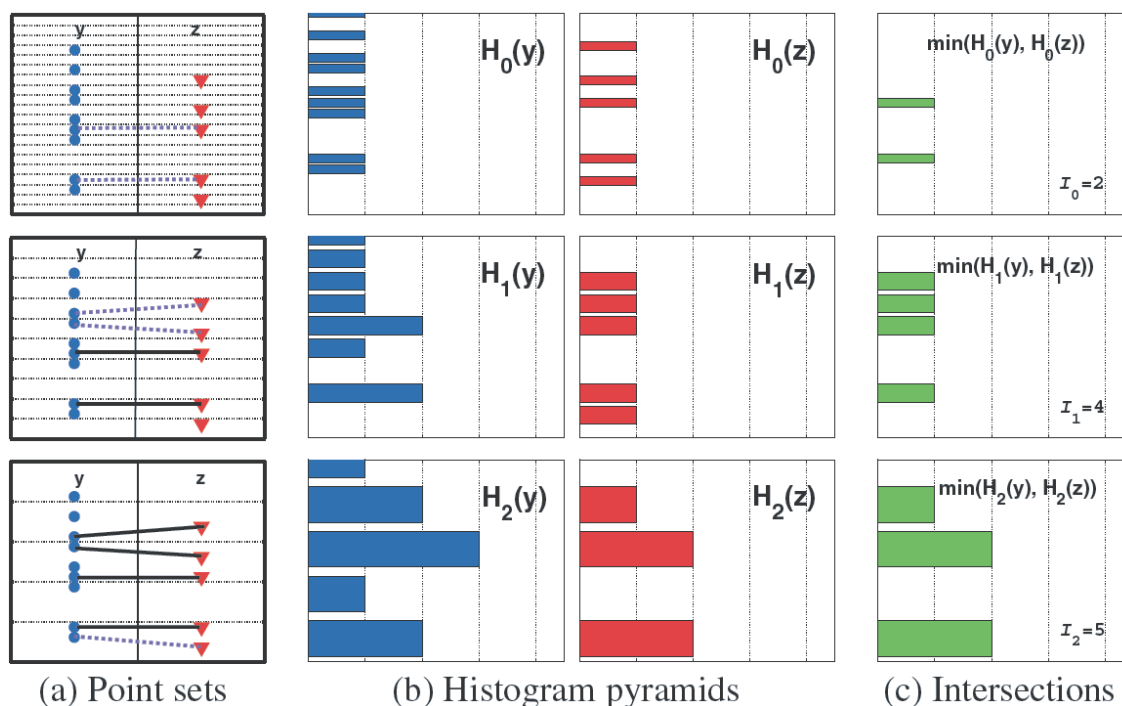


図 4.2: PMK による類似度計算 [29]. (a) 1 次元データ集合 y (左側), z (右側) の類似度計算例. 薄い破線はヒストグラムのビンの境界を意味し, 太い破線はそれぞれのレベルにおいて対応づいたペアを意味する. (b) は複数のビンの分解能を持つヒストグラムピラミッドであり, 水平方向の棒グラフはそれぞれヒストグラムの頻度を意味する. (c) は二つのヒストグラム $H_i(y)$, $H_i(z)$ のヒストグラム交差, $\mathcal{I}_i(H_i(y), H_i(z))$ の計算結果を表す. PMK は, それぞれのレベルにおいて新たに適合したデータのペア数を用いて類似度を計算する ((4.4.3) 式). 本計算例では, 各レベルで適合したデータのペア数はレベル 0 から $\mathcal{I}_i(H_i(y), H_i(z)) = 2, 4, 5$, 新たに適合したペア数は $N_i = 2, 2, 1$ となる. これより, 重みはそれぞれのレベルで $w_i = 1, \frac{1}{2}, \frac{1}{4}$ となる.

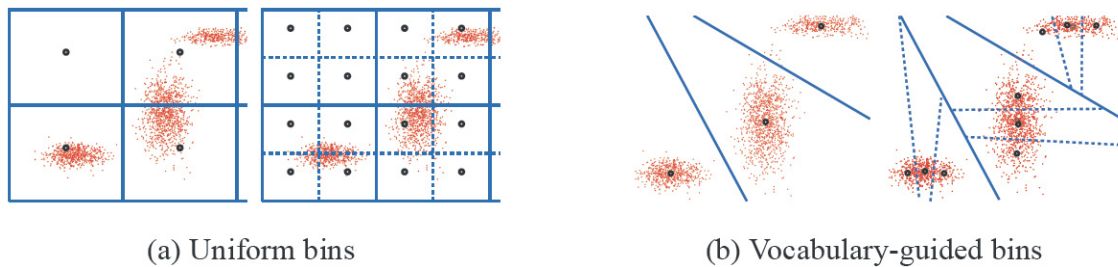


図 4.3: PMK [29] vs. VG-PMK [30]. 左図 (a) は PMK によるヒストグラムのビン幅の設定を表し, 右図 (b) は Vocabulary-guided PMK (VG-PMK) によるビン幅の設定を表す. VG-PMK の方がデータの分布に対して適応的なビン幅を設定できていることがわかる.

決定され, 1 レベルが上がることによりそのビン幅は 2 倍になるように設計されている.

しかしながら, ビン幅の変化率が一定であることに起因する問題が考えられる. 図 4.3-(a) に例を示す. 左図は PMK によるビン幅の分割方法を示しているが, データの分布に対して明らかに不適切なビンの設定をしていることがわかる. これにより, 類似度計算が正しく行われなことが懸念される.

Vocabulary-guided Pyramid Match Kernel (VG-PMK)

PMK が抱える問題に対処するために, Vocabulary-guided pyramid match kernel (VG-PMK) と呼ばれる, PMK を拡張した手法が提案されている [30]. VG-PMK では, データ集合に対し k -means クラスタリングを行うことにより, 分布に応じた適応的なビンの設定をすることができる. これにより, より効率的かつ高精度なデータ集合間の類似度計算が実現できる.

データの分布に応じた適応的なビンの設定を行うために, 階層的 k -means クラスタリングをデータ集合に対して行う. これを L 回再帰的に行うことにより, レベル $i < L$ において k^i 個のノードを持つ VG ピラミッドを構成することができる ((図 4.3-(b))).

階層的 k -means クラスタリングにより構成された VG ピラミッドについて, デー

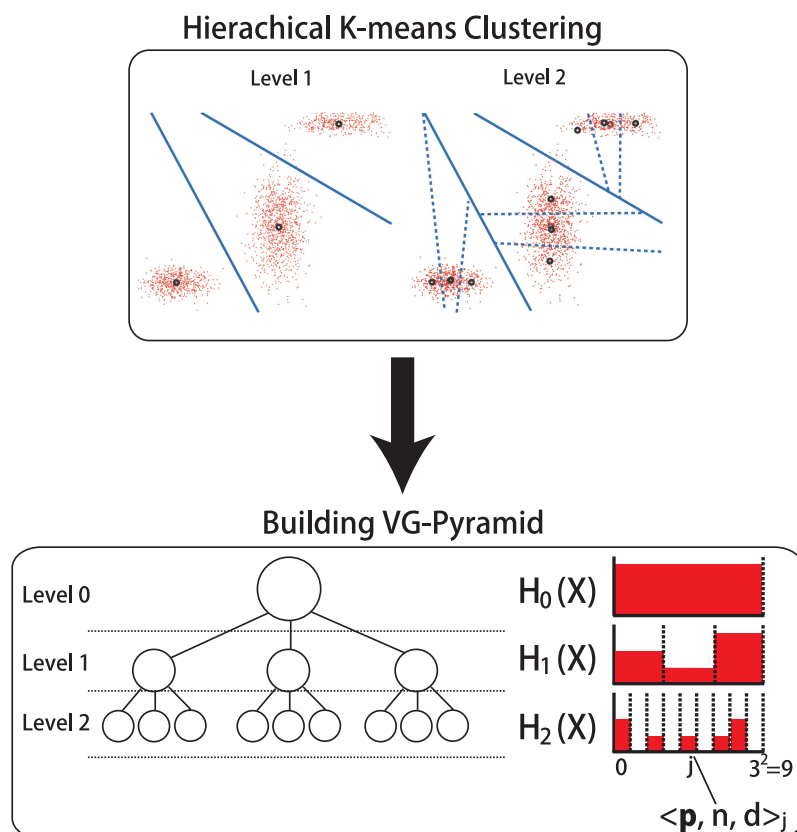


図 4.4: VG ヒストグラムピラミッドの構築．階層的 k -means クラスタリングを実行することにより，レベル i で k^i 個のビンを持つヒストグラムピラミッドを構築することができる．

タ集合 \mathbf{X} を以下のように VG ピラミッドに射影する．

$$\Phi(\mathbf{X}) = [H_0(\mathbf{X}), \dots, H_{L-1}(\mathbf{X})] . \quad (4.4.5)$$

ここで $H_i(\mathbf{X})$ は，レベル i におけるヒストグラムを表し， k^i 個のビンを持つ．それぞれのビンには，以下のような 3 種のデータを一組としたデータセットの頻度が格納される．

$$H_i(\mathbf{X}) = [\langle \mathbf{p}, n, d \rangle_1, \dots, \langle \mathbf{p}, n, d \rangle_{k^i}] . \quad (4.4.6)$$

ここで $\mathbf{p} = (p_1, \dots, p_i)$, $p_i \in [1, k]$ は，各レベルにおいて属したビンの ID を格納したベクトルを意味する．また n は頻度を表し， d はビン中心への最大距離を意味する．図 4.4 に概要を示す．

構築される VG ピラミッドに基づき，データ集合間の類似度計算を行う．PMK による類似度計算と同様に，それぞれの階層におけるヒストグラム交差に基づく処理により実現される． $n_{ij}(\mathbf{X})$ を，レベル i のヒストグラム $H_i(\mathbf{X})$ における j 番目のビン $\langle \mathbf{p}, n, d \rangle_j$ の頻度 n をとる作用素とし， $c_h(n_{ij}(\mathbf{X}))$ を，レベル i, j 番目のビンが持つ h ($1 \leq h \leq k$) 番目の子ノードが持つ頻度をとる作用素とする．これはすなわち，直近のレベル $i-1$ におけるヒストグラム $H_{i-1}(\mathbf{X})$ の h 番目のビンの頻度をとることを意味する．同様に， $d_{ij}(\mathbf{X})$ を，レベル i のヒストグラム $H_i(\mathbf{X})$ における j 番目のビンのエントリ $\langle \mathbf{p}, n, d \rangle_j$ のビン中心までの距離 d をとる作用素とする．対応付けるデータ集合をそれぞれ \mathbf{X}, \mathbf{Y} としたとき，データ集合間類似度 $\mathcal{C}(\Phi(\mathbf{X}), \Phi(\mathbf{Y}))$ は次式で与えられる．

$$\begin{aligned} & \mathcal{C}(\Phi(\mathbf{X}), \Phi(\mathbf{Y})) \\ &= \sum_{i=0}^{L-1} \sum_{j=1}^{k^i} w_{ij} \left[\min(n_{ij}(\mathbf{X}), n_{ij}(\mathbf{Y})) - \sum_{h=1}^k \min(c_h(n_{ij}(\mathbf{X})), c_h(n_{ij}(\mathbf{Y}))) \right]. \end{aligned} \quad (4.4.7)$$

ここで $w_{ij} = d_{ij}(\mathbf{X}) + d_{ij}(\mathbf{Y})$ は重みを意味する．これにより，サンプルデータの分布に即したデータ集合間のマッチングを行うことが可能となる．

VG-PMK は，データサンプルの次元が高次元になるときに効果的に働くことが期待される．本手法において取り扱うデータサンプルは，およそ百次元程度になる．このような高次元データ集合の高精度な類似度計算を行うために，本手法では PMK の拡張である VG-PMK を利用する．次節において，VG-PMK を用いた動線片間の類似度計算について詳しく述べる．

4.4.2 動線片間の連結確率の計算

人物の動線片間の連結確率 $P_{link}(T_i^{in}|T_j^{in})$ は，歩容特徴に基づく類似度 $P_g(T_j^{in}|T_i^{in})$ ，局所的見えに基づく類似度 $P_a(T_j^{in}|T_i^{in})$ ，動きに基づく類似度 $P_m(T_j^{in}|T_i^{in})$ ，時間関係に基づく類似度 $P_{tg}(T_j^{in}|T_i^{in})$ ，そして動線片間の時間の依存性 $P_{td}(T_j^{in}|T_i^{in})$ により，次式により計算される．

$$\begin{aligned} P_{link}(T_j^{in}|T_i^{in}) &= P_g(T_j^{in}|T_i^{in}) \cdot P_a(T_j^{in}|T_i^{in}) \cdot P_m(T_j^{in}|T_i^{in}) \cdot \\ & P_{tg}(T_j^{in}|T_i^{in}) \cdot P_{td}(T_j^{in}|T_i^{in}). \end{aligned} \quad (4.4.8)$$

これらの類似度は，動線片 T_i^{in} を持つ人物に属する特徴点の動きの軌跡群を用いて計算される．以下本節では (4.4.8) 式におけるそれぞれの類似度の計算方法について詳しく述べる．

歩容特徴に基づく類似度

歩行の周期成分の抽出

文献 [88] にあるように，人物の歩容特徴は，人間の身体的な構造から，歩行に伴う周期的な動きに表れると仮定する．文献 [88] では，垂直方向の周期的な動きにのみ注目しており，画像平面上における特徴点の動きの垂直成分について直線当てはめを行うことにより，周期成分を抽出している．しかしながら，人物の歩く方向（特にカメラの光軸方向）に伴い，垂直方向ではなく，水平方向に主だった周期性が観測されることがある．このような問題に対し，本手法では，垂直，水平の両方向について歩容に相当する歩行の周期成分を求める．

動線片 T_i^{in} を持つ人物に属する k 番目の特徴点の動きの軌跡 $\text{tr}_i^{(k)}$ を，位置，時間の三次元データ (x, y, t) の集合と見なし，主成分分析を適用する．第一主成分の主成分ベクトル $\text{tr}_i^{(k), PCA}$ は，変位の大きい人物の進行方向成分に相当すると考えられるため，特徴点軌跡との差分を計算することにより，歩行に伴う特徴点の動きの周期成分 $\text{tr}_i^{(k), p}$ を抽出することができる，すなわち，

$$\text{tr}_i^{(k), p} = \text{tr}_i^{(k)} - \text{tr}_i^{(k), PCA}, \quad (4.4.9)$$

となる．図 4.5，4.6 に，本処理により実際に得られた結果の一例を示す．

歩容特徴の周波数特性の類似度計算

歩容特徴に相当する特徴点の動きの周期成分 $\text{tr}_i^{(k), p}$ の各成分に対し，高速フーリエ変換（FFT）を行うことにより周波数特性を獲得する．また FFT に用いる窓サイズ w は，人間の歩行周期はおおよそ 1 秒程度である [63] ことに基づき，映像のフレームレートに応じて決定される．FFT により得られる水平，垂直方向それぞれの振幅特性 $\mathbf{a}_{i,x}^{(k)} = \{a_{i,x}^{(k),0}, \dots, a_{i,x}^{(k),w}\}$ ， $\mathbf{a}_{i,y}^{(k)} = \{a_{i,y}^{(k),0}, \dots, a_{i,y}^{(k),w}\}$ は，人物に属する特徴点軌跡群それぞれについて計算される．

動線片 T_i^{in} を持つ人物に属する特徴点軌跡群から得られる，人物の歩容特徴に相当する振幅特性の集合 $\mathbf{a}_{i,x} = \{\mathbf{a}_{i,x}^{(k)}\}$ ， $\mathbf{a}_{i,y} = \{\mathbf{a}_{i,y}^{(k)}\}$ を用い，歩容特徴に基づく類似度 $P_g(T_j^{in} | T_i^{in})$ は計算される．対応付けられる候補の動線片 T_j^{in} へ連結する類似度は，VG-PMK により次式のように計算される．

$$P_g(T_j^{in} | T_i^{in}) = \mathcal{C}(\Phi_{amp,x}(\mathbf{a}_{j,x}), \Phi_{amp,x}(\mathbf{a}_{i,x})) \cdot \mathcal{C}(\Phi_{amp,y}(\mathbf{a}_{j,y}), \Phi_{amp,y}(\mathbf{a}_{i,y})). \quad (4.4.10)$$

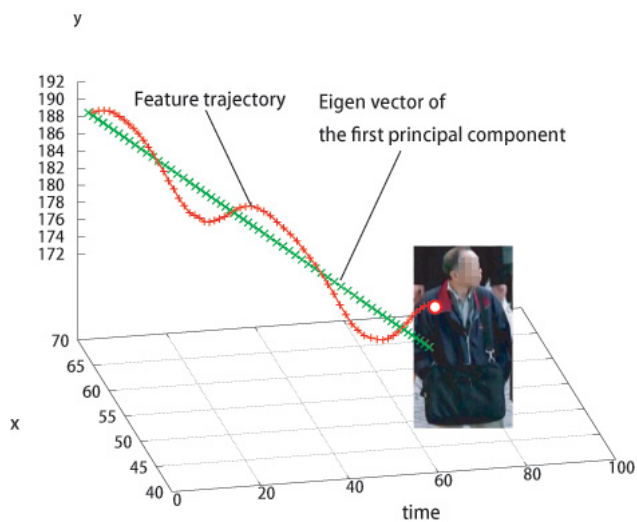
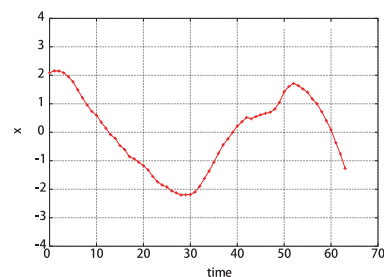
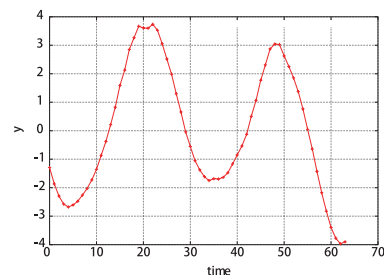


図 4.5: 主成分分析による周期運動の抽出 . 赤色の軌跡は特徴点の動きの軌跡 $\text{tr}_i^{(k)}$ を表し , 緑色の点列は第一主成分の主成分ベクトル $\text{tr}_i^{(k), PCA}$ である .



(a) Extracted periodic component along with horizontal axis



(b) Extracted periodic component along with vertical axis

図 4.6: 得られた周期成分 $\text{tr}_i^{(k), P}$ の例 . (a) 水平方向の周期成分 , (b) 垂直方向の周期成分 .

ここで $\mathcal{C}(\mathbf{P}, \mathbf{Q})$ は正規化された類似度であり ,

$$\mathcal{C}(\mathbf{P}, \mathbf{Q}) = \frac{1}{\sqrt{N_c}} \tilde{\mathcal{C}}(\mathbf{P}, \mathbf{Q}) , \quad (4.4.11)$$

である . また正規化定数 N_c は ,

$$N_c = \tilde{\mathcal{C}}(\mathbf{P}, \mathbf{P}) \tilde{\mathcal{C}}(\mathbf{Q}, \mathbf{Q}) , \quad (4.4.12)$$

で与えられる .

局所的見えに基づく類似度

遮蔽に伴い , 人物の見えは不安定に観測される . そのため , 追跡対象全体の見えを手掛かりに類似度を測ることは , 動線の対応付けの性能低下を引き起こす可能性がある . そこで本手法では , 局所領域における見えを手掛かりに動線片間の類似度を計算する . 具体的には , 動線片 T_i^{in} を持つ人物に属する特徴点軌跡周りの微小領域における二次元 HS (Hue-Saturation) 色ヒストグラムを利用する .

動線片 T_i^{in} が持つ時刻 $t = t'$ における k 番目の特徴点軌跡の位置を $\text{tr}_i^{(k)}|_{t=t'}$ とする．この位置を中心とした矩形局所パッチ内において，HS 色ヒストグラム $\mathbf{h}_{i,t'}^{(k)}$ を求める．人物に属する特徴点軌跡それぞれについて局所パッチ内色ヒストグラムは計算されるため，時刻 t' における動線片 T_i^{in} を持つ人物の見えは，局所パッチ内色ヒストグラム集合 $\mathbf{h}_{i,t'} = \{\mathbf{h}_{i,t'}^{(k)}\}$ により表現することができる．

これらの色ヒストグラムのデータ集合を用い，VG-PMK により類似度 $P_a(T_j^{in}|T_i^{in})$ を計算する．具体的には次式により与えられる．

$$P_a(T_j^{in}|T_i^{in}) = \mathcal{C} \left(\Phi_{app} \left(\mathbf{h}_{j,t_j^{head}} \right), \Phi_{app} \left(\mathbf{h}_{i,t_i^{tail}} \right) \right) . \quad (4.4.13)$$

ここで t_j^{head} , t_i^{tail} は，動線片 T_j^{in} の始端位置における時刻，動線片 T_i^{in} の終端位置における時刻をそれぞれ意味する．

これにより，遮蔽に起因した見えの観測の不安定性に頑健な類似度計測を行うことができる．

動きに基づく類似度

二つの動線片 T_i^{in} , T_j^{in} それぞれの位置，速度を手掛かりに動線片間の連結類似度を測る．

動線片 T_i^{in} の終端位置を \mathbf{x}_i^{tail} , 終端位置における速度を \mathbf{v}_i^{tail} とし，対応付ける候補であるもう一つの動線片 T_j^{in} の始端位置を \mathbf{x}_j^{head} , 速度を \mathbf{v}_j^{head} とする．また T_i^{in} の終端位置における時刻を t_i^{tail} , T_j^{in} の始端位置における時刻を t_j^{head} とし，その時間差を $\Delta t = |t_j^{head} - t_i^{tail}|$ とする．このとき，人物は等速直線運動すると仮定することにより，それぞれの時刻 t_i^{tail} , t_j^{head} において，動線片 T_i^{in} , T_j^{in} が互いにどの位置にいる（いた）のかを推定することができる．推定されるそれぞれの動線片の位置 $\tilde{\mathbf{x}}_i$, $\tilde{\mathbf{x}}_j$ は次式により与えられる．

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i^{tail} + \mathbf{v}_i^{tail} \Delta t , \quad (4.4.14)$$

$$\tilde{\mathbf{x}}_j = \mathbf{x}_j^{head} - \mathbf{v}_j^{head} \Delta t . \quad (4.4.15)$$

図 4.7 に概念図を示す．

推定される動線片の位置と実際の位置のペア $\tilde{\mathbf{x}}_i$ と \mathbf{x}_j , $\tilde{\mathbf{x}}_j$ と \mathbf{x}_i それぞれが互いに近接した関係にあるならば，動線片間の類似度は高いと見なすことができる．特徴

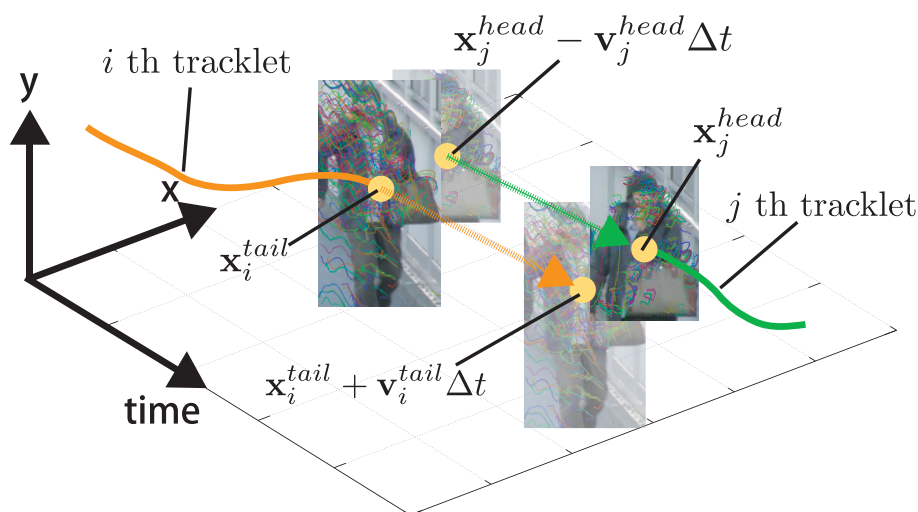


図 4.7: 動きに基づく類似度計算. i 番目の動線片の終端における位置 x_i^{tail} と速度 v_i^{tail} から, Δt 後にどの位置にいるか推定する. また j 番目の動線片の始端における位置 x_j^{head} と速度 v_j^{head} から, Δt 前にどの位置にいたか推定する. 推定される位置に基づき, 特徴点軌跡群の位置の類似度を測る.

点軌跡群の位置を用い, VG-PMK に基づき次式により類似度は計算される.

$$P_m(T_j^{in}|T_i^{in}) = \mathcal{C}\left(\Phi_{m_i}\left(\tilde{\mathbf{tr}}_j|_{t=t_i^{tail}}\right), \Phi_{m_i}\left(\mathbf{tr}_i|_{t=t_i^{tail}}\right)\right) \cdot \mathcal{C}\left(\Phi_{m_j}\left(\mathbf{tr}_j|_{t=t_j^{head}}\right), \Phi_{m_j}\left(\tilde{\mathbf{tr}}_i|_{t=t_j^{head}}\right)\right). \quad (4.4.16)$$

ここで $\mathbf{tr}_i|_{t=t_i^{tail}}$, $\mathbf{tr}_j|_{t=t_j^{head}}$ は, 時刻 t_i^{tail} , t_j^{head} における動線片 T_i^{in} , T_j^{in} がそれぞれ持つ特徴点軌跡群の位置の集合を表す. また $\tilde{\mathbf{tr}}_j|_{t=t_i^{tail}} = \left\{(\mathbf{tr}_j^{(0)}|_{t=t_j^{head}} - \mathbf{v}_j^{head} \Delta t), \dots, (\mathbf{tr}_j^{(n_j)}|_{t=t_j^{head}} - \mathbf{v}_j^{head} \Delta t)\right\}$ は, 二つの動線片の関係と等速直線運動により推定される, 時刻 t_i^{tail} における j 番目の動線片が持つ特徴点軌跡群の推定位置の集合を表す. 同様に, 時刻 $t = t_j^{head}$ における i 番目の動線片が持つ推定位置の集合が $\tilde{\mathbf{tr}}_i|_{t=t_j^{head}} = \left\{(\mathbf{tr}_i^{(0)}|_{t=t_i^{tail}} + \mathbf{v}_i^{tail} \Delta t), \dots, (\mathbf{tr}_i^{(n_i)}|_{t=t_i^{tail}} + \mathbf{v}_i^{tail} \Delta t)\right\}$ である.

時間関係に基づく類似度

対応付けるべき動線片間の時間差が大きくなればなるほど連結しづらいつらいつらと考えられる. そこで, 時間関係に基づく類似度 $P_t(T_j^{in}|T_i^{in})$ を以下のように定義する.

$$P_{tg}(T_j^{in}|T_i^{in}) = \exp(-\alpha \Delta t), \quad (4.4.17)$$

ここで α はパラメータであり, Δt は軌跡片 T_i^{in} の終端における時刻と T_j^{in} の先頭における時刻の差分を意味する.

動線片の時間関係に基づく対応付けの制約

動線片の対応付けは, マルコフ連鎖により表現されるため ((4.3.4) 式), 対応付ける二つの動線片の時間関係を考慮する必要がある. 対応付ける動線片 T_i^{in} の終端における時刻を t_i^{tail} , 対応付けられる動線片候補 T_j^{in} の始端における時刻を t_j^{head} としたとき, 動線片間の時間関係に基づく制約を以下のように定義する.

$$P_{td}(T_j^{in}|T_i^{in}) = \begin{cases} 1, & \text{if } t_j^{head} > t_i^{tail} \\ 0, & \text{otherwise} \end{cases} \quad (4.4.18)$$

これにより, 時間の依存関係に反する対応付けを防ぐことができる.

4.5 実験・考察

本手法の有用性を確認するために, 実際の映像を用いて動線の対応付けの実験を行った. 実験に用いた環境は, OS: Windows XP, CPU: Intel Core2 Quad 2.66GHz, RAM: 3.00GB である. 本実験では, 動線の対応付け処理の性能について注目するために, 事前に映像から人物領域を切り出し, 人工的に動線片を生成した. 図 4.8 にその概要を示す.

実験に用いた映像は全部で 2 つであり, 解像度は 640×360 , フレームレートは 60fps のものを用いた. また全ての実験で同じパラメータを用いた. 具体的には, 初期状態確率, 終了状態確率の対数値を $\ln P_{init}(T_i^{in}) = -10$, $\ln P_{term}(T_i^{in}) = -10$, 時間差に基づく類似度計算のパラメータを $\alpha = 0.002$ とした. また各 VG-PMK のパラメータ (クラスタ数 k , ピラミッドのレベル数 L) を次のようにそれぞれ設定した. 歩容特徴に基づくものを $k = 3$, $L = 5$, 局所的見えに基づくものを $k = 4$, $L = 3$, そして動き特徴に基づくものを $k = 2$, $L = 2$ とした.

歩容特徴の利用の性能への寄与を確認するために, 歩容特徴に基づく類似度計算を除いたものを基準システムとして用いた. これはすなわち, 動線片間の連結確率を, $P_{link}(T_j^{in}|T_i^{in}) = P_a(T_j^{in}|T_i^{in}) \cdot P_m(T_j^{in}|T_i^{in}) \cdot P_{tg}(T_j^{in}|T_i^{in}) \cdot P_{td}(T_j^{in}|T_i^{in})$ というように計算することを意味する.

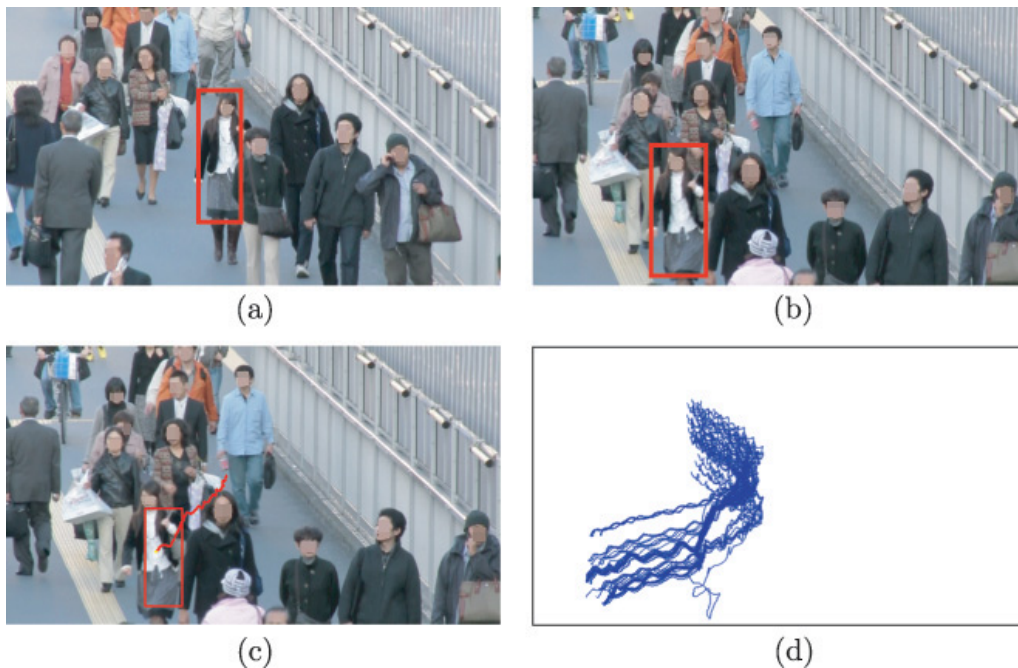


図 4.8: 動線片の生成．映像から人物領域を切り出すことにより，人工的に動線片を生成する．動線片の始端，終端それぞれにおいて人物領域を切り出し（(a)，(b)），両方の領域内に存在する特徴点の追跡結果に基づき，動線片を生成する（(c)）．図 (d) は追跡対象に属する特徴点の動きの軌跡群の抽出例を表す．

4.5.1 実験結果

最初に，比較的人物の動きの観測が良好なシーケンスについて実験を行った．図 4.9 に用いた映像の一部を示す．本実験において生成した動線片の数は 9 つであり，実際は 5 人の人物の動きの軌跡である．

表 4.1 に対応付けの結果を真値と併せて示す．また図 4.11 に歩容特徴を用いた動線の対応付け結果を，図 4.12 に歩容特徴を用いなかった場合の対応付け結果を示す．各人物に属する線は動線片を表しており，赤矢印が対応づいた方向を示している．これらの結果を比較するとわかるように，歩容特徴を用いた場合（図 4.11）では正しく対応付けが行われていることに対し，歩容特徴を用いなかった場合（図 4.12）では，2 人の人物が近接しているときにおいて，図中に示すように誤った対応付けを行ってしまっていることが見て取れる．図 4.10 にそのときの様子を示す．この様子からわかるように，動きと見えの指標では対処が難しい状況であると言える．



図 4.9: シーケンス (1) の一部 . 本映像のうち, 赤丸で囲った 5 人の人物から 9 つの動線片を生成した .



図 4.10: 歩容特徴を用いない場合の対応付け失敗例 . 赤丸で囲った 2 名の人物は互いに近接しており, 対処が難しい .

表 4.1: シーケンス (1) における動線片対応付け結果 . 表中の数字は tracklet の番号を表しており, 矢印の順番で対応づいたことを意味する .

Using gait features	Without gait features	Ground truth
Trajectory 0 : 0 → 1	Trajectory 0 : 0 → 3	Trajectory 0 : 0 → 1
Trajectory 1 : 7 → 8	Trajectory 1 : 4 → 5	Trajectory 1 : 7 → 8
Trajectory 2 : 2 → 3	Trajectory 2 : 2 → 1	Trajectory 2 : 2 → 3
Trajectory 3 : 4 → 5	Trajectory 3 : 7 → 8	Trajectory 3 : 4 → 5
Trajectory 4 : 6	Trajectory 4 : 6	Trajectory 4 : 6

歩容特徴の性能向上への寄与を定量的に確認するために, 本実験において得られた歩容特徴に基づく動線片間の類似度 $P_g(T_j^{in} | T_i^{in})$ を表 4.2 に示す . この結果からわかるように, 対応付けるべき動線片間 (図 4.11 参照) での類似度は, 他の動線片との類似度に比べおよそ $10^5 \sim 10^7$ のオーダーで大きな値を示していることが見て取れる . この結果から, 歩容特徴は動線片の対応付けにおいて効果的な指標であることを確認できた .

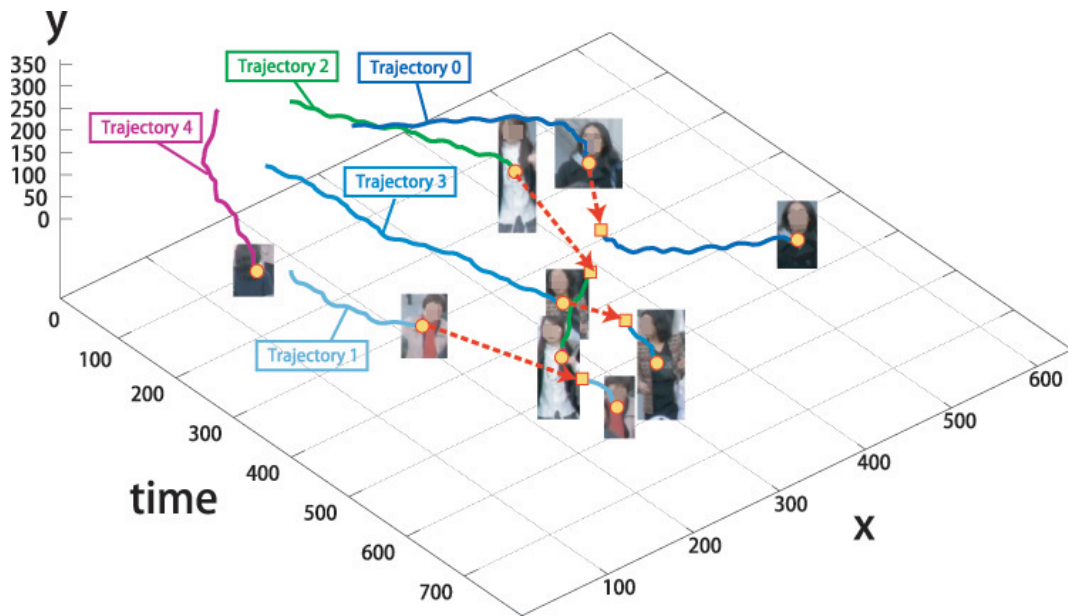


図 4.11: シーケンス (1) における動線の対応付け結果 (歩容特徴の利用)。人物に属する軌跡は動線片を表しており、赤矢印が対応付いた結果を示す。下段の Ground Truth の結果と併せて確認すると、正しく対応づいていることが見て取れる。

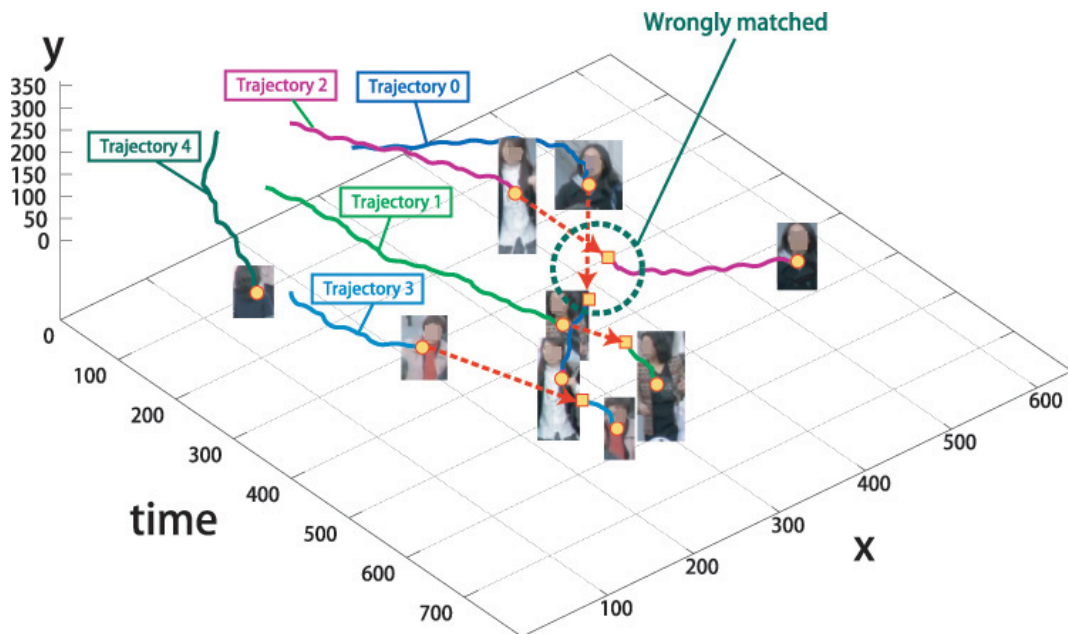


図 4.12: シーケンス (1) における動線の対応付け結果 (歩容特徴無し)。図中に示すように、時刻 530 - 630 フレームあたりにおいて、誤った対応付けを行っていることがわかる。

表 4.2: 歩容特徴に基づく動線片間の類似度．縦横方向に記される0から8までの数字は動線片番号を意味する．また，本手法では自分自身への対応は許可していないため，自分自身の間の類似度は計算されない．

	0	1	2	3	4	5	6	7	8
0		0.35	8.23E-07	2.26E-07	2.66E-07	2.41E-06	7.60E-07	3.71E-06	7.06E-07
1	0.36		1.42E-08	8.56E-08	5.27E-09	6.49E-08	2.58E-07	1.02E-07	1.93E-08
2	4.78E-07	1.88E-08		0.3	3.04E-07	1.07E-06	5.59E-08	2.23E-06	1.77E-06
3	1.81E-07	8.74E-08	0.29		4.76E-08	1.40E-06	5.89E-08	1.24E-06	4.92E-06
4	1.48E-06	5.49E-09	1.17E-06	8.24E-08		0.13	3.32E-08	2.02E-05	1.35E-06
5	7.45E-07	7.84E-08	7.20E-06	9.52E-07	0.11		8.55E-08	1.79E-05	7.30E-07
6	1.30E-06	2.26E-07	3.67E-08	5.20E-07	3.06E-08	1.17E-07		8.92E-08	4.69E-08
7	1.09E-06	5.08E-08	3.20E-06	1.10E-06	3.56E-06	1.34E-05	7.74E-08		0.07
8	9.26E-07	2.01E-08	8.09E-07	2.72E-06	3.60E-07	7.12E-07	5.09E-08	0.02	



(a) 周りの群集と同調した動き

(b) 進路方向の変更

図 4.13: シーケンス (2) の一部．シーケンス (1) に比べ非常に混雑しており，部分遮蔽，完全遮蔽が頻繁に起きるシーンである．赤丸で示す人物は，はじめ周りの群集と同様の動きをしている（左図(a)）が，しばらくすると進路を変更して異なる方向へ移動した（右図(b)）．これは，動線の対応付けにおいて難しいケースであると考えられる．

次に，動線片の対応付けにおいて困難な状況のうちの一つである，観測視野内に多数の人物が存在する混雑したシーケンスについて実験を行った．図 4.13 に用いた映像の一部を示す．これらの図からわかるように，本シーケンスでは部分遮蔽，完全遮蔽が頻繁に起きていることがわかる．また図に赤丸で示した人物は，途中で進行方向を変えるなどの複雑な行動をしており，動線対応付けにおいて難しいケースであると言える．本実験において生成した動線片の数は 23 であり，実際は 15 人の人物の動きの軌跡である．

表 4.3 に対応付け結果を真値と併せて示す．また図 4.14 に歩容特徴を用いた動線の対応付け結果を，図 4.15 に歩容特徴を用いなかった場合の対応付け結果を示す．歩容特徴を利用しない場合の結果（図 4.15）では，多数の人物が存在する領域では正しい対応付けが行えていない．それに比して歩容特徴を利用した結果（図 4.14）では，混雑状況であるにも拘らず，ほぼ全ての人物の対応付けが行うことができていることが見て取れる．

以上の結果から，歩容特徴を用いることにより，混雑環境下においても頑健な動線の対応付けを実現することができると言える．

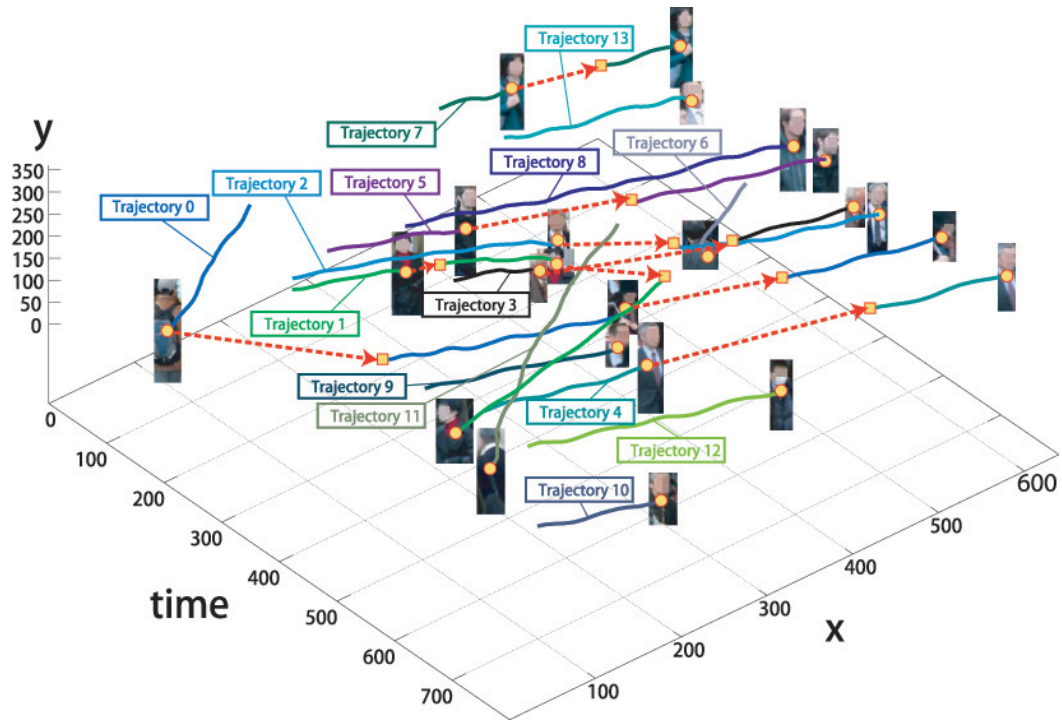


図 4.14: シーケンス (2) における動線の対応付け結果 (歩容特徴の利用) .

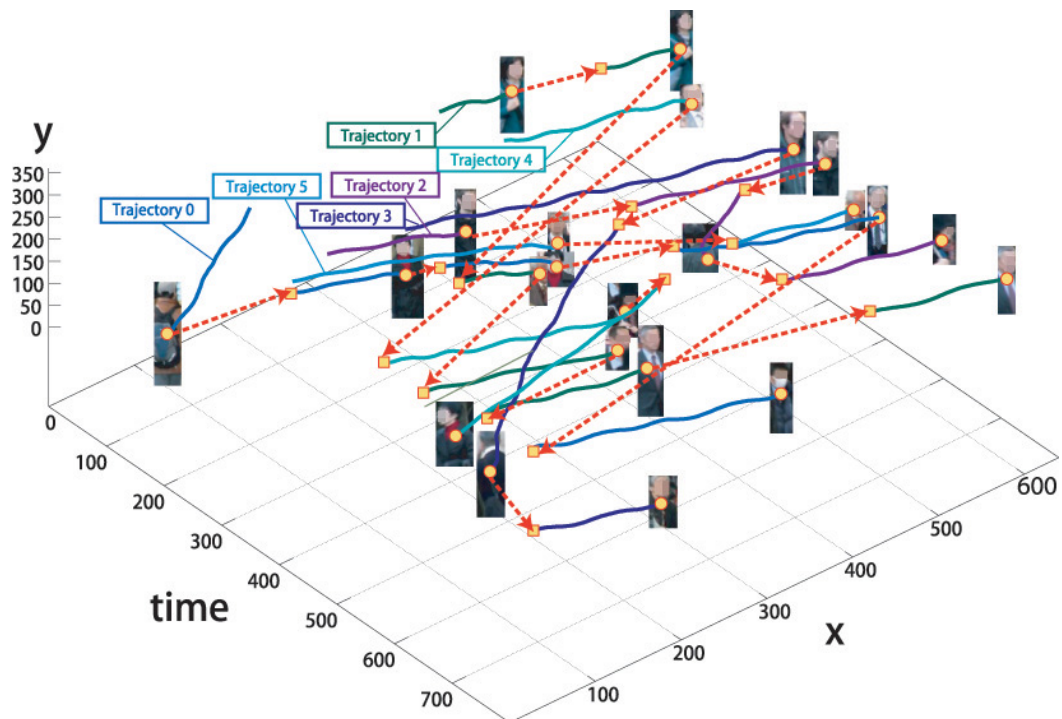


図 4.15: シーケンス (2) における動線の対応付け結果 (歩容特徴無し) . 明らかに誤った対応付けを行ってしまっていることが見て取れる .

表 4.3: シーケンス (2) における動線片対応付け結果 .

Using gait features	Without gait features	Ground truth
Trajectory 0 : 0 → 14 → 15	Trajectory 0 : 0 → 1 → 2 → 13 → 21	Trajectory 0 : 0
Trajectory 1 : 1 → 2 → 3	Trajectory 1 : 7 → 8 → 17 → 16 → 4 → 5	Trajectory 1 : 14 → 15
Trajectory 2 : 12 → 13	Trajectory 2 : 9 → 10 → 6 → 15	Trajectory 2 : 1 → 2 → 3
Trajectory 3 : 17 → 18	Trajectory 3 : 11 → 20 → 19	Trajectory 3 : 12 → 13
Trajectory 4 : 4 → 5	Trajectory 4 : 22 → 14 → 3	Trajectory 4 : 17 → 18
Trajectory 5 : 9 → 10	Trajectory 5 : 12 → 18	Trajectory 5 : 4 → 5
Trajectory 6 : 6		Trajectory 6 : 9 → 10
Trajectory 7 : 7 → 8		Trajectory 7 : 6
Trajectory 8 : 11		Trajectory 8 : 7 → 8
Trajectory 9 : 16		Trajectory 9 : 11
Trajectory 10 : 19		Trajectory 10 : 16
Trajectory 11 : 20		Trajectory 11 : 19
Trajectory 12 : 21		Trajectory 12 : 20
Trajectory 13 : 22		Trajectory 13 : 21
		Trajectory 14 : 22

4.5.2 対応付け失敗に関する考察

シーケンス(2)における動線片の対応付けの結果に注目すると、歩容特徴を用いた場合(図4.14)、用いない場合(図4.15)両者において、視野内の端際付近での誤対応を確認できる。一方で、観測視野の端際では、新たな人物が視野内に現れる、もしくは視野内を移動していた人物が視野外に出る、といった行動が頻繁に観測されると考えられる。これらのことから、マルコフ連鎖によりモデル化した動線の対応付けにおける初期状態確率 $P_{init}(T_i^{in})$ 、終了状態確率 $P_{term}(T_i^{in})$ が動線片の誤対応に影響を及ぼしていると推測できる。前述しているように、本手法では、初期状態確率と終了状態確率を実験的に決定しており、ある一定の値を利用している。しかしながら、これらの失敗結果からもわかるように、視野内の端の領域では動線片が初期状態、終了状態となる確率は視野内中心位置に比べて高いと考えられる。

このような問題へ対処するために、観測領域のうちどの領域が出入口でどの領域が遮蔽物のある領域なのかを理解する、シーン認識の技術が必要となると考えられる。観測領域に関する事前知識を反映させることにより、適切な初期状態・終了状態確率の計算が行うことができると考えられる。

4.6 本手法のまとめ

本手法では、人物固有の性質である歩容特徴を利用することにより、人物動線の対応付けを実現した。追跡対象である人物に属する特徴点の動きの軌跡から、歩容特徴に相当する周波数特徴を求め、特徴量集合間の類似度を測る Pyramid Match Kerne により軌跡片間の類似度を計算する。このように計測される動線片間の類似度をもとに、二部グラフの最大重みマッチング手法の一つであるハンガリアン法を適用することで、最適な動線の対応付けを行うことができる。実際の映像を用いた実験を通じ、本手法の有効性を確認した。

第5章 結論

5.1 本研究のまとめと貢献

本研究では、実世界における人間の行動の理解に向けた、カメラ映像を用いた行動計測技術に関する研究を行った。特に、行動計測技術のうち重要な役割を果たす人物追跡技術についての技術開発を行った。

実環境において頑健に対象の人物を追跡するためには、複雑な背景、追跡対象の見えの変化、障害物による遮蔽など、行動計測の性能低下を引き起こす様々な問題へ対処しなければならない。本研究では、人物の行動特徴に基づく人物追跡手法を開発することにより、その問題へ対処した。

人間の行動特徴とは、行動に関する個人が持つ特性のことであり、これは個々の人物を追跡するための特徴的な指標と見なすことができる。本研究では、観測環境に基づく行動特徴、人物の個人性に基づく行動特徴の二つの行動特徴を考えた。このような行動特徴を用いることにより、本研究では次の三つの技術を開発した：

- (1) 行動履歴に基づく人物存在確率の利用による人物追跡の安定化、
- (2) 歩容特徴と局所的な見えに基づく混雑環境下人物追跡、
- (3) 歩容特徴を用いた人物動線の対応付け。

以下、それぞれの研究についてまとめ、本研究を遂行することにより得られた貢献、ならびに今後の課題について述べる。

5.1.1 行動履歴に基づく人物存在確率の利用による人物追跡の安定化

コンビニエンスストアやスーパーマーケットといった屋内環境を想定した場合、決まった通路の頻繁な通行、立ち読みのための本棚付近での滞留といった人物の行動は、対象空間の特定の領域で頻繁に観測される。このような人物の行動を長時間観測することにより、行動履歴に基づいた人物の存在確率分布を得ることができる。この人物

存在確率（本研究では環境属性と呼ぶ）を，追跡安定化のための情報源（importance function）として時系列フィルタの一つであるパーティクルフィルタの枠組みに組み込む．また環境属性は，追跡器により毎フレーム推定される追跡結果を用いて逐次的に更新される．これにより，安定な人物追跡が実現される．

本研究により得られた貢献，残された課題について以下に示す．

貢献

本研究の貢献は，人物の行動履歴に基づいた存在確率（環境属性）を追跡の枠組みに導入することにより，人物三次元追跡の安定化を図ることができる点にある．

観測環境の物理的な制約を追跡に利用する手法 [91] は提案されているものの，人物の行動履歴に基づいた人物存在確率の提案とその人物追跡への応用法は，これまでにないアプローチである．また人物追跡と環境属性の更新は同時に実行することが可能であるため，机や棚の移動といった観測環境の変動に柔軟に対応することができる．これにより，静的な観測環境（屋内環境）における人物追跡の安定化を図ることができる．具体的には，観測環境の事前知識（障害物や出入口の位置）を必要としない柔軟な追跡初期化，追跡失敗からの高速な追跡再開を実現することが可能となる．

今後の課題

追跡性能の評価の節（2.5.6 節）において考察しているように，観測環境において人物が普段ほとんど行わないような動きをする場合，すなわち環境属性の確率密度値が小さい領域に移動する場合に起き得る問題へ対処しなければならない．環境属性の確率密度値が小さい領域では，システムは人物は存在しにくいとみなすため，追跡初期化用の仮説のサンプリングを重点的に行わない．そのため，このような領域において追跡が不安定化・失敗した場合，速やかな追跡再開を行うことは難しい．

このような問題を解決するために，人物存在可能性が低い領域を集中的に監視する別のシステムとの併用が考える．また，最低限の追跡初期化性能を残すために，環境属性の極端な偏りを吸収する枠組みの提案が考えられる．例を挙げると，本研究でも用いた追跡対象の行動状態の考慮，画像の色情報を用いた人物存在確率の安定化手法 [32] の利用といったことが考えられる．

二つ目の課題は，環境属性のノンパラメトリックモデルへの拡張である．本手法では，環境属性をパラメトリックモデルである混合正規分布で表現した．しかしな

から，混合正規分布の分布の個数がハイパーパラメータであるため，分布の個数の設定次第で大きく性能が変化することが懸念される．このような問題に対し，パルツェン密度推定 [15,23,24] といったノンパラメトリックな確率密度分布推定手法の適用が効果的であると考えられる．

本研究を基にした今後の発展について簡単に述べる．観測環境における人間の行動は，時間帯によって大きく変化すると考えられる．例えば店舗内における人物の行動を考えると，昼食の時間帯などは弁当や飲み物を購入するために顧客は弁当売場付近に存在しやすいと考えられる．これに対し深夜の時間帯では，立ち読みのために来店する顧客が増えることが予想される．このように，特定の時間帯に即した環境属性を構築することにより，更なる性能の改善が期待される．

観測環境内における人物の動きの流れの学習もまた興味深いトピックである．例えば，買い物を目的とした顧客は，店舗内を左回りに行動しやすい傾向があり，その一方で ATM のみを利用する顧客は ATM に直接向かう行動をとりやすい [90]．存在可能性の学習だけでなく，このような動きの傾向を学習することにより，更なる追跡の安定化が期待される．近年報告されている粉体力学の考え方に基づいた場の流れの学習手法 [1,2] が一つの参考になるのではないだろうか．

5.1.2 歩容特徴と局所的な見えに基づく混雑環境下人物追跡

観測視野内に多数の人物が存在するような混雑環境（朝のラッシュ時における駅の構内，イベント会場など）において，頻繁に発生する遮蔽や，複数の人物が非常に近接していることにより，個々の人物を正しく追跡することが難しい．そこで本手法では，人物の個人性に相当する歩容特徴と局所的な見えの時間変動の一貫性という二つの指標を，特徴点軌跡のクラスタリングに基づく追跡の枠組みへ組み入れた．

周波数空間における歩容特徴は，生体認証の分野において頻繁に利用されている指標であり，個人を識別するための重要な手掛かりであることが知られている．また，局所領域における見えの時間的な変化は，人物の動きが周りとは類似する傾向のある混雑環境下において個々の人物を区別するための効果的な指標となる．このような動きと見えの異なる種類の指標を利用することにより，混雑環境下において頑健な人物追跡を実現した．

本研究を遂行することにより得られた貢献，残された課題について以下に示す．

貢献

本研究の主な貢献は、歩容特徴という人物の個人性に基づく行動特徴と局所的な見えを利用することにより、動的に変動する観測環境（混雑環境）において、頑健な人物追跡を実現できる点にある。

人間が生来持つ個人性である歩容特徴を、特徴点軌跡クラスタリングの考え方のもと人物追跡に応用することは、これまでにない斬新な方法である。個人認証などにも利用されていることからわかるように、混雑環境下における人物追跡においても追跡対象を他と区別するための指標として重要な役割を果たす。

局所的な見えの情報を利用する考え方は、これまでも提案されているものの、隣接する三本の特徴点軌跡により張られる三角パッチ内の見えの時間変動を考える方法は類を見ない。また本手法では色ヒストグラムを利用したものの、HoG 特徴量などの強力な見えの指標を応用することは容易であり、その拡張性も有用であると言える。

今後の課題

本研究における大きな仮定は、特徴点と同じ人物に属するならば、その動きの軌跡に類似した性質が表れるという点である。しかしながら、実験の節（3.5 節）において考察しているように、同じ人物に属していても、胴体や腕などの異なる部位に属する特徴点の動きは大きく異なる。これにより、正しくは一人の人物であるにも拘らず、システムは複数の人物が存在するとみなしてしまう。このような同一人物内の動き特徴の差異を吸収するような枠組みの導入が必要となる。

特徴点軌跡の生成の高度化も対処すべき大きな課題である。本追跡アルゴリズムは、特徴点の動きの軌跡が生成されていることを前提としているため、特徴点軌跡自体が存在しない場合、クラスタリング処理を行うことができない。このような問題に対処するために、軌跡の生成に利用する特徴点の種類をより頑健なものに変更することが考えられる。現在のシステムで利用する特徴点は、“ Good Features to Track ”[84,93] と呼ばれるものであるが、この代わりに、近年物体認識や行動認識の分野で利用されている SIFT 特徴 [60]、SURF 特徴 [4] などを用いることにより、性能の改善が期待される。また 3.5 節でも触れたように、Mean Shift アルゴリズムと SIFT 特徴量を組み合わせた特徴点追跡手法 [94, 108] などの利用も考えられる。

本手法のクラスタリング処理は、閾値処理に基づく単純なエッジ除去であり、またエッジ除去のためのハイパーパラメータ th_p が存在することから、必ずしも頑健

に動作するとは限らない．このような問題に対し，Normalized Cut [83] などのグラフカットのアルゴリズムの適用が考えられる．これにより追跡性能の向上が期待される．

本手法は人物の動きに基づいた追跡手法である．そのため，人物が停留してしまうと，動き情報が得られなくなり追跡することができなくなるといった本質的な問題がある．そのため，第三章の関連研究において述べたような見えに基づいた追跡手法や背景差分に基づく手法などを組み合わせることにより，このような問題へ対処することが必要である．

5.1.3 歩容特徴を用いた人物動線の対応付け

障害物や他の人物により追跡対象が遮蔽されるとき，対象の人物追跡は失敗する．これに伴い，人物の動線はいくつかの動線片に分断されることが考えられる．観測視野内における一貫した人物追跡を実現するために，分断化された動線片を対応付ける枠組みが必要となる．これに対し本手法では，人物固有の性質である歩容特徴に基づく動線の対応付けを実現した．歩容特徴は個々の人物特有の性質であるため，人物動線の対応付けにおいて効果的な指標になることが期待される．

具体的には，追跡対象である人物に属する特徴点の動きの軌跡群により得られる手掛かりを，動線の対応付けの指標として利用した．得られる特徴点軌跡群より，属する人物の歩容特徴に相当する周波数特徴を求める．特徴点軌跡群から得られる歩容特徴を用い，特徴量集合間の類似度を測る Pyramid Match Kernel (PMK) を適用することにより，動線間の類似度を計測する．得られる動線片間の類似度をもとに，ハンガリアン法に基づく最適化を行うことにより，最適な動線の対応付けを実現した．

本研究を遂行することにより得られた貢献，残された課題について以下に示す．

貢献

本研究の貢献は，人物の個人性である歩容特徴を利用することにより，頑健な人物動線の対応付けを実現することが可能となる点にある．本手法により，混雑環境下においても頑健な対応付けを実現することが可能となった．

観測視野内における一貫した人物追跡を実現するために，人間の個人性である歩容特徴を，属する特徴点軌跡群から抽出し，動線の対応付けに用いる考え方は，こ

れまででない新しいアプローチである．生体認証の研究分野において歩容特徴は利用されているが，その多くは人物のシルエットを用いたパターンマッチングによるものが多く，特徴点の動きの軌跡から歩容特徴を取得する方法は類を見ない．

また，Pyramid Match Kernel と呼ばれるデータ集合間の類似度計測手法の適用により，対象の動線片間の類似時計算を高精度に行うことができる．このようなPMKの利用は，一般物体認識や行動認識の研究分野で活発に利用されているものの，人物追跡の安定化に向けた動線片の対応付けへの応用例は見られない．

今後の課題

実験の考察の節(4.5.2節)において議論しているように，マルコフ連鎖によりモデル化した動線の対応付けにおける初期状態確率 $P_{init}(T_i^{in})$ ，終了状態確率 $P_{term}(T_i^{in})$ の設定が性能に大きな影響を及ぼすことが考えられる．本研究では，これらの確率を実験的に決定し，ある一定の値を利用している．そのため，対応付けが複雑になりうる混雑環境では，誤った対応付けを引き起こす可能性がある．

このような問題へ対処するために，観測領域のうちどの領域が出入口でどの領域が遮蔽物のある領域なのかを理解する，シーン認識の技術が必要となると考えられる．人物が通行しうる視野内の端の領域では，動線片が初期状態，終了状態となる確率は視野内中心位置に比べて高いと見なすことができる．このような観測領域に関する事前知識を反映させることにより，適切な初期状態・終了状態確率の計算を行うことができると考えられる．

第三章の特徴点軌跡のクラスタリングによる追跡の研究においても議論しているように，特徴点軌跡群を利用する手法において，特徴点軌跡の生成の高度化は考慮すべき重要な課題である．そのため，5.1.2節でも述べたように，より密に検出でき，より頑健に軌跡群を獲得できる特徴点の利用が求められる．

本研究では，人物動線の対応付けのみについて議論しており，実際の人物追跡の枠組みへ適用する段階にない．観測視野内の一貫した人物追跡を実現するために，上述した問題点について対処し，人物追跡の枠組みへ応用することが求められる．本手法は，特徴点軌跡群の性質を用いていることから，特徴点軌跡のクラスタリングに基づく人物追跡手法 [7, 58, 74, 88] に組み込むことが効果的なのではなからうか．

5.2 将来展望

カメラ映像を用いた人間の行動計測技術とは、「誰が・いつ・どこで」の自動認識を行うことである。この課題のうち、本研究では、人物追跡、すなわち「いつ・どこに」の課題について研究し、議論した。本研究成果に基づいた行動計測技術の今後の発展に期待したい。

しかしながら、実世界における頑健な行動計測技術の確立には、未だ多くの課題が残されている。以下の節では、本研究では議論しなかった課題について言及する。それに加え、行動計測技術を利用した応用技術の今後について展望する。

複数の観測領域を跨ぐ広域行動計測

本研究では、単一の観測視野内における人物追跡について研究してきた。しかしながら、安心・安全の社会環境の実現のためには、施設全体、更には町全体を包括するような広域監視システムの確立が望まれる。直感的にわかるように、広域をカメラで観測するためには、複数のカメラの設置が必須である。しかしながら、プライバシーの問題、カメラの設置位置の物理的な問題などから、全ての領域を観測できるようにカメラを配置することは難しい。

このような問題に対処するために、視野を共有しない複数の観測領域を跨ぐことを可能とする、広域行動計測技術の確立が求められる。これを実現するために、以下の技術が必要になると考えられる：

- 観測視野内における人物追跡，
- 複数の観測領域において観察される人物の同一性推定，
- 観測視野外における人物の動線推定．

観測視野内における人物追跡については、本研究において議論した。また複数の観測領域間での人物の同一性推定については、第四章の関連研究の節（4.2 節）において概観しているため、そこを参照していただきたい。ここでは、広域行動計測技術の確立のための最後の課題である、「観測視野外での人物の動線推定」について言及する。

カメラで観察されていない領域における人物の振る舞いを推定することは、人物の行動を一貫して追跡するために重要な技術である。近年、この課題に対していくつかの研究がなされている。一つは、観測領域における対象の振る舞いから、視野外

における動きを推定する方法である [52,75] . このような手法は , 観測視野内における振る舞いと同様の振る舞いを観測視野外においても行うという強い仮定をしているため , 追跡対象が観測視野外において不規則な行動をした場合 , 推定が不安定になる問題が残っている .

二つに , カメラだけでなくレーザや光電センサなどの異なる種類のセンサを同時に利用する , センサ統合のアプローチが考えられる [53,67] . レーザや光電センサは , カメラに比べ設置に関する制約が弱いいため , カメラで観測できない領域を補助することが容易であると考えられる .

このように , 広域監視システムの実現に向けた要素技術は広く研究されている . 今後の技術の発展が期待される .

“誰”の認識のための個人認証

“誰”の認識とはすなわち , 追跡している人物の名前 , 年齢 , 性別といった属性を認識することを意味する . カメラ画像による個人認証を行うためには , 個人性が現れる歩容特徴 , 顔特徴を手がかりに行うことが一般的である .

歩容特徴に関する個人認証は , 本研究でも利用しているように , 本追跡システムと親和性が高いと考えられる . このような認証技術を併用することにより , 個人の識別情報と紐付けされた人物行動計測が可能となると考えられる .

近年の顔認識技術の発展に伴い , 顔画像に基づく個人認証は実用に近い技術になってきていると考えられる . 店舗内などの屋内環境においては , 顔検出 , 認識が比較的容易であると考えられるが , 動的に変動するような複雑環境においては顔領域を安定に検出することは難しい . そのため , 実環境 (例えば多数の人物が存在するような混雑シーン) における顔検出の不安定性に関する対処が必要となると考えられる .

顔認証と歩容認証の同時利用が実現できれば , 頑健な個人認証を行うことができると考えられる . 今後の技術発展に期待したい .

人間の行動理解のための行動認識

人間の行動計測技術を基盤とした上位技術の一つに , 人間の行動認識が挙げられる . 行動認識とは , 観測映像中の人物は一体何をしているのか , という行動の意味を理解するための技術であり , 近年注目されている研究分野である . 本研究により得られた人物動線は , このような行動認識において必要不可欠な情報である .

行動計測により得られた動線データを用いる応用例として、不審者人物の検出が考えられる。鈴木らは、対象空間における人物の動きには特定の傾向があることに着目し、人物追跡により得られる動線データを分類することで、傾向から逸脱した非定常行動を検出している [90]。また Hu らは、店舗内において購買に関連する特定の行動（商品を手取るなど）を検出することを実現している [37]。

このような不審者人物検出、特定の行動の検出以外にも、一般的な行動認識に関する研究は数多く行われており [27, 50, 78]、その発展は目覚ましい。実世界における人間の任意の行動理解の実現のために、行動計測、認識技術両分野の技術の発展を期待したい。

謝辞

本研究を遂行するにあたり、日頃から熱心にご指導くださいました佐藤洋一准教授に心から厚く御礼申し上げます。

また、本研究を進めるにあたり、貴重なご指摘ならびにご助言いただきました、岡部孝弘助教、国立情報学研究所 杉本晃宏教授、佐藤いまり准教授に深く御礼申し上げます。

佐藤研究室の卒業生である、埼玉大学 小林貴訓助教には、研究に関する議論のみならず、機器の利用方法やプログラムの実装方法など様々な面でお世話になりました。深く御礼申し上げます。同じく佐藤研究室の卒業生である、電気通信大学 木谷クリス真実助教には、理論に関する議論以外にも、学会論文の校正など多大なご指導をいただきました。ここに深く御礼申し上げます。

佐藤研究室の皆様には、日頃の研究生生活において大変お世話になりました。島野美保子 JST さきがけ研究員には、研究の理論の部分で熱心に議論をしていただきました。ここに御礼申し上げます。同学年の菅野裕介氏、小林理弘氏には、日頃より研究に関する議論をはじめとした様々な相談に乗っていただき、多くのご協力をいただきました。ここに感謝申し上げます。また、秘書の鈴木咲恵さん、新木真理子さん、今川洋子さん、薄井千緒さんには、出張や研究費の運用など数多くの事務手続きにおいて大変お世話になりました。心より感謝申し上げます。

最後に、貴重な学生生活を支援してくださった両親に深く感謝します。

平成 22 年 1 月

杉村 大輔

参考文献

- [1] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–6, 2007.
- [2] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. In *Proc. of European Conference on Computer Vision (ECCV)*, pp. 1–14, 2008.
- [3] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf. Parametric correspondence and chamfer matching: Two new techniques for image matching. In *Proc. of International Joint Conference on Artificial Intelligence (IJCAI)*, Vol. 2, pp. 659–663, 1977.
- [4] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *International Journal of Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346–359, 2008.
- [5] J. Beis and D. Lowe. Shap indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1000–1006, 1997.
- [6] S. Birchfield and S. Rangarajan. Spatiograms versus histograms for region-based tracking. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 1158–1163, 2005.
- [7] G. Brostow and R. Cipolla. Unsupervised bayesian detection of independent motion in crowds. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 594–601, 2006.
- [8] R. Duraiswami C. Yang and L. Davis. Fast multiple object tracking via a hierarchical particle filter. In *IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 212–219, 2005.

- [9] Y. Cai, K. Huang, and T. Tan. Human appearance matching across multiple non-overlapping cameras. In *Proc. of IAPR International Conference on Pattern Recognition (ICPR)*, pp. 1–4, 2008.
- [10] J. R. Casas, A. P. Sitjes, and P. P. Folch. Mutual feedback scheme for face detection and tracking aimed at density estimation in demonstrations. *Vision, Image and Signal Processing*, Vol. 152, No. 3, pp. 334–346, 2005.
- [11] T. Chateau, V. G. Belille, F. Chausse, and J. T. Lapreste. Real-time tracking with classifiers. In *Proc. of International Workshop on Dynamical Vision*, pp. 218–231, 2006.
- [12] E. D. Cheng and M. Piccardi. Matching of objects moving across disjoint cameras. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, pp. 1769–1772, 2006.
- [13] M. Cheng, M. Ho, and C. Huang. Gait analysis for human identification through manifold learning and hmm. In *Proc. of IEEE Workshop on Motion and Video Computing (WMVC)*, pp. 1–8, 2007.
- [14] L. P. Chew. Constrained delaunay triangulations. In *Proc. of Annual Symposium on Computational Geometry*, pp. 215–222, 1987.
- [15] D. Comaniciu. An algorithm for data-driven bandwidth selection. *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 25, No. 2, pp. 281–288, 2003.
- [16] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 24, No. 5, pp. 603–619, 2002.
- [17] D. N. T. Cong, C. Achard, L. Khoudour, and L. Douadi. Video sequences association for people re-identification across multiple non-overlapping cameras. In *Proc. of IAPR International Conference on Image Analysis and Processing (ICIAP)*, pp. 179–189, 2009.
- [18] G. Csurka, C. R. Dance, L. Fan, and C. Bray. Visual categorization with bags of keypoints. In *Proc. of European Conference on Computer Vision (ECCV)*, pp. 1–22, 2004.

- [19] N. Dalal, B. Triggs, and C. Schmid. Histograms of oriented gradients for human detection. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893, 2004.
- [20] D. Demirdjian, K. Tollmar, K. Koike, N. Checka, and T. Darrell. Activity maps for location-aware computing. In *Proc. of IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 70–75, 2002.
- [21] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *Journal of Royal Statistical Society B*, Vol. 39, pp. 1–22, 1977.
- [22] L. Dong, V. Parameswaran, V. Ramesh, and I. Zoghlami. Fast crowd segmentation using shape indexing. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 1–8, 2007.
- [23] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2001.
- [24] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *Proc. of European Conference on Computer Vision (ECCV)*, Vol. 2, pp. 751–767, 2000.
- [25] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *The Journal of Machine Learning Research*, Vol. 4, pp. 933–969, 2003.
- [26] N. Gheissari, T. B. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 1528–1535, 2006.
- [27] A. Gilbert, J. Illingworth, and R. Bowden. Fast realistic multi-action recognition using mined dense spatio-temporal features. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 925–931, 2009.
- [28] M. Gill and A. Spriggs. Assessing the impact of cctv. *Home Office Research Study*, 252, 2005.

- [29] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 1458–1465, 2005.
- [30] K. Grauman and T. Darrell. Approximate correspondences in high dimensions. In *Proc. of Advances in Neural Information Processing Systems (NIPS)*, pp. 505–512, 2006.
- [31] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proc. of European Conference on Computer Vision (ECCV)*, pp. 262–275, 2008.
- [32] H. Habe, H. Nakagawa, and M. Kidode. Efficient acquisition of human existence priors from motion trajectories. In *IEEE Workshop on Visual and Contextual Learning from Annotated Images and Videos*, pp. 85–91, 2009.
- [33] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux. Interest points harvesting in video sequences for efficient person identification. In *Proc. of IEEE International Workshop on Visual Surveillance (VS)*, pp. 1–7, 2008.
- [34] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *Proc. of ACM/IEEE International Conference on Distributed Smart Cameras*, pp. 1–6, 2008.
- [35] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothes-invariant gait identification using part-based adaptive weight control. In *Proc. of IAPR International Conference on Pattern Recognition (ICPR)*, pp. 1–4, 2008.
- [36] L. Hu, Y. Wang, S. Jiang, Q. Huang, and W. Gao. Human reappearance detection based on on-line learning. In *Proc. of IAPR International Conference on Pattern Recognition (ICPR)*, pp. 1–4, 2008.
- [37] Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang. Action detection in complex scenes with spatial and temporal ambiguities. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 128–135, 2009.

- [38] C. Huang, B. Wu, and R. Nevatia. Robust object tracking by hierarchical association of detection responses. In *Proc. of European Conference on Computer Vision (ECCV)*, Vol. 2, pp. 788–801, 2008.
- [39] Y. Hyodo, K. Fujimura, S. Yuasa, T. Naito, and S. Kamijo. Pedestrian tracking through camera network for wide area surveillance. In *Proc. of International Conference on Systems, Man and Cybernetics*, pp. 656–661, 2008.
- [40] Ipa 「教育用画像素材集サイト」, <http://www2.edu.ipa.go.jp/gz/>.
- [41] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision (IJCV)*, Vol. 29, No. 1, pp. 5–28, 1998.
- [42] M. Isard and A. Blake. Icondensation: Unifying low-level and high-level tracking in a stochastic framework. In *Proc. of European Conference on Computer Vision (ECCV)*, Vol. 1, pp. 893–908, 1998.
- [43] M. Isard and J. MacCormick. Bramble: A bayesian multiple-blob tracker. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 34–41, 2001.
- [44] O. Javed, K. Shafique, Z. Rasheed, and M. Shah. Modeling inter-camera space-time and appearance relationships for tracking across non overlapping views. *International Journal of Computer Vision and Image Understanding (CVIU)*, Vol. 109, No. 2, pp. 146–162, 2008.
- [45] K. Jeong and C. Jaynes. Object matching in disjoint cameras using a color transfer approach. In *Proc. of IAPR Workshop on Machine Vision and Applications (MVA)*, pp. 443–455, 2008.
- [46] Y. Jin and F. Mokhtarian. Data fusion for robust head tracking by particles. In *Proc. of IEEE Workshop on Visual Surveillance and Performance Evaluation and Tracking (VS-PETS)*, pp. 33–40, 2005.
- [47] R. Kaucic, A. G. A. Perera, G. Brooksby, J. Kaufhold, and A. Hoogs. A unified framework for tracking through occlusions and across sensor gaps. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 990–997, 2005.

- [48] S. M. Khan and M. Shah. A multiview approach to tracking people in crowded scenes using a planar homography constraint. In *Proc. of European Conference on Computer Vision (ECCV)*, Vol. 4, pp. 133–146, 2006.
- [49] Z. Khan, T. R. Balch, and F. Dellaert. An mcmc-based particle filter for tracking multiple interacting targets. In *Proc. of European Conference on Computer Vision (ECCV)*, Vol. 4, pp. 279–290, 2004.
- [50] K. M. Kitani, Y. Sato, and A. Sugimoto. Recovering the basic structure of human activities from noisy video-based symbol strings. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 22, No. 8, pp. 1621–1646, 2008.
- [51] Y. Kobayashi, D. Sugimura, Y. Sato, K. Hirasawa, N. Suzuki, H. Kage, and A. Sugimoto. 3d head tracking using the particle filter with cascaded classifiers. In *Proc. of British Machine Vision Conference (BMVC)*, pp. 37–46, 2006.
- [52] 小林貴訓, 佐藤洋一, 杉本晃宏. 視野を共有しないカメラ群を用いた人物行動軌跡の推定. 情報処理学会研究報告, 2005-CVIM-150-21, pp. 169–176, 2005.
- [53] 小林貴訓, 杉村大輔, 関真規人, 平澤宏祐, 鈴木直彦, 鹿毛裕史, 佐藤洋一, 杉本晃宏. 分散カメラとレーザ測域センサの統合によるエリア内人物追跡. 画像の認識・理解シンポジウム (MIRU), pp. 1–6, 2007.
- [54] 小出治. 防犯カメラの効果と利用について. 予防時報, Vol. 224, pp. 30–35, 2006.
- [55] S. Lee, Y. Liu, and R. Collins. Shape variation-based frieze pattern for robust gait recognition. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 1–8, 2007.
- [56] B. Leibe, K. Schindler, and L. V. Gool. Coupled detection and trajectory estimation for multi-object tracking. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 1–8, 2007.
- [57] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 878–885, 2005.

- [58] Y. Li and H. Ai. Fast detection of independent motion in crowds guided by supervised learning. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, Vol. 3, pp. 341–344, 2007.
- [59] Y. Li, C. Huang, and R. Nevatia. Learning to associate: Hybridboosted multi-target tracker for crowded scene. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 2953–2960, 2009.
- [60] D. G. Lowe. Object recognition from local scale invariant features. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 1150–1157, 1998.
- [61] G. Loy, L. Fletcher, N. Apostoloff, and A. Zelinsky. An adaptive fusion architecture for target tracking. In *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 261–265, 2002.
- [62] C. Madden and M. Piccardi. A framework for track matching across disjoint cameras using robust shape and appearance features. In *Proc. of IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 188–193, 2007.
- [63] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of European Conference on Computer Vision (ECCV)*, Vol. 3, pp. 151–163, 2006.
- [64] 松元郁祐, 加藤丈和, 和田俊和. Network augmented multisensor association-condensation: Condensation の自然な拡張による 3 次元空間内での人物頭部の実時間追跡. 情報処理学会研究報告, 2005-CVIM-150-21, pp. 161–168, 2005.
- [65] 松元郁祐, 加藤丈和, 和田俊和. 複数カメラを用いた condensation によるオクルージョンにロバストな人物追跡. 画像の認識・理解シンポジウム (MIRU), pp. 501–506, 2006.
- [66] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *International Journal of Computer Vision (IJCV)*, Vol. 65, No. 1–2, pp. 43–72, 2005.
- [67] 中村克行, 趙卉菁, 柴崎亮介. レーザスキャナと画像センサの融合による歩行者追跡. 画像センシングシンポジウム講演論文集, pp. 177–180, 2005.

- [68] P. Nillius, J. Sullivan, and S. Carlsson. Multi-target tracking-linking identities using bayesian network inference. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 2187–2194, 2006.
- [69] K. Nummiaro, E. Koller-Meier, and L. Van Gool. An adaptive color-based particle filter. *Image and Vision Computing*, Vol. 21, No. 1, pp. 99–110, 2003.
- [70] K. Okuma, A. Taleghani, N. Freitas, J. Little, and D. Lowe. A boosted particle filter: Multitarget detection and tracking. In *Proc. of European Conference on Computer Vision (ECCV)*, Vol. 1, pp. 28–39, 2004.
- [71] 大澤文孝. 書籍版 DirectX9 実践プログラミング. 工学社, 2005.
- [72] A. G. A. Perera, C. Srinivas, A. Hoogs, G. Brooksby, and W. Hu. Multi-object tracking through simultaneous long occlusions and split-merge conditions. In *Proc. of CS Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 666–673, 2006.
- [73] B. Prosser, S. G. Gong, and T. Xiang. Multi-camera matching using bi-directional cumulative brightness transfer functions. In *Proc. of British Machine Vision Conference (BMVC)*, pp. 1–10, 2008.
- [74] V. Rabaud and S. Belongie. Counting crowded moving objects. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 705–711, 2006.
- [75] A. Rahimi, B. Dunagan, and T. Darrell. Tracking people with a sparse network of bearing sensors. In *Proc. of European Conference on Computer Vision (ECCV)*, pp. 507–518, 2004.
- [76] C. E. Rasmussen and C. K. I. Williams. *Gaussian Process for Machine Learning*. The MIT Press, 2006.
- [77] M. D. Rodriguez and M. Shah. Detecting and segmenting humans in crowded scenes. In *Proc. of ACM International Conference on Multimedia (ACMMM)*, pp. 353–356, 2007.

- [78] M. S. Ryoo and J. K. Aggarwal. Recognition of high-level group activities based on activities of individual members. In *Proc. of IEEE Workshop on Motion and Video Computing (WMVC)*, pp. 1–8, 2008.
- [79] M. Sato and S. Ishii. On-line em algorithm for the normalized gaussian network. *Neural Computation*, Vol. 12, No. 2, pp. 407–432, 2000.
- [80] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. In *Proc. of Annual Conference on Computational Learning Theory*, pp. 297–336, 1999.
- [81] Y. Shan, H. S. Sawhney, and R. Kumar. Vehicle identification between non-overlapping cameras without direct feature matching. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 378–385, 2005.
- [82] J. Sherrah and S. Gong. Fusion of perceptual cues for robust tracking of head pose and position. *Pattern Recognition*, Vol. 32, No. 8, 2001.
- [83] J. Shi and J. Malik. Normalized cuts and image segmentation. *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 22, No. 8, pp. 888–905, 2000.
- [84] J. Shi and C. Tomasi. Good features to track. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 593–600, 1994.
- [85] O. Sidla, Y. Lypetsky, N. Brandle, and S. Seer. Pedestrian detection and tracking for counting applications in crowded situations. In *Proc. of IEEE International Conference on Video and Signal Based Surveillance (ICVSS)*, pp. 70–75, 2006.
- [86] X. Song, J. Cui, H. Zha, and H. Zhao. Vision-based multiple interacting targets tracking via on-line supervised learning. In *Proc. of European Conference on Computer Vision (ECCV)*, pp. 642–655, 2008.
- [87] 杉本晃宏, 谷内清剛, 松山隆司. 確信度付き仮説群の相互作用に基づく複数対象追跡. *情報処理学会論文誌コンピュータビジョンとメディア*, Vol. 43, pp. 69–84, 2002.
- [88] D. Sugimura, K. M. Kitani, T. Okabe, Y. Sato, and A. Sugimoto. Using individuality to track individuals: Clustering individual trajectories in crowds using local

- appearance and frequency trait. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 1467–1474, 2009.
- [89] 杉村大輔, 小林貴訓, 佐藤洋一, 杉本晃宏. 行動履歴に基づく人物存在確率の利用による人物三次元追跡の安定化. *情報処理学会論文誌コンピュータビジョンとメディア*, Vol. 1, No. 2, pp. 100–110, 2008.
- [90] N. Suzuki, K. Hirasawa, K. Tanaka, Y. Kobayashi, Y. Sato, and Y. Fujino. Learning motion patterns and anomaly detection by human trajectory analysis. In *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, pp. 498–503, 2007.
- [91] 鈴木達也, 岩崎慎介, 小林貴訓, 佐藤洋一, 杉本晃宏. 環境モデルの導入による人物追跡の安定化. *電子情報通信学会論文誌 DII*, Vol. J88-DII, No. 8, pp. 1592–1600, 2005.
- [92] L. F. Teixeira and L. Corte-Real. Video object matching across multiple independent views using local descriptors and adaptive learning. *Pattern Recognition Letters*, Vol. 30, pp. 157–167, 2009.
- [93] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University, 1991.
- [94] Yuji Tsuduki and Hironobu Fujiyoshi. A method for visualizing pedestrian traffic flow using sift feature point tracking. In *Proc. of Pacific-Rim Symposium on Image and Video Technology (PSIVT)*, pp. 25–36, 2009.
- [95] P. Tu, T. Sebastian, G. Doretto, N. Krahnstoeber, J. Rittscher, and T. Yu. Unified crowd segmentation. In *Proc. of European Conference on Computer Vision (ECCV)*, 2008.
- [96] D. Tweed and A. Calway. Tracking many objects using subordinated condensation. In *Proc. of British Machine Vision Conference (BMVC)*, pp. 283–292, 2002.
- [97] J. R. Vacca. *Biometric Technologies and Verification Systems*. Elsevier Inc., 2007.
- [98] J. Vermaak, A. Doucet, and P. Perez. Maintaining multi-modality through mixture tracking. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, Vol. 2, pp. 1110–1116, 2003.

- [99] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 511–518, 2001.
- [100] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu. Shape and appearance context modeling. In *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 1–8, 2007.
- [101] Y. Wang, J. Wu, and A. Kassim. Particle filter for visual tracking using multiple cameras. In *Proc. of IAPR Conference on Machine Vision Applications (MVA)*, pp. 298–301, 2005.
- [102] B. C. Welsh and D. P. Farrington. Crime prevention effects of closed circuit television: A systematic review, 2002.
- [103] B. Wu and R. Nevatia. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *International Journal on Computer Vision (IJCV)*, Vol. 75, No. 2, pp. 247–266, 2007.
- [104] J. Xing, H. Ai, and S. Lao. Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses. In *Proc. of IEEE CS conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1200–1207, 2009.
- [105] 横田秀次郎. 32 ビット通信プログラム作成法. CQ 出版社, 1997.
- [106] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.
- [107] T. Zhao and R. Nevatia. Bayesian human segmentation in crowded situations. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 459–466, 2003.
- [108] H. Zhou, Y. Yuan, and C. Shi. Object tracking using sift features and mean shift. *Computer Vision and Image Understanding (CVIU)*, Vol. 113, No. 3, pp. 345–352, 2009.

発表文献

論文誌

1. 杉村大輔, 木谷クリス真実, 岡部孝弘, 佐藤洋一, 杉本晃宏, “ 歩容特徴と局所的見えを考慮した特徴点軌跡のクラスタリングによる混雑環境下人物追跡 ”, 電子情報通信学会論文誌 D-II, 2010 (条件付採録)
2. 大野宏, 中嶋信生, 佐藤洋一, 小林貴訓, 杉村大輔, 加納梢, “ 不審行動検知による万引き防止策の一提案 ”, 日本セキュリティ・マネジメント学会論文誌, Vol. 23, No. 1, pp. 26-38, 2009 .
3. 杉村大輔, 小林貴訓, 佐藤洋一, 杉本晃宏, “ 行動歴に基づく人物存在確率の利用による人物三次元追跡の安定化 ”, 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol. 1, No. 2, pp. 100-110, 2008 .
4. 小林貴訓, 杉村大輔, 平澤宏祐, 鈴木直彦, 鹿毛裕史, 佐藤洋一, 杉本晃宏, “ パーティクルフィルタとカスケード型識別器の統合による人物三次元追跡 ”, 電子情報通信学会論文誌 D-II, Vol.J90-D, No.8, pp.2049-2059, 2007 .

国際会議

5. D. Sugimura, K. M. Kitani, T. Okabe, Y. Sato and A. Sugimoto, “ Using Individuality to Track Individuals: Clustering Individual Trajectories in Crowds using Local Appearance and Frequency Trait ”, *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 1467-1474, 2009.
6. D. Sugimura, Y. Kobayashi, Y. Sato and A. Sugimoto, “ Incorporating Long-Term Observations of Human Actions for Stable 3D People Tracking ”, *Proc. of IEEE Workshop on Motion and Video Computing (WMVC)*, pp. 1-7, 2008.

7. Y. Kobayashi, D. Sugimura, Y. Sato, K. Hirasawa, N. Suzuki, H. Kage and A. Sugimoto, " 3D Head Tracking using the Particle Filter with Cascaded Classifiers ", *Proc. of British Machine Vision Conference (BMVC)*, pp.37-46, 2006.

国内会議（査読あり）

8. 新川祥史, 杉村大輔, 佐藤洋一, 小池英樹, " 人の視覚特性に基づく大型ディスプレイによる複数人物への選択的情報提示 ", *インタラクシオン 2010*, 2010 . (発表予定)
9. 杉村大輔, 木谷クリス真実, 岡部孝弘, 佐藤洋一, 杉本晃宏, " 歩容特徴と局所的な見えを用いた特徴点軌跡クラスタリングによる混雑環境下人物追跡 ", *画像の理解・認識シンポジウム (MIRU2009)*, pp. 135-142, 2009 .
10. 杉村大輔, 小林貴訓, 佐藤洋一, 杉本晃宏, " 行動履歴に基づいた環境属性の自動構築を伴う三次元人物追跡 ", *画像の認識・理解シンポジウム (MIRU2007)*, pp.405-411, 2007 .
11. 小林貴訓, 杉村大輔, 平澤宏祐, 鈴木直彦, 鹿毛裕史, 佐藤洋一, 杉本晃宏, " カスケード型識別器を用いたパーティクルフィルタによる人物三次元追跡 ", *画像の認識・理解シンポジウム (MIRU2006)*, pp. 222-228, 2006 .

国内会議（査読なし）

12. 小林貴訓, 杉村大輔, 関真規人, 平澤宏祐, 鈴木直彦, 鹿毛裕史, 佐藤洋一, 杉本晃宏, " 分散カメラとレーザ測域センサの統合によるエリア内人物追跡 ", *画像の認識・理解シンポジウム (MIRU2007)*, pp. 839-844, 2007 .
13. 杉村大輔, 小林貴訓, 佐藤洋一, 杉本晃宏, " 行動履歴を反映させた適応的環境属性を伴う三次元人物追跡 ", *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*, 2006-CVIM-156-21, pp.171-178, 2006 .

特許・その他

14. 小林貴訓，佐藤洋一，杉村大輔，関真規人，平澤宏祐，鈴木直彦，鹿毛裕史，杉本晃宏，“パーティクルフィルタとカスケード型識別器の統合による人物三次元追跡～人物追跡の頑健化・高精度化に向けて～”，画像ラボ，日本工業出版，Vol. 18，No. 12，2007．
15. 佐藤洋一，小林貴訓，杉村大輔，鹿毛裕史，平澤宏祐，鈴木直彦，“人物追跡装置”，特願 2006-195877，2006．