

博 士 論 文

分散センサの統合による
エリア内人物追跡と動線推定

People Tracking and Trajectory Estimation by
Integrating Observations from Distributed Sensors for
Local Area Surveillance

東京大学大学院 情報理工学系研究科
電子情報学専攻

小林 貴訓

YOSHINORI KOBAYASHI

指導教員 佐藤洋一 准教授

2007年6月

概要

社会的な防犯意識の高まりと共に、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ものの、大量の監視カメラ映像を常に人手で監視することはもはや現実的ではなく、監視カメラ映像は、事後の参照のために一定期間保存された後、破棄される場合が多い。これに対して、監視カメラ映像を自動解析し、人物行動の計測が可能となれば、不審者の検出・発報や、POS データでは分からない顧客の購買行動などの情報を抽出できると考えられる。また、人物行動の計測結果に基づいたインタラクティブ広告や空調器機の制御など、広範な分野への応用が期待され、社会的意義も大きい。

一方で、公共施設や商業施設などで人物行動を計測する場合、センサの設置の制約や、障害物や人物相互による遮蔽、プライバシーの問題などから、エリア全体で安定した観測を得ることは難しい。そのため、環境に疎密に分散配置されたセンサ群を用いて特定エリア全体の人物行動を計測する技術が求められている。

そこで、本研究では、スーパーやコンビニエンスストアなどの数人～20人程度の人物が同時に存在する空間に着目し、監視カメラ映像に基づいて人物頭部の三次元位置及び向きを計測する手法を提案する。また、監視カメラで観測されない領域を含む特定エリア全体においては、人物動線を計測する手法について取り組む。

第2章では、監視カメラ映像を想定し、視野を共有した複数のカメラにより人物頭部の三次元位置と向きを追跡する手法を提案する。監視カメラなどの画像センサによる人物追跡では、人物が必ずしもカメラに対して一定方向を向いていないために起こる見えの変化の問題や、広い範囲を観察することにより頭部が低解像度で観察される問題、昼間と夜間などの照明変動の問題、さらに店舗内の商品棚などの複雑な背景の問題などの課題がある。

これらの課題に対して、識別器を用いた顔検出技術を時系列フィルタによる追跡の枠組みに統合することで、頑健かつ高精度に人物頭部を追跡する手法を提案する。矩形特徴を用いた AdaBoost 学習による識別器は、その精度と頑健性から、静止画像中の顔検出に広く用いられている。しかし、識別対象の大きさをさまざまに変化させながら、画面全体を探索することは追跡の枠組みにおいては効率的ではない。そこで、時系列フィルタの一つであるパーティクルフィルタに AdaBoost 学習による識別器を統合し、人物頭部の三次元位置と向きを追跡する。具体的には、パーティクルフィルタによって予測された仮説に対応した画像領域を、識別器によって評価する。このとき、人物頭部は実際の向きやカメラとの位置関係により、その見えが変化するため、頭部の各方向に対応した識別器をあらかじめ準備し、複数のカメラによりさまざまな方向で観察される人物頭部に対し、識別器を適応的に選択して評価する。これにより、人物頭部の三次元位置と同時に、向きを推定できる。多くの検証実験から、提案手法の優位性を確認した。

第3章では、スーパーやコンビニエンスストアなどの実環境を想定し、複数人物の追跡を行う手法を提案する。実環境での人物追跡では、人物の出現に対する追跡初期化という本質的な課題がある。また、監視カメラなどのセンサの設置には位置や姿勢などの物理的な制約により、理想的な設置状況を実現できない場合も多く、障害物や人物相互の遮蔽などによる観測不能状態が頻繁に起こると予想される。そのため、実環境では、観測不能により追跡が失敗した場合でも、良好な観測が得られた時点で、速やかに追跡を再開する枠組みが求められる。

これに対しては、新たにレーザ測域センサを導入し、エリア内での人物の出現を検出することで、頑健に追跡初期化を行う手法を提案する。複数の人物頭部は、パーティクルフィルタの拡張である Mixture Particle Filter により追跡し、状態遷移モデルとレーザ測域センサによる人物検出位置を統合した確率密度分布から仮説を生成することで、カメラ画像による追跡にレーザ測域センサを統合する。これにより、エリア内に人物が出現した時点で速やかに追跡を開始し、人物同士がすれ違うなどの遮蔽により追跡が失敗した場合でも、良好な観測が得られた時点で、速やかに追跡を再開できる。さらに、レーザ測域センサを導入することで、単眼カメラによる観測においても、安定な三次元追跡を実現し、分散配置されたセンサ群のさまざまな組合せによる三次元人物追跡を可能とした。

第4章では、複数観測領域間で観察される人物動線の対応付けの問題を扱う。エリア内全体を通じて、連続した人物動線を計測するためには、障害物や人物相互による遮蔽により断片化した人物動線や、複数観測領域間における人物動線に対応付ける必要がある。

環境に分散した観測領域では、異なる2地点で同時に同じ人物は観察されないという事実や、ある程度離れた観測領域間では、その間の旅行時間を予想できるなどの時間的空間的な尤もらしさ(時空間的尤度)を考えることができる。そこで、先行研究にならい、想定される多くの人物動線の対応から、尤もらしい人物動線の組を、観察された見えの類似性と、時空間的尤度から、グラフの最適化手法を利用して探索する。具体的には人物動線の出現・消失をノードとし、対応をエッジとして扱うことで、尤もらしい人物動線の組を重み付き2部グラフの最大マッチングにより取得した。

第5章では、エリア内全体における連続した人物動線の取得のため、観測領域外の人物動線を補間する手法を提案する。先行研究において、観測情報と運動モデルから、観測領域外を含む人物動線を最大事後確率推定により求める手法が提案されているが、室内の机や壁などの障害物の配置を考慮していないため、推定した人物動線が壁を突き抜けてしまうなどの問題があった。

これに対して、環境の制約条件を導入することで、観測領域外においても、尤もらしい人物動線を推定する手法を提案する。具体的には、人物の運動モデルとカメラによる観測モデルに加えて、環境形状から予測される人物の存在確率モデルを最大事後確率推定の枠組みで統合し、人物動線の尤もらしさを評価する関数を構築する。そして観測された一部の人物動線を用いて、評価関数の最適化を行うことで、観測領域外を含む人物動線を推定した。

第6章では、全体を総括し、今後の課題と展望について述べる。

目次

第 1 章	序論	1
1.1	本研究の背景	1
1.2	本研究の目的	2
1.3	本研究の課題	3
1.4	本研究の位置づけ	5
1.5	本論文の構成	6
第 2 章	時系列フィルタと識別器の統合による三次元人物頭部追跡	7
2.1	はじめに	7
2.2	関連研究	9
2.3	パーティクルフィルタ	11
2.4	AdaBoost 学習による識別器	13
2.4.1	ブースティング法	13
2.4.2	Viola と Jones の顔識別器	15
2.5	時系列フィルタと識別器の統合による三次元人物頭部追跡	17
2.5.1	人物頭部モデル	17
2.5.2	状態遷移モデルに基づく仮説の伝播	18
2.5.3	カメラ画像への射影	19
2.5.4	仮説の評価手順	20
2.5.5	カスケード型識別器による仮説の評価	21
2.5.6	非カスケード型識別器による仮説の評価	22
2.6	実験	23
2.6.1	カスケード型識別器を用いた追跡器による実験	25
2.6.2	非カスケード型識別器を用いた追跡器による実験	27
2.7	考察	32
2.7.1	識別器による仮説の評価	32
2.7.2	複数識別器の適応的利用	34
2.7.3	状態遷移モデルによる予測	36
2.7.4	カメラの台数と追跡性能	37
2.7.5	人物頭部の向き の 推定	38

2.8	まとめ	40
第3章	カメラとレーザ測域センサの統合によるエリア内複数人物追跡	41
3.1	はじめに	41
3.2	関連研究	44
3.3	Mixture Particle Filter	45
3.4	カメラとレーザ測域センサの統合によるエリア内複数人物追跡	47
3.4.1	レーザ測域センサ	47
3.4.2	レーザ測域センサによる人物頭部候補位置の検出	48
3.4.3	レーザ測域センサの統合による人物頭部追跡	49
3.5	実験	53
3.5.1	エリア内における複数人物の追跡	53
3.5.2	人物相互の遮蔽	56
3.5.3	単眼カメラによる三次元人物頭部追跡	59
3.6	考察	60
3.6.1	レーザ測域センサによる追跡初期化	60
3.6.2	追跡対象の部分的な遮蔽	62
3.7	まとめ	63
第4章	複数観測領域間の人物動線の対応付け	65
4.1	はじめに	65
4.2	関連研究	66
4.2.1	複数カメラの設置状況の推定	66
4.2.2	複数観測領域間の人物動線の対応付け	67
4.2.3	観測領域外を含む人物動線の推定	68
4.3	複数観測領域間の人物動線の対応付け	69
4.3.1	事前確率	71
4.3.2	時空間的尤度	71
4.3.3	見えの尤度	72
4.3.4	グラフによるモデル化	73
4.4	実験	75
4.4.1	シミュレーション	75
4.4.2	実環境における実験	77
4.5	まとめ	80
第5章	観測領域外を含む人物動線の推定	81
5.1	はじめに	81
5.2	観測領域外を含む人物動線の推定	82
5.2.1	人物頭部の移動軌跡	82
5.2.2	運動モデルと観測モデルによる評価	82

5.2.3	環境モデルの導入	85
5.2.4	各モデルによる評価の統合と最適化	86
5.3	実験	88
5.3.1	シミュレーション	88
5.3.2	実環境における実験	89
5.4	まとめ	93
第 6 章	結論	95
6.1	本研究の成果	95
6.2	今後の課題	97
6.3	将来展望	99
謝辞		101
参考文献		103
発表文献		111

目次

1.1	監視カメラの普及	1
1.2	監視カメラ映像を用いた人物の行動計測	2
1.3	本研究がめざすエリア内人物行動計測	3
1.4	人物頭部追跡の難しさ	3
1.5	観測できない領域を含む人物動線計測の課題	4
2.1	画像を用いた人物追跡の難しさ	8
2.2	識別器とパーティクルフィルタの併用による人物追跡	10
2.3	パーティクルフィルタにおける仮説の更新	12
2.4	ブースティング法による顔識別器の学習	13
2.5	カスケード型識別器	15
2.6	弱識別器で用いる矩形特徴	16
2.7	人物頭部モデル	17
2.8	人物頭部周辺のオプティカルフロー	19
2.9	入力画像領域の調整	20
2.10	カスケード型識別器による評価	21
2.11	IEEE1394 カメラ Flea	23
2.12	複数の PC による処理の分散	24
2.13	実験で得られたカメラ画像の例	24
2.14	カスケード型識別器を用いた追跡器による人物頭部の追跡結果	25
2.15	カスケード型識別器を用いた追跡器による人物頭部の追跡結果の軌跡	26
2.16	非カスケード型識別器を用いた追跡器による人物頭部の追跡結果	27
2.17	非カスケード型識別器を用いた追跡器による人物頭部の追跡結果の軌跡	28
2.18	尤度のキャリブレーションに用いた顔画像の例	29
2.19	尤度キャリブレーション結果 (Front)	29
2.20	尤度キャリブレーション結果 (Left45)	30
2.21	尤度キャリブレーション結果 (Left90)	30
2.22	尤度キャリブレーション結果 (Left135)	30
2.23	尤度キャリブレーション結果 (Right45)	31
2.24	尤度キャリブレーション結果 (Right90)	31

2.25	尤度キャリブレーション結果 (Right135)	31
2.26	カスケード段数に基づく評価	32
2.27	弱識別器出力に基づく評価	33
2.28	輪郭の輝度勾配の類似性に基づく評価	33
2.29	仮説が通過するカスケード段数	34
2.30	頭部の回転とカスケード型識別器の評価値	35
2.31	頭部の回転と非カスケード型識別器の尤度	35
2.32	識別器の数による追跡精度の比較	36
2.33	カメラ台数による追跡精度の比較	37
2.34	複数カメラによる追跡の頑健性	37
2.35	仮説群の期待値による頭部向きの推定誤差	39
2.36	識別器評価値の分布に基づく頭部向きの推定誤差	39
3.1	特定エリア内における複数人物追跡の難しさ	41
3.2	Merge-Split アプローチ	42
3.3	Mixture Particle Filter の概念	45
3.4	レーザ測域センサ (SICK 社 S300)	47
3.5	レーザ測域センサによる計測	48
3.6	レーザ測域センサの追跡の枠組みへの統合	49
3.7	レーザ測域センサによる観測に基づく追跡の開始	50
3.8	混合分布からの仮説のサンプリング	51
3.9	混合分布からのサンプリングによる仮説の更新	52
3.10	複数人物の追跡の様子	53
3.11	検出された人物頭部候補	53
3.12	複数人物頭部の追跡結果の軌跡 (非カスケード型識別器)	54
3.13	複数人物頭部の追跡結果の軌跡 (カスケード型識別器)	55
3.14	人物相互遮蔽 (カメラ画像上での頭部の遮蔽)	57
3.15	人物相互遮蔽 (レーザ測域センサによる観測の遮蔽)	58
3.16	単眼カメラによる三次元人物頭部追跡	59
3.17	追跡手法と画像による対象検出	60
3.18	空間を分割した立方体グリッドの画像への射影	61
3.19	立方体グリッドの観察に基づいた頭部検出	61
3.20	立方体グリッドの観察に基づいた追跡	62
4.1	カメラトポロジの推定	66
4.2	カメラ間における観測の対応付け	67
4.3	視野外を含む移動軌跡推定	68
4.4	時空間的尤度	69
4.5	見えの尤度	69

4.6	観測領域間の遷移確率による事前確率の定義	71
4.7	カーネル密度推定による時空間的尤度の推定	72
4.8	ヒストグラム距離に基づく見えの尤度の推定	72
4.9	有向グラフによるモデル化	73
4.10	重み付き 2 部グラフによる観測対応のモデル化	74
4.11	シミュレータ	75
4.12	人物を模擬した CG オブジェクトの移動経路	75
4.13	CG オブジェクトの追跡	76
4.14	観測エリア	77
4.15	観測領域 1 における追跡の様子	78
4.16	観測領域 2 における追跡の様子	78
4.17	対応付けの推定結果	79
5.1	環境モデルの概念	85
5.2	距離とペナルティの関係	86
5.3	最適化開始時に与える初期軌跡	87
5.4	シミュレータ	88
5.5	シミュレーション	89
5.6	室内レイアウト図と推定された軌跡	90
5.7	実験で得られたカメラ画像の例	91
5.8	軌跡の推定結果	92

表目次

2.1	勾配法とブロックマッチング法	18
2.2	カスケード型識別器を用いた追跡器による人物頭部の追跡における検出誤差	26
2.3	非カスケード型識別器を用いた追跡器による人物頭部の追跡における検出誤差	28
3.1	SICK 社レーザ測域センサ S300 の主な仕様	47
3.2	複数人物頭部の追跡における検出誤差 (非カスケード型識別器)	55
3.3	複数人物頭部の追跡における検出誤差 (カスケード型識別器)	56

第1章

序論

1.1 本研究の背景

社会的な防犯意識の高まりと共に、店舗や駅、商店街などで監視カメラの普及が進んでいる(図1.1)。イギリスの監視カメラ設置台数は、400万台を超え、ロンドンの街中では、歩行者は一日に300回、平均して5分に1回は監視カメラに撮影されるという。また、スコットランドでは、大量の監視カメラの設置により、実際に犯罪発生率が飛躍的に減少したとの報告もあり、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。

一方で、大量の監視カメラで撮影された映像を、常に人手で監視することはもはや現実的ではなく、多くの場合、監視カメラ映像は、事後の参照のために一定期間保存された後、破棄されている。また、監視カメラの画質の向上やさらなる台数の増加により、近い将来には、映像の長期的な保存も困難になると予想される。



図 1.1 監視カメラの普及

監視対象について考えると，店舗での万引き防止や駅での安全確認など，人物行動の観察を目的として監視カメラが設置されている場合が圧倒的に多いことは明らかである．そこで，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，人物行動の計測が可能となれば，不審者の検出・発報や，POS データからでは分からない顧客の購買行動など，より有益な情報が抽出可能となると考えられる．また，人物行動の計測結果に基づいたインタラクティブ広告や，環境に配慮した空調器機の制御など，広範な分野への応用も期待でき，社会的意義は大きい(図 1.2)．

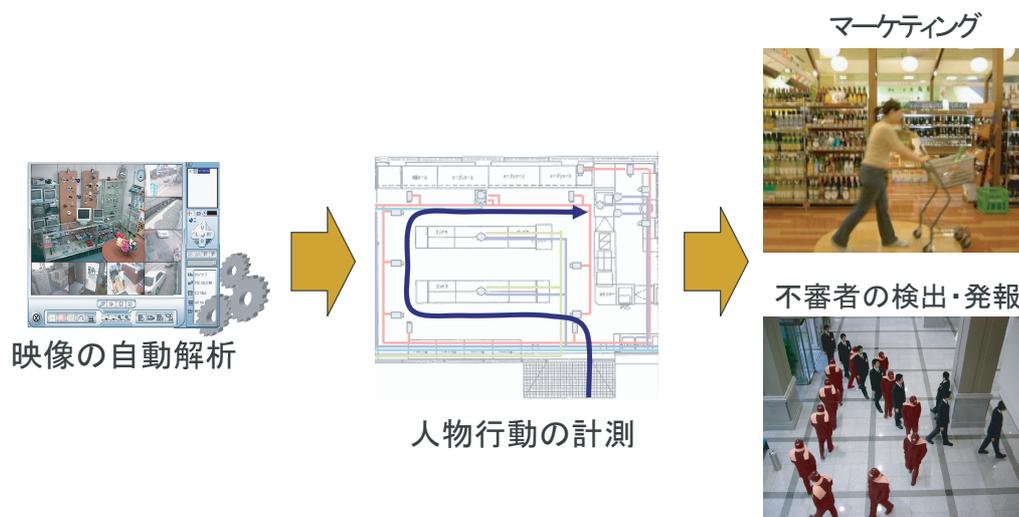


図 1.2 監視カメラ映像を用いた人物の行動計測

1.2 本研究の目的

本研究では，監視カメラシステムから得られる映像を用い，特定のエリア内における人物行動の計測を行うことを目的とする．公共施設や商業施設などで人物行動を計測することを考えると，人物相互や障害物などによる遮蔽や，プライバシーの問題などから，エリア全体で安定した観測を得ることは難しい．そこで，本研究では，エリア内に疎密に分散配置されたセンサ群を用いて，特定のエリア内全体を通して人物行動を計測する技術の確立をめざす．この実現には観測可能な領域内における人物の追跡技術の確立と，観測できない領域を含む人物行動の計測技術が必要となる．

具体的には，本研究ではスーパーやコンビニエンスストアなどの数人～20人程度の人物が同時に存在する空間に着目する．そして，監視カメラなどの画像センサで観測可能な領域については，人物頭部の三次元位置及び向きを実時間で計測する手法を開発する(図 1.3(a))．また，観測できない領域を含む特定エリア全体においては，人物動線を計測することをめざす(図 1.3(b))．

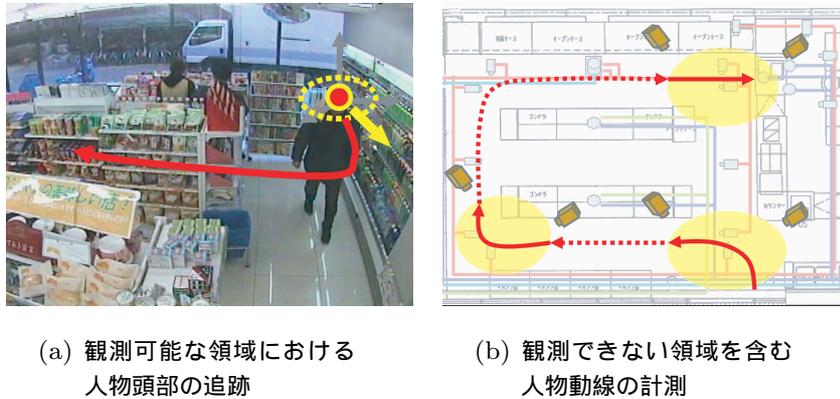


図 1.3 本研究がめざすエリア内人物行動計測

1.3 本研究の課題

監視カメラなどの画像センサによる人物追跡では、人物が必ずしもカメラに対して一定方向を向いていないために起こる見えの変化の問題や、広い範囲を観察することにより頭部が低解像度で観察される問題、昼間と夜間などの照明変動の問題、さらに店舗内の商品棚などの複雑な背景の問題などの課題がある。

また、実際のスーパーやコンビニエンスストアなどへの適用を考えると、監視カメラなどのセンサ配置には物理的な制約があることから、障害物や人物相互による遮蔽により、センサによる観測が得られない領域が生じることが予想される。このような状況を検知し、速やかに追跡を再開するため、人物の出現や消失の頑健な検出も重要な課題である (図 1.4)。



図 1.4 人物頭部追跡の難しさ

さらに，監視カメラで観測できない領域を含む特定エリア内全体における人物動線の計測では，以下の課題が予想される（図 1.5）．

1) 複数カメラの設置状況の推定

カメラの設置位置や姿勢が未知の場合，実世界と得られた観測データの対応付けが困難となる．そこで，実世界での位置やダイナミクスが既知の観測対象から得られる情報を基に，カメラの配置やトポロジなどを推定する．

2) 複数観測領域間の人物動線の対応付け

複数の人物がエリア内を移動する場合，複数のカメラで観察された人物が同一人物であるか検討し，観測データ同士を対応付ける必要がある．そこで，人物が一方のカメラで観察された後に，他方のカメラでどのように観察されるかを手掛りとして人物の同一性を推定する．

3) 観測領域外を含む人物動線の推定

一本の連続した動線としてエリア内の人物動線を得るためには，複数の観測領域で計測された動線を補間する必要がある．そこで，カメラで観察されない領域における人物動線を，観察された少ない情報から補間し，カメラで観察できない領域を含む尤もらしい人物動線を推定する．

ただし，1) の複数カメラの設置状況は，事前の計測とキャリブレーションにより獲得できる場合も多い．そのため，本論文ではカメラの設置位置や姿勢は既知として議論を進め，2)，3) の課題を解決することにより，特定エリア内における人物動線の計測を可能とすることをめざす．

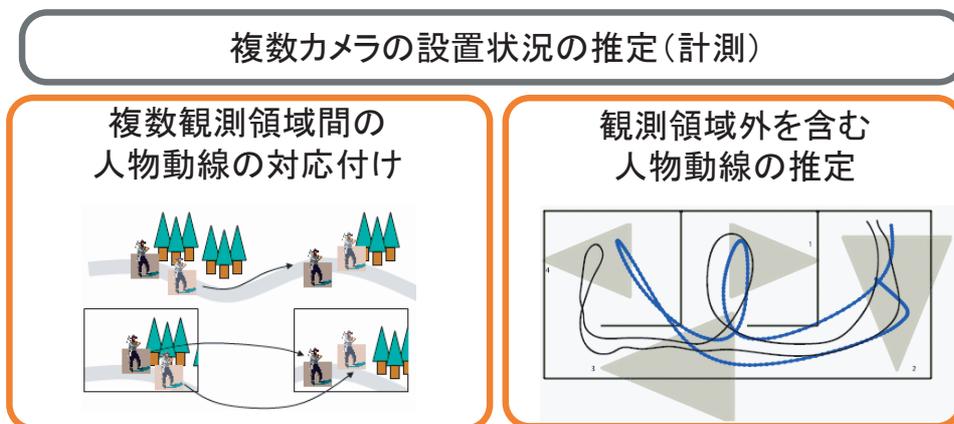


図 1.5 観測できない領域を含む人物動線計測の課題

1.4 本研究の位置づけ

人物センシングでは、動作の計測によりジェスチャの認識を行う例 [70, 107] や、顔や動作の解析から表情や行動の認識をめざす取組み [2, 85] など報告されているが、本研究では、人物センシングの基盤技術となる人物の検出と追跡に着目する。

人物センシングに用いられるセンサを大別すると、人体に計測機器を装着する接触型センサと環境に計測機器を埋め込む非接触型センサがある。接触型センサでは、マーカやデバイスを人体に装着することで、人物の位置や動きなどの計測を行う^{*1}が、公共施設や商業施設などでは、訪れた人物にセンサをまねなく装着させることは難しい。一方、監視カメラなど環境に計測機器を埋め込む非接触型センサでは、人物にセンサを装着しないことから、本研究で着目する不特定人物の追跡に適している。

画像センサを用いる例では、監視カメラなどの一般的なカメラのほかに、全方位カメラを用いる例 [47, 97] や、能動カメラを用いる例 [110, 111] などがある。少ないカメラ台数で広い観測領域をカバーしたい場合や、特定箇所を詳細に観測したい場合などは、全方位カメラや能動カメラの利用は有効であると考えられるが、本研究では、既設監視カメラの有効活用という観点から、これらの画像センサは用いない。

画像センサ以外では、オープンスペースでの人物追跡にレーザ測域センサを用いる手法 [17, 62] が提案され、有効性が報告されている^{*2}。監視カメラに代表される環境に埋め込む非接触型センサでは、センサの配置が課題となるが、本研究では、カメラとレーザ測域センサを相補的に用いることで、この問題を低減することをめざす。

同時計測人数に着目すると、一人を対象とした計測手法の報告が非常に多い。また、数人～20人の計測を行う例では、スポーツ映像などの解析で多くの例が報告されている (例えば [12, 91] など)。さらに、多人数を対象とした例では、駅のコンコースなどで、数十人の群集を計測する取り組みも報告されている [11]。本研究では、同時追跡人数を数人～20人程度として、人物頭部の三次元位置と向きを高精度に追跡することをめざす。

計測対象空間の大きさに着目すると、自動車内でのドライバの計測など限定された空間を対象とする例 [1] もあるが、多くは室内などの比較的人物が自由に行動できる空間を対象とし、視野を共有した複数のカメラで人物を計測する例が多い (例えば [56, 61, 93, 102] など)。また、最近では、視野を共有しないカメラ群を用いてより広い領域を観測する試みも報告されている (例えば [39, 83] など)。本研究では、スーパーやコンビニエンスストアなどに代表される大きさの空間に着目し、遮蔽などの観測不能状態を織り込んだアプローチをとることで、実用的な人物動線の計測手法を確立する。

^{*1} タグを用いたシステムでは Active Badge [33] などがよく知られている。

^{*2} このほかにも、人感センサ [32] を用いる例や、床圧力センサ [79] を用いる例などが報告されているが、精度や設置工事の簡素化など課題も多い。

1.5 本論文の構成

本論文の構成は以下の通りである．

- 1 章 序論
- 2 章 時系列フィルタと識別器の統合による三次元人物頭部追跡
- 3 章 カメラとレーザ測域センサの統合によるエリア内複数人物追跡
- 4 章 複数観測領域間の人物動線の対応付け
- 5 章 観測領域外を含む人物動線の推定
- 6 章 結論

2 章では，監視カメラ映像を想定し，映像中の人物頭部の位置と向きを追跡を行う手法について詳しく述べる．3 章では，カメラとレーザ測域センサを相補的に用いることで，さまざまなセンサ配置を可能とした，より実用的な人物追跡手法について述べる．4 章では，複数の観測領域で観測される人物の同一性を推定することで人物動線の対応付けを行う手法について述べ，5 章では，観測領域外を含む人物動線の推定手法について述べる．最後に 6 章にて，全体を総括し，残された課題について議論する．

第2章

時系列フィルタと識別器の統合による三次元人物頭部追跡

2.1 はじめに

監視カメラの小型化や低価格化などに伴い、監視カメラの普及が著しい。一方で、大量の監視カメラで撮影された映像を、常に人手で監視することはもはや現実的ではなく、多くの場合、監視カメラ映像は、事後の参照のために一定期間保存された後、破棄されている。これに対して、監視カメラ映像の自動解析により、人物行動の計測が可能となれば、不審者の検出・発報や、POS データでは分からない顧客の購買行動などの情報を抽出可能となり、セキュリティやマーケティングなどの分野への応用が期待できる。このような監視カメラ映像の事後利用から事前利用への転換は、監視カメラ利用における一つのパラダイムシフトであり、監視カメラ映像による人物追跡技術は、監視カメラ市場において、製品の競争力を大きく向上させる基盤技術である。

監視カメラ映像を用いた人物追跡では、人物が必ずしもカメラに対して一定方向を向いていないために起こる見えの変化の問題や、広範囲の観察により人物が低解像度で観察される問題、昼夜の照明変動や複雑な背景の問題などの課題がある (図 2.1)。

カメラ画像を用いた対象追跡では、これまでに多くの手法が提案されている^{*1}が、なかでも、パーティクルフィルタ^{*2}の有効性が数多く報告されている ([7, 14, 37, 49, 55, 56, 63, 65, 72, 84] など)。パーティクルフィルタは、状態量と尤度を持つ多数の仮説群により離散的な確率密度として追跡対象を表現し、状態遷移モデルを用いて伝播させることで、動きの変動や観測の雑音に対して頑健な追跡を実現する手法である。特に、観測値が非ガウス型となる場合においても頑健な追跡を実現できるとして注目されている。

パーティクルフィルタによる人物追跡では、仮説の人物らしさをカメラ画像を用いて評価するが、これにはカラーヒストグラムや輪郭の輝度変化の類似性などが用いられることが多い ([7, 37, 49, 56, 65, 72, 84] など)。しかし、このような比較的単純な評価手法は、照明変動や複雑な背景下における人物の追跡では必ずしも十分ではなく、追跡性能の向上には頑健かつ高精度な評価手法が求められる。

^{*1} 対象追跡については文献 [44] が参考になる。

^{*2} Sequential Monte Carlo, Condensation などとも呼ばれる。



図 2.1 画像を用いた人物追跡の難しさ

一方で 静止画像などから人物の顔を検出する手法が数多く提案されている．なかでも AdaBoost 学習による識別器を用いた顔検出手法がよく知られており，特に，矩形特徴を用いた識別器による顔検出手法は，照明変動や複雑背景下での観察にも頑健であり，検出精度が高いことから，広く利用されるようになってきている．AdaBoost 学習による識別器では多数の弱識別器を線形結合することで高精度な識別器を構成するが，Viola と Jones はこれをカスケード型とし，学習，検出時に用いる矩形特徴を高速に計算する手法 [105] を提案している．このカスケード型識別器は，非検出対象がカスケードの初期に棄却されるため，単一の識別対象に関して，より高速な処理が可能である．

静止画像からの顔検出では，この識別器を画像全体を網羅するように識別対象の大きさや位置をさまざまに変化させて適用するが，時系列画像を用いて対象を追跡する枠組みでは，画面全体を逐次探索することは効率的ではない．そのため，パーティクルフィルタの仮説の評価に識別器を応用し，探索範囲を限定することは有効であると考えられる．

このような考えに基づいて，本章では，視野を共有した複数のカメラを用い，状態量として三次元位置と向きを持つ人物頭部を三次元空間内で追跡する手法を提案する．

特に新規性の高い本手法の特徴を以下に挙げる．

- 1) パーティクルフィルタを用いた追跡の枠組みにおいて，仮説の評価に矩形特徴を用いた AdaBoost 学習による識別器を応用する．
- 2) 人物頭部は実際の向きやカメラとの位置関係により，その見えが変化するため，人物頭部の方向毎に学習させた複数の識別器を準備し，さまざまな方向で観察される人物頭部に対し，識別器を適応的に選択して評価する．

本章では，2.2 節にて関連研究について概観する．2.3 節では，パーティクルフィルタについて述べ，2.4 節では，AdaBoost 学習による識別器について述べる．そして，2.5 節にて時系列フィルタと識別器の統合による三次元人物頭部追跡について詳細に述べる．2.6 節では検証実験について述べ，2.7 節にて考察を行う．最後に，2.8 節にてまとめる．

2.2 関連研究

カメラ画像を用いた対象追跡では、フレーム間の類似度に着目し、その勾配から対象の移動先を推定する手法と、対象の動きをモデル化し、移動先を予測する手法がある。前者の代表的なアルゴリズムとして MeanShift[16] が挙げられ、後者の代表としてカルマンフィルタ [43] やパーティクルフィルタ [37] がある。姿勢により大きく見えの異なる対象や、対象の三次元姿勢を含めた追跡を行う場合には、対象の動き予測を行う時系列フィルタが適している。

識別器を用いた追跡の枠組みでは、Support Vector Tracking(SVT)[4] がよく知られている。SVT では画面上で識別器の評価値を最大にする探索を行うが、姿勢変動を伴う三次元物体追跡では、姿勢による見えの変化を1つの画面上での評価値の勾配によって追跡することは難しい。また、パーティクルフィルタを用いて対象の見えを詳細に検討することで仮説の評価を行う例として、疎テンプレートマッチングを用いて仮説を評価する手法 [55] があるが、これは対象の解像度が低い場合に適した方法ではない。

一方、本研究で提案する AdaBoost 学習による識別器をパーティクルフィルタによる追跡に用いる例は少ない。

パーティクルフィルタによる人物追跡に AdaBoost 学習による識別器を併用する例を図 2.2 に示す。AdaBoost 学習による識別器をパーティクルフィルタと併用した例では、Okuma らの手法 [67] がよく知られている。Okuma らは追跡対象の検出に AdaBoost 学習による識別器を用い、仮説の生成に検出結果を利用している。しかし、識別器は、追跡ではなく、従来からの利用法である検出に用いられるに留まり、識別器を仮説の評価に用いる本手法とは、アプローチが大きく異なる。また、識別器を画面全体に逐次適用するため、計算コストが高く、実時間での追跡には至っていない。Yang ら [113] は、Coarse-to-Fine により段階的に行う仮説の評価の枝刈りに AdaBoost 学習による識別器を用いているが、2値分類器としての補助的な利用に留まっている。また、本研究と同時期に、Chateau らは仮説の評価に AdaBoost 学習による識別器を利用する手法 [14] を提案している。しかし、単一の識別器を用いた画面上での追跡に留まっており、姿勢による見えの違いを考慮して三次元空間内で対象を追跡するものではない。

三次元空間内で追跡を行う例では、複数のカメラによる仮説の評価を統合して人物を追跡する手法 [56, 106] や、さらに環境モデルを併用して頑健な追跡を行う手法 [93] が提案されているが、各カメラにおける仮説の評価には、比較的単純な手法が用いられるに留まっている。Nickel ら [63] は、複数のカメラを用いた仮説の評価に AdaBoost 学習による識別器を利用しているが、識別器による検出結果の二次的な利用であり、対象の向きとカメラとの関係は考慮していない。

複数の AdaBoost 学習による識別器を一つの対象に併用した例では、複数の AdaBoost 学習による識別器の出力をニューラルネットワークを用いて統合することで、低解像度画像での検出性能を向上させる手法 [26] が提案されているが、対象の追跡を目的としたものではない。

これらに対して本手法では、視野を共有した複数のカメラを用い、三次元空間内で人物頭部を頑健に追跡する。具体的には、パーティクルフィルタの仮説の評価に AdaBoost 学習による識別器を応用することで、頑健かつ高精度に人物頭部を追跡する。また、人物の向きとカメラとの関係に基づいて、識別器を適応的に選択して用いることで、高精度な追跡と同時に人物頭部の向きを推定する。

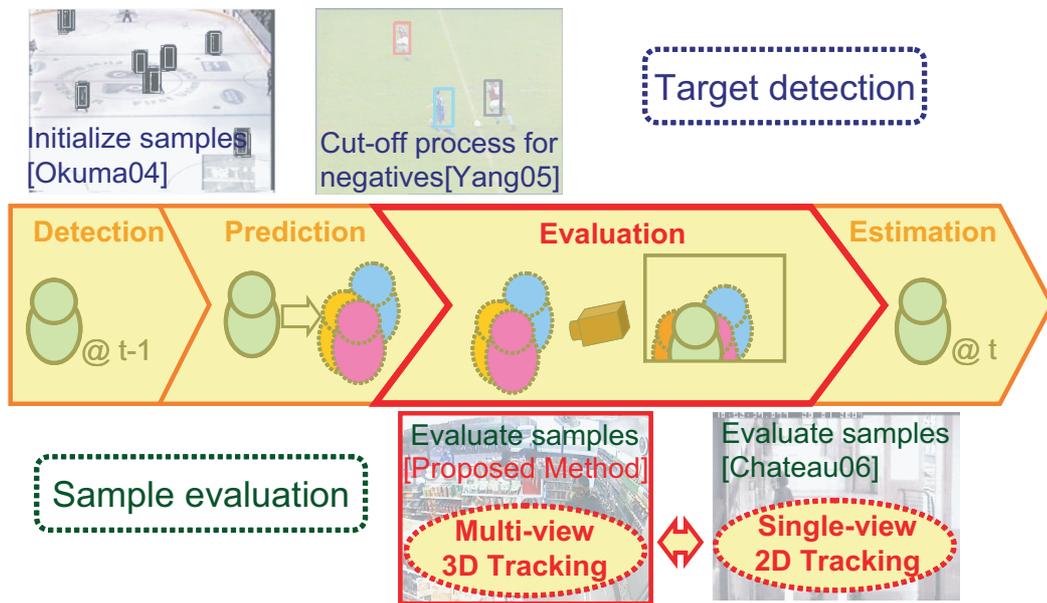


図 2.2 識別器とパーティクルフィルタの併用による人物追跡

一方、人物頭部の向きを推定する例では、顔の構造に基づいたモデルを用いるものが多い。頭部領域重心と肌色領域重心との水平距離に基づく簡易なモデルで顔の回転角を推定し、正面顔画像を獲得する能動的カメラシステム [23] や、目や口の両端などを特徴点とする顔モデルを用い、頭部姿勢を追跡する手法 [58, 66] などが報告されている。特に、目や口の両端などを特徴点とするモデルベースの手法では高精度な推定結果が報告されているが、顔部品を観察する必要から、人物頭部が高解像度で観察される必要があり、追跡可能な顔向き範囲にも制限がある。また、追跡の開始のためには、正面顔の観察などの初期キャリブレーションを必要とする場合が多い。

本研究で着目する監視カメラ映像では、人物がさまざまな位置、姿勢で視野内に出現することが想定され、初期キャリブレーションを必要としない、アピランススペースの手法 (例えば [59] など) が有効であると考えられる。これには、高次局所自己相関特徴を用いる手法 [116] や、四方向面特徴に基づいた顔向きごとの識別器の出力を統合して推定する手法 [115] などがあるが、いずれも比較的高解像度で人物頭部が観察される場合に限られる。また、顔認証の分野では、顔向きに依存せず識別を行うため、カメラに対する顔向きを副次的に推定するものもあるが、低解像度での頭部の観察は想定されていない [6, 48]。

これらに対して、低解像度で観察される頭部の向きを追跡と同時に推定する例は少ない。Robertsonら [76] は、20~40ピクセル程度で観察される頭部の向きを、追跡対象領域の肌色画素の量と対象の移動方向の学習結果に基づいて8方向に分類しているが、複数カメラによる三次元追跡の枠組みで試みられたものではなく、本研究で用いている識別器などの高度な画像特徴量を用いたものではない。

2.3 パーティクルフィルタ

本節では、パーティクルフィルタ [37] について、アルゴリズムの概略を説明する。

パーティクルフィルタは逐次ベイズ推定の理論に基づいている。時刻 t における対象の状態量を \mathbf{x}_t 、画像による観測を \mathbf{z}_t とし、時刻 t までに得られる観測を $\mathbf{Z}_t = \{\mathbf{z}_1, \dots, \mathbf{z}_t\}$ とする。このとき、時刻 t における対象の事前確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_{t-1})$ は、マルコフ過程を仮定することにより、時刻 $t-1$ における事後確率密度分布 $p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$ と時刻 $t-1$ から t への状態遷移確率密度分布 $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ を用いて以下のように表すことができる。

$$p(\mathbf{x}_t | \mathbf{Z}_{t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1})p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})d\mathbf{x}_{t-1} \quad (2.1)$$

ここで、 $p(\mathbf{z}_t | \mathbf{Z}_{t-1})$ を一定とすると、時刻 t における事後確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_t)$ は、ベイズの法則より、時刻 t における尤度 $p(\mathbf{z}_t | \mathbf{x}_t)$ と事前確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_{t-1})$ により次式のように表すことができる。

$$p(\mathbf{x}_t | \mathbf{Z}_t) \propto p(\mathbf{z}_t | \mathbf{x}_t)p(\mathbf{x}_t | \mathbf{Z}_{t-1}) \quad (2.2)$$

逐次ベイズ推定では、この事後確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_t)$ の期待値を逐次求めることで、対象の追跡を実現する。

パーティクルフィルタでは、時刻 t における事後確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_t)$ を、状態量 \mathbf{x}_t の仮説群 $\{\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(N)}\}$ と各仮説に対応する重み $\{\pi_t^{(1)}, \dots, \pi_t^{(N)}\}$ により離散的に近似し、次のプロセスを経て、逐次的に更新する。

1) 仮説の選択

時刻 $t-1$ における事後確率密度分布 $p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$ を離散的に近似した N 個の仮説 $\{\mathbf{s}_{t-1}^{(1)}, \dots, \mathbf{s}_{t-1}^{(N)}\}$ の重み $\{\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}\}$ の比に従い、仮説群 $\{\mathbf{s}_{t-1}^{(1)}, \dots, \mathbf{s}_{t-1}^{(N)}\}$ を選択する。

2) 状態遷移確率密度分布に基づく伝播

選択された仮説群 $\{\mathbf{s}_{t-1}^{(1)}, \dots, \mathbf{s}_{t-1}^{(N)}\}$ を、状態遷移確率密度分布 $p(\mathbf{x}_t | \mathbf{x}_{t-1} = \mathbf{s}_{t-1}^{(n)})$ に従い伝播し、 $p(\mathbf{x}_t | \mathbf{Z}_{t-1})$ に相当する時刻 t における N 個の仮説群 $\{\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(N)}\}$ を生成する。

3) 画像による重み $\pi_t^{(n)}$ の推定

画像から尤度の評価を行うことで、仮説群 $\{\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(N)}\}$ の重み $\pi_t^{(n)} \approx p(\mathbf{z}_t | \mathbf{x}_t = \mathbf{s}_t^{(n)})$ を推定する。

パーティクルフィルタにおける仮説の更新の様子を図 2.3 に示す。

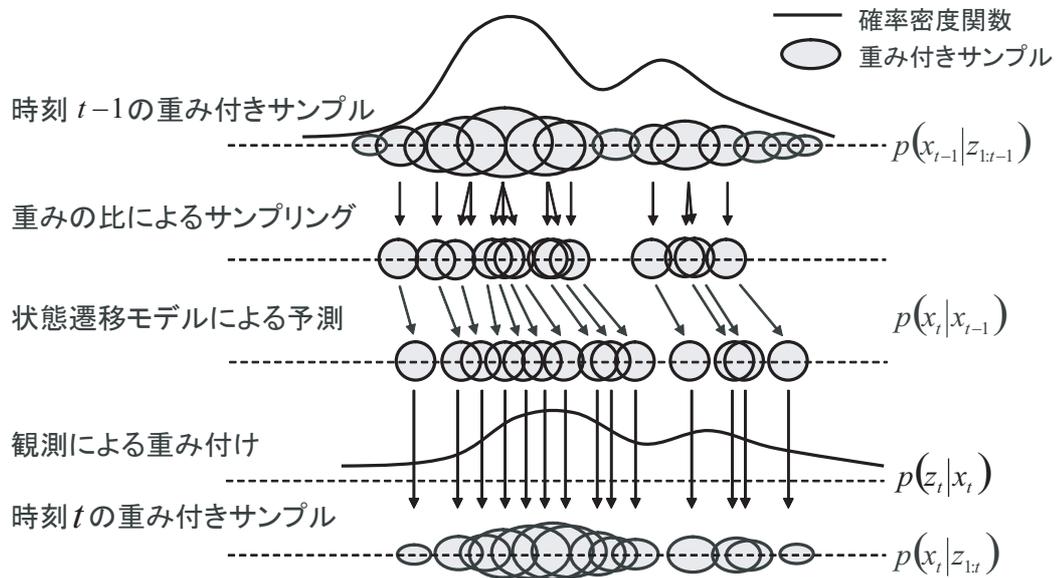


図 2.3 パーティクルフィルタにおける仮説の更新

パーティクルフィルタの追跡性能は、仮説数を同じとすれば、状態遷移確率密度分布による仮説の伝播と、観測による仮説の重み付けに大きく依存する。そのため、頑健かつ高精度な追跡の実現のためには、追跡対象の動きを精度良く予測する状態遷移モデルの利用と、追跡対象の尤度を精度良く推定する評価手法の開発が重要となる。これに対して、本研究では、追跡対象の動き予測にオプティカルフローに基づいた手法を用い、追跡対象の尤度の推定に AdaBoost 学習による識別器を応用した手法を用いることで、頑健かつ高精度な人物頭部の追跡を行う。

次節では、本研究で新たに提案する追跡手法の議論に向けた準備として、AdaBoost 学習による識別器について述べる。

2.4 AdaBoost 学習による識別器

AdaBoost 学習は、複数の弱識別器出力の多数決に基づいて判別を行う識別器を構築するための学習アルゴリズムである。本節では AdaBoost 学習による識別器について、概略を説明する。

一般に、複数の識別器の結合により判別を行う識別器の学習手法をアンサンブル学習^{*3}と呼ぶ。アンサンブル学習は、複数の識別器の結合によって、単一の識別器では達成できない高い性能を持つ識別器を構築する手法である。特に、バギング法^{*4} やブースティング法 [20] がよく知られている。

2.4.1 ブースティング法

ブースティング法は、学習データに重みを導入し、逐次的に識別器を学習させることで識別性能を向上させる手法である (図 2.4)。ブースティング法では、以前の学習によって得た識別器が誤って判別した学習データに対して、指数的に大きな重みを付与する。これにより、次の識別器の学習では、先の識別器が誤った重みの大きな学習データに対して、正しい判別を行うように識別器の学習が行われる。

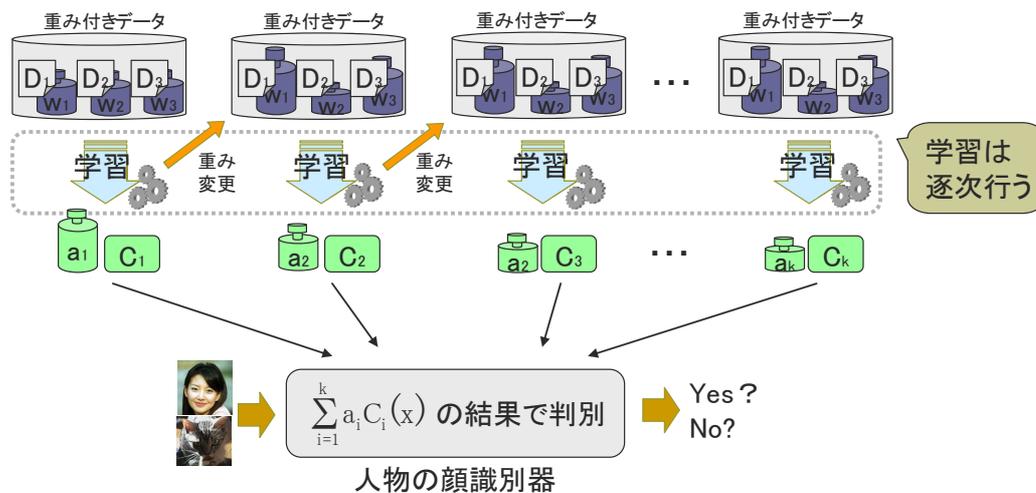


図 2.4 ブースティング法による顔識別器の学習

^{*3} 最も簡単なアンサンブル学習は、単純に複数の識別器を個別に学習し、新たな入力に対する識別はそれらの識別器の多数決をとるものである。直観的には、次の例でアンサンブルの効果を知ることができる。K 個の識別器が統計的に独立であり、識別器の誤り確率を一律の θ とすると K 個の識別器のうち k 個の識別器の判定が誤る確率密度分布 $p(k)$ は 2 項分布となり $p(k) = {}_K C_k \theta^k (1 - \theta)^{K-k}$ と表すことができる。例えば、 $K = 21$, $\theta = 0.3$ とすると、アンサンブルによりクラス判定が誤る確率は $P(11) + \dots + P(21) \approx 0.018$ であり、 $\theta = 0.3$ に比べてはるかに小さい。このように、アンサンブル学習により得られた識別器は、一つの識別器を独立に用いた場合に比べ、はるかに高い識別性能を示す。アンサンブル学習についての詳細は、文献 [3, 75, 101]などを参考されたい。

^{*4} バギング法 [10](bagging は bootstrap paggregating の略) は、学習データから復元抽出した K 組の学習データを用い、識別器を K 通りの方法で独立に学習させ、得られた K 個の識別器の多数決で判別を行う識別器の学習手法である。ニューラルネットのような自由度の高い識別器をアンサンブルする場合に、効果が大きいとされる手法である [25]。

ブースティング法の代表的なアルゴリズムとして AdaBoost^{*5}アルゴリズムがある．AdaBoostアルゴリズムは，単独では識別性能が低い弱識別器でも，多数を組合せることにより，高い識別性能を示すことから，近年，特に注目が集まっている．以下に AdaBoost アルゴリズム [20] の学習手順を示す．

AdaBoost アルゴリズム

Step 1. [初期化]

N 個の学習データ $(x_1, y_1), \dots, (x_N, y_N)$, $y_n \in +1, -1$ を準備し，重みを一様に初期化する．

$$\omega_{t=1,n} = \frac{1}{N} \quad (2.3)$$

Step 2. [ブースティング]

For $t = 1 \dots T$:

Step 2.1

弱識別器 $h_j(x)$ を適用し重みつき識別誤り率 ϵ_j を計算する． $j = 1 \dots J$ で誤り率が最小の弱識別器 $h_t(x)$ を選択する．

$$\epsilon_j = \sum_{n=1}^N \omega_{j,n} I(y_n \neq h_j(x_n)) \quad (2.4)$$

ただし， $I(u)$ は u が真 (偽) のとき， $I(u) = 1(0)$ をとる．

Step 2.2

選択された弱識別器 $h_t(x)$ の重み α_t を計算する．

$$\alpha_t = \frac{1}{2} \log \frac{1 - \epsilon_t}{\epsilon_t} \quad (2.5)$$

Step 2.3

学習データの重みを更新する．

$$\omega_{t+1,n} = \frac{\omega_{t,n} \exp(-\alpha_t y_n h_t(x_n))}{\sum_{n=1}^N \omega_{t,n} \exp(-\alpha_t y_n h_t(x_n))} \quad (2.6)$$

Step 3. [出力]

得られた T 個の弱識別器を用いてアンサンブル識別器を構成する．

$$H(x) = \text{sgn} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad (2.7)$$

*5 AdaBoost の Ada は adaptive の意味

2.4.2 Viola と Jones の顔識別器

Viola と Jones により提案された顔識別器 [105] は、検出時間の短縮のため、複数の AdaBoost 識別器を組み合わせたカスケード構造 (図 2.5) をなしている。図 2.5 において H_i は識別器を表す。入力画像に対し、各段で顔、非顔の判定を行い、顔と判定された画像だけが次の段へ進む。最後の段まで通過したものが顔と判定される。

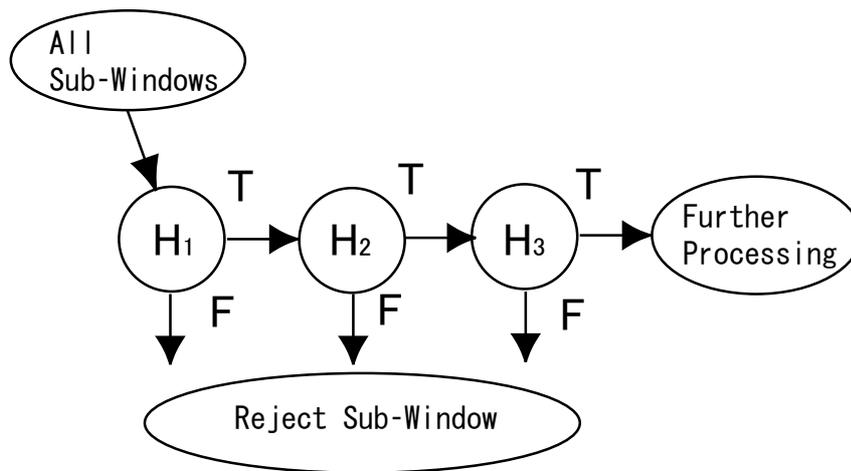


図 2.5 カスケード型識別器

カスケードの各段において、学習用顔画像を通過させる割合 (学習用顔画像通過率) を Dr ($0 < Dr < 1$)、学習用非顔画像を通過させる割合 (学習用非顔画像通過率) を Fp ($0 < Fp < 1$) とすると、カスケード n 段通過後は、学習用顔画像は Dr^n 、学習用非顔画像は Fp^n だけ通過していることになり、例えば $Dr = 0.999$ 、 $Fp = 0.5$ とした場合、 $n = 40$ のカスケード型識別器は学習用顔画像通過率は $0.999^{40} \approx 0.96$ 、学習用非顔画像通過率は $0.5^{40} \approx 9.1 \times 10^{-13}$ となり、学習用顔画像をほとんど通過させ、学習用非顔画像をほとんど通過させない顔識別器となる。

カスケードの各段を構成する識別器 H_i は、図 2.6 のような特徴を用いて評価を行う多数の弱識別器 $h_{i,t}(x)$ の線形結合により、以下のように表される。

$$H_i(x) = \text{sgn} \left(\sum_{t=1}^{T_i} \alpha_{i,t} h_{i,t}(x) \right). \quad (2.8)$$

ここで、 T_i は i 段目で用いられる弱識別器の数であり、 $\alpha_{i,t}$ は学習時に決まる弱識別器の重みである。

弱識別器 $h_{i,t}(x)$ は、それぞれの特徴の白矩形内の合計画素値と黒矩形内の合計画素値の差分に対し、しきい値との大きさを比較して +1 又は -1 を出力する。矩形特徴の位置と大きさを画像内でどのようにとるかによって膨大な種類の特徴が存在する^{*6}が、これらの中から学習用顔画像を良く識別する特徴が AdaBoost アルゴリズムにより選択され、各段の識別器が準備される。

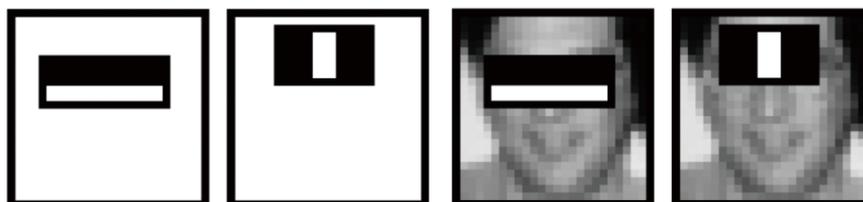


図 2.6 弱識別器で用いる矩形特徴

本研究で用いるすべての識別器は、この矩形特徴に基づいて学習を行ったものである。

次節では、2.3 節で述べたパーティクルフィルタによる追跡の枠組みに、本節で述べた AdaBoost 学習による識別器を統合することで、人物頭部を頑健に追跡する手法について述べる。

^{*6} 24×24 ピクセルの画像領域ではおよそ 12 万種の特徴を考えることができる。

2.5 時系列フィルタと識別器の統合による 三次元人物頭部追跡

本節では、時系列フィルタと識別器の統合による三次元人物頭部追跡について詳細に述べる。本手法では、視野を共有した複数のカメラを用いて、三次元位置と向きを状態量に持つ人物頭部を、パーティクルフィルタにより追跡する。本節で述べる提案手法は以下の特徴を持つ。

- 1) パーティクルフィルタの仮説の伝播にオプティカルフローに基づいた状態遷移モデルを適用する。
- 2) パーティクルフィルタの仮説の評価に AdaBoost 学習による識別器を応用する。
- 3) 人物頭部は実際の向きやカメラとの位置関係により、その見えが変化するため、人物頭部の各方向毎に学習を行った複数の識別器を準備し、さまざまな方向で観察される人物頭部に対し、識別器を適応的に選択して用いる。
- 4) 状態量に三次元位置と向きを持つ人物頭部を、視野を共有した複数のカメラを用いて観察し、三次元空間内で人物を追跡する。

2.5.1 人物頭部モデル

室内空間に床面を XY 平面と一致させ、高さ方向を Z 軸とした三次元世界座標系 XYZ をとる。人物頭部モデルとして、中心座標が (x, y, z) である一定の大きさの楕円体を仮定する。人物は頭部を傾けて室内を移動することは少ないと仮定すると、人物頭部の向きは Z 軸回りの回転 θ のみで表せる。これより、人物頭部の時刻 t における i 番目の仮説を以下で表す。

$$\mathbf{s}_t^{(i)} = [x_t^{(i)} \quad y_t^{(i)} \quad z_t^{(i)} \quad \theta_t^{(i)}]^\top \quad (2.9)$$

また、この様子を図 2.7 に示す。

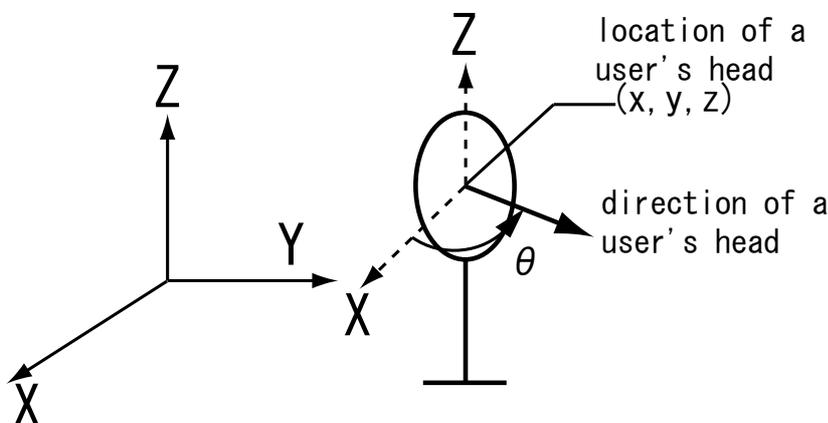


図 2.7 人物頭部モデル

2.5.2 状態遷移モデルに基づく仮説の伝播

状態遷移確率 $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ として、画面上での追跡対象領域のオプティカルフロー [34] に基づいた状態遷移モデルを仮定する。

オプティカルフローには、映像の時空間微分の拘束条件からフローベクトルを推定する勾配法 [50] と、画像中のある大きさの領域をテンプレートブロックとし、次フレームの画像中を正規化相関に基づいて探索するブロックマッチング法がある。勾配法とブロックマッチング法は、以下の特徴を持つ。

表 2.1 勾配法とブロックマッチング法

手法	利点	欠点
勾配法	計算コストが低い	雑音や姿勢変動などによる急激な輝度変化に弱い
ブロックマッチング法	雑音や姿勢変動などによる急激な輝度変化に強い	計算コストが高い

本研究で着目する監視カメラ映像では、画像全体に占める人物頭部領域が比較的小さく、正規化相関による探索領域を頭部周辺に限定することで、計算コストを抑えられることから、雑音や姿勢変動などによる急激な輝度値の変化に強いブロックマッチング法を用いる。

本手法では、画面上での追跡対象領域から一定の大きさの 16 個のグリッドを抽出し、それぞれに対してブロックマッチングアルゴリズムにより求めたオプティカルフローの平均ベクトルを追跡対象領域の画面上での変位として用いる (図 2.8)。追跡対象領域をグリッドに分割することにより、頭部姿勢の変動などにより、一部に大きな見えの変化が生じてても、安定したフローベクトルを得ることができ、追跡対象領域サイズが大きい場合にも、テンプレートブロックが一定の大きさであるため、計算コストが増加しない。また、三次元空間における追跡対象の変位ベクトルは、複数のカメラ画像上で観察された変位に基づいて移動した画像座標を三次元空間に逆投影し、時刻 $t-1$ で推定された三次元位置と差分をとることで得られる。

このような考えに基づき、仮説 $\mathbf{s}_{t-1}^{(n)} = \begin{bmatrix} x_{t-1}^{(n)} & y_{t-1}^{(n)} & z_{t-1}^{(n)} & \theta_{t-1}^{(n)} \end{bmatrix}^\top$ を $\mathbf{s}_t^{(n)}$ に次のように伝播する。

$$\mathbf{s}_t^{(n)} = \mathbf{s}_{t-1}^{(n)} + \mathbf{v}_t + \boldsymbol{\omega}. \quad (2.10)$$

ここで、 \mathbf{v}_t はオプティカルフローから予測された三次元空間における変位ベクトルであり、 $\mathbf{v}_t = \begin{bmatrix} \hat{x}_t & \hat{y}_t & \hat{z}_t & \hat{\theta}_t \end{bmatrix}^\top$ である (ただし $\hat{\theta}_t = 0$)。これは、時刻 t において、2つのカメラの画像上で、人物頭部周辺のオプティカルフローにより予測された画像座標 $(u_{cam1,t}, v_{cam1,t}), (u_{cam2,t}, v_{cam2,t})$ を三次元空間に逆投影し、時刻 $t-1$ で推定された人物頭部の三次元位置と差分をとることで得る。 $\boldsymbol{\omega}$ は平均 0、共分散行列 Σ_ω を持つガウス雑音であり、 Σ_ω は分散 $\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_\theta^2$ を対角要素に持つ。ただし、視野を共有するカメラがない場合は、 $\hat{x}_t = 0, \hat{y}_t = 0, \hat{z}_t = 0, \hat{\theta}_t = 0$ とする。

次に、このように伝播した仮説 $\mathbf{s}_t^{(n)}$ のカメラ画像への射影を考える。

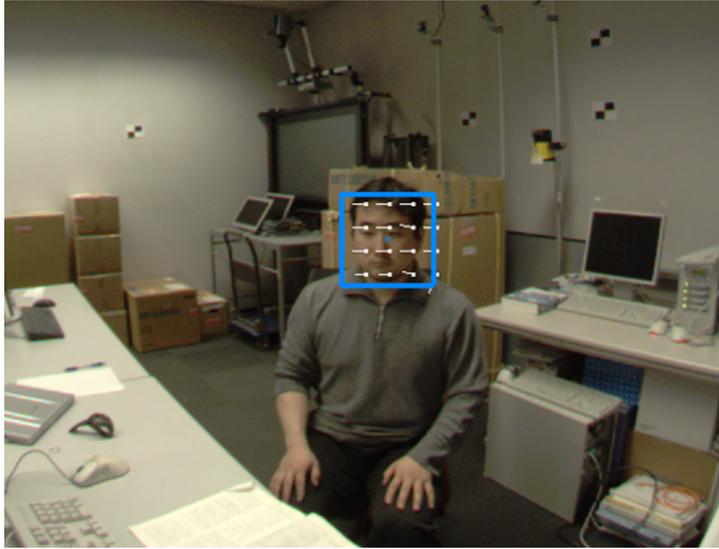


図 2.8 人物頭部周辺のオプティカルフロー

2.5.3 カメラ画像への射影

時刻 t の n 番目の仮説 $\mathbf{s}_t^{(n)} = \begin{bmatrix} x_t^{(n)} & y_t^{(n)} & z_t^{(n)} & \theta_t^{(n)} \end{bmatrix}^\top$ は、校正済みの i 番目のカメラ画像への射影関数を F_i とすると、以下のように射影することができる。

$$\mathbf{p}_{i,t}^{(n)} = F_i \left(\mathbf{s}_t^{(n)} \right). \quad (2.11)$$

ここで $\mathbf{p}_{i,t}^{(n)}$ は、仮説 $\mathbf{s}_t^{(n)}$ の三次元位置を i 番目のカメラ画像へ射影したものである。

このとき、 i 番目のカメラに観察される相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ は以下のように表される。

$$\theta_{i,t}^{(n)} = \theta_t^{(n)} - \tan^{-1} \left(\frac{\left[\mathbf{C}_i - \mathbf{K} \mathbf{s}_t^{(n)} \right]^y}{\left[\mathbf{C}_i - \mathbf{K} \mathbf{s}_t^{(n)} \right]^x} \right). \quad (2.12)$$

ここで \mathbf{C}_i は i 番目のカメラの XY 座標であり、 \mathbf{K} は仮説 $\mathbf{s}_t^{(n)}$ から XY 位置成分を取り出す行列である。 $[\]^x$ は X 軸に対応する要素を取り出すことを表す。

最後に、人物頭部モデルの楕円体の幅をカメラ画像へ射影し、 i 番目のカメラで観察される人物頭部の幅 l_i を得る。

このように、人物頭部モデルをカメラ画像へ射影することで、カメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ 、相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ 、人物頭部の幅 l_i を準備する。

次節では、これらを用いた仮説 $\mathbf{s}_t^{(n)}$ の評価について述べる。

2.5.4 仮説の評価手順

各仮説に対応した人物頭部候補領域画像 $g_{i,t}^{(n)}$ の人物頭部らしさを AdaBoost 学習による識別器と輪郭の輝度勾配の類似性を用いて評価する。

一般に、識別器を静止画像に適用して対象検出に利用する場合、画像中の対象周辺に発生する多くの候補領域をマージして最終的に一つの対象として検出する。しかし、本手法では、マージに相当する処理はパーティクルフィルタの枠組みにより提供されるため、識別器による評価は、各仮説に対応した人物頭部候補領域画像 $g_{i,t}^{(n)}$ について独立して考えれば良い。

具体的には、以下の手順により時刻 t における n 番目の仮説 $s_t^{(n)}$ の識別器による評価を行い、重み $\pi_t^{(n)}$ を得る。

- 1) 時刻 t の n 番目の仮説 $s_t^{(n)}$ を i 番目のカメラ画像に射影し、カメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ 、相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ 、カメラ画像上での人物頭部の幅 $l_{i,t}^{(n)}$ を得る。ただし、仮説を射影した際にカメラの視野外となる場合、以降の手順は行わず、処理を終了する。
- 2) カメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ を中心として、人物頭部の幅 l_i に従い人物頭部の上部輪郭上 6 点の輝度勾配ベクトルとモデルの輝度勾配ベクトルの内積をとることで、輪郭輝度勾配の類似性に基づく重み $\pi_{edge,i,t}^{(n)}$ を取得する。図 2.9 の頭部画像に示した点はこの評価点を表している。
- 3) 仮説の射影によって得られた相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ に基づいて、識別器を選択する。例えば、正面、 90° 右向き、 90° 左向きの 3 方向の識別器を用いた場合、相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ が $-45^\circ \sim 45^\circ$ の場合は人物頭部正面の識別器が選択され、 $45^\circ \sim 135^\circ$ の場合は 90° 左向きの識別器が選択され、 $-45^\circ \sim -135^\circ$ の場合は 90° 右向きの識別器が選択される。
- 4) 仮説 $s_t^{(n)}$ を射影したカメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ を中心に、カメラ画像上での人物頭部の幅 $l_{i,t}^{(n)}$ を一辺とする領域を切り出す。このとき、選択された識別器の種類によって切り出す領域を図 2.9 のように調整する*7。
- 5) 切り出した画像のサイズを変更*8し、識別器に入力可能な人物頭部候補領域画像 $g_{i,t}^{(n)}$ を生成する。
- 6) 人物頭部候補領域画像 $g_{i,t}^{(n)}$ を選択された識別器に入力し、重み $\pi_{ada,i,t}^{(n)}$ を取得する。
- 7) 上記手順を各カメラに対して行い、得られた重みを次式により統合する。

$$\pi_t^{(n)} = \prod_i \pi_{edge,i,t}^{(n)} \pi_{ada,i,t}^{(n)} \quad (2.13)$$

すべての仮説に対して、式 (2.13) による重みを計算し、期待値をとることで各時刻の人物頭部の状態量を推定する。



図 2.9 入力画像領域の調整

*7 識別器の種類により学習用画像の切り出し位置を調整しているため。

*8 各カメラの観測情報の精度は一定と仮定し、評価に用いる識別対象画像サイズを固定としているため。

以降では識別器による仮説の評価を、「カスケード型識別器による仮説の評価」と「非カスケード型識別器による仮説の評価」のそれぞれの場合について述べる。

2.5.5 カスケード型識別器による仮説の評価

各仮説に対応した人物頭部候補領域画像 $g_{i,t}^{(n)}$ の人物頭部らしさをカスケード型識別器を用いて評価する。カスケードの各段の識別器は、階層が進むにしたがって、より多くの弱識別器 $h_t(x)$ を用いて判定を行う。そのため、より多くの識別器を通過した人物頭部候補領域画像 $g_{i,t}^{(n)}$ は、より多くの人物頭部の特徴を保持していると考えられる。つまり、カスケードの通過段数と人物頭部らしさには正の相関があると考えられる。そこで、人物頭部候補領域画像 $g_{i,t}^{(n)}$ をカスケード型識別器に入力した際に通過した識別器の数(カスケード段数)を人物頭部らしさの評価値とする。これは、仮説が実際の人物頭部の状態と大きく離れて生成された場合、対応する人物頭部候補領域画像 $g_{i,t}^{(n)}$ はカスケードの初期に棄却されるため、計算コストの点からも都合が良い。

一方、各段の識別器の識別性能が一定である場合、識別器が直列に配列されていることから、評価値を通過段数に応じて指数的に増加させることも考えられる。しかし、カスケードの各段に配置される識別器が持つ識別性能は一定ではなく、カスケードの初期を通過した人物頭部候補領域画像 $g_{i,t}^{(n)}$ の多くが、カスケードの最後まで通過する傾向があるため、評価値を指数的に増加させると、カスケードの中盤で棄却された人物頭部候補領域画像 $g_{i,t}^{(n)}$ の評価が追跡に反映されにくい。そのため、カスケード型識別器による仮説の評価では、人物頭部候補領域画像 $g_{i,t}^{(n)}$ が通過したカスケード段数により仮説の評価を行う。ただし、正面、 90° 左向き、 90° 右向きなどの人物頭部の各向きのカスケード型識別器は、同じカスケード段数とし、人物頭部と非人物頭部で通過する識別器の数(カスケード段数)に十分な差がつくように学習を行う。

仮説の評価は、具体的には、人物頭部候補領域画像 $g_{i,t}^{(n)}$ を選択したカスケード型識別器に入力し、人物頭部候補領域画像 $g_{i,t}^{(n)}$ が通過した識別器の数(カスケード段数)を取得し、得られたカスケード段数を対応する仮説の重み $\pi_{ada,i,t}^{(n)}$ とする。例えば、カスケードの最初の段で棄却された場合、重みは1となり、全カスケード段数が40段で、すべての識別器を通過した場合、重みは41となる(図2.10)。

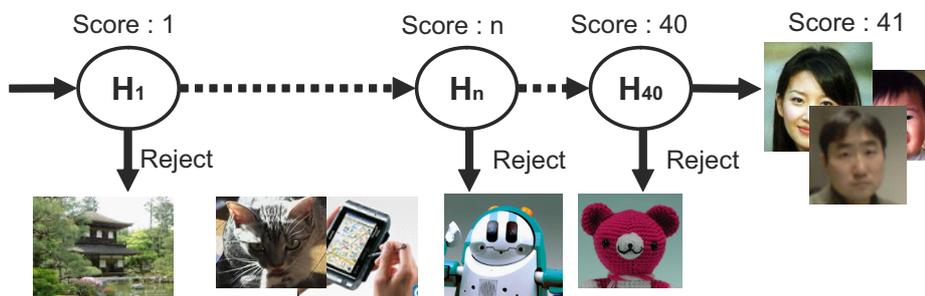


図 2.10 カスケード型識別器による評価

2.5.6 非カスケード型識別器による仮説の評価

各仮説に対応した人物頭部候補領域画像 $g_{i,t}^{(n)}$ の人物頭部らしさを AdaBoost 識別器の式 (2.7) の判定前の値を用いることで人物頭部らしさを評価する．すなわち以下の $H(x)$ を用いる．

$$H(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (2.14)$$

$H(x)$ の値は弱識別器である矩形特徴量の出力をそのまま保持しているため， $H(x)$ の値が大きいほど入力画像 x がより多くの矩形特徴を保持していると考えられる．

学習時にカスケード段数をあらかじめ決定しているカスケード型識別器では，真のデータを入力した場合の出力は最大で“カスケードの全段数 + 1”となり，正面，90° 右向き，90° 左向きなどの各識別器の最大評価値は整合性が保たれている．しかし，弱識別器出力の線形結合である $H(x)$ の値は，正面，90° 右向き，90° 左向きなどの各識別器の間で整合性は無い．そこで， $H(x)$ の値を調整し，各方向の識別器で同一の尺度での評価を可能とするため， $H(x)$ を尤度へ変換する関数を導入する．このような識別器の出力の調整は尤度のキャリブレーションと呼ばれ，文献 [14, 71, 96] などで議論されている．

Platt は，入力画像 x に対する識別器出力 $H(x)$ が得られたとき，入力画像 x が識別対象である確率密度分布 $p(\text{Positive} | H(x))$ を以下のようにシグモイド関数で近似する手法を示している [71]．

$$p(\text{Positive} | H(x)) = \frac{1}{1 + \exp(A \cdot H(x) + B)} \quad (2.15)$$

このシグモイド関数のパラメータ A, B は，学習画像セット (H_k, y_k) を用いて，以下のように損失関数^{*9}を最小化することで推定する．

$$\arg \min_{A, B} \left[- \sum_i y_i \log(p_k) + (1 - y_k) \log(1 - p_k) \right]$$

$$p_i = \frac{1}{1 + \exp(A \cdot H_k + B)} \quad (2.16)$$

ここで， $H_k = H(x_k)$ であり， y_k は *Positive* と *Negative* の別 $y_k \in \{0, 1\}$ を表す．

仮説の評価は，具体的には，人物頭部候補領域画像 $g_{i,t}^{(n)}$ を選択した非カスケード型識別器に入力し，識別器の出力 $H(x)$ を式 (2.15) で調整し，仮説の重み $\pi_{ada,i,t}^{(n)}$ とする．

^{*9} 相互情報量に基づいて定義される．詳細は文献 [69] などを参照されたい．

2.6 実験

本手法の有効性を確認するため、人物頭部の追跡実験を行った。実験は室内天井に設置した視野を共有する IEEE1394 カラーカメラ 2 台 (Point Grey Research 社製 Flea(図 2.11)) を用いて行った。また、各カメラの映像は 640×480 ピクセルで取得した。



図 2.11 IEEE1394 カメラ Flea

カメラキャリブレーションは Zhang の手法 [118] により行った^{*10}。また、Zhang の手法により内部パラメータを取得後、カメラを環境に設置し、実世界座標 (x, y, z) とカメラ画像座標 (u, v) の対応を 10 点程度計測することで、位置と姿勢を表す外部パラメータ (\mathbf{T}, \mathbf{R}) を求めた^{*11}。

また、カメラはネットワークで接続された複数台のクライアント PC (Petium4 3.2GHz, Memory 1GByte) にそれぞれ接続し、追跡は仮説の生成や尤度の統合を行うサーバ PC (Petium4 3.2GHz, Memory 1GByte) と通信することにより行った (図 2.12)。具体的には、以下の手順によりサーバ PC とクライアント PC の間で仮説 $s_t^{(n)}$ と重み $\pi_{i,t}^{(n)}$ が送受信される。

- 1) サーバ PC で新たに生成された仮説 $s_t^{(n)}$ は、送信キューに格納される。
- 2) サーバ PC の送信キューに格納された仮説 $s_t^{(n)}$ は、クライアント PC に順次送信される。
- 3) クライアント PC で受信された仮説 $s_t^{(n)}$ は、受信キューに一時的に格納される。
- 4) クライアント PC の受信キューに格納された仮説 $s_t^{(n)}$ は、接続されたカメラ画像に基づいて順次評価され、重み $\pi_{i,t}^{(n)}$ がクライアント PC の送信キューに格納される。
- 5) クライアント PC の送信キューに格納された重み $\pi_{i,t}^{(n)}$ はサーバ PC に順次送信される。
- 6) サーバ PC で受信された重み $\pi_{i,t}^{(n)}$ は、受信キューに一時的に格納され、他のクライアント PC から受信した重みと統合される。

送受信される仮説と重みはネットワークの輻輳を回避するため、20~30 個のデータを一つのバケットとし、1Gbps のイーサネットワークで送受信した。

^{*10} Zhang の手法によるカメラキャリブレーションは MATLAB[54] 上で動作するツールボックスとして公開されている [9]。

^{*11} 外部パラメータの算出は文献 [81] を参照されたい。

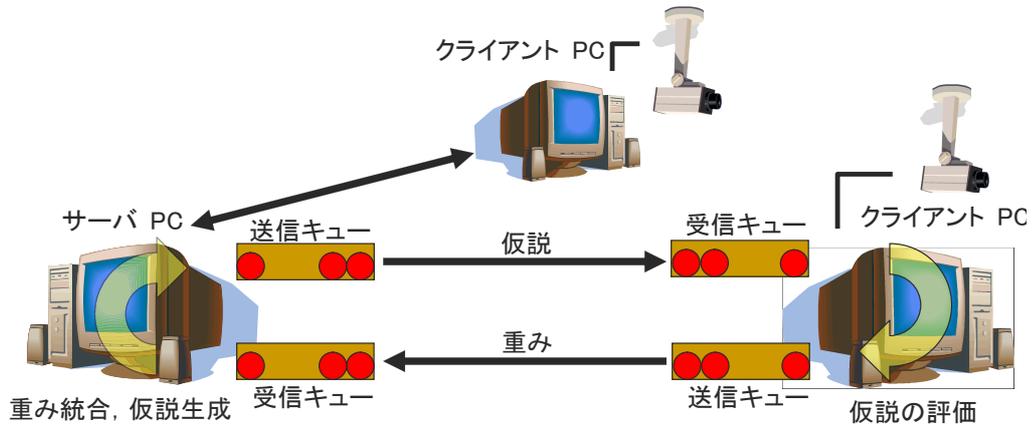


図 2.12 複数の PC による処理の分散

実験では、およそ $2\text{m} \times 2\text{m}$ の床上を観測領域とし、パーティクルフィルタの仮説数を 200 として人物一人の頭部追跡実験を行った。 Σ_{ω} の分散 $\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_{\theta}^2$ の標準偏差は、それぞれ $\sigma_x = 4\text{cm}$, $\sigma_y = 4\text{cm}$, $\sigma_z = 2\text{cm}$, $\sigma_{\theta} = 45^\circ$ とした。人物は頭部の向きを変えながら、観測領域内を歩き、屈伸動作を行った。実験で利用したカメラから得られた画像の例を図 2.13 に示す。この実験では、人物頭部は 1 フレーム (30ms) で平均 1.2cm、最大 6.6cm 移動した。また、カメラ画像では平均 3.3 ピクセル、最大 21.1 ピクセルの変位が観察された。



(a) Camera 1

(b) Camera 2

図 2.13 実験で得られたカメラ画像の例

2.6.1 カスケード型識別器を用いた追跡器による実験

人物頭部の正面, 90° 右向き, 90° 左向きをそれぞれ検出するように学習を行った 3 種類のカスケード型識別器を用いて追跡した. 各識別器のカスケード段数は 40 段とし, 識別器に入力する画像サイズは 24×24 ピクセルとした.

カスケード型識別器を用いた追跡器による追跡結果を図 2.14 に示す. 頭部位置の推定結果は矩形で画像に重ねて表示した. 図 2.14 より, ほぼ正確に人物頭部の中心を推定できていることが分かる. また, 一人の人物頭部を 2 台のカメラを用いて追跡した場合, 仮説 200 個の処理時間の合計は 30ms 程度であった.

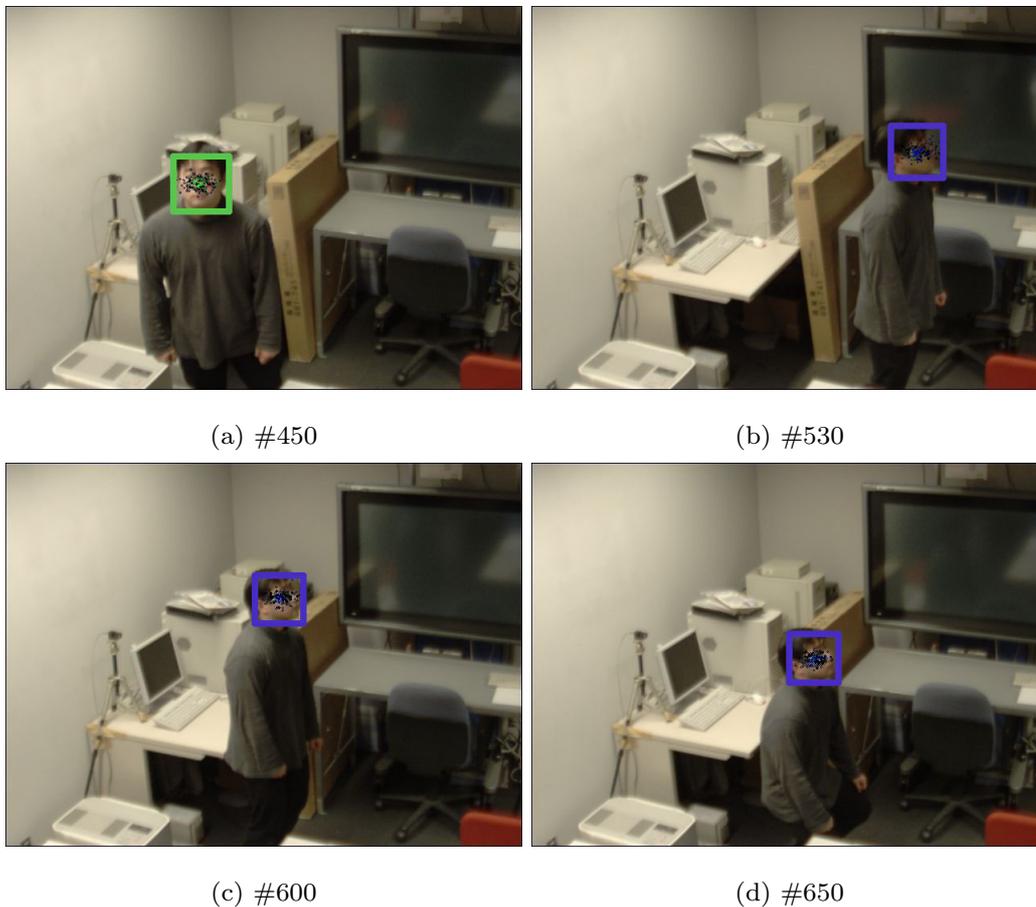


図 2.14 カスケード型識別器を用いた追跡器による人物頭部の追跡結果

提案手法による追跡の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して求めた三次元座標を真の位置と見なし, 推定結果と比較した. 図 2.15 に, 推定結果と対応する人物頭部の真の位置の XY 平面, 及び XZ 平面での軌跡を示す. また, Z 軸方向及び XY 平面上での平均誤差及び標準偏差を表 2.2 に示す. XY 平面上の平均誤差, Z 軸方向の平均誤差は共に 2cm 程度であり, 高い精度で追跡できている. なお, 図 2.15 において, 人物頭部の真の位置の軌跡上に示した灰色の矩形は, その周辺での空間解像度 (画面上の 1 ピクセルの実世界での広がり) を表している.

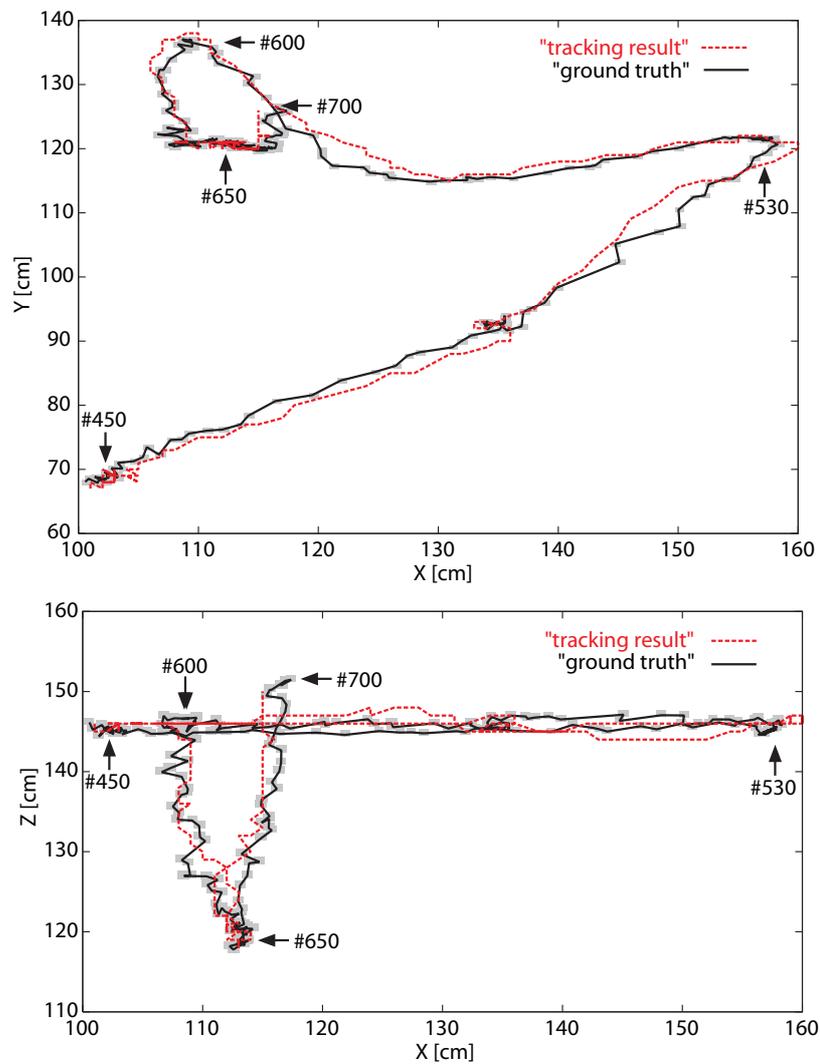


図 2.15 カスケード型識別器を用いた追跡器による人物頭部の追跡結果の軌跡

表 2.2 カスケード型識別器を用いた追跡器による人物頭部の追跡における検出誤差

	平均 [cm]	標準偏差 [cm]
Z 軸方向	1.02	0.65
XY 平面	1.98	1.46

2.6.2 非カスケード型識別器を用いた追跡器による実験

人物頭部の正面，45° 右向き，90° 右向き，135° 右向き，45° 左向き，90° 左向き，135° 左向きをそれぞれ検出するように学習及びキャリブレーションを行った 7 種類の非カスケード型識別器を用いて追跡した．識別器に入力する画像サイズは 24×24 ピクセルとした．

非カスケード型識別器を用いた追跡器による追跡結果を図 2.16 に示す．頭部位置の推定結果は矩形で画像に重ねて表示した．図 2.16 より，ほぼ正確に人物頭部の中心を推定できていることが分かる．また，1 人の人物頭部を 2 台のカメラを用いて追跡した場合，仮説 200 個の処理時間の合計は 50ms 程度であった．

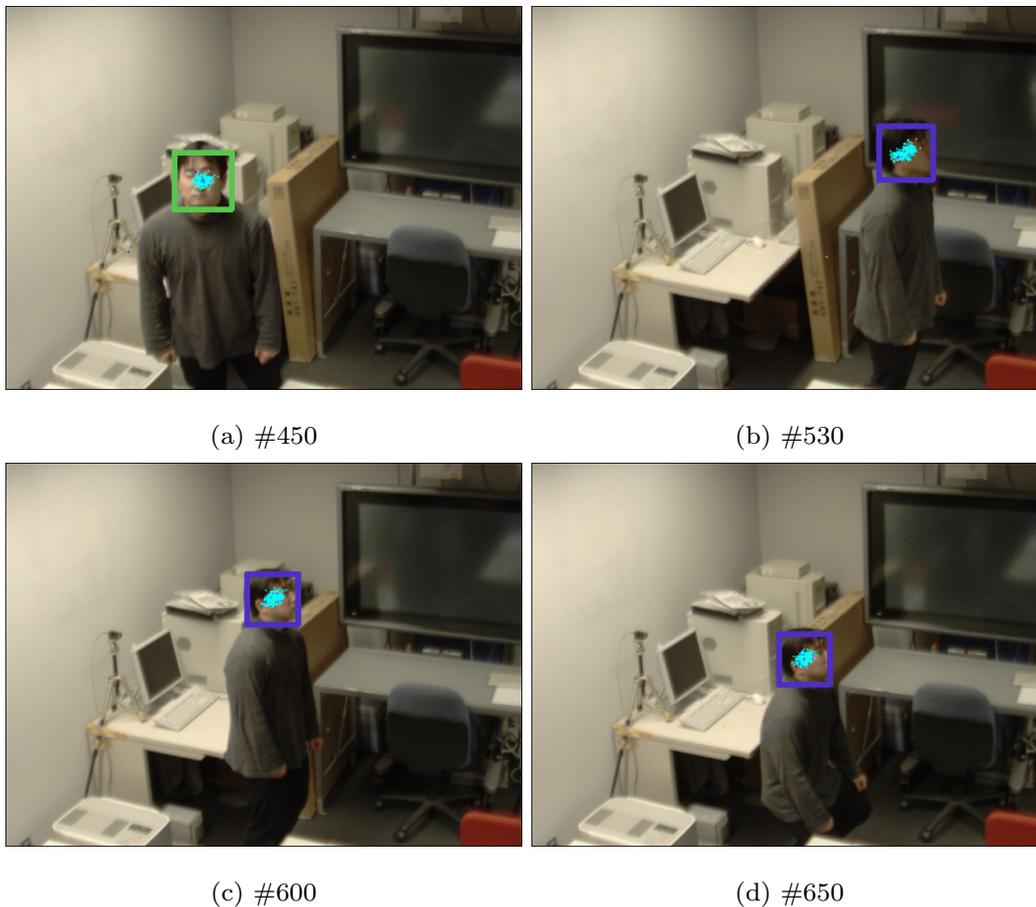


図 2.16 非カスケード型識別器を用いた追跡器による人物頭部の追跡結果

提案手法による追跡の精度を定量的に調べるために，2.6.1 節のカスケード型識別器を用いた追跡器による実験と同様に，人物頭部の真の位置と推定結果と比較した．図 2.17 に，推定結果と対応する人物頭部の真の位置の XY 平面，及び XZ 平面での軌跡を示す．また， Z 軸方向及び XY 平面上での平均誤差及び標準偏差を表 2.3 に示す． XY 平面上の平均誤差， Z 軸方向の平均誤差は共に 2cm 程度であり，高い精度で追跡できている．図 2.17 の灰色の矩形は空間解像度を表している．

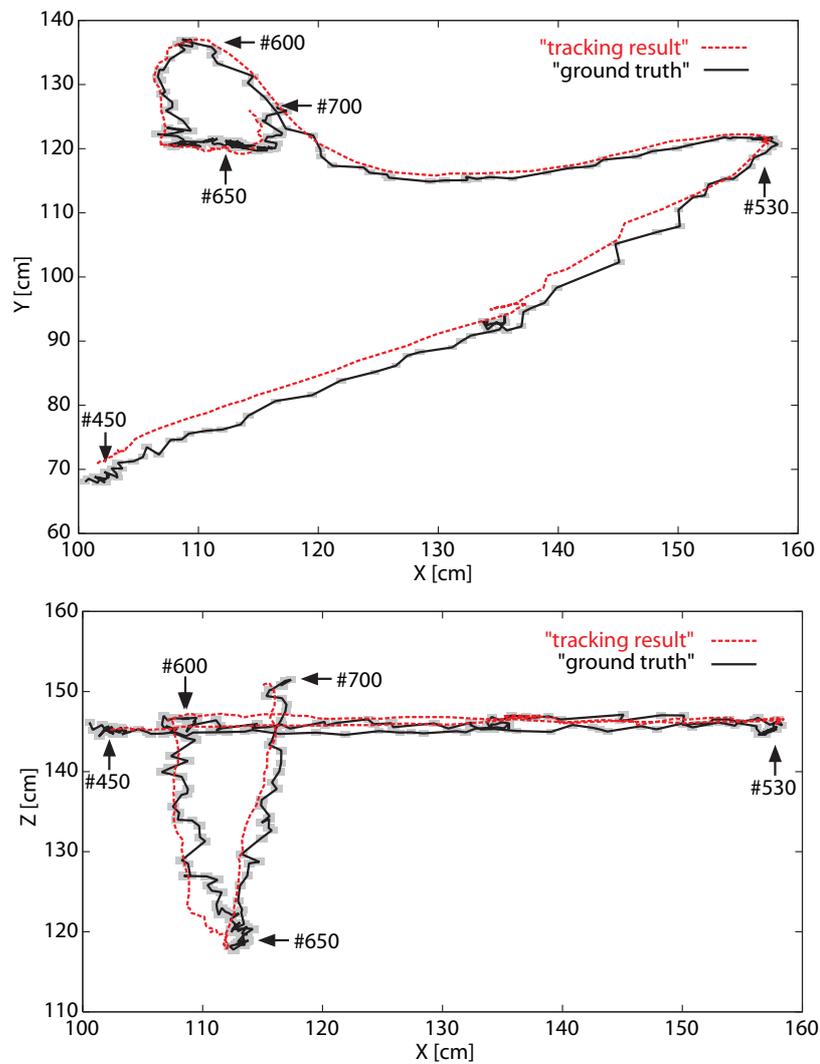


図 2.17 非カスケード型識別器を用いた追跡器による人物頭部の追跡結果の軌跡

表 2.3 非カスケード型識別器を用いた追跡器による人物頭部の追跡における検出誤差

	平均 [cm]	標準偏差 [cm]
Z 軸方向	1.23	0.66
XY 平面	2.13	1.09

識別器の尤度のキャリブレーションは、8100枚の頭部画像と61418枚の非頭部画像を用いて行った^{*12}。頭部画像は顔画像データベース^{*13}の識別器に対応する各方向の300人の頭部画像を頭部位置の切り出しを変動させて準備した。このとき、人物頭部領域以外の背景はランダムな背景画像に置き換えた(図2.18)。また、非頭部画像は、人物頭部が撮影されていない室内写真や風景写真など約400枚からランダムに位置と大きさを変化させて切り出すことで準備した。



図2.18 尤度のキャリブレーションに用いた顔画像の例

このようにして準備した頭部画像と非頭部画像を用いて尤度のキャリブレーションを行った。結果を図2.19～図2.25に示す。式(2.16)の最小化にはMATLAB[54]の最適化関数を利用した。

図2.19～図2.25より、すべての方向の識別器において、適切に尤度のキャリブレーションが行われていることがわかる。

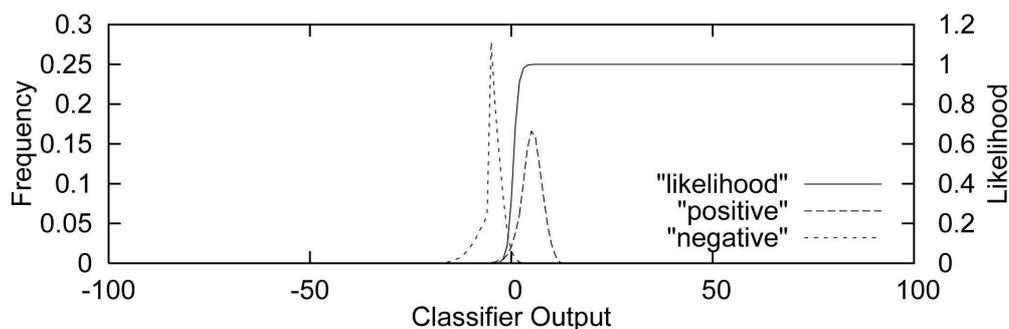


図2.19 尤度キャリブレーション結果 (Front)

^{*12} 識別器の学習に用いた画像とは別に準備する。

^{*13} (財) ソフトピアジャパンにより提供されている顔画像データベース [31] を用いた。10代～60代の日本人男女300人の頭部全周が15度間隔で撮影されている。

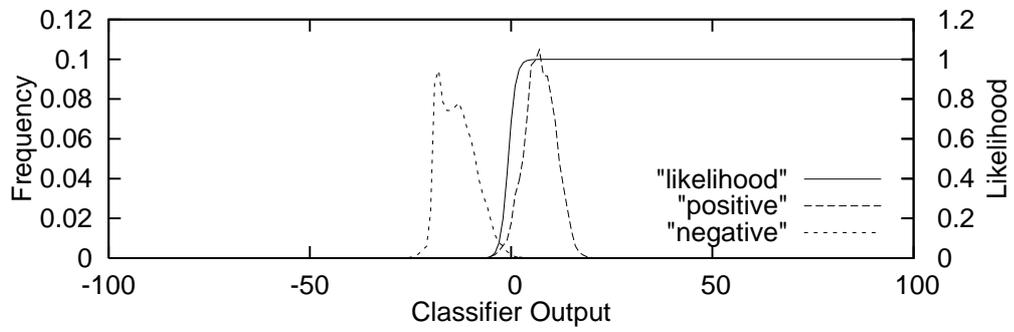


図 2.20 尤度キャリブレーション結果 (Left45)

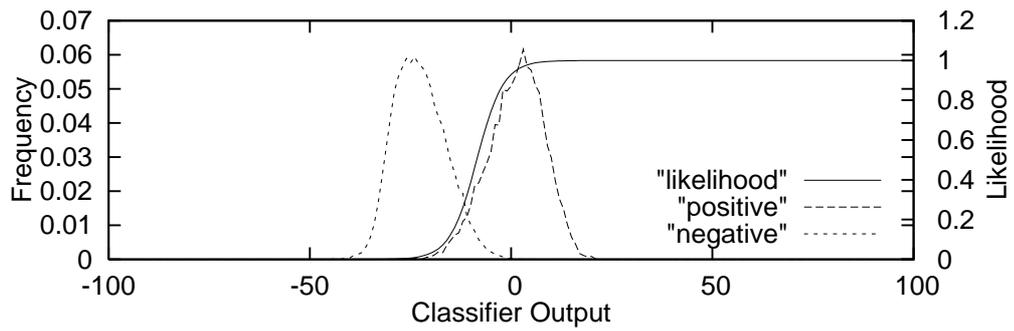


図 2.21 尤度キャリブレーション結果 (Left90)

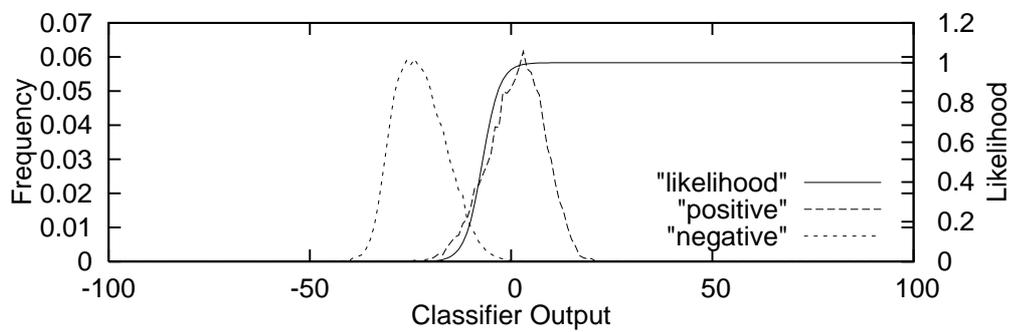


図 2.22 尤度キャリブレーション結果 (Left135)

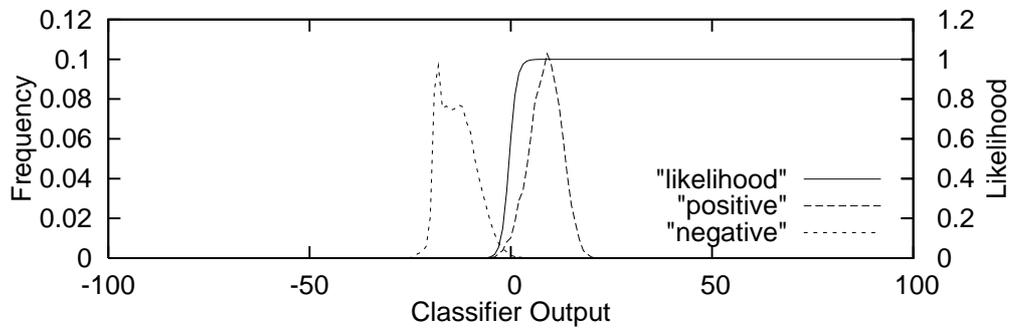


図 2.23 尤度キャリブレーション結果 (Right45)

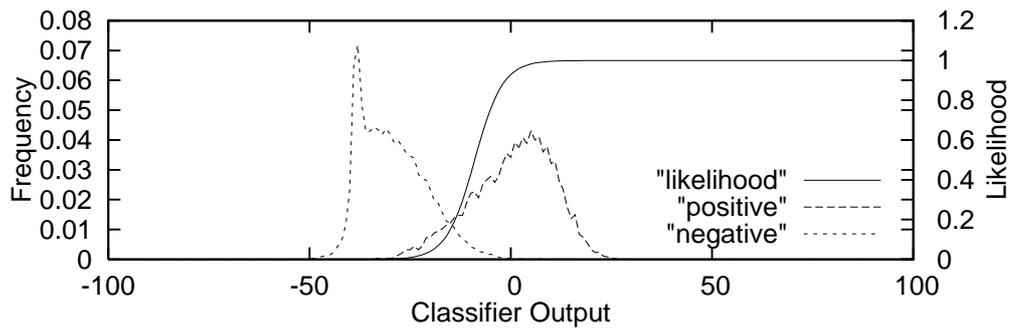


図 2.24 尤度キャリブレーション結果 (Right90)

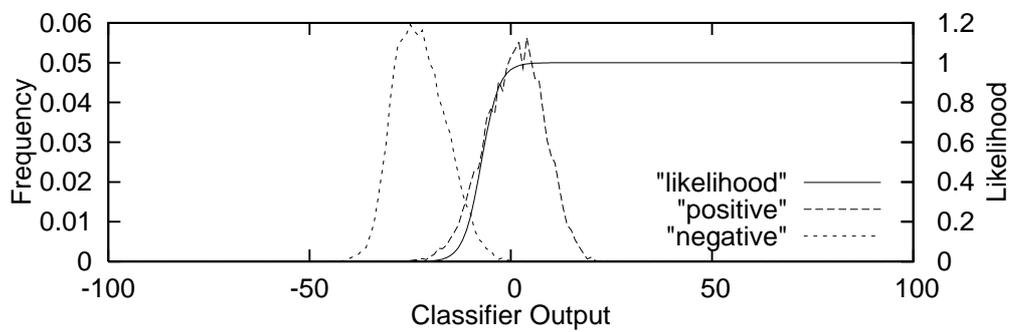


図 2.25 尤度キャリブレーション結果 (Right135)

2.7 考察

2.7.1 識別器による仮説の評価

パーティクルフィルタを用いた追跡では、各フレームにおいて、仮説の尤度を高精度に推定できることが追跡性能の向上に大きく寄与する。本手法の枠組みにおいては、人物頭部の仮説を画像平面上に射影した際の人物頭部画像領域中心でピークを持つ関数が理想的である。そこで、実際の人物頭部の位置周辺で、カメラから一定距離の平面を一边 1cm のグリッドで分割し、それぞれの三次元位置での人物頭部らしさの評価値を求めた。カスケード型識別器を用いた場合の評価を図 2.26 に示す。また、非カスケード型識別器を用いた場合の評価を図 2.27 に示す。さらに、輪郭の輝度勾配の類似性に基づく評価を図 2.28 に示す。なお、図 2.26 と図 2.27 の識別器は同程度の識別性能を示すように、同じ学習画像を用いて学習を行った。

図 2.26、図 2.27、図 2.28 より、識別器による評価 (図 2.26、図 2.27) は、輪郭の輝度勾配の類似性に基づく評価 (図 2.28) に比べ、人物頭部の中心で大きなピークを持ち、高精度に評価を行えることがわかる。

また、図 2.26 より、カスケード段数に基づく評価が、人物頭部の中心で鋭いピークを持ち、頭部以外の場所では一様に低くなっていることがわかる。

さらに、図 2.27 より、弱識別器出力に基づく評価が頭部中心で高く、頭部中心から離れるに従って評価値が緩やかに低くなり、カスケード段数に基づく評価に比べ分布の裾野が広いことが分かる。

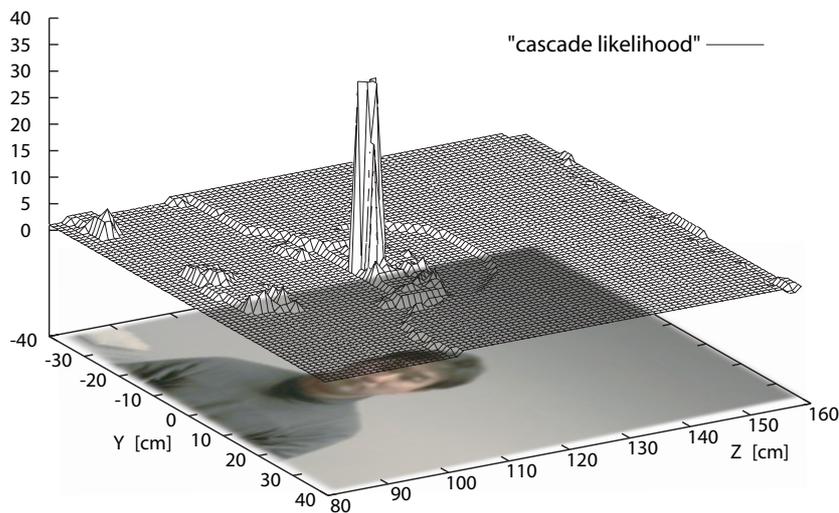


図 2.26 カスケード段数に基づく評価

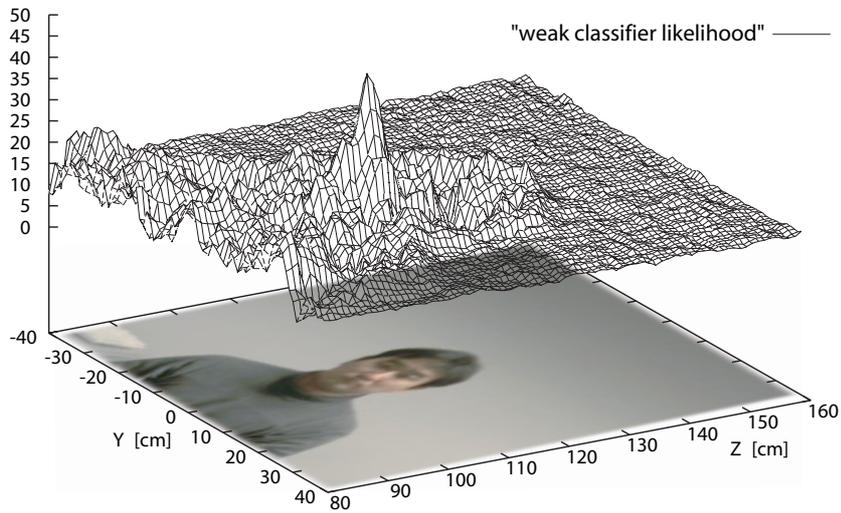


図 2.27 弱識別器出力に基づく評価

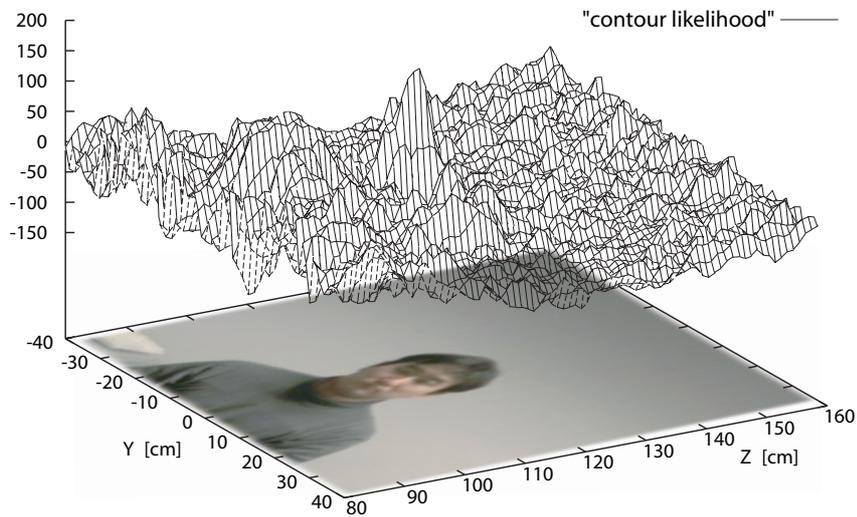


図 2.28 輪郭の輝度勾配の類似性に基づく評価

カスケード段数に基づく評価では、ピークが鋭いため、状態遷移モデルによる予測が外れた場合に追跡を失敗する可能性が高いが、実際の人物頭部の状態から大きく離れて生成された仮説はカスケードの初期で棄却されるため、計算コストの点で効率が良い。一方、弱識別器出力に基づく評価を用いた場合、状態遷移モデルによる予測が頭部中心を外れた場合でも追跡を失敗しにくい、実際の人物頭部の状態から離れて生成された多くの仮説に対しても、多数の弱識別器による判定を行うこととなり、計算コストの点では効率が悪い。

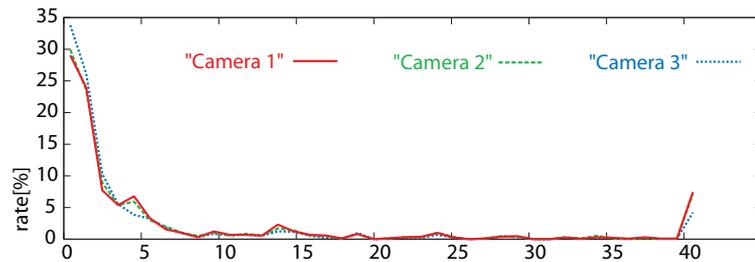


図 2.29 仮説が通過するカスケード段数

図 2.29 は人物 1 人の頭部を 3 台のカメラで 1200 フレームの間追跡した際のカスケードの各段で棄却される仮説の割合を示したものである。各カメラにおいて、カスケードの 5 段目までに全体の 7 割以上の仮説が棄却されていることがわかる。

パーティクルフィルタでは、仮説を増やすことによって、状態予測精度の低さを補うことができるが、計算コストは増大する。一方、少ない仮説数で状態予測精度が低い場合でも、分布の裾野の広い評価関数を用いることで、追跡の失敗を低減できると考えられる。本研究では追跡頑健性の観点から、弱識別器出力に基づく評価、すなわち、非カスケード型識別器を用いた評価を主として用いる。

2.7.2 複数識別器の適応的利用

人物頭部は向きやカメラとの位置関係によりその見えが異なるため、本手法では、複数の識別器を仮説とカメラの関係に基づいて選択的に用いている。そこで、相対的な頭部の向きと各方向の識別器の評価の関係を図 2.30、図 2.31 に示す。

図 2.30 は頭部の回転に伴い、カスケード型識別器による評価がどのように変化するかを示したものであり、横軸はフレーム番号、縦軸は頭部周辺を 2.7.1 節と同様に一辺 1cm のグリッドで分割した 266 点の評価値の和を表している。人物はカメラから見て、始めは左を向いており、フレームが進むにしたがって、徐々に正面を向き、最後には右を向いた状態となる。図 2.30 上部の写真はグラフの対応するフレーム付近の頭部の画像である。図 2.30 より、頭部がカメラから見て左向きの場合には 90° 左向きの識別器の評価が高く、正面、90° 右向きの識別器による評価は低い。頭部が正面を向いている場合は、正面の識別器の評価が最も高く、右を向いている場合は 90° 右向きの識別器の評価が最も高い。このように、正しい向きの識別器の評価が高くなることから、頭部の向きの仮説に基づいて選択的に識別器を用いても、頭部の向きを正しく推定でき、追跡を継続できる。

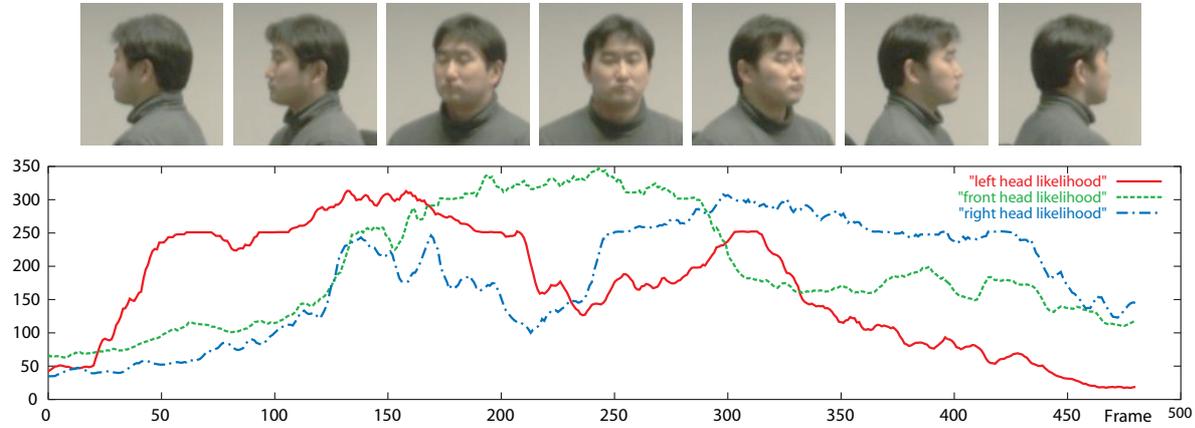


図 2.30 頭部の回転とカスケード型識別器の評価値

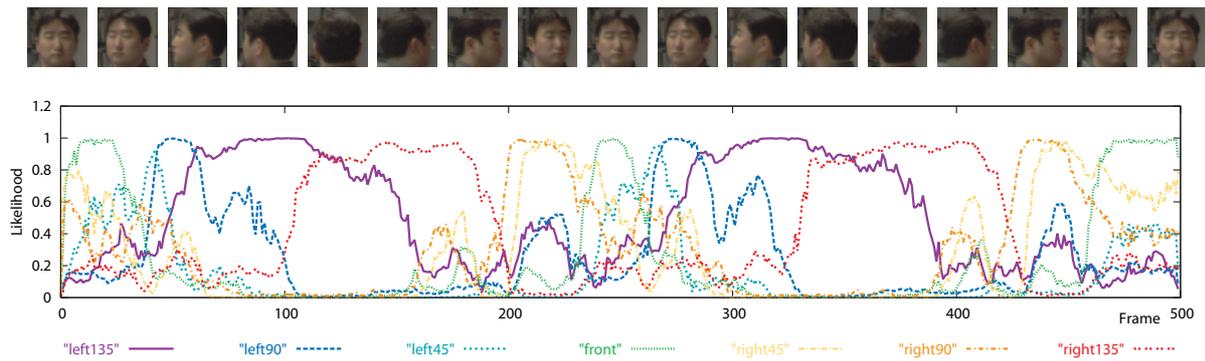


図 2.31 頭部の回転と非カスケード型識別器の尤度

図 2.31 は頭部の回転に伴い，非カスケード型識別器による評価がどのように変化するかを示したものである．図 2.30 と同様に横軸はフレーム番号，縦軸は尤度を示している．また，図 2.31 上部の写真はグラフの対応するフレーム付近の頭部の画像である．図 2.31 より，7 方向の識別器を用いることで，頭部のあらゆる方向において，いずれかの識別器の評価が高くなることから，頭部の向きの仮説に基づいて選択的に識別器を用いても，頭部の向きを正しく推定でき，人物頭部を継続して追跡できる．

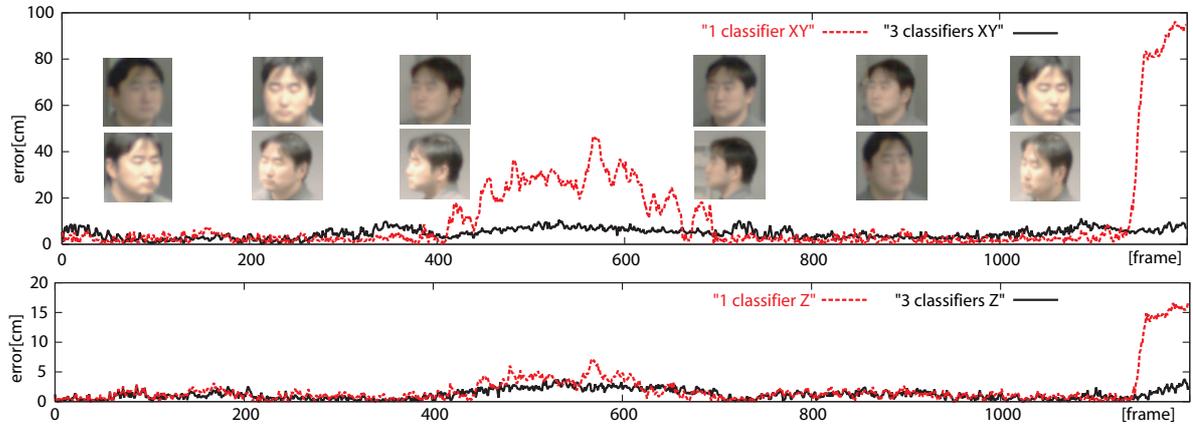


図 2.32 識別器の数による追跡精度の比較

複数識別器を適応的に用いることによる頭部の姿勢変動に対する追跡安定性を調べるため、2台のカメラで1人の人物頭部を観察し、人物頭部の正面の識別器1つを用いた場合と、90°右向き、90°左向きを加えた識別器3つを用いた場合で追跡精度を比較した。人物に対して2台のカメラはおよそ45°の角度を隔てて設置し、人物はどちらか一方のカメラで頭部の正面が観察されるように観測領域内を移動した。図2.32に1200フレームの間のXY平面上とZ軸方向の追跡誤差を示す。図中の写真は、2台のカメラによるグラフの対応するフレーム付近の頭部の画像である。

400フレームから700フレームの間で識別器1つを用いた場合の誤差が大きくなっている。これは一方のカメラで人物頭部の正面が観察できない状況となり、人物頭部の正面が観察されないカメラにおいて、正しい人物頭部らしさの評価ができていないためと考えられる。これに対して、識別器3つを用いた場合では、頭部の姿勢が変動しても、高精度な追跡が継続できていることがわかる。

2.7.3 状態遷移モデルによる予測

パーティクルフィルタによる追跡では、高精度な評価と共に、状態遷移モデルによる予測精度が重要となる。

多くの例では、人物頭部の急な動きの変化は少ないと仮定し、仮説 $\mathbf{s}_{t-1}^{(n)} = [x_{t-1}^{(n)} \ y_{t-1}^{(n)} \ z_{t-1}^{(n)} \ \theta_{t-1}^{(n)}]^\top$ を $\mathbf{s}_t^{(n)}$ に以下のように伝播する等速直線運動を仮定したモデルが用いられる。

$$\mathbf{s}_t^{(n)} = \mathbf{s}_{t-1}^{(n)} + \bar{\mathbf{v}}_t + \omega \quad (2.17)$$

ここで、 $\bar{\mathbf{v}}_t$ は過去一定時間の人物頭部の各状態量の平均速度であり、 $\bar{\mathbf{v}}_t = [\bar{x}_t \ \bar{y}_t \ \bar{z}_t \ \bar{\theta}_t]^\top$ である。 ω は平均 $\mathbf{0}$ 、共分散行列 Σ_ω を持つガウス雑音であり、 Σ_ω は分散 $\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_\theta^2$ を対角要素に持つ。また、 $\bar{\mathbf{v}}_t = \mathbf{0}$ として、ランダムウォークを仮定したモデルも用いられることが多い。

しかし、これらの手法では、人物頭部の動きの急激な変化に追従できず、追跡を失敗するケースが多い。例えば、等速直線運動を仮定したモデルでは、状態遷移モデルによる予測がオーバーシュートしてしまう場合がある。また、ランダムウォークを仮定したモデルでは、対象の高速な移動に対して追従性が悪い。

これらに対して、本手法では、人物の頭部位置周辺のオプティカルフローに基づき次の時刻での頭部位置を予測している。そのため、等速直線運動やランダムウォークを仮定したモデルに比べ、予測精度が高く、より少ない仮説での追跡が可能である。また、急激な動きの変化に対しても高い追従性を示す。

2.7.4 カメラの台数と追跡性能

人物を観察するカメラの台数による追跡精度の変化を調べるために、1 人の人物頭部の追跡にカメラ 2 台を用いた場合とカメラ 3 台を用いた場合で追跡精度を比較した。図 2.33 に 1200 フレームの間の XY 平面上と Z 軸方向の追跡誤差を示す。

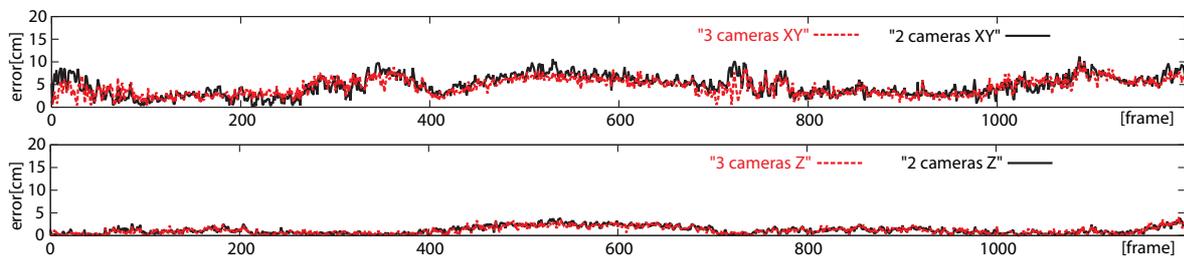


図 2.33 カメラ台数による追跡精度の比較

カメラ 2 台を用いた場合と、カメラ 3 台を用いた場合では追跡精度に大きな差は観察されなかった。これは、2 台のカメラによる仮説の評価により十分高精度に追跡できているためと考えられる。

また、本手法では仮説の状態量として三次元位置を持つため、図 2.34(a) のように人物が一部のカメラの視野から外れた場合においても、仮説をカメラに射影することで、視野から外れた仮説を判断することができるため、安定して追跡を継続することができる。

また、カスケード型識別器を用いた追跡器においては、追跡時に 1 台のカメラにおいて障害物による遮蔽が起きた場合でも追跡は安定していた (図 2.34(b))。これは、一部のカメラで障害物により頭部が遮蔽された場合、遮蔽の起きたカメラでの識別器による評価が一様な低い値となるため、頭部が観察されるカメラでの評価に、その影響が及びにくいと考えられる。



(a) 視野外への移動

(b) 障害物による遮蔽

図 2.34 複数カメラによる追跡の頑健性

2.7.5 人物頭部の向きの推定

本手法により、人物頭部の向きを推定することが可能であるが、正面、 90° 右向き、 90° 左向きの3方向のカスケード型識別器を用いた実験では、正面、右向き、左向きが判別できるに留まり、用いた識別器の数以上の分解能は観察されなかった。

7方向の非カスケード型識別器を用いた追跡では、より精度の高い人物頭部向きの推定が期待できる。そこで、人物頭部を水平方向 15° 刻みの視点で撮影した画像を用いて頭部向きの推定を行った。実験では、画像上の頭部の大きさが 30×30 ピクセル程度の1視点の静止画像に対して、30フレームの追跡処理を行い平均誤差を算出した。人物 A~J の10人の画像を対象として実験を行い、頭部の真の向きと推定方向の誤差を調べた。結果を図 2.35 に示す。平均誤差は 24.7° であるが、左右 90° 横顔付近で誤差が大きい。これは、頭部向きが左右 90° 付近のとき、左右 45° や左右 135° の識別器が比較的高い評価値を出力するのとためと考えられる。

そこで、各時刻において、頭部位置を推定した後、7方向すべての識別器による評価の分布を調べ、あらかじめ学習によって得た分布と比較することで頭部の向きを推定することを試みた。学習データとして、水平 24 方向 (15° 刻み) の5人の人物頭部に対して、識別器による評価 (7方向) を記録した。テスト画像を7方向の識別器により評価した値の分布と、学習によって得た24方向の値の分布のマハラノビス距離を算出し、最も近いクラスと2番目に近いクラスを取得し、距離に基づいて内挿することで頭部の向きを推定した。

このようにして、人物 A~D の4人の頭部の向きを推定した場合の誤差を図 2.36 に示す。平均誤差は 13.4° となり、追跡時に得られる期待値よりも高精度に向きを推定できている。

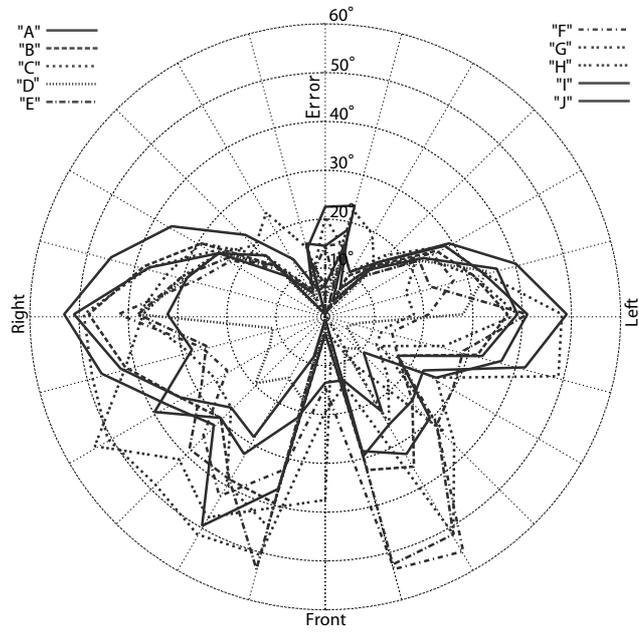


図 2.35 仮説群の期待値による頭部向き の推定誤差

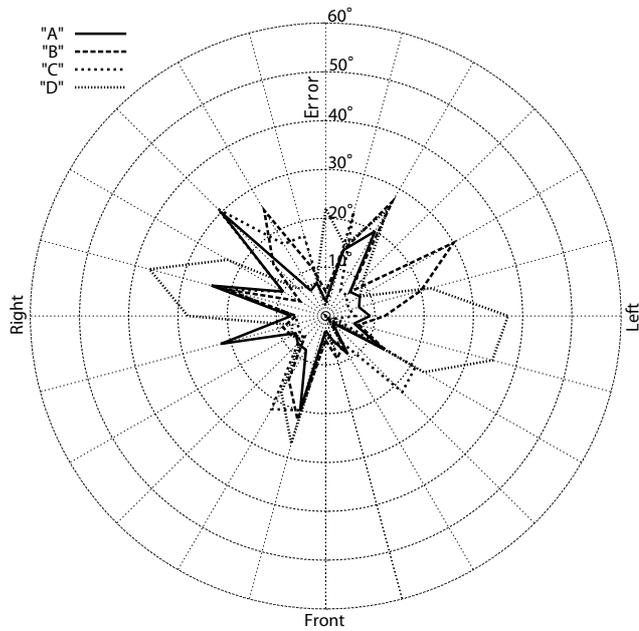


図 2.36 識別器評価値の分布に基づく頭部向き の推定誤差

2.8 まとめ

本章では、視野を共有した複数のカメラを用い、人物頭部を三次元空間内で頑健に追跡する手法を提案した。

パーティクルフィルタにおける仮説の評価に AdaBoost 学習による識別器を応用し、複数の識別器を仮説とカメラの関係に基づいて適応的に用いることで、人物が頭部の向きを変えながら移動する場合でも、高精度に人物頭部を追跡できることを示した。

AdaBoost 学習による識別器は、カスケード型と非カスケード型の両方を用いて、検証実験を行い、どちらの識別器を用いた場合でも、高精度に追跡できることを確認した。非カスケード型識別器では、各方向の識別器の出力を調整することで、人物頭部の全周において評価を可能とした。

オプティカルフローに基づく状態遷移モデルを用いることで、従来のランダムウォークや等速直線運動を仮定したモデルに比べて高精度な状態予測を可能とし、頑健な追跡を実現した。

追跡時の頭部向きの推定では、識別器の評価値の分布を考慮することで、追跡と同時に高精度に人物頭部の向きを推定可能とした。

第3章

カメラとレーザ測域センサの 統合によるエリア内複数人物追跡

3.1 はじめに

スーパーやコンビニエンスストアなどの特定エリア内において、人物の動線や各時刻での人物の注目方向を計測することができれば、動線の分類による不審者の検知や、顧客の購買行動の分析、インタラクティブ広告などへの応用が期待できる。

店舗などの特定エリア内において、監視カメラなどの画像を用いて複数人物の追跡を行うには、人物一人の追跡に関する課題に加えて、複数人物の追跡への拡張や、人物の出現に対する追跡初期化、人物相互の遮蔽などの課題がある(図 3.1)。



図 3.1 特定エリア内における複数人物追跡の難しさ

カメラ画像に基づく追跡初期化では、複雑背景、低解像度、照明変動、多様な見えなどの要因から、高精度かつ効率的な追跡対象の検出は難しい。また、複数対象の三次元追跡では、追跡開始時の三次元位置の獲得のため、複数カメラによる同一対象の観測を正しく対応づける必要がある。

人物相互の遮蔽に対しては、視野を共有した複数カメラを用い、理想的なカメラ設置状況を仮定することで人物相互の遮蔽の問題を回避している例も多い ([56, 61] など) が、実際の店舗などでは、監視カメラなどのセンサを設置する位置や姿勢などに制約があり、理想的な設置状況を実現できない場合も多い。例えば、店舗内の壁ぎわ通路などでは、壁側からの観察が難しいことは容易に予想される。また、直上からの観察も、天井の高さが十分ではないなどの要因により困難な場合が多く、人物相互の遮蔽による観測不能状態が頻繁に起きると予想される。

このように、実環境においては、良好な観測が得られないことが原因で、追跡を継続できない場合も多い。そこで、追跡が失敗した場合でも、良好な観測が得られた時点で、速やかに追跡を再開する枠組みが求められる。これには、観測不能な状態においても追跡対象の存在を内部状態として保持するアプローチ (例えば [35, 38] など) と、遮蔽により観測不能になった場合に追跡を終了し、観測可能となった時点で再度追跡を開始するアプローチ (例えば [53, 91, 100, 114] など) がある*1。後者は、観測領域内で追跡対象の検出を行う検出器を、追跡器と並列に動作させることで実現され、検出器は追跡初期化にも適用できることから、実環境での利用に適した効率的なアプローチといえる。このアプローチは、Merge-Split アプローチと呼ばれる (図 3.2)。

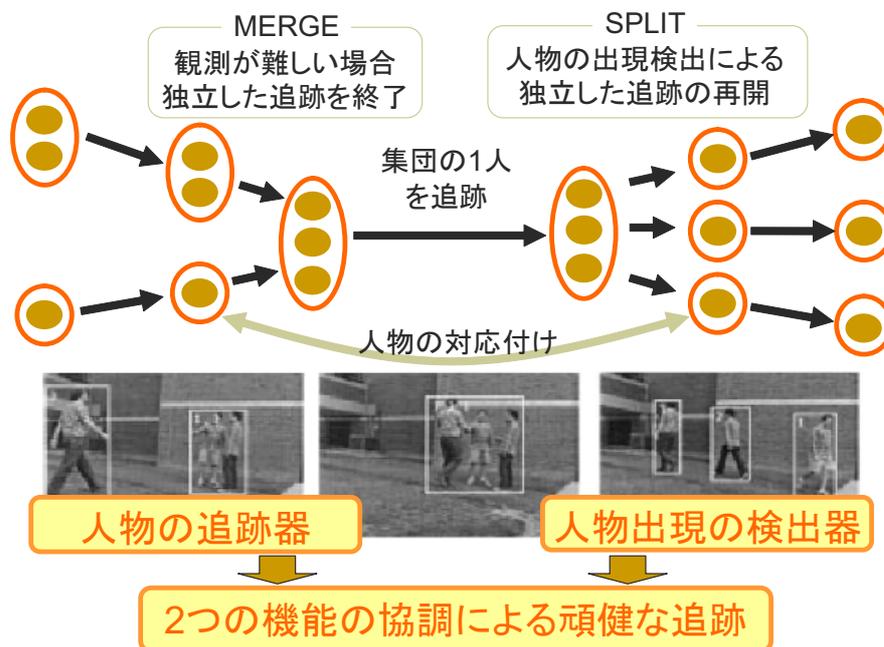


図 3.2 Merge-Split アプローチ

そこで、本章では、2章で述べた内容を踏まえ、検出器として新たにレーザ測域センサを導入し、エリア内に配置された分散カメラと相補的に統合することで、複数の人物頭部の三次元位置と向きを頑健に追跡する手法を提案する。

*1 これらの手法やアプローチについては、文献 [22] にまとめられている。

特に新規性の高い本手法の特徴は以下である。

- 1) レーザ測域センサによる観測に基づいて、エリア内での人物の出現と消失を検出することで、追跡の開始と終了を制御し、頑健に追跡初期化を行う。
- 2) 追跡は Mixture Particle Filter [104] の枠組みに基づいて行い、オプティカルフローに基づいた状態遷移とレーザ測域センサによる観測に基づいた分布を混合した確率密度分布から仮説を生成することで、カメラ画像による追跡にレーザ測域センサを統合する。

これにより、追跡初期化においては、エリア内に人物が出現した時点で速やかに追跡を開始し、人物同士がすれ違うなどの遮蔽により追跡が失敗した場合でも、良好な観測が得られた時点で、速やかに追跡初期化を行う。さらに、追跡においては、レーザ測域センサを導入することで、単眼カメラによる観測においても安定な三次元追跡を実現し、分散配置されたセンサ群のさまざまな組合せによる三次元人物追跡を可能とする。具体的には、追跡初期化時は、レーザ測域センサと一つ以上のカメラによる観測が得られればよく、追跡時は、単眼カメラとレーザ測域センサ、視野を共有した複数カメラ、視野を共有した複数カメラとレーザ測域センサ、のいずれかの組合せによる観測が得られればよい。

本章では、3.2 節にて関連研究について概観する。3.3 節では、パーティクルフィルタの拡張である Mixture Particle Filter について述べる。そして、3.4 節にて、カメラとレーザ測域センサの統合によるエリア内複数人物追跡について詳細に述べる。3.5 節では検証実験について述べ、3.6 節にて考察を行う。最後に、3.7 節にてまとめる。

3.2 関連研究

2章で述べたように、カメラ画像を用いた追跡では、パーティクルフィルタ [37] に基づいた手法が数多く提案され、複数対象追跡についても、その有効性が報告されている ([46, 51, 89, 91, 99, 113, 117] など)。

人物の出現・消失に対して、Isard ら [38] は、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を提案している。この手法は、人物の出現・消失や人物相互の遮蔽を扱うことができる反面、計算コストが高く、人物と背景を事前に学習する必要があるため、不特定人物の追跡には適していない。また、Huang ら [35] は、遮蔽を対象の重なりとして扱う手法を提案しているが、対象の重なりをレイヤとして記憶し、その相互関係を考慮するため、計算コストは高い。

一方、Merge-Split アプローチによる複数対象追跡では、追跡対象の検出を行う検出器を追跡器と並列に動作させることで、頑健に追跡初期化を行う手法が Okuma ら [67] により提案されている。この手法では、追跡対象の検出器に AdaBoost 学習による識別器を用い、追跡器に色ヒストグラムによる尤度評価に基づいた Mixture Particle Filter を用いている。さらに Cai ら [12] は、能動カメラを用いる場合に問題となる、検出器による検出結果と追跡対象の対応づけを高度化した手法を提案している。しかし、三次元空間での追跡を考える場合、画像平面と実世界の対応を逐次求める必要があることから、対象の検出と追跡に同じカメラから得られる画像情報を用いることは効率的ではない。

視野を共有する複数カメラによる人物の三次元追跡では、松元ら [57] により、視野を共有する複数カメラの一部で遮蔽が起きた場合にも、尤度の統合を工夫することで、追跡が継続できることが示されている。しかし、追跡初期化に関しては詳細に議論されておらず、複雑な背景などに対して、どの程度頑健に動作するのが明らかでない。これに対して、三次元空間での追跡対象の探索に、観測履歴に基づいた環境情報を利用することで、追跡初期化を効率的に行う手法 [90] も提案されている。しかし、追跡初期化には、視野を共有する複数カメラでの良好な観測が必要であり、複数対象の三次元位置の獲得では誤対応の課題が残る。

Cui ら [17] は、足元に設置した複数のレーザ測域センサによる観測に基づいて歩行者を追跡している。追跡初期化は、レーザ測域センサによる計測点の軌跡を観察することで実現される。この手法では、多数の歩行者をオープンスペースで高精度に追跡可能であるが、レーザ測域センサによる観測では見えの情報が得られないため、人物の同定や注目方向の計測などは難しい。また、カメラとレーザ測域センサの併用では、中村らの例 [62] があるが、カメラによる追跡はレーザ測域センサによる追跡が寸断された場合の補助的利用に留まり、カメラとレーザ測域センサによる観測を統合して追跡するものではない。

これらに対し、本手法では、レーザ測域センサの導入により、追跡対象の検出を容易にし、検出精度の向上と複数カメラによる観測の誤対応を回避した、頑健な追跡初期化を行う。また、三次元空間でカメラとレーザ測域センサによる観測を統合することにより、追跡の枠組みを変更することなく、さまざまなセンサの組合せで、人物頭部の三次元位置と向きを追跡する。

このような、カメラによる見えの観測とレーザ測域センサによる距離の観測という、異なるセンサモダリティを相補的に統合した追跡手法の例はない。

3.3 Mixture Particle Filter

2章で述べたように、カメラ画像を用いた人物追跡では、パーティクルフィルタが有効である。パーティクルフィルタは、追跡対象を状態量と尤度を持つ多数の仮説群により離散的な確率密度として表現することで、観測値が非ガウス型となる場合においても頑健な追跡を行う手法である。しかし、パーティクルフィルタでは追跡対象をノンパラトリックな分布で表現するため、複数対象を扱う場合には、単一対象に対する分布か、複数対象に対する分布かを区別することが難しい。

これに対して、Vermaarkらは、シーン全体の確率密度分布が個々の追跡対象の確率密度分布の重み付け和で表現できることを示し、実質的に複数のパーティクルフィルタを独立に駆動し、仮説の重みを調整することで、複数対象の追跡が可能であることを示した [104]。彼らの手法は Mixture Particle Filter として知られている。この概念を図 3.3 に示す。

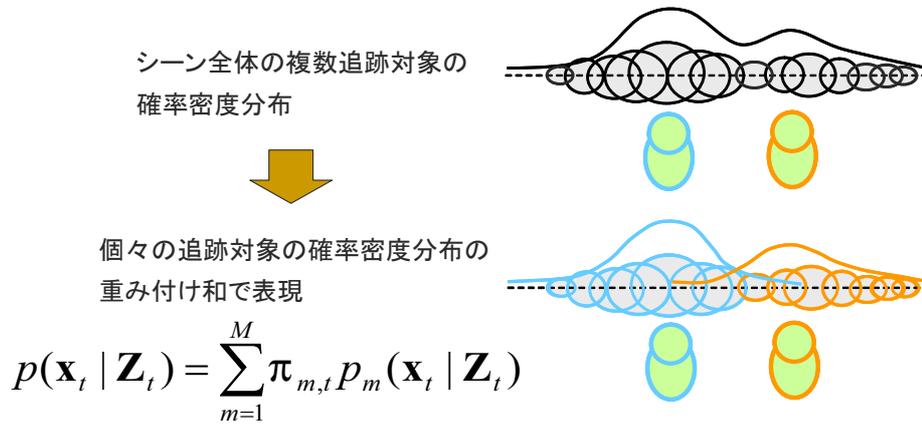


図 3.3 Mixture Particle Filter の概念

Mixture Particle Filter について、アルゴリズムの概略を以下に説明する。

時刻 t における対象の状態量を \mathbf{x}_t 、画像による観測を \mathbf{z}_t とし、時刻 t までに得られる観測を $\mathbf{Z}_t = \{\mathbf{z}_1, \dots, \mathbf{z}_t\}$ とする。このとき、時刻 t における対象の事前確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_{t-1})$ は、マルコフ過程を仮定することにより、時刻 $t-1$ における事後確率密度分布 $p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$ と時刻 $t-1$ から t への状態遷移確率密度分布 $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ を用いて以下のように表すことができる。

$$p(\mathbf{x}_t | \mathbf{Z}_{t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1}) d\mathbf{x}_{t-1} \quad (3.1)$$

時刻 t における事後確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_t)$ は、ベイズの法則より、時刻 t における尤度 $p(\mathbf{z}_t | \mathbf{x}_t)$ と事前確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_{t-1})$ により次のように表すことができる。

$$p(\mathbf{x}_t | \mathbf{Z}_t) = \frac{p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{Z}_{t-1})}{\int p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{Z}_{t-1}) d\mathbf{x}_t} \quad (3.2)$$

さらに，複数対象の追跡のため， p_m を m 番目の追跡対象の確率密度分布として，式 (3.1)，式 (3.2) を以下のように拡張する．

$$p(\mathbf{x}_t | \mathbf{Z}_{t-1}) = \sum_{m=1}^M \Pi_{m,t-1} p_m(\mathbf{x}_t | \mathbf{Z}_{t-1}) \quad (3.3)$$

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{Z}_t) &= \frac{\sum_{m=1}^M \Pi_{m,t-1} p_m(\mathbf{z}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{Z}_{t-1})}{\sum_{n=1}^M \Pi_{n,t-1} \int p_n(\mathbf{z}_t | \mathbf{x}_t) p_n(\mathbf{x}_t | \mathbf{Z}_{t-1}) d\mathbf{x}_t} \\ &= \sum_{m=1}^M \left[\frac{\Pi_{m,t-1} \int p_m(\mathbf{z}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{Z}_{t-1}) d\mathbf{x}_t}{\sum_{n=1}^M \Pi_{n,t-1} \int p_n(\mathbf{z}_t | \mathbf{x}_t) p_n(\mathbf{x}_t | \mathbf{Z}_{t-1}) d\mathbf{x}_t} \right] \\ &\quad \times \left[\frac{p_m(\mathbf{z}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{Z}_{t-1})}{\int p_m(\mathbf{z}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{Z}_{t-1}) d\mathbf{x}_t} \right] \\ &= \sum_{m=1}^M \Pi_{m,t} p_m(\mathbf{x}_t | \mathbf{Z}_t) \end{aligned} \quad (3.4)$$

ここで， $\Pi_{m,t}$ は以下で表される個々の追跡対象の重みであり， $\sum_{m=1}^M \Pi_{m,t} = 1$ を満たす．

$$\Pi_{m,t} = \frac{\Pi_{m,t-1} \int p_m(\mathbf{z}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{Z}_{t-1}) d\mathbf{x}_t}{\sum_{n=1}^M \Pi_{n,t-1} \int p_n(\mathbf{z}_t | \mathbf{x}_t) p_n(\mathbf{x}_t | \mathbf{Z}_{t-1}) d\mathbf{x}_t} \quad (3.5)$$

これより，個々の対象の追跡は事後確率密度分布 $p_m(\mathbf{x}_t | \mathbf{Z}_t)$ の期待値を逐次求めることで実現される．

本手法では，次節にて述べるように，事前確率密度分布 $p_m(\mathbf{x}_t | \mathbf{Z}_{t-1})$ として，オプティカルフローに基づいた状態遷移とレーザ測域センサによる観測に基づいた分布を混合したモデルを仮定する．また，尤度 $p_m(\mathbf{z}_t | \mathbf{x}_t)$ は，2章で述べた手法に従い，輪郭の輝度勾配の類似性と AdaBoost 識別器による評価に基づいてカメラ画像から得る．

Mixture Particle Filter では，時刻 t における事後確率密度分布 $p(\mathbf{x}_t | \mathbf{Z}_t)$ を，状態量 \mathbf{x}_t の仮説群 $\{\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(I)}\}$ と各仮説に対応する重み $\{\pi_t^{(1)}, \dots, \pi_t^{(I)}\}$ により次のように離散的に近似する．

$$p(\mathbf{x}_t | \mathbf{Z}_t) \approx \sum_{m=1}^M \Pi_{m,t} \sum_{i \in I_m} \pi_t^{(i)} \delta(\mathbf{s}_t^{(i)}) \quad (3.6)$$

ここで， I_m は m 番目の追跡対象に所属する仮説群を表す．また， $\delta(u)$ はデルタ関数である．

また， $\Pi_{m,t}$ は次式によって更新する．

$$\Pi_{m,t} = \frac{\Pi_{m,t-1} \sum_{i \in I_m} \pi_t^{(i)}}{\sum_{n=1}^M \Pi_{n,t-1} \sum_{i \in I_n} \pi_t^{(i)}} \quad (3.7)$$

この枠組みにより複数対象の追跡が実現されるため，以降は個々の対象の追跡について議論する．Mixture Particle Filter の詳細については文献 [104] を参照されたい．

3.4 カメラとレーザ測域センサの統合による エリア内複数人物追跡

3.4.1 レーザ測域センサ

レーザ測域センサは，レーザ光を内部のミラーの回転により円周上に投光し，反射光が観測される遅延時間（光飛翔時間）に基づいて，センサから物体までの距離を計測するセンサである（図 3.4）．本研究で用いたレーザ測域センサ（SICK 社製 S300[86]）では，センサを中心に周囲 270°，距離 15m の計測範囲を，計測周波数 24Hz，角度分解能 0.5°，計測誤差 ±1cm で計測可能である．主な仕様を表 3.1 にまとめて示す．

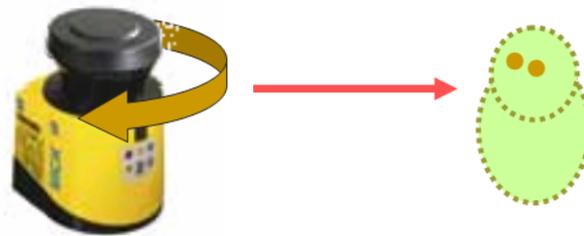


図 3.4 レーザ測域センサ (SICK 社 S300)

表 3.1 SICK 社レーザ測域センサ S300 の主な仕様

計測距離	1500 [cm]
距離分解能	1 [cm]
スキャン角度	270 [°]
角度分解能	0.5 [°]
測定周波数	24 [Hz]
インタフェース	RS422
レーザクラス	クラス 1(アイセーフ)
外形寸法	102x152x105 [cm]
重量	1.2 [Kg]
電源	DC24V

本手法では，床面と平行な平面を計測面とする．この場合，事前にセンサの位置 $(\Delta m, \Delta n)$ と向き β を世界座標系で計測しておくことで，レーザ測域センサで計測される点の座標 (m, n) の二次元のヘルマート変換（式 (3.8)）により，計測点の世界座標 (x, y) を得ることができる．

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} m \\ n \end{pmatrix} + \begin{pmatrix} \Delta m \\ \Delta n \end{pmatrix} \quad (3.8)$$

3.4.2 レーザ測域センサによる人物頭部候補位置の検出

本手法では，レーザ測域センサを人物の平均的な頭部の高さに設置する^{*2}ことで，観測領域内の人物頭部候補を検出する．

具体的には，レーザ測域センサによる観測に基づいて，時刻 t における人物頭部候補 $y_t = [x_t \ y_t \ z_t \ \theta_t]^T$ を得る．このとき， x_t, y_t をレーザ測域センサによる観測から取得し， z_t, θ_t は，以降で述べる通り，追跡初期化時と追跡時で異なる値を設定する．

人物が存在しないときの計測点群を背景データとして事前に記録した後，以下の手順により人物頭部候補位置 x_t, y_t を得る．

- 1) 時刻 t で得られた計測点群から，背景データとの差分を抽出する．
- 2) 差分計測点群を距離 10cm を閾値としてクラスタリングする．
- 3) クラス内に差分計測点が一定数^{*3}以上存在するクラスを人物頭部候補とし，クラスの平均を x_t, y_t とする．

図 3.5 に人物頭部候補位置を検出している様子を示す．

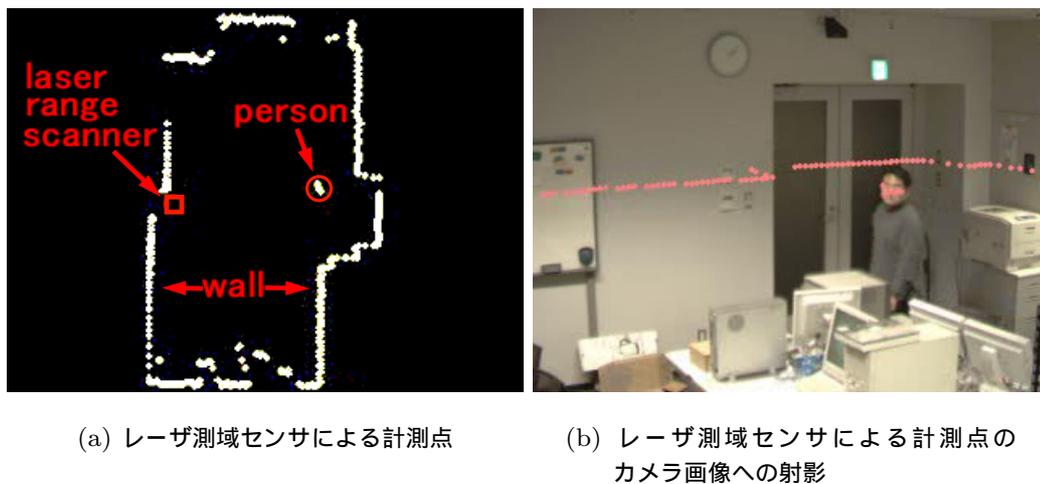


図 3.5 レーザ測域センサによる計測

図 3.5(a) は，レーザ測域センサにより計測された点を二次元の地図として表示したものである．矩形で示した位置にレーザ測域センサが設置されており，壁面などの室内の形状がスキャンされていることが分かる．また，赤円は検出された人物頭部候補位置である．図 3.5(b) のカメラ画像に重ねて表示した点は，レーザ測域センサにより計測された点を校正済みのカメラ画像座標に射影したものである^{*4}．人物頭部付近の一定の高さに計測点が存在していることが分かる．

^{*2} レーザ測域センサはレーザクラス 1 (アイセーフ) であり，また空間中の同じ位置にはパルス的に投光されるため，人間の目に影響はない．

^{*3} センサから最も遠い位置で頭部が検出されるよう決定する．

^{*4} 通常のカメラではレーザが投光された点は観察できない．

3.4.3 レーザ測域センサの統合による人物頭部追跡

本節では，レーザ測域センサによる観測に基づいて得られる人物頭部候補を，追跡の枠組みに統合する手法について述べる．レーザ測域センサによる観測から得られる人物頭部候補位置 x_t, y_t は，以下の3つのフェーズで追跡の枠組みに統合される．

1) 追跡初期化への統合

レーザ測域センサによる観測から得られた人物頭部候補位置の近傍に既追跡対象がなければ，新規追跡対象として追跡を開始する．

2) 追跡への統合

既追跡対象の近傍にレーザ測域センサによる観測から得られた人物頭部候補位置があれば，仮説を状態遷移モデルとレーザ測域センサによる観測を統合した混合分布からサンプリングする．

3) 追跡終了への統合

既追跡対象の近傍にレーザ測域センサによる観測から得られた人物頭部候補位置がなく，識別器により評価された全仮説の累積尤度が一定のしきい値より低い場合，追跡を終了する．

各フェーズでの人物頭部候補位置と既追跡対象の対応付けは， XY 平面上の距離に基づいて行う．本手法では，水平距離 50cm 以内の人物頭部候補位置と既追跡対象を同一人物として対応付けた．図 3.6 にレーザ測域センサが統合された追跡の手順を示す．

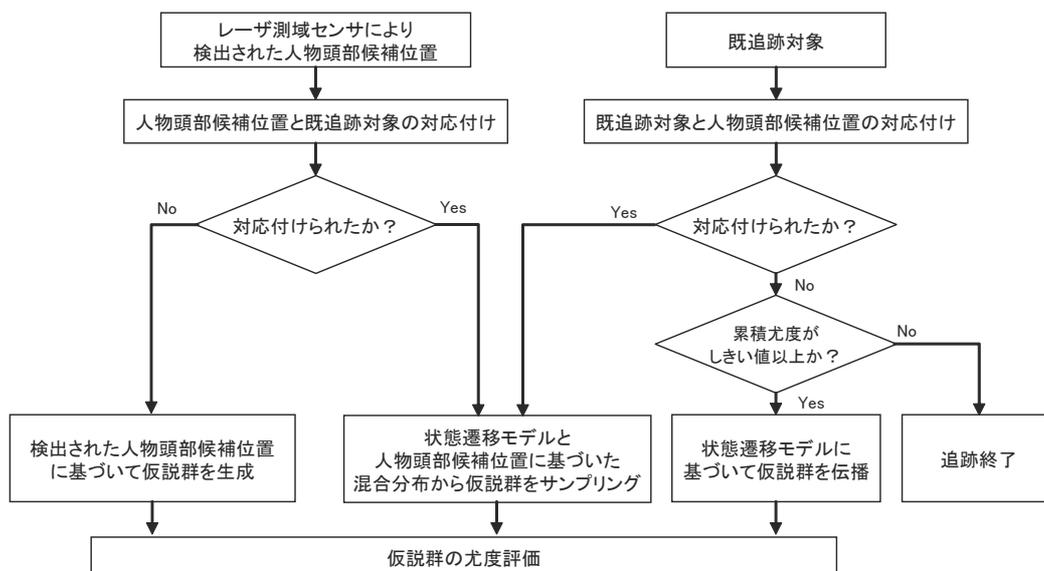


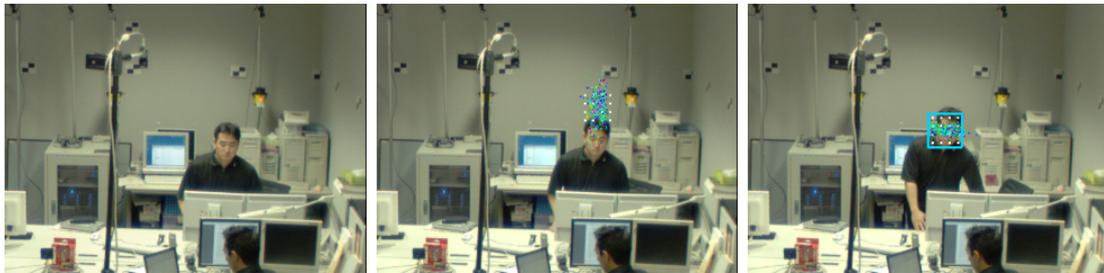
図 3.6 レーザ測域センサの追跡の枠組みへの統合

以下にそれぞれのフェーズについて詳細を述べる．

追跡初期化への統合

レーザ測域センサによる観測から得られた人物頭部候補位置の近傍に既追跡対象がない場合，人物頭部候補を新規追跡対象として追跡を開始する．追跡初期化時は，レーザ測域センサにより検出された人物頭部候補位置 x_t, y_t の周辺に，向き θ_t を一様に分布させ，センサを設置した高さ z_t から， Z 軸方向に大きくばらつきをとった仮説群を生成する．そして，識別器により評価された全仮説の累積尤度が一定のしきい値以上であれば，追跡を開始する．

図 3.7 に，追跡初期化の様子を示す．人物が立ち上がり，頭部が計測面の高さに到達すると，レーザ測域センサにより捕捉される (図 3.7(a))．次に，人物頭部候補位置の近傍に既追跡対象がない場合，追跡初期化のための仮説が生成される (図 3.7(b))．そして，識別器により評価された全仮説の累積尤度が一定のしきい値以上であれば，追跡が開始される (図 3.7(c))．



(a) 人物の捕捉

(b) 仮説の散布

(c) 追跡の開始

図 3.7 レーザ測域センサによる観測に基づく追跡の開始

追跡への統合

既追跡対象の近傍にレーザ測域センサによる観測から得られた人物頭部候補位置がある場合，オプティカルフローに基づいた状態遷移モデル $p_{opt}(\mathbf{x}_t | \mathbf{x}_{t-1})$ によって伝播させた分布と，レーザ測域センサによる観測に基づいた分布 $p_{laser}(\mathbf{x}_t | \mathbf{y}_t)$ を次のように混合した分布 $q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}_t)$ から仮説のサンプリングを行うことで，レーザ測域センサによる観測を追跡に統合する．

$$\begin{aligned} q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}_t) \\ = (1 - \alpha)p_{opt}(\mathbf{x}_t | \mathbf{x}_{t-1}) + \alpha p_{laser}(\mathbf{x}_t | \mathbf{y}_t) \end{aligned} \quad (3.9)$$

ここで， α は混合比パラメータであり，レーザ測域センサによる観測が得られない場合は $\alpha = 0$ とすることで，カメラ画像のみを用いた追跡となる．

図 3.8 に混合分布からのサンプリングの概念図を示す．

オプティカルフローに基づいた状態遷移モデルによって伝播させた分布 $p_{opt}(\mathbf{x}_t | \mathbf{x}_{t-1})$ は 2.5.2 節と同様に，仮説 $\mathbf{s}_{t-1}^{(n)} = [x_{t-1}^{(n)} \ y_{t-1}^{(n)} \ z_{t-1}^{(n)} \ \theta_{t-1}^{(n)}]^\top$ を $\mathbf{s}_t^{(n)}$ に次のように伝播することで得る．

$$\mathbf{s}_t^{(n)} = \mathbf{s}_{t-1}^{(n)} + \mathbf{v}_t + \boldsymbol{\omega}. \quad (3.10)$$

ここで， \mathbf{v}_t はオプティカルフローから予測された三次元空間における変位ベクトルであり， $\mathbf{v}_t = [\dot{x}_t \ \dot{y}_t \ \dot{z}_t \ \dot{\theta}_t]^\top$ である (ただし $\dot{\theta}_t = 0$)．これは，時刻 t において，2つのカメラの画像上で，人

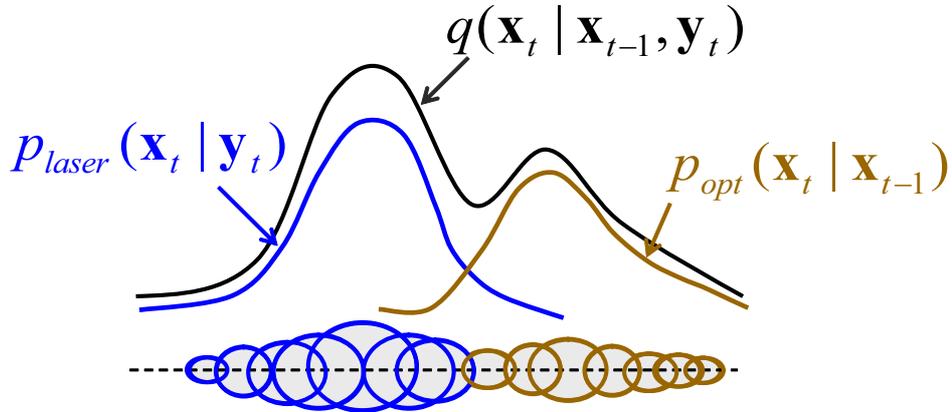


図 3.8 混合分布からの仮説のサンプリング

物頭部周辺のオプティカルフローにより予測された画像座標 $(u_{cam1,t}, v_{cam1,t}), (u_{cam2,t}, v_{cam2,t})$ を三次元空間に逆投影し、時刻 $t - 1$ で推定された人物頭部の三次元位置と差分をとることで得る。 ω は平均 0 、共分散行列 Σ_ω を持つガウス雑音であり、 Σ_ω は分散 $\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_\theta^2$ を対角要素に持つ。ただし、視野を共有するカメラがない場合は、 $\dot{x}_t = 0, \dot{y}_t = 0, \dot{z}_t = 0, \dot{\theta}_t = 0$ とする。

レーザ測域センサにより検出された人物頭部候補位置に基づく分布 $p_{laser}(\mathbf{x}_t | \mathbf{y}_t)$ は、計測位置の誤差が正規分布に従うと仮定して以下のように定義する。

$$p_{laser}(\mathbf{x}_t | \mathbf{y}_t) \approx N(\mathbf{x}_t | \mathbf{y}_t, \Sigma_\omega) \tag{3.11}$$

ここで $N(\mathbf{x}_t | \mathbf{y}_t, \Sigma_\omega)$ は、 \mathbf{x}_t の確率密度分布が平均ベクトル \mathbf{y}_t 、共分散行列 Σ_ω の多変量正規分布に従うことを表す。ただし、 \mathbf{y}_t の状態量 z_t, θ_t は $z_t = z_{t-1}, \theta_t = \theta_{t-1}$ とする。

混合分布からサンプリングを行う場合の仮説の更新の様子を図 3.9 に模式的に示す。

追跡終了への統合

既追跡対象の近傍でレーザ測域センサにより人物頭部候補位置が検出されなくなった場合で、識別器により評価された全仮説の累積尤度が一定のしきい値より低くなった場合に、追跡が失敗したとみなし、追跡を終了する^{*5}。

また、追跡対象同士が空間的に非常に近い位置となった場合も、一方の追跡を終了する。これは、複数の人物頭部は空間内の同じ位置に存在しないことに基づき、同一人物の頭部を複数の追跡器で追跡することを防ぐために行う。

^{*5} 有効仮説数を評価する指標により追跡失敗を判断する手法 [27] など提案されているが、本研究では、仮説の評価に識別器を用いていることから、累積尤度による判断を行っても、十分頑健に動作する。

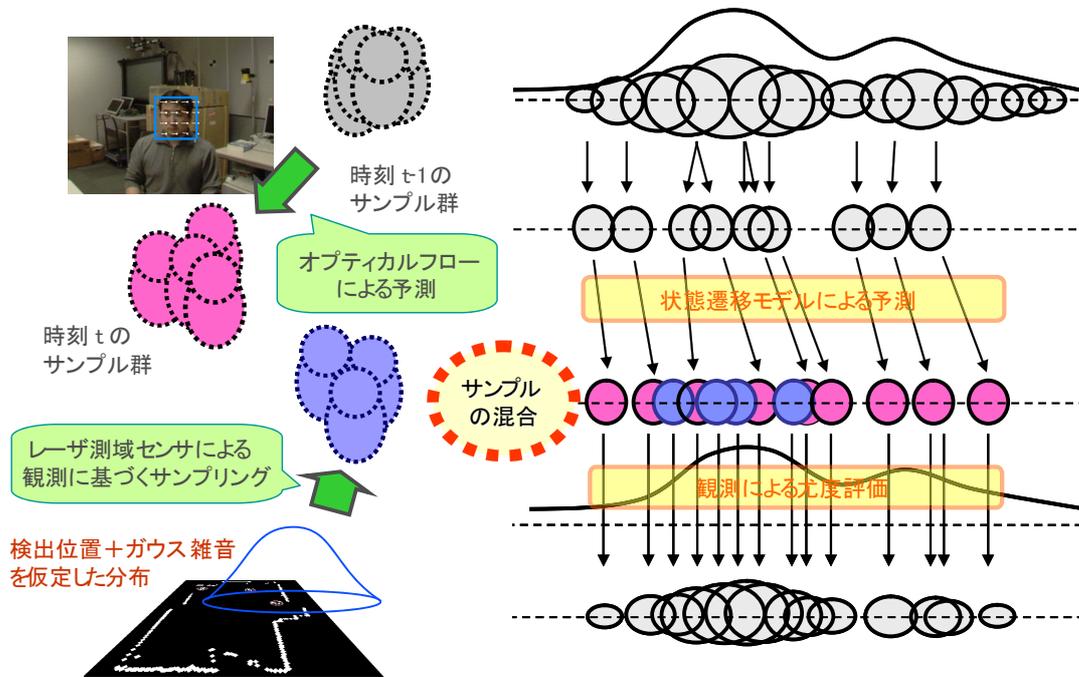


図 3.9 混合分布からのサンプリングによる仮説の更新

3.5 実験

これまでの提案手法に基づき、およそ $5\text{m} \times 5\text{m}$ の床上を観測エリアとし、室内天井に IEEE1394 カラーカメラ (Point Grey Research 社製 Flea) を設置して実験を行った。校正済みのカメラ映像は 20fps で録画し、追跡処理はオフラインで行った。人物一人を追跡するための仮説は 250 個とした。

3.5.1 エリア内における複数人物の追跡

3 人の人物がエリア内に出現し、頭部の向きを変えながらエリア内を通過した。この実験では、視野を共有した 2 台のカメラとレーザ測域センサを用いた。また、混合比パラメータは $\alpha = 0.2$ とした。この実験では、人物頭部は 1 フレーム (50ms) で平均 5.1cm、最大 16.5cm 移動した。また、カメラ画像では平均 2.8 ピクセル、最大 7.1 ピクセルの変位が観察された。実験の様子を図 3.10 に示す。



図 3.10 複数人物の追跡の様子

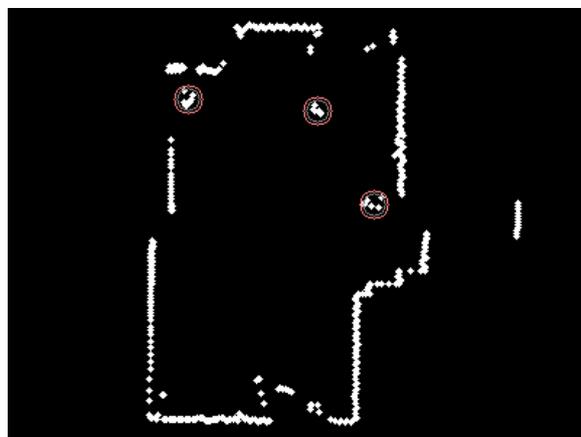


図 3.11 検出された人物頭部候補

図 3.10 の人物頭部の矩形は追跡結果を重ねて表示したものである．また，矩形の色は人物頭部の向きを表している．図 3.11 はレーザ測域センサによる人物頭部候補位置の検出の様子である．図 3.10 より，複数の人物を，広い観測範囲においても追跡できていることがわかる．また，識別器に入力される画像領域が 20×20 ピクセルと小さい場合でも，追跡を継続することができた．

非カスケード型識別器による追跡

仮説の評価に非カスケード型識別器を用いた場合の人物頭部位置の推定結果と真の位置の XY 平面及び XZ 平面での軌跡を図 3.12 に示す．

また，提案手法による追跡の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して求めた三次元座標を真の位置と見なし，推定結果と併せて示した．図 3.12 より，室内の人物の出現と消失を捕らえ，複数の人物頭部を追跡できていることがわかる．また， Z 軸方向及び XY 平面上での平均誤差及び標準偏差を表 3.2 に示す． XY 平面上の平均誤差は 6cm 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 程度に対応する領域が多く含まれることを考慮すると，高い精度で追跡できている．

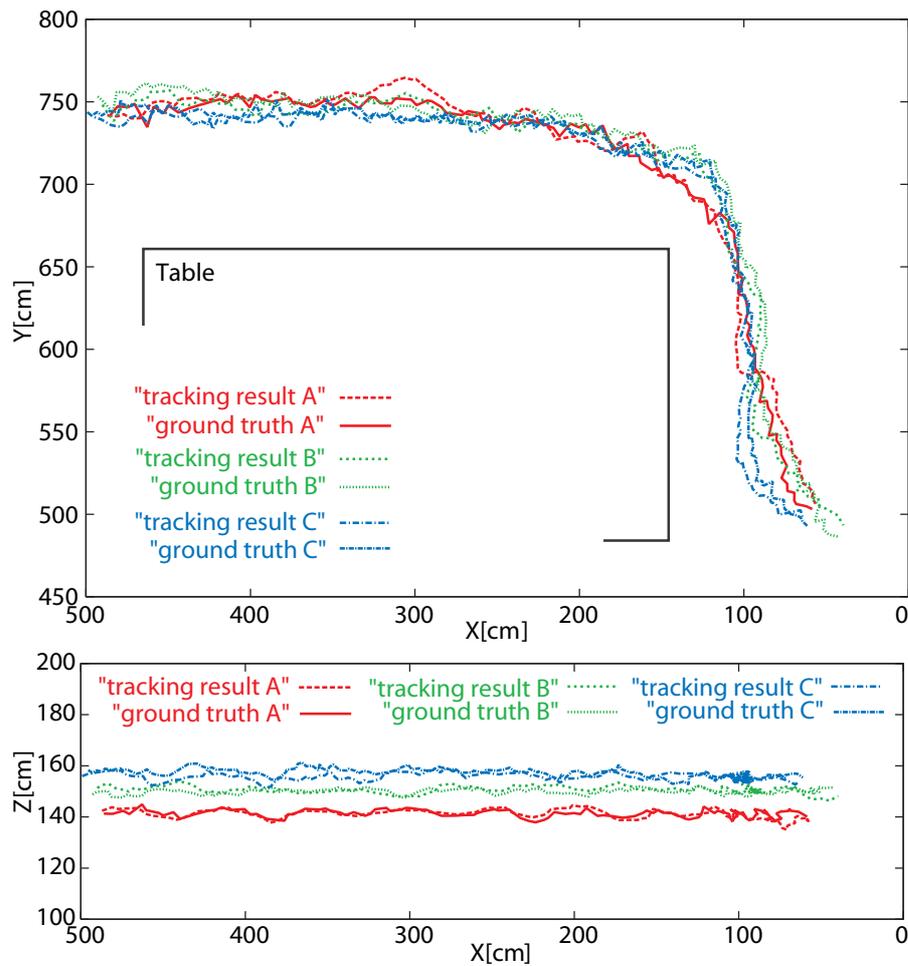


図 3.12 複数人物頭部の追跡結果の軌跡 (非カスケード型識別器)

表 3.2 複数人物頭部の追跡における検出誤差 (非カスケード型識別器)

		平均 [cm]	標準偏差 [cm]
人物 A	Z 軸方向	1.37	1.10
	XY 平面	6.68	4.35
人物 B	Z 軸方向	1.41	1.03
	XY 平面	6.27	2.69
人物 C	Z 軸方向	1.58	1.37
	XY 平面	5.68	2.69

カスケード型識別器による追跡

2 章での議論を踏まえ、仮説の評価にカスケード型識別器を用いた場合の人物頭部位置の推定結果と真の位置の XY 平面及び XZ 平面での軌跡を図 3.13 に示す。

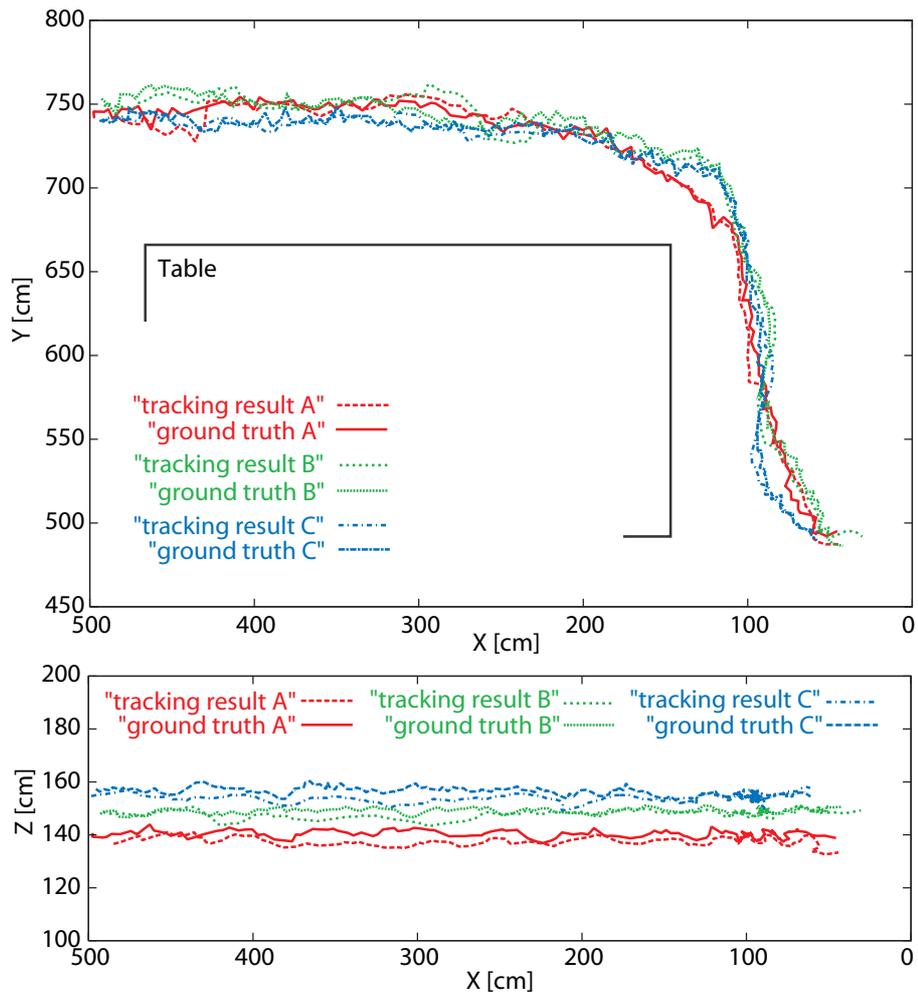


図 3.13 複数人物頭部の追跡結果の軌跡 (カスケード型識別器)

図 3.13 より，カスケード型識別器を用いた場合でも，室内の人物の出現と消失を捕らえ，複数の人物頭部を追跡できていることがわかる．また， Z 軸方向及び XY 平面上での平均誤差及び標準偏差を表 3.3 に示す． XY 平面上の平均誤差は 5cm 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 程度に対応する領域が多く含まれることを考慮すると，高い精度で追跡できている．

表 3.3 複数人物頭部の追跡における検出誤差 (カスケード型識別器)

		平均 [cm]	標準偏差 [cm]
人物 A	Z 軸方向	2.59	1.73
	XY 平面	5.08	3.29
人物 B	Z 軸方向	2.14	1.60
	XY 平面	5.63	3.99
人物 C	Z 軸方向	3.77	2.29
	XY 平面	5.03	3.08

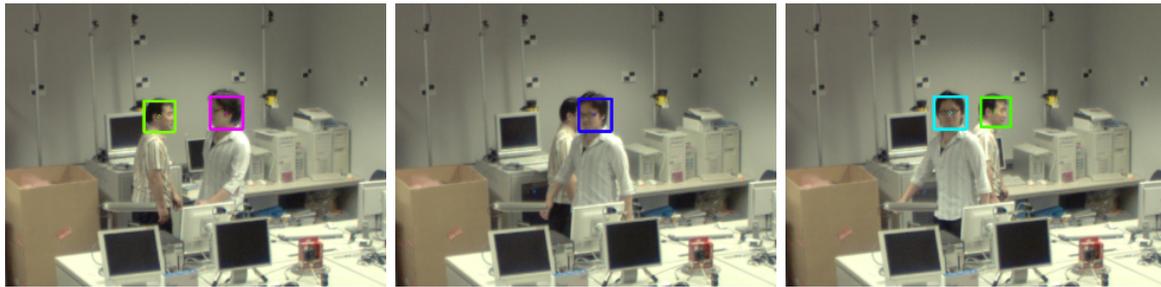
図 3.12, 図 3.13 より，エリア内の人物の出現と消失を捕らえ，複数の人物頭部を高精度に三次元追跡できていることがわかる．3 人の人物には身長差があるが，頑健に追跡初期化が行われており，図 3.12(下)，図 3.13(下) の追跡結果にも身長差が観察できる．

さらに，本手法では，追跡と同時に頭部の向きを推定を行うことが可能である．レーザ測域センサによる観測のみでは，このような身長差や頭部の向きは計測できない．

3.5.2 人物相互の遮蔽

二人の人物がすれ違い，一方の人物が遮蔽される場合の実験を行った．この実験では，視野を共有した 2 台のカメラとレーザ測域センサを用いた．また，混合比パラメータは $\alpha = 0.2$ とした．

図 3.14 において、二人の人物がすれ違い、人物 B が人物 A によって一時的に遮蔽される。追跡結果の軌跡から、人物 B は、人物 A に遮蔽されている間追跡されていないが、人物 A の影から出現し、レーザ測域センサによる検出とカメラ画像による良好な観測が得られた時点で速やかに初期化され、追跡が頑健に再開されていることが分かる。このように、本手法では、人物相互の遮蔽などにより追跡が一時的に失敗した場合でも、レーザ測域センサによる観測に基づいた頑健な追跡初期化により、速やかな追跡の再開が可能である。



(a) #800

(b) #823

(c) #835

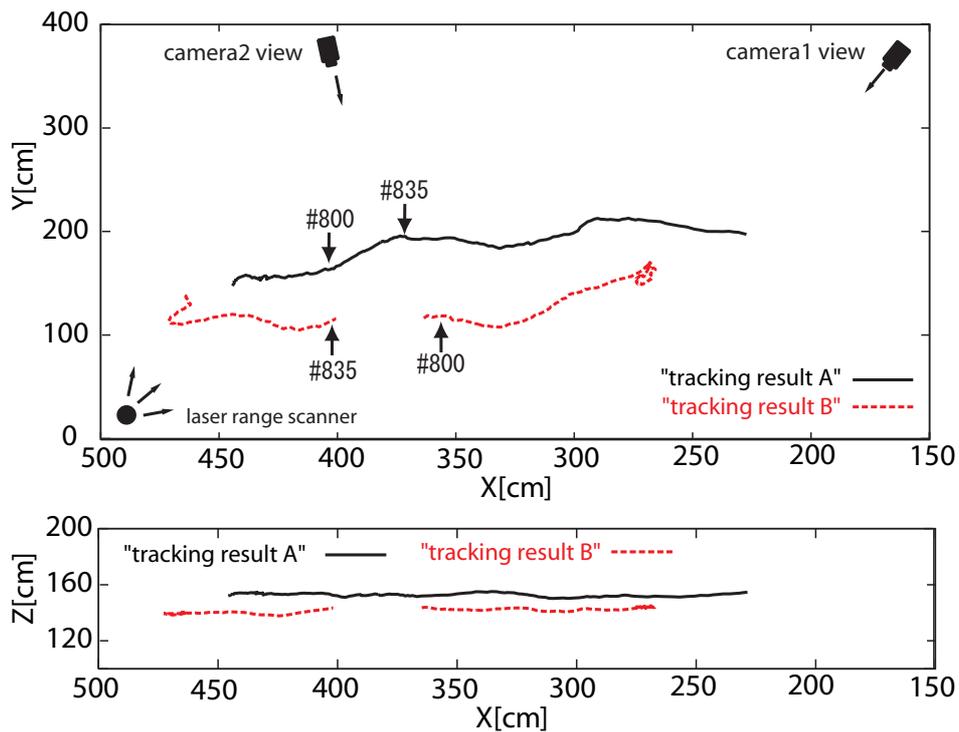
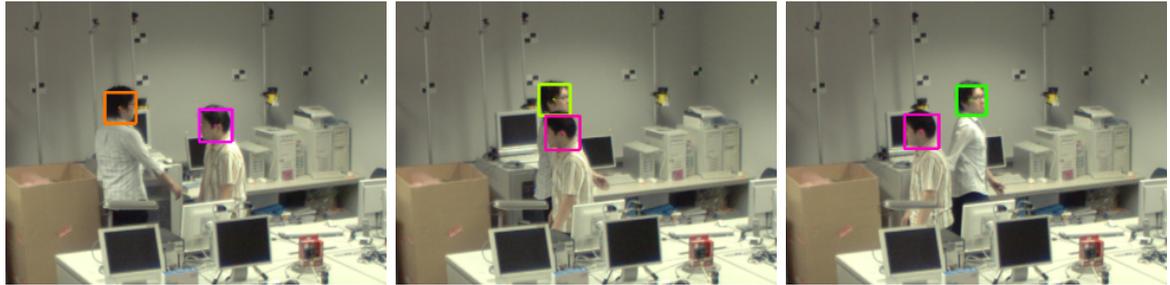


図 3.14 人物相互遮蔽 (カメラ画像上での頭部の遮蔽)

図 3.15 では、二人の人物はすれ違うが、身長差によりカメラ画像上で頭部の遮蔽が起きない。この場合、識別器の評価値が低下しないため、追跡が継続できる。また、実験では、二人の人物が交差する際に、人物 B のレーザ測域センサによる観測が、人物 A に遮蔽され得られない状況であった。しかし、視野を共有した 2 台のカメラによる観測により人物 B の追跡は継続される。このように、カメラ画像による評価に基づいて追跡を行うことから、レーザ測域センサによる観測が得られない場合でも追跡を継続できる。



(a) #1060

(b) #1082

(c) #1095

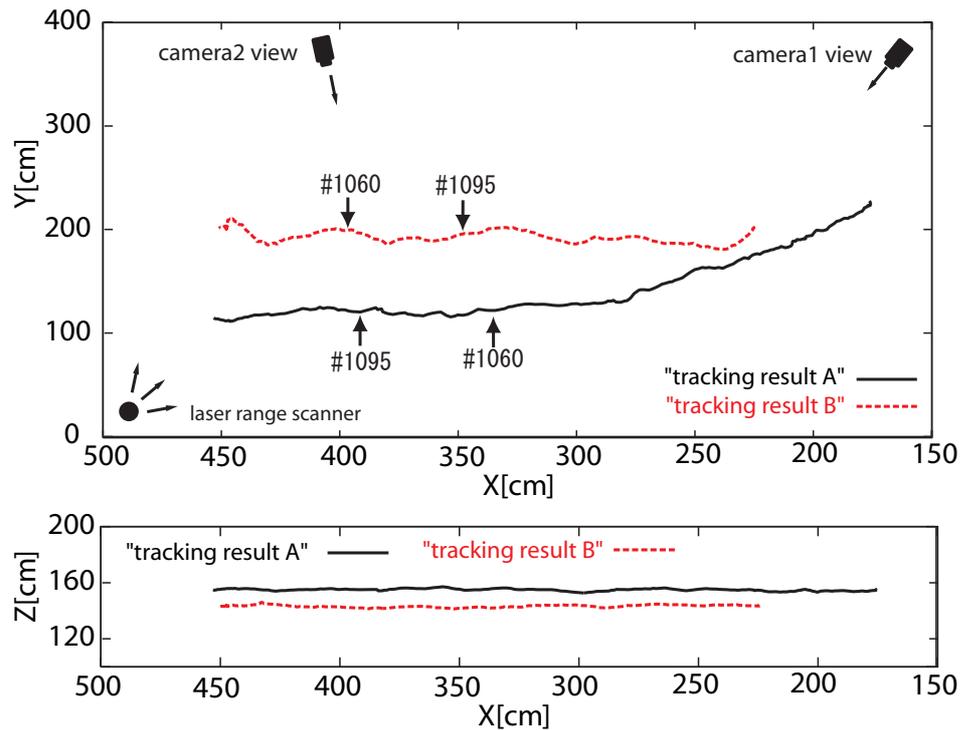


図 3.15 人物相互遮蔽 (レーザ測域センサによる観測の遮蔽)

3.5.3 単眼カメラによる三次元人物頭部追跡

視野を共有した複数のカメラによる観察が困難である場合を想定し、単眼カメラとレーザ測域センサの組合せによる三次元人物頭部追跡実験を行った。この実験では、混合比パラメータを $\alpha = 0.5$ とした。

図 3.16 に単眼カメラとレーザ測域センサを用いて追跡した人物頭部の軌跡を示す。また、視野を共有した2台のカメラとレーザ測域センサを用いた場合、単眼カメラでレーザ測域センサを用いない場合、の追跡結果も併せて示した。

図 3.16 より、単眼カメラの場合でも、レーザ測域センサによる観測を統合することで、視野を共有した2台のカメラを用いた場合に近い良好な三次元追跡が実現できていることが分かる。また、レーザ測域センサによる観測を利用しない場合、人物頭部の画像上の大きさのみを手掛かりにカメラからの距離が推定されるため、カメラ光軸方向に大きな誤差が生じている。このように、本手法では、レーザ測域センサによる観測を追跡に統合することで、単眼カメラでも安定な三次元追跡が可能である。

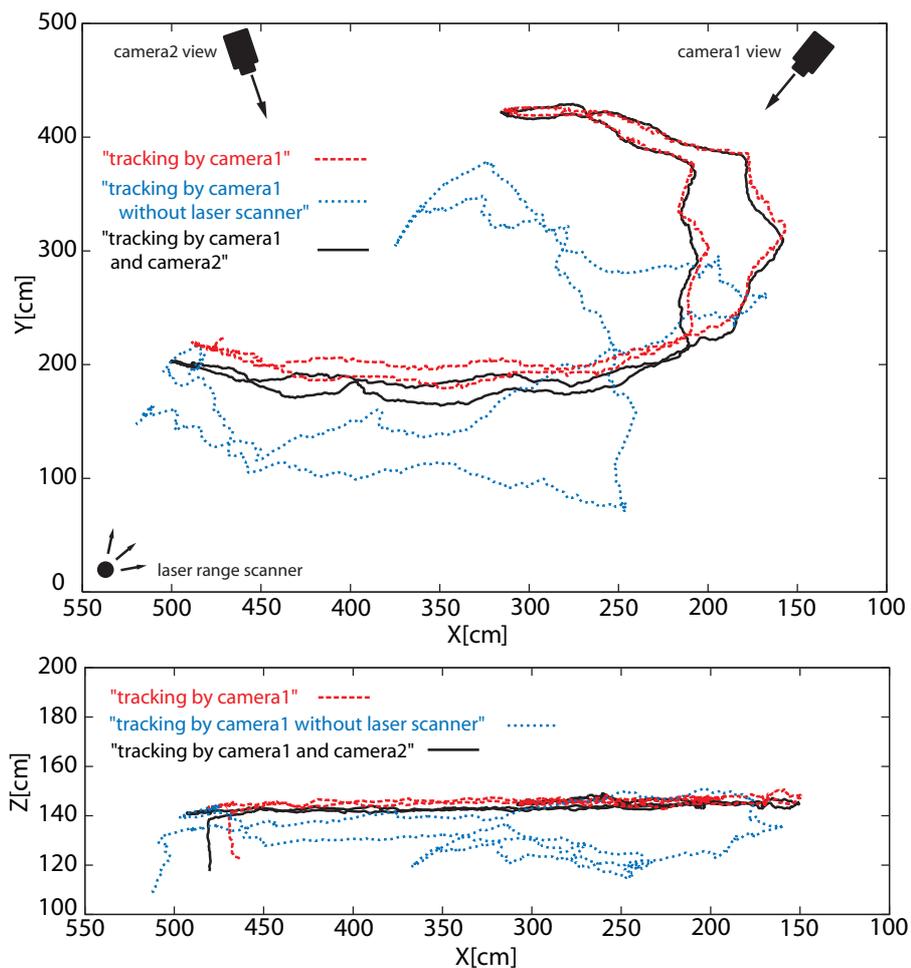


図 3.16 単眼カメラによる三次元人物頭部追跡

3.6 考察

3.6.1 レーザ測域センサによる追跡初期化

本手法では，レーザ測域センサによる観測に基づいて，観測領域内での人物の出現を検出し，追跡初期化を頑健に行うことができる．これは，検出と追跡を同じカメラ画像に基づいて行う場合に比べ，多くの利点がある．

画像の網羅的な精査

本研究では，三次元位置を状態量に持つ人物頭部の仮説をカメラ画像に射影することで，カメラ画像全体を網羅的に精査することを避けている．そのため，追跡初期化のために，カメラ画像を網羅的に精査して人物頭部の検出を行うことは，効率が悪いことは明らかである (図 3.17)．

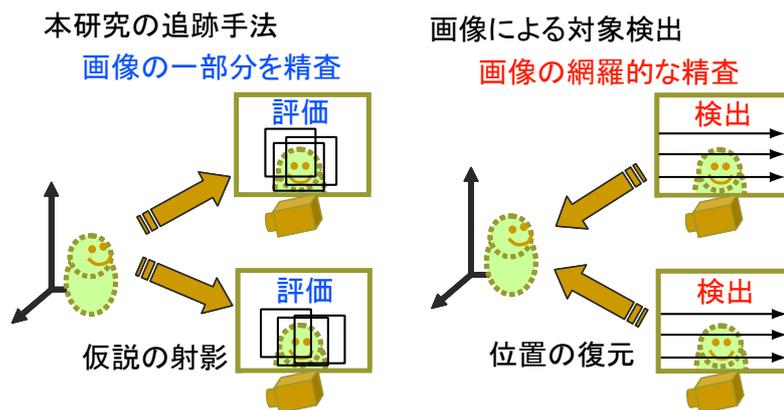


図 3.17 追跡手法と画像による対象検出

各カメラ画像上での複数検出対象の対応付け

視野を共有した複数のカメラ画像に基づいて，複数追跡対象の検出を行う場合，追跡対象の三次元位置を同定するために，各カメラ画像で検出された対象をカメラ間で対応付ける必要がある．観測空間中の対象がもれなく各カメラで検出された場合でも，この対応付けは難しいが，各カメラにおいて検出された対象に過不足がある場合には，さらにこの対応付けが困難となる．文献 [13] では，この対応付けを行うために，ベイジアンネットワークを用いており，煩雑な処理が必要となる．また，文献 [45] では地平面を手掛かりに対応付けを行っているが，屋内環境では障害物により床面が観察できない場合も多い．さらに，実環境においては，そもそも対象の頑健な検出が難しいことを考えると，レーザ測域センサによる対象検出は実用性が高い．

空間分割による検出

検出においても三次元位置を仮定し，仮定された三次元位置をカメラ画像に射影し，変化を観察することで，追跡対象を検出することも考えられる．この考えに基づいて，視野を共有したカメラ

で観測可能な空間を一辺 25cm の立方体で分割し，一般的な背景差分^{*6}により変化が観察された領域のボリューム最上部を人物頭部候補位置として検出することを試みた．図 3.18 は空間を分割した立方体をカメラ画像に投影した様子である．

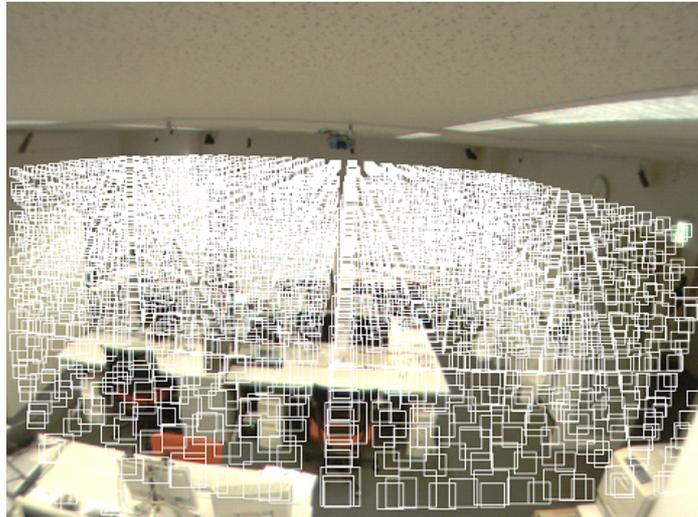


図 3.18 空間を分割した立方体グリッドの画像への射影

この立方体グリッドを複数のカメラで観察し，射影された画像領域の半分以上に差分が検出された場合に，分割された空間に変化があったと見なして，差分ボリュームを作成した^{*7}．そして，作成されたボリュームの最上部を人物頭部候補位置とした．図 3.19(左) は画像上で立方体グリッドの変化を検出している様子である．また，図 3.19(右) は変化のあった立方体グリッドをワイヤースケルトンで表示したものである．最上部の立方体グリッドが人物頭部候補位置として赤く表示されている．



図 3.19 立方体グリッドの観察に基づいた頭部検出

このようにして検出した人物頭部候補位置に基づいて追跡を試みた (図 3.20) ．

^{*6} [68] や [98] に基礎的な画像処理手法がまとめられている．

^{*7} これは古典的な視体積交差法として知られる Space Carving に似た処理である [5, 87, 108] ．

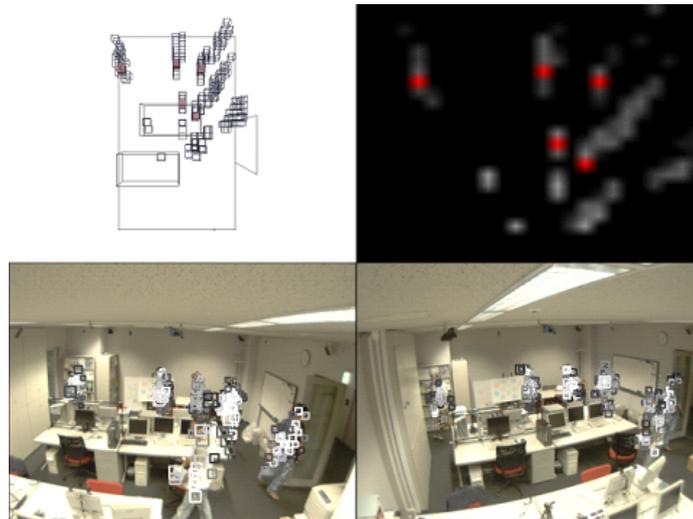


図 3.20 立方体グリッドの観察に基づいた追跡

しかし、この検出手法では追跡初期化に失敗するケースが多く見受けられた。主な原因として人物頭部候補位置の検出精度の低さが挙げられる。グリッドは一辺 25cm の立方体であるため、検出された人物頭部候補位置グリッドが一つずれると、人物頭部ひとつ分の誤差が生じることとなり、追跡の開始は難しい。

これに対しては、立方体グリッドを小さくすることも考えられるが、25cm よりも小さい立方体グリッドでは、背景差分が頑健でないことから、近傍の立方体グリッドで誤検出が生じるため、差分ボリュームが安定に検出できなかった*⁸。

また、複数の人物により、カメラ画像上で差分領域が複数観察されるようになると、差分領域の誤対応により、虚像ボリュームが生成される場合があった。文献 [90] では、観測履歴情報に基づいて、空間中に人物検出用の仮説を効率的に分布させる手法が提案されているが、この虚像ボリュームと同じ問題が残されていると考えられる。

これらのことから、人物頭部候補位置の検出にレーザ測域センサを用いるメリットは大きい。

3.6.2 追跡対象の部分的な遮蔽

追跡対象領域を分割し、分割された領域をそれぞれ追跡することで、部分的な遮蔽の影響を軽減する手法が提案されている。例えば、Wu ら [109] は、静止画像において、人物輪郭の一部を検出する Adaboost 識別器の検出結果の組合せを MAP 推定の枠組みで考え、人物領域の一部で遮蔽がある場合でも、人物の検出を可能としている。

これに対して、本手法では、明示的に部分遮蔽を扱っていないが、監視カメラ映像を想定し、人物頭部を追跡対象としていることから、胴体領域の遮蔽には影響を受けない。また、頭部画像領域のわずかな遮蔽は、識別器による評価で吸収されるため、部分遮蔽の影響は最小限に抑えられている。

*⁸ 画面の変化を検出するための高度な背景差分法も提案されている（例えば文献 [80] など）が、一般的な監視カメラの俯角では、差分ボリュームのあいまい性が大きく、高精度な差分ボリュームの取得は難しい。

3.7 まとめ

本章では、Mixture Particle Filter の枠組みに基づいて、分散カメラとレーザ測域センサを相補的に統合した人物追跡手法を提案した。レーザ測域センサによる観測に基づいて、エリア内での人物の出現を検出することで、頑健な追跡初期化を可能とした。これにより、人物同士がすれ違うなどの遮蔽により追跡が失敗した場合でも、良好な観測が得られた時点で、速やかに追跡を再開することができた。また、オプティカルフローに基づいた状態遷移とレーザ測域センサによる観測に基づく分布を混合した確率密度分布から仮説を生成することで、レーザ測域センサを追跡に統合し、単眼カメラを用いた場合でも良好な三次元追跡を実現した。これにより、分散配置されたセンサ群のさまざまな組合せによる三次元人物追跡を可能とした。

第4章

複数観測領域間の人物動線の 対応付け

4.1 はじめに

監視カメラ普及を背景に，安全・安心な社会環境の実現のため，監視カメラの活用への期待は大きい．一方で，プライバシーの問題や，カメラの物理的な配置の制限，障害物による遮蔽などの問題により，人物の行動すべてをカメラなどのセンサで直接観察することは困難である場合が多い．このような状況で，人物の行動を広範囲にわたり計測するためには，環境に疎に分散配置された視野を共有しないカメラ群からの情報を統合し，人物の行動を計測する技術が重要となる．

環境に疎に分散配置されたカメラ群による人物の行動計測には，大きく次の二つの課題がある．一つは，複数のカメラで観察される人物の同一性推定の課題である．これは，複数のカメラで観察された人物が同一人物であるかを，人物が一方のカメラで観察された後に，他方のカメラでどのように観察されるかを手掛りとして推定するものである．もう一つは，カメラで観察できない箇所における人物の行動推定の課題である．これは，カメラで観察されない箇所における人物の行動を，観察された少ない情報から補間し，推定するものである．

視野を共有しないカメラ群による人物の行動計測の実現には，これらの課題を解決する必要があるが，本章では，複数カメラで観察される人物の同一性推定の問題に着目し，カメラによる観測から得られる複数の人物動線に対応付ける手法について述べる．

3章までに述べたように，カメラで観測可能な領域においても，人物相互の遮蔽などにより，追跡が失敗し人物動線が分断されてしまう場合がある．このような分断された動線同士に対応付けることができれば，特定の観測エリア内において一人の人物動線を一本につなぎ合わせることができ，動線分類による行動パターンの認識や，逸脱行動の検出にも応用できると考えられる．

本章では，4.2節にて，疎に分散した観測領域を扱う関連研究について概観する．そして，4.3節にて，複数観測領域間の人物動線に対応付け手法について，Javedらの手法に基づいて詳しく述べる．4.4節では，検証実験について述べ，最後に，4.5節にてまとめる．

4.2 関連研究

視野を共有しないカメラ群を利用した監視システムでは、それぞれのカメラからの情報を独立に扱い、情報の収集や提示に着目したシステムと、カメラ群から得られた情報を統合的に扱うことで、新たな知見を得ようとするものがある。情報の収集や提示に着目したシステムでは CMU の VSAM Project[15] などがよく知られているが、本節では環境に疎密に分散配置されたカメラ群を統合的に利用する観点から、観測領域間を横断した情報を獲得するための課題と関連研究について述べる。

4.2.1 複数カメラの設置状況の推定

実世界におけるカメラの配置状況の推定は、大量かつ広範囲にカメラを配置した場合など、全てのカメラの設置位置や姿勢の計測が困難な場合に有効な技術である。それぞれのカメラの実世界における物理的な位置や姿勢が分からない場合に、カメラで観察される対象から、カメラの位置や姿勢、観察される空間同士の間隔を推定する。

カメラの配置やトポロジを推定する手法として、Makris ら [8, 52] は、それぞれのカメラの位置や姿勢が分からない場合に、カメラで観察される対象から、カメラで観察している空間同士の間隔を推定する手法を提案している (図 4.1)。視野を共有しない複数のカメラで撮影された映像において、自動車などの対象が出現する点、消失する点をノードとして扱い、各ノードにおいて、対象の出現、消失のイベントが、どの程度の頻度で発生しているかを考慮してノード間の間隔を推定し、最終的にカメラのトポロジを推定している。さらに Eric ら [19] は、この考えを統計的依存性に基づいて一般化し、GPS の情報を併用することで、実際の位置関係を推定する手法を示している。また、浮田 [103] は複数のカメラで観察される対象を投票ベースの手法で対応づけ、対象の対応付けとカメラ群の配置を同時に推定する手法を提案している。Rahimi ら [73] はカメラで観察する対象の移動速度などが詳しく分かっている場合に、カメラ画像中で観察される対象の様子からカメラの状態を推定できるとし、天井から下方を観察しているカメラ群の配置を、対象の既知の動きから推定する手法を提案している。

本研究では、人物動線の実世界座標系での計測をめざすことから、複数カメラの設置位置や姿勢は事前の校正により既知であるとして議論する。

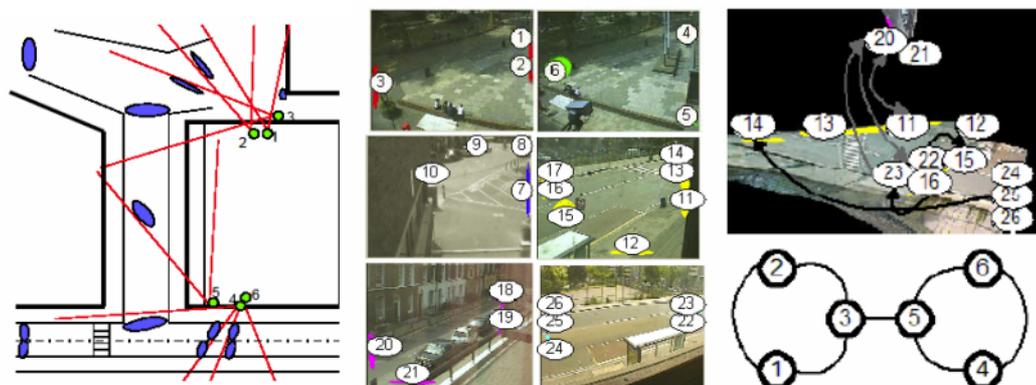


図 4.1 カメラトポロジの推定

4.2.2 複数観測領域間の人物動線の対応付け

複数観測領域間における観測の対応付けは、視野を共有しないカメラ群を利用し、複数の対象を観察する場合に必要な技術である。高速道路の2地点における自動車の観測の対応付け(例えば[82])などでは、車線や車種などの多くの制約条件を考えることができるため、高精度な対応付けが可能である。しかし、人物はエリア内を自由に移動する機会が多いため、観測の対応付けは容易ではない。

池亀ら[36]やJavedら[40, 41]は視野を共有しない複数のカメラで観察される人物を対応づけ、同一性を推定する手法を提案している(図4.2)。推定には、人物の出現・消失時の位置、移動方向・速度、色、及びカメラ間の平均移動時間などの情報を用いており、同一人物であることがあらかじめ分かっている状況で事前に学習したデータに基づいて、確率的な推定を行っている。また、Javedら[39]は一方のカメラの視野境界が他方のカメラの視野とどのような関係にあるかを利用して、人物を追跡する手法を提案している。これは、一方のカメラの視野境界線を他方のカメラの視野に関係づけ、人物がどの視野境界線を越えてフレームアウトしたかを考慮することで、次にどのカメラに出現するかを予測して対応付けを行うものである。また、Shetら[83]は、対応付けを人物の観測から得られる情報だけではなく、同時に観察されるオブジェクトとの関係(例えば、個人所有の自動車から人物が降りてきた。など)を考慮し、多値デフォルト論理^{*1}を用いて推定する手法を提案している。

本研究では、Javedらの手法[40]に基づき、次節にてこの問題を議論する。

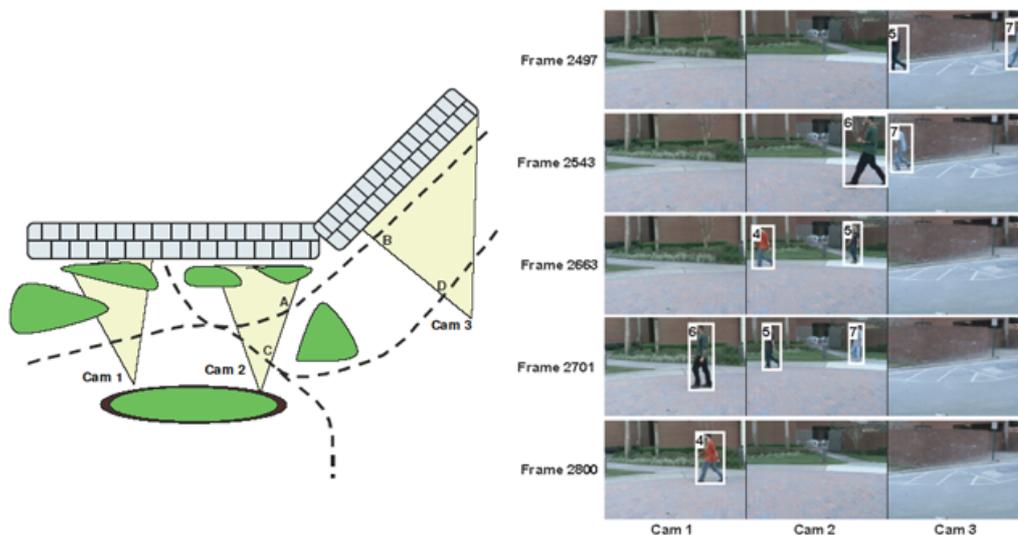


図 4.2 カメラ間における観測の対応付け

^{*1} 常識的なルールを覆す重要な情報が得られない限り、デフォルトと呼ばれる一般法則を適用する推論法。

4.2.3 観測領域外を含む人物動線の推定

カメラで観察されない箇所における対象の振る舞いの推定は、エリア内に観測不能領域がある場合や、より広いエリアを少ないカメラで観測する際に有効となる技術である。

Rahimi ら [74] は屋内に分散配置した視野を共有しないカメラ群を利用して人物の移動軌跡を推定する手法を提案している (図 4.3)。対象の移動速度が既知である場合、カメラ画像中で観察される対象の様子から、対象とカメラの距離が推定でき、2 台のカメラ間の距離が既知である場合、2 台のカメラによる観察によって、対象の移動速度が推定できる。この考えを一般化し、人物の移動がモデル化できる場合に、複数のカメラによる観察から、カメラで観察されない領域を含む人物動線を推定する手法を提案している。しかし、Rahimi らの手法では、二次元の移動軌跡推定のみで、三次元空間における人物の移動軌跡推定は議論されていない。

本研究では、Rahimi らの手法 [74] に基づき、5 章にてこの問題について議論する。

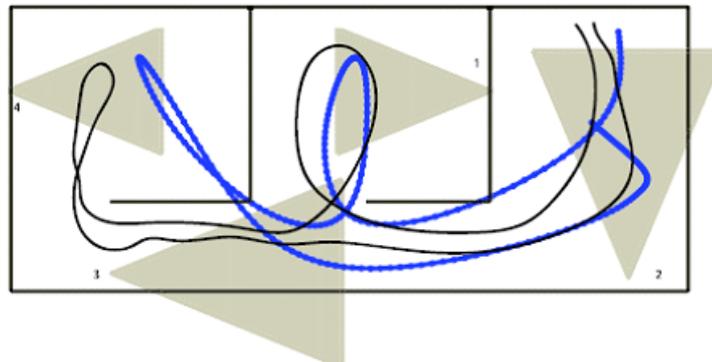


図 4.3 視野外を含む移動軌跡推定

4.3 複数観測領域間の人物動線の対応付け

複数観測領域間の人物動線の対応付けでは、異なる 2 地点で同時に同じ人物が観察されることはないといった事実や、ある程度離れた観測領域間では、その間の旅行時間を予測することができるなどの時間的空間的な尤もらしさ (時空間的尤度) を考えることができる (図 4.4)。また、同じ人物であれば、異なるカメラで観察された場合でも、その見えが大きく異ならないという仮説から、見えの類似性 (見えの尤度) を考えることができる (図 4.5)。

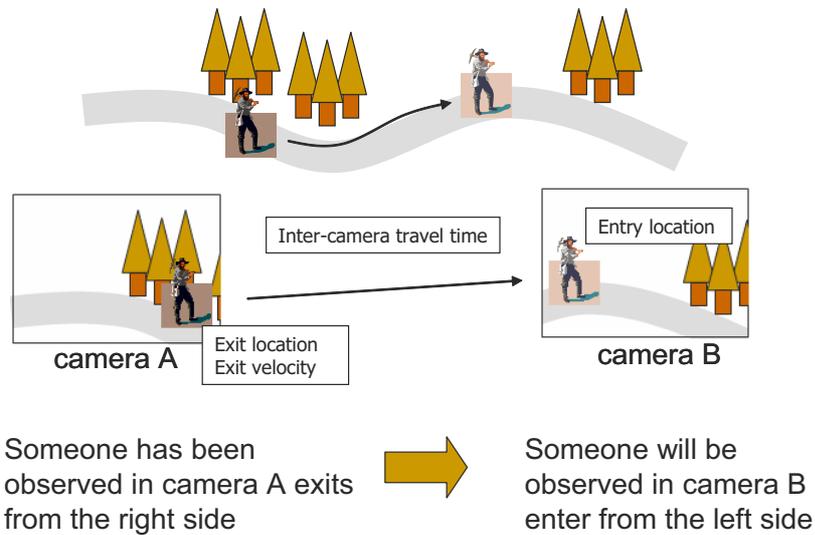


図 4.4 時空間的尤度

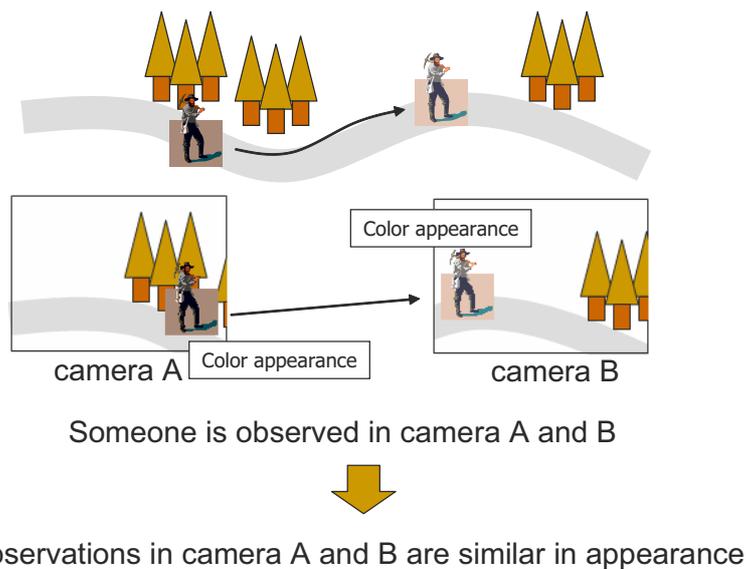


図 4.5 見えの尤度

本節では、このような考えに基づいて、複数観測の同一性推定を Javed ら [40] の手法に倣い定式化する。

カメラを $C_1, \dots, C_i, C_j, \dots, C_n$, 人物動線を $T_1, \dots, T_a, T_b, \dots, T_m$ とし、カメラ C_i で観察された一つの動線 T_a の観測を $O_{i,a}$ とする。同様に、カメラ C_j で観察された一つの動線 T_b の観測を $O_{j,b}$ とし、 $O_{i,a}$ と $O_{j,b}$ が同一人物の観測である仮説を $k_{i,a}^{j,b}$ とすると、 $O_{i,a}$ と $O_{j,b}$ が同一人物である確率密度分布は以下のように表すことができる。

$$p_{i,j}(k_{i,a}^{j,b} | O_{i,a}, O_{j,b}) \quad (4.1)$$

各カメラでの観測は独立であるとし、仮説も互いに独立であるとする、ある観測集合 \mathbf{O} が得られたときの仮説集合 \mathbf{K} の条件付き確率は以下となる。

$$p(\mathbf{K} | \mathbf{O}) = \prod_{k_{i,a}^{j,b} \in \mathbf{K}} p_{i,j}(k_{i,a}^{j,b} | O_{i,a}, O_{j,b}) \quad (4.2)$$

$p(\mathbf{K} | \mathbf{O})$ は、ベイズの定理から、以下のように事前確率と尤度で表すことができる。

$$\begin{aligned} p(\mathbf{K} | \mathbf{O}) &= \prod_{k_{i,a}^{j,b} \in \mathbf{K}} p_{i,j}(k_{i,a}^{j,b} | O_{i,a}, O_{j,b}) \\ &= \prod_{k_{i,a}^{j,b} \in \mathbf{K}} \frac{p_{i,j}(O_{i,a}, O_{j,b} | k_{i,a}^{j,b}) p_{i,j}(k_{i,a}^{j,b})}{p_{i,j}(O_{i,a}, O_{j,b})} \end{aligned} \quad (4.3)$$

これが最大となる仮説集合 $\mathbf{K}^* = \arg \max_{\mathbf{K} \in \Sigma} p(\mathbf{K} | \mathbf{O})$ を求めることで、尤もらしい観測の対応を得ることができる。観測の組の生起確率 $p_{i,j}(O_{i,a}, O_{j,b})$ が一様であるとする、

$$p(\mathbf{K} | \mathbf{O}) \propto \prod_{k_{i,a}^{j,b} \in \mathbf{K}} p_{i,j}(O_{i,a}, O_{j,b} | k_{i,a}^{j,b}) p_{i,j}(k_{i,a}^{j,b}) \quad (4.4)$$

であり、観測 $O_{i,a}$ の特徴量を見えの特徴と時空間的特徴に分けて考えると、式 4.4 は以下のように表すことができる。

$$\begin{aligned} p(\mathbf{K} | \mathbf{O}) &\propto \prod_{k_{i,a}^{j,b} \in \mathbf{K}} p_{i,j}(O_{i,a}(app), O_{j,b}(app) | k_{i,a}^{j,b}) \\ &\quad p_{i,j}(O_{i,a}(st), O_{j,b}(st) | k_{i,a}^{j,b}) \\ &\quad p_{i,j}(k_{i,a}^{j,b}) \end{aligned} \quad (4.5)$$

ここで、 $p_{i,j}(O_{i,a}(app), O_{j,b}(app) | k_{i,a}^{j,b})$ は二つの観測の見えの類似性に基づく尤度であり、 $p_{i,j}(O_{i,a}(st), O_{j,b}(st) | k_{i,a}^{j,b})$ は二つの観測の時空間的整合性に基づく尤度である。また、 $p_{i,j}(k_{i,a}^{j,b})$ は事前確率である。

見えの尤度は、二つの観測の見えが似ている場合に高い尤度となるような指標であり、人物であれば、服装の類似性や顔認証の結果などを考えることができる。時空間的尤度は、ある地点で消失した人物は、ある時間が経過した後特定の場所で出現することが多いという時空間的な尤もらしさを表す尤度である。通路の入口と出口に監視カメラが設置されている場合などでは、入口から出口までの旅行時間が人物の同一性推定の重要な手掛りとなることは、容易に想像できる。

以降では、それぞれの確率密度分布を実際に観測可能な値によって定義し、観測の組み合わせの尤もらしさを評価する関数を構築する。

4.3.1 事前確率

事前確率 $p_{i,j}(k_{i,a}^{j,b})$ は、観測の対応の生起確率を表す。観測領域間の人物の遷移確率に偏りがあれば、それをを用いても良い (図 4.6) が、人物がどちらに移動するか予想できない場合などでは、一定と仮定してよく、本研究でも一定と仮定する。

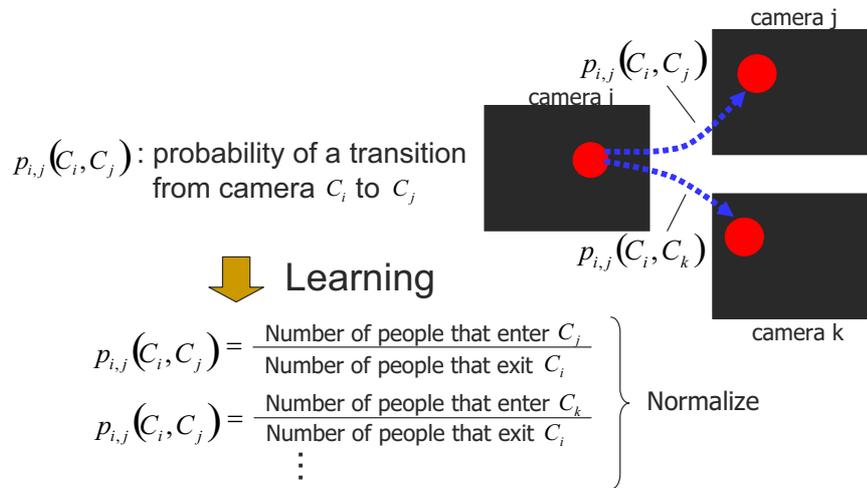


図 4.6 観測領域間の遷移確率による事前確率の定義

4.3.2 時空間的尤度

事前に対応関係が既知の人物動線データを記録しておき、多変量カーネル密度推定^{*2}により、時空間的尤度を推定する。

学習は、同一人物の観測 $O_{i,a}(st)$ と $O_{j,b}(st)$ から得られる学習データを以下のようなベクトルとして記録することで行う。

$$\mathbf{x}_{i,j} = [u_i \ v_i \ \dot{u}_i \ \dot{v}_i \ x_j \ y_j \ t_{i,j}]^T \tag{4.6}$$

ここで、ベクトル $\mathbf{x}_{i,j}$ の要素は、消失位置 (u_i, v_i) 、消失時速度 (\dot{u}_i, \dot{v}_i) 、出現位置 (x_j, y_j) 、旅行時間 $t_{i,j}$ である。

時空間的尤度は、次式を用いてテストデータの近傍で学習データの密度を計算することで推定する (図 4.7)。

$$\hat{p}(\mathbf{x}) = \frac{1}{n} |\mathbf{H}|^{-\frac{1}{2}} \sum_{i=1}^n K(\mathbf{H})^{-\frac{1}{2}} (\mathbf{x} - \mathbf{x}_i) \tag{4.7}$$

ここで、 $K(u)$ はプロダクトカーネルであり、 $\mathbf{H} = \text{diag}[h_1^2 \ h_2^2 \ \dots \ h_7^2]$ はカーネルサイズを表す。

^{*2} Parzen Window(パルツェン窓) 法とも言われる。詳細は文献 [75] に詳しい。

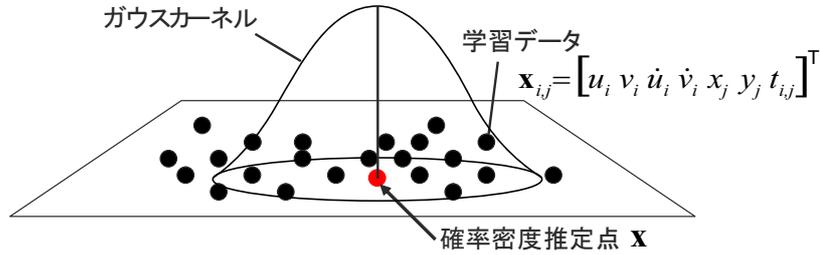


図 4.7 カーネル密度推定による時空間的尤度の推定

4.3.3 見えの尤度

見えの尤度は 2 つの観測 $O_{i,a}(st)$ と $O_{j,b}(st)$ の色ヒストグラムを計算し、ヒストグラム間の Bhattacharyya 距離 $D(a,b)$ を以下のようにガウス分布でスケールング^{*3}して用いる (図 4.8)。

$$p_{i,j}(D(a,b)) = \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \exp\left(-\frac{1}{2}\left(\frac{D(a,b) - \mu_{i,j}}{\sigma_{i,j}}\right)^2\right) \quad (4.8)$$

ガウス分布の平均 $\mu_{i,j}$ と分散 $\sigma_{i,j}^2$ は事前に学習データから取得する。

m 個の bin を持つヒストグラム k, q 間の Bhattacharyya 距離 $D(k, q)$ は以下で表される。

$$D(k, q) = \sqrt{1 - \sum_{i=1}^m \sqrt{\hat{k}_n \hat{q}_n}} \quad (4.9)$$

ここで、 \hat{k}_n, \hat{q}_n は n 番目の bin の値を表す。

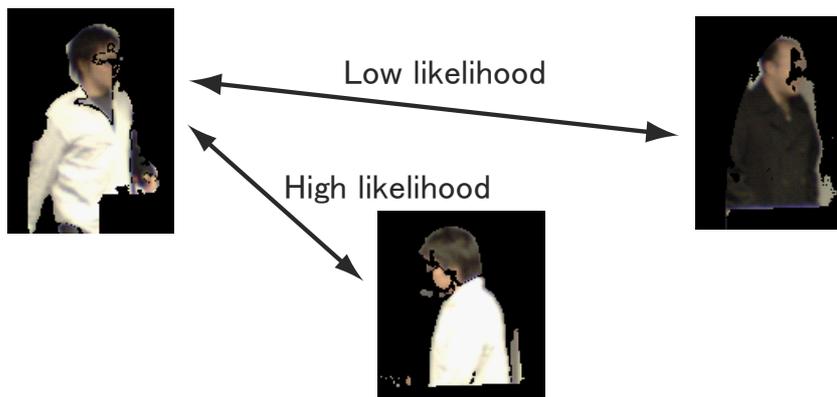


図 4.8 ヒストグラム距離に基づく見えの尤度の推定

^{*3} 同じ人物を複数の観測領域で観察すると、観測間の色ヒストグラムの Bhattacharyya 距離がガウス分布に従う状況を仮定しているが、このような状況は限定的である。文献 [41] では、見えの尤度推定を高度化し、観測領域間の照明条件の違いを学習により獲得する手法が提案されている。

4.3.4 グラフによるモデル化

事前確率，時空間的尤度，見えの尤度を統合し，観測の組み合わせの尤もらしさを評価する関数を以下のように構築する．

$$\mathbf{K}^* = \arg \max_{\mathbf{K} \in \Sigma} \sum_{k_{i,a}^{j,b} \in \mathbf{K}} \log(p_{i,j}(O_{i,a}(app), O_{j,b}(app) | k_{i,a}^{j,b}) \tag{4.10}$$

$$p_{i,j}(O_{i,a}(st), O_{j,b}(st) | k_{i,a}^{j,b}) p_{i,j}(k_{i,a}^{j,b}))$$

これを最大化するため，観測の対応を有向グラフでモデル化する．観測をノード，対応付けの仮説をエッジとし，評価関数の値をコストとして割り当てる (図 4.9) ．

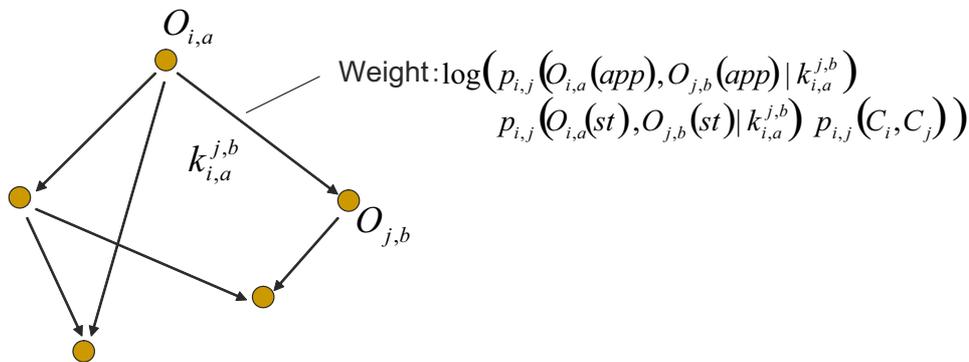


図 4.9 有向グラフによるモデル化

この有向グラフは，重み付き 2 部グラフに変換することができ，重み付き 2 部グラフの最大マッチング^{*4}を求めることで，複数観測領域間における人物動線の対応付けを得ることができる．この手順を図 4.10 に示す．

図 4.10(a) は，真の観測対応の有向グラフである．この真の対応付けを重み付き 2 部グラフの最大マッチングにより求めたい．そのため，まず観測対応を有向グラフによりモデル化し，次に重み付き 2 部グラフへ変換する．

図 4.10(b) は，各カメラで得られた観測に ID を割当てたものである．次に，観測対応の仮説を有向グラフのエッジとして記述し (図 4.10(c))，エッジにコスト (重み) を割当てる (図 4.10(d)) ．さらに，カメラに入ってくる観測とカメラから出て行く観測に分けて考えるため，一つの観測を二つのノードに分けて新たに ID を割当てる (図 4.10(e)) ．この変換により，有向グラフを無向グラフで考えることができ (図 4.10(f))，カメラ ID も考える必要が無くなる (図 4.10(g)) ．このような変換により，最終的に観測対応の重み付き 2 部グラフが得られる (図 4.10(h)) ．

^{*4} 重み付き 2 部グラフの最大マッチングは，Edmonds の手法 [21] などで行うことができる．

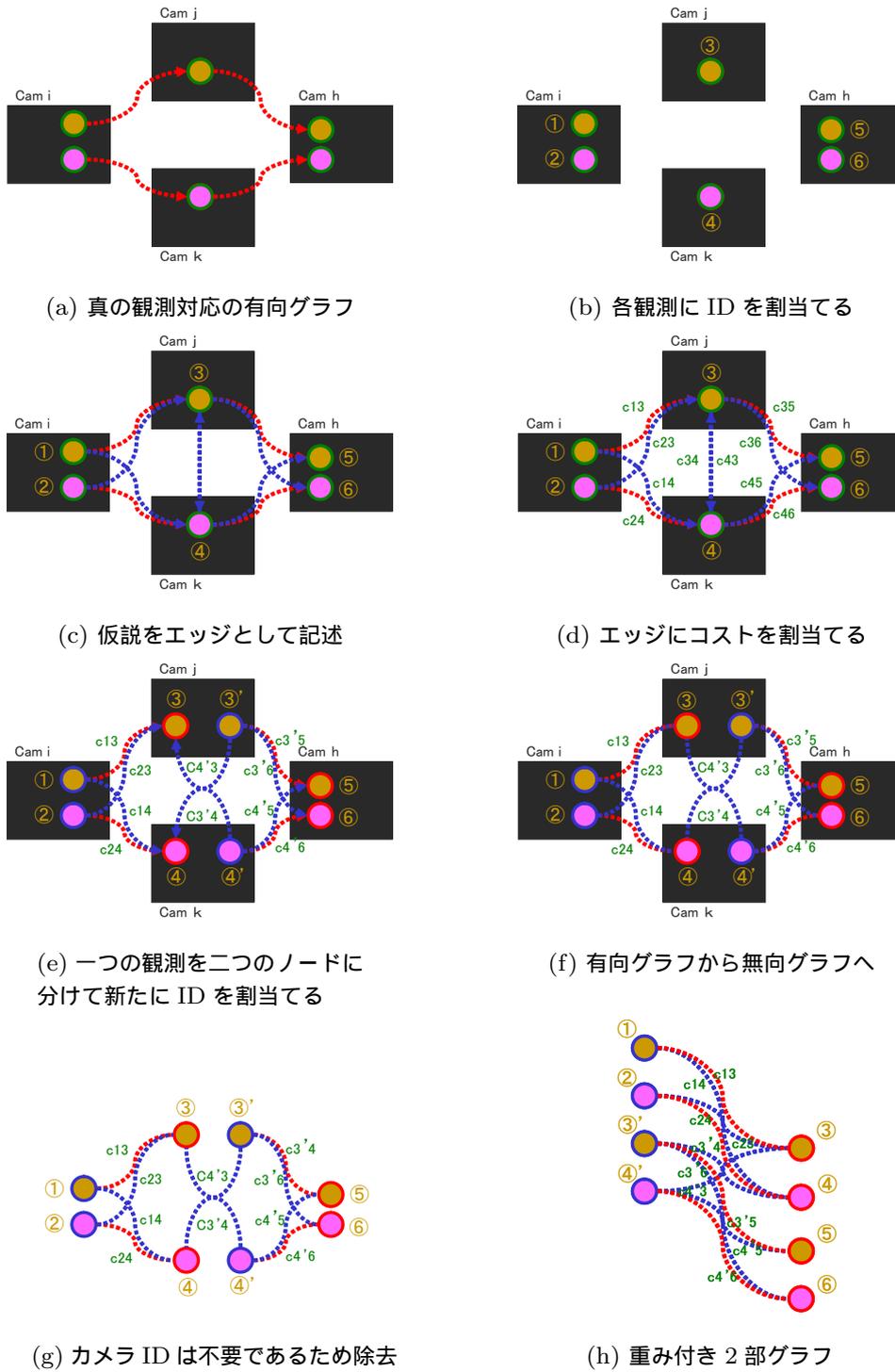


図 4.10 重み付き 2 部グラフによる観測対応のモデル化

4.4 実験

4.4.1 シミュレーション

シミュレータ

予備実験として、PC上で動作するシミュレータ(図4.11)を開発してシミュレーションを行った。通路に障害物が置かれた状況を想定し、人物を模擬した球形のCGオブジェクトをランダムに通過させた。シミュレータでは、あらかじめ設定した観測領域の映像をファイルに録画することができる。シミュレーションによる実験では、図4.11に示したように視野を共有しないように3箇所の観測領域 cam1, cam2, cam3 を設定した。

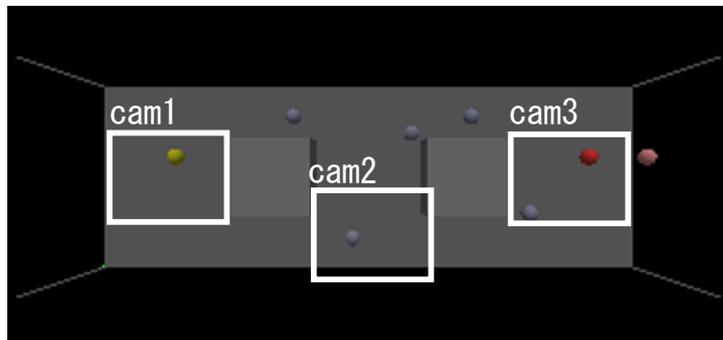


図 4.11 シミュレータ

オブジェクトが通過する経路は、図4.12に示したA,B,C,Dの4通りであり、いずれの経路を通過するかは、出発時にランダムに決定される。また、各オブジェクトの移動速度もランダムに設定されるため、観測するカメラによりオブジェクトの観測順序が異なる状況も模擬される。また、オブジェクトは図4.12(左)のような見えを持つ。いずれの見えとなるかは、出発時にランダムに選ばれる。

このシミュレータを用いて録画した映像に基づいて、CGオブジェクトの対応付けを行った。

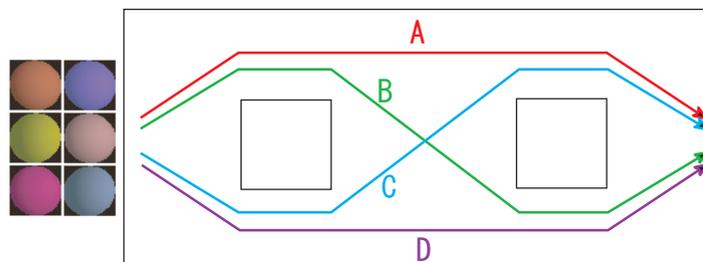


図 4.12 人物を模擬したCGオブジェクトの移動経路

時空間的尤度と見えの尤度の学習

シミュレータで 100 個の CG オブジェクトを通過させ、録画した映像を用いて時空間的尤度の学習を行った。録画した映像から、CG オブジェクトの追跡を行い(図 4.13)、出現位置 (x_i, y_i) 、消失位置 (u_i, v_i) 、消失速度 (\dot{u}_i, \dot{v}_i) 、見えの画像を各観測に対して記録した。そして、正しいオブジェクト同士の対応から旅行時間 $t_{i,j}$ を計算し、学習データ $x_{i,j}$ として記録することで、時空間的尤度の学習を行った。

また、シミュレーションでは、同じ CG オブジェクトの見えは同一であるため、見えの尤度を推定するパラメータは $\mu_{i,j} = 0, \sigma_{i,j}^2 = 1$ と仮定した。

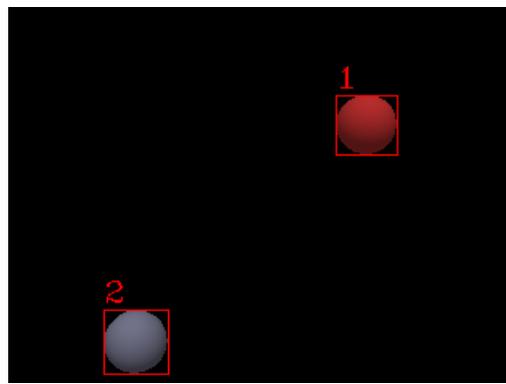


図 4.13 CG オブジェクトの追跡

CG オブジェクトの対応付け

シミュレータで 100 個の CG オブジェクトを通過させ、記録した複数の映像で観察された CG オブジェクトの対応付けを行った。

記録された映像から、CG オブジェクトの追跡を行い、出現位置 (x_j, y_j) 、消失位置 (u_i, v_i) 、消失速度 (\dot{u}_i, \dot{v}_i) を計測し、見えの画像を保存した。このようにして記録した観測データを用い、時間的に遡らない観測データの組合せを仮説として、重み付き 2 部グラフを構築した。そして、最大マッチングを Edmonds の手法を用いて求めることで観測の対応付けを行った。

このシミュレーション実験では、通過させたすべてのオブジェクトに対して、正しい対応付けを推定することができた。

4.4.2 実環境における実験

観測領域の設定

実環境における実験は、2章で述べた手法に基づき、室内天井に IEEE1394 カラーカメラ (Point Grey Research 社製 Flea) を設置して行った。図 4.14 に示すように、隣り合った 2 つの部屋の出入り口付近を観測領域として、観測された複数人物動線の対応付けを行った。2 つの観測領域における人物の追跡の様子を図 4.15、図 4.16 に示す。

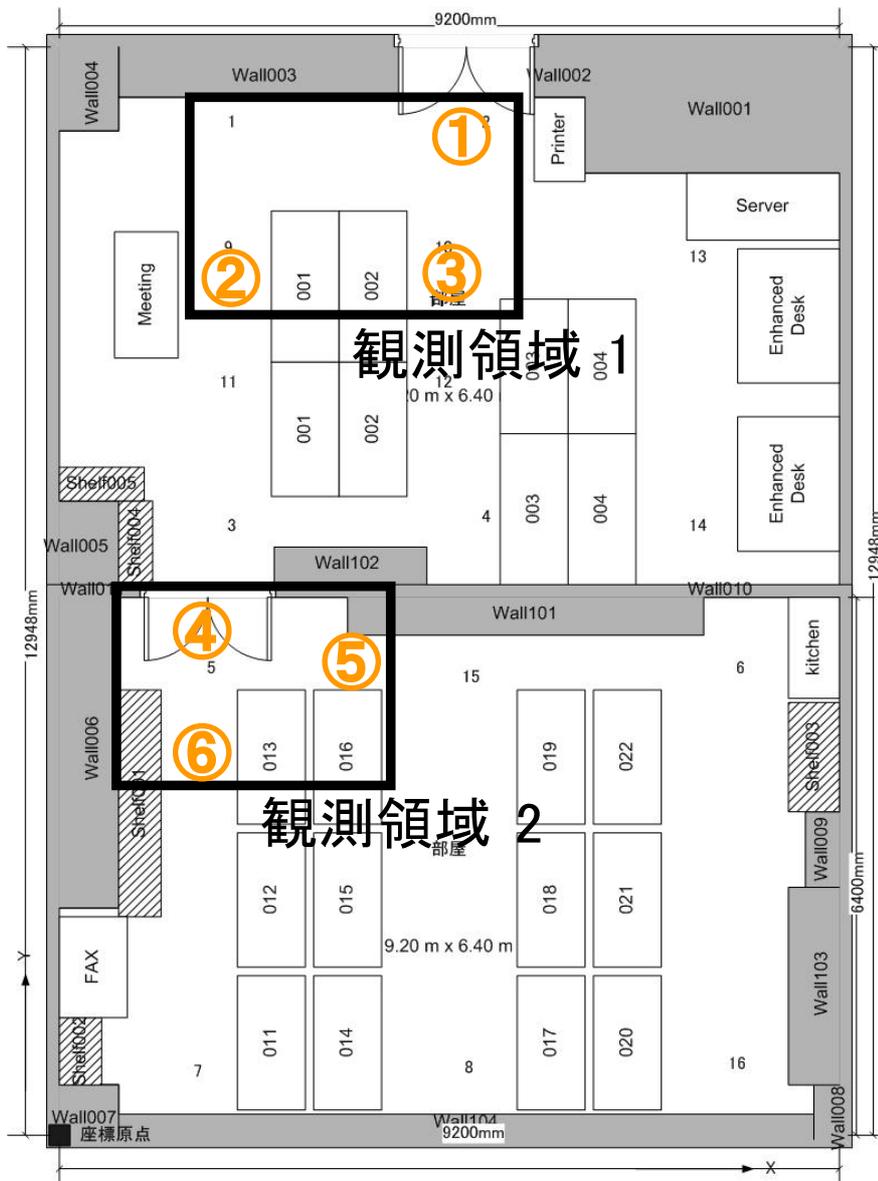


図 4.14 観測エリア

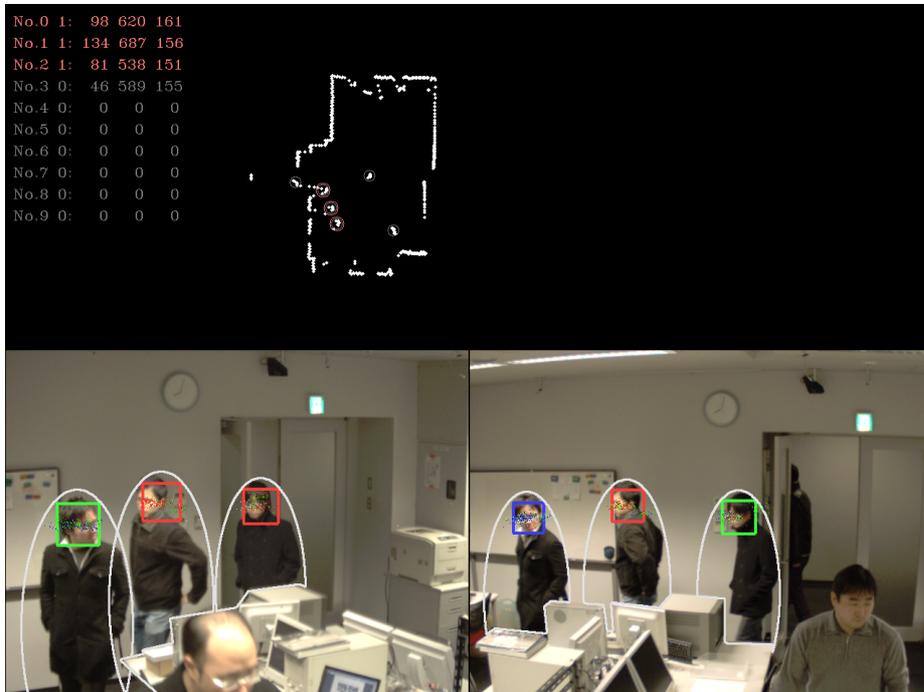


図 4.15 観測領域 1 における追跡の様子

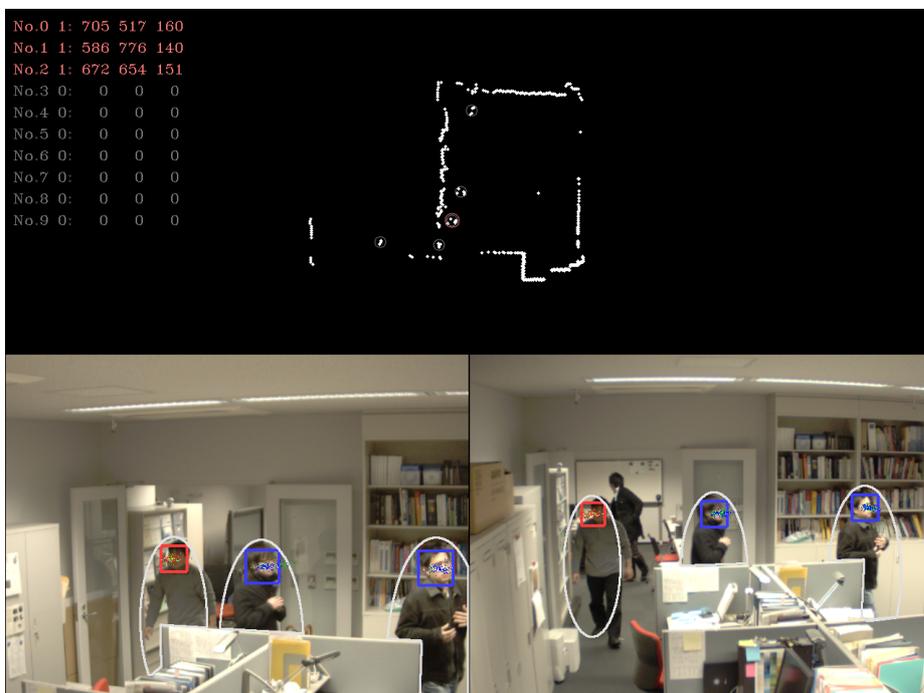


図 4.16 観測領域 2 における追跡の様子

時空間的尤度と見えの尤度の学習

時空間的尤度の学習は、事前に 5 人の人物が 2 つの観測領域間を移動した際の人物頭部の出現位置 (x_j, y_j, z_j) 、消失位置 (u_i, v_i, w_j) 、消失速度 $(\dot{u}_i, \dot{v}_i, \dot{w}_i)$ 、旅行時間 $t_{i,j}$ のデータを 2 回記録し、データ数を増やすため、便宜的に雑音を付与することで 50 個の学習データとして記録した。ここで、本実験では、人物頭部を三次元追跡しているため、時空間的尤度も三次元座標に基づいて学習していることに注意されたい。

人物の見えの尤度は、画像上で観察される人物を床と頭部の高さに基づいて楕円形でモデル化し、背景差分により得られる人物領域との論理積をとった面積が、複数カメラで得られるすべての観測のなかで、最大となるときの色ヒストグラムを記録し、見えの尤度を推定するパラメータ $\mu_{i,j}, \sigma_{i,j}^2$ を取得した。

実験結果

実験では、観測領域 1 と観測領域 2 の間を 5 人の人物が移動した。図 4.14 の観測領域内の番号の位置で人物の出現・消失が観測され、観測領域 1 では 5 個の観測、観測領域 2 では 8 個の観測が得られた。

本章で述べた手法により、それぞれの観測の対応付けを行った結果を図 4.17 に示す。図 4.17 より、一組を除いて正しい観測の対応付けが得られている。観測の誤対応が一組推定されているが、これは長い消失時間の後に観察された人物についての誤対応である。このような誤対応は、一定時間が経過した観測同士は対応付けを行わない (仮説を立てない) などの制約を導入することで、回避できると考えられる*5。

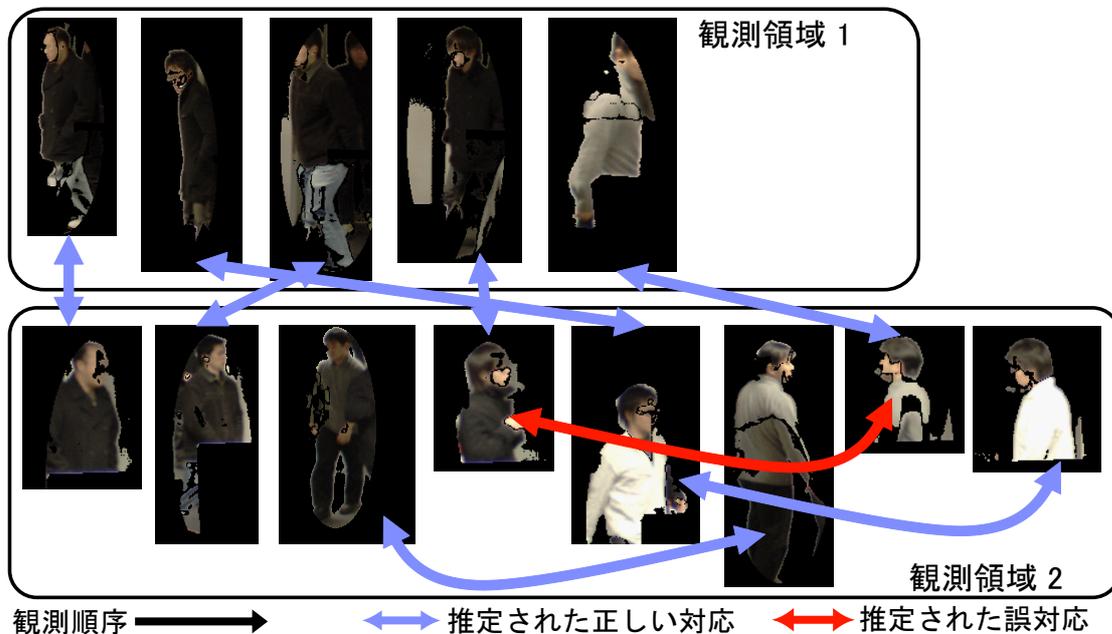


図 4.17 対応付けの推定結果

*5 長い時間、複数の人物が観測領域外にいる場合、時空間的尤度による評価が困難となる。そのため、観測領域外に長い時間滞在した後に観測される人物を、以前の観測に対応付けることは非常に難しい。

4.5 まとめ

本章では、複数観測領域間における人物動線の対応付けを行った。

複数観測領域間における人物動線の対応付けでは、異なる 2 地点で同時に同じ人物が観察されることは無いといった事実や、ある程度離れた観測領域間では、その間の旅行時間を予想することができるなどの時間的空間的な尤もらしさ(時空間的尤度)を考えることができる。そこで、想定される多くの人物動線の組から、尤もらしい人物動線の組を、グラフの最適化手法を利用して探索した。

人物動線の出現・消失をノードとし、対応をエッジ、尤もらしさを重みとして、有向グラフで表現し、重み付き 2 部グラフへ変換した。そして、変換された重み付き 2 部グラフの最大マッチングを Edmonds の手法により求めることで、複数観測領域間における人物動線の対応付けを行った。実環境における実験により、人物動線の正しい対応付けが得られることを確認した。

第 5 章

観測領域外を含む人物動線の推定

5.1 はじめに

安全・安心な社会環境の実現に向け、監視カメラ活用への期待は大きい。しかし一方で、センサの物理的な配置の制限や、障害物による遮蔽、プライバシーの問題などにより、人物の行動すべてをカメラなどのセンサで直接観察することは困難である場合が多い。このような状況で、人物の行動を計測するためには、環境に疎に分散配置された視野を共有しないカメラ群からの情報を統合し、人物の行動を計測する技術が求められる。

環境に密に配置された視野を共有するカメラ群による観察から、人物の行動を計測する手法は従来から提案されている (例えば [56, 61] など)。しかし、環境に疎に分散配置された視野を共有しないカメラ群からの情報を統合し、人物の動線を計測している例は少ない。

本章では、4 章までの議論に基づき、環境に疎に分散された観測領域間において、人物の同一性は既知であるものとして、観測領域外を含めた人物の動線の推定を行う。特に、机や壁の配置などの環境の情報をシステムに与えることで、尤もらしい人物動線を観測エリア全体で推定する手法について述べる。このような特定の観測エリアにおける、入室から退室までの人物動線の推定は、店舗や公共施設などでの不審人物の検出・発報やマーケティングなどへの応用が期待され、社会的意義も大きいと考えられる。

本章では、5.2 節にて観測領域外を含む人物動線の推定手法について述べる。そして、5.3 節にて、検証実験について述べ、5.4 節にてまとめる。

5.2 観測領域外を含む人物動線の推定

Rahimi らの手法 [74] では、机や壁などの障害物の配置状況を考慮していないため、推定された人物動線が壁を通過してしまうなどの問題があった。また、センサとしてカメラを壁面に水平に設置する必要があり、一般的な監視カメラのように天井付近に配置したカメラを用いた議論はなされていないなど、十分ではない点も多い。

そこで、本節では室内天井に設置された監視カメラ映像を想定し、人物が室内を歩行している場合を仮定して、観測領域外を含む尤もらしい人物動線を推定する手法を提案する。4 章までの議論に基づき、人物の同一性は既知として、人物一人の三次元頭部移動軌跡を推定する。また、観測には位置や姿勢が既知である環境に疎に分散配置された視野を共有しないカメラ群を用い、環境の情報を考慮することで、人物動線をカメラで観測されない箇所においても良好に推定する。

以降では Rahimi ら [74] と同様の考えに基づいて、観測領域外を含む人物頭部の三次元移動軌跡推定を定式化する。

5.2.1 人物頭部の移動軌跡

ある時刻 t における人物頭部の状態量 \mathbf{x}_t を式 (5.1) のように定義する。

$$\mathbf{x}_t = \begin{bmatrix} u_t & \dot{u}_t & v_t & \dot{v}_t & w_t & \dot{w}_t \end{bmatrix}^\top \quad (5.1)$$

ここで、 u_t, v_t, w_t は世界座標系における人物頭部の位置、 $\dot{u}_t, \dot{v}_t, \dot{w}_t$ はそれぞれの速度成分を表す。

また、観測開始から時刻 T まで ($t = 1 \dots T$) の軌跡 \mathbf{X} を \mathbf{x}_t の状態集合として表現すると、軌跡 \mathbf{X} は式 (5.2) のような $1 \times 6T$ の行列として表現される。

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top & \mathbf{x}_2^\top & \dots & \mathbf{x}_T^\top \end{bmatrix}^\top \quad (5.2)$$

ここで、軌跡 \mathbf{X} は時間経過に伴う位置の履歴を含んでいるため、人物頭部の移動軌跡の推定は、この軌跡 \mathbf{X} を推定する問題に帰着される。

5.2.2 運動モデルと観測モデルによる評価

運動モデル

人物頭部の動きは急激に変化しないと仮定し、運動モデルを式 (5.3) のように準備する。

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{v}_t \quad (5.3)$$

ここで \mathbf{A} は時刻 t から時刻 $t+1$ に遷移するときの状態量の変化をモデル化した状態遷移行列である。また、 \mathbf{v}_t は平均 $\mathbf{0}$ 共分散行列 Σ_v を持つ雑音である。

\mathbf{A} , Σ_v は例えば以下のような行列である .

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Sigma_v = \text{diag} [0.1 \quad 1 \quad 0.1 \quad 1 \quad 0.1 \quad 1]$$

これより , \mathbf{x}_{t-1} が与えられたときの \mathbf{x}_t の確率分布を式 (5.4) のように表現することができる .

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = N(\mathbf{x}_t | \mathbf{A}\mathbf{x}_{t-1}, \Sigma_v) \quad (5.4)$$

ここで $N(\mathbf{x}_t | \mathbf{A}\mathbf{x}_{t-1}, \Sigma_v)$ は , \mathbf{x}_t の確率分布が平均ベクトル $\mathbf{A}\mathbf{x}_{t-1}$ 共分散行列 Σ_v の多変量正規分布に従うことを表す .

さらに , \mathbf{x}_t の状態集合である軌跡 \mathbf{X} についても確率分布で表現でき , 事前確率として式 (5.5) のように定義する .

$$p(\mathbf{X}) = \prod_{t=1}^T p(\mathbf{x}_t | \mathbf{x}_{t-1}) = N(\mathbf{X} | \mathbf{0}, \Lambda_{\mathbf{X}}) \quad (5.5)$$

ここで $\Lambda_{\mathbf{X}}$ は以下で表される行列である .

$$\Lambda_{\mathbf{X}} = \begin{bmatrix} \mathbf{A}^\top \Sigma_v \mathbf{A} & -\mathbf{A}^\top \Sigma_v & 0 & \dots \\ -\Sigma_v \mathbf{A} & \Sigma_v + \mathbf{A}^\top \Sigma_v \mathbf{A} & -\mathbf{A}^\top \Sigma_v & \dots \\ 0 & -\Sigma_v \mathbf{A} & \Sigma_v + \mathbf{A}^\top \Sigma_v \mathbf{A} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

観測モデル

時刻 t における人物頭部の 3 次元位置の観測を , 各カメラ画像への射影と雑音の和で記述し , 観測モデルを式 (5.6) のように準備する .

$$\mathbf{y}_t^i = \pi^i(\mathbf{x}_t) + \omega_t \quad (5.6)$$

ここで $\pi^i(\mathbf{x}_t)$ は , 時刻 t における状態量 \mathbf{x}_t から i 番目のカメラの画像座標 $[m_t^i \ n_t^i]^\top$ への射影を行う関数である . また , ω_t は平均ベクトル $\mathbf{0}$, 共分散行列 Σ_ω を持つ雑音であり , $\Sigma_\omega = \text{diag} [\sigma_\omega^2 \ \sigma_\omega^2]$ である . σ_ω^2 はカメラ画像における観測の水平 , 垂直方向の分散であるが , ここでは , 簡単のために同じ値を採用している .

これより，時刻 t における状態量 \mathbf{x}_t を持つ人物頭部が i 番目のカメラに画像座標 \mathbf{y}_t^i で観測される確率 $p(\mathbf{y}_t^i|\mathbf{x}_t)$ を式 (5.7) のように表すことができる．

$$p(\mathbf{y}_t^i|\mathbf{x}_t) = \begin{cases} N(\mathbf{y}_t^i|\pi^i(\mathbf{x}_t), \Sigma_\omega) & \text{if } \ell^i(\mathbf{x}_t) \\ \delta(\mathbf{y}_t^i - \phi) & \text{otherwise} \end{cases} \quad (5.7)$$

ここで， $\ell^i(\mathbf{x}_t)$ は， i 番目のカメラ視野内に状態量 \mathbf{x}_t を持つ人物頭部が観測される場合に真となる関数であり， i 番目のカメラ視野内で人物頭部が観測されないとき， \mathbf{y}_t^i は ϕ とする．

全てのカメラ ($i = 1 \dots N$) の観測開始から時刻 T まで ($t = 1 \dots T$) の観測の集合を \mathbf{Y} とすると，尤度を以下のように定義できる．

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{t=1}^T \prod_{i=1}^N p(\mathbf{y}_t^i|\mathbf{x}_t)$$

これを時刻 t で i 番目のカメラ視野内で観測される場合 $(t, i) \in o$ と観測されない場合 $(t, i) \notin o$ で分け，ある軌跡 \mathbf{X} に対して観測 \mathbf{Y} が得られる確率を式 (5.8) のように表す．

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{(t,i) \in o} N(\mathbf{y}_t^i|\pi^i(\mathbf{x}_t), \Sigma_\omega) \prod_{(t,i) \notin o} (1 - \ell^i(\mathbf{x}_t)) \quad (5.8)$$

運動モデルと観測モデルによる評価関数

ベイズの法則に基づき，運動モデル $p(\mathbf{X})$ ，観測モデル $p(\mathbf{Y}|\mathbf{X})$ を用いて，観測 \mathbf{Y} が得られたときの軌跡 \mathbf{X} の条件付き確率 $p(\mathbf{X}|\mathbf{Y})$ を求める．

$$p(\mathbf{X}|\mathbf{Y}) \propto p(\mathbf{X})p(\mathbf{Y}|\mathbf{X}) = N(\mathbf{X}|\mathbf{0}, \Lambda_{\mathbf{X}}) \prod_{(t,i) \in o} N(\mathbf{y}_t^i|\pi^i(\mathbf{x}_t), \Sigma_\omega) \prod_{(t,i) \notin o} (1 - \ell^i(\mathbf{x}_t))$$

これを整理して，確率 $p(\mathbf{X}|\mathbf{Y})$ が高いほど，尤もらしい軌跡として高い評価を与える評価関数 $O_o(\mathbf{X})$ を導出する (式 (5.9)) ．

$$O_o(\mathbf{X}) = \mathbf{X}^\top \Lambda_{\mathbf{X}}^{-1} \mathbf{X} + \frac{1}{\sigma_\omega^2} \sum_{(t,i) \in o} (|\mathbf{y}_t^i - \pi^i(\mathbf{x}_t)|)^2 \quad (5.9)$$

ここでは簡単のため，観測モデルにおいて観測されない場合の評価を無視している．また，整理の途中で，正負の変換を行っているため，評価関数 $O_o(\mathbf{X})$ の値は小さいほど尤もらしい評価を与える．

ここまでの式変形の詳細は文献 [74] を参照されたい．

5.2.3 環境モデルの導入

本節では，これまでのモデルに加え，さらに環境モデルとして，人物頭部の存在確率に基づく軌跡の尤もらしさを評価する指標を導入する．通常，人物は壁を通過することはなく，机や棚などの障害物を越えて移動することもほとんどないと考えられる．また，経験的に，人物が通路やドアを通過する際は，壁面や障害物の直近ではなく，通りやすい箇所を通過するケースが多いと考えられる(図5.1)．このことから，人物が壁面や障害物に近づくほど大きくなり，壁面や障害物の位置で非常に大きなペナルティを与える関数を環境モデルとして準備する．これを，状態量 \mathbf{x}_t を持つ人物頭部と障害物との距離を $d_h(\mathbf{x}_t)$ として，ペナルティを与える関数を $e_h(d_h(\mathbf{x}_t))$ で表す．

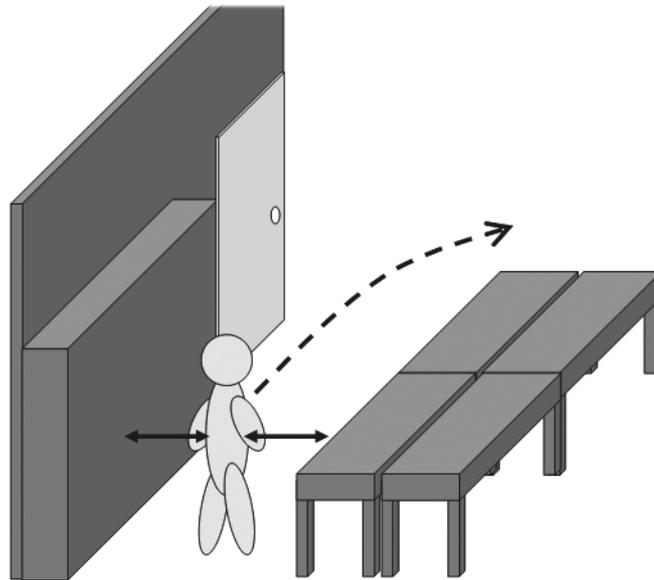


図 5.1 環境モデルの概念

また，人物頭部の高さについても，歩いている間は大きな変化は起こらないと予測されることから，人物頭部の高さは，平均的な身長から予測される人物頭部の高さから大きな乖離を持たないと仮定し，平均的な身長から予測される人物頭部の高さからの乖離に対してペナルティを与える関数を準備する．これを，状態量 \mathbf{x}_t を持つ人物頭部と，平均的な身長から予測される人物頭部の高さとの距離を $d_v(\mathbf{x}_t)$ として，ペナルティを与える関数を $e_v(d_v(\mathbf{x}_t))$ で表す．

時刻 t における人物頭部の状態量 \mathbf{x}_t に対する環境モデルによる評価は， $e_h(d_h(\mathbf{x}_t))$ と $e_v(d_v(\mathbf{x}_t))$ の和で表し，観測開始から時刻 T まで ($t = 1 \dots T$) の軌跡 \mathbf{X} に対する環境モデルによる評価関数を $O_e(\mathbf{X})$ として，式(5.10)のように与える．

$$O_e(\mathbf{X}) = \lambda \sum_t (e_h(d_h(\mathbf{x}_t)) + e_v(d_v(\mathbf{x}_t))) \quad (5.10)$$

ここで, λ はペナルティ調整用のパラメータである.

関数 $e_h(d_h(\mathbf{x}_t))$ と関数 $e_v(d_v(\mathbf{x}_t))$ における水平方向及び鉛直方向の距離と, 与えるペナルティの関係を図 5.2 に示す.

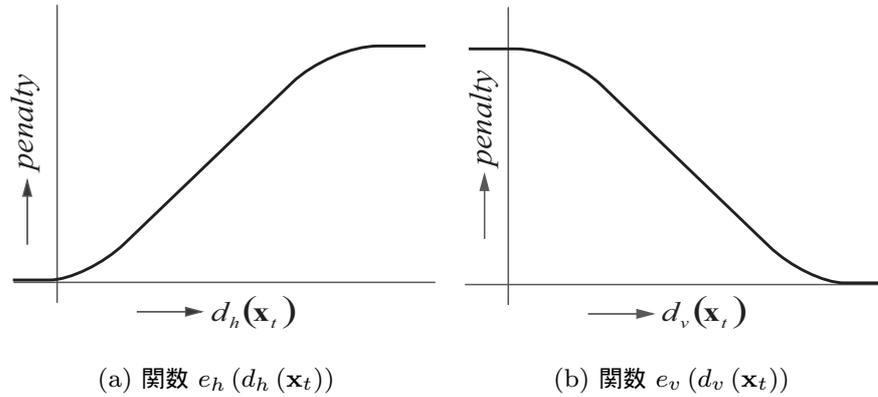


図 5.2 距離とペナルティの関係

このようにして準備した環境モデルをあわせて評価することで, 人物頭部の軌跡が障害物を通過したり, 平均的な身長から大きく乖離してしまう問題を回避する.

5.2.4 各モデルによる評価の統合と最適化

軌跡 \mathbf{X} が $\mathbf{x}_1 \dots \mathbf{x}_T$ の状態集合として与えられたとき, 運動モデル, 観測モデル, 環境モデルによる評価を行い, 尤もしい軌跡 \mathbf{X}^* を推定するため, これまでの評価関数を統合する. 評価関数の統合は式 (5.11) のようにそれぞれの評価関数の和で表現し, この評価関数の値が最も小さくなる軌跡 \mathbf{X}^* を推定する.

$$\begin{aligned} \mathbf{X}^* &= \arg \min_{\mathbf{X}} (O_o(\mathbf{X}) + O_e(\mathbf{X})) = \\ & \arg \min_{\mathbf{X}} \left(\mathbf{X}^\top \Lambda_{\mathbf{X}}^{-1} \mathbf{X} + \frac{1}{\sigma_\omega^2} \sum_{(t,i) \in o} (|\mathbf{y}_t^i - \pi^i(\mathbf{x}_t)|)^2 \right. \\ & \quad \left. + \lambda \sum_t (e_h(d_h(\mathbf{x}_t)) + e_v(d_v(\mathbf{x}_t))) \right) \end{aligned} \quad (5.11)$$

この評価関数の最小化には非線形計画法を用いる. 本手法で用いる評価関数は凸関数ではないため, 非線形計画法を適用した場合, 局所的な最適解に収束してしまう可能性が考えられる. この問題を回避するために, 評価関数の最適化を始めるための初期軌跡 \mathbf{X}_0 を経験的に複数与え, それぞれに対して最適化を行い, 得られた評価関数の値が最小となる軌跡を採用する. 例えば, 人物頭部がカメラ 1 とカメラ 2 で観測された場合, 初期軌跡 \mathbf{X}_0 は次のように与える.

a) カメラ 1 とカメラ 2 の間に障害物がない場合

カメラ 1 及びカメラ 2 の光軸が平均的な身長から予測される人物頭部の高さで交わる点を A, B とし, A-B をカメラ 1 で人物頭部が観測されてからカメラ 2 で観測されるまでの観測フレーム数 T で等間隔に分割した点をそれぞれの時刻における初期人物頭部位置とする. 分割した 2 点間の差分をそれぞれの時刻における速度として初期軌跡 X_0 を生成する (図 5.3(a)).

b) カメラ 1 とカメラ 2 の間に障害物がある場合

カメラ 1 及びカメラ 2 の光軸が平均的な身長から予測される人物頭部の高さで交わる点を A, B とし, 経験的に経路として人物が通過する位置にホールドポイント H1, H2 を置く. A-H1, H1-B の距離に応じて, カメラ 1 で人物頭部が観測されてからカメラ 2 で観測されるまでの観測フレーム数 T を配分し, 配分されたフレーム数によって A-H1, H1-B を等間隔で分割した点をそれぞれの時刻における初期人物頭部位置とする. また, 分割した 2 点間の差分をそれぞれの時刻における速度として初期軌跡 X_0 を生成する. H2 を通過する場合も同様に生成する (図 5.3(b)).

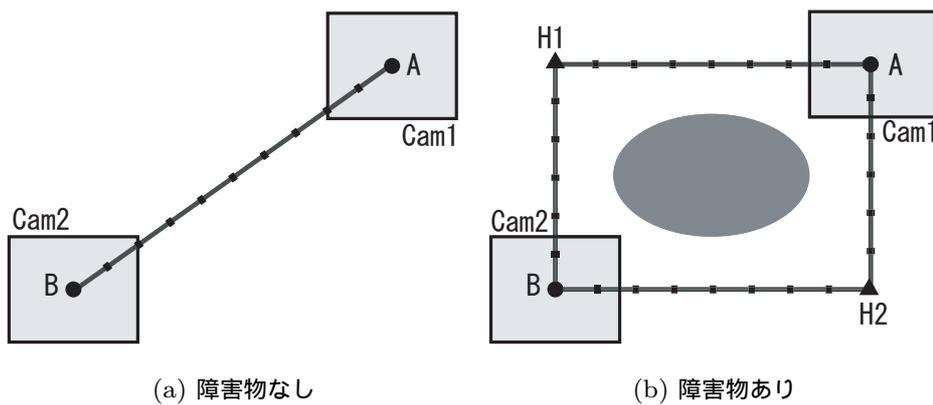


図 5.3 最適化開始時に与える初期軌跡

5.3 実験

5.3.1 シミュレーション

これまで述べた手法に基づき，人物の尤もらしい移動軌跡推定のシミュレーションを行った．

観測データの取得

シミュレーションはPC上で動作するシミュレータを開発して行った(図5.4)．マウスのドラッグによる軌跡を人物の模擬移動軌跡として入力に使用した．ただし，マウスによるドラッグでは鉛直方向の高さを与えることができないため，状態量の生成では，一定間隔でサンプリングされた水平方向位置に対して平均的な身長から予測される一定の人物頭部位置の高さを付与した．また，模擬移動軌跡の入力では，実際の人物の移動を模擬する観点から，室内に配置された壁や机などの障害物を避けて入力を行った．用いた室内モデルは実際の室内を模擬したもので，広さは920cm×1290cmである．

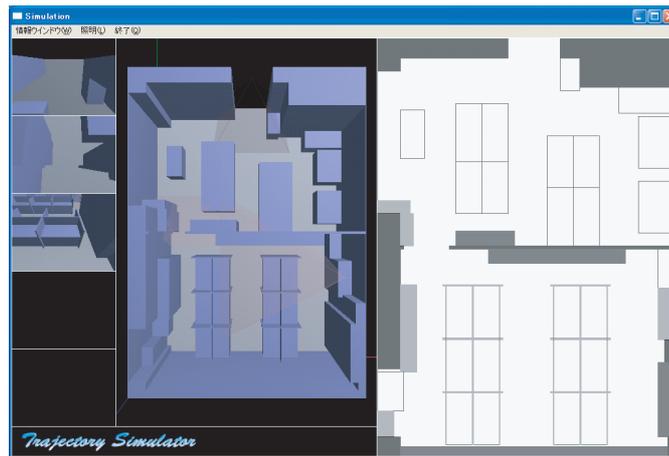


図 5.4 シミュレータ

シミュレーション結果

シミュレータへの入力により記録された観測データを用いて，軌跡の推定を行った．評価関数の最小化には MATLAB[54] の非線形計画法による最適化関数を利用した．推定結果の例を図 5.5(a) に示す．図 5.5 において実線が，マウスによる入力によって生成された人物の模擬移動軌跡であり，人物の右上方から左下方への移動を模擬している．点線は推定された軌跡である．灰色の矩形は壁や棚，机の配置を示しており，右上方と左下方の矩形領域はカメラの視野を表している．

図 5.5(b) は環境モデルによる評価を行わずに推定を行った結果である．図より明らかに軌跡が机の上を通過していることが分かる．日常的に人物が机の上を通過することがない状況において，このような推定結果は好ましくない．

また，推定時に入力した初期軌跡は三種類であり，図 5.5(c)(d)(e) にそれぞれの初期軌跡を入力した場合に推定された軌跡と評価関数の値を示した．初期軌跡は図 5.5(c)(d)(e) に黒三角形で示したホールドポイントを通る軌跡である．評価関数の値から，最終的な推定軌跡として図 5.5(c) の軌跡が採用されている．

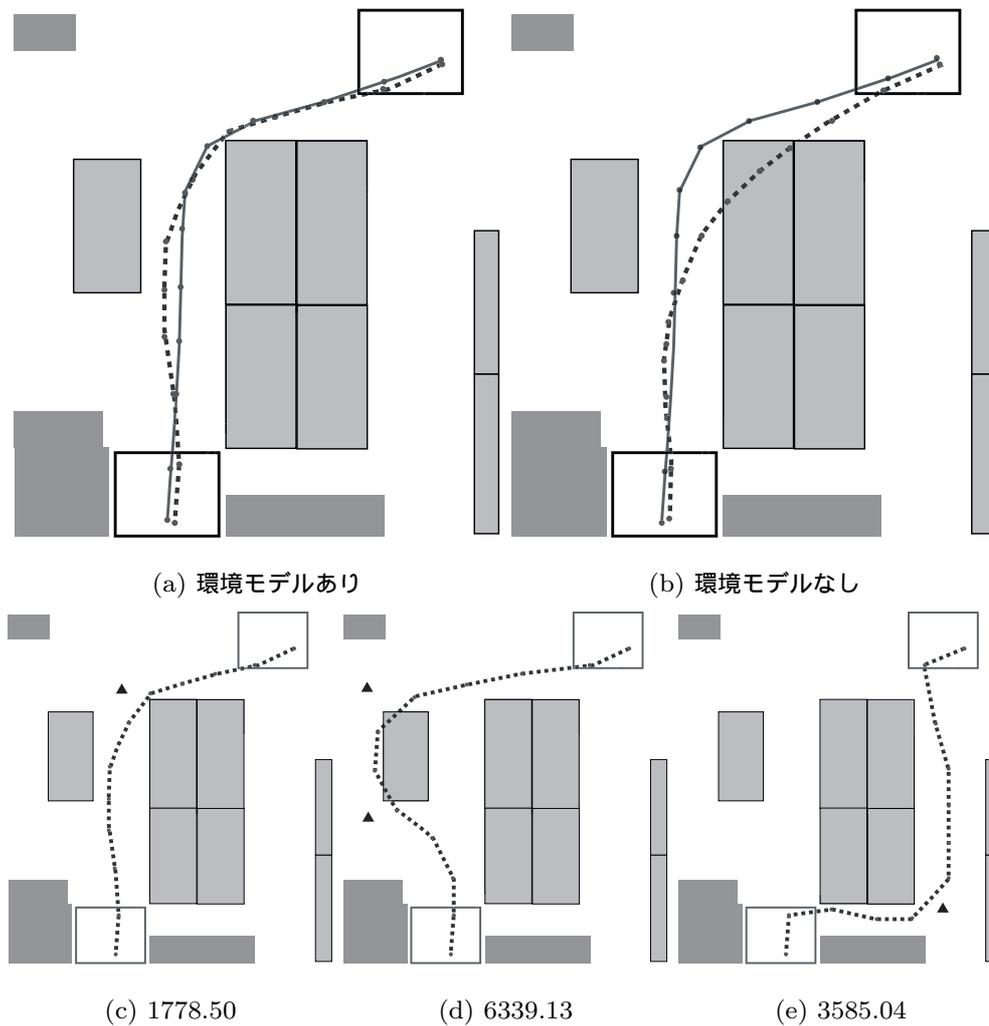


図 5.5 シミュレーション

5.3.2 実環境における実験

実環境における実験は、室内天井に IEEE1394 カラーカメラ (Point Grey Research 社製 Flea) を設置して行った。カメラの設置箇所は図 5.6 に示す廊下側室入口 (A)、窓側室への通路 (B)、窓側室の窓際 (C)、窓側室給湯設備前 (D) の 4 箇所であり、4 台のカメラは視野を共有していない。また、カメラ A、B と C、D は焦点距離の異なるレンズを使用しており、各カメラはそれぞれ異なる姿勢で下方を撮影している。各カメラの映像は 320×240 画素、毎秒 15 フレームの映像で得ることとし、4 台のカメラの映像を、汎用 PC (CPU Intel Pentium4 3.2GHz, Memory 2GByte, OS WindowsXP) 1 台で記録した。

実験ではあらかじめ決められた軌跡を人物が移動した際の映像を記録した。人物は、室外から廊下側室に入り (A)、通路を通って窓側室へ入り (B)、窓側室の窓際で向きを変え (C)、窓側室の給湯設備の前 (D) まで移動する。

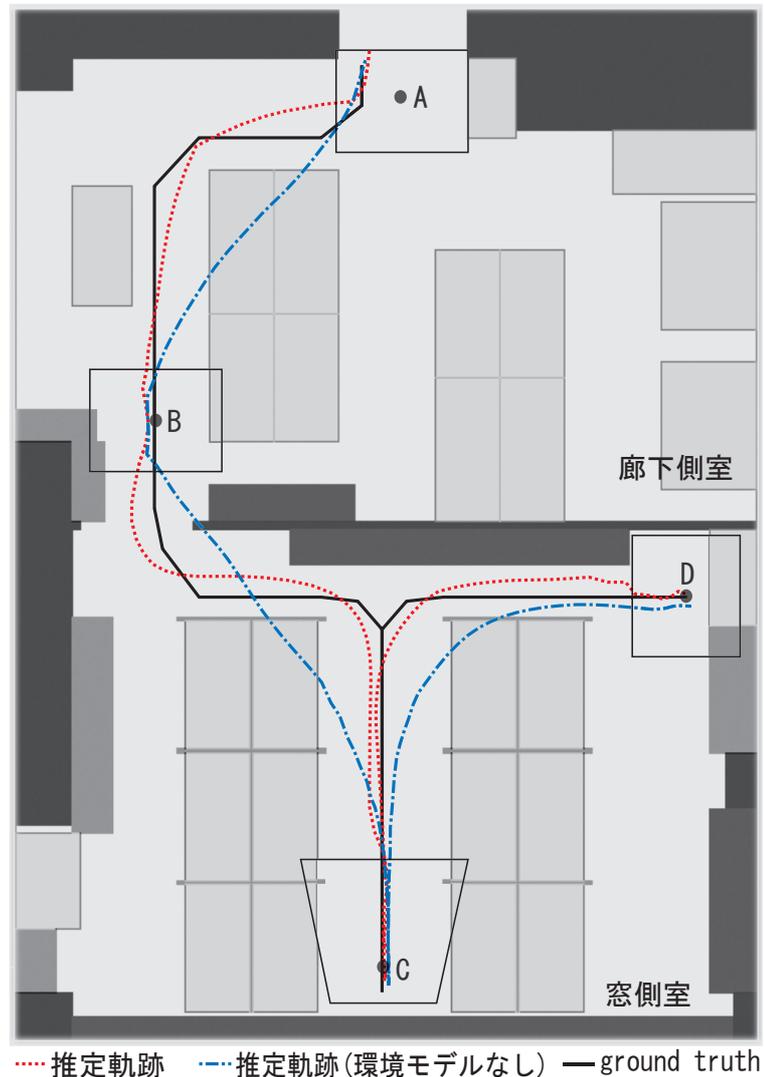


図 5.6 室内レイアウト図と推定された軌跡

観測データの取得

実験で得られた各カメラの画像の例を図 5.7 に示す。図 5.7 の赤い円は検出された人物頭部の位置を示している。

各カメラで得られた人物頭部の位置データは、最適化の際の計算時間を考慮してすべては使用せず、5 フレーム毎のデータを観測データとして使用した。本実験では、カメラ 4 台の各 101 フレームの観測データを使用して、人物が室内を約 40 秒間移動した際の軌跡の推定を行った。

また、推定に用いる環境モデルは室内レイアウト図からあらかじめ作成しておき、室内の座標系は図 5.6 の左下床面を原点として、横方向を X 軸，縦方向を Y 軸，床面からの高さを Z 軸とした。

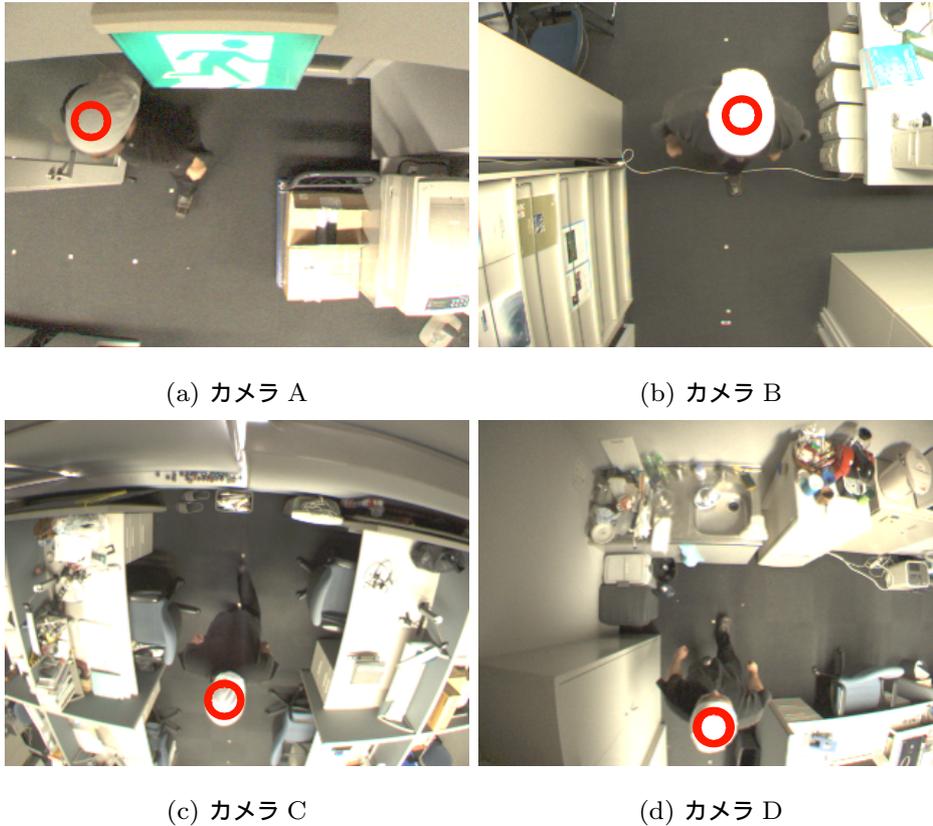


図 5.7 実験で得られたカメラ画像の例

実験結果

実験で記録した映像を利用して推定した人物の移動軌跡の推定結果を図 5.8 に示す．また，図 5.6 にも推定結果を重ねて表示した．図 5.8 及び図 5.6 において，赤色の点線が本手法を用いて推定を行った結果である．黒色の実線は，人物が実験で移動したあらかじめ決められた軌跡であり，ground truth として表示している．また，環境モデルの導入による効果を明示するため，同じ観測データを用いて環境モデルなしで推定を行った結果を青色の一点鎖線で示している．

図 5.6 においてカメラの設置箇所付近で推定軌跡が ground truth に近づいている．これは評価関数の観測モデルにより，推定軌跡が拘束されているためと考えられる．一方，カメラで観測されない部分においては ground truth と差が大きくなっているが，軌跡は滑らかに接続されており，運動モデルによる拘束が有効に働いていると考えられる．また，環境モデルなしで推定した結果と比較して，机や壁などの障害物を避けた人物の移動軌跡を推定できていることから，環境モデルによる拘束が有効に働いていると考えられる．また，鉛直方向について乖離が少ないことから環境モデルによる拘束が有効に働いていることが分かる．

カメラで人物頭部が観測されたフレームは全体の 3 割程度であるが，運動モデルや環境モデルにより軌跡を拘束することで，人物の移動について得られる情報が少ない場合でも，尤もらしい軌跡が推定できている．また，人物が特異な行動をとった場合には，本手法の評価関数の値が特異なものになると予想される．このことを利用して，観測情報に基づいた人物行動の尤もらしさを評価する手法として本手法を利用することも考えられる．

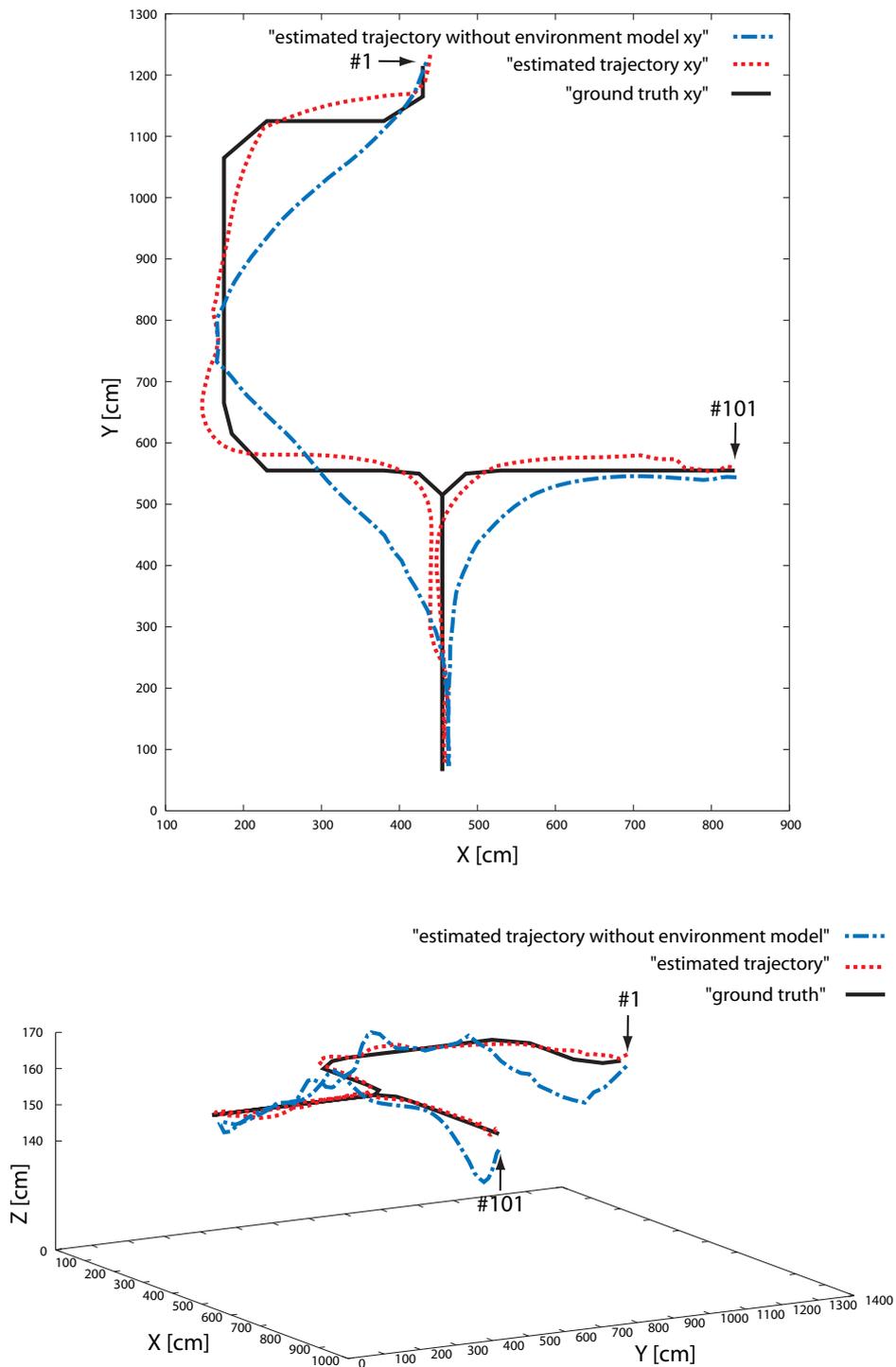


図 5.8 軌跡の推定結果

5.4 まとめ

本章では、視野を共有しないカメラ群を用いて、人物の移動軌跡を、運動モデルと観測モデルに環境モデルを加えた評価関数の最適化を行うことで推定する手法を提案した。

運動モデルは、人物の動きは急激に変化しないと仮定してモデル化した。環境モデルは、レイアウト図などから壁や机の位置を取得し、それらの位置には人物は存在しないなどの情報をペナルティとして経験的に付与することでモデル化した。カメラによる観測から、いくつかの経験に基づいた移動軌跡を初期軌跡として入力し、運動モデル及び観測モデル、環境モデルに基づく評価関数を、非線形計画法を用いて最適化することで尤もらしい人物の移動軌跡を推定した。実環境における実験の結果、机や壁などの環境を考慮することにより、人物の移動軌跡が机や壁を通過してしまう問題を回避することができた。

第6章

結論

6.1 本研究の成果

本研究では、分散配置されたセンサ群からの情報を統合することで、実環境における人物頭部の頑健かつ高精度な追跡を実現し、観測領域外を含む特定エリア全体における人物動線の推定を可能とした。本論文を構成する各章について、その成果をまとめる。

時系列フィルタと識別器の統合による三次元人物頭部追跡

カメラからの入力画像を用いた人物追跡では、人物が必ずしもカメラに対して一定方向を向いていないために起こる見えの変化や、広い範囲を視野とすることによる低解像度での観察、昼間と夜間などの照明変動、さらに商品棚などの複雑な背景などの課題がある。

これらの課題に対して、識別器を用いた顔検出技術を時系列フィルタによる追跡の枠組みに統合することで、頑健かつ高精度な三次元人物追跡を実現した。具体的には、パーティクルフィルタによる追跡の枠組みにおいて、オプティカルフローに基づいて予測された仮説に対応した複数カメラの画像領域を識別器によって評価する。このとき、人物頭部は実際の向きやカメラとの位置関係により見えが変化するため、各方向の頭部の向きごとに学習を行った識別器をあらかじめ準備し、複数のカメラによりさまざまな方向で観察される人物頭部に対し、識別器を選択的に用いることで、見えの変化に対応した。また、各方向に対応した複数の識別器による評価から、頭部の三次元位置と同時に向きを推定可能とし、多くの検証実験から、本手法の優位性を確認した。

カメラとレーザ測域センサの統合によるエリア内複数人物追跡

エリア内での複数人物の追跡では、人物相互の遮蔽を含む人物の出現・消失に対する追跡初期化が大きな課題となる。

これに対して、エリア内に配置された分散カメラとレーザ測域センサを相補的に統合することで、エリア内での人物の出現を検出し、頑健に追跡初期化を行う手法を開発した。また、オプティカルフローに基づいた状態遷移とレーザ測域センサによる観測に基づいた分布を混合した確率密度分布から仮説を生成することで、カメラ画像による追跡にレーザ測域センサを統合した。これにより、単眼カメラによる観測においても安定な三次元追跡を可能とし、分散配置されたセンサ群のさまざまな組合せによる三次元人物追跡を可能とした。

複数観測領域間の人物動線の対応付け

エリア内全体を通じて、連続した人物動線を計測するためには、障害物や人物相互による遮蔽により断片化した人物動線や、複数観測領域間における人物動線に対応付ける必要がある。これに対して、異なる 2 地点で同時に同じ人物が観察されることはないといった事実や、ある程度離れた観測領域間では、その間の旅行時間を予想することができるなどの時空間的尤度を、見えの類似性と併せて考えることで、人物動線の対応付けを行った。具体的には、観測された多くの人物動線から、尤もらしい人物動線の組を、観察された見えの類似性と時空間的尤度から、グラフの最適化手法を利用して推定した。人物動線の出現・消失をノードとし、対応をエッジとして扱うことで、尤もらしい人物動線の組を、重み付き 2 部グラフの最大マッチングを求めることで取得した。

観測領域外を含む人物動線の推定

連続した人物動線の取得のため、観測された情報を用いて観測領域外の人物動線の補間を行った。先行研究では室内の机や壁などの障害物の配置を考慮していないため、推定した人物動線が壁を突き抜けてしまうなどの課題があった。

これに対して、環境の制約条件を導入することで、良好な人物動線を推定する手法を開発した。具体的には、人物の運動モデルとカメラの観測モデル、環境形状から予測される人物の存在確率モデルを最大事後確率推定の枠組みで統合し、人物動線の尤もらしさを評価する関数を構築した。そして、観測された一部の人物動線を用いて、評価関数の最適化を行うことで、観測領域外を含む人物動線を推定した。

6.2 今後の課題

本研究で議論した技術の実用化に向け、実際の店舗などでの実証実験を重ねることが今後の課題である。これに関しては、都内コンビニエンスストアでの実証実験が既に予定されており、今後の報告が待たれる。また、更なる技術的な検討課題としては以下が挙げられる。

センサキャリブレーションの簡素化

実際の店舗などへの適用では、センサ設置作業の簡素化やセンサキャリブレーションの自動化が望まれる。特に、センサの位置や姿勢を表すパラメータは、現地にセンサを設置した後に計測する必要があるため、24時間営業のコンビニエンスストアなどでは、業務の妨げにならない簡易な手法が求められる。

本研究での検証実験では、実世界のある座標が、カメラ及びレーザ測域センサでどのように観測されるかの対応を実測することでセンサの位置や姿勢のキャリブレーションを行った。これには、土木作業用のトータルステーションを用いたが、作業者の負荷は小さくない。視野を共有する複数カメラで、レーザポインタなどの識別が容易なマーカを観察することで複数カメラ間のキャリブレーションを行う手法も提案されている [64, 95]。しかし、カメラ間で十分な視野の重なりが必要であることや、障害物などにより理想的な位置にマーカを移動できない環境も考えられるため、レーザ測域センサによる計測点の併用などによる、作業者の負荷を軽減するツールの開発が望まれる。

分散センサ間の同期

本研究では、観測に用いるセンサ群は同期して動作するという仮定に基づいているが、非同期のセンサ群を用いる場合には、非同期センサ情報の整合性をとりつつ情報統合を行う手法を検討する必要があると考えられる^{*1}。なお、2章で行った検証実験では、30fpsで駆動されるカメラに同期させるため、24Hzで観測されたレーザ測域センサの観測データを直近の観測データで補填している。そのため、厳密には計測データ間に数msの微小な時刻のずれが生じている。しかし、レーザ測域センサによる観測情報を確率的に用いる本研究の枠組みでは、ほとんど問題とならなかった。

処理速度の向上

本研究では、複数人物の追跡の場合に30fpsのビデオフレームレートでの処理が実現できていない。これに対しては、カメラから遠い人物の仮説を削減し、仮説の評価を効率化することや、人物位置とカメラ画像との射影関係を参照テーブルなどを用いて高速化するなど、実装レベルでの最適化を進める必要がある。

^{*1} 例えば、文献 [60] などでは非同期センサの統合利用が議論されている。

追跡性能のさらなる向上

人物頭部の追跡性能のさらなる向上にむけては、Online Boosting[28] 技術の検討が挙げられる。Online Boosting は識別器の学習をオンラインで行う枠組みであり、特定の環境や対象に特化した効率的な識別器を構築することができるとして注目されている。また、条件によっては、より少ない弱識別器により、安定な仮説の評価を行うことができる可能性もあり、計算コストの削減の観点からも、将来的な検討が望まれる。しかし、不特定人物の追跡においては、識別器の汎化性能を低下させる恐れもあるため、導入には慎重な検討が必要であろう。

人物動線対応付けの高度化

複数観測領域間の人物の対応付けでは、より高精度な見えに基づいた同一性推定手法の導入が望まれる。監視カメラ映像では、頭部の解像度が低く、頭部向きなどの姿勢が一定でないことから、従来の顔認証技術をそのまま導入することは難しいが、同一人物らしさを顔認証技術を応用して推定することも考えられる。例えば、監視カメラを想定した枠組みではないが、従来の顔認証技術に時空間的無矛盾性の概念を導入し、認証精度を向上させる取組み [24, 29] も報告されており、今後の発展が期待される。また、歩容認証 [18, 42, 77, 78, 112] も興味深い技術である。歩容認証は人物の歩行シルエットなどを周波数解析することで認証を行う方法であるが、本研究のアプローチから、弱識別器の評価値の時間変化を周波数解析することなども考えられる。

観測領域外の動線補間の簡素化

観測領域外を含む動線補間では、非線形最適化に時間がかかるため、多人数の動線補間は難しい。また、観測領域外の人物動線としては、本研究で実現したほどの精度が必要でない場合も多いと考えられる。そこで、観測領域外の人物動線は経路を推定する程度にとどめ、スプライン曲線などで、断片的な動線を接続することも有用であろう。

6.3 将来展望

人を中心とした実世界事象のコンピュータによる理解では、「いつ」、「どこで」、「誰が」、「何を」、「どうしたか」を認識することが大きな課題となる。本研究で議論した、エリア内における人物の検出・追跡及び動線の推定は、「いつ」、「どこで」の認識に相当する重要な基盤技術であり、本研究の成果に基づいた今後の発展が期待される。

「誰が」を認識するための個人認証

本研究の成果から、エリア内において、人物の顔画像を得られることから、顔認証技術の併用による、個人の識別情報と紐付けされた人物動線計測が可能になると考えられる。一般に、顔による個人識別では、人物の正面顔を用いて辞書内を探索するが、本研究の成果により、人物頭部の位置だけではなく、向きも同時に追跡できるため、人物追跡の枠組みのなかで、人物の正面顔を抽出し、個人識別に利用することも容易である。例えば、文献 [30] などのベストショット画像を保存して個人識別に応用する取り組みなどは、本研究と親和性が高いと考えられる。

「何を」を認識するための物体認識

コンピュータビジョンにおいて、物体認識は重要なテーマとして研究が継続されている分野であり、今後の研究成果が期待される。一方、実際の店舗などでのシステムを考えると、人物に ID タグを取り付けることは難しい場合が多いが、商品にタグをつけることは比較的容易であると考えられる。例えば、レンタル CD ショップなどでは、すでに商品に万引き防止タグがつけられており、棚から無くなった商品を検出、管理するスマートシェルフも実用化されつつある (例えば、[88] など)。これらの技術を併用すれば、人物と注目商品の紐付けや、より詳細な購買行動の計測も実現できると考えられる。

「どうしたか」を認識するための行動認識

例えば、本研究の成果として得られた人物動線を Stauffer ら [92] のように分類することで、人物の行動認識に向けた有益な情報を抽出できると考えられる。また、人物動線の学習により非定常行動の検出などを行うこともできると考えられる。文献 [94] では、顧客の動線データを分類し、そのどれにも当てはまらない逸脱行動を検出している。

謝辞

本研究を遂行するにあたり，日頃から熱心な御指導を賜り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，貴重なご指摘とアドバイスを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。

定期的なミーティングを通じて，電気通信大学 小池英樹教授，慶應義塾大学 中西泰人准教授には，様々な視点で貴重なご指導，ご助言を頂きました。心より御礼申し上げます。

三菱電機株式会社 関真規人氏，平澤宏祐氏，鈴木直彦氏，鹿毛裕史氏，佐藤和也氏とは，産学連携プロジェクトを通じて議論を重ね，多くの貴重なご助言を頂きました。また，検証実験に欠くことのできない機器やソフトウェアなどを貸与頂き，本研究の遂行に多大なるご協力を頂きました。ここに深く御礼申し上げます。

研究生生活全般においては，佐藤研究室の皆様に変にお世話になりました。皆様のご厚意に深く感謝申し上げます。

最後に，貴重な学生生活を支援してくれた両親，弟，義妹，そして，惜しみない協力と理解により，支えてくれた妻に心より感謝します。

平成 18 年 6 月 小林 貴訓

参考文献

- [1] 足立和正, 濱田尊裕, 中野倫明, 山本新, “ドライバの意識低下検知のための動画像処理によるまばたき計測,” 電気学会論文誌 C, vol.124, no.3, pp.776-783, 2004.
- [2] J. Aggarwal and S. Park, “Human Motion: Modeling and Recognition of Actions and Interactions,” *Proc. of 3D Data Processing Visualization and Transmission(3DPVT)*, pp640-647, 2004.
- [3] 麻生英樹, 津田宏治, 村田昇, “パターン認識と学習の統計学,” 岩波書店, pp.141-220, 2003.
- [4] S. Avidan, “Support Vector Tracking,” *IEEE Trans. Pattern Analysis and Machine Intelligence(PAMI2004)*, vol.26, no.8, pp.1064-1072, 2004.
- [5] H. Baker, “Three-dimensional Modelling,” *Proc. 5th International Joint Conference on Artificial Intelligence*, pp.649-655, 1977.
- [6] D. Beymer and T. Poggio, “Face Recognition From One Example View,” *A.I. Memo*, no.1536, Artificial Intelligence Laboratory, MIT, 1995.
- [7] S. Birchfield, “Elliptical Head Tracking Using Intensity Gradients and Color Histograms,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR98)*, pp.232-237, 1998.
- [8] J. Black, D. Makris and T. Ellis, “Validation of Blind Region Learning and Tracking,” *Proc. of International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, vol.1, pp9-16, 2005.
- [9] J.Y. Bouguet, “Camera Calibration Toolbox for Matlab,”
http://www.vision.caltech.edu/bouguetj/calib_doc/.
- [10] L. Breiman, “Bagging Predictors,” *Machine Learning*, vol.24, no.2, pp.123-140, 1996.
- [11] G.J. Brostow and R. Cipolla, “Unsupervised Bayesian Detection of Independent Motion in Crowds,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR2006)*, vol.1, pp.594-601, 2006.
- [12] Y. Cai, N. Freitas and J. Little, “Robust Visual Tracking for Multiple Targets,” *Proc. of European Conference on Computer Vision(ECCV2006)*, vol.4, pp.107-118, 2006.
- [13] T.H. Chang and S. Gong, “Tracking Multiple People with a Multi-Camera System,” *Proc. of IEEE Workshop on Multi-Object Tracking*, pp.19-26, 2001.
- [14] T. Chateau, V.G. Belille, F. Chausse and J.T. Lapreste, “Real-Time Tracking with Classifiers,” *Proc. of International Workshop on Dynamical Vision in conjunction with*

- ECCV*, 2006.
- [15] R. Collins, A. Lipton, and T. Kanade, "A System for Video Surveillance and Monitoring," *Proc. of American Nuclear Society(ANS) 8th International Topical Meeting on Robotic and Remote Systems*, pp.25-29, 1999.
- [16] D. Comaniciu, V. Ramesh and P.Meer, "Kernel-based Object Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence(PAMI2003)*, vol.25, no.5, pp.564-575, 2003.
- [17] J. Cui, H. Zha, H. Zhao and R. Shibasaki, "Fusion of Detection and Matching Based Approaches for Laser Based Multiple People Tracking," *Proc. of International Conference on Computer Vision and Pattern Recognition(CVPR2006)*, vol.1, pp.642-649, 2006.
- [18] D. Cunado, M.S. Nixon, and J.N. Carter, "Automatic Extraction and Description of Human Gait Models for Recognition Purposes," *Computer Vision and Image Understanding*, vol.90, no.1, pp.1-41, 2003.
- [19] W. Eric and L. Grimson, "Inference of Non-Overlapping Camera Network Topology by Measuring Statistical Dependence," *Proc. of International Conference on Computer Vision(ICCV2005)*, vol.2, pp1842-1849, 2005.
- [20] Y. Freund and R. Schapire, "Decision-theoretic Generalization of On-line Learning and an Application to Boosting," *Journal of Computer and Systems Sciences*, vol.55, pp.119-139, 1997.
- [21] H. Gabow, "An Efficient Implementation of Edmonds' Algorithm for Maximum Matching on Graphs," *Journal of Association for Computing Machinery*, vol.23, pp.221-234, 1976.
- [22] P.F. Gabriel, J.G. Verly, J.H. Piater and A. Genon, "The State of the Art in Multiple Object Tracking Under Occlusion in Video Sequences," *Proc. of 6th International Conference on Information Processing in Sensor Networks*, pp.509-518, 2003.
- [23] 橋本猛, 呉海元, 谷内田正彦, "能動的カメラシステムによる顔認識のための最適顔画像の獲得," 画像の認識・理解シンポジウム (MIRU96) 講演論文集, vol.II, pp.37.42, 1996.
- [24] 橋本学, "人物の多面的動線認識とその応用," 電気学会論文誌, vol.125, no.11, pp.705-707, 2005.
- [25] L. K. Hansen and P. Salamon, "Neural Network Ensembles," *IEEE Trans. Pattern Analysis and Machine Intelligence(PAMI1990)*, vol.12, no.10, pp.993-1001, 1990.
- [26] S. Hayashi and O. Hasegawa, "Detecting Faces from Low Resolution Images," *Proc. of Asian Conference on Computer Vision(ACCV2006)*, vol.34, pp.787-796, 2006.
- [27] F. Heijden, "Consistency Checks for Particle Filters," *IEEE Trans. on Pattern Analysis and Machine Intelligence(PAMI2006)*, vol.28, no.1, pp.140-145, 2006.
- [28] G. Helmut, G. Michael and B. Horst, "Real-Time Tracking via On-line Boosting," *Proc. of British Machine Vision Conference(BMVC2006)*, vol.1, pp.47-56, 2006.
- [29] 平澤宏祐, 鹿毛裕史, 三輪祥太郎, 橋本学, "時空間的無矛盾性を考慮したマルチポイント顔認証による動線認識," 画像センシングシンポジウム (SSII2005), 2005.

- [30] 平澤宏祐, 鹿毛裕史, 三輪祥太郎, 羽島一夫, 橋本学, “ベストショット顔画像記録システムの開発,” 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU), vol.104, no.447, pp.61-66, 2004.
- [31] 財団法人ソフトピアジャパン, “顔画像データベース,” <http://www.softopia.or.jp>.
- [32] S. Honda, K. Fukui, K. Moriyama, S. Kurihara and M. Numao, “Extracting Human Behaviors with Infrared Sensor Network,” *Proc. of IEEE 4th International Conference on Networked Sensing Systems(INSS)*, pp.122-125, 2007.
- [33] A. Hopper, A. Harter and T. Blackie, “The Active Badge System,” *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems*, pp.533-534, 1993.
- [34] B. Horn and B. Schunck, “Determining Optical Flow,” *Artificial Intelligence*, vol.17, pp.185-203, 1981.
- [35] Y. Huang and I. Essa, “Tracking Multiple Objects Through Occlusions,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR2005)*, vol.2, pp.1182-1190, 2005.
- [36] 池亀幸久, 平野真, 玉木徹, 山本正信, “視野重複のないマルチカメラを用いた映像中の人物抽出と人物の確率的歩行経路推定,” 電子情報通信学会言語理解とコミュニケーション (NLC) 研究会技報, pp.43-48, 2005.
- [37] M. Isard and A. Blake, “Condensation - Conditional Density Propagation for Visual Tracking,” *International Journal of Computer Vision*, vol.29, no.1, pp.5-28, 1998.
- [38] M. Isard and J. MacCormick, “BramBLE: a Bayesian multiple-blob tracker,” *Proc. of IEEE International Conference on Computer Vision(ICCV2001)*, vol.2, pp.34-41, 2001.
- [39] O. Javed, Z. Rasheed, O. Alatas and M. Shah, “Knight: A Real Time Surveillance System for Multiple Overlapping and Non-Overlapping Cameras,” *Proc. of IEEE International Conference on Multimedia and Expo(ICME2003)*, vol.1, pp.649-652, 2003.
- [40] O. Javed, Z. Rasheed, K. Shafique, and M. Shah, “Tracking Across Multiple Cameras With Disjoint Views,” *Proc. of International Conference on Computer Vision(ICCV2003)*, vol.2, pp.952-958, 2003.
- [41] O. Javed, Z. Rasheed, K. Shafique, and M. Shah, “Appearance Modeling for Tracking in Multiple Non-overlapping Cameras,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR2005)*, pp.26-33, 2005.
- [42] 榎原靖, 佐川立昌, 向川康博, 越後富夫, 八木康史, “周波数領域における方向変換モデルを用いた歩容認証,” 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.48, no.SIG1(CVIM17), pp.78-87, 2007.
- [43] R. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Trans. of ASME Journal of Basic Engineering*, pp.35-45, 1960.
- [44] 加藤丈和, 深尾隆則, 羽下 哲司, “対象追跡-フレーム間の類似度に着目した手法から動きのモデルに着目した手法まで,” 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, vol.2005-CVIM-150-20, pp.185-198, 2005.
- [45] S. M. Khan and M. Shah, “A Multiview Approach to Tracking People in Crowded Scenes

- using a Planar Homography Constraint,” *Proc. of European Conference on Computer Vision(ECCV2006)*, vol.4, pp.133-146, 2006.
- [46] Z. Khan, T.R. Balch and F. Dellaert, “An MCMC-based Particle Filter for Tracking Multiple Interacting Targets,” *Proc. of European Conference on Computer Vision(ECCV2004)*, pp.279-290, 2004.
- [47] 小阪勇氣, 平山高嗣, 岩井儀雄, 谷内田正彦, “全方位画像を用いた歩行軌跡による確率的人物行動認識システム,” 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, vol.2005-CVIM-149, pp.163-170, 2005.
- [48] 栗田多喜夫, 島井博行, 高橋隆史, 重原孝臣, “向きに依存しない顔画像認識のための向きに依存した複数の識別器の統合,” 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU99-170), pp.35-42, 1999.
- [49] G. Loy, L. Fletcher, N. Apostoloff and A. Zelinsky, “An Adaptive Fusion Architecture for Target Tracking,” *Proc. of 5th IEEE International Conference on Automatic Face and Gesture Recognition(FG2002)*, pp.261-265, 2002.
- [50] B. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision,” *Proc. of 7th International Joint Conference on Artificial Intelligence(IJCAI)*, pp.674-679, 1981.
- [51] J. MacCormick and A. Blake, “A Probabilistic Exclusion Principle for Tracking Multiple Objects,” *International Journal of Computer Vision*, vol.39, no.1 pp.57-71, 2000.
- [52] D. Makris, T. Ellis, and J. Black, “Bridging the Gaps between Cameras,” *Proc. of Computer Vision and Pattern Recognition(CVPR2004)*, vol.2, pp.205-210, 2004.
- [53] J.S. Marques, P.M. Jorge, A.J. Abrantes, J.M. Lemos, “Tracking Groups of Pedestrians in Video Sequences,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR2003)*, vol.9, pp.101-107, 2003.
- [54] 株式会社サイバネットシステム “MATLAB,” <http://www.cybernet.co.jp/matlab/>.
- [55] 松原康晴, 尺長健, “疎テンプレートマッチングとその実時間物体追跡への応用,” 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.46, no.SIG9(CVIM11), pp.60-71, 2005.
- [56] 松本郁佑, 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 コンピュータビジョンとイメージメディア研究会, vol.2005-CVIM-150-20, pp.161-168, 2005.
- [57] 松本郁佑, 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,” 画像の理解・認識シンポジウム (MIRU2006), pp.501-506, 2006.
- [58] Y. Matsumoto and A. Zelinsky, “An Algorithm for Realtime Stereo Vision Implementation of Head Pose and Gaze Direction Measurement,” *Proc. of 4th International Conference on Automatic Face and Gesture Recognition(FG2000)*, pp.499-504, 2000.
- [59] L. Morency, A Rahimi and T.Darrell, “Adaptive View-based Appearance Models,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recogni-*

- tion(CVPR2003), vol.1, pp.803-810, 2003.
- [60] 森大樹, 内海章, 大谷淳, 谷内田正彦, 中津良平, “非同期多視点画像による人物追跡システムの構築,” 電子情報通信学会論文誌 D-II, vol.J84-D-II no.1, pp.102-110, 2001.
- [61] 中澤篤志, 加藤博一, 日浦慎作, 井口征士, “分散視覚エージェントを用いた複数人物追跡システム,” 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.42, no.11, pp.2699-2710, 2001.
- [62] 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,” 第11回画像センシングシンポジウム 講演論文集, pp.177-180, 2005.
- [63] K. Nickel, T. Gehrig, R. Stiefelhagen and J. McDonough, “A Joint Particle Filter for Audiovisual Speaker Tracking,” *Proc. of 7th International Conference on Multimodal Interfaces*, pp.61-68, 2005.
- [64] 野口真身, 加藤丈和, “マーカの軌跡を用いた非同期カメラの幾何学的・時間的キャリブレーション,” 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.47, no.SIG5(CVIM13), pp.59-68, 2006.
- [65] 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.
- [66] 岡兼司, 菅野裕介, 佐藤洋一, “頭部変形モデルの自動構築をともなう実時間頭部姿勢推定,” 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.47, no.SIG10(CVIM15), pp.185-194, 2006.
- [67] K. Okuma, A. Taleghani, N. Freitas, J. Little and D. Lowe, “A Boosted Particle Filter: Multitarget Detection and Tracking,” *Proc. of European Conference on Computer Vision(ECCV2004)*, vol.3021 of LNCS, pp.28-39, 2004.
- [68] 大沢裕ほか, “Image processing : 画像処理標準テキストブック,” 画像情報教育振興協会, 1997.
- [69] 小沢一雅, “情報理論の基礎,” 国民科学社, pp.3-33, 1980.
- [70] V. Pavlovic, R. Sharma, and T. Huang, “Visual Interpretation of Hand Gestures for Human-computer Interaction: A Review,” *IEEE Trans. on Pattern Analysis and Machine Intelligence(PAMI1997)*, vol.19, no.7, pp.677-695, 1997.
- [71] J.C. Platt, “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods,” *Advances in Large Margin Classifiers*, 1999.
- [72] P. Prez, J. Vermaak and A. Blake, “Data Fusion for Visual Tracking with Particles,” *Proc. of the IEEE*, vol.92, no.3, pp.495-513, 2004.
- [73] A. Rahimi, B. Dunagan, and T. Darrell, “Simultaneous Calibration and Tracking with a Network of Non-Overlapping Sensors,” *Proc. of Computer Vision and Pattern Recognition(CVPR2004)*, pp.187-194, 2004.
- [74] A. Rahimi, B. Dunagan, and T. Darrell, “Tracking People with a Sparse Network of Bearing Sensors,” *Proc. of European Conference on Computer Vision(ECCV2004)*, pp.507-518, 2004.
- [75] O. D. Rishard, E. H. Peter, and G. S. David, “パターン識別,” 新技術コミュニケーション

- ンズ, pp.161-172, pp.483-488, 2001.
- [76] N. Robertson and I. Reid, "Estimating Gaze Direction from Low-Resolution Faces in Video," *Proc. of European Conference on Computer Vision(ECCV2006)*, vol.2 pp.402-415, 2006.
- [77] 佐川立昌, 越後富夫, 八木康史, "人物追跡のための周波数解析による歩行画像列マッチング," 画像の認識・理解シンポジウム (MIRU2005), pp.1373-1379, 2005.
- [78] S. Sarkar, J. P. Phillips, Z. Vega, P. Grother and K. W. Bowyer, "The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis," *Trans. of Pattern Analysis and Machine Intelligence(PAMI2005)*, vol.27, no.2, pp.162-177, 2005.
- [79] 佐藤哲, 和田俊和, 加藤丈和, "MCMC/EM アルゴリズム/MDL を用いた床圧力センサからの複数人物位置追跡," 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, vol.2005-CVIM-150-19, pp.153-160, 2005.
- [80] 佐藤雄隆, 坂上勝彦, "Bi-polar Radial Reach Correlation によるロバスト背景差分," 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU), vol.104 no.741, pp.73-78, 2005.
- [81] 佐藤淳, "コンピュータビジョン-視覚の幾何学-", コロナ社, pp.118-156, 1999.
- [82] Y. Shan, H.S. Sawhney and R. Kumar, "Vehicle Identification between Non-Overlapping Cameras without Direct Feature Matching," *Proc. of International Conference on Computer Vision(ICCV2005)*, vol.1, pp378-385, 2005.
- [83] V. D. Shet, D. Harwood and L. S. Davis, "Multivalued Default Logic for Identity Maintenance in Visual Surveillance," *Proc. of European Conference on Computer Vision(ECCV2006)*, vol.4, pp.119-132, 2006.
- [84] J. Sherrah and S. Gong, "Fusion of Perceptual Cues for Robust Tracking of Head Pose and Position," *Pattern Recognition*, vol.34, no.8, 2001.
- [85] 下田宏, 國弘威, 吉川榮和, "動的顔画像からのリアルタイム表情認識システムの試作," ヒューマンインタフェース学会論文誌, vol.1, no.2, pp.25-32, 1999.
- [86] ジック株式会社, "セイフティレーザスキャナ S300," <http://www.sick.jp>.
- [87] G. Slabaugh B. Culbertson T. Malzbender and R. Schafer, "A Survey of Methods for Volumetric Scene Reconstruction from Photographs," *Technical Report 1, Center for Signal and Image Processing, Georgia Institute of Technology*, pp.81-100, 2001.
- [88] 三菱電機株式会社, "RFID 応用スマートシェルフシステム," 三菱電機技報 2005 年 1 月号, 2005.
- [89] 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イメージメディア, vol.43 no.SIG04(CVIM4), pp.69-84, 2002.
- [90] 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), 2007.
- [91] J. Sullivan and S. Carlsson, "Tracking and Labelling of Interacting Multiple Targets," *Proc. of European Conference on Computer Vision(ECCV2006)*, vol.3, pp.619-632, 2006.

- [92] C. Stauffer, W. Eric and L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking" *IEEE Trans. on Pattern Analysis and Machine Intelligence(PAMI2000)*, vol.22, no.8, pp.747-757, 2000.
- [93] 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 D-II, vol.J88-D-II no.8, pp.1592-1600, 2005.
- [94] N. Suzuki, K. Hirasawa, K. Tanaka, Y. Kobayashi, Y. Sato and Y. Fujino, "Learning Motion Patterns and Anomaly Detection by Human Trajectory Analysis," *Proc. of IEEE International Conference on Systems, Man and Cybernetics(SMC2007)*, 2007
- [95] T. Svoboda, D. Martinec and T. Pajdla, "A convenient multi-camera self-calibration for virtual environments," *PRESENCE: Teleoperators and Virtual Environments*, vol.14 no.4, pp.407-422, 2005.
- [96] 高塚皓正, 田中正行, 奥富正敏, "カスケード型識別器のキャリブレーションと尤度分布を利用した顔検出," 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU), vol.106, no.539, pp.45-54, 2007.
- [97] 寺沢征彦, 山澤一誠, 竹村治雄, 横矢直和, "3台の全方位画像センサを使った重なり判定による人物追跡," 電子情報通信学会ソサイエティ大会, no.D-12-58, 2000.
- [98] E. Trucco and A. Verri, "Introductory Techniques for 3-D Computer Vision," Prentice Hall, 1998.
- [99] D. Tweed and A. Calway, "Tracking Many Objects Using Subordinated CONDENSATION," *Proc. of British Machine Vision Conference(BMVC2002)*, pp.283-292, 2002.
- [100] 内田恭輔, 三浦純, 白井良明, "雑踏における複数歩行者の追跡," システム制御情報学会論文誌, vol.14, no.4, pp.180-185, 2001.
- [101] 上田修功, "アンサンブル学習," 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.46, no.SIG15(CVIM12), pp.11-20, 2005.
- [102] 浮田宗伯, 松山隆司, "能動視覚エージェント群による複数対象の実時間協調追跡," 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.43, no.SIG11(CVIM5), pp.64-79, 2002.
- [103] 浮田宗伯, "広域分散カメラ群の連結関係推定法," 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, vol.2005, no.88 2005-CVIM-150, pp.177-184, 2005.
- [104] J. Vermaak, A. Doucet and P. Perez, "Maintaining Multi-Modality through Mixture Tracking," *Proc. of IEEE International Conference on Computer Vision(ICCV2003)*, vol.2, pp.1110-1116, 2003.
- [105] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR2001)*, vol.1, pp.511-518, 2001.
- [106] Y. Wang, J. Wu and A. Kassim, "Particle Filter for Visual Tracking Using Multiple Cameras," *Proc. of IAPR Conference on Machine Vision Applications*, pp.298-301, 2005.
- [107] T. Watanabe and M. Yachida "Real Time Gesture Recognition Using Eigenspace

- from Multi Input Image Sequences,” *Proc. of 3rd IEEE International Conference on Automatic Face and Gesture Recognition(FG1998)*, pp.428-433, 1998.
- [108] 宇小軍, 和田俊和, 東海彰吾, 松山隆司, “平面間透視投影を用いた並列視体積交差法,” 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.42, no.SIG6(CVIM2), pp.33-43, 2001.
- [109] B. Wu and R. Nevatia, “Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR2005)*, vol.1, pp.90-97, 2005.
- [110] 藪内勉, 岩井儀雄, 谷内田正彦, “複数の能動的カメラを用いた人物の抽出と追跡,” 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, vol.1998-CVIM-109-3, pp.17-24, 1998.
- [111] K.Yachi, T. Wada and T. Matsuyama, “Human head tracking using adaptive appearance models with a fixed-viewpoint pan-tilt-zoom camera,” *Proc. of IEEE 4th International Conference on Automatic Face and Gesture Recognition(FG2000)*, pp.150-155, 2000.
- [112] C. Yam, M. S. Nixon, and J. N. Carter, “Automated Person Recognition by Walking and Running via Model-based Approaches,” *Pattern Recognition*, vol.37, no.5, pp.1057-1072, 2004.
- [113] C. Yang, R. Duraiswami and L. Davis, “Fast Multiple Object Tracking via a Hierarchical Particle Filter,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, vol.1, pp.212-219, 2005.
- [114] T. Yang, S.Z. Li, Q. Pan, J. Li, “Real-time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR2005)*, pp.970-975, 2005.
- [115] 安本護, 本郷仁志, 渡辺博己, 山本和彦, 輿水大和, “マルチカメラ統合を用いた人物識別と顔方向推定,” 電子情報通信学会論文誌 D-II, vol.J84-D-II, pp.145-150, 2001.
- [116] 依田育士, 坂上勝彦, “高次局所自己相関特徴を使った多眼画像からの顔とその向き認識,” 第5回画像センシングシンポジウム講演論文集, pp.203-208, 1999.
- [117] T. Yu and Y. Wu, “Collaborative Tracking of Multiple Targets,” *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR2004)*, vol.1, pp.834-841, 2004.
- [118] Z. Zhang, “Flexible Camera Calibration By Viewing a Plane From Unknown Orientations,” *Proc. of International Conference on Computer Vision(ICCV99)*, pp.666-673, 1999.

発表文献

本論文の内容に関連する発表文献および特許申請は以下の通りである。

学術雑誌

- 1) 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 D-II, Vol.J88-D-II, No.8, pp.1592-1600, August 2005.
- 2) T. Suzuki, S. Iwasaki, Y. Kobayashi, Y. Sato and A. Sugimoto, “Incorporating Environment Models for Improving Vision-Based Tracking of People,” *Systems and Computers in Japan*, Vol.38, No.1, January 2007.
- 3) 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(採録決定), Vol.J90-D-II, No.8, August 2007.
- 4) 鈴木 直彦, 平澤 宏祐, 田中 健一, 小林 貴訓, 佐藤洋一, 藤野陽三, “人物動線データ群における逸脱行動人物検出および行動パターン分類,” 電子情報通信学会論文誌 D(条件付採録).

国際会議

- 1) Y. Kobayashi, D. Sugimura, Y. Sato, H. Hirasawa, N. Suzuki, H. Kage and A. Sugimoto, “3D Head Tracking using the Particle Filter with Cascaded Classifiers,” *Proc. of the British Machine Vision Conference (BMVC2006)*, pp.37-46, September 2006.
- 2) N. Suzuki, K. Hirasawa, K. Tanaka, Y. Kobayashi, Y. Sato and Y. Fujino, “Learning Motion Patterns and Anomaly Detection by Human Trajectory Analysis,” *to appear in Proc. of IEEE International Conference on Systems, Man and Cybernetics (SMC2007)*, 2007

国内会議

- 1) 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 コンピュータビジョンとイメージメディア研究会, 2005-CVIM-150-21, pp.169-176, September 2005.
- 2) 鈴木 直彦, 平澤 宏祐, 田中 健一, 小林 貴訓, 佐藤 洋一, 藤野 陽三, “Hidden Markov Model を用いた逸脱行動人物検出,” 電子情報通信学会 パターン認識・メディア理解研究会, vol.106, no.99, PRMU2006-46, pp.43-48, June 2006.
- 3) 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), July 2006.
- 4) 小林 貴訓, 佐藤 洋一, 杉本 晃宏, “疎分散カメラ群を用いた人物行動軌跡の推定,” 画像の認識・理解シンポジウム (MIRU2006), July 2006.
- 5) 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, November 2006.
- 6) 鈴木 直彦, 平澤 宏祐, 田中 健一, 小林 貴訓, 佐藤 洋一, 藤野 陽三, “人物動線データ分析による逸脱行動人物の検出,” 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, 2007-CVIM-158-15, pp.109-115, March 2007.
- 7) 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), July 2007.
- 8) 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), July 2007.

特許

- 1) 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.
- 2) 鈴木 直彦, 小林 貴訓, 佐藤 洋一: 移動体行動解析装置, 特願 2007-061560, 2007.