

NAIST-IS-DD0461030

博士論文

実環境における実時間状態推定のための
適応的サンプリング法

坂東 誉司

2007年3月23日

奈良先端科学技術大学院大学
情報科学研究科 情報生命科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学) 授与の要件として提出した博士論文である。

坂東 誉司

審査委員：

石井 信 教授 (主指導教員)

木戸出 正繼 教授 (副指導教員)

柴田 智広 助教授 (副指導教員)

実環境における実時間状態推定のための 適応的サンプリング法*

坂東 誉司

内容梗概

直接観測することができない隠れ状態の時系列推定問題において、パーティクルフィルタは各時刻における隠れ状態の分布をサンプルとその重みにより近似することで計算困難な時系列ベイズ推定を効率良く行うことができる。ただし、その推定精度と計算量の間にはトレードオフの関係があり、特に高次元状態変数について精度の良い推定を行うためには、一般に状態空間の次元の指数オーダーのサンプル数が必要となる。従って、実環境における頑健な隠れ状態推定を実時間でを行うためには、少数に制限されたサンプルを如何に有効に分布の近似に用いることができるかが重要な課題となる。そのためには推定対象や環境の正確なモデルが必要とされるが、これらは一般に誤差（モデル化誤差）を含むため、モデル化誤差の存在を前提とした推定アルゴリズムを開発する必要がある。本研究では、幾つかのモデルに基づいて構築した相補的な予測分布をサンプル生成のためのサンプリング分布として採用し、これらを適応的に選択/混合することにより頑健な実時間隠れ状態推定を行うことを考える。観測重視/予測重視の、もしくは問題の階層構造を利用して異なるモデルにより作成した相補的なサンプリング分布を、推定状態の確信度や推定への寄与の大きさを規準として混合することで、サンプル数を少数に抑えながらモデル化誤差を吸収して頑健な実時間状態推定の実現を図る。ここでは、実環境における視覚追跡課題、画像列からのドライバの頭部姿勢推定等の実問題を通して提案手法の有効性を確認した。

キーワード

パーティクルフィルタ, 変分ベイズ推定, EM アルゴリズム, 視覚追跡, 姿勢推定, 実時間処理

*奈良先端科学技術大学院大学 情報科学研究科 情報生命科学専攻 博士論文, NAIST-IS-DD0461030, 2007年3月23日.

Adaptive Sampling Method for Real-time State Estimation in the Real World*

Takashi Bando

Abstract

Particle Filters (PFs) is a sample-based approach to Bayesian estimation of intractable posterior distributions from time series signals with non-Gaussian noises. One of the major drawbacks of PFs is that a large number of samples are generally required for accurate estimation of state variables lying in a high dimensional space, which is time-consuming. An implementation of effective approximation with a few number of samples is key issue for solving this difficulty, and it requires the proposal distribution based on accurate models of estimation target and environment, which is generally disturbed with modelling error under the real environment.

This study presents two kinds of adaptive proposal distribution for absorbing the modelling error. One is a dynamic-switching scheme of two proposal distributions, which emphasis on prediction or observation, according to the variance of the current estimated state as confidence in estimation. Another is a mixture of complementary prediction densities computed with some individual models which are constructed based on hierarchy of high-dimensional state variables, and its mixture ratio corresponding to the reliability of each layer is determined by means of an on-line EM algorithm. These adaptive proposal distributions enable robust and accurate estimation with a few numbers of samples, which is demonstrated by computer simulations as well as real experiments of read-ball tracking and driver's head pose estimation in a real car, respectively.

Keywords:

Particle Filtering, Variational Bayes, EM algorithm, Visual Tracking, Pose Estimation, Real-Time Processing

*Doctoral Dissertation, Department of Bioinformatics and Genomics, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0461030, March 23, 2007.

目次

第1章	序論	1
1.	研究背景	1
1.1	隠れ状態推定問題	1
1.2	高次元隠れ状態の実時間推定問題	4
2.	本論文の構成	8
第2章	時系列ベイズ推定	10
1.	時系列ベイズ推定	10
2.	解法 1: カルマンフィルタ	13
2.1	定式化	13
2.2	数値実験	14
2.3	まとめ	15
3.	解法 2: パーティクルフィルタ	16
3.1	定式化	16
3.2	数値実験	22
3.3	まとめ	25
第3章	サンプリング分布の動的な切り替えによる頑健な視覚追跡	27
1.	いくつかのパーティクルフィルタ	28
2.	Switching Particle Filters	30
2.1	動的なサンプリング分布の切り替えと柔らかな切り替え手法	30
2.2	観測モデル	33
2.3	状態遷移モデルとそのオンライン学習	34
3.	数値実験	35

3.1	追跡頑健性	36
3.2	追跡精度	38
3.3	サンプル数	39
3.4	議論	40
4.	応用: 実環境内でのボール追跡	43
4.1	実験環境	43
4.2	追跡対象のダイナミクス学習	44
4.3	遮蔽や類似物体の存在する環境での追跡	45
5.	まとめ	47
第4章 適応的サンプリング分布とその実環境における高次元状態推定への 応用		49
1.	階層的隠れ状態推定	51
2.	適応的混合サンプリング分布	52
3.	数値実験:剛体の三次元姿勢推定問題	55
3.1	上位層における剛体の姿勢推定	57
3.2	下位層における特徴点位置推定	60
3.3	実験結果	62
3.4	まとめ	64
4.	応用:実際の車内で撮影されたドライバの頭部画像列からの頭部姿 勢推定	67
4.1	適応的サンプリング分布を採用したドライバ頭部画像列か らの頭部姿勢推定	69
4.2	三次元顔モデル	71
4.3	画像からの顔特徴点観測モデル	72
4.4	実験結果	74
4.5	まとめ	76
5.	ドライバの三次元顔モデル推定	80
5.1	因子分解法	81
5.2	変分ベイズを用いた確率論的因子分解法	82

5.3	数値実験	91
5.4	実画像を用いた実験	96
5.5	まとめ	97
第 5 章	議論	99
第 6 章	結言	105
	謝辞	108
	参考文献	110
	付録	123
A.	略語表	123
B.	EM アルゴリズムを用いたダイナミクスパラメータのオンライン 推定	123
C.	EM アルゴリズムを用いた混合比の推定	128
D.	アフィン変換パラメータからの頭部姿勢の復元	131
D.1	パーティクルフィルタによる推定	131
D.2	拡張カルマンフィルタを用いた手法	132
E.	テイラー展開を用いた射影行列の逐次近似	137
F.	変分ベイズを用いた因子分解法	141
F.1	VB-E ステップ	141
F.2	VB-M ステップ	144
F.3	自由エネルギー	146
	研究業績	148

目次

1.1	SPFs の枠組み	4
1.2	ASD の枠組み	8
2.1	観測変数 a と隠れ変数 b の関係を表すグラフィカルモデル	10
2.2	隠れ変数 x_t と観測変数 z_t の関係を表すグラフィカルモデル	11
2.3	カルマンフィルタによる時系列状態推定結果	15
2.4	モンテカルロ法を用いた π の近似計算	17
2.5	重点サンプリング法	18
2.6	逐次的な再サンプリングを伴うパーティクルフィルタ	21
2.7	パーティクルフィルタを用いた時系列状態推定結果	23
2.8	推定に用いたサンプル数と試行内平均絶対推定誤差	24
2.9	状態空間の次元数とサンプル数, 推定誤差の関係	24
3.1	Condensation と APF によりサンプリングされたサンプル	29
3.2	h-SPFs のアルゴリズム	31
3.3	s-SPFs における APF サンプル数	33
3.4	実環境をモデル化したシミュレーション画像	36
3.5	遮蔽の幅, 類似物体の大きさについての追跡成功率	36
3.6	χ^2 検定による追跡成功率の差の検定結果	37
3.7	遮蔽の幅, 類似物体の大きさについての平均絶対追跡誤差	38
3.8	Wilcoxon の順位和検定による平均絶対追跡誤差の差の検定結果	39
3.9	サンプル数の変化に伴う, 遮蔽の幅, 類似物体の大きさについての追跡成功率の変化	39

3.10	サンプル数の変化に伴う，遮蔽の幅，類似物体の大きさについての平均絶対追跡誤差の変化	40
3.11	h-SPFs と s-SPFs におけるサンプル数の変化に伴う追跡成功率と成功試行内平均絶対誤差の変化	41
3.12	追跡対象のダイナミクス学習実験で用いた実画像例	44
3.13	ダイナミクス学習の進行の様子	45
3.14	実験環境と追跡成功を定義する領域	46
3.15	追跡対象のダイナミクス学習と追跡頑健性	47
3.16	追跡対象のダイナミクス学習と推定精度	48
4.1	階層モデルのグラフィカルモデル	51
4.2	数値実験環境	55
4.3	剛体の姿勢推定のグラフィカルモデル	56
4.4	階層的な剛体の姿勢推定手法の実行手順	58
4.5	シミュレーション画像	63
4.6	推定された剛体の姿勢	64
4.7	各時刻で推定された混合比 $\alpha_{a,t}$	65
4.8	各手法の平均推定誤差	66
4.9	ドライバの頭部姿勢推定の実験環境	70
4.10	実環境中で計測された人物の頭部画像例	70
4.11	ドライバの平均三次元顔モデル	72
4.12	テンプレートとの正規化相関値とサンプルの重みの関係	73
4.13	頭部姿勢の平均推定誤差	74
4.14	提案手法に推定された顔特徴点位置	75
4.15	各時刻で推定された混合比 $\alpha_{a,t}$	77
4.16	平均顔モデルと個人顔モデルを用いた頭部姿勢推定法の追跡成功率の比較	79
4.17	確率的因子分解法の実行手順	86
4.18	確率的因子分解法の実行手順（続き）	87
4.19	自由エネルギーの計算手順	88

4.20	推定された特徴点位置の時系列データ例	92
4.21	因子分解法により復元された三次元構造	93
4.22	確率的因子分解法により復元された三次元構造	93
4.23	確率的逐次因子分解法により復元された三次元構造	94
4.24	形状推定誤差の 50 試行の平均値と分散	94
4.25	計算時間の比較	95
4.26	三次元顔の推定誤差の平均値と分散 (50 試行)	96
4.27	確率的逐次因子分解法により復元された三次元構造	97

表目次

4.1 各特徴点テンプレートの大きさ	73
A.1 略語表	124

第1章 序論

1. 研究背景

1.1 隠れ状態推定問題

直接観測することの出来ない隠れ状態の逐次推定問題に対して、時系列ベイズ推定法が近年広く適用され、その有効性が示されてきた。ただし、時系列ベイズ推定法は計算困難な隠れ状態についての積分を含むため、隠れ状態が連続変数である場合には何らかの近似が用いられることが一般的である。状態変数が線形ガウス過程に従う時には確率分布にガウス分布を仮定することで、上記積分計算を解析的に行うことができる [42] (カルマンフィルタ)。しかし、視覚追跡等多くの実環境における状態推定問題においては、隠れの存在など多くの非ガウス性非定常雑音の存在する影響で、線形ガウス過程を仮定することは困難である場合が多く、こうした場合にカルマンフィルタを用いて頑健な状態推定を実現することは困難である [35]。この様な問題に対していくつかの改善手法が提案され [39, 78]、視覚追跡等の非線形/非ガウス性の強い実環境問題に適用されているが [10, 49, 60, 61]、近年の計算機の進捗に伴い、先述の積分をサンプル近似することで非ガウスな分布も扱うことが可能なパーティクルフィルタ [28] が広く用いられるようになってきた。しかし、パーティクルフィルタはサンプル近似による手法であるため、その推定精度、計算時間はサンプル数についてトレードオフの関係にある。更に問題が高次元の状態推定である場合には、推定精度を維持するために大量のサンプルが必要となり、実時間での状態推定の実現は非常に困難な課題となる。この点はパーティクルフィルタの中心的な問題点として、様々な分野で研究が行われ、多くの解決手法が提案されている [16, 86, 41]。これらの手法の多くでは、サンプル数が十分に準備されていることを前提としてその性能を議論しているが、実際

に実環境における実時間隠れ状態推定問題への適用を考えた場合，十分なサンプルを準備することは難しい．近年の計算機の進捗に伴い，実時間で扱うことのできるサンプル数も増加しているため，パーティクルフィルタは多くの実問題に適用されてきているが，特に状態空間が高次元になる場合には状態空間の次元数の増加に対して推定精度を維持するために必要なサンプル数が指数的に増加することから，効率の良いサンプリングを行うことは依然として重要な課題として残っている．

パーティクルフィルタは，実環境における状態推定問題としてロボット位置推定問題 [20, 24, 23, 81, 82, 73, 89] や視覚追跡など広い分野への適用が報告されているが，本論文では実時間視覚追跡問題を中心に，その頑健な実現を図るための手法について考察を行う．

実時間視覚追跡

視覚追跡はコンピュータビジョンの分野で中心的な課題として，ロボット視覚 [94, 95] などマンマシンインタフェースへの応用 [49] や交通監視システム [100, 101]，ビデオ会議システム [4, 11, 59, 90]，スポーツ映像中継システム [65] など幅広い分野への応用が考えられることから，非常に多くの研究が行われてきた．追跡対象も剛体 [56] や昆虫 [43] など様々であるが，とりわけ，主にマンマシンインタフェースへの応用を考えた，画像列からの人物の顔追跡や頭部姿勢推定は多くの研究が行われている [4, 6, 7, 12, 13, 29, 30, 55, 80, 99]．テンプレートマッチングに代表される伝統的な手法を用いた視覚追跡手法も提案されているが [40, 56]，近年，不確実な雑音等を頑健に扱うことのできるカルマンフィルタを応用した確率論的手法が多く用いられるようになり，その有効性も報告されている [10, 49, 61, 60]．さらに近年では，実環境における実時間視覚追跡問題の様に，遮蔽の存在や照明条件の変化，複雑な背景など，非ガウス雑音の多く存在する環境での頑健な追跡の実現のためにパーティクルフィルタを用いた手法が数多く提案されている．パーティクルフィルタを実時間視覚追跡に適用した最も基本的なアルゴリズムであり，Sampling Importance Resampling (SIR) 法 [28] を用い，標準的な PC での実時間処理を可能としていた Condensation [35] を始めとして，追跡

対象の形状に注目して追跡を行う手法 [35, 105, 50] , 追跡対象の周辺領域のヒストグラムに注目して追跡を行う手法 [18, 19, 64, 65] , 追跡対象を主成分分析等により特定の部分空間へと射影することで、「見え方」の変化する対象についても頑健な追跡を実現する手法 [9, 43, 51, 52, 69, 92] , もしくは形状や色など複数の情報を混合することで、唯一の情報を用いた手法と比べて頑健な追跡を実現する手法 [11, 14, 36, 74, 75] , 追跡対象として複数の物体を同時に追跡するための手法 [34, 65, 85] などが提案されている。

本論文第 3 章では、追跡対象の観測モデル等を可能な限り簡略化し、パーティクルフィルタにおけるサンプリング分布の違いが実時間視覚追跡の様な実環境における課題に対してどのような影響を与えるかを調べ、パーティクルフィルタを一般の実問題に適用する上で頑健で推定精度の良いサンプリング分布を構築することを考える。以下で具体的なアプローチについて説明する。

アプローチ

先述の通り、時系列ベイズ推定において計算困難であった積分をサンプルを用いて近似するパーティクルフィルタは、近年、様々な分野で広く用いられ、様々なサンプリング法を利用した手法が報告されている。ここでは、この中で特に Condensation [35] と Auxiliary Particle Filter [41] (APF) の二つの手法に注目する。Condensation はパーティクルフィルタを実時間視覚追跡に適用した最も基本的なアルゴリズムであり、Sampling Importance Resampling (SIR) 法 [28] を用いて現時刻の観測情報と独立なサンプリング分布からサンプルを生成するため、予測分布と観測情報との間のズレが大きい時には事後分布の推定精度が著しく低下するという問題点がある。APF は、この点をサンプリング分布に観測情報を導入することで改善した手法である。現時刻の観測と独立にサンプルを得るため、予測を重視していると考えられる Condensation と、サンプリング分布に観測情報を導入することで観測情報を重視してサンプルを得る APF は、想定できなかった雑音によるモデル化誤差が予測モデルや観測モデルに存在する実時間視覚追跡特有の状況下では、それぞれのモデル化誤差に対する頑健さの点から相補的であると考えられる。そこで、ここではこの二つの手法を状況に合わせて補い合

い，両手法の長所を活用するために単純な規準に従ってこれらの手法を動的に切り替える (Switching Particle Filters, SPFs) ことを提案する (cf. 図 1.1) . また，こ

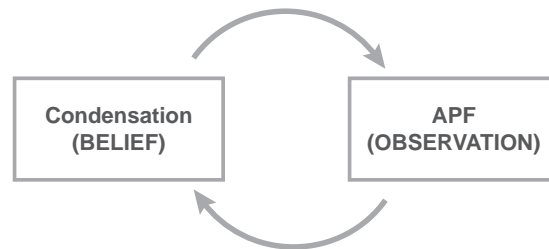


図 1.1 SPFs の枠組み．ここでは，自身の予測を重視する Condensation と観測を重視する APF の動的な切り替えを考える．

の SPFs において，各時刻において採用するサンプリング法を完全に切り替える手法 (h-SPFs) に対して，各サンプリング法でサンプルするサンプル数を徐々に変化させる手法 (s-SPFs) と比較することで，計算資源の極端に限られた状態において，どのような切り替え手法が適しているのか考察を行う．更に，画像列から追跡対象のダイナミクスを学習する手法 [62, 63, 67] をオンライン EM アルゴリズム [72] を利用してオンライン化することでサンプリング分布を逐次的に改善していく手法についても提案し，SPFs と組み合わせることでより正確な状態推定の実現が可能であることについて述べる．ここでは，追跡対象として赤いボールと単純な対象を採用し，最も簡単な観測モデルの一つである単純な色情報を観測モデルに採用することで，追跡対象のモデルの複雑さが状態推定に与える影響を極力排除し，純粋なサンプリング法の違いによる追跡性能の違いを数値実験と実際の環境における実実験を通して確認する．更に，各手法の推定精度と頑健性について統計的手法を用いて評価することで，より一般的な性能の評価を試みる．

1.2 高次元隠れ状態の実時間推定問題

前節ではサンプリング分布の性能を純粋に比較するため，赤いボールという単純な対象を追跡対象として扱い，可能な限り単純化した観測モデルを仮定していた．ところが，例えば人間の頭部や手の様な，より複雑な対象を追跡対象とする

実環境における応用課題を考えると，追跡対象のモデルが高次元となる場合が多い．先述の通り，パーティクルフィルタはサンプル近似を利用した手法であるため，近似する状態空間が高次元である場合には，精度の良い近似を行うためには大量のサンプルを利用する必要がある．しかしながら，実時間での処理の実現のためには十分なサンプルを準備することは困難である場合が多い．以下では，実際の車内におけるドライバの頭部姿勢推定をこの様な高次元状態推定課題の例として，前節と同様に複数のサンプリング分布を利用することで，状態空間の大きさと比較して遥かに少数のサンプルを有効活用し，実環境に多く存在するモデル化誤差に対して頑健で精度の良い実時間状態推定の実現を図る手法について説明する．

ドライバの頭部姿勢推定

近年，より安全な車社会の実現へ向けて，ドライバの認知能力を補助するため，ドライバの視線方向推定等のドライバの状態推定を盛り込んだドライバ支援システムの研究が盛んに行われるようになってきている [3, 17, 77, 57, 98, 106]．今後高齢化社会の進行と共に，体調や年齢に応じて低下することのある認知能力の補助機構はますます重要になってくると考えられるが，実際の車内で直接観測することのできるドライバの情報は少なく，視線方向など重要な情報は直接観測できない場合も多い．これを実時間で推定することが出来れば，安全な車社会の実現に向けて大きな一歩となると考えられるため，多くの取り組みがなされてきた．しかし，西日や高速道路の照明など照明条件の厳しい変化や，ドライバ自身の手や顔向きの変化による顔特徴点の隠れなど大きな雑音に加え，極端に制限された計算資源などにも関わらず，正確に視線方向を推定するためには頭部姿勢と顔特徴点をそれぞれ数 $^{\circ}$ ，数ピクセルの精度で同時推定する必要がある．この様に，大きな非ガウス雑音の存在や計算資源の制限など厳しい環境にも関わらず，高次元状態空間を持つ対象を高精度で推定しなければならず，非常に困難な課題となっている．

人物の（頭部）姿勢推定 [4, 6, 12, 46, 48, 47, 55, 99]，顔検出 [7, 53, 88]/追跡等 [11, 29, 30, 31, 90, 91] や近年はマンマシンインタフェースとりわけ Attentive

User Interfaces[87] としても重要な情報である視線方向推定技術 [93, 102, 66, 96] は、広い応用分野を持つため過去多くの研究がなされてきた。先述の通り、ドライバの視線方向推定を行うためには、非ガウス雑音の多々存在する実環境においても正確に頭部姿勢や顔特徴点位置を知る必要があるが、複雑な背景を持つ実環境においての人物の（頭部）姿勢推定や顔検出/追跡をマルコフ連鎖モンテカルロ (MCMC) 法 [46, 48, 47] やパーティクルフィルタ [4, 12, 55, 75, 99] などの確率論的手法を用いることで頑健に行う手法が多く提案されている。

しかしながら、これらのドライバ状態推定手法や人物の頭部姿勢推定手法は、カラー情報を必要とするもの [46, 48, 47, 77]、ステレオ視のためカメラを複数台必要とするもの [3, 57] など、実際の車内での使用を考える際には車内に乗せることの出来るカメラの制約や、更には極端に制限された計算資源の問題などから直接適用することは困難な場合がある。そこで、本研究では照明条件の変化に比較的頑健であることの知られている近赤外カメラにより実際の車内で観測された単眼グレー画像列から、ドライバの視線方向推定のためのドライバの頭部姿勢と複数の顔特徴点の画像面上における位置を、実際の車内で処理を行うために極端に制限された計算資源にも関わらず、実時間で精度の良い同時推定を頑健に行うことを考える。ここでは、非ガウス雑音を扱うことのできるパーティクルフィルタを頭部姿勢と複数の顔特徴点位置の同時推定手法として採用し、実環境における頑健で高速な推定の実現を図るが、このためには高次元状態空間を扱う必要がある。パーティクルフィルタを用いて高次元状態空間で精度の良い推定をするためには、状態空間の次元数の指数オーダーのサンプルが必要とされており、極端に制限された計算資源では精度の良い推定を実現することは困難である。この問題に対して、ここでは、頭部姿勢と顔特徴点位置の間に存在する問題の階層構造を利用することにより、サンプルを有効に活用し、実時間での頑健な推定の実現を図る。次節で具体的なアプローチについて簡単に説明する。

アプローチ

本研究では、頭部姿勢と画像面上での顔特徴点位置がカメラによる射影を通じて頭部姿勢を上位、顔特徴点位置を下位とす階層構造を持つモデルで表現できる

ことを利用して、頭部姿勢と顔特徴点位置の頑健な推定の実現を図る。

頭部姿勢と顔特徴点位置の関係に見られるように、物体姿勢が変化すれば物体の特徴点の画像面上における位置が変化するという様に画像の生成モデルが階層構造を持つことはごく一般的であるが、実環境の中では状態遷移モデルや観測モデルに用いた仮定や近似の崩れ、想定できなかった雑音等により、推定精度が低下する場合がしばしば考えられる。とりわけ、パーティクルフィルタの様なサンプル近似に基づいた手法において、実時間処理のためサンプル数が厳しく制限される場合には、上位層から階層構造を通じて得られる下位層のサンプリング分布の精度が上記の要因により低下し、推定に寄与するサンプル数が極端に低下することにより生ずる推定精度の低下が階層構造を通して他の階層へと伝播してしまうことがある。これに対して、各階層に独立なダイナミクスを考え、階層構造を通じて得られる上位層の影響を排除し、下位層のサンプリング分布は下位層における予測のみを用いて得られるモデルを考えると、今度は逆に下位層のダイナミクスモデルの精度が低下がサンプリング分布の精度低下に直結しているため、結局、推定の頑健性を低下させることになる。このように、状態推定に重要な役割を果たすサンプリング分布を一つのモデルのみから得る手法では、様々な仮定や近似の崩れ、大きな雑音の存在する実環境において、モデル化誤差が大きくなった場合には著しく推定の頑健性を損なってしまう。

そこで本研究では、各階層に並列なダイナミクスを仮定し、ある層のサンプリング分布を、その他の階層から階層構造を通じて得られた予測分布の混合分布として表現することを考える。この混合サンプリング分布の混合比は、推定への寄与を規準として、オンライン EM アルゴリズム [72] により逐次推定する。この様に、サンプリング分布を適応的に生成する (Adaptive Sampling Distribution, ASD) ことにより、各階層で仮定したモデルが十分な精度を持つ層からの予測分布の混合比は適応的に大きくなり、逆にモデルの精度が低下している層からの予測分布は適応的に小さくなり、頑健で正確な推定の実現が可能となる。この枠組みは図 1.2 に示す様に、異なる階層のモデルを用いて作成した予測分布を適応的に選択/混合する仕組みと表現することが可能で、SPFs と同様の構造と考えることが出来る (cf. 図 1.1)。ここでは、特徴点の隠れを含む画像列からの剛体の三次元姿勢

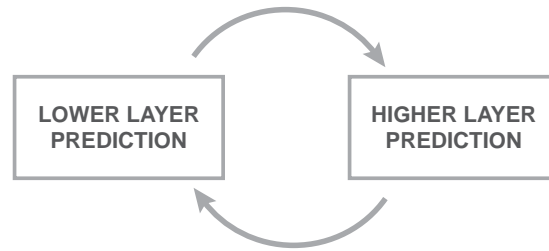


図 1.2 ASD の枠組み．ここでは，各階層に仮定したモデルにより得られた複数の予測分布をそれぞれの信頼度に応じて適応的に使い分けることを考える．

推定と特徴点位置推定問題の数値実験，実際の車内で撮影された画像列からのドライバの頭部姿勢と顔特徴点の位置推定問題を通して，適応的サンプリング分布を用いた状態推定が頑健で精度の良い推定の実現に有効であることを示す．

2. 本論文の構成

本論文は以下で構成される．

第 1 章：本章 第 1 章では，研究背景として隠れ状態推定問題に対する時系列ベイズ推定の適用と，その際用いられる幾つかの近似について説明した．加えて，実環境における実時間状態推定課題に取り組むに当たって問題となる点を明確にし，本論文で提案する手法の採用するアプローチについて簡単に説明した．

第 2 章 第 2 章では一般のベイズ推定から時系列ベイズ推定の説明を行い，その計算困難性と近似解法としてカルマンフィルタ，パーティクルフィルタを紹介する．また，カルマンフィルタ，パーティクルフィルタを実問題に適用する上での問題点を明確にし，本論文で提案する手法の立場を明確にする．

第 3 章 第 3 章では，予測重視のサンプリング手法と観測重視のサンプリング手法を適応的に切り替えることで，実環境中でより頑健な状態推定（ここでは視覚追跡）を行う手法の提案を行う．ここでは，予測重視のサンプリング法として，現時刻の観測と独立なサンプリング分布を採用した Condensation を用い，観測重

視のサンプリング法として，サンプリング分布に観測情報を導入した Auxiliary Particle Filter を用いる．この適応的なサンプリング分布の切り替え手法が，実環境における視覚追跡課題において有効であることを確認するために，実環境を模した数値実験を通して，追跡頑健性とその精度を統計的に評価する．また，追跡対象のダイナミクスを逐次学習する手法の提案を行い，サンプリング分布の切り替え手法と組み合わせてその性質について議論する．

第 4 章 第 4 章では，高次元状態推定問題におけるパーティクルフィルタの問題点を解決するために問題の持つ階層性を考慮した適応的サンプリング分布の紹介と，その高次元状態推定問題への適用について説明する．ここでは高次元状態推定問題として，剛体姿勢と画像面上における特徴点位置の推定問題を採用し，数値実験により階層構造を考慮した適応的サンプリング分布を採用した手法といくつかの従来手法を比較することで，提案手法の有効性を示す．更に，提案手法を実際の環境で撮影された画像を入力とする，実車内におけるドライバの頭部姿勢と目や鼻など顔特徴点位置の同時推定問題に適用する．加えて，より厳しい実環境における課題として，先術のドライバの頭部姿勢と顔特徴点位置の同時推定問題においてドライバの三次元顔モデルが未知の場合について述べる．

第 5 章 第 5 章では，実環境における実時間状態推定問題に対して，広く用いられてきたパーティクルフィルタの問題点をもう一度明確にし，これに対して本論文で述べてきたサンプリング分布の切り替え手法の有効性について定性的な議論を行う．

第 6 章 最後に第 6 章で本論文をまとめる．

なお，本論文で用いた略語については付録 A に略語表を示したので参照されたい．

第2章 時系列ベイズ推定

直接観測することのできない隠れ状態の逐次推定問題に対して、時系列ベイズ推定が近年広く適用され、その有効性が示されてきた。本章では一般の時系列ベイズ推定について簡単に紹介し、その計算困難性と、その各種解法の内、本論文で扱うカルマンフィルタとパーティクルフィルタについて最も簡単なものを紹介する。各手法の隠れ状態推定に対する有効性を数値実験によって示し、各手法の持つ特徴と問題点について紹介し、3章以降で考える実問題への準備とする。

1. 時系列ベイズ推定

今、 a, b を1次元連続確率変数とする時、確率の乗法公式より a, b の同時確率密度分布 $p(a, b)$ は

$$p(a, b) = p(a|b)p(b) = p(b|a)p(a) \quad (2.1)$$

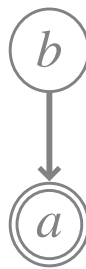


図 2.1 観測変数 a と隠れ変数 b の関係を表すグラフィカルモデル。 b が隠れ変数を、 a が観測変数を示す。

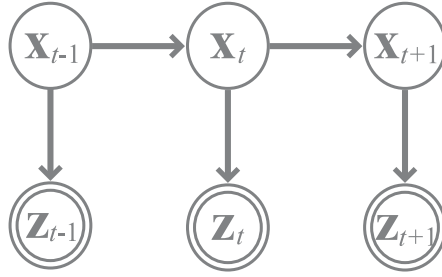


図 2.2 隠れ変数 x_t と観測変数 z_t の関係を表すグラフィカルモデル． x_t のダイナミクスにマルコフ性を，観測 z_t の x_t についての条件付き独立性を仮定している．

で得ることができる．ここで， $p(b), p(a)$ をそれぞれ b, a の事前分布， $p(a|b), p(b|a)$ をそれぞれ b, a の尤度と呼ぶ．この時上式 (2.1) を変形すれば

$$p(b|a) = \frac{p(a|b)p(b)}{p(a)} \quad (2.2)$$

を得る．これをベイズの定理と呼び， b が直接観測することのできな隠れ変数である場合にも a が観測されれば， a が観測された時の b の尤度 $p(a|b)$ と b の事前分布 $p(b)$ ，周辺尤度 $p(a)$ を用いれば確率変数 b の a を与件とする事後分布 $p(b|a)$ が得られることを示している．

さて，ここまでは単純な 1 次元連続確率変数についてのベイズ推定を説明したが，以降では特に時系列の多次元連続確率変数についてのベイズ推定について説明する．今，直接観測することの出来ない隠れ変数ベクトル $\mathbf{x}_t \in \mathcal{R}^{N_x}$ の時系列を $\mathbf{x}_{0:t} = \{\mathbf{x}_0, \dots, \mathbf{x}_t\}$ とし，観測ベクトル $\mathbf{z}_t \in \mathcal{R}^{N_z}$ の時系列を $\mathbf{z}_{1:t}$ とする．この時，式 (2.2) より，観測系列 $\mathbf{z}_{1:t}$ を得た下での $\mathbf{x}_{0:t}$ の事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ は，ベイズの定理をそのまま用いて

$$p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{\int p(\mathbf{z}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})d\mathbf{x}_{0:t}} \quad (2.3)$$

で得ることができる．

更に図 2.2 に示すように，観測 z_t が各時刻の状態 x_t について条件付き独立であるとして，

$$p(\mathbf{z}_{1:t}|\mathbf{x}_{0:t}) = \prod_{i=1}^t p(z_i|\mathbf{x}_i)$$

が成り立つとし，また，隠れ変数 x_t の状態遷移にマルコフ性を仮定して，

$$p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t}) = p(\mathbf{x}_{t+1}|\mathbf{x}_t)$$

が成り立つとすれば，式 (2.3) の時系列ベイズ推定は以下の通り再帰的な形で

$$\begin{aligned} p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t}) &= \frac{p(\mathbf{z}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{\int p(\mathbf{z}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})d\mathbf{x}_{0:t}} \\ &= p(\mathbf{x}_{0:t-1}|\mathbf{z}_{1:t-1}) \frac{p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{p(\mathbf{z}_t)} \end{aligned}$$

と書き直すことが出来る．加えて，各時刻で $x_{1:t}$ の事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ を求め，時刻 0 から t までの系列について推定を行うことは時刻の経過と共に計算量の増大を招くため，時刻 t においては周辺分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ を考える形に書き直せば，以下の予測と推定の 2 ステップにより記述することが出来る．予測ステップは x_t の事前分布 $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$ を得るステップであり，

$$p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{x}_{t-1} \quad (2.4)$$

と，前時刻 $t-1$ における事後分布 $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ をダイナミクス $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ に従って遷移させる形で記述することができる．推定ステップは尤度 $p(\mathbf{z}_t|\mathbf{x}_t)$ を用いて事前分布 $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ を修正し，事後分布 $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ を得るステップであり，以下で書くことができる．

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1})}{\int p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1})d\mathbf{x}_t} \quad (2.5)$$

但し，式 (2.4) 中の x_{t-1} に関する積分，式 (2.5) 中の x_t に関する積分は，共に多次元連続状態変数に関する積分となっており，一般には計算困難である．このため，実際の問題に適用するにあたって多くの解法が提案されており，以降では代表的な例として，カルマンフィルタとパーティクルフィルタについて簡単に説明する．

2. 解法 1: カルマンフィルタ

2.1 定式化

前節で説明した通り，時系列ベイズ推定は計算困難な隠れ状態についての積分を含むため，隠れ状態が連続変数である場合には何らかの仮定や近似を用いることが一般的である．本節では，線形ガウスシステムを仮定することで上記計算困難な時系列ベイズ推定を行うカルマンフィルタについて簡単に説明する．

ここでは時刻 t における隠れ状態変数ベクトル x_t ，観測変数ベクトル z_t に関して，式 (2.6),(2.7) で示す線形システム方程式を考える．

$$x_t = Ax_{t-1} + Bu_t, \quad u_t \sim \mathcal{N}_{N_x}(\mathbf{0}, I_{N_x}) \quad (2.6)$$

$$z_t = Cx_t + Dv_t, \quad v_t \sim \mathcal{N}_{N_z}(\mathbf{0}, I_{N_z}) \quad (2.7)$$

ここで， A は $N_x \times N_x$ 次状態遷移行列， B はプロセスノイズの大きさを制御する $N_x \times N_x$ 次ノイズ標準偏差行列， C は $N_z \times N_x$ 次観測行列， D は観測ノイズの大きさを制御する $N_z \times N_z$ 次観測ノイズ標準偏差行列である．また， $\mathcal{N}_N(\mu, \Sigma)$ は平均 μ ，分散 Σ の N 次元ガウス分布を表し， I_N は $N \times N$ 次の単位行列を表す．この時，状態遷移過程は

$$p(x_t | x_{t-1}) = \mathcal{N}_{N_x}(Ax_{t-1}, BB^T)$$

で，観測過程は

$$p(z_t | x_t) = \mathcal{N}_{N_z}(Cx_t, DD^T)$$

と，ガウス分布で表現することができ，式 (2.4) で表される時系列ベイズ推定の予測ステップは

$$p(x_t | z_{1:t-1}) = \mathcal{N}_{N_x}(\bar{x}_t, \bar{V}_t) \quad (2.8)$$

$$\bar{x}_t = A\hat{x}_{t-1}$$

$$\bar{V}_t = BB^T + A\hat{V}_{t-1}A^T$$

と、式(2.5)で表される推定ステップは

$$\begin{aligned}
 p(\mathbf{x}_t | \mathbf{z}_{1:t}) &= \mathcal{N}_{N_x}(\hat{\mathbf{x}}_t, \hat{V}_t) & (2.9) \\
 \hat{\mathbf{x}}_t &= \bar{\mathbf{x}}_t + K_t(\mathbf{z}_t - C\bar{\mathbf{x}}_t) \\
 \hat{V}_t &= (I - K_t C)\bar{V}_t \\
 K_t &= \bar{V}_t C^T (D D^T + C \bar{V}_t C^T)^{-1}
 \end{aligned}$$

と変形することが出来る．ここで、 K_t はカルマンゲインと呼ばれ、予測と観測のバランスを取る重要なパラメータである．この様に、システム方程式に線形を、プロセス/観測ノイズにガウスノイズを仮定することで、一般には計算困難であった積分計算を解析的に解くことができ、上記の通り効率良く計算することができる．

2.2 数値実験

カルマンフィルタによる隠れ状態推定の有効性を確認するため、以下の設定で数値実験を行った．時刻 t における真の隠れ状態 x_t 、観測 z_t は

$$\begin{aligned}
 x_t &= \sin\left(\frac{2\pi t}{360}\right) \\
 z_t &= x_t + 0.1v_t, \quad v_t \sim \mathcal{N}_1(0, 1)
 \end{aligned}$$

に従って生成されるとする．この時、以下のシステム方程式を仮定して隠れ状態 x_t の時系列推定をカルマンフィルタを用いて行った．実際には状態変数 x_t は1次元の連続状態変数ベクトルであるが、ここでは状態遷移過程に線形2次のマルコフ過程を仮定するため、擬似的に状態変数の次元を1次元増やし、

$$\mathbf{x}_t = \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix}$$

と新たに置きなおす．この新たな隠れ状態変数 \mathbf{x}_t について以下の線形システム方程式を仮定する．

$$\begin{aligned}
 \mathbf{x}_t &= A\mathbf{x}_{t-1} + B\mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}_2(\mathbf{0}, I_2) \\
 z_t &= C\mathbf{x}_t + Dv_t, \quad v_t \sim \mathcal{N}_1(0, I_1)
 \end{aligned}$$

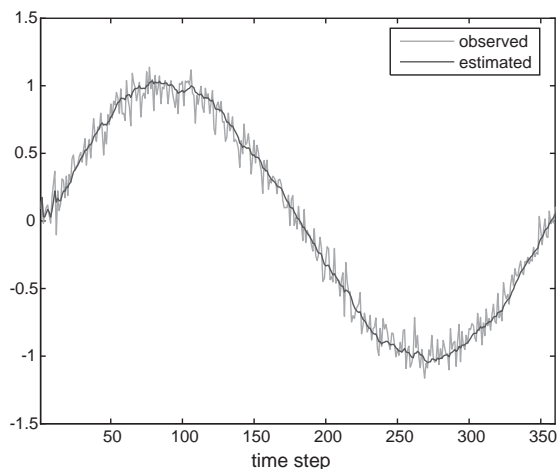


図 2.3 カルマンフィルタによる時系列状態推定結果．雑音のモデルが既知であれば，雑音で汚れた観測信号から現信号を非常に高い精度で復元することが可能．

ここで，

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0.001 & 0 \\ 0 & 0 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad D = 0.1$$

とした．観測 z_t と推定された状態 x_t を図 2.3 に示す．大きなノイズの存在にも関わらず，線形システムにおいて雑音のモデルが既知であれば，非常に高い精度で隠れ状態の推定を行うことが可能であることが解る．

2.3 まとめ

以上，ここでは時系列ベイズ推定を，線形ガウス過程を仮定することで解析的に行うことのできるカルマンフィルタを簡単に説明した．カルマンフィルタは，ここで数値実験を通して確認したように，線形ガウス過程を仮定できる問題に対しては非常に精度よく隠れ状態推定を行うことができる．ただし，本研究で主に扱うような視覚追跡問題等，実環境における隠れ状態推定問題には非線形過程であったり非ガウス非定常雑音が存在することが多く，そうした場合，隠れ状態の

分布は非ガウス分布となるため，カルマンフィルタの直接的な適用は困難である．この点はカルマンフィルタの中心的な課題として多くの研究が行われてきた．

非線形過程への適用については，テイラー展開を利用してシステム方程式の線形近似を行った上でカルマンフィルタを適用する拡張カルマンフィルタや，シグマポイントと呼ばれる代表点の遷移を通して線形過程を近似する Unscented Kalman Filter 等改善手法が多く提案されており，それぞれ様々な分野で利用されている．

非ガウス雑音の存在に対しては，複数のカルマンフィルタを同時に適用する混合カルマンフィルタや分布をサンプルで近似することにより分布の形状についての仮定を用いないパーティクルフィルタ等が提案されている．パーティクルフィルタに関しては，状態遷移/観測過程に線形性を仮定する必要も無く，近年視覚追跡をはじめ，多くの応用分野で適用されている．次節ではこのパーティクルフィルタについて簡単な説明を行う．

3. 解法 2: パーティクルフィルタ

3.1 定式化

前節では，式 (2.4),(2.5) で表される時系列ベイズ推定に存在する，一般には計算困難な隠れ状態 x_t, x_{t-1} に関する積分を，分布の形状にガウス分布を，システム方程式に線形を仮定することにより，解析的に計算を行うカルマンフィルタについて簡単に説明した．本節では計算困難であった積分を多数のサンプルを用いてモンテカルロ近似することで，時系列ベイズ推定を実現するパーティクルフィルタについて簡単に説明する．

多数のサンプルを用いた近似計算は古くから多く行われており，代表的な例としては図 2.4 に示す様に，サンプル数の比で面積比を近似し円周率 π を近似的に得る方法が知られている．今， $x_{0:t}$ の事後分布 $p(x_{0:t}|z_{1:t})$ からサンプル $\{x_{0:t}^{(i)}, i = 1, \dots, N\}$ が得られたとすれば，事後分布 $p(x_{0:t}|z_{1:t})$ は

$$p(x_{0:t}|z_{1:t}) \approx \frac{1}{N} \sum_{i=1}^N \delta_{x_{0:t}^{(i)}}(x_{0:t})$$

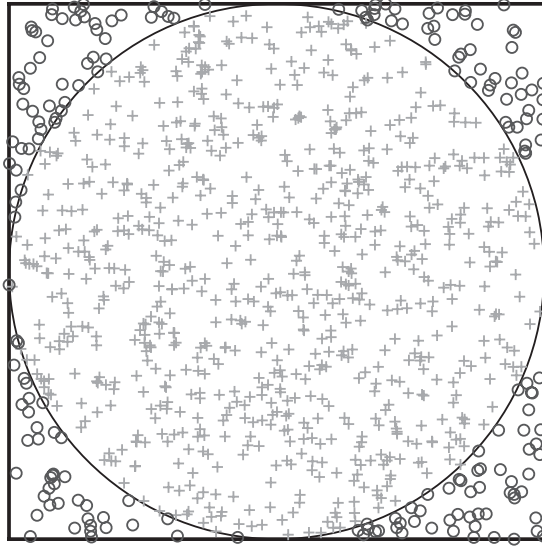


図 2.4 モンテカルロ法を用いた π の近似計算．一様分布からサンプルを生成し，正方形に内接する円に含まれるサンプル（+のサンプル）数と正方形に含まれるサンプル（+のサンプルと のサンプル）数の比が面積比と等しいとすることで π を近似的に知ることができる．

とサンプルを用いて近似することができる．ここで， $\delta_{\mathbf{x}_{0:t}}^{(i)}$ はディラクのデルタ関数であり

$$\delta_{\mathbf{x}_{0:t}}^{(i)} = \begin{cases} \infty & (\mathbf{x}_{0:t} = \mathbf{x}_{0:t}^{(i)}) \\ 0 & (\mathbf{x}_{0:t} \neq \mathbf{x}_{0:t}^{(i)}) \end{cases}$$

を満たす．この時，ある関数 $f(\mathbf{x}_{0:t})$ の $\mathbf{x}_{0:t}$ の事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ に関する期待値は

$$\begin{aligned} I(f) &= \int f(\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t}) d\mathbf{x}_{0:t} \\ &\approx \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_{0:t}^{(i)}) \end{aligned} \quad (2.10)$$

で近似することが出来る．しかし一般には，任意の確率分布からのサンプル生成は困難である場合が多く，そうした場合には上記の近似法を直接適用することは困難である．そのため，正規分布等サンプルの生成が簡単な確率分布から生成したサンプルを利用して任意の確率分布からのサンプル生成を行う手法は多く提案

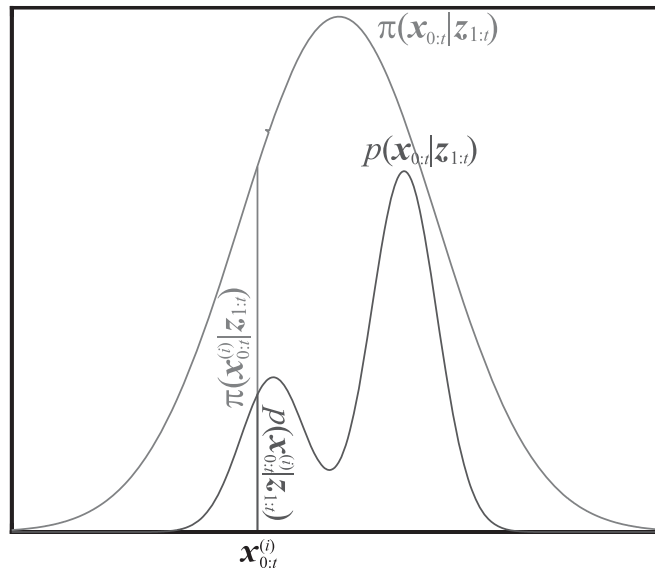


図 2.5 重点サンプリング法．直接 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ からサンプルを得ることが困難な場合，新たにサンプリングが容易なサンプリング分布 $\pi(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ を導入し，得られたサンプル $\mathbf{x}_{0:t}^{(i)}$ を $p(\mathbf{x}_{0:t}^{(i)}|\mathbf{z}_{1:t})$ ， $\pi(\mathbf{x}_{0:t}^{(i)}|\mathbf{z}_{1:t})$ により重みづけすることで，近似的に $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ に従うサンプルを得る．

されており，代表的なものとしては棄却サンプリング，適応的棄却サンプリング，重点サンプリング，ギブスサンプリング，マルコフ連鎖モンテカルロ法などが挙げられる．ここでは，計算コスト，時系列ベイズ推定との親和性などの点から重点サンプリングを利用した手法について説明する．

今， $\mathbf{x}_{0:t}$ の事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ に従うサンプル $\{\mathbf{x}_{0:t}^{(i)}, i = 1, \dots, N\}$ を得たいが，事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ から直接サンプルを得ることは困難だとする．そこで，図 2.5 に示す様に容易にサンプル可能な分布 $\pi(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ をサンプリング分布として導入し，得られたサンプルを重み

$$w(\mathbf{x}_{0:t}) = \frac{p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})}{\pi(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})}$$

を用いて修正することで，事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ とサンプリング分布 $\pi(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ の誤差を修正し，近似的に事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ に従うサンプルを得ることを考える．こ

の時，式 (2.10) の事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ に関する期待値は上記重みを用いて

$$\begin{aligned} I(f) &= \frac{\int f(\mathbf{x}_{0:t})w(\mathbf{x}_{0:t})\pi(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})d\mathbf{x}_{0:t}}{\int w(\mathbf{x}_{0:t})\pi(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})d\mathbf{x}_{0:t}} \\ &\approx \frac{\frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_{0:t}^{(i)})w(\mathbf{x}_{0:t}^{(i)})}{\frac{1}{N} \sum_{j=1}^N w(\mathbf{x}_{0:t}^{(j)})} \\ &= \sum_{i=1}^N f(\mathbf{x}_{0:t}^{(i)})\tilde{w}(\mathbf{x}_{0:t}^{(i)}) \end{aligned}$$

と書くことができる．ここで， $\tilde{w}(\mathbf{x}_{0:t}^{(i)})$ はサンプル $\mathbf{x}_{0:t}^{(i)}$ が持つ正規化重みで，

$$\tilde{w}(\mathbf{x}_{0:t}^{(i)}) = \frac{w(\mathbf{x}_{0:t}^{(i)})}{\frac{1}{N} \sum_{j=1}^N w(\mathbf{x}_{0:t}^{(j)})}$$

である．また，事後分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ はサンプルと正規化重みを用いて

$$p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t}) \approx \sum_{i=1}^N \tilde{w}(\mathbf{x}_{0:t}^{(i)})\delta_{\mathbf{x}_{0:t}^{(i)}}(\mathbf{x}_{0:t})$$

とディラクのデルタ関数の重みつき重ね合わせで近似することができる．

この重点サンプリング法を周辺分布 $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ のベイズ推定に適用して， $\pi(\mathbf{x}_t|\mathbf{z}_{1:t}) = p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t-1})$ とすれば，式 (2.5) より

$$\begin{aligned} w(\mathbf{x}_t) &= \frac{p(\mathbf{x}_t|\mathbf{z}_{1:t})}{\pi(\mathbf{x}_t|\mathbf{z}_{1:t})} \\ &\propto \frac{p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t-1})}{\pi(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})} \\ &\propto p(\mathbf{z}_t|\mathbf{x}_t) \end{aligned}$$

となる．この時，式 (2.4),(2.5) の 2 ステップは

$$p(\mathbf{x}_t^{(i)}|\mathbf{z}_{1:t-1}) = p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})p(\mathbf{x}_{t-1}^{(i)}|\mathbf{z}_{1:t-1}) \quad (2.11)$$

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \approx \sum_{i=1}^N \tilde{w}(\mathbf{x}_t^{(i)})\delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t) \quad (2.12)$$

となり，事前分布 $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t-1})$ から生成したサンプルと，尤度に比例した正規化重み $\tilde{w}(\mathbf{x}_t)$ を用いて，周辺事後分布 $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ に従うサンプルを得ることができる．但し，

$$\tilde{w}(\mathbf{x}_t^{(i)}) \propto \tilde{w}(\mathbf{x}_{t-1}^{(i)})p(\mathbf{z}_t|\mathbf{x}_t^{(i)}) \quad (2.13)$$

とする．この様に，重点サンプリングを繰り返すことにより，各時刻での事後分布 $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ を近似するサンプルを得る手法を逐次重点サンプリング法 (Sequential Importance Sampling, SIS) と呼ぶ．但し，SIS 法には以下の大きな問題点があることが知られている．式 (2.13) で示される SIS 法の正規化重みは各時刻で前時刻の正規化重みに尤度 $p(\mathbf{z}_t|\mathbf{x}_t^{(i)})$ を掛け，更に正規化を行う形で定義されるため，数十ステップも時刻が過ぎれば $\tilde{w}(\mathbf{x}_t^{(i)})$ の分布が尖り，ある一つのサンプルに対応する重みのみが 1 で，残りのサンプルに対応する重みが 0 になってしまう．これは分布を一つのサンプルで近似することになり，近似精度の著しい低下を招く．この問題に対して，

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{z}_{1:t}) &\approx \sum_{i=1}^N \tilde{w}(\mathbf{x}_t^{(i)})\delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t) \\ &\approx \frac{1}{N} \sum_{i=1}^N N_t^{(i)}\delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t) \end{aligned}$$

の様に，各時刻で小さな重みを持ち分布の近似に貢献しないサンプルを淘汰し，大きな重みを持つサンプルを多数生成して（再サンプリング），それぞれのサンプルに対応する重みを正規化することで，多数のサンプルの重みが小さい値になること防ぎ，近似精度を維持する手法が Gordon らにより提案されている [28]．上記の他にも様々な再サンプリングアルゴリズムが提案されているが [32, 44, 54]，その多くは上述の手法と同様に

$$E[N_t^{(i)}] = N\tilde{w}(\mathbf{x}_t^{(i)})$$

が成り立つように再サンプリングを行う手法である．逐次的な再サンプリングを伴うパーティクルフィルタアルゴリズムを図 2.6 に示しておく．

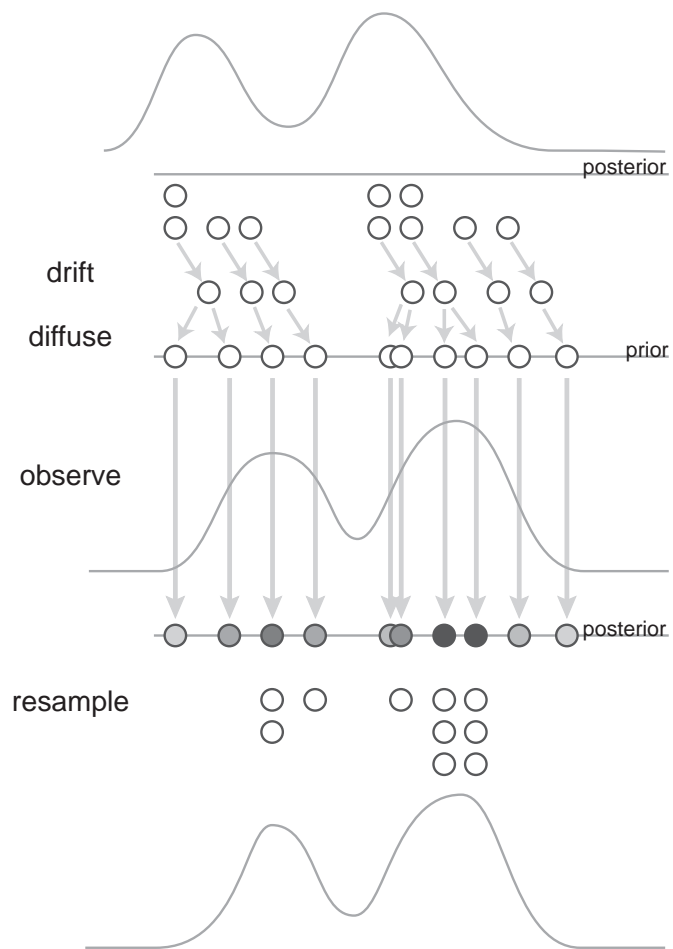


図 2.6 逐次的な再サンプリングを伴うパーティクルフィルタ。 はサンプルを，色の濃さは重みの大きさを示す。

3.2 数値実験

パーティクルフィルタによる隠れ状態推定の有効性を確認するため、以下の設定で数値実験を行った。時刻 t における真の隠れ状態 x_t 、観測 z_t が

$$x_t = \cos\left(\frac{2\pi t}{360}\right)$$
$$z_t = x_t + 0.1v_t, \quad v_t \sim \mathcal{N}_1(0, 1)$$

に従って生成されるとする。この時、以下のシステム方程式を仮定して隠れ状態 x_t の時系列推定をパーティクルフィルタを用いて行った。実際には状態変数 x_t は 1 次元の連続状態変数ベクトルであるが、ここでは状態遷移過程に線形 2 次のマルコフ過程を仮定するため、2.2 節と同様に擬似的に状態変数の次元を 1 次元増やし、

$$\mathbf{x}_t = \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix}$$

と新たに置きなおす。この新たな隠れ状態変数 \mathbf{x}_t について以下のシステムを仮定する。

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + B\mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}_2(\mathbf{0}, I_2)$$
$$z_t = C\mathbf{x}_t + Dv_t, \quad v_t \sim \mathcal{N}_1(\mathbf{0}, I_1)$$

ここで、

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0.001 & 0 \\ 0 & 0 \end{pmatrix},$$
$$C = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad D = 0.1$$

とした。また数値実験に用いるサンプル数は $N = 10, 100$ とした。観測 z_t と推定された状態 x_t を図 2.7 に示す。ここでは非常に簡単な数値実験を扱ったため、サンプル数が少数の場合でも比較的精度の良い推定を行うことができているが、サンプル数が大きい場合の方が精度良く推定できている。更に詳しくサンプル数と推定

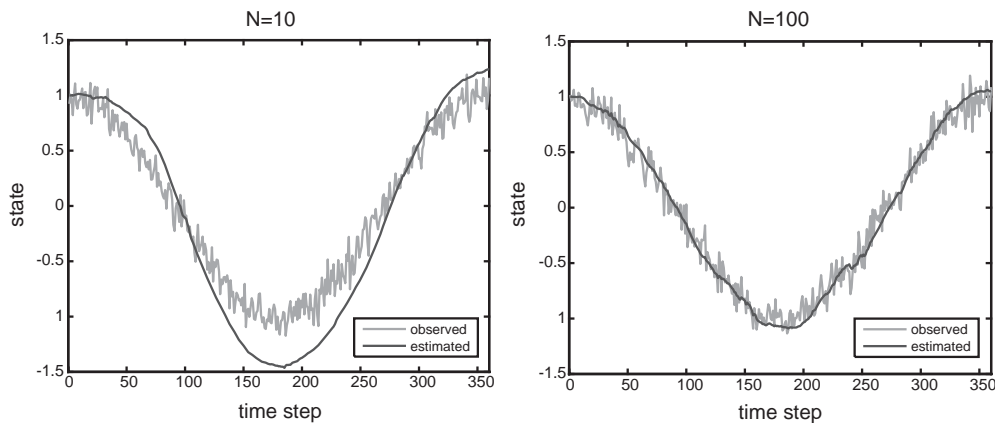


図 2.7 パーティクルフィルタを用いた時系列状態推定結果．左図は $N = 10$ の場合，右図は $N = 100$ の場合．

精度の関係を調べるために，サンプル数を $N = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100$ として，それぞれ先の数値実験に 100 試行ずつ適用した．図 2.8 にサンプル数毎の試行内平均絶対誤差の分布を示す．サンプル数の増加と共に推定誤差が低くなり，その分布の分散も小さくなる傾向が見えており，安定して正確な状態推定を行うことができていることがわかる．

次に，状態空間の次元数と推定精度，サンプル数の関係を調べるために，先の実験での状態 x_t の次元を一次元ずつ増加させながら同様の状態推定課題を行った．状態遷移モデル，観測モデル，その他のパラメータ等は先の実験と同じものを用いた．図 2.9 に状態空間の広さとサンプル数，推定誤差の関係を示す．サンプル数を固定して考えた場合，状態空間の次元数の増加に対して一試行内での平均推定誤差が上昇していくのがわかる．サンプル数を $N = 10, 100, 1000$ と増加させれば推定精度はある程度維持できるが， $N = 1000$ の場合でも状態空間の次元が 5 次元にもなれば，推定誤差が大きくなってきている．この様に，状態空間の高次元化に対して，精度の維持に必要なサンプル数の増加が指数的事であることはパーティクルフィルタにおける大きな問題点の一つである．

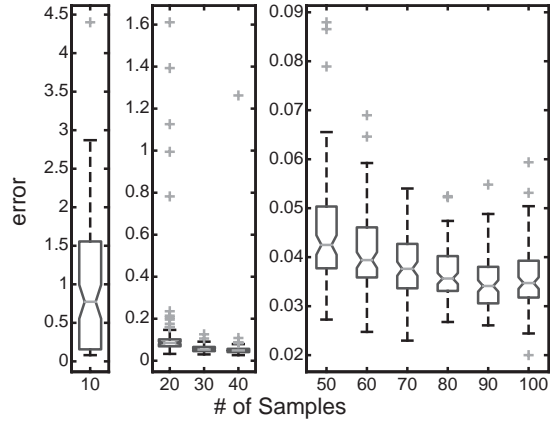


図 2.8 推定に用いたサンプル数と試行内平均絶対推定誤差（100 試行）. サンプル数が大きくなれば試行内平均絶対推定誤差は小さくなり，その分布の幅も小さくなって頑健な推定が実現できている .

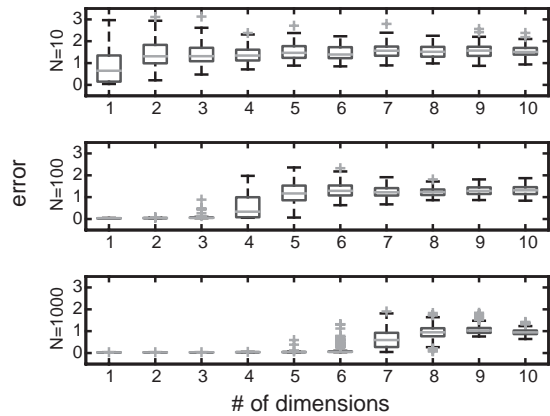


図 2.9 状態空間の次元数とサンプル数，推定誤差の関係 . 状態空間の次元数と試行内平均絶対推定誤差（100 試行）の関係をサンプル数 $N = 10, 100, 1000$ それぞれに対して箱ヒゲ図で表示する .

3.3 まとめ

以上，ここでは時系列ベイズ推定を，分布をサンプルで近似することにより計算可能とするパーティクルフィルタを簡単に説明した．ここで行った数値実験では状態遷移過程，観測過程共に線形を仮定していたが，カルマンフィルタと異なりパーティクルフィルタは分布の形状に仮定を置かないため，非ガウスもしくは非線形な過程も扱うことが可能である．このためパーティクルフィルタは，カルマンフィルタが適用困難であった非ガウス雑音を扱う必要のある問題に多く適用され，コンピュータビジョンの分野においても例えば視覚追跡などに多く適用され，その有効性が示されている [35, 85, 70, 16]．ただし，パーティクルフィルタはサンプルにより分布を近似する手法であるため，推定したい変数（隠れ状態）の推定分散を小さくするには，多くのサンプルを準備する必要がある．サンプル数が多い場合に最適なサンプリング分布は理論的に求めることが可能である [2]．ただし，現実的にはサンプル数は有限であり，多くの応用問題において十分と言えるサンプル数を用意することは容易ではない．特に，コンピュータビジョンの分野では，状態空間が高次元となる対象を扱うことが普通であり，推定精度を高めるためには一般に大量のサンプルが必要となる．しかし，大雑把に言って計算時間はサンプル数に比例するため，実問題への適用において，実時間処理のためには，サンプル数を抑えることが必要である．

サンプル数を抑えたまま精度の良い推定を実現することは，パーティクルフィルタにおける重要な課題として，サンプリング分布に観測情報を導入することでその改善を図る手法 [86, 41]，マルコフ連鎖モンテカルロ法やカーネル密度推定法と組み合わせた手法 [16, 27] 等，様々な手法が提案されている．しかし一方で，現実世界の観測情報は，モデル化困難な大きな雑音で汚れる場合があることが一般的であり，そうした場合に観測情報を導入したサンプリング分布は，かえって推定の頑健性を損なう恐れがある．第3章ではこうした実問題特有の問題に対処するため，観測情報を導入したサンプリング分布を利用する手法として APF [41] を，観測情報を導入せず，予測にのみ基づいたサンプリング分布を利用する手法として Condensation [35] を採用し，これらの手法を自身の推定に対する確信度を規準として動的に切り替える手法の提案を行う．提案手法では自身の推定に確信

を持っている間には予測にのみ基づいたサンプリング法を用い，自身の推定に確信を持たない場合には観測情報を導入したサンプリング法を用いることで，実環境における大きな雑音の影響を低減する．また，状態遷移モデルの学習 [62] をオンラインで実現し，提案手法に導入することで，推定された隠れ状態に基づいて予測にのみ基づいたサンプリング分布の精度を改善することで更なる推定精度の上昇を図る．

第3章 サンプルング分布の動的な切り替えによる頑健な視覚追跡

前章では、時系列ベイズ推定と、その近似解法であるカルマンフィルタ、パーティクルフィルタについて簡単に説明した。システムに線形過程を仮定し、分布の形状にガウス分布を仮定するカルマンフィルタに比べ、パーティクルフィルタはシステムの線形性、分布のガウス性共に仮定する必要が無いため、シグナルノイズのみならず、遮蔽や障害物によっても加えられる非ガウス雑音で汚された画像を扱う必要のある実時間視覚追跡を含む様々な分野から注目を集めている。この手法は有限のサンプルを用いたサンプルング法により時系列ベイズ推定の近似を行う手法であるため、様々なサンプルング法を用いた手法が提案されているが、実環境に存在する様々な非ガウス雑音によるモデル化誤差が状態遷移モデルや観測モデルに頻繁に生じる実時間視覚追跡に特有の状況の下では、評価されていない。

ここでは Condensation [35] と Auxiliary Particle Filter [41] (APF) の二つの手法に注目する。Condensation はパーティクルフィルタを実時間視覚追跡に適用した最も基本的なアルゴリズムであり、Sampling Importance Resampling (SIR) 法 [28] を用い、標準的な PC での実時間処理を可能としていた。Condensation には予測分布と観測情報との間のズレが大きい時に事後分布の推定精度が著しく低下するという問題点があるが、APF はこの点を現時刻の観測情報を用いて改善した手法である。この二つの手法は、予測モデルや観測モデルに想定できなかった雑音によるモデル化誤差の存在する実時間視覚追跡特有の状況下では、それぞれのモデル化誤差に対する頑健さの点から相補的であると考えることができ、ここでは、この二つの手法の長所を活用するために単純な規準に従ってこれらの手法を動的に切り替える (Switching Particle Filters, SPFs) ことを提案する。ここでは、各時刻で用いるサンプルング法を完全に切り替えてしまう手法 (Hard Switching Particle

Filters, h-SPFs) に対して, 別の切り替え手法としてより柔らかい切り替え手法 (Soft Switching Particle Filters, s-SPFs) を提案し, 数値実験を通してその性能を比較する.

本章は以下の様に構成される. まず 1 節では, パーティクルフィルタを簡潔に説明し, 代表的な手法として Condensation と APF を紹介する. 2 節では SPFs の提案を行い, h-SPFs と s-SPFs それぞれの枠組みについて説明し, 3 節で SPFs の推定精度と頑健性を数値実験を通して比較し, その性質について議論する. 4 では実環境で撮影された画像を用いた応用課題において, SPFs が頑健な追跡を維持するだけでなく, 追跡対象のダイナミクスモデルのオンライン学習に対しても有効であることを示す. 最後に 5 でこの章をまとめ, 今後の課題について述べる.

1. いくつかのパーティクルフィルタ

パーティクルフィルタはサンプルとその重みにより, 確率分布を近似する手法である. 時刻 t における追跡対象の状態変数ベクトルを $\mathbf{x}_t \in \mathcal{R}^{N_x}$, 観測変数ベクトルを $\mathbf{z}_t \in \mathcal{R}^{N_z}$ とすると, 分布をサンプルと重み $\{\mathbf{x}_t^{(n)}, w_t^{(n)}, n = 1, \dots, N\}$ を用いて近似するとベイズの定理より, 推定された事後分布は

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}) \approx \sum_{n=1}^N w_t^{(n)} \delta_{\mathbf{x}_t^{(n)}}(\mathbf{x}_t) \quad (3.1)$$

予測ステップは

$$p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) \approx \sum_n w_{t-1}^{(n)} p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(n)}) \quad (3.2)$$

と書くことができる. ここで, $\delta_{\mathbf{x}_t^{(n)}}(\mathbf{x}_t)$ はディラクのデルタ関数とし, サンプルの重み $w_{t-1}^{(n)}$ は $w_{t-1}^{(n)} \propto p(\mathbf{z}_{t-1} | \mathbf{x}_{t-1}^{(n)})$, $\sum_{n=1}^N w_{t-1}^{(n)} = 1$ の様に決定される.

もし, 十分多くのサンプルを準備する事が出来れば式 (3.1), 式 (3.2) の計算を正確に行うことができるが, 一般に実時間で処理を行うためには, それは困難である.

Condensation ではサンプリング分布として各時刻での観測情報 \mathbf{z}_t と独立な $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ が用いられるため, 状態空間は \mathbf{z}_t とは無関係に探索される. このため, 追跡対

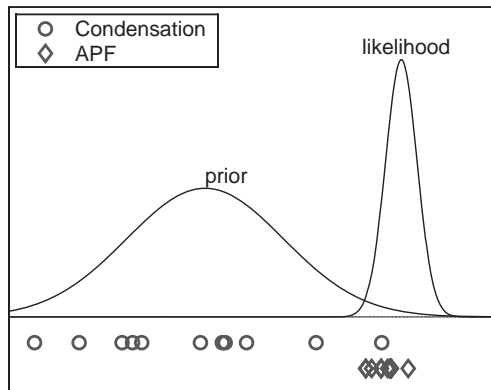


図 3.1 Condensation と APF によりサンプリングされたサンプル．Condensation では予測分布をサンプリング分布として利用し，APF では観測情報を利用したサンプリングを行う．

象のダイナミクスの変化など，予測モデルに存在するモデル化誤差により予測分布と観測情報に大きなズレが生じた時には外れ値の問題 [41] が生じる．この時，図 3.1 に示すように，追跡対象の動きを予測する事前分布からサンプルされたほとんどのサンプルが低い尤度を持つことになり，事後分布の推定に寄与するサンプル数の減少により近似精度は著しく低下する．

Pitt と Shepherd によって提案された Auxiliary Particle Filter [41] (APF) は Condensation の外れ値の問題を改善した手法である．この手法では，時刻 t におけるサンプル $x_t^{(n)}$ を得るためのサンプリング分布として $p(x_t|x_{t-1}^{(n)})$ を採用する．ここで $x_{t-1}^{(n)}$ は $p(z_t|\mu_t^{(n)})w_{t-1}^{(n)}$ に応じてサンプルされた時刻 $t-1$ のサンプル $x_{t-1}^{(n)}$ を示し， $\mu_t^{(n)}$ は平均値や最頻値などのダイナミクス $p(x_t|x_{t-1}^{(n)})$ を近似するための代表点である．この手法は代表点を用いてダイナミクスを近似し，代表点での尤度を計算することで，より「ありそうな」代表点の周りに多くのサンプルを分布させる．この様に現在の観測情報に注目することで，図 3.1 に示すように，大きな重みを持つサンプルを多数生成し，外れ値の問題が生じる状況においても事後分布の近似精度を保つ．但し， z_t を用いて状態空間の探索範囲を限定するため，保持するサンプルの多様性は Condensation と比べて低くなる．

2. Switching Particle Filters

2.1 動的なサンプリング分布の切り替えと柔らかな切り替え手法

APF は現在の観測情報 z_t も利用して事前分布からのサンプリングを行うが、Condensation は z_t とは無関係にサンプリングを行う。その結果、APF と比較して Condensation は事前知識を重視して状態推定を行うのに対し、APF は観測情報を重視して状態推定を行う。観測モデルにモデル化誤差が少ない場合には観測情報を重視することで精度良く推定を行うことができるが、遮蔽や追跡対象によく似た特徴が複数存在するような複雑な実環境における視覚追跡においては、状況によってはモデル化誤差の影響を強く受け精度の著しく低下した観測よりも、予測を重視した方が頑健に追跡を行うことができると考えられる。二つの手法を並列に用いて推定を行い、各サンプルの重みの平均が大きな手法の推定値を採用するような手法も考えられる。しかし、APF の各サンプルの方が高い尤度を持つ領域に集まる傾向があるため、この手法ではほとんどの状況で APF が採用され、観測情報を重視した追跡になりがちで二つの手法の長所を統合することができない。そこで、この二つの相補的な手法の長所を活かすため、状況に応じてサンプリング法の動的な切り替えを行う (Switching Particle Filters, SPFs) ことを考える。動的にサンプリング法を切り替えることで、追跡対象の状態をうまく推定できている時には事前知識を重視したサンプリング法を用いて、不確かな観測情報を過信するのではなく自分の信念に基づいた追跡を行い、推定がうまく行かなくなると、観測情報を重視したサンプリング法を用いて周囲の確からしい領域への素早い収束を図る。以下では、どの様にサンプリング法の切り替えを行うかについて、二つの手法を提案する。

一つ目の切り替え手法は推定の精度を示す推定状態の分散 V_{est} を用いて、完全に各時刻で使用するサンプリング法を切り替える手法 (Hard Switching Particle Filters) である。各時刻の状態の期待値を

$$\hat{x}_t = \sum_n \pi_t^{(n)} x_t^{(n)} \quad (3.3)$$

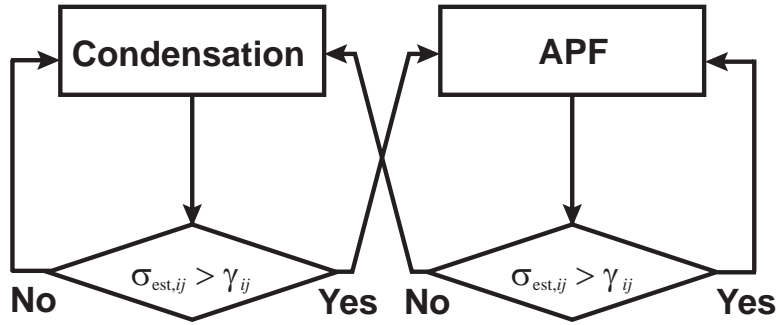


図 3.2 h-SPFs のアルゴリズム . h-SPFs では $\sigma_{est,ij}$ の値に応じて Condensation と APF のどちらのサンプリング法を用いるかを動的に決定する .

とする時 , 推定状態の分散共分散行列 V_{est} は

$$V_{est}(t) = \begin{pmatrix} \sigma_{est,11}^2 & \cdots & \sigma_{est,1N_x}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{est,N_x1}^2 & \cdots & \sigma_{est,N_xN_x}^2 \end{pmatrix} \quad (3.4)$$

$$\sigma_{est,ij}^2 = \frac{\sum_n \pi_t^{(n)} (\hat{x}_{t,i} - x_{t,i}^{(n)}) (\hat{x}_{t,j} - x_{t,j}^{(n)})}{N - 1}, \quad (3.5)$$

で定義される . これと閾値行列 γ を用いれば h-SPFs の枠組みは以下で書くことができる (cf. 図 3.2) .

if $(\exists i, j) \sigma_{est,ij} > \gamma_{ij}$ then use APF,
otherwise, use Condensation.

閾値行列 γ は , 自身の推定に自信を持っているか持っていないのかを分けるためのパラメータであり , ここでは以下の様にして定める . 追跡対象をボールとした視覚追跡において , ここでは簡単のため画像上でのボールの大きさは変化しないとし , 更に $x_i, (i = 1, 2)$ は独立であるとする . この仮定の下で分散共分散行列 $V_{est}(t)$ は対角行列となるため , 切り替えの閾値は $\gamma_{ii}, (i = 1, 2)$. についてのみ考えればよい . 予め行った実験において , 追跡対象の状態推定が成功している時の推定状態の標準偏差が , 追跡対象のボールの半径とほぼ一致していたため , ここ

では切り替えの閾値としてボールの半径を用いた．これにより，推定値の標準偏差がボールの半径よりも小さい時は，正しい推定を行うことができているとして Condensation を用いた予測を重視したサンプリングを行い，推定値の標準偏差がボールの半径よりも大きくなったときには，推定が不正確になったとして，APF を用いて観測情報を重視したサンプリングを行うことのできる枠組みが実現可能である．以下の実験では全てこの閾値を採用している．

h-SPFs は各時刻で完全に用いるサンプリング法を切り替える手法であるが，よりソフトな切り替え手法として，サンプリング法の異なるサンプルを混在させる手法 (Soft Switching Particle Filters, s-SPFs) が考えられる．この手法では $\sigma_{est,ij}$ の値に応じて，Condensation, APF でサンプリングするサンプルの割合を変化させる (cf. 図 3.3) ．

$$N_{cond} = N - N_{apf}, \quad (3.6)$$

$$N_{apf} = \max_{ij} N_{apf,ij}, \begin{cases} N_{apf,ij} = 0 & (\sigma_{est,ij} < \gamma_{min,ij}) \\ N_{apf,ij} = \frac{N(\sigma_{est,ij} - \gamma_{min,ij})}{\gamma_{max,ij} - \gamma_{min,ij}} & (\gamma_{min,ij} \leq \sigma_{est,ij} \leq \gamma_{max,ij}) \\ N_{apf,ij} = N & (\sigma_{est,ij} > \gamma_{max,ij}) \end{cases} \quad (3.7)$$

ここで， N_{cond} , N_{apf} は各サンプリング法でサンプルされるサンプル数である．つまり，自身の推定に対する確信度が高いほど Condensation でサンプリングされるサンプルが多くなり，確信度が低くなれば APF でサンプリングされる数が多くなる．s-SPFs における閾値も h-SPFs と同様にボールの半径を用いて決定する．s-SPFs の場合は γ_{min} と γ_{max} をボールの半径を挟むように適当な勾配と共に設定した．

これらの手法を用いて，Condensation と APF を動的に切り替えることで，各アルゴリズムの長所を融合できる．この2つの手法，h-SPFs と s-SPFs に共通する考え方は，自身の推定に確信があるならば観測情報を重視しないで，逆に，推定結果に自信が無ければ観測情報を重視した推定を行うということである特に実環境における視覚追跡のような，複雑な雑音が多々存在するために観測情報にあまり信用がおけない状況が考えられる場合には，観測情報を重視することは追跡に失敗する主要な要因になりうる．

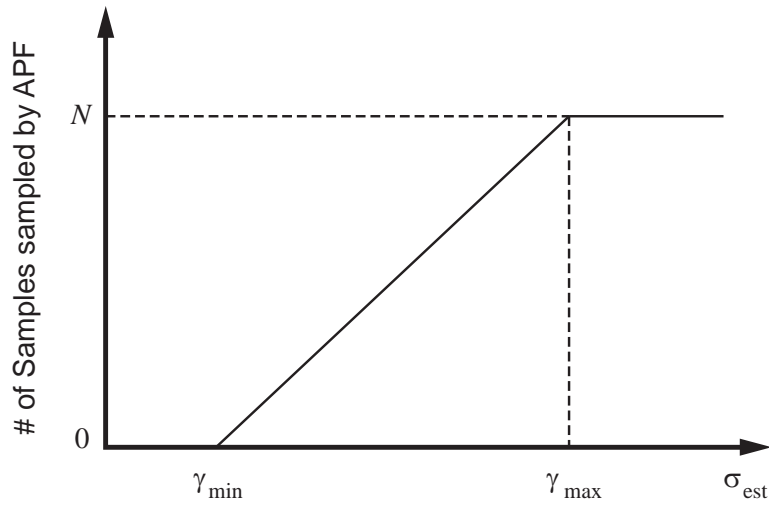


図 3.3 s-SPFs における APF サンプル数． $\sigma_{est,ij}$ が大きくなるに従い，APF サンプルの数は線形に増加する．

2.2 観測モデル

パーティクルフィルタによる視覚追跡を行うために，追跡対象の観測モデル（または尤度関数）を決める必要がある．一般に，より正確で複雑なモデルを用いる程，尤度関数が非常に尖った形状を持つため，より頑健では無くなってしまふ．更に，各サンプルについて複雑なモデルを仮定することは，計算量の増大を招き，実時間での処理を困難にしてしまう場合がある．そこで，ここでは色情報に基づく簡単なモデルを採用する．

色情報を観測モデルに用いた研究は以前からいくつか行われており (e.g. [18, 64, 65])，これらは Bhattacharyya similarity coefficient を用いてヒストグラム間の距離を定義し，尤度を求めているが，ここでは，推定の頑健さと計算量の観点から最も単純な観測モデルを採用する．状態ベクトル x が二次元画像平面上での位置として定義されているとし，画像 z を観測画像として，以下で定義される x におけるボールの赤色度を用いて観測モデルの定義を行う．

$$r(z, x) \equiv \frac{R(x)}{R(x) + G(x) + B(x)} \quad (3.8)$$

ここで， $R(x_i), G(x_i), B(x_i)$ はそれぞれ x_i における RGB の強度を示す．以上の赤

色度の定義の下で，観測の確率過程はガウスクーネルを用いた色の拡散に外れ値のための背景赤色度項 ρ を加えたものとして以下の様に定義する．

$$\begin{aligned} p(z_t | \mathbf{x}_t^{(n)}) &\propto \rho + k(\mu_r, r(z_t, \mathbf{x}_t^{(n)})) \\ &= \rho + \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(r(z_t, \mathbf{x}_t^{(n)}) - \mu_r)^2}{2\sigma_r^2}\right) \end{aligned} \quad (3.9)$$

ここで， μ_r, σ_r はガウシアンカーネルの平均と分散であり，これは事前にボールの周りのヒストグラムを正規分布で近似することで求め，先に示した ρ の値もこのガウシアン最大の 10% とする．

2.3 状態遷移モデルとそのオンライン学習

予測ステップ $p(x_t | z_{t-1})$ が $\sum_n \pi_{t-1}^{(n)} p(x_t | \mathbf{x}_{t-1}^{(n)})$ の様に近似できるが，適切な追跡のためには追跡対象のダイナミクスについての正しい知識が必要である．North らはバッチ EM アルゴリズムにより追跡対象の正しいダイナミクスの学習を行った [62] が，バッチ EM アルゴリズムは固定された画像列に対して繰り返し適用されるため学習が遅く，実時間視覚追跡においては実用的ではない．そこで，ここではダイナミクス学習にオンライン EM アルゴリズム [72] を用いる．オンライン学習は古い情報を忘却する枠組みを持っているため，動的に変化する環境においても有効に働くと考えられる．

以降の実験では，追跡対象のダイナミクスとして線形二次システム

$$\mathbf{x}_{t+2} = A_1 \mathbf{x}_{t+1} + A_0 \mathbf{x}_t + \mathbf{d} + B \mathbf{u}_t \quad (3.10)$$

を仮定する．ここで， \mathbf{u}_t は平均 0，分散 1 の白色ガウス雑音であり，パラメータ A_1, A_0 と \mathbf{d} をオンライン学習により決定する．実時間処理の制約からモンテカルロスムーザ [44, 37] を用いる事が出来なかったため，ここでは固定ラグカルマンスムーザを用いて，学習に必要なスムージング推定値を求める．詳細は付録 B に示す．

3. 数値実験

SPFs の有効性を確認するために実環境をモデル化したシミュレーション画像を用いたボールの追跡を行う。比較する手法は、h-SPFs と s-SPFs, サンプルング法を切り替える代わりにプロセスノイズの大きさを切り替える手法 (Process-noise controlled Particle Filter, PC-PF) の三つである。PC-PF は Condensation をベースにした手法であり、ダイナミクスモデルにおけるプロセスノイズの標準偏差 (式 (3.10) に示す B) を推定に対する確信度に応じて動的に制御する手法である。つまり、h-SPFs と同様に PC-PF の枠組みは以下で書くことが出来る。

$$\begin{aligned} & \text{if } (\exists i, j) \sigma_{\text{est},ij} > \gamma_{ij}, \text{ then use } B_{\text{large}} \\ & \text{otherwise, use } B, \end{aligned}$$

本論文で扱う視覚追跡課題の様にサンプルが二次元平面に分布している場合には、このサンプルの分布の広がりや視野の広さと考えることが出来る。PC-PF において大きなプロセスノイズを用いることは、追跡対象を見失いそうになった場合に視野を広げてより広い範囲から追跡対象を探そうとしていることに相当する。

これらの手法が遮蔽や類似物体の存在するような複雑な状況において追跡頑健性と精度の面でどのように振る舞うかを調べるために実環境をモデル化したシミュレーション画像列を用意し、遮蔽の幅、類似物体の大きさ・位置を変化させながら実験を行った (図 3.4)。ここでは追跡対象の大きさを半径 5pixels とし、対象は周期 1Hz, 振幅 200 ピクセルの正弦運動を行うとする。遮蔽の幅は 5 ~ 100pixel とし、類似物体の大きさは 1 ~ 20pixels とし、数は 3 個で位置は 6ヶ所の候補の中から選択する (類似物体の配置で ${}_6C_3 = 20$ 種類の画像列ができる)。画像列の長さは 60 フレームとする。以上の条件の下で、各手法を各条件 10 回ずつ適用し、手法の追跡頑健性と精度を評価する。つまり、各手法はそれぞれ (20 種類の類似物体の配置) \times (20 種類の類似物体の大きさ) \times (20 種類の遮蔽の幅) \times (10 試行) = 80,000 回適用される。各手法のパラメータは、 $\gamma = 5$ pixels (h-SPFs), $\gamma_{\min} = 4, \gamma_{\max} = 20$ pixels (s-SPFs), $\gamma = 5, B_{\text{large}} = 5$ pixels (PC-PF) として適用した。

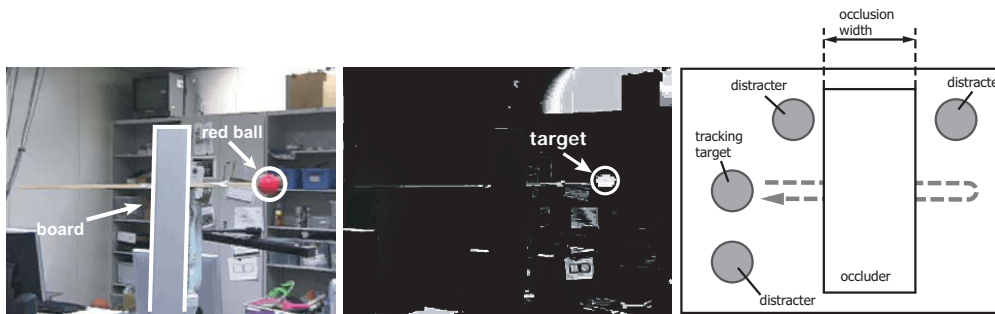


図 3.4 実環境（左図）をモデル化したシミュレーション画像（右図）。中央図は各ピクセルで計算した尤度を 0~255 に正規化したもの。白円で示した領域が追跡対象を示し、その他にも高輝度領域が画像全体に分布していることから非常に多くの類似物体が実環境には存在していることがわかる。シミュレーション画像ではこれらの類似物体を単純に円でモデル化した。

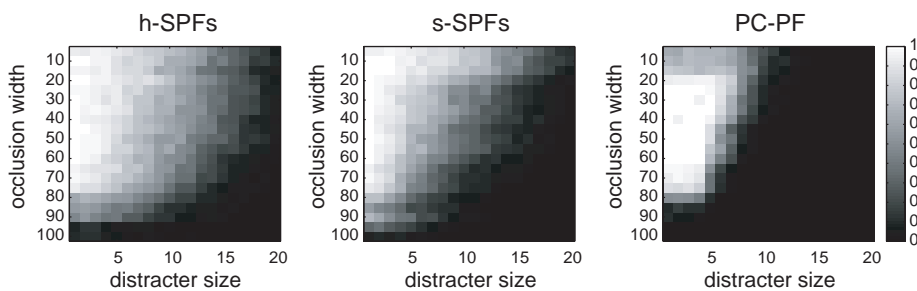


図 3.5 遮蔽の幅 (occlusion width)、類似物体の大きさ (distracter size) についての h-SPFs, s-SPFs, PC-PF の追跡成功率。

3.1 追跡頑健性

ここでは遮蔽の幅や類似物体の大きさについての各手法の追跡頑健性を計るため、シミュレーション画像の各条件における追跡成功率を用い、より広い条件において高い追跡成功率を維持できた手法が、遮蔽や類似物体の存在に対して頑健であるとする。追跡成功を“推定した位置が追跡対象の周りに設定した領域から外れたフレーム数が 15 フレーム以下であった時”と定義し、追跡成功率を求めたものを図 3.5 に示す。PC-PF は非常に成功率の高い領域がある反面、それ以外の領域での成功率は低く、この手法が遮蔽の幅や類似物体の大きさに大きく依存した

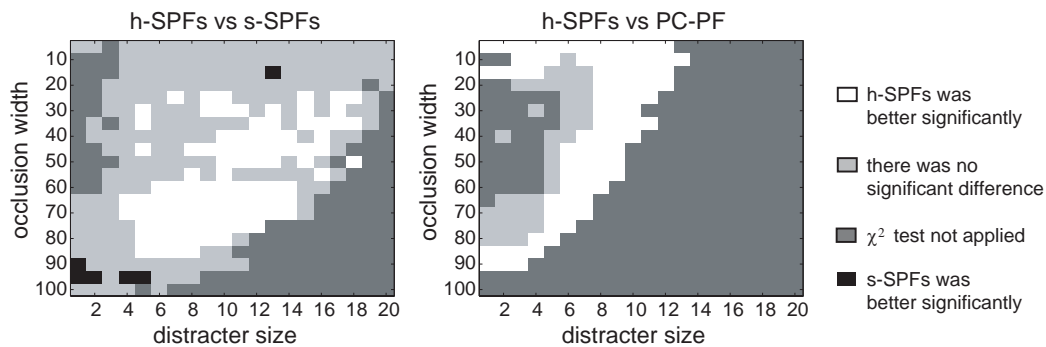


図 3.6 χ^2 検定 ($p < 0.05$) による追跡成功率の差の検定結果．左図は h-SPFs と s-SPFs の間の、右図は h-SPFs と PC-PF の間での検定結果．

手法であることがわかる．特に類似物体が大きい時に失敗の回数が極端に増える．大きなプロセスノイズを設定し、広い視野を持って追跡を行うことは、大雑把なダイナミクスモデルであるにも関わらずターゲットのダイナミクスの変化に対して非常に頑健な追跡を行うことができるが、周辺に大きなノイズ源があるような場合それに釣られる可能性も高くなってしまう．それに対して h-SPFs や s-SPFs では、広い領域 (特に遮蔽の幅について) で頑健な追跡が可能であるが、h-SPFs の方が類似物体の大きさに対して、より頑健な追跡を行うことができている．

また、これらの成功率に手法ごとの差があるかどうかを χ^2 検定を用いて検定を行う．各手法 (h-SPFs: $i = 1$, s-SPFs / PC-PF: $i = 2$) の追跡成功試行数を n_{i1} 、失敗試行数を n_{i2} 、期待度数を E_{ij} とした時、

$$\sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (3.11)$$

が自由度 1 の χ^2 分布に近似的に従う．有意水準 1% の χ^2 検定を行った結果を図 3.6 に示す．図中縦軸が遮蔽の幅、横軸が類似物体の大きさを示し、白は h-SPFs の成功率が有意に高い条件、黒は s-SPFs の成功率が有意に高い条件、薄いグレーはどちらも有意とはいえなかった条件、濃いグレーは成功もしくは失敗試行が少なすぎて χ^2 検定を適用することができなかった条件である．ほとんど雑音の無いサンプルな状況 (遮蔽の幅が小さく、類似物体も小さいような) においてや、逆に厳しすぎる状況 (大きな遮蔽、類似物体) においては、成功率に差は無いものの、

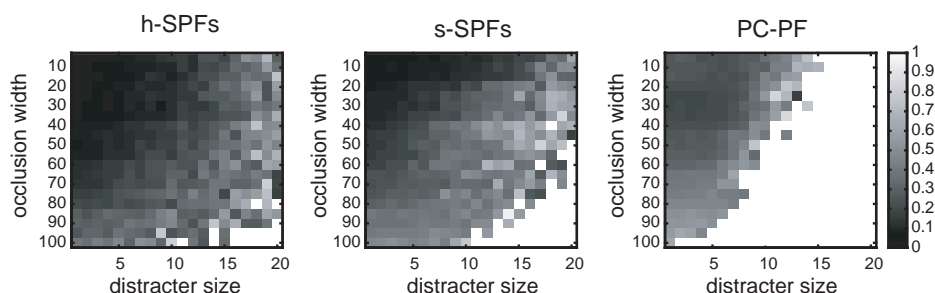


図 3.7 遮蔽の幅 (occlusion width), 類似物体の大きさ (distracter size) についての h-SPFs, s-SPFs, PC-PF の成功試行内平均絶対追跡誤差 .

それ以外の領域の広い範囲で h-SPFs の成功率が有意に高くなっており, h-SPFs がこれらの雑音に対して頑健であることを示している .

3.2 追跡精度

次に各手法の追跡精度を計るために, 上述の成功試行内での誤差の比較を行う . 前節で成功と判断した試行におけるの平均絶対誤差

$$\text{tracking error} = \frac{1}{L} \sum_{i=1}^L |\hat{x}_i - x_i^*| \quad (3.12)$$

を図 3.7 に示す . ここで, L は画像列の長さとし, x_i^* は追跡対象の真の状態とする . h-SPFs は s-SPFs と比べ広い範囲で小さな誤差に抑えており, 正確な追跡を行うことができている . この誤差の大きさに手法ごとに差があるか検定を行う . ここでは, 各条件での誤差の大きさに外れ値がある, 成功率の低い条件ではサンプル数が小さいなどの理由により Wilcoxon の順位和検定を用いた . 図 3.8 に検定結果を示す . 図中白は h-SPFs の誤差が有意に小さい条件, 黒は s-SPFs の誤差が有意に小さい条件, 薄いグレーはどちらも有意とはいえない条件, 濃いグレーはサンプル数が 0 (一度も追跡に成功していない) で検定を行うことが不可能だった条件を示している . h-SPFs は他の手法と比べ, 広い範囲で高い精度を維持できている .

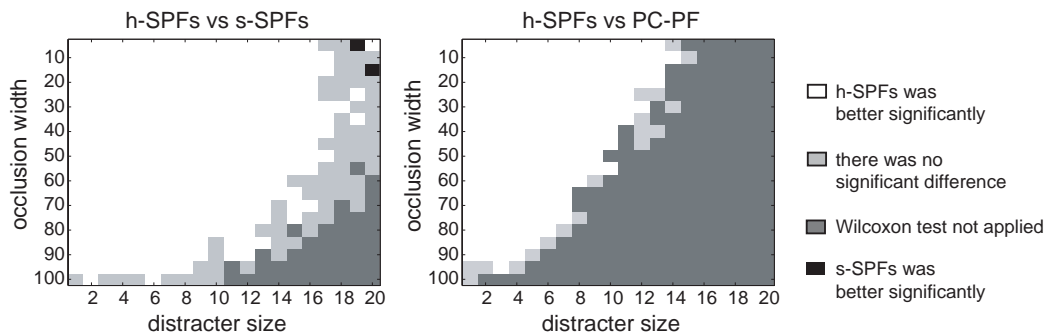


図 3.8 Wilcoxon の順位和検定 ($p < 0.01$) による平均絶対追跡誤差の差の検定結果．左図は h-SPFs と s-SPFs の間の，右図は h-SPFs と PC-PF の間での検定結果．

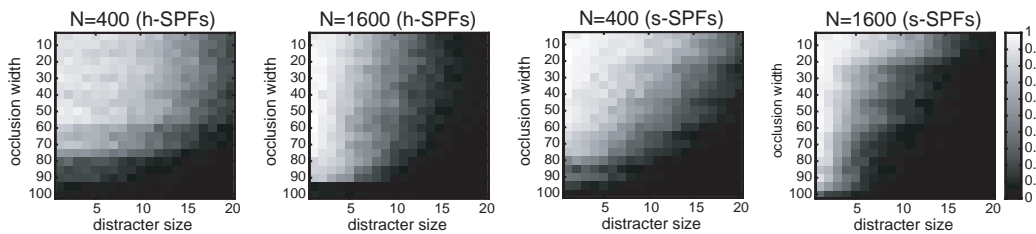


図 3.9 サンプル数の変化 ($N = 400, 1600$) に伴う，遮蔽の幅 (occlusion width)，類似物体の大きさ (distracter size) についての h-SPFs, s-SPFs の追跡成功率の変化

3.3 サンプル数

h-SPFs と s-SPFs について，サンプル数を増減させて性質の変化を調べる．ここでは前節までのサンプル数の 2 倍，1/2 倍の数における成功率，誤差を調べておく．図 3.9 に成功率，図 3.10 に成功試行内での平均誤差を示す．成功試行内の誤差にさほどの差は見えないものの，成功率の高い範囲がサンプル数の少ない場合の方が広がっている．その傾向は特に類似物体の大きさの方向に顕著に見える．これは，今回の様に画像面上にサンプルを分布させた場合，サンプルの広がりには視野を表現しており，サンプルの数が多ければ多いほど分布の裾の方まで表現でき，見落としが少ないために余計な類似物体まで発見してしまい，逆に追跡がうまく行かなかった場合があったためである．更に細かくサンプル数を変化させながら成功率と成功試行内での誤差を調べる．ここでは h-SPFs と s-SPFs の性質の差を見るために，あまりに簡単な/厳しい条件は取り扱わず，遮蔽の幅を 30

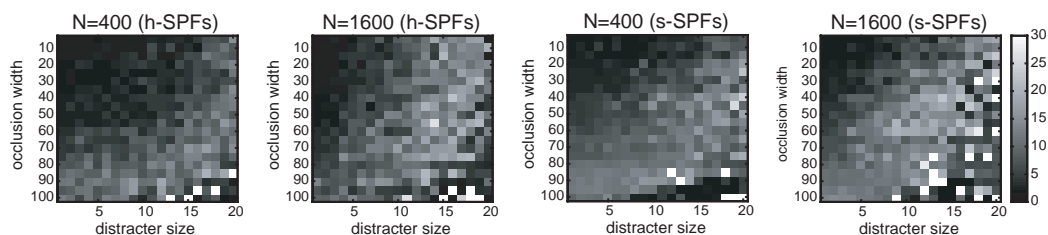


図 3.10 サンプル数の変化 ($N = 400, 1600$) に伴う，遮蔽の幅，類似物体の大きさについての h-SPFs, s-SPFs の成功試行内平均絶対追跡誤差の変化

~75 に，遮蔽の大きさを 5~15 ピクセルの 100 条件に限定した時の，成功率，誤差とサンプル数の関係を図 3.11 に示す．サンプル数が減少するにつれて，全体的に成功率，成功試行内の平均絶対誤差共に改善傾向にある．先の結果と合せてみると，実環境における視覚追跡の様に遮蔽や類似物体の存在する複雑な状況において，サンプル数を多くして分布の近似精度を上げることが必ずしも推定精度の向上に有効ではないことを示している．但し，当然少なければ少ないほど良いわけではなく，あまりに少なすぎると成功率が極端に低下してしまう (cf. 図 3.11) .

3.4 議論

以上の結果を踏まえて h-SPFs, s-SPFs の特徴について考察を行う．PC-PF で行ったように大きなプロセスノイズを用いてサンプルの遷移先の分布をぼかすことは，ダイナミクスがわかっていない追跡対象の安定した追跡には非常に有効であると考えられる．しかし，実環境における視覚追跡のように，複雑で大きな非ガウス雑音が多々入るような状況において，大きなプロセスノイズを用いて追跡を行うことは，追跡対象の周りに分布する類似物体に惑わされる危険性を大きく上げることになってしまう．この結果は図 3.7 でも見たように，仮定したプロセスノイズの大きさに最適な遮蔽の幅や，追跡対象の予期せぬ動きに対しては非常に有効に働くものの，周囲にある類似物体の大きさに非常に敏感な手法である．

それに対して，h-SPFs, s-SPFs はサンプリング法を切り替えることにより，プロセスノイズは小さいまま観測情報に対するサンプルの収束性を加減する手法である．この操作により，状態遷移モデル，観測モデルそれぞれにモデル化誤差の

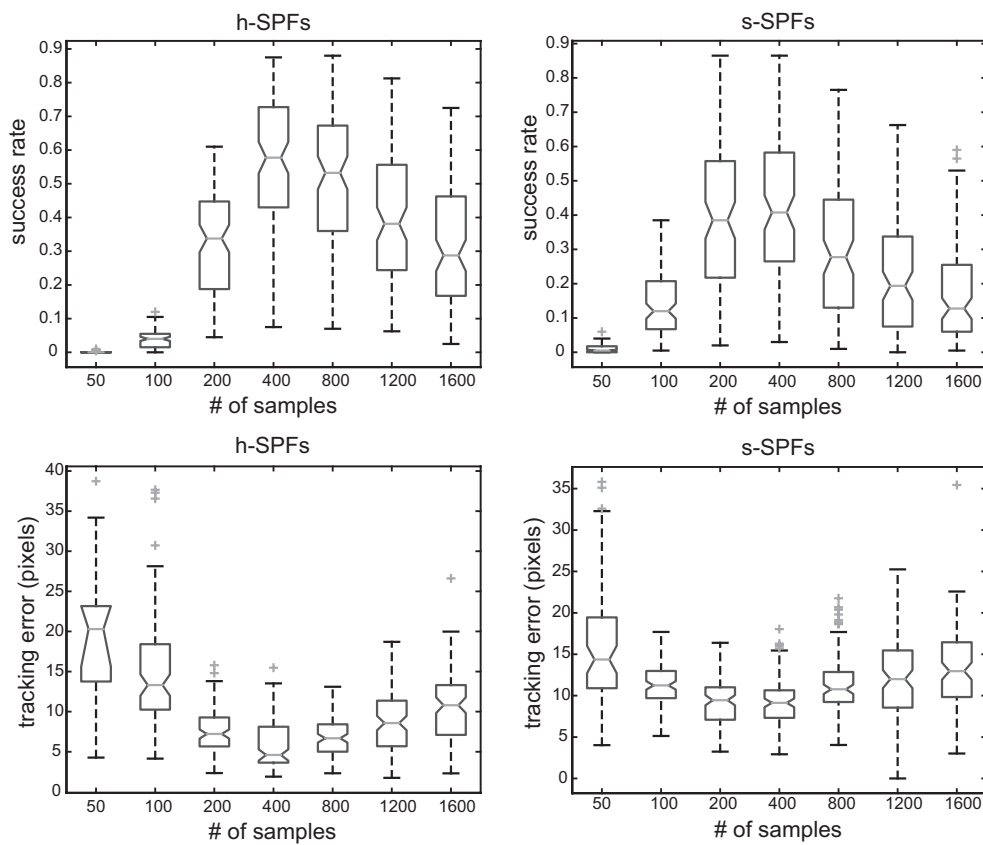


図 3.11 h-SPFs と s-SPFs におけるサンプル数の変化 ($N = 50, 100, 200, 400, 800, 1200, 1600$) に伴う追跡成功率と誤差の変化

存在する実環境において，サンプルの多様性と収束性をうまく両立させ，状態遷移モデルに存在するモデル化誤差に起因する外れ値問題のみならず，遮蔽，類似物体などに起因する観測モデルに存在するモデル化誤差にも頑健で正確な追跡を可能としている．

これらの手法を用いる際に必要なパラメータとして，切り替えの閾値，つまり“うまく追跡できている時の推定値の分散”がある．このパラメータは，状況に応じて事前知識を重視したサンプリング法を採用するか観測を重視するサンプリング法を採用するかを決めるパラメータであり，つまり，自身の推定に確信を持って良いのかどうかを決めるためのパラメータである．このパラメータの設定に失敗すると，サンプリング法の切り替えのタイミングを逃し，モデル化誤差の多く存在するモデルを重視することにつながり，追跡に失敗することも多々あった．SPFsはこのパラメータに，h-SPFsの場合は特に，大きく依存している．その点に関してはh-SPFsよりもs-SPFsの方が頑健である．先述の通り切り替えの閾値は，うまく追跡できている時と追跡できていない時を切り分けるパラメータであるため，ここでの問題設定の下ではボールの大きさや遮蔽の幅などに影響を受ける．ここでは追跡に成功している時の推定値の分散がボールの半径とほぼ一致していたため，簡単のため画像面上でのボールの大きさは変化しないという仮定の下，ボールの半径をh-SPFsの切り替えの閾値として採用した．s-SPFsの場合は，ボールの半径を挟むように，適当な勾配と共に γ_{\min} と γ_{\max} を決めてやれば比較的うまく追跡を行ってくれる．数値実験における各時刻での各サンプリング法でサンプリングされるサンプル数を見ると，傾きを急に設定した場合は特に，h-SPFsとほぼ同様であった．しかし，ロボットの視覚としてこの様な視覚追跡システムを用いる場合には，サンプル数を多く取ることはその他にリアルタイムで行う必要のある処理に大きな負担をかけるため不可能である．この様に，少ない計算資源をいかにうまく使うかということに重点を置いて考えた場合，h-SPFsの方がs-SPFsよりもより良い手法なのかもしれない．

4. 応用: 実環境内でのボール追跡

この節では、実環境で撮影された実画像を用いて SPFs と Condensation, APF, Unscented Particle Filter [86] (UPF) の性能を比較する。Merwe らによって提案された UPF は、APF 同様に現在の観測を用いてサンプリング分布の修正を行う手法である。UPF におけるサンプリング分布 $p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t})$ は Unscented Kalman Filter[39] (UKF) を用いて以下で推定される。

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t}) = \mathcal{N}_{N_x}(\bar{\mathbf{x}}_t^{(n)}, \bar{P}_t^{(n)}) \quad (3.13)$$

ここで、 $\bar{\mathbf{x}}_t^{(n)}, \bar{P}_t^{(n)}$ はそれぞれ UKF により推定されたサンプル $\mathbf{x}_{t-1}^{(n)}$ の遷移先の平均と分散であり、この分布からのサンプリングにより、時刻 t におけるサンプル $\mathbf{x}_t^{(n)}$ を得る。ここでの実験においては、このサンプリング分布の計算の際に必要な観測情報として、式 (3.8) で定義した赤色度の画像全体に渡る重みつき平均を用いた。以下では、前節と同様に各手法の頑健性と精度を推定に失敗したフレーム数と、推定成功時の推定誤差で評価する。

4.1 実験環境

ここでは人間の手やマニピュレータによる正弦運動を行う赤いボールの追跡を行う。実画像列はデジタルビデオカメラを用いて撮影され、Pentium 4 (2GHz) の Linux PC により処理される。画像列は Video for Linux を用いて PC に解像度 640×480 で取り込まれ、 320×240 にダウンサンプリングされる。また、サンプル数はシミュレーションの結果より実時間処理を実現でき、ある程度の精度を維持できる 800 個とした。本システムは 1 フレームあたり約 8msec で処理を行うことができた。

数値実験と同様に、ここでも SPFs の閾値 γ にはボールの半径を用いた。今回の実験では追跡対象であるボールは三次元空間内を動くため、画像面上でのボールの半径は、カメラとの距離に応じて数ピクセルの範囲で変化するが、先述の通り、ここでは不変であると仮定して適当な時刻でのボールの半径を用いて γ を設定する。

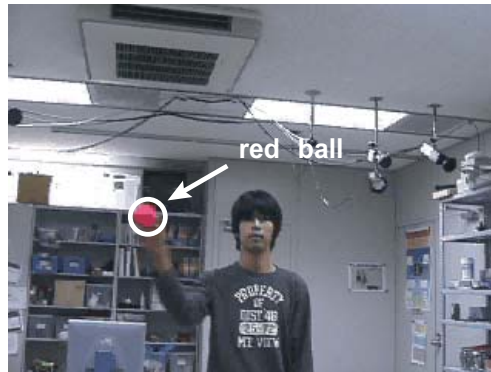


図 3.12 追跡対象のダイナミクス学習実験で用いた実画像例．

4.2 追跡対象のダイナミクス学習

タスク環境を示したサンプル画像を図 3.12 左図に示す．ここでは赤いボールは人間によって左右に動かされる．

図 3.13 左図に学習の進行と推定状態の標準偏差との関係を示す．Condensation 追跡器にオンライン学習を導入した追跡器では，学習が進むにつれて σ_{est} が小さくなり，追跡器は確信を持って推定を行うようになることがわかる．標準偏差は 200 フレーム辺りで収束しているが，これはボールの正弦運動の振動数の推定値が収束するフレームとほぼ一致している．これに対して h-SPFs では，学習の初期から小さい標準偏差を維持することが出来ている．

ボールの正弦運動 1 周期あたりに APF が選択されたフレーム数と切り替えの回数を図 3.13 右図に示す．ここで，APF のフレーム数が 2 サイクル目で上がっているのは学習初期の精度の悪いダイナミクスを予測分布を作るのに用いているので，推定した状態に確信が持てない為である．

学習の進行と共に APF の選択されるフレーム数が減少していることは，学習の初期は観測情報を重視して推定を行っていたが，学習が進むにつれ事前知識を重視した推定を行うようになってきていることを示している．このように自動的に観測情報重視から事前知識重視の推定へ遷移することで，次節で説明するような厳しい状況においては特に，学習を非常に安定して行うことができるようになっていく．

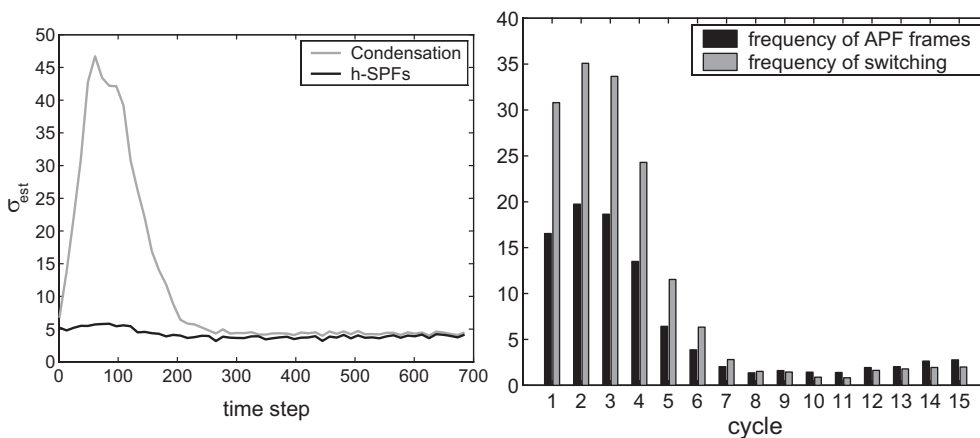


図 3.13 ダイナミクス学習の進行の様子．推定状態の標準偏差（左図）と正弦運動の周期毎の APF でサンプルされたサンプル数の平均とサンプリング分布の切り替えの回数（右図）．

4.3 遮蔽や類似物体の存在する環境での追跡

視覚システムを実環境で用いるに当たって，普通に起こり得る状況の下で，各パーティクルフィルタの性能を統計的に比較するために，図 3.14 左図に示すようにカメラとボールの間に板を置いて人工的に遮蔽のある環境を作った．また，ボール以外の赤い物体は全て障害物となる．

図 3.15 にボールを“見失った”フレーム数のヒストグラムを示す（ビデオの長さは 423 フレーム，約 14 秒である）．“見失った”とはボールの周りに半径の 3 倍の領域を設定し，その領域から推定値が外れることとする（cf. 図 3.14 右図）．ボールが遮蔽の後ろに隠れている時の真のボールの位置は，その前後の位置から線形補完で求めた．また，図 3.16 に追跡成功試行内での平均誤差を示しておく．追跡成功は数値実験の時と同様に，先ほど設定したボールの周りの領域から推定値が外れたフレーム数が 100 フレーム以内であったとする．

h-SPFs, s-SPFs そして UPF はダイナミクス学習あり/なしに関わらず良い性能を見せている．APF はダイナミクス学習と共に適用しなければ良い性能を見せるが，ダイナミクスの学習を行うと極端に見失ったフレーム数が上昇する．

この様にダイナミクス学習の初期から遮蔽により大きく追跡対象が隠されてし

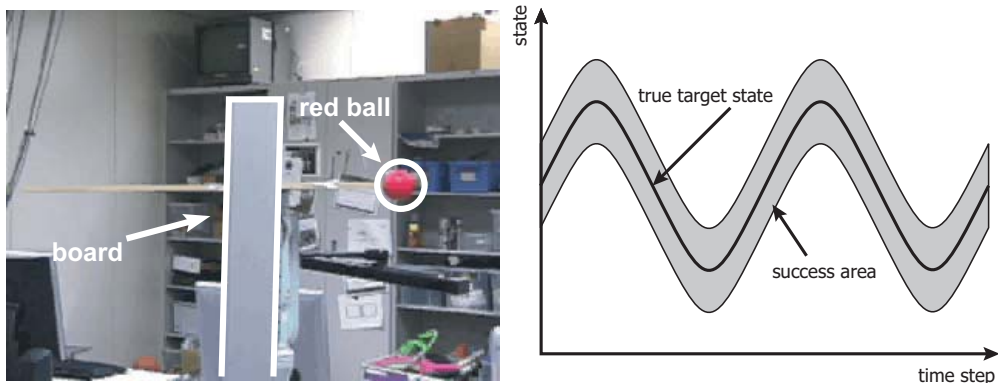


図 3.14 遮蔽や追跡対象の周囲に類似物体が多数存在する実験環境（左図）と，追跡手法の頑健性を定義するために追跡成功/失敗を定義する領域（右図）．真の状態の周囲に対象の大きさに依存したグレーの領域を定義し，この領域に状態の推定値があるフレームを追跡成功フレーム，無いフレームを追跡失敗フレームと定義する．

まう様な厳しい状況において，安定したダイナミクスの学習を行うことは非常に困難である．実際，学習タスクにおける Condensation と APF の追跡成功回数は共に 0 回であった（それ故図 3.16 において，Condensation と APF については誤差を計算することが出来なかった）．にも関わらず，SPFs はよく働いている．これは恐らく，SPFs が観測と予測の間のバランスをうまく制御することで，学習の進行を安定なものにしているためでないかと考えられる (cf. 4.2 節)．

以上の結果より，SPFs は UPF を含む他の手法と比べて良い性能を示していることが確認された．APF は観測情報を重視するため，ボールが遮蔽の後ろに隠れてしまった時など，障害物によって追跡に失敗する場合があった．Condensation は遮蔽を超えることに対しては頑健であったが，その後追跡対象に対しての収束が非常に遅く，その間に追跡に失敗してしまうことがあった．UPF はこれら二つの手法と比べると良い性能を見せたが，推定の正確さでは SPFs と比べて悪かった．ここでは z_t から状態 x_t を得ることができないため，UPF を使用する際に必要となる観測値には赤色度によるピクセルの重み付け平均を用いた．このような粗悪な観測情報に関わらず，特にダイナミクス学習を行う場合には UPF は非常に良く働いているが，やはり観測情報を用いてサンプルの遷移先を決める手法は遮蔽，類似物体に対して，観測情報が信用できない場合は特に，誤差が大きくなっ

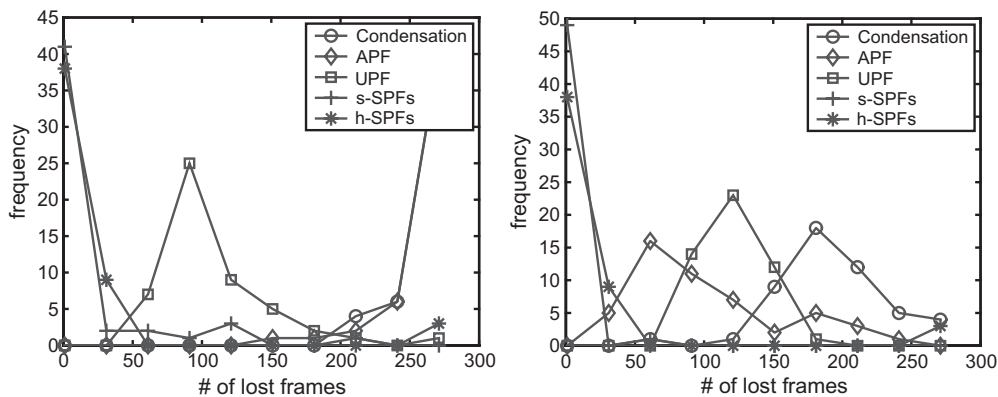


図 3.15 追跡対象のダイナミクス学習と追跡頑健性．ダイナミクス学習を行った場合の“見失った”フレーム数を左図に，ダイナミクス学習を行わなかった場合の“見失った”フレーム数を右図に示す．

てしまう．

5. まとめ

ここでは実時間視覚システムのための新しいパーティクルフィルタの枠組みとして，SPFs (h-SPFs, s-SPFs) を提案した．これらの手法は二つのパーティクルフィルタ (Condensation と APF) を組み合わせたものである．これらの切り替え手法では障害物や遮蔽の存在するような状況において正確さと頑健性の点で相補的な特徴を持つ Condensation と APF を，推定に対する確信度に応じて動的に切り替えることで，これらの長所を活用する．実画像を用いた視覚追跡実験により，切り替え手法が他のよく知られたパーティクルフィルタと比べて良い性能を持つことを示した．また，この切り替え手法は追跡の精度上昇のための対象のダイナミクス学習を安定に行えるようにすることを示した．ここでは切り替えの閾値を追跡対象のボール半径を用いて事前に決定したが，これをハイパーパラメータと見て学習するなど，閾値の自動的な決定についてはこれからの課題である．

本章では純粋なサンプリング分布の性質を調べるために，低次元の状態空間と単純な観測モデルを仮定したが，実際の環境における応用課題を考えれば，高次

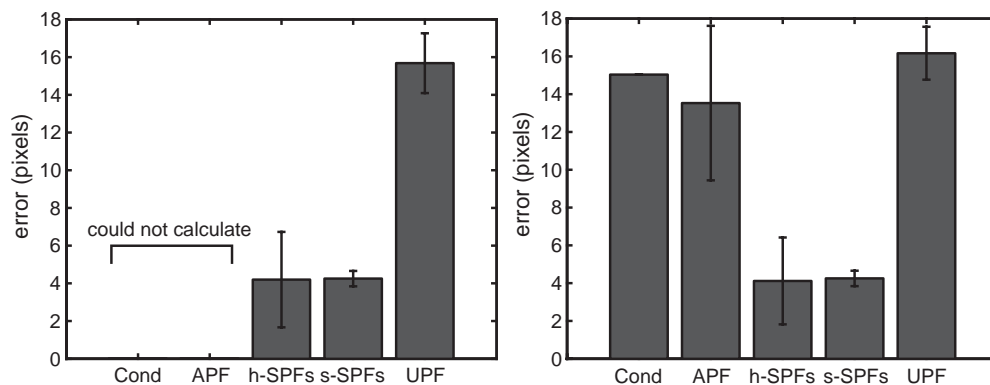


図 3.16 追跡対象のダイナミクス学習と推定精度．ダイナミクス学習を行った場合の平均絶対追跡誤差を左図に，ダイナミクス学習を行わなかった場合の平均絶対追跡誤差を右図に示す．

元な状態空間を扱う必要がある場合も多く考えられる．パーティクルフィルタは状態空間にサンプルを生成し，分布の近似を行う手法であるため，状態空間が高次元化した場合には推定精度の維持に必要なサンプル数は指数関数的に増加すると考えられる．この問題については次章で考えることとしたい．

第4章 適応的サンプリング分布とその実環境における高次元状態推定への応用

パーティクルフィルタはサンプル近似に基づく手法であるため、近年の計算機の進捗により、より多くのサンプルを用いることが出来るようになり、近似精度の上昇と共に多くの応用研究がなされてきた。前節ではサンプリング分布の性能を純粋に比較するため、赤いボールという単純な対象を追跡対象として扱い、可能な限り単純化した観測モデルを仮定していたが、例えば人間の頭部や手の様な、より複雑な対象を追跡対象とする実環境における応用課題を考えると、追跡対象のモデルが高次元となる場合が多い。先述の通り、パーティクルフィルタはサンプル近似を利用した手法であるため、近似する状態空間が高次元である場合には、精度の良い近似を行うためには大量のサンプルを利用する必要がある。しかし高次元な隠れ状態の推定問題においては、次元数の増加に対して精度を維持した推定に必要なサンプル数は指数的に増加するため、依然としてサンプリングの効率の改善はパーティクルフィルタにおける大きな課題である。

前節でも紹介した通り、サンプル数を抑えたまま精度の良い推定を実現することは、パーティクルフィルタにおける重要な課題として、サンプリング分布に観測情報を導入することでその改善を図る手法 [86, 41]、マルコフ連鎖モンテカルロ法 (MCMC) やカーネル密度推定法と組み合わせた手法 [16, 27] 等、様々な手法が提案されている。ここでは、階層構造を持つ高次元隠れ状態推定問題に対して、階層構造を上手く利用することでサンプリング分布を改善し、少数のサンプルであっても精度を維持することを考える。

コンピュータビジョンの分野において、例えば図 4.1 (ただし点線で示した矢

印は除く)に示すように,物体の姿勢が変化すれば物体の特徴点の画像面上における位置が変化するというように画像の生成モデルが階層構造を持つことはごく一般的であるが,実環境の中では状態遷移モデルや観測モデルに用いた仮定や近似の崩れ,想定できなかった雑音等により,推定精度が低下する場合がしばしば考えられる.とりわけ,実時間処理のためサンプル数が厳しく制限される状況では,上記の要因で起る精度の低下が激しく,階層間に強い依存関係を仮定することは,ある層での精度の低下を他の層へと伝播させていく結果を招く.また,各階層に独立な状態遷移を考え,他の層の精度低下の影響を極力排除するモデルを考えると,他の層の情報を用いて推定精度が低下した層の推定を修正することが出来ず,結局,推定の頑健性を低下させることになる.このように,様々な仮定や近似の崩れ,大きな雑音の存在する実環境において,頑健で精度の良い状態推定を実現するためには,階層間に適度な独立性と相互作用を実現し,その大きさを状況に合わせて制御する必要がある.そこで,ここでは,ある層のサンプリング分布を,その他の階層からの予測分布の混合分布として表現し,その混合比をオンラインEMアルゴリズム[72]により逐次推定する手法を提案する.精度の低下した階層からの影響は混合比を適応的に小さくすることで最小限に止め,逆に精度を維持している階層の混合比を大きく設定することで,頑健で正確な推定の実現が可能となる.ここでは,剛体の姿勢推定とその特徴点の画像面上での視覚追跡を例にとり,下位層で画像面上における特徴点の視覚追跡を,上位層で剛体の三次元モデルの画像面へのアフィン射影を仮定し,これらの過程の推定により剛体の三次元姿勢推定を行う.数値実験により提案手法の有効性を確認すると共に,提案手法を単眼画像からの自家用車内におけるドライバの頭部姿勢推定と顔特徴点追跡に応用し,実際の車内を想定した,極端に制限された計算資源環境にも関わらず頑健で精度の良い推定の実現が可能であることを示す.

本章は以下のように構成される.まず,1節で,階層構造をもつ隠れ状態の時系列ベイズ推定について簡単に説明し,2節で階層間の相互作用の大きさ(混合サンプリング分布における混合比)を適応的に決定する手法(Particle Filtering using Adaptive Sampling Distribution, PF-ASD)の提案を行う.次に,3節で提案手法の剛体の姿勢推定への適用について説明し,数値実験を通してその有効性の確認を

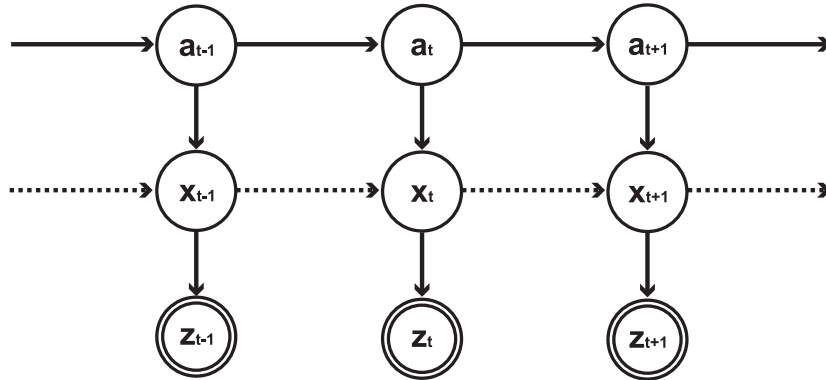


図 4.1 階層モデルのグラフィカルモデル．ここで， a_t は隠れ変数， x_t は観測変数であり， a_t が上位の隠れ状態変数， x_t が下位の隠れ状態変数である．点線の矢印は実際には存在しないが，ここでは，各層での推定精度に応じてこの矢印の寄与の大きさを制御する事で頑健な推定を図る．

行い，4節で実際の車内で撮影された画像を用いたドライバの頭部姿勢と顔特徴点位置の同時推定問題の実実験を行う．加えて5節では，ドライバの頭部姿勢推定の精度を更に上昇させるためのドライバの顔モデルのオンライン推定手法の紹介を行う．

1. 階層的隠れ状態推定

本節では，上位の隠れ状態（三次元姿勢） $a_t \in \mathcal{R}^{N_a}$ ，下位の隠れ状態（特徴点の画像面上での位置） $x_t \in \mathcal{R}^{N_x}$ が図 4.1 に示すようなグラフィカルモデルで表現される階層構造を持つとして，観測系列（観測画像） $z_{1:t}$ から，これらの隠れ状態変数の時系列推定を行うことを考える．観測系列 $z_{1:t}$ に対する a_t と x_t の同時事後分布は，ベイズの定理より，

$$p(a_t, x_t | z_{1:t}) \propto p(z_t | a_t, x_t) p(a_t, x_t | z_{1:t-1})$$

と書け，尤度関数を表す $p(z_t | a_t, x_t)$ と，同時事前分布 $p(a_t, x_t | z_{1:t-1})$ の積の形で表すことができる． $p(a_t, x_t | z_{1:t-1})$ は前時刻の同時事後分布 $p(a_{t-1}, x_{t-1} | z_{1:t-1})$ と状態

遷移モデル $p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{a}_{t-1}, \mathbf{x}_{t-1})$ を用いて

$$p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{z}_{1:t-1}) = \iint p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{a}_{t-1}, \mathbf{x}_{t-1}) p(\mathbf{a}_{t-1}, \mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}) d\mathbf{a}_{t-1} d\mathbf{x}_{t-1}$$

と書くことができる．ここで，図 4.1 の階層モデルにより $\mathbf{a}_t, \mathbf{x}_t$ が生成されているとし， \mathbf{a}_t の状態遷移にマルコフ性を， \mathbf{x}_t の \mathbf{a}_t の状態遷移に対する条件付き独立性を仮定すれば，

$$p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{a}_{t-1}, \mathbf{x}_{t-1}) = p(\mathbf{x}_t | \mathbf{a}_t) p(\mathbf{a}_t | \mathbf{a}_{t-1})$$

が成り立ち，同時事前分布 $p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{z}_{1:t-1})$ は，

$$p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{z}_{1:t-1}) = p(\mathbf{x}_t | \mathbf{a}_t) p(\mathbf{a}_t | \mathbf{z}_{1:t-1})$$

と書くことができる．これにより，パーティクルフィルタを用いた同時事後分布 $p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{z}_{1:t})$ の推定を， \mathbf{a}_t の事前分布 $p(\mathbf{a}_t | \mathbf{z}_{1:t-1})$ から \mathbf{a}_t のサンプルを得る， $p(\mathbf{x}_t | \mathbf{a}_t)$ を用いて \mathbf{x}_t の事前分布 $p(\mathbf{x}_t | \mathbf{z}_{1:t-1})$ から \mathbf{x}_t のサンプルを得る，尤度関数 $p(\mathbf{z}_t | \mathbf{a}_t, \mathbf{x}_t)$ を用いて各サンプルの重みを計算，という手順で階層的に行うことが可能となる．

2. 適応的混合サンプリング分布

前節で説明した階層構造を持つ隠れ状態推定法を，実問題へ適用する際，用いた近似や仮定の崩れ，想定できなかった雑音の影響等により，上位の推定精度が低下した場合には，その精度低下が下位に伝播し，さらに上位の推定に悪影響を与えるということが起り得る．特にパーティクルフィルタを用いた推定を行うに当たって，隠れ状態変数の次元数に対してサンプル数が少数に制限される場合には，サンプルを得るためのサンプリング分布の精度低下が最終的な推定精度に与える影響は大きい．本研究では，これを回避するために各層に独立な状態遷移（図 4.1 では点線で描かれた矢印で表現）を仮定し，さらに各層の推定への寄与を適応的に決定することで，各層間の相互作用のバランスを取り，実環境での推定精度の維持を図る． \mathbf{x}_t の状態遷移についてもマルコフ性を仮定すれば，状態遷移モデル $p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{a}_{t-1}, \mathbf{x}_{t-1})$ を

$$p(\mathbf{a}_t, \mathbf{x}_t | \mathbf{a}_{t-1}, \mathbf{x}_{t-1}) = p(\mathbf{x}_t | \mathbf{a}_t, \mathbf{x}_{t-1}) p(\mathbf{a}_t | \mathbf{a}_{t-1}) \quad (4.1)$$

と書くことができる．

ここで，下位層の状態遷移モデルとして

$$p(\mathbf{x}_t|\mathbf{a}_t, \mathbf{x}_{t-1}) = \alpha_{a,t}p(\mathbf{x}_t|\mathbf{a}_t) + \alpha_{x,t}p(\mathbf{x}_t|\mathbf{x}_{t-1}) \quad (4.2)$$

を仮定する．ただし， $\alpha_{a,t} + \alpha_{x,t} = 1$ とする．上式のように上位からの予測 $p(\mathbf{x}_t|\mathbf{a}_t)$ と，上位に独立な下位からの予測 $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ の混合分布として下位層の状態遷移 $p(\mathbf{x}_t|\mathbf{a}_t, \mathbf{x}_{t-1})$ をモデル化し，混合比 $\alpha_t = \{\alpha_{a,t}, \alpha_{x,t}\}$ を，各層の推定精度を考慮して適応的に決定することで，推定精度の低下した層からの影響を低く抑えつつ，推定精度の高い層の情報を重視した，頑健な推定の実現が期待される．具体的な混合比 α_t の決定法については後述する．

式(4.1)，(4.2)より，同時事前分布 $p(\mathbf{a}_t, \mathbf{x}_t|z_{1:t-1})$ は

$$\begin{aligned} p(\mathbf{a}_t, \mathbf{x}_t|z_{1:t-1}) &= p(\mathbf{x}_t|\mathbf{a}_t, z_{1:t-1})p(\mathbf{a}_t|z_{1:t-1}) \\ &= \pi(\mathbf{x}_t|\alpha_t)p(\mathbf{a}_t|z_{1:t-1}) \end{aligned}$$

と書くことができる．ここで， \mathbf{x}_t のサンプリング分布

$$\pi(\mathbf{x}_t|\alpha_t) = \alpha_{a,t}p(\mathbf{x}_t|\mathbf{a}_t) + \alpha_{x,t}p(\mathbf{x}_t|z_{1:t-1}) \quad (4.3)$$

は，上位からの予測分布 $p(\mathbf{x}_t|\mathbf{a}_t)$ と下位からの予測分布 $p(\mathbf{x}_t|z_{1:t-1})$ を用いて適応的に決定される混合サンプリング分布 (Adaptive Sampling Distribution, ASD) である．

混合比 α_t は ASD $\pi(\mathbf{x}_t|\alpha_t)$ の形状を決定するためのパラメータであり，これを \mathbf{x}_t の事後分布 $p(\mathbf{x}_t|\mathbf{a}_t, z_{1:t})$ に近づけるように α_t を設定することで，隠れ状態変数の次元数に比べてサンプル数が少数に制限される場合にも精度を保った推定を実現することができる．ここで， α_t の決定は，コンポーネントが二つである混合モデルにおける混合比の逐次最尤推定であり，コンポーネント選択のための指標変数が隠れ変数となるため，オンライン EM アルゴリズム [72] を用いてこれを解く．時刻 t において推定した画像面上での \mathbf{x}_t の事後分布 $p(\mathbf{x}_t|\mathbf{a}_t, z_{1:t})$ からリサンプルすることで $p(\mathbf{x}_t|\mathbf{a}_t, z_{1:t})$ に従う N 個のサンプル $\tilde{\mathbf{x}}_t^{(i)}$ を得る． $m = \{m_a, m_x\}$ をそれぞれサンプル $\tilde{\mathbf{x}}_t^{(i)}$ が $p(\mathbf{x}_t|\mathbf{a}_t)$ と $p(\mathbf{x}_t|z_{1:t-1})$ のどちらからサンプルされたかを示す指標変

数（隠れ変数）とすると，オンライン対数尤度は，

$$\begin{aligned}\mathcal{L}(\alpha_t) &= \eta_t \sum_{\tau=1}^t \left(\prod_{s=\tau+1}^t \lambda_s \right) \sum_{i=1}^N \log \pi(\tilde{\mathbf{x}}_\tau^{(i)} | \alpha_t) \\ &= \eta_t \sum_{\tau=1}^t \left(\prod_{s=\tau+1}^t \lambda_s \right) \sum_{i=1}^N \log \left(\sum_m p(\tilde{\mathbf{x}}_\tau^{(i)}, m | \alpha_\tau) \right)\end{aligned}$$

と定義することができ，これを最大にする α_t^* を逐次求めることで，上記事後分布 $p(x_t | \mathbf{a}_t, \mathbf{z}_{1:t})$ との KL 距離を最小とする ASD $\pi(x_t | \alpha_t^*)$ を求めることが出来る．ここで， λ_s は過去の不正確な推定による影響を徐々に忘却するための忘却係数であり， $0 \leq \lambda(s) \leq 1$ とする．また， $\eta_t = \left(\sum_{\tau=1}^t \left(\prod_{s=\tau+1}^t \lambda_s \right) \right)^{-1}$ は正規化のための定数であり，学習係数の役割を果たす．また，

$$\begin{aligned}p(\tilde{\mathbf{x}}_t^{(i)}, m_a | \alpha_t) &= \alpha_{a,t} p(\tilde{\mathbf{x}}_t^{(i)} | \mathbf{a}_t) \\ p(\tilde{\mathbf{x}}_t^{(i)}, m_x | \alpha_t) &= \alpha_{x,t} p(\tilde{\mathbf{x}}_t^{(i)} | \mathbf{z}_{1:t-1})\end{aligned}$$

である．この時， α_t^* は

$$\alpha_{m,t}^* = \frac{\langle\langle m \rangle\rangle_t}{\sum_{m' \in m} \langle\langle m' \rangle\rangle_t}$$

で求めることができ，この α_t^* を α_{t+1} として，次時刻の ASD $\pi(x_{t+1} | \alpha_{t+1})$ を得る．ここで，

$$\begin{aligned}\langle\langle m \rangle\rangle_t &= \eta_t \sum_{\tau=1}^t \left(\prod_{s=\tau+1}^t \lambda_s \right) \sum_{i=1}^N p(m | \tilde{\mathbf{x}}_\tau^{(i)}, \alpha_\tau) \\ p(m | \tilde{\mathbf{x}}_t^{(i)}, \alpha_t) &= \frac{p(\tilde{\mathbf{x}}_t^{(i)}, m | \alpha_t)}{\sum_{m' \in m} p(\tilde{\mathbf{x}}_t^{(i)}, m' | \alpha_t)}\end{aligned}$$

である．これは以下のように再帰的に書くことができる．

$$\langle\langle m \rangle\rangle_t = (1 - \eta_t) \langle\langle m \rangle\rangle_{t-1} + \eta_t \sum_{i=1}^N p(m | \tilde{\mathbf{x}}_t^{(i)}, \alpha_t)$$

計算の詳細については付録 C に示す．

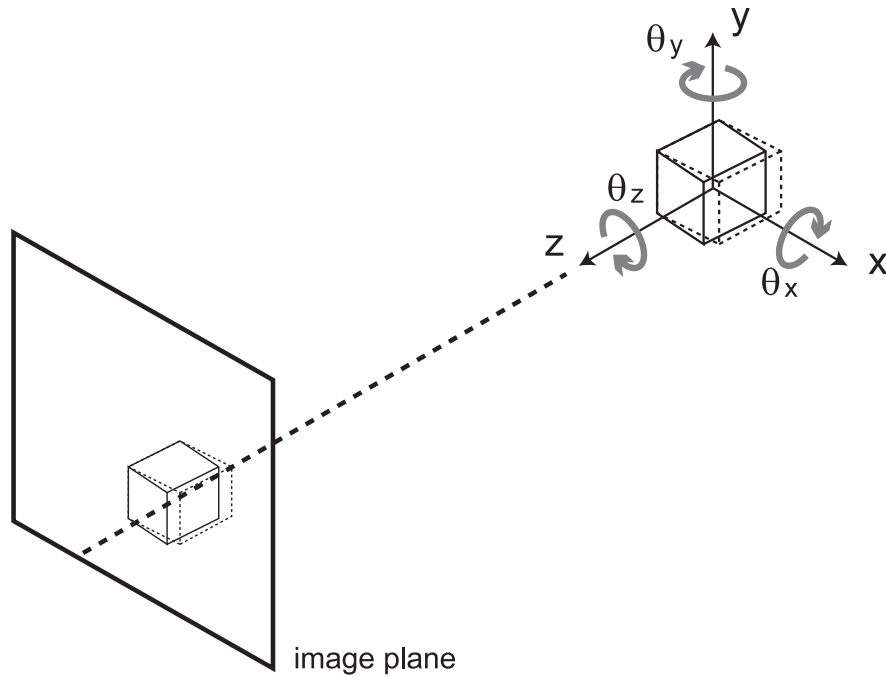


図 4.2 数値実験環境．剛体は重心を中心として X,Y,Z 軸周りにそれぞれ $\theta_x, \theta_y, \theta_z$ 回転し，画像面上に射影された特徴点位置を観測し，剛体の姿勢と特徴点位置の同時推定を図る．

3. 数値実験:剛体の三次元姿勢推定問題

前節で導入した ASD を利用したパーティクルフィルタ (Particle Filtering using Adaptive Sampling Distribution, PF-ASD) の有効性を確認するために，図 4.2 に示す様な剛体の重心を中心とした回転を行う剛体の三次元姿勢推定問題に適用する．単眼画像列からの剛体の三次元姿勢推定問題は，剛体の三次元姿勢を上位に，特徴点の画像面上での位置を下位にと，階層的な隠れ変数構造を持つ問題と考えることができる．今，時刻 t における画像面上での N_f 個の特徴点位置をそれぞれ $\mathbf{x}_{n,t} = (x_{n,t}, y_{n,t})^T, (n = 1, \dots, N_f)$ ，あらかじめ得られている剛体の三次元モデルを画像面に射影するアフィンカメラ行列を $P_{a,t}$ とすれば，各特徴点 $\{\mathbf{x}_{n,t}\}_{n=1}^{N_f}$ について

$$\mathbf{x}_{n,t} = P_a \tilde{\mathbf{X}}_{n,0}$$

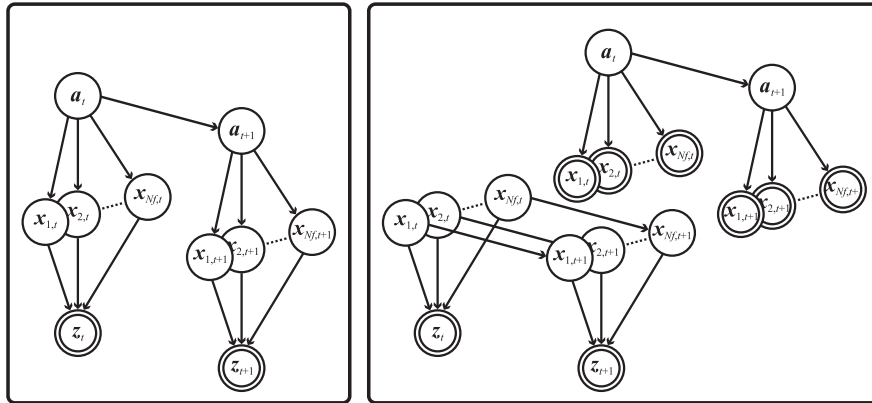


図 4.3 剛体の姿勢推定のグラフィカルモデル．左図は真の階層モデルを表し，右図はここで導入した実際には存在しない特徴点毎に独立なダイナミクスを仮定した並列ダイナミクスモデル．

が成り立つ．但し，

$$P_{a,t} = \begin{pmatrix} a_{1,t} & a_{2,t} & a_{3,t} & a_{4,t} \\ a_{5,t} & a_{6,t} & a_{7,t} & a_{8,t} \end{pmatrix}$$

とし， $\tilde{X}_{n,0} = (X_{n,0}^T, 1)^T$ ， $X_{n,0} = (X_{n,0}, Y_{n,0}, Z_{n,0})$ は剛体の三次元モデルから得られる n 番目の特徴点の三次元位置とする．このとき，アフィンカメラ行列 $P_{a,t}$ の各要素を並べたベクトルを $a_t = (a_{1,t}, \dots, a_{8,t})^T$ ，観測された画像を z_t とすると，これらは図 4.3 左図に示す階層グラフィカルモデルで表現される関係を持つ．これに同図右図に示す下位層に並列なダイナミクスを持つモデルを，その信頼度に応じて混合することで，少数のサンプル数を維持したまま頑健な状態推定の実現を目指す．ここでは，上位で頭部姿勢を表すアフィンカメラ行列の各要素 a_t を隠れ状態として推定することで，上位層の観測過程は線形となるのでカルマンフィルタを用いてこの推定を行う．なお，三次元空間における頭部姿勢 $\theta_t = (\theta_{x,t}, \theta_{y,t}, \theta_{z,t})^T$ のアフィンカメラ行列からの復元についての詳細は付録 D を参照．ここで， $\theta_{x,t}, \theta_{y,t}, \theta_{z,t}$ は X, Y, Z 軸を中心とした回転角である．実画像を入力とする下位層は，隠れ等の大きな雑音の影響で非ガウス過程となるため，パーティクルフィルタを用いて隠れ状態の分布推定を行う．このようにパーティクルフィルタにより推定する状態空間を x_t のみに限定することで，より少数のサンプルを用いた頑健な推定の実時

間での実現を図る．ここで，上位層では擬似的な観測変数 x_t に対してガウス分布を仮定し，一方，下位層では，隠れ変数である x_t は観測の非ガウス性により，その事後分布は非ガウス分布になる．すなわち，上位層での隠れ変数 a_t の推定時には入力分布をガウスとして近似することになる．また，一部に隠れが生じる場合に，隠れていない特徴点の推定精度が低下するのを防ぐため，ここでは各特徴点 $\{x_{n,t}\}_{n=1}^{N_f}$ の間に独立性を仮定し，各特徴点について混合比 $\alpha_{n,t}$ の決定を考える．三次元姿勢推定アルゴリズムを図 4.4 に示す．なお，ここで上位の隠れ変数 a_t をカルマンフィルタを用いて推定したのと同様にして，高次元な状態空間を解析的に周辺化可能な領域とそうでない領域に分割し，解析的に周辺化不可能な領域のみパーティクルを用いて推定を行う (Rao-Blackwellisation, RB[15]) 事で，パーティクルフィルタを用いて推定する状態空間を削減し，少数のサンプルで推定精度の維持を図る Rao-Blackwellised Particle Filter (RB-PF) [21] が提案され，その有効性が報告されている [22, 1]．ただし，RB-PF はサンプリング分布に階層間の相互作用を導入せず，ここでいう下位の推定精度が低下した際に上位からの情報を用いてこれを修正することは出来ない．

3.1 上位層における剛体の姿勢推定

前節で述べたように，上位層では隠れ状態変数として頭部姿勢を表すアフィンカメラ行列の各要素 a_t を採用する．これにより線形の観測過程を仮定することができるため，上位層ではカルマンフィルタを用いて頭部姿勢の時系列ベイズ推定を行う． a_t の初期分布はガウス分布であるとし，状態遷移過程を

$$\mathbf{a}_{t+1} = A_a \mathbf{a}_t + B_a \mathbf{u}_{a,t}$$

観測過程を

$$\mathbf{x}_t = C_a \mathbf{a}_t + D_{a,t} \mathbf{v}_{a,t}$$

1. 下位の隠れ状態 x_t の推定 .

- $p(x_t|a_t)$ からの $x_{a,t}^{(i)}$ を $\alpha_{a,t}N$ 個得る .
- $p(x_t|z_{1:t-1})$ からの $x_{z,t}^{(i)}$ を $\alpha_{z,t}N$ 個得る .
- $x_{z,t}^{(i)}$ と $x_{a,t}^{(i)}$ をまとめて , N 個のサンプル $x_t^{(i)}$ とし , パーティクルフィルタにより x_t の期待値 \hat{x}_t と推定分散 $\Sigma_x(t)$ を得る .

以上を各特徴点について行い , $\{\hat{x}_{n,t}, \Sigma_{x,n}(t)\}$ を得る .

2. 上位の隠れ状態 a_t の推定 .

- $\hat{x}_{n,t}$ と $\Sigma_{x,n}(t)$ を上位における観測値と観測ノイズ分散として , カルマンフィルタにより $p(a_t|z_{1:t})$ を推定 .

3. 混合比 α_{t+1} の決定 .

- $p(x_{n,t}|a_t, z_{1:t})$ からのリサンプルにより , N 個のサンプル $\tilde{x}_t^{(i)}$ を得る .
- $\mathcal{L}(\alpha_t)$ を最大化する α_t^* を求め , 新しく α_{t+1} とする .

以上を各特徴点について行い , $\{\alpha_{n,t+1}\}$ を得る .

4. 手順 1 に戻る .

図 4.4 階層的な剛体の姿勢推定手法の実行手順

と仮定する．状態遷移モデルとして線形二次のモデルを仮定するため，ここでは2.2節，3.2節の数値実験と同様に擬似的に状態変数を

$$\mathbf{a}_t = \begin{pmatrix} a_{1,t} \\ a_{1,t-1} \\ \vdots \\ a_{8,t} \\ a_{8,t-1} \end{pmatrix}$$

と置きなおし，

$$A_a = \begin{pmatrix} 2 & -1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad B_a = \begin{pmatrix} b_{a1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & b_{a8} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

とし， $b_{a1} = b_{a2} = b_{a3} = b_{a5} = b_{a6} = b_{a7} = 0.02$ ， $b_{a4} = b_{a8} = 1$ とする．観測行列 C_a は三次元の剛体モデルから得られる各特徴点の初期三次元位置 $\mathbf{X}_0 = \{\mathbf{X}_{1,0}, \dots, \mathbf{X}_{N_f,0}\}$ ， $\mathbf{X}_{n,0} = (X_{n,0}, Y_{n,0}, Z_{n,0})^T$ を用いて

$$C_a = \begin{pmatrix} C_{a,1} & \mathbf{0} \\ \mathbf{0} & C_{a,1} \\ \vdots & \vdots \\ C_{a,N_f} & \mathbf{0} \\ \mathbf{0} & C_{a,N_f} \end{pmatrix}, \quad C_{a,n} = \begin{pmatrix} X_{n,0} & 0 & Y_{n,0} & 0 & Z_{n,0} & 0 & 1 & 0 \end{pmatrix}$$

である．一般に実環境においての視覚追跡では，画像面上での特徴点位置 $\mathbf{x}_t = \{\mathbf{x}_{1,t}, \dots, \mathbf{x}_{N_f,t}\}$ の分布はガウス分布ではないが，ここでは非ガウス性を時間依存の雑音行列 $D_{a,t}$ で近似表現するものとして，ガウス分布を仮定する．すなわち， $D_{a,t}$ は \mathbf{x}_t の事後分布 $p(\mathbf{x}_t | \mathbf{a}_t, \mathbf{z}_{1:t})$ を $\hat{\mathbf{x}}_t$ と $D_{a,t}$ を用いて

$$p(\mathbf{x}_t | \mathbf{a}_t, \mathbf{z}_{1:t}) \approx \mathcal{N}_{N_x}(\mathbf{x}_t | \hat{\mathbf{x}}_t, D_{a,t} D_{a,t}^T)$$

の様にガウス分布で近似されるとき、 $v_{a,t}$ は白色ガウス雑音と仮定する．ここで、 \hat{x}_t と

$$D_{a,t} = \begin{pmatrix} \text{Std}(x_{1,t}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \text{Std}(x_{N_f,t}) \end{pmatrix}$$

は各々の顔特徴点 $x_{1,t}, \dots, x_{N_f,t}$ について、式(4.10)により求められる事後分布 $p(x_{n,t} | a_t, z_{1:t})$ の期待値と標準偏差を用いて表される．この時、事後分布 $p(a_t | z_{1:t})$ はカルマンフィルタを用いて以下で推定することができる．

$$p(a_t | z_{1:t}) = \mathcal{N}_{N_a}(a_t | \hat{a}_t, \hat{V}_{a,t}) \quad (4.4)$$

ここで、

$$\begin{aligned} \hat{a}_t &= \bar{a}_t + K_t(\hat{x}_t - C_a \bar{a}_t) \\ \hat{V}_{a,t} &= (I - K_t C_a) \bar{V}_{a,t} \\ K_t &= \bar{V}_{a,t} C_a^T (D_{a,t} D_{a,t}^T + C_a \bar{V}_{a,t} C_a^T)^{-1} \end{aligned}$$

である．ここで、 \hat{x}_t は $z_{1:t}$ に依存していることに注意する．また、これを用いて a_{t+1} の事前分布 $p(a_{t+1} | z_{0:t})$ は以下で与えられる．

$$p(a_{t+1} | z_{1:t}) = \mathcal{N}_{N_a}(a_{t+1} | \bar{a}_{t+1}, \bar{V}_{a,t+1}) \quad (4.5)$$

ただし、

$$\begin{aligned} \bar{a}_{t+1} &= A_a \hat{a}_t \\ \bar{V}_{a,t+1} &= B_a B_a^T + A_a \hat{V}_{a,t} A_a^T \end{aligned}$$

である．

3.2 下位層における特徴点位置推定

下位層では、画像面上での特徴点位置 $\{x_{n,t}\}_{n=1}^{N_f}$ を隠れ状態変数にとり、この時系列ベイズ推定を行う．先述の通り、実世界に存在する非ガウス雑音の影響を直接受ける下位層ではパーティクルフィルタを用いてこの推定を行う．

ここでは剛体の一部に隠れが生じる場合に、隠れていない特徴点の推定精度が低下するのを防ぐため、各特徴点に独立性を仮定し、以下の状態遷移/観測過程を仮定する。状態遷移過程には、上位層の場合と同様に、線形二次のモデルを仮定するため、 n 番目の特徴点の隠れ状態変数 $\mathbf{x}_{n,t} = (x_{n,t}, y_{n,t})^T$ を

$$\mathbf{x}_{n,t} = \begin{pmatrix} x_{n,t} \\ x_{n,t-1} \\ y_{n,t} \\ y_{n,t-1} \end{pmatrix}$$

で置きなおし、

$$\mathbf{x}_{n,t} = A_f \mathbf{x}_{n,t-1} + B_f \mathbf{u}_{f,t} \quad (4.6)$$

とする。ここで、

$$A_f = \begin{pmatrix} 2 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad B_f = \begin{pmatrix} b_{fx} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b_{fy} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

とし、 $b_{fx} = b_{fy} = 5$ とした。また、観測画像 z_t において観測された n 番目の特徴点位置を $y_{n,t}$ とすれば、 i 番目のサンプル $\mathbf{x}_{n,t}^{(i)}$ の重みは

$$w_{n,t}^{(i)} \propto \exp\left(-\frac{\|y_{n,t}(z_t) - \mathbf{x}_{n,t}^{(i)}\|^2}{2\sigma_x^2}\right) \quad (4.7)$$

とした。但し、ここでは $\sigma_x = 1$ とした。

ASD $\pi(\mathbf{x}_{n,t}|\alpha_{n,t})$ からのサンプル $\mathbf{x}_{n,t}^{(i)}$ を得るために必要な分布は以下のように書くことができる。まず、剛体の三次元モデルを画像面上に射影するアフィン変換過程を通して他の特徴点との位置関係の情報を含む、上位からの予測分布 $p(\mathbf{x}_{n,t}|\mathbf{a}_t)$ に関しては

$$p(\mathbf{x}_{n,t}|\mathbf{a}_t) = \mathcal{N}_{N_x}(\mathbf{x}_{n,t} | C_{a,n} \bar{\mathbf{a}}_t, C_{a,n} \bar{V}_{a,t} C_{a,n}^T) \quad (4.8)$$

により得られる．下位層からの予測分布 $p(x_{n,t}|z_{1:t-1})$ に関しては前時刻の事後分布 $p(x_{n,t-1}|z_{1:t-1})$ と下位層の状態遷移モデル $p(x_{n,t}|x_{n,t-1})$ とを用いて， $x_{n,t-1}$ を周辺化することにより得ることができる．実環境に存在する非ガウス雑音により $x_{n,t-1}$ の分布は非ガウスとなるため，これを解析的に求めることは困難であるが，前時刻の事後分布 $p(x_{n,t-1}|\mathbf{a}_{t-1}, z_{1:t-1})$ よりリサンプルした $\tilde{x}_{n,t-1}^{(i)}$ を用いて，式 (4.6) で仮定した状態遷移モデルにより

$$\mathbf{x}_{n,t}^{(i)} = A_f \tilde{\mathbf{x}}_{n,t-1}^{(i)} + B_f \mathbf{u}_{f,t} \sim p(\mathbf{x}_{n,t}^{(i)}|\tilde{\mathbf{x}}_{n,t-1}^{(i)}) \quad (4.9)$$

と時間発展させることにより， $p(x_{n,t}|z_{1:t-1})$ に従うサンプル $\mathbf{x}_{n,t}^{(i)}$ を得ることができる．

式 (4.8),(4.9) より，それぞれの予測分布から得られた N 個のサンプル $\mathbf{x}_{n,t}^{(i)}$ と式 (4.7) で示された重みを用いて， n 番目の特徴点 $\mathbf{x}_{n,t}$ の事後分布 $p(\mathbf{x}_{n,t}|\mathbf{a}_t, z_{1:t})$ は式 (2.12) より，

$$p(\mathbf{x}_{n,t}|\mathbf{a}_t, z_{1:t}) \approx \sum_{i=1}^N \tilde{w}_{n,t}^{(i)} \delta_{\mathbf{x}_{n,t}^{(i)}}(\mathbf{x}_{n,t}) \quad (4.10)$$

で得ることができる．ここで

$$\tilde{w}_{n,t}^{(i)} = \frac{w_{n,t}^{(i)}}{\sum_{j=1}^{(j)} w_{n,t}^{(j)}}$$

とする．

3.3 実験結果

ここでは，図 4.5 に示すような上底 20mm，下底 30mm，奥行き 20mm の台形柱の姿勢推定を行う．カメラには内部パラメータが既知のピンホールカメラを仮定し，焦点距離を 40mm，各ピクセルの大きさは 0.1mm 正方とした．この台形柱の姿勢を変化させ，画像面上に透視投影された頂点の座標値に標準偏差 1pixel のガウスノイズを付加し，観測値として用いた．隠れている頂点に関しては標準偏差 100pixel のガウスノイズを付加して精度の低い観測値として用いる．8 つの頂点の内，後側にある 4 つには隠れが存在し，前側の 4 つは完全に観測可能であ

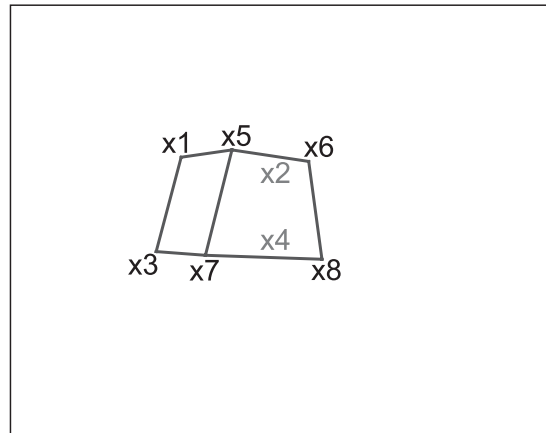


図 4.5 シミュレーション画像例．図中に示した特徴点 x_1, \dots, x_8 にノイズが付加され観測される．薄いグレーでしめした特徴点 (x_2, x_4) は剛体自身によって隠れるため，大きな観測ノイズが付加される．

るように，台形柱の姿勢を変化させることで，625 フレームの頂点の画像面上での座標値からなるデータを生成し，これを元に台形柱の姿勢の復元を行った．

ここでは，PF-ASD の他に，(i) $\alpha_t = \{1, 0\}$ で固定し，完全に上位からの予測分布を信頼して推定を行う手法，(ii) $\alpha_t = \{0, 1\}$ で固定し，完全に下位からの予測分布を信頼して推定を行う手法，(iii) 厳密な RB-PF の三つの手法も合わせて適用し，PF-ASD の有効性を確認する．PF-ASD 及び，手法 (i)，手法 (ii) においては，各頂点の位置推定に用いるサンプル数は各 50 個，計 400 個とし，手法 (iii) の RB-PF については各頂点を同時に扱う必要があるため， $N = 25,000$ として適用した．

図 4.6 に提案手法により推定された剛体の姿勢を，図 4.7 に提案手法においてオンライン EM アルゴリズムによって推定されたサンプリング分布の混合比を，図 4.8 に各手法の 100 試行での平均推定誤差を示す．図 4.8 に見るように， $\alpha_t = \{1, 0\}$ と固定した手法 (i) は，ここでは上位層に用いている三次元モデルが正確であるため，精度良く推定を行うことができているが，画像面上への射影にアフィン射影を仮定している影響で誤差が若干上昇している．手法 (ii) は，上位層からの情報を下位層では全く用いず，各頂点を完全に独立に扱うため，隠れの存在する頂点の推定精度が大きく低下し，ひいては上位層で推定する台形柱の姿勢にも悪影響を及ぼす．また，(iii) の RB-PF では下位層で各頂点をまとめて扱うため，依然

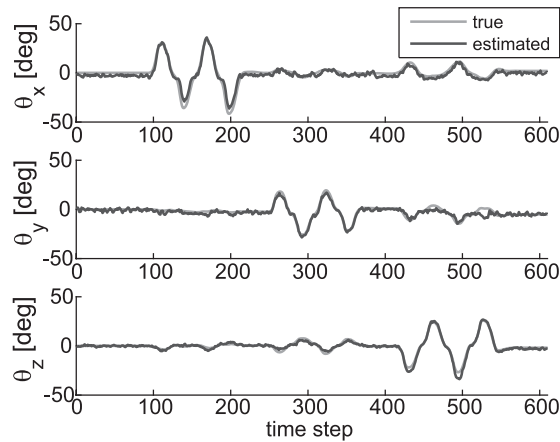


図 4.6 推定された剛体の姿勢

として高次元な状態空間を扱う必要があり，サンプル数の不足から来る精度の低下が，台形柱の姿勢推定にも悪影響を与えている．PF-ASD によれば，オンライン EM アルゴリズムで推定された混合比（図 4.7）からも分かるように，図中グレーの領域で示されるような頂点が隠れている期間には上位層からの情報を重視してサンプリング分布を生成し，隠れていない特徴点に関しては，隠れている頂点の影響や用いている近似の崩れなどで精度の低下した上位からの予測分布を重視せず，下位からの予測分布を重視してサンプリング分布を生成する．このように，適応的に混合比の制御を上手く行うことができおり，そのため，長期の隠れや近似の崩れなどに関わらず，頭部姿勢，画像面上での頂点位置共に頑健で正確な追跡を実現できている．

3.4 まとめ

上位層の強い仮定（対象の三次元モデルやアフィン射影の仮定等）により精度の低下した下位層の推定に必要な事前分布 $p(x_t|a_t, z_{1:t-1})$ を修正するために，本研究では下位層にも独立な状態遷移を考え，上位層からの予測分布 $p(x_t|a_t)$ と下位層からの予測分布 $p(x_t|z_{1:t-1})$ の混合分布 $\pi(x_t|\alpha_t)$ を下位層のサンプリング分布として用いた．ここで用いた上位層からの予測分布 $p(x_t|a_t)$ は対象の三次元モデルのアフィン変換を通して他の特徴点との位置関係の情報を含んでおり，それに対

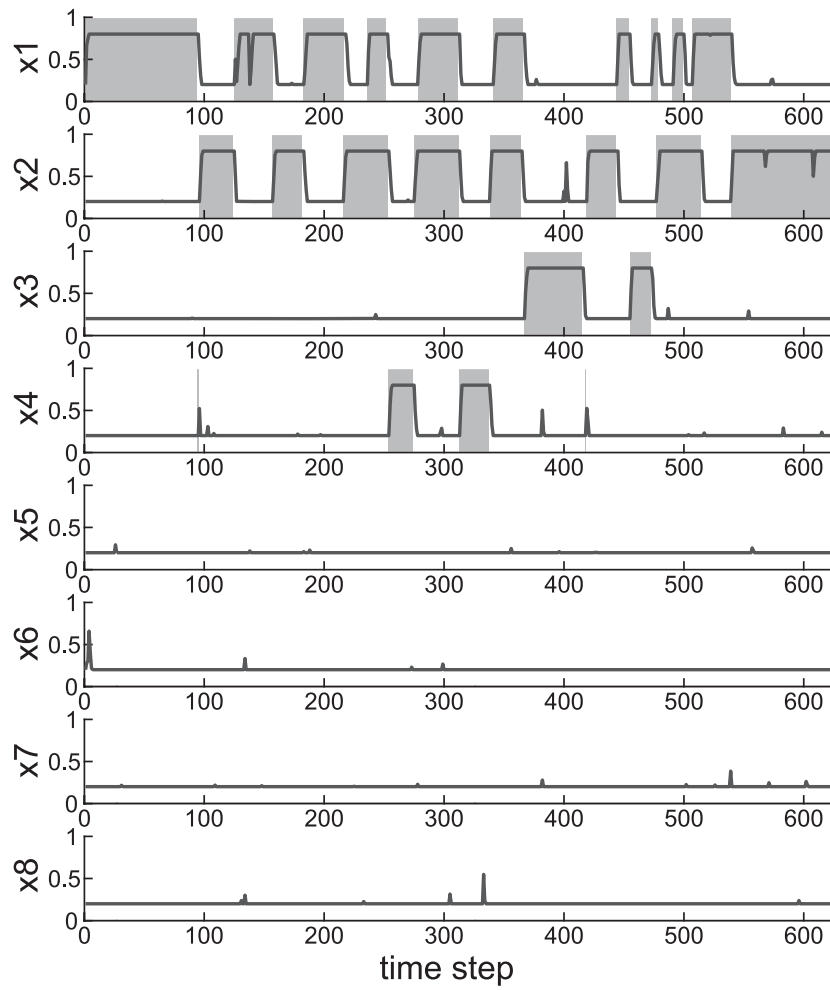


図 4.7 各時刻で推定された混合比 $\alpha_{a,t}$. 図中グレーの領域は特徴点が隠れてしまう時刻を示している .

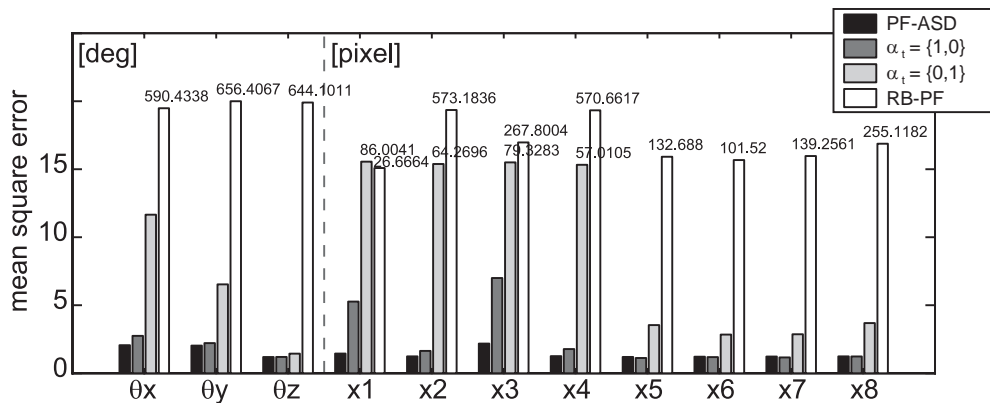


図 4.8 PF-ASD, $\alpha_{a,t} = 1$ で固定したもの, $\alpha_{a,t} = 0$ で固定したもの, RB-PF による 100 試行の平均推定誤差

して $p(x_t|z_{1:t-1})$ は前時刻に推定された事後分布 $p(x_{t-1}|z_{1:t-1})$ から下位層の状態遷移モデルを用いて, 特徴点ごとに独立に得られる予測分布である. 上位のモデルでは各特徴点が三次元モデルに従い存在すること, アフィン変換により対象の姿勢が復元可能であることなどの近似をおいたため, 特に, 顔など個人間のバラつきを含む対象を扱う際には, 上位からの予測の精度には限界がある. それに対して下位層の状態遷移モデルを用いて各特徴点を独立に予測すれば, 上位で仮定した近似の影響を受けずに予測が可能である. ただし, 実世界では照明条件の変化や, 遮蔽の存在など, 特徴点が観測できない, もしくは大きな雑音の存在する場合が多々考えられる. こうした場合に, 各特徴点を独立に予測, 推定することは追跡の頑健性を著しく損なう要因となる. このように相補的な性質を持つ二つの予測分布を式 (4.3) により混合し, 状況に合わせて混合比を制御すれば, 二つの予測分布の長所を結合し, サンプル数を多く用意することなしに頑健で正確な推定を行うことができた.

相補的な二つの手法を切り替えることで頑健な追跡の実現を図る手法はいくつか提案されている [81, 5, 68] が, 多くは切り替えの規準になるパラメータを事前に決めておく必要があった. ここで提案した PF-ASD では, 混合予測分布 $\pi(x_t|\alpha_t)$ を事後分布 $p(x_t|a_t, z_{1:t})$ に近づけるように混合比 α_t をオンライン EM アルゴリズム [72] により逐次推定することで, このパラメータを自動的に決定した. オンラ

イン EM アルゴリズムの忘却係数 λ_t は、ここでは $\lambda_t = 0.5$ と非常に大きく設定した。一般に学習の安定化には λ_t を小さく設定するのが良いとされるが、小さな λ_t では急激に変化する状況への追従が困難である。小さな λ_t は時間方向の推定誤差の除去に有効であるが、比較的大きな λ_t を用いつつ、それに応じてパーティクル数を増やすことで α_t の不安定さを吸収することができる。ここで設定したパーティクル数 N は 50 個と少数であったが、 α_t の不安定さを吸収するには十分であると考えられる。

また、ここでは剛体の回転運動のみを考え、剛体とカメラの相対的な位置は変化しないことを仮定することで、上位層で推定する状態変数としてアフィンカメラ行列を採用することが出来た。しかし、現実には追跡対象とカメラの相対位置は変化するのが一般的であり、上位層での状態変数として並進成分を含む三次元姿勢を考える必要がある。この場合には本節で述べたアフィンカメラ行列を上位層の状態変数と仮定することは出来ないが、拡張カルマンフィルタを用いて逐次的にカメラ行列を近似しなおすことで、並進成分も含む三次元姿勢を推定することが可能である。拡張カルマンフィルタを用いた上位層での並進成分も含む三次元姿勢の直接推定に関する詳細は付録 E に示す。

4. 応用:実際の車内で撮影されたドライバの頭部画像列からの頭部姿勢推定

近年、快適で安全な車社会の実現のためにドライバ支援システムの研究が盛んに行われるようになってきている [3, 17, 77, 57, 98, 106]。今後高齢化社会の進行と共に、体調や年齢に応じて低下することのある認知能力の補助機構はますます重要になってくると考えられるが、実際の車内で直接観測することのできるドライバの情報は少なく、視線方向など重要な情報は直接観測できない場合も多い。特に、視線方向はドライバの状態を知る上で重要な情報であり、視線方向や、視線方向を知るためのドライバの頭部姿勢推定の試みも多く行われてきた [77, 57, 98, 106]。Smith ら [77] は単眼カメラを用いて、色情報により顔特徴点を検出し、頭部姿勢を顔特徴点の位置関係により復元し、視線方向の検出を図った。Matsumoto と

Zelinsky[57] は，ステレオカメラを用いて作成したステレオ画像と三次元座標からなる三次元顔モデルを用いて，テンプレートマッチングにより得られた顔特徴点の三次元座標位置に重みつき最小二乗法を適用することで頭部姿勢の推定を行い，視線方向を検出した．林ら [98] は，分離度フィルタ [107] 及び更新型テンプレートマッチングの両方を用いて，それらのマッチ位置を重み付け統合する手法を採用し，この重みパラメータを適切に設定することにより，頑健で高精度な顔特徴点の追跡を実現し，これに基づいてドライバの顔方向の推定を行った．白井ら [106] は，撮像時にのみ近赤外光のパルス光を投射するため周囲の照度変化に影響を受けにくい近赤外パルス光投射方式を用いた撮像システムを用いて視線方向の推定を行った．顔の検出にはニューラルネットを使用し，顔方向の検出は両眼と鼻孔の情報を用いて行う．視線方向の推定は検出された眼領域画像からテンプレートにより検出した瞳孔位置と眼頭位置を用いて行う．

また，頭部姿勢推定は上記の例以外にも，ロボットの視覚システムなど広い分野への応用が考えられるため，人物の（頭部）姿勢推定 [4, 6, 12, 46, 48, 47, 55, 99]，顔検出 [7, 53, 88]/追跡等 [11, 29, 30, 31, 90, 91] 多くの研究が行われている．頭部姿勢推定問題へのアプローチとして，近年その頑健性から確率論的な時系列フィルタリングが多く用いられており，前述の通り頭部姿勢推定問題の持つ非線形過程や雑音の非ガウス性などからマルコフ連鎖モンテカルロ (MCMC) 法 [46, 48, 47] やパーティクルフィルタ [4, 12, 55, 75, 99] などの非ガウス分布を扱うことのできる確率論的手法を用いることで頑健に頭部姿勢推定を行うことのできる手法が多く提案されている．

しかしながら，これらのドライバ状態推定手法や人物の頭部姿勢推定手法は，カラー情報を必要とするもの [46, 48, 47, 77]，ステレオ視のためカメラを複数台必要とするもの [3, 57] など，実際の車内での使用を考える際には車内に乗せることの出来るカメラの制約や，更には極端に制限された計算資源の問題などから直接適用することは困難な場合がある．そこで，本研究では照明条件の変化に比較的頑健であることの知られている近赤外カメラにより実際の車内で観測された単眼グレー画像列から，ドライバの視線方向推定のためのドライバの頭部姿勢と複数の顔特徴点の画像面上における位置を，実際の車内で処理を行うために極端に

制限された計算資源にも関わらず，実時間で精度の良い同時推定を頑健に行うことを考える．本研究でも確率的な時系列フィルタリングの一種であるパーティクルフィルタにより，頭部姿勢の時系列推定問題を解決する．加えて，既に提案されている手法の多くが問題の階層構造を利用した一つのダイナミクスモデルに基づくものであるのに対して，提案手法は先述の通り，複数のダイナミクスモデルを用意し，各階層でのモデル化誤差や大きな雑音の存在に応じて，各層のダイナミクスの推定への寄与の大きさを制御することにより，より頑健な推定の実現を目指す．また，線形ガウス過程を仮定できる階層が存在すれば，その層は解析的な確率的フィルタリングアルゴリズムであるカルマンフィルタを用いて推定することにより，より少数のサンプルにより精度を維持した推定の実現が可能であると期待される．この枠組みは先述の通り Rao-Blackwellisation [15] (RB) を利用した RB-PF [21] と関連が深い．但し，RB-PF は階層間に相互作用を明に記述することが困難で，大きな雑音の存在や非定常にモデル化誤差が変化するような状況においては，推定精度の悪化が考えられる．

4.1 適応的サンプリング分布を採用したドライバ頭部画像列からの頭部姿勢推定

本節では PF-ASD を採用した階層的状態推定法を，自家用車内で撮影された単眼画像からのドライバの頭部姿勢推定問題に適用し，著しく計算資源の制限されるような状況においても安定してドライバの頭部姿勢と顔特徴点位置の実時間推定が可能であることを示す．ここでは，図 4.9 に示す様に実際の車のハンドル奥に設置された近赤外カメラを用いて撮影された大きさ 640×480 の画像を入力画像として用いる．近赤外カメラは照明条件の変化に頑健であるが，輝度情報しか得られないため，色の情報を推定に用いることは出来ない．図 4.10 に観測された画像例を示しておく．

ここでは 3 節での数値実験と同様に，上位の隠れ状態として頭部姿勢を，下位の隠れ状態として両目中心，目尻，目頭，口尻，眉の内端と鼻の計 11 点の顔特徴点を採用し，3 節での数値実験と同様の手順で推定を行う．状態遷移過程に関

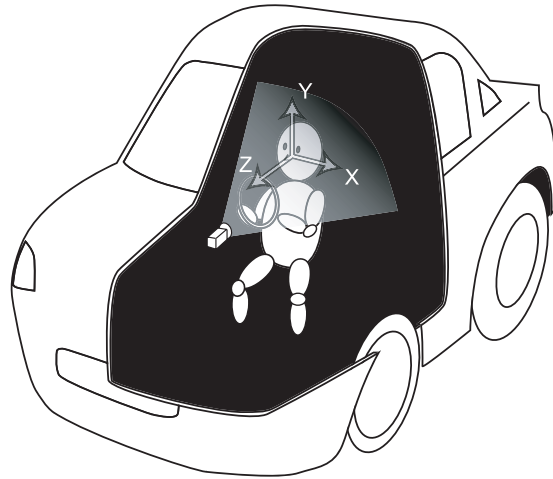


図 4.9 ドライバの頭部姿勢推定の実験環境．ハンドル奥に設置した単眼カメラを用いてドライバの頭部周辺画像を撮影し，頭部姿勢と顔特徴点位置の推定を図る．



図 4.10 実環境中で計測された人物の頭部画像例．顔特徴点は頭部姿勢に応じてその「見え」を大きく変化させ，目尻など一部隠れてしまう特徴点も存在する．

しては，上位下位共に3節で仮定したモデルと同一のモデルを仮定することができるが，観測過程に関しては実環境における問題を扱うため，同様のモデルを仮定することが出来ない．数値実験においては，上位の観測過程において剛体の三次元構造が既知であることを仮定したが，車内でドライバの頭部を三次元計測することが出来ないため，何らかのモデルを導入する必要がある．また，下位の観測モデルに関しては実際の画像を扱う上で，直接顔特徴点の位置を計測することが出来ないため，画像から何らかのモデルを用いて顔特徴点位置の観測値を得る必要がある．以下ではドライバの三次元顔モデル獲得法と，顔特徴点の観測モデルについて簡単に説明する．

4.2 三次元顔モデル

本実験で想定している環境では，実際の車内でドライバの三次元顔モデルを計測することが出来ないため，上位層における観測過程において必要なドライバの三次元顔モデルを得るためには，何らかの手法を用いてこれを推定する必要があるが，ここでは簡単のため，あらかじめ三次元計測器で計測された男女13人の顔モデルを位置あわせ後に平均した平均顔モデルを用いた．13人の顔モデルと平均顔モデルを図4.11に示す．なお，この13人中に後の実験で用いる画像列のドライバは含まれていないことを付記しておく．

一般に画像からの三次元構造の獲得法としては，古くからステレオビジョン等を始め様々な研究がなされているが，ここでは車内に設置可能なカメラ数の制限から単眼の画像列からの三次元構造復元を考える必要がある．単眼カメラ画像列からの三次元構造復元手法としてはTomasiとKanadeの因子分解法[84]が有名であるが，この手法は複数枚に渡って特徴点の時系列データを得る必要がある．そこで，ここでは推定初期においては平均顔モデルを用いて頭部姿勢推定を実行し，得られた特徴点時系列を用いてドライバ個人に適応した三次元顔モデルをオンラインで獲得することを考える．本章では，適応的サンプリング分布を採用することによる推定精度の改善について考察するため，この話題については5節で述べることとする．

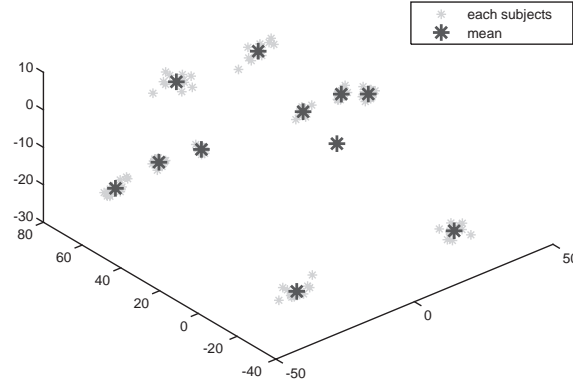


図 4.11 ドライバの平均三次元顔モデル．薄い* はそれぞれ計測された個人の顔特徴点の三次元位置を，濃い* は平均特徴点位置を示す．

4.3 画像からの顔特徴点観測モデル

階層での隠れ状態推定において，真の顔特徴点位置 $x_{n,t}$ が直接観測できないことに対応して，真の観測過程 $p(z_t|x_{n,t})$ は未知であるため，サンプルの重み $w_{n,t}^{(i)}$ の計算には何らかのモデルが必要である．ここでは，テンプレートとの正規化相関を用いてパーティクルの重み $w_{n,t}^{(i)}$ を決定する．時刻 t において画像 z_t が観測されたとすると，特徴点 x_n の標準テンプレート r_n と時刻 t における位置 $x_{n,t}$ の周辺で作った観測画像 z_t からの切り出し画像 $s_{n,t}$ の間の正規化相関値 $r(r_n, s_{n,t})$ を用いて，サンプルの重み $w_{n,t}^{(i)}$ は

$$w_{n,t}^{(i)} \propto \exp\left(-\frac{(r(r_n, s_{n,t}) - 1.0)^2}{2D_f^2}\right) \quad (4.11)$$

と決定する．この $w_{n,t}^{(i)}$ の計算が陰に観測過程 $p(z_t|x_{n,t})$ を規定していることになる． D_f は正規化相関値の変化がパーティクルの重みに与える影響の大小を決めるパラメータであり，ここでは正規化相関値が 0.5 以下でパーティクルの重みがほぼ 0 になるように， $D_f = 0.15$ と設定した（図 4.12 参照）．

ここでは，特徴点の標準テンプレート $r_{n,t}$ として，追跡開始時に正面向きのドライバの頭部画像から検出した特徴点位置周りで切り出した画像を用いた．特徴

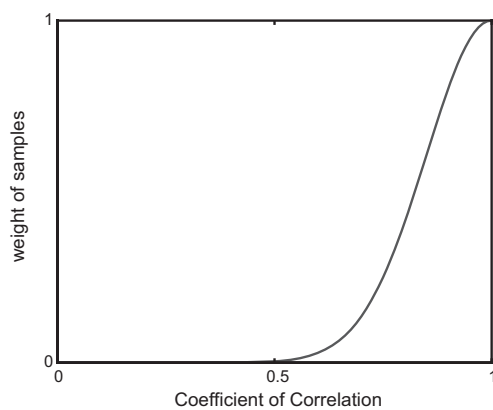


図 4.12 テンプレートとの正規化相関値とサンプルの重みの関係

表 4.1 各特徴点テンプレートの大きさ

	テンプレートの大きさ ($x \times y$)
鼻	64 × 42
目中心	64 × 32
目頭, 目尻	36 × 36
口端	24 × 24

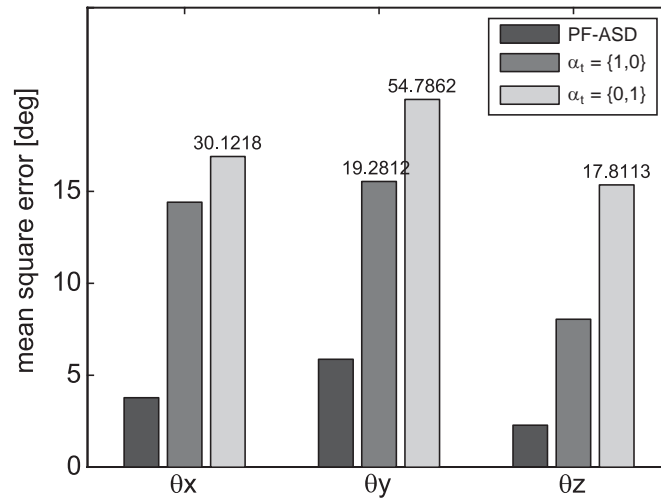


図 4.13 PF-ASD, $\alpha_{a,t} = 1$ で固定したもの, $\alpha_{a,t} = 0$ で固定したものによる頭部姿勢の平均推定誤差 (50 試行) .

点の検出手法としては, 頑健で高速な特徴点抽出が期待される Viola and Jones の AdaBoost を利用した手法 [88] を用いた. 用いた各特徴点テンプレートの大きさは表 4.1 に示す.

4.4 実験結果

PF-ASD の他に, (i) $\alpha_t = \{1,0\}$ で固定し, 完全に上位からの予測分布を信頼して推定を行う手法, (ii) $\alpha_t = \{0,1\}$ で固定し, 完全に下位からの予測分布を信頼して推定を行う手法を適用し, PF-ASD の有効性を確認する. 各特徴点追跡に用いるパーティクル数は 50 個とした. RB-PF については必要とするサンプル数が非常に多く, 実時間処理の実現が困難であったので, ここでは比較対象から除外した. 図 4.13 に各手法の頭部姿勢の推定誤差を示す. 誤差を求めるための頭部姿勢はドライバの頭部にジャイロセンサを着けて計測した. 剛体の姿勢推定の数値実験と同様に提案手法を用いることで, 従来手法では困難であった, 頭部姿勢推定が頑健に行うことが可能となる. また, 図 4.15 に提案手法で推定された $\alpha_{a,t}$ の時系列変化を示しておく.

ここでは, 正面向きの初期画像から獲得された標準テンプレートを用いて顔特



図 4.14 提案手法に推定された顔特徴点位置 . + の大きさは推定分散の逆数 (推定の確信度) を示し , 大きいほど確信を持って推定していることを表す .

徴点を検出しているため、横向き時など正面向きから頭部姿勢が変化すれば検出精度は著しく低下する。その影響を受けて、図 4.14 に示す様に、頭部姿勢の変化が大きく、隠れや見えの変化により不確かな観測しか得ることのできなかつた特徴点に関しては、推定精度が低下している。この様に観測モデルのモデル化誤差に起因した観測情報の精度低下により、各特徴点に独立性を仮定したため頭部姿勢を通して他の特徴点の情報を得ることの出来ない (ii) の手法では、こうした場合に頑健な特徴点位置推定を継続することが出来ない。

特徴点位置は頭部姿勢にのみ依存するとする (i) の手法もまた、上位層において幾つかの仮定（線形のカメラモデルの採用、頭部が純粋な回転運動のみを行うとした仮定、アフィンカメラ行列の各要素に仮定した独立性など）が崩れた場合には、状態遷移モデルのモデル化誤差に起因した雑音により、画像面上での特徴点位置の予測を精度良く行なうことが出来ず、特徴点位置推定に有効なサンプル数が少なくなり、頭部姿勢推定の精度も大きく低下してしまう。

それに対して提案手法によれば、このような精度の低い観測・予測情報にも関わらず、二つの階層の予測分布の信頼度を適応的に変化させる (cf. 図 4.15) ことで、重視すべき予測分布を状況に合わせて適切に決定し頑健な推定の実現が可能となっている。なお提案手法によれば、一枚の入力画像を処理するために必要な計算時間は Pentium4 2.8G 1048Mb RAM の Windows2000 が動作する PC で画像の入力も含めて 29.05 msec であり、十分に実時間で動作が可能であった。

4.5 まとめ

本章では、適応的サンプリング分布 (ASD) を採用した高次元状態推定法を剛体の姿勢と特徴点位置の同時推定問題、実際の車内で撮影された画像列からのドライバの頭部姿勢と顔特徴点の同時推定問題に適用し、その有効性を確認した。ここでは頭部姿勢と顔特徴点の同時推定問題を、その物理的關係から上位層で頭部姿勢を、下位層で顔特徴点を推定するように階層化し、下位層で顔特徴点を推定することによる実環境中の非ガウス性雑音の影響が小さいことと、処理時間の低減のために、上位層と下位層の間にアフィンカメラモデルを採用し、上位層で頭部姿勢を表現するアフィンカメラ行列の各要素を推定することで実時間で処理

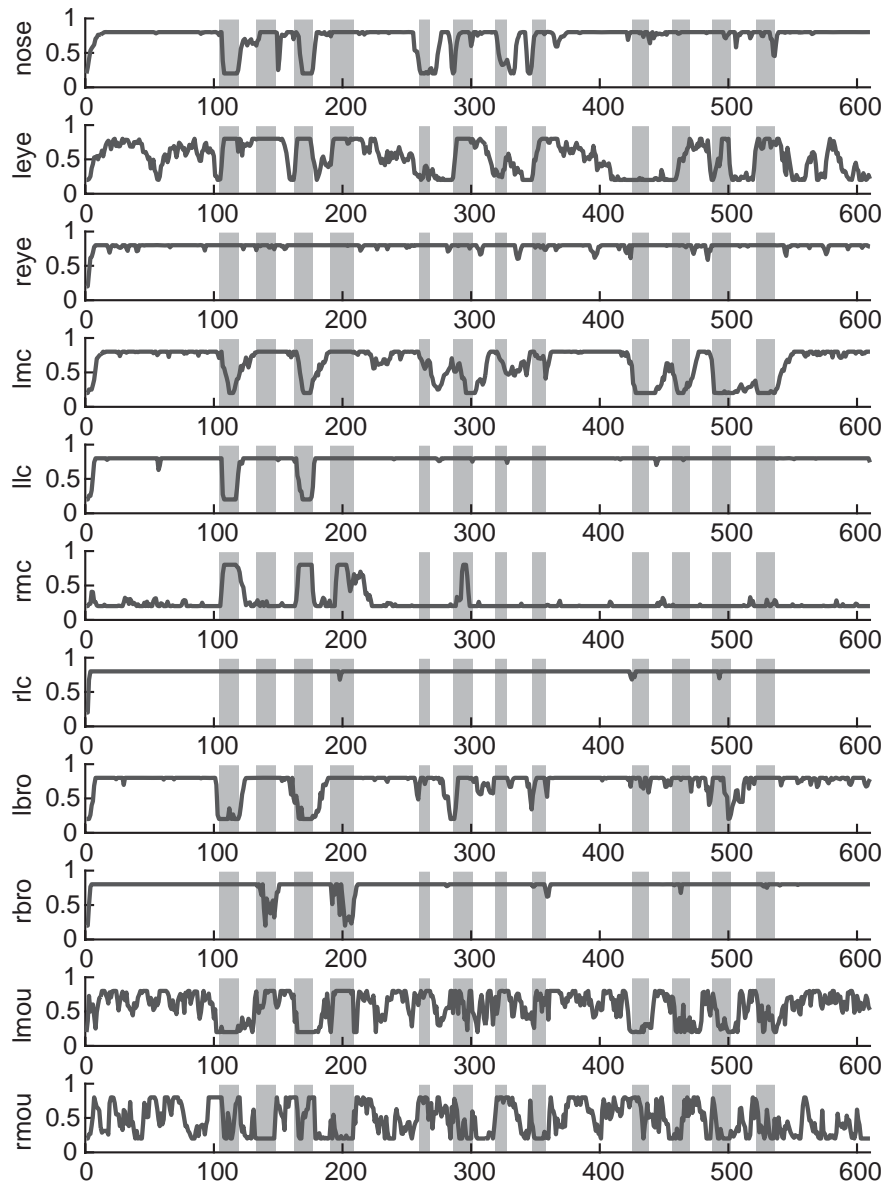


図 4.15 各時刻で推定された混合比 $\alpha_{a,t}$. 图中グレーの領域は顔向きが正面から右左右左，上下上下，右左右左回転の順に 20 度以上回転した時刻を示す．また，“nose” は鼻中心，“leye” は左目中心，“reye” は右目中心，“lmc” は左目頭，“llc” は左目尻，“rmc” は右目頭，“rlc” は右目尻，“lbro” は左眉，“rbro” は右眉，“lmou” は左口尻，“rmou” は右口尻を，それぞれ示す．

を実現した。また，下位層において各特徴点の間に独立性を仮定することで，部分的な隠れに対してある特徴点の推定精度の低下が他の特徴点に伝播するのを防ぎ，逆に上位の頭部姿勢を通して得られる他の特徴点の情報を利用して，隠れが生じている特徴点に関しても頑健な推定を実現した。

実環境で考えられる，西日や街灯の照明の影響など厳しい照明条件の変化に対する頑健性を考慮して，ここでは頭部姿勢の撮影用のカメラに近赤外光カメラを採用した。このため，画像からの顔特徴点の観測の再にも，色の情報を用いた特徴点検出手法を採用することが出来ず，ここでは観測モデルとして最も単純なものの一つである正面向き顔特徴点テンプレートとの正規化相関を用いた。この単純な観測モデルは頭部姿勢の変化によって非常に大きなモデル化誤差を含み，ここではPF-ASDを用いることによって大きなモデル化誤差による推定精度の低下を抑えることを実現したが，更なる推定精度の向上のためには観測モデルの改善は不可欠である。ここで採用したテンプレートを用いた視覚追跡手法としても，顔の様にその姿勢によって見え方を変える物体の視覚追跡手法として，テンプレートを変形させることで頑健な視覚追跡の実現を考える手法 [94, 38, 40]，更に遮蔽などによる隠れを考慮する手法 [95, 61, 60, 104] など，様々な手法が提案されている。特にテンプレートの変形と追跡対象の姿勢の変化による見えの変化との間の関係を線形近似し，これを予め学習しておくことで頑健な追跡を実現する手法 [40] は，三次元剛体の姿勢推定への有効性も報告されている [56]。こうした見えの変化を伴う対象の状態推定手法として，基底画像を用いた手法も提案されており [8]，パーティクルフィルタを用いた視覚追跡手法においても，その有効性が示されている [43, 30, 51, 52]。

また，ここでは頭部姿勢推定に必要な三次元顔モデルには，予め計測された複数人の顔モデルを位置・スケールを合わせた後平均した平均顔モデルを採用した。平均顔モデルを用いることによるモデル化誤差も，先述の観測モデルの含む大きなモデル化誤差同様，PF-ASDを用いることで影響を最小限に抑えることができたが，更なる推定精度の向上のためには，ドライバ個人に適応した三次元顔モデルを得ることが必要となる。図 4.16 に平均顔モデルと予め計測しておいたドライバ個人の三次元顔モデルを用いて PF-ASD による頭部姿勢の推定を行った場合

の推定成功率を示す．推定成功を定義するために，ここでは試行内の平均絶対誤差を用い，この閾値を変化させながら 100 試行中の推定成功率の変化を $\theta_x, \theta_y, \theta_z$ について調べた．図中で明らかな様に，計測した個人の三次元顔モデルを用いることで，平均顔モデルを用いた場合よりも平均絶対誤差が小さい試行が増えており，推定精度を高く維持するために精度の良い三次元顔モデルが重要であることがわかる．ステレオ視による手法等，画像からの三次元構造の復元課題は古くか

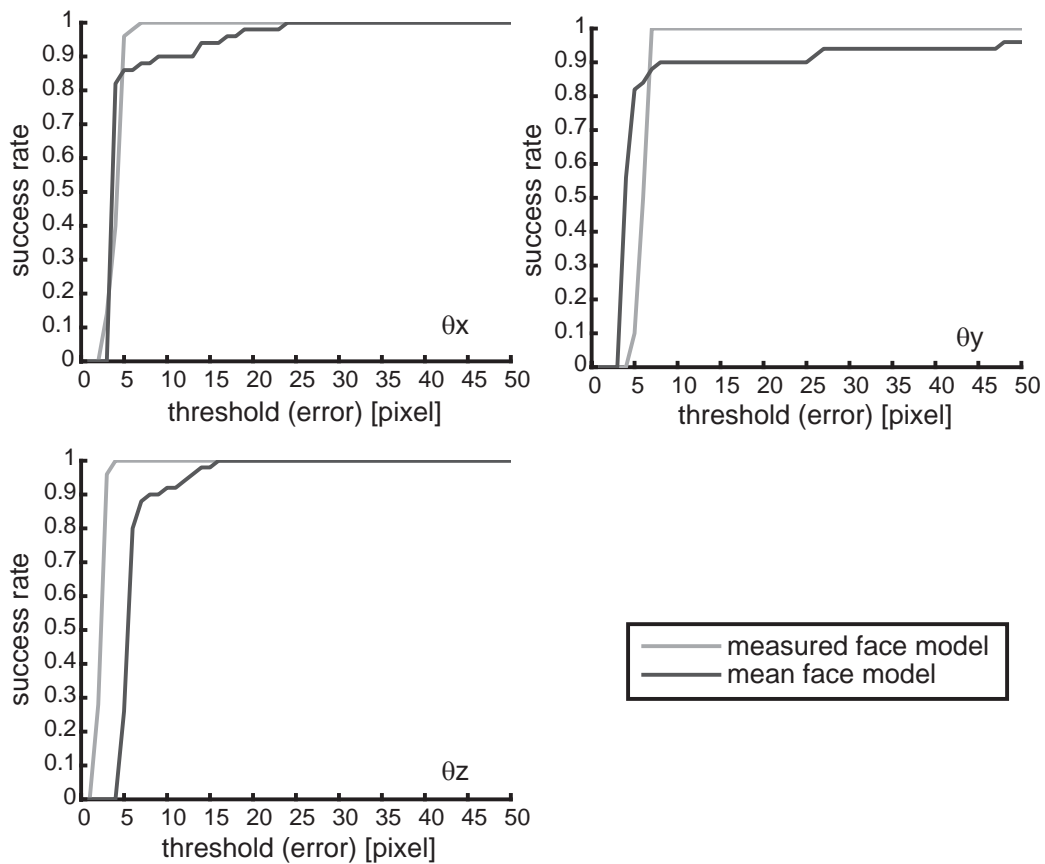


図 4.16 平均顔モデルと個人顔モデルを用いた頭部姿勢推定法の推定成功率 (100 試行) の比較．推定成功を定義する閾値には試行内における平均絶対誤差を採用した．

らコンピュータビジョンの分野における中心的な課題として多く研究されてきたが，ここではカメラの設置台数の制約から単眼カメラを用いた場合について考える必要がある．この話題に関しては 5 節で詳しく述べる．

5. ドライバの三次元顔モデル推定

前節では車内で撮影された単眼カメラ画像列から，ドライバの頭部姿勢と顔特徴点位置をドライバの三次元顔モデルを用いて頑健に推定する手法を提案した．しかし，実際のドライバの三次元顔モデルを予め得ておくことは現実的に困難であり，前節ではドライバ個人の三次元顔モデルの代わりに，複数人の三次元顔モデルを平均化した平均顔モデルを用いていた．PF-ASD を用いることにより，平均顔モデルを用いることによるモデル化誤差の影響は最低限に抑えて頑健な頭部姿勢推定が実現可能であったが，更なる推定精度向上のためには，この平均顔モデルの含むモデル化誤差を低く抑えることが重要な課題の一つである．そこで，ここでは推定初期においては平均顔モデルを用いて頭部姿勢推定を実行し，得られた特徴点時系列を用いてドライバ個人に適応した三次元顔モデルをオンラインで獲得することを考える．一般に画像からの三次元構造の獲得法としては，古くからステレオビジョン等を始め様々な研究がなされているが，ここでは車内に設置可能なカメラ数の制限から単眼の画像列からの三次元構造復元を考える必要がある．単眼カメラ画像列からの三次元構造復元手法としては Tomasi と Kanade の因子分解法 [84] が有名であるが，この因子分解法は，カメラモデルをアフィン射影で近似することにより，複数の画像における特徴点の対応の組から線形計算のみを用いてそれぞれの特徴点とカメラの三次元位置を復元することのできる手法であり，その数値計算における安定性と，十分実用的な精度を持つことから広く研究されてきた．しかし，検出された特徴点の対応の組に，特徴点の隠れや誤検出から生じる欠損やズレが生じた場合には因子分解法を単純に適用することはできない．そのような状況は実問題における適用に際して頻繁に起こり得るため，いくつかの改善手法が提案されてきた [76, 45, 79, 103]．その多くは特徴点検出法には主眼をおかず決定論的な特徴点検出法を用いていたが，本論文で提案する手法では，特徴点検出法として前節で説明した PF-ASD の様な確率的手法を考える．確率的特徴点検出法では，検出された特徴点位置と共に，その検出精度も得ることが可能であるため，因子分解法を変分ベイズを用いた確率的成分分析法 [83] を用いて拡張し，特徴点位置と検出精度を考慮することで隠れなどにより特徴点をはっきり検出できない場合にも頑健に三次元構造の復元を行うことができると

考えられる．

さらに，ロボットの視覚システム等，実世界の中で環境と相互作用を行いな
がら動作するシステムにおいては，時刻毎に観測される画像から時々刻々と変化
する対象の構造を逐次的に知ることが重要であるが，最も基本的な因子分解法は
バッチ型であり，観測された画像列全にわたって得られた特徴点位置系列から三
次元構造を復元するため，画像の枚数が増加するに従い計算量は著しく増大する
問題があった．

そこで，いくつかの逐次化手法が提案されてきたが [58, 25, 26]，本論文でもオ
ンライン自由エネルギー [71] を逐次最小化して確率的因子分解を行うことより，逐
次的な時系列データからの三次元構造復元を容易に行うことができる．本論文は
以下で構成される．5.1 節で因子分解法について簡単に説明し，5.2 節で確率論的
因子分解法とその逐次解法の導出を行う．5.3 節では剛体の三次元構造復元に関
する数値実験を行い，5.4 節では顔画像の時系列データを用いた実験を行う．最
後に 5.5 節でまとめと今後の課題について述べる．

5.1 因子分解法

因子分解法は，ごく簡単に書けば， N_f 個の特徴点を N_{im} 枚の画像より検出し
て作られたベクトル（軌跡ベクトル） $\mathbf{x}^{(n)} = (x_1, y_1, \dots, x_T, y_T)^T \in \mathcal{R}^{2N_{im}}, n = 1, \dots, N_f$
を並べた観測データ行列 $W = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_f)}) \in \mathcal{R}^{2N_{im} \times N_f}$ を，特徴点の三次元位置
 $\mathbf{s}^{(n)} = (X^{(n)}, Y^{(n)}, Z^{(n)})^T$ を並べた行列 $S = (\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N_f)}) \in \mathcal{R}^{3 \times N_f}$ と回転運動を表す行
列 $M = (\mathbf{m}^{(1)}, \dots, \mathbf{m}^{(3)}) \in \mathcal{R}^{2N_{im} \times 3}$ ，平行移動を表すベクトル \mathbf{t} を用いて，

$$W = MS + \mathbf{t} = \tilde{M}\tilde{S}$$

の様に対象とする物体の運動（回転と平行移動）を示す行列（運動行列） $\tilde{M} \equiv$
 $(\mathbf{m}^{(1)}, \dots, \mathbf{m}^{(4)}) \in \mathcal{R}^{2N_{im} \times 4}$ と， N_f 個の特徴点の相対的な三次元位置関係を示す行列
（形状行列） $\tilde{S} = (\tilde{\mathbf{s}}^{(1)}, \dots, \tilde{\mathbf{s}}^{(N_f)}) \in \mathcal{R}^{4 \times N_f}$ に分解する手法である．ここで， $\tilde{\mathbf{s}}^{(n)} =$
 $(\mathbf{s}^{(n)}, 1)^T = (X^{(n)}, Y^{(n)}, Z^{(n)}, 1)^T$ の同次座標とする．ただし，この分解には自由度が
あるため，カメラモデルに基づく制約を考慮することで，これを一意に決定する．
ここでは最も簡単なカメラモデルに平行投影を仮定した場合について説明する．

真の回転運動行列を M^* , 形状行列を S^* とすると, 先述の通り,

$$\begin{aligned} M^* &= MA \\ S^* &= A^{-1}S \end{aligned}$$

を満たす A が存在する. 例えば, カメラモデルに平行投影を仮定すると, A の満たすべき条件は回転運動行列が正規直行基底をなすことより

$$\begin{aligned} \theta_{t,x}^T AA^T \theta_{t,x} &= 1 \\ \theta_{t,y}^T AA^T \theta_{t,y} &= 1 \\ \theta_{t,x}^T AA^T \theta_{t,y} &= 0 \end{aligned}$$

を満たすように決定すればよい. ただし, $\theta_{t,x}, \theta_{t,y}$ はそれぞれ M^T の第 $(2t-1)$ 列, 第 $2t$ 列である. 詳細, その他のカメラモデルについては [97] が詳しい.

5.2 変分ベイズを用いた確率論的因子分解法

因子分解法は, 数値計算上の安定性やカメラモデルの近似の枠内で十分な精度を持つことから平行投影を仮定する基本的なもののみならず, 様々なカメラモデルに対応した拡張手法が提案されている. しかし, これらの因子分解法では, 特徴点の軌跡ベクトルが全ての画像に渡って欠損無く得られていなければならない, 特徴点の隠れやフレームアウト, 特徴点を画像間で誤って対応付けてしまうことなどが頻繁に考えられる実問題への適用における大きな問題点として広く研究されてきた [76, 45, 79, 103]. これらの手法は特徴点検出法には主眼をおかず, 特徴点の隠れ等大きな雑音で汚されているデータの復元結果に与える影響をロバスト ρ 関数 [45], RANSAC [79] 等を用いて小さく抑えることにより, 頑健な三次元構造復元を行おうとするものである. それに対して本稿で提案する手法は, 特徴点検出法として確率的手法を用いることで, 検出された特徴点位置のみならず, その検出精度も同時に評価し, 確率的に拡張した因子分解法を用いることで, 上記の様な大きな雑音に対しても頑健な三次元構造復元を図る.

これ以降は表記を簡単にするため, $D = 2N_{im}, N_b = 3, \tilde{N}_b = 4$ とする.

確率論的因子分解法

対角行列 $\Sigma_x^{(n)} = \text{diag}((\sigma_{x,1}^{(n)})^2, \dots, (\sigma_{x,D}^{(n)})^2) \in \mathcal{R}^{D \times D}$ を既知のノイズ分散行列として、各特徴点の軌跡ベクトル $\mathbf{x}^{(n)} = (x_1^{(n)}, \dots, x_D^{(n)})^T$ が

$$\mathbf{x}^{(n)} = \tilde{M}\tilde{\mathbf{s}}^{(n)} + \mathbf{v}^{(n)}, \quad \mathbf{v}^{(n)} \sim \mathcal{N}_D(\mathbf{v}^{(n)}|\mathbf{0}, \Sigma_x^{(n)})$$

の形で得られるとすると、

$$p(\mathbf{x}^{(n)}|s^{(n)}, \tilde{M}) = \mathcal{N}_D(\mathbf{x}^{(n)}|\tilde{M}\tilde{\mathbf{s}}^{(n)}, \Sigma_x^{(n)})$$

が得られる。ただし、 $\mathcal{N}_D(\mathbf{x}|\mu, \Sigma)$ は平均 μ 、分散 Σ の D 次元ガウス分布とする。ここで $s^{(n)}$ の事前分布を

$$p(s^{(n)}) = \mathcal{N}_{N_b}(s^{(n)}|\mathbf{0}, I_{N_b})$$

と仮定すれば、ベイズの定理より

$$p(W, S|\tilde{M}) = \prod_{n=1}^{N_f} p(s^{(n)})p(\mathbf{x}^{(n)}|s^{(n)}, \tilde{M})$$

が成り立つ。今、運動行列の自然共役事前分布を

$$p(\tilde{M}) = \prod_{d=1}^D \mathcal{N}_{\tilde{N}_b}(\boldsymbol{\theta}_d|\mathbf{e}_d, \Sigma_\theta)$$

とする。ここで、 $\boldsymbol{\theta}_d = (m_d^{(1)}, \dots, m_d^{(4)})^T \in \mathcal{R}^{\tilde{N}_b}$ 、 $\mathbf{e}_d = (\delta_{1d}, \dots, \delta_{4d})^T \in \mathcal{R}^{\tilde{N}_b}$ 、 $\Sigma_\theta = \gamma^{-1}I_{\tilde{N}_b} \in \mathcal{R}^{\tilde{N}_b \times \tilde{N}_b}$ とし、 $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_D)^T = \tilde{M}$ である。但し、 δ_{ij} はクロネッカのデルタ関数であり、

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

である。この時、試験事後分布

$$q(S, \tilde{M}) = q(S)q(\tilde{M}), \quad q(S) = \prod_{n=1}^{N_f} q_n(s^{(n)})$$

に関する自由エネルギーは以下で書くことができる。

$$\begin{aligned}\mathcal{F}[q] &= \log p(W) - KL(q(S, \tilde{M}) \| p(S, \tilde{M}|W)) \\ &= N_f \mathcal{L} - \mathcal{H}\end{aligned}\quad (4.12)$$

ここで，

$$\begin{aligned}\mathcal{L} &= \frac{1}{N_f} \sum_{n=1}^{N_f} \int q_n(s^{(n)}) q(\tilde{M}) \log \left(\frac{p(\mathbf{x}^{(n)}, s^{(n)} | \tilde{M})}{q_n(s^{(n)})} \right) ds^{(n)} d\tilde{M} \\ \mathcal{H} &= \int q(\tilde{M}) \log \frac{q(\tilde{M})}{p(\tilde{M})} d\tilde{M}\end{aligned}$$

であり，この自由エネルギーの最大化は $q(S, \tilde{M})$ と $p(S, \tilde{M}|W)$ の KL 距離を $q(S), q(\tilde{M})$ に関して最小化することと等価である。

変分ベイズ推定法では，以下の VB-E ステップ，VB-M ステップを繰り返すことで，上記自由エネルギーを最大化する $q(S), q(\tilde{M})$ を求めることができる。なお，以上で得られるのは形状行列 S と運動行列 \tilde{M} のアフィン近似解である。ユークリッド近似解を得るためには前節と同様にカメラモデルに基づいた最適化を行う必要がある。

VB-E ステップ VB-E ステップでは $q_n(s^{(n)})$ を以下により得る。

$$q_n(s^{(n)}) = \mathcal{N}_{N_b}(s^{(n)} | \hat{\mathbf{s}}^{(n)}, \hat{\Sigma}_s^{(n)}) \quad (4.13)$$

ここで，

$$\begin{aligned}\hat{\mathbf{s}}^{(n)} &= \bar{\mathbf{s}}^{(n)} + K^{(n)}(\mathbf{x}^{(n)} - \hat{\mathbf{x}}^{(n)}) \\ \hat{\Sigma}_s^{(n)} &= (I_{N_b} - K^{(n)} \hat{M}) \bar{\Sigma}_s^{(n)} \\ K^{(n)} &= \bar{\Sigma}_s^{(n)} \hat{M} (\hat{\Sigma}_x^{(n)})^{-1} \\ \hat{\mathbf{x}}^{(n)} &= \langle \tilde{M} \rangle_{q_m} \begin{pmatrix} \bar{\mathbf{s}} \\ 1 \end{pmatrix}, \quad \hat{\Sigma}_x^{(n)} = \Sigma_x^{(n)} + \hat{M} \bar{\Sigma}_s^{(n)} \hat{M}^T \\ \bar{\mathbf{s}}^{(n)} &= -(\Xi_{ss}^{(n)})^{-1} \Xi_{s1}^{(n)}, \quad \bar{\Sigma}_s = (\Xi_{ss}^{(n)})^{-1} \\ (\hat{V}^{(n)})^{-1} &= \sum_{d=1}^D (\sigma_{x,d}^{(n)})^{-2} \hat{V}_d^{-1}\end{aligned}$$

である。ただし, $\Xi_{ss}^{(n)} \in \mathcal{R}^{N_f \times N_f}$, $\Xi_{s1}^{(n)} \in \mathcal{R}^{N_f}$ は, $\Xi^{(n)} \in \mathcal{R}^{\tilde{N}_f \times \tilde{N}_f}$ を以下の様に分割して得られる行列とベクトルである。

$$\Xi^{(n)} = \tilde{I}_{\tilde{N}_b} + (\hat{V}^{(n)})^{-1} = \left(\begin{array}{c|c} \Xi_{ss}^{(n)} & \Xi_{s1}^{(n)} \\ \hline \Xi_{1s}^{(n)} & \Xi_{11}^{(n)} \end{array} \right)$$

VB-M ステップ VB-M ステップでは $q(\tilde{M})$ を以下により得る。

$$q(\tilde{M}) = \prod_{d=1}^D \mathcal{N}_{\tilde{N}_b}(\boldsymbol{\theta}_d | \hat{\boldsymbol{\theta}}_d, \hat{V}_d^{-1}) \quad (4.14)$$

ここで

$$\hat{\boldsymbol{\theta}}_d^T = \left(\mathbf{e}_d^T \Sigma_\theta^{-1} + \sum_{n=1}^{N_f} (\sigma_{x,d}^{(n)})^{-2} \langle x_d^{(n)} (\tilde{\mathbf{s}}^{(n)})^T \rangle_{q_s} \right) \hat{V}_d^{-1}$$

$$\hat{V}_d = \Sigma_\theta^{-1} + \sum_{n=1}^{N_f} (\sigma_{x,d}^{(n)})^{-2} \langle \tilde{\mathbf{s}}^{(n)} (\tilde{\mathbf{s}}^{(n)})^T \rangle_{q_s}$$

である。ここで用いられる統計量はそれぞれ

$$\langle x_d^{(n)} (\tilde{\mathbf{s}}^{(n)})^T \rangle_{q_s} = x_d^{(n)} ((\hat{\mathbf{s}}^{(n)})^T, 1)$$

$$\langle \tilde{\mathbf{s}}^{(n)} (\tilde{\mathbf{s}}^{(n)})^T \rangle_{q_s} = \begin{pmatrix} \hat{\mathbf{s}}^{(n)} (\hat{\mathbf{s}}^{(n)})^T + \Sigma_s^{(n)} & \hat{\mathbf{s}}^{(n)} \\ (\hat{\mathbf{s}}^{(n)})^T & 1 \end{pmatrix}$$

であり, ここで $\langle f(\mathbf{a}) \rangle_{q_b}$ は $f(\mathbf{a})$ の $q(\mathbf{b})$ に関する期待値である。この時, 運動行列と形状行列の期待値は, 式 (4.13), (4.14) より

$$\langle S \rangle_{q_s} = (\hat{\mathbf{s}}^{(1)}, \dots, \hat{\mathbf{s}}^{(N_f)})$$

$$\langle \tilde{M} \rangle_{q_m} = (\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_D)^T$$

で得ることができる。

VB法を用いた確率的因子分解法の実行手順を図 4.17, 4.18 に, 自由エネルギーの計算法を図 4.19 に, 導出の詳細に関しては付録 F に示す。

三次元形状行列: $S \equiv (s^{(1)}, \dots, s^{(N_f)}) \in \mathcal{R}^{3 \times N_f}$, $s^{(n)} = (X^{(n)}, Y^{(n)}, Z^{(n)})^T \in \mathcal{R}^3$,

特徴点数: N_f , 同次形状行列: $\tilde{S} \equiv \begin{pmatrix} s^{(1)} & \dots & s^{(N_f)} \\ 1 & \dots & 1 \end{pmatrix}$

運動行列: $\tilde{M} \equiv (\mathbf{m}_1, \dots, \mathbf{m}_{\tilde{N}_d}) \in \mathcal{R}^{D \times \tilde{N}_d}$, $M = (\mathbf{m}_1, \dots, \mathbf{m}_{\tilde{N}_d}) \in \mathcal{R}^{D \times 3}$, \mathbf{m}_i : 基底ベクトル

軌跡ベクトル: $\mathbf{x}_{1:T}^{(n)} \equiv (x_1^{(n)}, y_1^{(n)}, \dots, x_T^{(n)}, y_T^{(n)})^T \in \mathcal{R}^D$, $D = 2T$

軌跡行列: $P_x \equiv (\mathbf{x}_{1:T}^{(1)}, \dots, \mathbf{x}_{1:T}^{(N_f)}) \in \mathcal{R}^{D \times N_f}$ (4.15)

とする. 今, 各特徴点の軌跡ベクトル $\mathbf{x}_{1:T}^{(n)}$ は対角行列 $\Sigma_x^{(n)}$ を既知の分散共分散行列として

$$\mathbf{x}_{1:T}^{(n)} = \tilde{M}\tilde{s}^{(n)} + \mathbf{v}^{(n)}, \quad \mathbf{v}^{(n)} \sim \mathcal{N}_D(\mathbf{v}^{(n)} | \mathbf{0}, \Sigma_x^{(n)})$$

で生成されるとすると, 以下の VB-E ステップと VB-M ステップを繰り返すことにより運動行列 M と三次元形状行列 S の推定を行うことが可能である.

初期化 パラメータ Θ を運動行列 \tilde{M} として, 初期値と事前分布を以下の様に設定する.

$$\Theta = \tilde{M} = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \in \mathcal{R}^{D \times \tilde{N}_b}$$

$$P(\Theta) = \prod_{d=1}^D \mathcal{N}_{\tilde{N}_b}(\boldsymbol{\theta}_d | \mathbf{e}_d, \gamma^{-1} I_{\tilde{N}_b})$$

$$\mathbf{e}_d = (\delta_{d1}, \dots, \delta_{d\tilde{N}_b}) \in \mathcal{R}^{\tilde{N}_b}, \boldsymbol{\theta}_d = (\mathbf{m}_{1d}, \dots, \mathbf{m}_{\tilde{N}_d}) \in \mathcal{R}^{\tilde{N}_b}$$

図 4.17 確率的因子分解法の実行手順

VB-E ステップ

$$\begin{aligned}
 q_n(\mathbf{s}^{(n)}) &= \mathcal{N}_{N_b}(\mathbf{s}^{(n)} | \hat{\mathbf{s}}^{(n)}, \hat{\Sigma}_s^{(n)}) \\
 \langle \tilde{\mathbf{s}}^{(n)} \tilde{\mathbf{s}}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} &= \begin{pmatrix} \hat{\mathbf{s}}^{(n)} \hat{\mathbf{s}}^{(n)T} + \hat{\Sigma}_s^{(n)} & \hat{\mathbf{s}}^{(n)} \\ \hat{\mathbf{s}}^{(n)T} & 1 \end{pmatrix} \\
 \hat{\mathbf{s}}^{(n)} &= \bar{\mathbf{s}}^{(n)} + K^{(n)}(\mathbf{x}^{(n)} - \hat{\mathbf{x}}^{(n)}) \\
 \hat{\Sigma}_s^{(n)} &= (I_{N_b} - K^{(n)} \hat{M}) \bar{\Sigma}_s^{(n)} \\
 K^{(n)} &= \bar{\Sigma}_s^{(n)} \hat{M} \hat{\Sigma}_x^{(n)-1} \\
 \hat{\mathbf{x}}^{(n)} &= \hat{\Theta} \begin{pmatrix} \bar{\mathbf{s}} \\ 1 \end{pmatrix}, \quad \hat{\Sigma}_x^{(n)} = \Sigma_x^{(n)} + \hat{M} \bar{\Sigma}_s^{(n)} \hat{M}^T \\
 \bar{\mathbf{s}}^{(n)} &= -\bar{\Xi}_{ss}^{(n)-1} \bar{\Xi}_{s1}^{(n)}, \quad \bar{\Sigma}_s = \bar{\Xi}_{ss}^{(n)-1}, \quad \bar{\Xi} = \tilde{I}_{N_b} + \hat{V}^{(n)-1} \\
 \hat{V}^{(n)-1} &= \sum_{d=1}^D \sigma_{x,d}^{(n)-2} \hat{V}_d^{-1}
 \end{aligned}$$

VB-M ステップ

$$\begin{aligned}
 q(\Theta) &= \prod_{d=1}^D \mathcal{N}_{N_b}(\theta_d | \hat{\theta}_d, \hat{V}_d^{-1}) \\
 \hat{\theta}_d^T &= \left(\mathbf{e}_d^T \Sigma_\theta^{-1} + \sum_{n=1}^{N_f} \sigma_{x,d}^{(n)-2} \langle x_d^{(n)} \tilde{\mathbf{s}}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} \right) \hat{V}_d^{-1} \\
 \hat{V}_d &= \Sigma_\theta^{-1} + \sum_{n=1}^{N_f} \sigma_{x,d}^{(n)-2} \langle \tilde{\mathbf{s}}^{(n)} \tilde{\mathbf{s}}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})}
 \end{aligned}$$

図 4.18 確率的因子分解法の実行手順 (続き)

自由エネルギー 自由エネルギー \mathcal{F} は以下で計算することができる。

$$\mathcal{F} = N_f \mathcal{L} - \mathcal{H}$$

$$\mathcal{L} = \mathcal{L}_{xs} - \mathcal{H}_s$$

$$\begin{aligned} \mathcal{L}_{xs} = & -\frac{N_b + D}{2} \log(2\pi) + \frac{1}{N_f} \sum_{n=1}^{N_f} \left\{ \frac{1}{2} \langle \mathbf{s}^{(n)T} \mathbf{s}^{(n)} \rangle_{q_n(\mathbf{s}^{(n)})} - \frac{1}{2} \log |\Sigma_x^{(n)}| \right. \\ & - \frac{1}{2} \text{Tr} \left[\langle \mathbf{x}^{(n)} \mathbf{x}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} \Sigma_x^{(n)-1} \right] + \text{Tr} \left[\langle \tilde{\mathbf{s}}^{(n)} \mathbf{x}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} \Sigma_x^{(n)-1} \langle \tilde{\mathbf{M}} \rangle_{q(\Theta)} \right] \\ & \left. - \frac{1}{2} \text{Tr} \left[\langle \tilde{\mathbf{s}}^{(n)} \tilde{\mathbf{s}}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} \langle \tilde{\mathbf{M}}^T \Sigma_x^{(n)-1} \tilde{\mathbf{M}} \rangle_{q(\Theta)} \right] \right\} \end{aligned}$$

$$\mathcal{H}_s = -\frac{N_b}{2} (1 + \log(2\pi)) - \frac{1}{N_f} \sum_{n=1}^{N_f} \frac{1}{2} \log |\hat{\Sigma}_s^{(n)}|$$

$$\begin{aligned} \mathcal{H} = & \frac{\gamma}{2} \text{Tr} \left[(\hat{\Theta} - E)^T (\hat{\Theta} - E) \right] - \frac{D \tilde{N}_b}{2} (1 + \log \gamma) \\ & + \frac{1}{2} \sum_{d=1}^D \log |\hat{V}_d| + \frac{\gamma}{2} \sum_{d=1}^D \text{Tr} \left[\hat{V}_d^{-1} \right] \end{aligned}$$

ここで用いた期待値はそれぞれ

$$\langle \mathbf{s}^{(n)T} \mathbf{s}^{(n)} \rangle_{q_n(\mathbf{s}^{(n)})} = \langle \mathbf{s}^{(n)} \rangle_{q_n(\mathbf{s}^{(n)})}^T \langle \mathbf{s}^{(n)} \rangle_{q_n(\mathbf{s}^{(n)})} = \hat{\mathbf{s}}^{(n)T} \hat{\mathbf{s}}^{(n)}$$

$$\langle \mathbf{x}^{(n)} \mathbf{x}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} = \mathbf{x}^{(n)} \mathbf{x}^{(n)T}$$

$$\langle \tilde{\mathbf{s}}^{(n)} \mathbf{x}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} = \langle \tilde{\mathbf{s}}^{(n)} \rangle_{q_n(\mathbf{s}^{(n)})} \mathbf{x}^{(n)T}$$

$$\langle \tilde{\mathbf{s}}^{(n)} \tilde{\mathbf{s}}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} = \begin{pmatrix} \hat{\mathbf{s}}^{(n)} \hat{\mathbf{s}}^{(n)T} + \hat{\Sigma}_s^{(n)} & \hat{\mathbf{s}}^{(n)} \\ \hat{\mathbf{s}}^{(n)T} & 1 \end{pmatrix}$$

$$\langle \tilde{\mathbf{M}} \rangle_{q(\Theta)} = \hat{\Theta}$$

$$\langle \tilde{\mathbf{M}}^T \Sigma_x^{(n)-1} \tilde{\mathbf{M}} \rangle_{q(\Theta)} = \hat{\Theta}^T \Sigma_x^{(n)-1} \hat{\Theta} + \hat{V}^{(n)-1}$$

図 4.19 自由エネルギーの計算手順

確率的因子分解法のオンライン化

前節で説明した確率的因子分解法は全てのデータが得られた後に推定を行うバッチ手法であった。しかし、現実には、計算資源の制限等により、長いデータ系列に対してバッチ処理を行うことは困難である。本節では、オンライン自由エネルギーの逐次最大化により、再帰的な復元結果の更新に加え、徐々に形状を変化させるような対象についても同様に扱うことが可能であることを示す。

時刻 t におけるオンライン自由エネルギー \mathcal{F}_t^λ は、過去の不正確な推定による影響を徐々に忘却するための忘却係数を $\lambda_s (0 \leq \lambda_s \leq 1)$ 、忘却係数の影響を正規化するための係数を $\eta_t = \left(\sum_{\tau=1}^t \Lambda(\tau, t)\right)^{-1}$ 、 $\Lambda(\tau, t) = \prod_{s=\tau+1}^t \lambda_s$ として、前節のバッチ自由エネルギー(式(4.12))を以下の様に修正することで得ることができる。

$$\begin{aligned}\mathcal{F}_t^\lambda[q] &= \eta_t \sum_{\tau=1}^t \Lambda(\tau, t) \left\{ \log p(W_\tau) - KL(q(S_\tau, \tilde{M}_\tau) \| p(S_\tau, \tilde{M}_\tau | W_\tau)) \right\} \\ &= N_f \mathcal{L}_t^\lambda - \mathcal{H}_t^\lambda\end{aligned}\quad (4.16)$$

ただし、

$$\begin{aligned}\mathcal{L}_t^\lambda &= \eta_t \sum_{\tau=1}^t \Lambda(\tau, t) \left\{ \frac{1}{N_f} \sum_{n=1}^{N_f} \int q_n(s_\tau^{(n)}) q(\tilde{M}_\tau) \log \left(\frac{p(\mathbf{x}_\tau^{(n)}, s_\tau^{(n)} | \tilde{M}_\tau)}{q_n(s_\tau^{(n)})} \right) ds_\tau^{(n)} d\tilde{M}_\tau \right\} \\ \mathcal{H}_t^\lambda &= \eta_t \sum_{\tau=1}^t \Lambda(\tau, t) \left\{ \int q(\tilde{M}_\tau) \log \frac{q(\tilde{M}_\tau)}{p(\tilde{M}_\tau)} d\tilde{M}_\tau \right\}\end{aligned}$$

とする。ここで、 $W_t = (\mathbf{x}_t^{(1)}, \dots, \mathbf{x}_t^{(N_f)}) \in \mathcal{R}^{2 \times N_f}$ は時刻 t において得られる特徴点の画像面上での位置を並べたもので、 $\mathbf{x}_t^{(n)} = (x_t^{(n)}, y_t^{(n)})^T = (x_{1,t}^{(n)}, x_{2,t}^{(n)})^T \in \mathcal{R}^2$ である。また、 $\tilde{M}_t = (\theta_t^{(1)}, \dots, \theta_t^{(N_b)}) \in \mathcal{R}^{2 \times N_b}$ は特徴点 $\tilde{s}^{(n)}$ を画像面上の点 $\mathbf{x}_t^{(n)} = \tilde{M}_t \tilde{s}^{(n)}$ に射影するカメラ行列である。このオンライン自由エネルギー \mathcal{F}_t^λ の逐次最大化は、前節で説明した VB-E ステップ、VB-M ステップと同様の以下の手順で実行可能である。

オンライン VB-E ステップ バッチ VB-E ステップと同様に、ここでは $q_n(s_t^{(n)})$ を以下により得る。

$$q_n(s_t^{(n)}) = \mathcal{N}_{N_b}(s_t^{(n)} | \hat{s}_t^{(n)}, \hat{\Sigma}_{s,t}^{(n)}) \quad (4.17)$$

ここで ,

$$\begin{aligned}
\hat{\mathbf{s}}_t^{(n)} &= \bar{\mathbf{s}}_t^{(n)} + K_t^{(n)}(\mathbf{x}_t^{(n)} - \hat{\mathbf{x}}_t^{(n)}) \\
\hat{\Sigma}_{s,t}^{(n)} &= (I_{N_b} - K_t^{(n)}\hat{M}_t)\bar{\Sigma}_{s,t}^{(n)} \\
K_t^{(n)} &= \bar{\Sigma}_{s,t}^{(n)}\hat{M}_t(\hat{\Sigma}_{x,t}^{(n)})^{-1} \\
\hat{\mathbf{x}}_t^{(n)} &= \langle\langle \tilde{M}_t \rangle\rangle_t \begin{pmatrix} \bar{\mathbf{s}}_t \\ 1 \end{pmatrix}, \quad \hat{\Sigma}_{x,t}^{(n)} = \Sigma_{x,t}^{(n)} + \hat{M}_t\bar{\Sigma}_{s,t}^{(n)}\hat{M}_t^T \\
\bar{\mathbf{s}}_t^{(n)} &= -(\Xi_{ss,t}^{(n)})^{-1}\Xi_{s1,t}^{(n)}, \quad \bar{\Sigma}_{s,t}^{(n)} = (\Xi_{ss,t}^{(n)})^{-1}, \\
\Xi_t &= \tilde{I}_{\tilde{N}_b} + (\hat{V}_t^{(n)})^{-1} \\
(\hat{V}_t^{(n)})^{-1} &= \sum_{d=1}^2 (\sigma_{x,d,t}^{(n)})^{-2} \hat{V}_{d,t}^{-1}
\end{aligned}$$

オンライン VB-M ステップ オンライン VB-M ステップでは $q(\tilde{M}_t)$ を以下により得る .

$$q(\tilde{M}_t) = \prod_{d=1}^2 \mathcal{N}_{\tilde{N}_b}(\boldsymbol{\theta}_d | \hat{\boldsymbol{\theta}}_{d,t}, \hat{V}_{d,t}^{-1}) \quad (4.18)$$

ここで

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{d,t}^T &= \left(\mathbf{e}_d^T \Sigma_{\theta}^{-1} + \sum_{n=1}^{N_f} (\sigma_{x,d,t}^{(n)})^{-2} \langle\langle \mathbf{x}_{d,t}^{(n)} (\tilde{\mathbf{s}}_t^{(n)})^T \rangle\rangle_t \right) \hat{V}_{d,t}^{-1} \\
\hat{V}_{d,t} &= \Sigma_{\theta}^{-1} + \sum_{n=1}^{N_f} (\sigma_{x,d,t}^{(n)})^{-2} \langle\langle \tilde{\mathbf{s}}_t^{(n)} (\tilde{\mathbf{s}}_t^{(n)})^T \rangle\rangle_t
\end{aligned}$$

であり , ここで用いられる統計量はそれぞれ

$$\begin{aligned}
\langle\langle \mathbf{x}_{d,t}^{(n)} (\tilde{\mathbf{s}}_t^{(n)})^T \rangle\rangle_t &= \eta_t \sum_{\tau=1}^t \Lambda(\tau, t) \mathbf{x}_{d,\tau}^{(n)} (\tilde{\mathbf{s}}_{\tau}^{(n)})^T \\
\langle\langle \tilde{\mathbf{s}}_{\tau}^{(n)} (\tilde{\mathbf{s}}_{\tau}^{(n)})^T \rangle\rangle_t &= \eta_t \sum_{\tau=1}^t \Lambda(\tau, t) \begin{pmatrix} \hat{\mathbf{s}}_{\tau}^{(n)} (\hat{\mathbf{s}}_{\tau}^{(n)})^T + \hat{\Sigma}_{s,\tau}^{(n)} & \hat{\mathbf{s}}_{\tau}^{(n)} \\ (\hat{\mathbf{s}}_{\tau}^{(n)})^T & 1 \end{pmatrix}
\end{aligned}$$

で得られる．特徴点の三次元位置と運動行列の時刻 t における期待値はそれぞれ

$$\begin{aligned}\langle\langle\hat{\mathbf{s}}_\tau^{(n)}\rangle\rangle_t &= \eta_t \sum_{\tau=1}^t \Lambda(\tau, t) \hat{\mathbf{s}}_\tau^{(n)} \\ \langle\langle\tilde{M}_t\rangle\rangle_t &= \left(\langle\langle\hat{\boldsymbol{\theta}}_{1,\tau}\rangle\rangle_t, \langle\langle\hat{\boldsymbol{\theta}}_{2,\tau}\rangle\rangle_t \right)^T \\ \langle\langle\hat{\boldsymbol{\theta}}_{d,\tau}\rangle\rangle_t &= \eta_t \sum_{\tau=1}^t \Lambda(\tau, t) \hat{\boldsymbol{\theta}}_{d,\tau}\end{aligned}$$

これらの統計量は，以下の形で再帰的に計算することが可能である．

$$\begin{aligned}\langle\langle x_{d,\tau}^{(n)} (\tilde{\mathbf{s}}_\tau^{(n)})^T \rangle\rangle_t &= (1 - \eta_t) \langle\langle x_{d,\tau}^{(n)} (\tilde{\mathbf{s}}_\tau^{(n)})^T \rangle\rangle_{t-1} + \eta_t x_{d,t}^{(n)} ((\hat{\mathbf{s}}_t^{(n)})^T, 1) \\ \langle\langle \tilde{\mathbf{s}}_\tau^{(n)} (\tilde{\mathbf{s}}_\tau^{(n)})^T \rangle\rangle_t &= (1 - \eta_t) \langle\langle \tilde{\mathbf{s}}_\tau^{(n)} (\tilde{\mathbf{s}}_\tau^{(n)})^T \rangle\rangle_{t-1} + \eta_t \begin{pmatrix} \hat{\mathbf{s}}_\tau^{(n)} (\hat{\mathbf{s}}_\tau^{(n)})^T + \hat{\Sigma}_{s,\tau}^{(n)} & \hat{\mathbf{s}}_\tau^{(n)} \\ (\hat{\mathbf{s}}_\tau^{(n)})^T & 1 \end{pmatrix} \\ \langle\langle \hat{\mathbf{s}}_\tau^{(n)} \rangle\rangle_t &= (1 - \eta_t) \langle\langle \hat{\mathbf{s}}_\tau^{(n)} \rangle\rangle_{t-1} + \eta_t \hat{\mathbf{s}}_t^{(n)} \\ \langle\langle \hat{\boldsymbol{\theta}}_{d,\tau} \rangle\rangle_t &= (1 - \eta_t) \langle\langle \hat{\boldsymbol{\theta}}_{d,\tau} \rangle\rangle_{t-1} + \eta_t \hat{\boldsymbol{\theta}}_t\end{aligned}$$

5.3 数値実験

提案手法の有効性を確認するために，図 4.5 に示すような，画像列からの剛体の三次元構造復元問題に提案手法を適用する．本手法では，画像面上での特徴点の位置の他に，その観測ノイズ分散も必要となるため，近年盛んに研究されている統計的特徴点追跡手法（例えば，[8, 35, 43] など）を特徴点追跡手法として採用し，特徴点の推定位置を観測値，特徴点位置の推定分散を観測ノイズ分散として扱う．ここでは，提案手法による純粋な三次元構造の復元能力を測るため，特徴点位置の推定値と推定分散 $\mathbf{x}_t^{(n)}, \Sigma_{x,t}^{(n)}$ は既知とする．図 4.5 に示すような剛体の頂点と平面の中心計 14 点を特徴点として採用し，時刻 t における特徴点の真の位置を \mathbf{y}_t ，観測ノイズを $\mathbf{v}_t \sim \mathcal{N}(\mathbf{v}_t | \mathbf{0}, \Sigma_{x,t}^{(n)})$ として

$$\mathbf{x}_t^{(n)} = \mathbf{y}_t^{(n)} + \mathbf{v}_t$$

により特徴点の推定値を得た．推定分散は特徴点が見えている時には 50 とし，それ以外は 1 と設定した．また，画像列の長さは 360 フレームとし，剛体を鉛直方

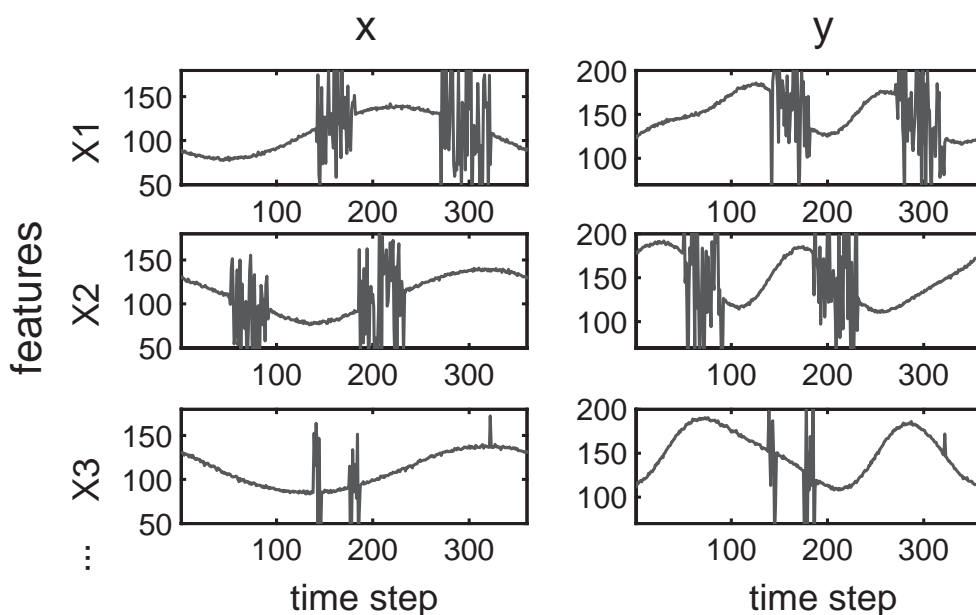


図 4.20 推定された特徴点位置の時系列データ例．特徴点が隠れにより観測されない場合には大きな観測ノイズが付加される．

向に $1[\text{deg/frame}]$ ，水平方向に $2[\text{deg/frame}]$ で等速で回転させた．以上により得られた特徴点の画像面上での位置の時系列例を図 4.20 に示す．この様に大きな雑音で汚された特徴点時系列データに確率的逐次因子分解法を適用する．構造復元の際に推定が不安定になるのを避けるため，ここでは窓枠を 90 フレームに設定し，1 時刻に前後 45 フレーム分のデータが観測されるとして，構造復元を行った．また，特徴点が隠れる期間が長くなれば，特徴点の推定精度も悪化するため，各時刻でオンライン自由エネルギーを計算し，自由エネルギーが最大となる時のみ，構造の推定値を更新した．三次元構造の推定結果を図 4.21-4.23 に示す．なお，ここではアフィン近似解を真の特徴点位置との二乗誤差が最小になる様にアフィン変換したものを表示している．通常の因子分解法では隠れによる大きな雑音の影響を受け，三次元構造の復元精度が大きく低下しているのに対して，確率的（逐次）因子分解法を用いた場合には，特に逐次手法において平面の中心に関しては多少の誤差を含むものの，隠れの影響による大きな雑音で汚された特徴点の系列にも関わらず，頑健な三次元構造復元を行うことができている．真の三次元構造

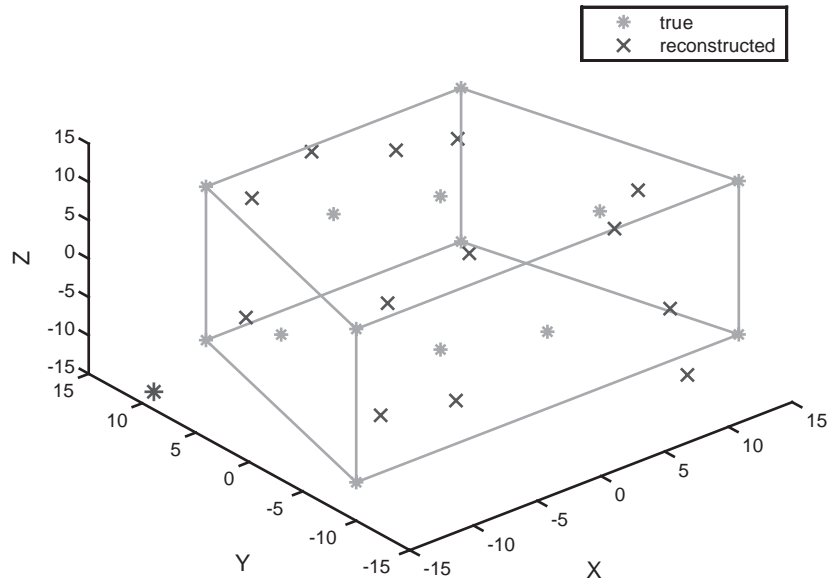


図 4.21 因子分解法により復元された三次元構造

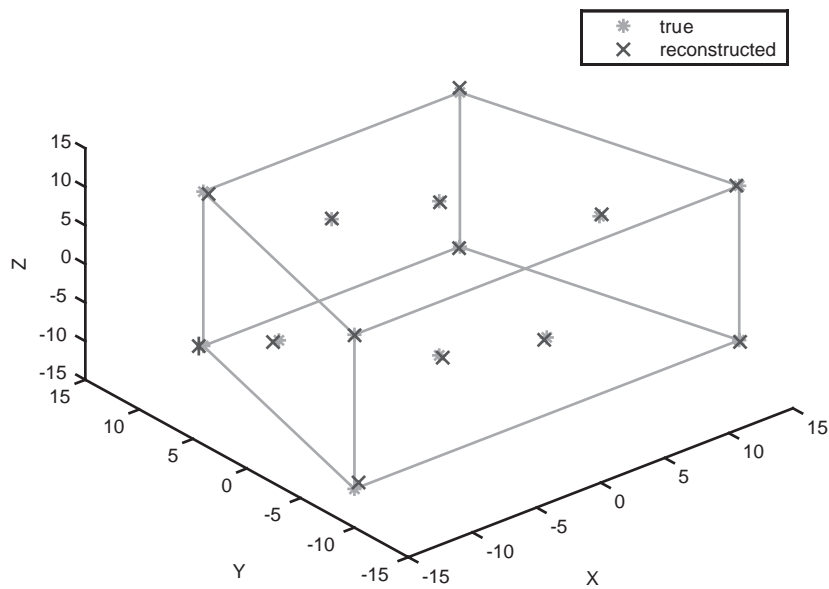


図 4.22 確率的因子分解法により復元された三次元構造

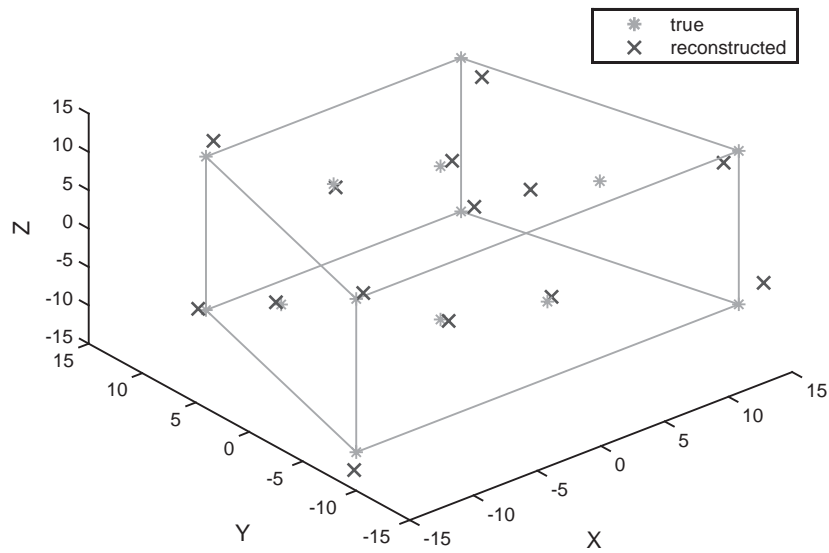


図 4.23 確率的逐次因子分解法により復元された三次元構造

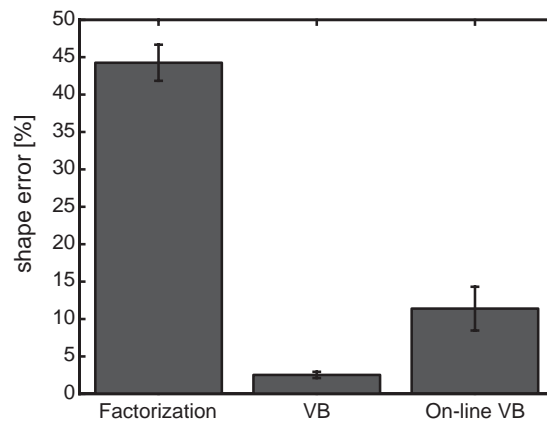


図 4.24 因子分解法 (Factorization), 確率的因子分解法 (VB), 確率的逐次因子分解法 (On-line VB) による形状推定誤差の 50 試行の平均値と分散

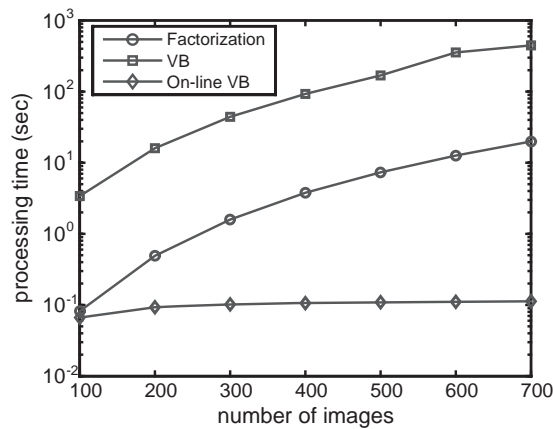


図 4.25 因子分解法，確率的因子分解法，確率的逐次因子分解法の計算時間の比較．確率的逐次因子分解法に関しては 1 フレーム辺りの処理時間を示す．

を S^* ，推定された三次元構造を \hat{S} として，推定誤差を

$$Shape\ Error = \frac{\|S^* - \hat{S}\|}{\|S^*\|} \quad (4.19)$$

とする時，因子分解法 (Factorization)，確率的因子分解法 (VB)，確率的逐次因子分解法 (On-line VB) の推定誤差の 50 試行の平均値と分散を図 4.24 に示す．確率的因子分解法は繰り返し手法であるため，自由エネルギーが収束するまで前述の VB-E ステップ，VB-M ステップを繰り返す必要があるが，ここでは 10 回の繰り返いで十分収束した．また，画像列の長さに応じた処理時間の 50 試行の平均値を図 4.25 右に示す．因子分解法を変分ベイズ法により確率的手法へと拡張し，特徴点の観測値を観測ノイズ分散と共に評価することで，大きな雑音を含む特徴点のデータ系列からも，頑健な推定を行うことが可能である．更に，確率的因子分解法のオンライン化を行う事で，精度を保ったまま 1 フレーム辺りの処理時間を軽減し，実時間処理を可能とすることができた．特徴点位置の系列データを全て観測してから処理を行うバッチ処理確率的因子分解法に比べて逐次処理を行う手法が推定誤差が増加するのは，全ての特徴点が同時に観測する時刻がほとんど無く（特に，平面の中心については全ての特徴点が同時に観測される時刻は存在しない），逐次的に処理を行う事で，ある時刻で観測されている特徴点の推定精度は上昇するが，観測できなかった特徴点の精度は次第に悪化していくためである．

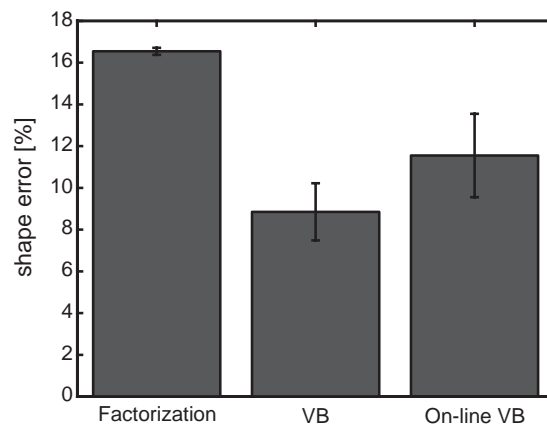


図 4.26 因子分解法 (Factorization), 確率的因子分解法 (VB), 確率的逐次因子分解法 (On-line VB) による三次元顔の推定誤差の平均値と分散 (50 試行)。

5.4 実画像を用いた実験

続いて, 両目中心, 目頭, 目尻, 鼻, 口尻, 眉の内端の計 11 点を顔特徴点として, 図 4.10 に示すような実際の人物の頭部画像からこれらの特徴点の三次元位置を復元する実験を行った。本実験では純粹に三次元構造復元の頑健性を比較するために以下の手順で観測値と観測ノイズ分散を得ることにより, 特徴点追跡処理は評価の対象外とした。

まず学習用の顔特徴点画像列 (上下・左右・左右回転運動を含む) から主成分分析により各顔特徴点の平均画像と基底画像, それぞれの基底画像に対応した固有値を得て, テスト用の画像列の顔特徴点位置周りで獲得された特徴点画像の生成確率を計算し (詳細は [83] 参照), この特徴点画像の生成確率の逆数を特徴点の観測ノイズ分散とした。なお, 各顔特徴点の真の位置は画像列より人の手で取得し, 真の特徴点位置を先に得た観測ノイズ分散を持つガウス雑音を加えて観測値とした。

以上で得られた特徴点データ系列から三次元顔の復元を行い, 数値実験と同様に式 (4.19) より求めた実際に計測された三次元形状との誤差の 50 試行についての平均と分散を図 4.26 に示す。また, 確率的逐次因子分解法により推定された三次元顔モデルを図 4.27 に示す。ここで行った実験は, 前節の数値実験と比べて特

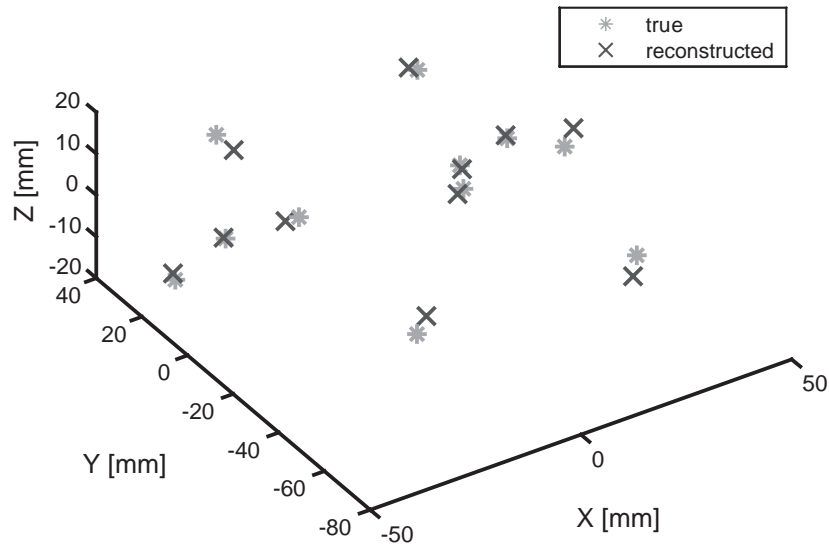


図 4.27 確率的逐次因子分解法により復元された三次元構造 .

微点で大きな隠れの存在は無いが，各フレームで観測されるノイズの分散が非定常に変化するため，確率的因子分解法では数値実験と同様の精度で復元を行うことができているが，通常の因子分解法では確率的因子分解法と比べて誤差が上昇している .

5.5 まとめ

本節では，画像面上における特徴点位置のデータ系列から対象の三次元構造と運動を復元することのできる因子分解法を，確率的手法を用いて拡張した確率的因子分解手法と，更に逐次処理化を行った確率的逐次因子分解法の説明を行った . 確率的に拡張した因子分解法を適応することにより，特徴点の隠れや誤追跡による大きな雑音の影響を低く抑え，剛体の三次元構造復元の数値実験と実際の顔画像からの三次元顔の構造復元実験において頑健な構造復元を行うことができた . また，確率的因子分解法を逐次手法へと修正することで，観測される画像枚数に従って大きく増加していた計算時間を，推定精度を余り低下させることなく削減することができた .

ここで示した確率的（逐次）因子分解法を用いれば、画像面上で推定した各顔特徴点位置の期待値とその分散から、追跡対象とした人の三次元顔モデルを確率的に得ることが可能である。ただし、逐次手法においては、確率的因子分解法を逐次手法へと修正した影響で、特徴点が長時間観測されない場合には復元精度の低下が生じている。この原因は、各特徴点で同一の忘却係数を用いてオンライン化を行ったため、隠れていて観測情報の精度が極端に低下している特徴点の影響を他の特徴点も同様に受けているためであると考えられる。こうした観測されない特徴点の推定精度低下が他の特徴点にも伝播するのを防ぐためには、忘却係数の制御 [33] を行う必要があると考えられる。つまり、観測情報が信頼できる場合には適応的に忘却係数を上げ、観測を重視して精度良い推定を行い、隠れなどにより観測情報の信頼度が低下している場合には適応的に忘却係数を下げ、精度の悪い観測情報を利用した学習の進行を抑える枠組みを考える必要がある。こうした逐次手法の欠点を改善することができれば、初期には平均顔モデルを利用して顔特徴点追跡と頭部姿勢推定を行う頭部姿勢推定システムにおいて、オンライン自由エネルギー等を指標とした平均顔モデルから推定したドライバ個人の三次元顔モデルへのスムーズな切り替えが実現できると考えられる。この点については今後の課題としたい。

第5章 議論

第3章ではサンプリング分布の動的な切り替えにより，実環境において実時間状態推定を少数のサンプルにも関わらず推定精度を保った頑健な推定を行う手法について述べ，第4章ではより複雑な高次元状態の階層的推定法について述べた．本章では，実環境における頑健な実時間状態推定を行うための，サンプリング分布，対象のモデルに関する一般的な議論を行う．

サンプリング分布

本論文でも述べてきたように，実環境での状態推定問題では，例えば視覚追跡における照明条件の変動や遮蔽，障害物の存在など，観測情報が大きな非ガウス性雑音で汚される状況が多々起りうる．第3章では，こうした実環境における実時間状態推定問題に，時系列ベイズ推定の近似解法であるパーティクルフィルタを適用する際に，推定精度を維持して，頑健な状態推定を行うための手法について述べた．

一般的にパーティクルフィルタで状態推定を行う際には，対象のダイナミクスモデルと観測モデルが必要である．パーティクルフィルタは非線形過程も扱うことが可能であるため，理想的なダイナミクスモデルが得られている場合には，それを用いることが出来る．しかし，実環境における実時間視覚追跡の様な課題において，追跡対象のダイナミクスが予め完全に既知であることは稀であり，多くの場合は線形のダイナミクスを仮定することが多い [35]．画像列から追跡対象のダイナミクスを学習することも試みられている [5, 62, 63, 67] が，バッチ学習であったり，オンラインであっても学習初期の精度の低いダイナミクスモデルを用いることで状態推定の精度自体が悪化することも多々ありえる．有限のサンプルによる分布の近似を行うパーティクルフィルタにおいて，ダイナミクスモデル由

来のサンプリング分布の精度低下の影響でサンプルの重みの分散が増大し、推定に寄与するサンプルとほとんど寄与しないサンプルが混在することで、分布の近似精度が低下し、ひいては全体の推定精度の低下を引き起こす問題は外れ値問題と呼ばれ、ダイナミクスモデルをサンプリング分布とするパーティクルフィルタ [28, 35] の大きな問題点として様々な改善手法が提案されてきた [86, 41] .

改善手法の多くはサンプリング分布に現時刻の観測モデルを導入することでサンプルの持つ重みの分散を小さくし、推定精度の維持を図る手法であり、サンプルが十分に多い場合にはその有効性が示されている [2] . 例えば、Auxiliary Particle Filter [41] (APF) はダイナミクスモデルの代表点での観測情報を用いて 1 ステップのスムージングに類似する処理を行うことで、サンプリング分布に観測情報を導入し、より大きな重みを持つサンプルのみを生成する手法であるし、Unscented Particle Filter [86] (UPF) は Unscented Kalman Filter [39] (UKF) を用いて、サンプリング分布の推定を行うことでサンプリング分布に観測情報を導入する。これらの手法は、非常に精度の良い観測モデルを用いることが出来る場合には有効であるが、実環境における実時間視覚追跡の様な課題においては、正確なダイナミクスモデルを用いることが困難であることと同様に、完全な観測モデルを用いることもまた困難である。照明条件の変化や、遮蔽/障害物の存在、追跡対象の形状変化、背景の変化など、観測モデルを構築する際に考慮すべき要因は多数あり、その全てをモデルに組み入れることは不可能に近い。このことが実問題における観測情報の非定常な精度低下を引き起こす大きな要因であり、そうした場合に観測情報を重視したサンプリング分布からのサンプルを用いて状態推定を行うことは推定精度を大きく低下させる。

第 3 章では、これを防ぐために観測情報を重視するタイミングを限定することで精度の低下した観測情報の影響を極力受けない手法 (Switching Particle Filters, SPFs) の提案を行った。SPFs では、従来の観測情報と独立なサンプリング分布と観測情報を導入したサンプリング分布がお互いに相補的な性質を持つことを利用し、現時点での推定状態の分散を推定精度として推定精度の高い時には不安定な観測情報の影響を受けないサンプリング分布を、推定精度が低下している時には (例え不安定であっても) 観測情報を重視したサンプリング分布を採用すること

で、実問題に対する推定精度の維持と頑健性の向上を図った。また、オンライン EM アルゴリズム [72] を用いて、逐次的に追跡対象のダイナミクスモデルのパラメータを学習することで、ダイナミクスモデルの改善を行い、予測分布の精度向上から推定精度の向上を図った。ダイナミクスモデルのオンライン学習は、特に学習初期に推定精度の低下を招くことが多かったが、提案手法に依れば、学習初期の間は観測情報を重視してダイナミクスモデルの改善を行い、ダイナミクスモデルの学習が進んだ後は観測モデルと独立なサンプリング分布からのサンプルを用いて頑健な状態推定を行う形に自動的に移行することが可能であった。

観測モデル

第 3 章で述べたサンプリング分布の動的な切り替え手法では、サンプリング分布の性能を純粹に比較するため、可能な限り簡単な観測モデルを仮定した。しかし先に述べた通り、視覚追跡の精度は観測モデルに大きく依存するため、従来、追跡対象に合わせて様々な観測モデルを用いた手法が提案されている。例えば、手や頭部、目など輪郭のはっきりした対象を扱う際には追跡対象の形状に基づいた観測モデルを構築して追跡を行う手法 [35, 105, 50] が多く用いられ、スポーツ中継映像におけるプレーヤ追跡の様に画像列中で激しく運動する人物を追跡対象とする場合には、大きな形状の変化を伴うため、追跡対象の周辺領域のヒストグラムに基づいた観測モデルにより追跡を行う手法 [18, 19, 64, 65] が多く用いられてきた、この様に追跡対象の見え方が大きく変化する場合には、追跡対象を主成分分析 (Principal Component Analysis, PCA) 等により特定の特徴部分空間へと射影することで頑健な追跡を実現する手法 [9, 43, 51, 52, 69, 92] も多く用いられている。また、形状や色、音声など複数の情報を混合して観測モデルを構築し頑健な追跡を実現する手法 [11, 14, 36, 74, 75] も幾つか提案されている。

幅広い応用分野から人物の頭部、顔等を追跡対象とする研究も多いが、複雑な形状を扱う場合、モデルの次元が高次元となる場合が多々ある。これは形状ではなく領域を対象とする手法でも同様で、追跡対象の姿勢変化などによる見え方の変化などを説明するためには非常に高次元なモデルが必要となる。状態空間の高次元化と共に、精度を保った推定を行うために必要なサンプル数は指数的に増加す

るため，大まかな形状を用いたり [35]，領域のヒストグラム [18, 19, 65] を用いた観測モデル等，低次元化を行う手法が多く採用されている．近年多く採用されるようになってきた追跡対象の基底画像を用いた手法 [51, 52, 69] も，状態空間の低次元化の手法の一つであるが，これに更に，カルマンフィルタなど解析的手法で推定不可能な状態空間のみをサンプル近似する手法である Rao-Blackwellisation [15] (RB) を採用した Rao-Blackwellised Particle Filter [1, 21, 22] (RB-PF) を適用した手法 [43] も提案されている．この様にサンプル数を少なく抑えたまま，頑健な状態推定の実現の為には，可能な限り状態空間を階層構造化してサンプル近似で推定する際のモデルを低次元化し，簡単なモデルを用いて推定を行うことが必要と考えられる．

階層的状態推定とサンプリング分布

本論文でも頭部姿勢と顔特徴点の同時推定の様に高次元の状態空間を持つ問題を第 4 章で扱ったが，パーティクルフィルタで推定する状態空間を各顔特徴点の画像面上での位置のみに限定し，簡単な観測モデル（正面テンプレートとの正規化相関値）を用いて推定を行った．この様に状態空間の次元を低次元化し，観測モデルを簡略化すれば，詳細な観測モデルを用いた場合と比べて推定精度は低下するであろうが，サンプル数を少数に制限される場合にも頑健な推定を行うことができる．第 4 章で扱った実験では，問題の階層構造を利用した適応的サンプリング分布を導入することで，顔特徴点の間に階層構造を通した柔らかい制約に加えて，それぞれの顔特徴点独自のダイナミクスによる独立性を実現した．これにより全ての特徴点位置を同時に推定する必要がなくなるため，サンプルを各特徴点で 50 個ずつと非常に少数に制限したにも関わらず頑健で精度の良い推定が可能であった．先述の RB-PF を適用した場合，線形のカメラモデルを仮定することで頭部姿勢の推定がカルマンフィルタで行うことができるため，提案手法と同様にパーティクルフィルタで推定する状態空間の低次元化を行うことができる．但し，この場合は各特徴点の間に独立なダイナミクスを導入することが困難であるため，画像面上での全ての特徴点位置を同時に扱う必要があり，特徴点数が多い場合（本研究では特徴点数が 8 点の台形柱もしくは特徴点数が 11 点の頭部を扱っ

た．この時，それぞれの特徴点が 2 次元の状態空間で表現されるとすれば，全ての
特徴点位置の同時推定を行う場合には 16 次元，22 次元の状態空間を扱う必要
がある)には依然として高次元の状態空間を扱う必要がある．また，前パラグラ
フで述べたとおり，観測モデルは単純であればあるほど頑健ではあるが，照明条
件の変化等，想定していない雑音の影響を受けるため，非定常に推定精度の低下
することが多々起りうる．第 4 章で提案した適応的サンプリング分布 (Adaptive
Sampling Distribution, ASD) を採用すれば，こうした問題を階層構造を利用して
サンプリング分布を修正することで低減することに成功し，少数のサンプルにも
関わらず頑健で精度の良い推定の実現が可能であった．第 3 章で提案した SPFs
は，階層化されていない単純な問題において，予測を重視するサンプリング分布
と観測を重視するサンプリング分布を動的に切り替える手法であり，その規準と
して推定精度の良さを示す推定状態の分散の閾値を予め設定する必要があった．
それに対して第 4 章で提案した PF-ASD では，階層構造を利用して各層からそれ
ぞれ独立な予測に基づいたサンプリング分布を構成し，各層からのサンプリング
分布の状態推定への寄与の大きさに基づいて，どのサンプリング分布を重視する
かを適応的に決定している．これは暗にどの階層で用いた仮定が正しいかを適応
的に決めていくことに相当し，サンプリング分布の選択規準となる値を予め設定
しておく必要は無い．また，ダイナミクスモデルのオンライン学習やテンプレ
ートの逐次更新手法の様に，対象のモデルを適応的に変更するような手法は，一般
的に誤差が蓄積しやすく，最終的には追跡が破綻してしまう場合があることが知
られている．この様な問題に対しても，階層構造を利用して適応的にサンプリン
グ分布の選択を行う提案手法では，サンプリング分布の選択のみを行っており，
各層からのサンプリング分布の精度の枠内で推定を行うことができるため，非常
に頑健な推定の実現を行うことが可能であった．

サンプリング分布の改善という観点から考えれば，PF-ASD は一時刻前の事後
分布を用いて，事後確率の高いサンプルを得ることのできる予測分布を選択する
枠組みである．これはある規準に基づいてサンプルを評価しサンプルの状態遷移
先を決定する点において APF と類似の構造であるが，APF が現時刻の観測を用
いて一時刻前のサンプルを遷移させる 1 ステップのスミージングと類似した操作

を行うことに対して，PF-ASD では予測分布の確信度は緩やかに変化するという仮定の下，一時刻前の事後分布を用いて予測分布の確信度を評価している．この様に PF-ASD では観測情報を間接的に用いることで，実環境中で不正確になりがちな観測情報にもかかわらず，頑健な推定を実現している．また，第 3 章における実験では，各時刻で採用するサンプリング分布を完全に切り替える h-SPFs が緩やかに切り替える s-SPFs と比べて，多くの場合で良い性能を示していた．これはここでの実験では遮蔽の存在が大きな鍵になっていたため，観測の精度が遮蔽に隠れているとき/隠れていないときで大きく変化し，こうした極端な変化に上手く対応できる h-SPFs の性能が非常に良かったと考えられる．それに対して第 4 章の実画像を用いた実験においては，二つの予測分布からサンプルされるサンプル数は h-SPFs のように大きく切り替わるものと，s-SPFs のように緩やかに切り替わるものが，特徴点に応じて混在していた (cf. 図 4.15)．これは，この実験では頭部姿勢に応じて特徴点の観測精度が動的に変化するため，PF-ASD によって自動的に精度の維持のために適切な切り替え方を特徴点毎に選択できていたためと考えられる．この様に非定常にその精度を大きく変える実環境における観測情報を，状況に合わせて適切に利用できる点が PF-ASD の大きな利点の一つである．

第6章 結言

本論文では実環境中における隠れ状態推定問題に対して，実環境中に存在する非定常な非ガウス性雑音の影響で隠れ状態の分布が非ガウスとなるような場合にも頑健で精度の良い推定を行うための手法についての議論を行った．

第2章では一般のベイズ推定から時系列ベイズ推定の説明を行い，その計算困難性と解法としてカルマンフィルタ，パーティクルフィルタを紹介した．また，カルマンフィルタは線形システムと隠れ状態の分布にガウス分布を仮定する必要があること，パーティクルフィルタはサンプル近似による手法であるため分布の形状を仮定する必要は無いが，サンプル数にその推定精度が大きく依存することを述べ，実環境中における実時間状態推定問題において，如何に少ないサンプルで効率の良い近似を行うことが重要な課題であることを述べた．

第3章では，予測重視のサンプリング手法と観測重視のサンプリング手法を動的に切り替えることで，実環境中でより頑健な視覚追跡を行う Switching Particle Filters (SPFs) の提案を行った．ここでは，予測重視のサンプリング法として，現時刻の観測と独立なサンプリング分布を採用した Condensation を用い，観測重視のサンプリング法として，サンプリング分布に観測情報を導入した Auxiliary Particle Filter を用いた．観測情報が不確実になることが多々考えられる実環境中における実時間視覚追跡の様な状態推定問題において，現時刻の観測と独立なサンプリング分布を採用することによる外れ値問題と，現時刻の観測をサンプリング分布に導入することによるサンプリング分布の精度低下はどちらも頻繁に起こり得る．こうした場合に，どちらか一方のサンプリング分布を利用するのではなく，お互いに相補的なサンプリング分布を使い分けて状態推定を行う事で，不確実な観測情報の影響を出来る限り抑え，より頑健な推定の実現が可能であった．また，追跡対象のダイナミクスを逐次学習する手法の提案も行い，SPFs と組み合

わせることで、ダイナミクス学習初期にサンプリング分布の精度が低下する影響で起る追跡頑健性の低下を抑え、頑健で正確な視覚追跡の実現の可能性を示した。

第3章では、極端に簡略化した観測モデルを用いることで、サンプリング分布の持つ純粋な性質を評価することを試みたが、第4章では実環境中におけるより高度な高次元隠れ状態の実時間推定問題において、パーティクルフィルタを採用した際の問題点を解決するために、問題の持つ階層性を考慮した適応的サンプリング分布の紹介を行い、剛体の三次元姿勢と画像面上における特徴点位置の同時推定問題への適用について説明した。数値実験により階層構造を考慮した適応的サンプリング分布を採用した Particle Filter using Adaptive Sampling Distribution (PF-ASD) といくつかの従来手法を比較することで、PF-ASD の有効性を確認すると共に、4節では実際の環境で撮影された画像を入力とする、実車内におけるドライバの頭部姿勢と目や鼻など顔特徴点位置の同時推定問題にも適用した。但し、実問題に適用するにあたり、顔特徴点位置が直接観測することが出来ないため、ここではもっとも単純なモデルの一つである正面テンプレートとの正規化相関値に基づいたモデルを用いて、適応的サンプリング分布を採用したパーティクルフィルタの有効性を確認した。更に、ドライバの三次元顔モデルは車内で直接計測することが不可能なため一般には未知であり、第4章4節では複数人のドライバから作成した平均三次元顔モデルを用いて頭部姿勢を行っていたが、5節では、平均三次元顔モデルを用いて検出された顔特徴点の一時系列から、ドライバ個人に適応した三次元顔モデルの再構成を行う手法（因子分解法）について述べた。ただし、通常因子分解法はデータの欠損がなく、精度良く検出された特徴点位置時系列を必要とするため、頭部姿勢の変化に応じて隠れや特徴点の見え方に大きな変化が存在する本課題には、因子分解法を直接適用することは困難であった。そのため、ここでは因子分解法を確率的手法に拡張し、さらに逐次処理可能な形に拡張することで、隠れや大きな非定常雑音が存在する場合においても、頑健で精度の良い三次元顔モデルの復元が可能であることを示した。

第5章では、実環境における実時間状態推定問題に対して広く用いられてきたパーティクルフィルタの持つ問題点を明確にし、これに対して本研究で取り組んできた手法の有効性を、サンプリング分布、観測モデル、階層的状態推定の観点

から議論した。

謝辞

本研究の機会を与えてくださり，本研究を進める上で示唆に富んだ多大な御助言を頂いた石井 信 教授，博士前期過程在籍中も含めて，日頃の直接的な御指導と研究生活全般に渡る多くの御助言を頂いた柴田 智広 助教授に深く感謝申し上げます．また，セミナーや研究発表等で常に有益な議論と提案を頂いた作村 諭一 助教授，大羽 成征 助手，前田 新一 助手，吉田 和子 助手，竹ノ内 高志 研究員，行縄直人 研究員に感謝致します．論理生命学講座の歴代の秘書である，谷本 史 さん，那谷 緑 さん，藤澤 祐子さんには，日頃の生活や事務的な手続き等で多大な助力を頂きました．どうもありがとうございました．木戸出 正継 教授には主に研究発表の場で，直接研究に携わっていると気づき難い考慮すべき問題点を指摘して頂くなど，有益な御助言を頂きました．深く感謝致します．

川人 光男 ATR 脳情報研究所所長，銅谷 賢治 ATR 脳情報研究所計算神経生物学研究室室長（奈良先端科学技術大学院大学神経計算学講座客員教授，沖縄科学技術大学院大学先行的研究事業神経計算ユニット研究代表兼務），佐藤 雅昭 ATR 脳情報研究所計算脳イメージング研究室室長には，博士前期過程に在籍中から多くの御助言を頂きました．

また，ドライバの視線方向推定という応用技術への適用の場を与えて頂いた，株式会社デンソー統合システム開発部の深谷 直樹 博士，清水 幹郎 氏，本学論理生命学講座 OB でもある人見 謙太郎 氏に深く感謝申し上げます．特に深谷，清水両氏には応用実験に用いた画像データの撮影等，本学では取得困難なデータの取得に関して非常にお世話になりました．

博士前期過程に入学して以来，朝から夜まで五年間に渡って最も近くで研究を続けてきた，川脇 大 君を始め，同じ学年であった金子 裕子さん，田辺 昭博 君，塚田 祐基 君，平山 潤一郎 君，藤田 肇 君，森 健 君には，毎日の生活の中で多く

の刺激と助言を頂きました。また、鹿内 学 君、為井 智也 君、植野 剛 君、殿谷 徳和 君、牧野 研司 君、伊達 虎三 君、橋崎 知 君には、特に同じ柴田 助教授の研究グループということで、日常の雑多なことから実験の手伝いまで広く多大な助力を頂きました。本当にお世話になりました。ありがとうございます。その他にも、日頃の研究を支えていただいた論理生命学講座スタッフ・学生の皆様には改めて御礼申し上げます。

最後に、ご多忙であるにも関わらず本論文の審査を引き受けて下さった、石井 信 教授、木戸出 正継 教授、柴田 智広 助教授に改めて感謝いたします。どうもありがとうございました。

参考文献

- [1] Christophe Andrieu, Nando de Freitas, and Arnaud Doucet. Rao-blackwellised particle filtering via data augmentation. In *Advances in Neural Information Processing Systems 13*, 2001.
- [2] Christophe Andrieu, Arnaud Doucet, and Elena Puskaya. *Sequential Monte Carlo Methods for Optimal Filtering*, chapter 4, pp. 79–95. Springer, 2001.
- [3] Nicholas Apostoloff and Alexander Zelinsky. Vision in and out of vehicles: Integrated driver and road scene monitoring. *The International Journal of Robotics Research*, Vol. 23, No. 4-5, pp. 513–538, 2004.
- [4] Siley O. Ba and Jean-Marc Odobez. A rao-blackwellized mixed state particle filter for head pose tracking. In *ACM ICMI Workshop on Multimodal Multiparty Meeting Processing (MMMP)*, pp. 9–16, 2005.
- [5] Takashi Bando, Tomohiro Shibata, Kenji Doya, and Shin Ishii. Switching particle filters for efficient real-time visual tracking. In *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 2, pp. 720–723, 2004.
- [6] Sumit Basu, Irfan Essa, and Alex Pentland. Motion regularization for model-based head tracking. In *13th International Conference on Pattern Recognition*, Vol. 3, pp. 611–616, 1996.
- [7] Md. Al-Amin Bhuiyan, Vuthichai Ampornaramveth, Shin yo Muto, and Haruki Ueno. Face detection and facial feature localization for human-machine interface. *NII Journal*, Vol. 5, pp. 25–39, 2003.

- [8] Michael J. Black and Allan D. Jepson. EigenTracking: Robust matching and tracking of articulated objects using a view-based representation. In *European Conference on Computer Vision*, Vol. 1, pp. 329–342, 1996.
- [9] Michael J. Black and Allan D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, Vol. 26, No. 1, pp. 63–84, 1998.
- [10] Andrew Blake, Rupert Curwen, and Andrew Zisserman. A frame work for spatio-temporal control in the tracking of visual contours. *International Journal of Computer Vision*, Vol. 11, No. 2, pp. 127–145, 1993.
- [11] Andrew Blake, Michel Gangnet, Patrick Perez, and Jaco Vermaak. Integrated tracking with vision and sound. In *11th International Conference on Image Analysis and Processing*, pp. 354–357, 2001.
- [12] Bjorn Braathen, Marian Stewart Bartlett, Gwen Littlewort-Ford, and Javier Movellan. 3-d head pose estimation from video by stochastic particle filtering. In *Proceedings of the 8th Annual Joint Symposium on Neural Computation*, 2001.
- [13] Lars Bretzner, Ivan Laptev, and Tony Lindeberg. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In *Proceedings 5th IEEE International Conference on Automatic Face Gesture Recognition*, pp. 423–428, 2002.
- [14] David J. Bullock and John S. Zelek. Real-time tracking for visual interface applications in cluttered and occluding situations. In *15th International Conference on Vision Interface*, 2001.
- [15] G. Casella and C.P. Robert. Rao-blackwellisation of sampling schemes. *Biometrika*, Vol. 83, No. 1, pp. 81–94, 1996.

- [16] Cheng Chang and R. Ansari. Kernel particle filtering for visual tracking. *IEEE Signal Processing Letters*, Vol. 12, No. 3, pp. 242–5, 2005.
- [17] Xavier Clady, Francois Collange, Frédéric Jurie, and Philippe Martinet. Object tracking with a pan-tilt-zoom camera: application to car driving assistance. In *IEEE International Conference on Robotics and Automation*, pp. 1653–1658, 2001.
- [18] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Real-time tracking of non-rigid objects using mean shift. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR2000)*, pp. 142–151, 2000.
- [19] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 5, pp. 564–577, 2003.
- [20] Frank Dellaert, Wolfram Burgard, Dieter Fox, and Sebastian Thrun. Using the condensation algorithm for robust, vision-based mobile robot localization. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 588–594, 1999.
- [21] A. Doucet, S. Godsil, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, Vol. 10, No. 3, pp. 197–208, 2000.
- [22] Arnaud Doucet, Nando de Freitas, Kevin Murphy, and Stuart Russell. Rao-blackwellised particle filtering for dynamic bayesian networks. In *Uncertainty in Artificial Intelligence*, 2000.
- [23] Dieter Fox. Kld-sampling: Adaptive particle filters. In *Advances in Neural Information Processing Systems 14*, 2001.
- [24] Dieter Fox, Wolfram Burgard, Hannes Kruppa, and Sebastian Thrun. A probabilistic approach to collaborative multi-robot localization. *Autonomous Robots*,

- Vol. 8, No. 3, pp. 325–344, 2000.
- [25] Jun Fujiki and Takeshi Kurata. Recursive factorization method for the paraperspective model based on the perspective projection. In *International Conference on Pattern Recognition*, Vol. 1, pp. 406–410, 2000.
- [26] Jun Fujiki and Takeshi Kurata. A recursive factorization method for the metric affine projection model. *Systems and Computers in Japan*, Vol. 37, No. 7, pp. 83–93, 2006.
- [27] W. R. Gilks and C. Berzuini. Following a moving target - monte carlo inference for dynamic bayesian models. *Journal of the Royal Statistical Society*, Vol. B63, No. 1, pp. 127–146, 2001.
- [28] N.J. Gordon, J.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear non-gaussian bayesian state estimation. In *IEEE Proceedings on Radar Signal Processing*, Vol. 140, pp. 107–113, 1993.
- [29] Sabri Gurbuz, Keisuke Kinoshita, and Shinjiro Kawato. Real-time human nose bridge tracking in presence of geometry and illumination changes. In *Second International Workshop on Man-Machine Symbiotic Systems*, 2004.
- [30] Soumya Hamlaoui and Franck Davoine. Facial action tracking using particle filters and active appearance models. In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence*, pp. 165–169, 2005.
- [31] Dan Witzner Hansen, Riad Hammoud, Ronald Satria, and Jakob Sorensen. Improved likelihood function in particle-based ir eye tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 3, pp. 5–12, 2005.
- [32] T. Higuchi. Monte Carlo filter using the generic algorithm operations. *Journal of Statistical Computation and Simulation*, Vol. 59, No. 1, pp. 1–23, 1997.

- [33] Junichiro Hirayama, Junichiro Yoshimoto, and Shin Ishii. Bayesian representation learning in the cortex regulated by acetylcholine. *Neural Networks*, Vol. 17, pp. 1391–1400, 2004.
- [34] M. Isard and J. MacCormick. Bramble: A bayesian multiple-blob tracker. In *International Conference on Computer Vision*, Vol. 2, pp. 34–41, 2001.
- [35] Michael Isard and Andrew Blake. Condensation-conditional density propagation for visual tracking. *International Journal of Computer Vision*, Vol. 29, No. 1, pp. 5–28, 1998.
- [36] Michael Isard and Andrew Blake. Icondensation: Unifying low-level and high-level tracking in a stochastic framework. In *5th European Conference on Computer Vision*, 1998.
- [37] Michael Isard and Andrew Blake. A smoothing filter for condensation. In *5th European Conference on Computer Vision*, 1998.
- [38] Anli K. Jain, Yu Zhong, and Sridhar Lakshmanan. Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 3, pp. 267–278, 1996.
- [39] Simon J. Julier and Jeffrey K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *International Symposium Aerospace/Defense Sensing, Simulations and Controls*, 1997.
- [40] Frédéric Jurie and Michel Dhome. Real time tracking of 3dobjects: an efficient and robust approach. *Pattern Recognition*, Vol. 35, pp. 317–328, 2002.
- [41] Michael k. Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filter. *Journal of American Statistical Assosiation*, Vol. 94, pp. 590–599, 1999.
- [42] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, Vol. 82, No. Series D, pp. 35–45, 1960.

- [43] Zia Khan, Tucker R. Balch, and Frank Dellaert. A Rao-Blackwellized particle filter for EigenTracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 980–986, 2004.
- [44] Genshiro Kitagawa. Monte Carlo filter and smoother for non-Gaussian non-linear state space models. *Journal of Computational and Graphical Statistics*, Vol. 5, No. 1, pp. 1–25, 1996.
- [45] Fernando De la Torre and Michael J. Black. Robust principal component analysis for computer vision. In *International Conference on Computer Vision*, Vol. 1, pp. 362–369, 2001.
- [46] Mun Wai Lee and Isaac Cohen. Proposal maps driven mcmc for estimating human body pose in static images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 334–341, 2004.
- [47] Mun Wai Lee and Isaac Cohen. A model-based approach for estimating human 3d poses in static images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 6, pp. 905–916, 2006.
- [48] Mun Wai Lee and Ramakant Nevatia. Dynamic human pose estimation using markov chain monte carlo approach. In *IEEE Workshop on Motion and Video Computing (WACV/MOTION'05)*, Vol. 2, pp. 168–175, 2005.
- [49] Peihua Li, Tianwen Zhang, and Bo Ma. Unscented kalman filter for visual curve tracking. *Image and Vision Computing*, Vol. 22, No. 2, pp. 157–164, 2004.
- [50] Peihua Li, Tianwen Zhang, and Arthur E. C. Pece. Visual contour tracking based on particle filters. *Image and Vision Computing*, Vol. 21, No. 1, pp. 111–123, 2003.
- [51] Jongwoo Lim, David A. Ross, Rwei-Sung Lin, and Ming-Hsuan Yang. Incremental learning for visual tracking. In *Advances in Neural Information Processing Systems 17*, pp. 793–800, 2005.

- [52] Ruei-Sung Lin, David A. Ross, Jongwoo Lim, and Ming-Hsuan Yang. Adaptive discriminative generative model and its applications. In *Advances in Neural Information Processing Systems 17*, pp. 801–808, 2005.
- [53] Chengjun Liu. A bayesian discriminating features method for face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 6, pp. 725–740, 2003.
- [54] J.S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, Vol. 93, No. 443, pp. 1032–1044, 1998.
- [55] Le Lu, Xiang-Tian Dai, and Gregory Hager. A particle filter without dynamics for robust 3d face tracking. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 70–77, 2004.
- [56] Lucie Masson, Michel Dhome, and Frédéric Jurie. Robust real time tracking of 3d objects. In *17th International Conference on Pattern Recognition*, Vol. 4, pp. 252–255, 2004.
- [57] Yoshio Matsumoto and Alexander Zelinsky. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition 2000*, pp. 499–504, 2000.
- [58] Toshihiko Morita and Takeo Kanade. A sequential factorization method for recovering shape and motion from image streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 8, pp. 858–867, 1997.
- [59] Hammadi Nait-Charif and Stephen J. McKenna. Head tracking and action recognition in a smart meeting room. In *4th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, International Conference on Computer Vision Systems (PETS-ICVS)*, 2003.

- [60] Hieu T. Nguyen and Arnold W.M. Smeulders. Fast occluded object tracking by a robust appearance filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 8, pp. 1099–1104, 2004.
- [61] Hieu T. Nguyen and A.W.M. Smeulders. Appearance kalman tracking under severe occlusions. Technical Report 7, Intelligent Sensory Information Systems, University of Amsterdam, 12 2002.
- [62] Ben North and Andrew Blake. Learning dynamical models using expectation-maximisation. In *Proceeding of 6th International Conference on Computer Vision*, 1998.
- [63] Ben North and Andrew Blake. Learning and classification of complex dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 9, pp. 1016–1034, 2000.
- [64] Katja Nummiaro, Esther Koller-Meier, and Luc Van Gool. Object tracking with an adaptive color-based particle filter. In *Symposium for Pattern Recognition of the DAGM*, pp. 355–360, 2002.
- [65] Kenji Okuma, Ali Taleghani, Nando De Freitas, James J. Little, and David G. Lowe. A boosted particle filter: Multitarget detection and tracking. In *European Conference on Computer Vision*, pp. 28–39, 2004.
- [66] A. Perez, M.L. Cordoba, A. Garcia, R. Mendez, M.L. Munoz, J.L. Pedraza, and F. Sanchez. A precise eye-gaze detection and tracking system. In *WSCG'2003: 11th International Conference in Central Europe on Computer Graphics, Visualization and Digital Interactive Media*, pp. 105–8, 2003.
- [67] David Raynard, Andrew Wildenberg, Andrew Blake, and John Marchant. Learning dynamics of complex motions from image sequence. In *5th European Conference on Computer Vision*, pp. 357–368, 1996.

- [68] Ehud Rivlin, Hector Rotstein, and Yehoshua Y. Zeeyi. Two-mode control: An oculomotor-based approach to tracking systems. *IEEE Transactions on Automatic Control*, Vol. 43, No. 6, pp. 833–842, 1998.
- [69] David Ross, Jongwoo Lim, and Ming-Hsuan Yang. Adaptive probabilistic visual tracking with incremental subspace update. In *European Conference on Computer Vision*, pp. 470–482, 2004.
- [70] Yong Rui and Yunqiang Chen. Better proposal distributions: Object tracking using unscented particle filter. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '01)*, 2001.
- [71] M. Sato. Online model selection based on the Variational Bayes. *Neural Computation*, Vol. 13, pp. 1649–1681, 2001.
- [72] Masaaki Sato and Shin Ishii. On-line EM algorithm for the normalized gaussian network. *Neural Computation*, Vol. 12, No. 2, pp. 407–432, 2000.
- [73] D. Schulz and D. Fox. Bayesian color estimation for adaptive vision-based robot localization. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 2, pp. 1884–1889, 2004.
- [74] Chunhua Shen, Anton van den Hengel, and Anthony Dick. Probabilistic multiple cue integration for particle filter based tracking. In *Seventh International Conference on Digital Image Computing: Techniques and Applications*, pp. 399–408, 2003.
- [75] Jamie Sherrah and Shaogang Gong. Fusion of perceptual cues for robust tracking of head pose and position. *Pattern Recognition*, Vol. 34, No. 8, pp. 1565–1572, 2001.
- [76] Heung-Yeung Shum, Katsushi Ikeuchi, and Raj Reddy. Principal component analysis with missing data and its application to polyhedral object modeling.

- IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 9, pp. 854–867, 1995.
- [77] Paul Smith, Mubarak Shah, and Niels da Vitoria Lobo. Determining driver visual attention with one camera. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 4, No. 4, pp. 205–218, 2003.
- [78] H. W. Sorenson and A. R. Stubberud. Non-linear filtering by approximation of the a posteriori density. *International Journal of Control*, Vol. 8, No. 1, pp. 33–51, 1968.
- [79] Yasuyuki Sugaya and Kenichi Kanatani. Outlier removal for motion tracking by subspace separation. *IEICE Transactions on Information and Systems*, Vol. E86-D, No. 6, pp. 1095–1102, 2003.
- [80] Toni Tamminen and Jouko Lampinen. Sequential monte carlo for bayesian matching of objects with occlusions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 6, pp. 930–941, 2006.
- [81] Sebastian Thrun, Dieter Fox, and Wolfram Burgard. Monte carlo localization with mixture proposal distribution. In *American Association for Artificial Intelligence*, pp. 859–865, 2000.
- [82] Sebastian Thrun, Dieter Fox, Wolfram Burgard, and Frank Dellaert. Robust monte carlo localization for mobile robots. *Artificial Intelligence*, Vol. 128, No. 1-2, pp. 99–141, 2000.
- [83] Michael E. Tipping and Christopher M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, Vol. 11, pp. 443–482, 1999.
- [84] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, Vol. 9, No. 2, pp. 137–154, 1992.

- [85] David Tweed and Andrew Calway. Tracking many objects using subordinated condensation. In Paul Rosin and David Marshall, editors, *Proceedings of the British Machine Vision Conference*, pp. 283–292. BMVA Press, October 2002.
- [86] Rudolph van der Merwe, Nando de Freitas, Arnaud Doucet, and Eric Wan. The unscented particle filter. In *Advances in Neural Information Processing Systems 13*, 2001.
- [87] Roel Vertegaal. Attentive user interfaces. *Communications of the ACM*, Vol. 46, No. 3, pp. 30–33, 2003.
- [88] Paul Viola and Michael Jones. Robust real-time face detection. *International Journal of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.
- [89] Nikos Vlassis, Bas Terwijn, and Ben Kröse. Auxiliary particle filter robot localization from high-dimensional sensor observations. In 7-12, editor, *Int. Conf. on Robotics & Automation*, 2002.
- [90] Frank Wallhoff, Martin Zobl, Gerhard Rigoll, and Igor Potucek. Face tracking in meeting room scenarios using omnidirectional views. In *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 4, pp. 933–936, 2004.
- [91] Hanzi Wang, David Suter, and Konrad Schindler. Effective appearance model and similarity measure for particle filtering and visual tracking. In *European Conference on Computer Vision*, Vol. 3, pp. 606–618, 2006.
- [92] Ming Yang and Ying Wu. Tracking non-stationary appearances and dynamic feature selection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 1059–1066, 2005.
- [93] 足立佳久, 中西知, 久野義徳, 島田伸敬, 白井良明. 顔の視覚情報処理を用いた知的車椅子. *日本ロボット学会誌*, Vol. 17, No. 3, pp. 423–431, 1999.

- [94] 伊藤健, 坂根茂幸. アフィン変換テンプレート群の動的遷移に基づくビジュアルトラッキング. 日本ロボット学会誌, Vol. 19, No. 1, pp. 100–108, 2001.
- [95] 伊藤健, 坂根茂幸. オクルージョンに対してロバストなビューベーストビジュアルトラッキング. 日本ロボット学会誌, Vol. 19, No. 8, pp. 1018–1026, 2001.
- [96] 竹上健, 後藤敏行, 大山玄. 視線方向計測のための高精度瞳孔検出アルゴリズム. 電子情報通信学会論文誌, Vol. J86-D-II, No. 2, pp. 252–261, 2003.
- [97] 金谷健一, 菅谷保之. 因子分解法の完全レシピ. PRMU 115-127, 電子情報通信学会技術報告, 2003.
- [98] 林健太郎, 橋本学, 鷲見和彦. 頑健性と精緻性を備えた顔特徴点追跡による顔方向推定. 電子情報通信学会論文誌, Vol. J84-D-II, No. 8, pp. 1762–1771, 8 2001.
- [99] 岡兼司, 佐藤洋一, 中西泰人, 小池英樹. 適応的拡散制御を伴うパーティクルフィルタを用いた頭部姿勢推定. 電子情報通信学会論文誌, Vol. J88-D-II, No. 8, pp. 1601–1613, 2005.
- [100] 上條俊介, 松下康之, 池内克史, 坂内正夫. 時空間 markov random field モデルによる隠れにロバストなトラッキングアルゴリズム. 電子情報通信学会論文誌, Vol. J83-D-II, No. 12, pp. 2597–2609, 2000.
- [101] 上條俊介, 原田将弘, 坂内正夫. 統計モデルと意味階層の結合による交通映像異常事象検出システム. 電子情報通信学会論文誌, Vol. J88-A, No. 2, pp. 152–163, 2005.
- [102] 足立佳久, 恒成裕行, 松本吉央, 小笠原司. ユーザの注意推定に基づく搭乗型移動ロボットの走行支援システム. ヒューマンインタフェース学会論文誌, Vol. 6, No. 3, pp. 9–18, 2004.
- [103] 網倉忍, 伊藤洋平, 工藤博幸. データ欠損を考慮した因子分解法. PRMU 202-232, 電子情報通信学会技術報告, 2006.

- [104] 佐藤雄隆, 金子俊一, 五十嵐悟. 選択的正規化相関法によるロバスト画像照合. Prmu, 電子情報通信学会技術報告, 2000.
- [105] 北川洋介, 加藤丈和, 呉海元, 和田俊和. アイモデルを用いた Condensation による黒目追跡. PRMU 198-211, 電子情報通信学会技術報告, 2005.
- [106] 白井了, 井東道昌, 乗松有, 足立和正, 中野倫明, 山本新. ドライバの運転状態検知のための視線方向検出. PRMU 151-170, 電子情報通信学会技術報告, 2003.
- [107] 福井和広, 山口修. 形状抽出とパターン照合の組み合わせによる顔特徴点抽出. 電子情報通信学会論文誌, Vol. J80-D-II, No. 8, pp. 2170–2177, 1997.

付録

A. 略語表

本論文で使用した略語まとめたものを表 A.1 に示す .

B. EM アルゴリズムを用いたダイナミクスパラメータのオンライン推定

時刻 t における追跡対象の状態を x_t , 観測を z_t , 時刻 T までの状態列を $\mathbf{x}_{1:T} = \{x_1, \dots, x_T\}$, 観測列を $\mathbf{z}_{1:T} = \{z_1, \dots, z_T\}$ とする時 , 尤度を最大にするパラメータ θ を求める . 対数尤度は

$$\begin{aligned} \log P(\mathbf{z}_{1:T}|\theta) &= \int Q(\mathbf{x}_{1:T}) \log \frac{P(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}|\theta)}{P(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \theta)} d\mathbf{x}_{1:T} \\ &= \int Q(\mathbf{x}_{1:T}) \log \frac{P(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}|\theta)}{Q(\mathbf{x}_{1:T})} \frac{Q(\mathbf{x}_{1:T})}{P(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \theta)} d\mathbf{x}_{1:T} \\ &= \int Q(\mathbf{x}_{1:T}) \log \frac{P(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}|\theta)}{Q(\mathbf{x}_{1:T})} d\mathbf{x}_{1:T} \\ &\quad + \int Q(\mathbf{x}_{1:T}) \log \frac{Q(\mathbf{x}_{1:T})}{P(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \theta)} d\mathbf{x}_{1:T} \\ &= \mathcal{F}(Q(\mathbf{x}_{1:T}), \theta) + \text{KL}(Q(\mathbf{x}_{1:T})||P(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \theta)) \\ &\geq \mathcal{F}(Q(\mathbf{x}_{1:T}), \theta) \end{aligned} \tag{A.1}$$

ここで $Q(\mathbf{x}_{1:T})$ は事後分布を近似するものとして定義された試験事後分布である . KL 距離は非負値の距離関数であるので , 対数尤度は自由エネルギー \mathcal{F} を用いて下から近似される . よって , この自由エネルギーを $Q(\mathbf{x}_{1:T})$ と θ について最大化することは $\mathbf{z}_{1:T}$ に対する最尤推定量 θ を求めることと等価である .

表 A.1 略語表

abbreviation	words
APF	Auxiliary Particle Filter
ASD	Adaptive Sampling Distribution
EKF	Extended Kalman Filter
EM	Expectation Maximization
KL	Kullback Leibler
MCMC	Markov Chain Monte Carlo
PCA	Principal Component Analysis
PC-PF	Process-noise Controlled Particle Filter
PF-ASD	Particle Filtering using Adaptive Sampling Distribution
PPCA	Probabilistic Principal Component Analysis
RB	Rao-Blackwellisation
RB-PF	Rao-Blackwellised Particle Filter
UKF	Unscented Kalman Filter
UPF	Unscented Particle Filter
SIR	Sampling Importance Resampling
SIS	Sequential Importance Sampling
SPFs	Switching Particle Filters
h-SPFs	Hard Switching Particle Filters
s-SPFs	Soft Switching Particle Filters
VB	Variational Bayes

Eステップ: 式(A.1)より

$$\mathcal{F}(Q(\mathbf{x}_{1:T}), \theta) = \log P(\mathbf{z}_{1:T}|\theta) - \text{KL}(Q(\mathbf{x}_{1:T})||P(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \theta)) \quad (\text{A.2})$$

となり, 右辺第二項のKL距離は非負なので \mathcal{F} が $Q(\mathbf{x}_{1:T})$ に関して最大となるのは, KL距離が0となる

$$Q(\mathbf{x}_{1:T}) = P(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \theta) \quad (\text{A.3})$$

の時である.

Mステップ: 自由エネルギー \mathcal{F} は

$$\begin{aligned} \mathcal{F}(Q(\mathbf{x}_{1:T}), \theta) &= \int Q(\mathbf{x}_{1:T}) \log \frac{P(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}|\theta)}{Q(\mathbf{x}_{1:T})} d\mathbf{x}_{1:T} \\ &= \int Q(\mathbf{x}_{1:T}) \log P(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}|\theta) d\mathbf{x}_{1:T} \\ &\quad - \int Q(\mathbf{x}_{1:T}) \log Q(\mathbf{x}_{1:T}) d\mathbf{x}_{1:T} \end{aligned} \quad (\text{A.4})$$

の様に変形でき, 右辺第二項は θ に依存しないので, $Q(\mathbf{x}_{1:T})$ が固定されている時, \mathcal{F} の θ についての最大化は期待対数尤度

$$\mathcal{L}(\theta) = \int Q(\mathbf{x}_{1:T}) \log P(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}|\theta) d\mathbf{x}_{1:T} \quad (\text{A.5})$$

を θ について最大にすれば良い.

$$\begin{aligned} \mathcal{L}(\theta) &= \int Q(\mathbf{x}_{1:T}) \log P(\mathbf{z}_{1:T}|\mathbf{x}_{1:T}, \theta) P(\mathbf{x}_{1:T}|\theta) d\mathbf{x}_{1:T} \\ &= \int Q(\mathbf{x}_{1:T}) \log P(\mathbf{z}_{1:T}|\mathbf{x}_{1:T}, \theta) d\mathbf{x}_{1:T} \\ &\quad + \int Q(\mathbf{x}_{1:T}) \log P(\mathbf{x}_{1:T}|\theta) d\mathbf{x}_{1:T} \\ &= \langle \log P(\mathbf{z}_{1:T}|\mathbf{x}_{1:T}, \theta) \rangle_Q + \langle \log P(\mathbf{x}_{1:T}|\theta) \rangle_Q \end{aligned} \quad (\text{A.6})$$

のように書け, この内ダイナミクスモデルのパラメータに関係するのは右辺第二項のみであるから, これを \mathcal{L}_D とすると,

$$\mathcal{L}_D = \langle P(\mathbf{x}_{1:T}|\theta) \rangle_Q \quad (\text{A.7})$$

となる．ここで，

$$\langle f(\mathbf{x}_{1:T}) \rangle = \int Q(\mathbf{x}_{1:T}) f(\mathbf{x}_{1:T}) d\mathbf{x}_{1:T} \quad (\text{A.8})$$

である．ダイナミクスモデルとして

$$\mathbf{x}_{t+2} = A_1 \mathbf{x}_{t+1} + A_0 \mathbf{x}_t + \mathbf{d} + B \mathbf{u}_t \quad (\text{A.9})$$

のようにバイアス項を含む線形二次方程式を仮定する．ここでプロセスノイズ \mathbf{u}_t は白色ガウスノイズとする．このとき

$$\mathbf{e}_t = \mathbf{x}_{t+2} - (A_1 \mathbf{x}_{t+1} + A_0 \mathbf{x}_t + \mathbf{d}) \quad (\text{A.10})$$

が平均 0，分散 $C = (BB^T)$ のガウシアンに従うのでダイナミクスのマルコフ性より，

$$P(\mathbf{x}_{1:T}|\theta) = \prod_{t=1}^{T-2} \left\{ \frac{1}{\sqrt{2\pi}} C^{-\frac{1}{2}} \exp\left(-\frac{1}{2} C^{-1} \mathbf{e}_t \mathbf{e}_t^T\right) \right\} \quad (\text{A.11})$$

のように書ける．よって

$$\begin{aligned} \mathcal{L}_D(\theta) &= \left\langle \frac{1}{2}(T-2) \log |C^{-1}| \right\rangle_Q \\ &\quad - \left\langle \frac{1}{2} \sum_{t=1}^{T-2} \text{tr}(C^{-1} \mathbf{e}_t \mathbf{e}_t^T) \right\rangle_Q \\ &\quad - \left\langle \frac{1}{2}(T-2) \log 2\pi \right\rangle_Q \end{aligned} \quad (\text{A.12})$$

ここで，

$$\begin{aligned} \text{tr}(\mathbf{e}_t \mathbf{e}_t^T) &= \mathbf{x}_{t+2}^T \mathbf{x}_{t+2} - \mathbf{x}_{t+2}^T (A_1 \mathbf{x}_{t+1} + A_0 \mathbf{x}_t + \mathbf{d}) \\ &\quad - (A_1 \mathbf{x}_{t+1} + A_0 \mathbf{x}_t + \mathbf{d})^T \mathbf{x}_{t+2} \\ &\quad + (A_1 \mathbf{x}_{t+1} + A_0 \mathbf{x}_t + \mathbf{d})^T (A_1 \mathbf{x}_{t+1} + A_0 \mathbf{x}_t + \mathbf{d}) \\ &= \mathbf{x}_{t+2}^T \mathbf{x}_{t+2} + \mathbf{d}^T \mathbf{d} \\ &\quad - \mathbf{x}_{t+2}^T \mathbf{x}_{t+1}^T A_1^T - \mathbf{x}_{t+2}^T \mathbf{x}_t^T A_0^T - \mathbf{x}_{t+2}^T \mathbf{d}^T \\ &\quad - A_1 \mathbf{x}_{t+1} \mathbf{x}_{t+2}^T + A_0 \mathbf{x}_t \mathbf{x}_{t+2}^T + \mathbf{d} \mathbf{x}_{t+2}^T \\ &\quad + A_1 \mathbf{x}_{t+1} \mathbf{x}_{t+1}^T A_1^T + A_1 \mathbf{x}_{t+1} \mathbf{x}_t^T A_0^T + A_1 \mathbf{x}_{t+1} \mathbf{d}^T \\ &\quad + A_0 \mathbf{x}_t \mathbf{x}_{t+1}^T A_1^T + A_0 \mathbf{x}_t \mathbf{x}_t^T A_0^T + A_0 \mathbf{x}_t \mathbf{d}^T \\ &\quad + \mathbf{d} \mathbf{x}_{t+1}^T A_1^T + \mathbf{d} \mathbf{x}_t^T A_0^T \end{aligned} \quad (\text{A.13})$$

であるから

$$S_i = \sum_{t=1}^{T-2} \mathbf{x}_{t+i}, \quad S_{ij} = \sum_{t=1}^{T-2} \mathbf{x}_{t+i} \mathbf{x}_{t+j}^T \quad (\text{A.14})$$

とすると,

$$\begin{aligned} \mathcal{L}_D(\theta) = & \frac{1}{2}(T-2) \log |C^{-1}| - \frac{1}{2}(T-2) \log 2\pi \\ & - \frac{1}{2} C^{-1} \left\langle \left(S_{22} + (T-2) \mathbf{d} \mathbf{d}^T \right. \right. \\ & - S_{21} \mathbf{A}_1^T - S_{20} \mathbf{A}_0^T - S_2 \mathbf{d}^T \\ & - A_1 S_{12} + A_0 S_{02} + \mathbf{d} S_2^T \\ & + A_1 S_{11} \mathbf{A}_1^T + A_1 S_{10} \mathbf{A}_0^T + A_1 S_1 \mathbf{d}^T \\ & + A_0 S_{01} \mathbf{A}_1^T + A_0 S_{00} \mathbf{A}_0^T + A_0 S_0 \mathbf{d}^T \\ & \left. \left. + \mathbf{d} S_1^T \mathbf{A}_1^T + \mathbf{d} S_0^T \mathbf{A}_0^T \right) \right\rangle_Q \end{aligned} \quad (\text{A.15})$$

各パラメータについて

$$\frac{\partial \mathcal{L}_D}{\partial \theta} = 0 \quad (\text{A.16})$$

を解くと

$$\langle S_{21} \rangle_Q - A_0 \langle S_{01} \rangle_Q - A_1 \langle S_{11} \rangle_Q = \mathbf{d} \langle S_1^T \rangle_Q \quad (\text{A.17})$$

$$\langle S_{20} \rangle_Q - A_0 \langle S_{00} \rangle_Q - A_1 \langle S_{10} \rangle_Q = \mathbf{d} \langle S_0^T \rangle_Q \quad (\text{A.18})$$

$$\langle S_2 \rangle_Q - A_0 \langle S_0 \rangle_Q - A_1 \langle S_1 \rangle_Q - (T-2) \mathbf{d} = 0 \quad (\text{A.19})$$

$$C(T-2) - \sum_{t=1}^{T-2} \langle \mathbf{e}_t \mathbf{e}_t^T \rangle_Q = 0 \quad (\text{A.20})$$

が得られる.

$\bar{\theta}$ を E ステップで固定したパラメータとすると

$$Q(\mathbf{x}_{1:T}) = P(\mathbf{x}_{1:T} | \mathbf{z}_{1:T}, \bar{\theta}) \quad (\text{A.21})$$

であるから, M ステップで必要な統計量はスムージングを適用することにより得ることができる.

C. EM アルゴリズムを用いた混合比の推定

今,

$$\begin{aligned} q(\mathbf{x}_t) &= p(\mathbf{x}_{0:t} | \mathbf{a}_{0:t}, \mathbf{z}_{1:t}) & (A.22) \\ \pi(\mathbf{x}_t | \alpha_t) &= \alpha_{a,t} p(\mathbf{x}_{0:t} | \mathbf{a}_{0:t}) + \alpha_{z,t} p(\mathbf{x}_{0:t} | \mathbf{z}_{1:t}) \end{aligned}$$

但し,

$$\alpha_{a,t} + \alpha_{z,t} = 1 \quad (A.23)$$

とすると, $q(\mathbf{x}_t)$ と $\pi(\mathbf{x}_t | \alpha_t)$ の KL 距離を最小とするような α_t を求め, この α_t を用いて次時刻の予測分布 $p(\mathbf{x}_{0:t+1} | \mathbf{a}_{0:t+1}, \mathbf{z}_{1:t})$ を

$$p(\mathbf{x}_{0:t+1} | \mathbf{a}_{0:t+1}, \mathbf{z}_{1:t}) \approx \alpha_{a,t} p(\mathbf{x}_{0:t+1} | \mathbf{a}_{0:t+1}) + \alpha_{z,t} p(\mathbf{x}_{0:t+1} | \mathbf{z}_{1:t})$$

の様に推定することを考える. 今, $q(\mathbf{x}_t)$ と $\pi(\mathbf{x}_t | \alpha_t)$ の KL 距離は

$$\begin{aligned} \text{KL}(q(\mathbf{x}_t) || \pi(\mathbf{x}_t | \alpha_t)) &= \int q(\mathbf{x}_t) \log \left(\frac{q(\mathbf{x}_t)}{\pi(\mathbf{x}_t | \alpha_t)} \right) d\mathbf{x}_t \\ &= \int q(\mathbf{x}_t) \log q(\mathbf{x}_t) d\mathbf{x}_t - \int q(\mathbf{x}_t) \log \pi(\mathbf{x}_t | \alpha_t) d\mathbf{x}_t \quad (A.24) \end{aligned}$$

のようになる. 式 (A.24) の右辺第一項は α_t に無関係であるため, 式 (A.24) の右辺第二項を最大にする α_t は $q(\mathbf{x}_t)$ と $\pi(\mathbf{x}_t | \alpha_t)$ の KL 距離を最小にする. ここで, $q(\mathbf{x}_t)$ から得られたサンプル $\mathbf{x}_t^{(i)}, i = 1, \dots, N$ が十分あれば, 式 (A.24) 右辺第二項の積分は以下の様に置き換えることが出来る.

$$\int q(\mathbf{x}_t) \log \pi(\mathbf{x}_t | \alpha_t) d\mathbf{x}_t \approx \frac{1}{N} \sum_{i=1}^N \log \pi(\mathbf{x}_t^{(i)} | \alpha_t)$$

よって

$$\begin{aligned} \mathcal{L}(\alpha_t) &= \sum_{i=1}^N \log \pi(\mathbf{x}_t^{(i)} | \alpha_t) \\ &= \sum_{i=1}^N \log \left(\alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) + \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t}) \right) \end{aligned}$$

とすると、これを最大にする α_t を求めればよい。但し、 α_t には式 (A.23) の様な制約があるので、ラグランジュの未定乗数を用いて、停留条件は

$$\begin{aligned}
\frac{\partial \mathcal{L}(\alpha_t)}{\partial \alpha_t} &= \frac{\partial}{\partial \alpha_t} \left(\sum_{i=1}^N \log (\alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) + \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})) - \kappa(\alpha_{a,t} + \alpha_{z,t} - 1) \right) \\
&= \sum_{i=1}^N \frac{1}{\alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) + \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})} \frac{\partial}{\partial \alpha_t} (\alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) + \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})) \\
&\quad - \frac{\partial}{\partial \alpha_t} \kappa(\alpha_{a,t} + \alpha_{z,t} - 1) \\
&= \sum_{i=1}^N \frac{\alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t})}{\alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) + \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})} \frac{\partial}{\partial \alpha_t} \log \alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) \\
&\quad + \sum_{i=1}^N \frac{\alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})}{\alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) + \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})} \frac{\partial}{\partial \alpha_t} \log \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t}) \\
&\quad - \frac{\partial}{\partial \alpha_t} \kappa(\alpha_{a,t} + \alpha_{z,t} - 1) = 0 \tag{A.25}
\end{aligned}$$

式 (A.25) は非線形になるため、ここでは EM アルゴリズムを用いてこれを満たす α_t を求める。

E-step 現在のパラメータを $\bar{\alpha}_t$ とすると、これを用いて、各サンプル $\mathbf{x}_{0:t}^{(i)}$ に対して

$$p(m_i = a | \mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) = \frac{\bar{\alpha}_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t})}{\bar{\alpha}_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) + \bar{\alpha}_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})} \tag{A.26}$$

$$p(m_i = z | \mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) = \frac{\bar{\alpha}_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})}{\bar{\alpha}_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) + \bar{\alpha}_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})} \tag{A.27}$$

を計算する。但し、 m はサンプル $\mathbf{x}_{0:t}^{(i)}$ が $p(\mathbf{x}_{0:t} | \mathbf{a}_{0:t})$ と $p(\mathbf{x}_{0:t} | \mathbf{z}_{1:t})$ のどちらからサンプルされたかを示すインデックスである（隠れ変数）。これを用いて、対数尤度の隠れ変数の予測事後分布についての期待値 $Q(\alpha_t | \bar{\alpha}_t)$ を以下の様に求める。

$$\begin{aligned}
Q(\alpha_t | \bar{\alpha}_t) &= \sum_{i=1}^N p(m_i = a | \mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) \log \alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{a}_{0:t}) \\
&\quad + \sum_{i=1}^N p(m_i = z | \mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) \log \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)} | \mathbf{z}_{1:t})
\end{aligned}$$

M-step 期待対数尤度 $Q(\alpha_t|\bar{\alpha}_t)$ を式 (A.23) の制約の下, α_t について最大化する. 停留条件は

$$\begin{aligned}\frac{\partial Q(\alpha_t|\bar{\alpha}_t)}{\partial \alpha_t} &= \sum_{i=1}^N p(m_i = a|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) \frac{\partial}{\partial \alpha_t} \log \alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)}|\mathbf{a}_{0:t}) \\ &\quad + \sum_{i=1}^N p(m_i = z|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) \frac{\partial}{\partial \alpha_t} \log \alpha_{z,t} p(\mathbf{x}_{0:t}^{(i)}|\mathbf{z}_{1:t}) \\ &\quad - \frac{\partial}{\partial \alpha_t} \kappa(\alpha_{a,t} + \alpha_{z,t} - 1) = 0\end{aligned}$$

の様になり, これを満たすパラメータ α_t を新たに $\bar{\alpha}_t$ として E-step に戻る. この場合は

$$\begin{aligned}\frac{\partial Q(\alpha_t|\bar{\alpha}_t)}{\partial \alpha_{a,t}} &= \sum_{i=1}^N p(m_i = a|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) \frac{p(\mathbf{x}_{0:t}^{(i)}|\mathbf{a}_{0:t})}{\alpha_{a,t} p(\mathbf{x}_{0:t}^{(i)}|\mathbf{a}_{0:t})} - \kappa \\ &= \frac{1}{\alpha_{a,t}} \sum_{i=1}^N p(m_i = a|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) - \kappa = 0 \\ \frac{\partial Q(\alpha_t|\bar{\alpha}_t)}{\partial \alpha_{z,t}} &= \frac{1}{\alpha_{z,t}} \sum_{i=1}^N p(m_i = z|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) - \kappa = 0 \\ \frac{\partial Q(\alpha_t|\bar{\alpha}_t)}{\partial \kappa} &= \alpha_{a,t} + \alpha_{z,t} - 1 = 0\end{aligned}$$

これを解いて

$$\begin{aligned}\alpha_{a,t} &= \frac{\sum_{i=1}^N p(m_i = a|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t)}{\sum_{i=1}^N p(m_i = a|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) + \sum_{i=1}^N p(m_i = z|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t)} \\ \alpha_{z,t} &= \frac{\sum_{i=1}^N p(m_i = z|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t)}{\sum_{i=1}^N p(m_i = a|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) + \sum_{i=1}^N p(m_i = z|\mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t)}\end{aligned}$$

を得る.

但し, 上記の手法はバッチ形式であるため, これをオンライン化し逐次的に得られるデータに対して適用できる形に拡張する. 本手法では各時刻に得られる N 個の

サンプル $\mathbf{x}_{0:t}^{(i)}$ を用いて α_t を更新するため,

$$\sum_{i=1}^N p(m_i = a | \mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) = (1 - \eta_t) \left(\sum_{i=1}^N p(m_i = a | \mathbf{x}_{0:t-1}^{(i)}, \bar{\alpha}_{t-1}) \right) + \eta_t \left(\sum_{i=1}^N p(m_i = a | \mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) \right) \quad (\text{A.28})$$

$$\sum_{i=1}^N p(m_i = z | \mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) = (1 - \eta_t) \left(\sum_{i=1}^N p(m_i = z | \mathbf{x}_{0:t-1}^{(i)}, \bar{\alpha}_{t-1}) \right) + \eta_t \left(\sum_{i=1}^N p(m_i = z | \mathbf{x}_{0:t}^{(i)}, \bar{\alpha}_t) \right) \quad (\text{A.29})$$

の様に修正することで, 各時刻で得られたサンプルを用いて逐次的に α_t の推定を行うことができる. 但し,

$$\eta_t = \frac{1}{1 + \frac{\lambda_t}{\eta_{t-1}}} \quad (\text{A.30})$$

は正規化のための定数であり, 学習係数の様な役割を果たす. ここで, λ_t は忘却係数であり, $0 \leq \lambda_t \leq 1$ とする.

D. アフィン変換パラメータからの頭部姿勢の復元

D.1 パーティクルフィルタによる推定

上位層で推定したアフィンカメラ行列 $P_{a,t}$ から頭部姿勢 $\theta_t = (\theta_X, \theta_Y, \theta_Z)^T$ を推定する. 初期姿勢を $\theta_0 = (0, 0, 0)^T$ とすると, 各時刻における $P_{a,t}$ は頭部姿勢 θ_t により得られる回転行列 $R(\theta_t)$ を用いて

$$P_{a,t} = P_{a,0} R(\theta_t) X_0 \quad (\text{A.31})$$

と表すことができる．ここで X_0 は顔特徴点の初期三次元位置であり，回転行列 $R(\theta_t)$ は $Y \rightarrow X \rightarrow Z$ 軸の順で回転させたとする，

$$R(\theta_t) = R_Y R_X R_Z$$

$$= \begin{pmatrix} \cos \theta_{Y,t} & 0 & \sin \theta_{Y,t} \\ 0 & 1 & 0 \\ -\sin \theta_{Y,t} & 0 & \cos \theta_{Y,t} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{X,t} & -\sin \theta_{X,t} \\ 0 & \sin \theta_{X,t} & \cos \theta_{X,t} \end{pmatrix} \begin{pmatrix} \cos \theta_{Z,t} & -\sin \theta_{Z,t} & 0 \\ \sin \theta_{Z,t} & \cos \theta_{Z,t} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{A.32})$$

と書くことができる．この回転行列により式 (A.31) は θ_t に関して非線形となるため，ここではパーティクルフィルタを用いてこれを推定することを考える．頭部姿勢を示すサンプルを $\theta_t^{(i)}$ ($i = 1, \dots, N_\theta$) とする．この時，状態遷移モデルは以下で書くことができる．

$$\theta_t^{(i)} = A_\theta \theta_{t-1}^{(i)} + B_\theta \mathbf{u}_{\theta,t}$$

ここで， $\mathbf{u}_{\theta,t}$ は白色ガウス雑音とする．観測過程は $\mathbf{v}_{\theta,t}$ を白色ガウス雑音とすると，

$$P_{a,t}^{(i)} = P_{a,0} R(\theta_t^{(i)}) X_0 + D_\theta \mathbf{v}_{\theta,t}$$

で書けるので，各パーティクルの重み $w_{\theta,t}^{(i)}$ は上位で推定されたアフィン変換行列の期待値 \hat{a}_t と分散 $\hat{V}_{a,t}$ を用いて

$$w_{\theta,t}^{(i)} \propto \exp \left(\sum_{j=1}^8 -\frac{(a_{j,t}^{(i)} - \hat{a}_{j,t})^2}{\sigma_{a_{j,t}}^2} \right)$$

で得ることができる．

D.2 拡張カルマンフィルタを用いた手法

前節ではパーティクルフィルタを用いて頭部姿勢 θ_t の推定を行う手法について説明したが，実際の車内におけるドライバの頭部姿勢推定には，計算量の観点から適用困難な場合がある．その様な場合には以下の様に非線形な観測過程をテイ

ラー展開により線形で近似することでカルマンフィルタを用いた頭部姿勢推定を行うことが可能である．式 (A.32) より回転行列 $R(\theta_t)$ を

$$R(\theta_t) = \begin{pmatrix} r_{1,t} & r_{2,t} & r_{3,t} \\ r_{4,t} & r_{5,t} & r_{6,t} \\ r_{7,t} & r_{8,t} & r_{9,t} \end{pmatrix}$$

の様に書けば，

$$\begin{aligned} r_{1,t} &= \cos \theta_{Y,t} \cos \theta_{Z,t} + \sin \theta_{Y,t} \sin \theta_{X,t} \sin \theta_{Z,t} \\ r_{2,t} &= -\cos \theta_{Y,t} \sin \theta_{Z,t} + \sin \theta_{Y,t} \sin \theta_{X,t} \cos \theta_{Z,t} \\ r_{3,t} &= \sin \theta_{Y,t} \cos \theta_{X,t} \\ r_{4,t} &= \cos \theta_{X,t} \sin \theta_{Z,t} \\ r_{5,t} &= \cos \theta_{X,t} \cos \theta_{Z,t} \\ r_{6,t} &= -\sin \theta_{X,t} \\ r_{7,t} &= -\sin \theta_{Y,t} \cos \theta_{Z,t} + \cos \theta_{Y,t} \sin \theta_{X,t} \sin \theta_{Z,t} \\ r_{8,t} &= \sin \theta_{Y,t} \sin \theta_{Z,t} + \cos \theta_{Y,t} \sin \theta_{X,t} \cos \theta_{Z,t} \\ r_{9,t} &= \cos \theta_{Y,t} \cos \theta_{X,t} \end{aligned}$$

である．各要素について θ_{t-1} の周りでテイラー展開を行えば， $r_{1,t}$ について，

$$\begin{aligned} r_{1,t} &= \cos \theta_{Y,t} \cos \theta_{Z,t} + \sin \theta_{Y,t} \sin \theta_{X,t} \sin \theta_{Z,t} \\ &\approx r_{1,t-1} + (\theta_{X,t} - \theta_{X,t-1}) \frac{\partial}{\partial \theta_X} r_{1,t-1} + (\theta_{Y,t} - \theta_{Y,t-1}) \frac{\partial}{\partial \theta_Y} r_{1,t-1} + (\theta_{Z,t} - \theta_{Z,t-1}) \frac{\partial}{\partial \theta_Z} r_{1,t-1} \\ &= \cos \theta_{Y,t-1} \cos \theta_{Z,t-1} + \sin \theta_{Y,t-1} \sin \theta_{X,t-1} \sin \theta_{Z,t-1} \\ &\quad + (\theta_{X,t} - \theta_{X,t-1}) \sin \theta_{Y,t-1} \sin \theta_{Z,t-1} \cos \theta_{X,t-1} \\ &\quad + (\theta_{Y,t} - \theta_{Y,t-1}) (-\cos \theta_{Z,t-1} \sin \theta_{Y,t-1} + \sin \theta_{Y,t-1} \sin \theta_{Z,t-1} \cos \theta_{X,t-1}) \\ &\quad + (\theta_{Z,t} - \theta_{Z,t-1}) (-\cos \theta_{Y,t-1} \sin \theta_{Z,t-1} + \sin \theta_{Y,t-1} \sin \theta_{X,t-1} \cos \theta_{Z,t-1}) \\ &= f_{1X} \theta_{X,t} + f_{1Y} \theta_{Y,t} + f_{1Z} \theta_{Z,t} + f_{1c} \end{aligned}$$

と書くことができる．ここで

$$f_{1X} = \sin \theta_{Y,t-1} \sin \theta_{Z,t-1} \cos \theta_{X,t-1}$$

$$f_{1Y} = -\cos \theta_{Z,t-1} \sin \theta_{Y,t-1} + \sin \theta_{X,t-1} \sin \theta_{Z,t-1} \cos \theta_{Y,t-1}$$

$$f_{1Z} = -\cos \theta_{Y,t-1} \sin \theta_{Z,t-1} + \sin \theta_{Y,t-1} \sin \theta_{X,t-1} \cos \theta_{Z,t-1}$$

$$f_{1c} = \cos \theta_{Y,t-1} \cos \theta_{Z,t-1} + \sin \theta_{Y,t-1} \sin \theta_{X,t-1} \sin \theta_{Z,t-1} - f_{1X}\theta_{X,t-1} - f_{1Y}\theta_{Y,t-1} - f_{1Z}\theta_{Z,t-1}$$

である．同様にして，

$$r_{2,t} = f_{2X}\theta_{X,t} + f_{2Y}\theta_{Y,t} + f_{2Z}\theta_{Z,t} + f_{2c}$$

$$f_{2X} = \sin \theta_{Y,t-1} \cos \theta_{Z,t-1} \cos \theta_{X,t-1}$$

$$f_{2Y} = \sin \theta_{Z,t-1} \sin \theta_{Y,t-1} + \sin \theta_{X,t-1} \cos \theta_{Z,t-1} \cos \theta_{Y,t-1}$$

$$f_{2Z} = -\cos \theta_{Y,t-1} \cos \theta_{Z,t-1} - \sin \theta_{Y,t-1} \sin \theta_{X,t-1} \sin \theta_{Z,t-1}$$

$$f_{2c} = -\cos \theta_{Y,t-1} \sin \theta_{Z,t-1} + \sin \theta_{Y,t-1} \sin \theta_{X,t-1} \cos \theta_{Z,t-1} - f_{2X}\theta_{X,t-1} - f_{2Y}\theta_{Y,t-1} - f_{2Z}\theta_{Z,t-1}$$

$$r_{3,t} = f_{3X}\theta_{X,t} + f_{3Y}\theta_{Y,t} + f_{3c}$$

$$f_{3X} = -\sin \theta_{Y,t-1} \sin \theta_{X,t-1}$$

$$f_{3Y} = \cos \theta_{X,t-1} \cos \theta_{Y,t-1}$$

$$f_{3c} = \sin \theta_{Y,t-1} \cos \theta_{X,t-1} - f_{3X}\theta_{X,t-1} - f_{3Y}\theta_{Y,t-1}$$

$$r_{4,t} = f_{4X}\theta_{X,t} + f_{4Z}\theta_{Z,t} + f_{4c}$$

$$f_{4X} = -\sin \theta_{Z,t-1} \sin \theta_{X,t-1}$$

$$f_{4Z} = \cos \theta_{X,t-1} \cos \theta_{Z,t-1}$$

$$f_{4c} = \cos \theta_{X,t-1} \sin \theta_{Z,t-1} - f_{4X}\theta_{X,t-1} - f_{4Z}\theta_{Z,t-1}$$

$$r_{5,t} = f_{5X}\theta_{X,t} + f_{5Z}\theta_{Z,t} + f_{5c}$$

$$f_{5X} = -\cos \theta_{Z,t-1} \sin \theta_{X,t-1}$$

$$f_{5Z} = -\cos \theta_{X,t-1} \sin \theta_{Z,t-1}$$

$$f_{5c} = \cos \theta_{X,t-1} \cos \theta_{Z,t-1} - f_{5X}\theta_{X,t-1} - f_{5Z}\theta_{Z,t-1}$$

$$\begin{aligned}
r_{6,t} &= f_{6X}\theta_{X,t} + f_{6c} \\
f_{6X} &= -\cos\theta_{X,t-1} \\
f_{6c} &= -\sin\theta_{X,t-1} - f_{6X}\theta_{X,t-1}
\end{aligned}$$

$$\begin{aligned}
r_{7,t} &= f_{7X}\theta_{X,t} + f_{7Y}\theta_{Y,t} + f_{7Z}\theta_{Z,t} + f_{7c} \\
f_{7X} &= \cos\theta_{Y,t-1} \sin\theta_{Z,t-1} \cos\theta_{X,t-1} \\
f_{7Y} &= -\cos\theta_{Z,t-1} \cos\theta_{Y,t-1} - \sin\theta_{X,t-1} \sin\theta_{Z,t-1} \sin\theta_{Y,t-1} \\
f_{7Z} &= \sin\theta_{Y,t-1} \sin\theta_{Z,t-1} + \cos\theta_{Y,t-1} \sin\theta_{X,t-1} \cos\theta_{Z,t-1} \\
f_{7c} &= -\sin\theta_{Y,t-1} \cos\theta_{Z,t-1} + \cos\theta_{Y,t-1} \sin\theta_{X,t-1} \sin\theta_{Z,t-1} - f_{7X}\theta_{X,t-1} - f_{7Y}\theta_{Y,t-1} - f_{7Z}\theta_{Z,t-1}
\end{aligned}$$

$$\begin{aligned}
r_{8,t} &= f_{8X}\theta_{X,t} + f_{8Y}\theta_{Y,t} + f_{8Z}\theta_{Z,t} + f_{8c} \\
f_{8X} &= \cos\theta_{Y,t-1} \cos\theta_{Z,t-1} \cos\theta_{X,t-1} \\
f_{8Y} &= \sin\theta_{Z,t-1} \cos\theta_{Y,t-1} - \sin\theta_{X,t-1} \cos\theta_{Z,t-1} \sin\theta_{Y,t-1} \\
f_{8Z} &= \sin\theta_{Y,t-1} \cos\theta_{Z,t-1} - \cos\theta_{Y,t-1} \sin\theta_{X,t-1} \sin\theta_{Z,t-1} \\
f_{8c} &= \sin\theta_{Y,t-1} \sin\theta_{Z,t-1} + \cos\theta_{Y,t-1} \sin\theta_{X,t-1} \cos\theta_{Z,t-1} - f_{8X}\theta_{X,t-1} - f_{8Y}\theta_{Y,t-1} - f_{8Z}\theta_{Z,t-1}
\end{aligned}$$

$$\begin{aligned}
r_{9,t} &= f_{9X}\theta_{X,t} + f_{9Y}\theta_{Y,t} + f_{9Z}\theta_{Z,t} + f_{9c} \\
f_{9X} &= -\cos\theta_{Y,t-1} \sin\theta_{X,t-1} \\
f_{9Y} &= -\cos\theta_{X,t-1} \sin\theta_{Y,t-1} \\
f_{9c} &= \cos\theta_{Y,t-1} \cos\theta_{X,t-1} - f_{9X}\theta_{X,t-1} - f_{9Y}\theta_{Y,t-1}
\end{aligned}$$

と書くことができ，これらを用いれば，観測過程は以下の様に線形の形で近似することが可能である．

$$\mathbf{a}_t = C_{\Theta,t} \Theta_t + D_{\Theta,t} \mathbf{v}_{\Theta,t} \tag{A.33}$$

ここで ,

$$C_{\Theta,t} = \begin{pmatrix} a_{1,0} & 0 & 0 & a_{2,0} & 0 & 0 & a_{3,0} & 0 & 0 & 0 \\ 0 & a_{1,0} & 0 & 0 & a_{2,0} & 0 & 0 & a_{3,0} & 0 & 0 \\ 0 & 0 & a_{1,0} & 0 & 0 & a_{2,0} & 0 & 0 & a_{3,0} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{4,0} \\ a_{5,0} & 0 & 0 & a_{6,0} & 0 & 0 & a_{7,0} & 0 & 0 & 0 \\ 0 & a_{5,0} & 0 & 0 & a_{6,0} & 0 & 0 & a_{7,0} & 0 & 0 \\ 0 & 0 & a_{5,0} & 0 & 0 & a_{6,0} & 0 & 0 & a_{7,0} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{8,0} \end{pmatrix} \begin{pmatrix} f_{1X} & f_{1Y} & f_{1Z} & f_{1c} \\ f_{2X} & f_{2Y} & f_{2Z} & f_{2c} \\ f_{3X} & f_{3Y} & 0 & f_{3c} \\ f_{4X} & 0 & f_{4Z} & f_{4c} \\ f_{5X} & 0 & f_{5Z} & f_{5c} \\ f_{6X} & 0 & f_{6Z} & f_{6c} \\ f_{7X} & f_{7Y} & f_{7Z} & f_{7c} \\ f_{8X} & f_{8Y} & f_{8Z} & f_{8c} \\ f_{9X} & f_{9Y} & 0 & f_{9c} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.34})$$

であり , $\Theta_t = (\theta_t^T, 1)^T$ である . Θ_t の状態遷移過程は , $u_{\Theta,t}$ を白色ガウス雑音として , 以下で書くことができる .

$$\Theta_t = A_{\Theta} \Theta_{t-1} + B_{\Theta} u_{\Theta,t}$$

この様に観測過程を線形近似することにより , 頭部姿勢の事後分布 $p(\Theta_t | z_{1:t})$ はカルマンフィルタを用いて以下で推定することができる .

$$p(\Theta_t | z_{1:t}) = \mathcal{N}(\Theta_t | \hat{\Theta}_t, \hat{V}_{\Theta,t}) \quad (\text{A.35})$$

ここで ,

$$\begin{aligned} \hat{\Theta}_t &= \bar{\Theta}_t + K_t(\hat{a}_t - C_{\Theta,t} \bar{\Theta}_t) \\ \hat{V}_{\Theta,t} &= (I - K_t C_{\Theta,t}) \bar{V}_{\Theta,t} \\ K_t &= \bar{V}_{\Theta,t} C_{\Theta,t}^T (D_{h,t} D_{h,t}^T + C_{\Theta,t} \bar{V}_{\Theta,t} C_{\Theta,t}^T)^{-1} \end{aligned}$$

である . また , これを用いて Θ_{t+1} の事前分布 $p(\Theta_{t+1} | z_{0:t})$ は以下で与えられる .

$$p(\Theta_{t+1} | z_{1:t}) = \mathcal{N}(\Theta_{t+1} | \bar{\Theta}_{t+1}, \bar{V}_{\Theta,t+1}) \quad (\text{A.36})$$

ただし,

$$\begin{aligned}\bar{\Theta}_{t+1} &= A_{\Theta} \hat{\theta}_t \\ \bar{V}_{\Theta,t+1} &= B_{\Theta} B_{\Theta}^T + A_{\Theta} \hat{V}_{\Theta,t} A_{\Theta}^T\end{aligned}$$

である.

E. テイラー展開を用いた射影行列の逐次近似

姿勢ベクトルを $\theta_t^{(rot)} = (\theta_{x,t}, \theta_{y,t}, \theta_{z,t})^T$, 並進ベクトルを $\theta_t^{(ms)} = (\tau_{x,t}, \tau_{y,t}, \tau_{z,t})^T$ とし, $\theta_t = (\theta_t^{(rot)T}, \theta_t^{(ms)T})^T$ とすれば, 状態遷移過程は

$$\theta_{t+1} = A_{\theta} \theta_t + B_{\theta} u_t, \quad u_t \sim \mathcal{N}_{N_{\theta}}(\mathbf{0}, I)$$

で書け, 観測過程は

$$\begin{aligned}\tilde{x}_t &\simeq C_{X,t} \tilde{X}_0 + D_X v_t \\ &= C_{X,t}^{(int)} C_{X,t}^{(ext)} \tilde{X}_0 + D_X v_t \\ &= \begin{pmatrix} f k_x & f k_x \cot \phi & x_0 \\ 0 & f k_y / \sin \phi & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_t(\theta_t^{(rot)}) & \theta_t^{(ms)} \end{pmatrix} \tilde{X}_0 + D_X v_t \quad v_t \sim \mathcal{N}_3(\mathbf{0}, I)\end{aligned}$$

で書くことができる. ここで, (x_0, y_0) は画像中心, ϕ は x, y 軸のなす角, (k_x, k_y) は 1 画素の大きさの逆数, f は焦点距離, $R_t(\theta_t^{(rot)})$ は回転行列とする. 簡単のためにピント調整等を行わないとすれば, $f, \phi, (k_x, k_y)$ は定数であるので, $C_{X,t}^{(int)} = C_X^{(int)}$ と内部パラメータ行列を時不変であると考えることができる. また, 簡単のために以降は $(x_0, y_0) = (0, 0), (k_x, k_y) = (1, 1)$ とする.

以上のように観測過程は回転行列 $R_t(\theta_t^{(rot)})$ が $\theta_t^{(rot)}$ に関して非線形であるため, これに直接カルマンフィルタを適用することは出来ない. そのため, 以下の手順でこれを線形化し, 拡張カルマンフィルタを用いて θ_t の時系列推定を行う.

X_t に関して線形化 x_t の各成分について書き下せば ,

$$\mathbf{x}_t = \begin{pmatrix} f \frac{X_t}{Z_t} \\ f \frac{Y_t}{Z_t} \end{pmatrix} + D_X \mathbf{v}_t$$

$(f \frac{X_t}{Z_t}, f \frac{Y_t}{Z_t})$ をそれぞれ X_{t-1} の周りでテイラー展開し , 2 次以上の項を無視すれば ,

$$\begin{aligned} f \frac{X_t}{Z_t} &\approx \frac{f}{Z_{t-1}} X_{t-1} + (X_t - X_{t-1}) \frac{f}{Z_{t-1}} - (Z_t - Z_{t-1}) \frac{f}{Z_{t-1}^2} X_{t-1} \\ &= \frac{f}{Z_{t-1}} X_t + \frac{f}{Z_{t-1}} X_{t-1} - \frac{f}{Z_{t-1}^2} X_{t-1} Z_t \\ f \frac{Y_t}{Z_t} &\approx \frac{f}{Z_{t-1}} Y_{t-1} + (Y_t - Y_{t-1}) \frac{f}{Z_{t-1}} - (Z_t - Z_{t-1}) \frac{f}{Z_{t-1}^2} Y_{t-1} \\ &= \frac{f}{Z_{t-1}} Y_t + \frac{f}{Z_{t-1}} Y_{t-1} - \frac{f}{Z_{t-1}^2} Y_{t-1} Z_t \end{aligned}$$

と書けるので , 観測過程は以下の形で近似できる .

$$\tilde{\mathbf{x}}_t = \begin{pmatrix} \frac{f}{Z_{t-1}} & 0 & -\frac{f}{Z_{t-1}^2} X_{t-1} & \frac{f}{Z_{t-1}} X_{t-1} \\ 0 & \frac{f}{Z_{t-1}} & -\frac{f}{Z_{t-1}^2} Y_{t-1} & \frac{f}{Z_{t-1}} Y_{t-1} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tilde{\mathbf{X}}_t + D_X \mathbf{v}_t \quad (\text{A.37})$$

θ_t に関して線形化 前パラグラフでは観測過程を X_t に関して線形化した . しかし X_t は

$$X_t = (R_t(\theta_t^{(rot)}) \theta_t^{(ms)}) \tilde{X}_0$$

で表現され , これが回転行列 $R_t(\theta_t^{(rot)})$ を通して $\theta_t^{(rot)}$ に関して非線形であるため , ここでは $\theta_t^{(rot)}$ に関する線形化を考える . ここで ,

$$\begin{aligned} R_t(\theta_t^{(rot)}) &= R_{y,t} R_{x,t} R_{z,t} \\ &= \begin{pmatrix} \cos \theta_{y,t} & 0 & \sin \theta_{y,t} \\ 0 & 1 & 0 \\ -\sin \theta_{y,t} & 0 & \cos \theta_{y,t} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{x,t} & -\sin \theta_{x,t} \\ 0 & \sin \theta_{x,t} & \cos \theta_{x,t} \end{pmatrix} \begin{pmatrix} \cos \theta_{z,t} & -\sin \theta_{z,t} & 0 \\ \sin \theta_{z,t} & \cos \theta_{z,t} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} r_{1,t} & r_{2,t} & r_{3,t} \\ r_{4,t} & r_{5,t} & r_{6,t} \\ r_{7,t} & r_{8,t} & r_{9,t} \end{pmatrix} \end{aligned}$$

更に ,

$$\begin{aligned}
r_{1,t} &= \cos \theta_{y,t} \cos \theta_{z,t} + \sin \theta_{y,t} \sin \theta_{x,t} \sin \theta_{z,t} \\
r_{2,t} &= -\cos \theta_{y,t} \sin \theta_{z,t} + \sin \theta_{y,t} \sin \theta_{x,t} \cos \theta_{z,t} \\
r_{3,t} &= \sin \theta_{y,t} \cos \theta_{x,t} \\
r_{4,t} &= \cos \theta_{x,t} \sin \theta_{z,t} \\
r_{5,t} &= \cos \theta_{x,t} \cos \theta_{z,t} \\
r_{6,t} &= \cos \theta_{x,t} \\
r_{7,t} &= -\sin \theta_{y,t} \cos \theta_{z,t} + \cos \theta_{y,t} \sin \theta_{x,t} \sin \theta_{z,t} \\
r_{8,t} &= \sin \theta_{y,t} \sin \theta_{z,t} + \cos \theta_{y,t} \sin \theta_{x,t} \cos \theta_{z,t}
\end{aligned}$$

である . $(\cos \theta)' = -\sin \theta$, $(\sin \theta)' = \cos \theta$ より , $r_{1,t}, \dots, r_{9,t}$ をそれぞれテイラー展開により $\theta_{t-1}^{(rot)}$ の周りで線形近似すれば ,

$$\begin{aligned}
r_{1,t} &= \cos \theta_{y,t} \cos \theta_{z,t} + \sin \theta_{y,t} \sin \theta_{x,t} \sin \theta_{z,t} \\
&\approx \cos \theta_{y,t-1} \cos \theta_{z,t-1} + \sin \theta_{y,t-1} \sin \theta_{x,t-1} \sin \theta_{z,t-1} \\
&\quad + (\theta_{x,t} - \theta_{x,t-1})(\sin \theta_{y,t-1} \sin \theta_{z,t-1} \cos \theta_{x,t-1}) \\
&\quad + (\theta_{y,t} - \theta_{y,t-1})(-\cos \theta_{z,t-1} \sin \theta_{y,t-1} + \sin \theta_{x,t-1} \sin \theta_{z,t-1} \cos \theta_{y,t-1}) \\
&\quad + (\theta_{z,t} - \theta_{z,t-1})(-\cos \theta_{y,t-1} \sin \theta_{z,t-1} + \sin \theta_{y,t-1} \sin \theta_{x,t-1} \cos \theta_{z,t-1}) \\
&= f_{1x,t} \theta_{x,t} + f_{1y,t} \theta_{y,t} + f_{1z,t} \theta_{z,t} + f_{1c,t}
\end{aligned}$$

$r_{2,t}, \dots, r_{9,t}$ に関しても同様にテイラー展開を適用して , $\mathbf{r}_t = (r_{1,t}, \dots, r_{9,t})^T$ とすれば ,

$$\mathbf{r}_t = F_t = \tilde{\boldsymbol{\theta}}_t^{(rot)} \quad (\text{A.38})$$

と書くことができる . ただし ,

$$F_t = \begin{pmatrix} f_{1x,t} & f_{1y,t} & f_{1z,t} & f_{1c,t} \\ \vdots & \vdots & \vdots & \vdots \\ f_{9x,t} & f_{9y,t} & f_{9z,t} & f_{9c,t} \end{pmatrix}, \quad \tilde{\boldsymbol{\theta}}_t^{(rot)} = (\theta_{x,t}, \theta_{y,t}, \theta_{z,t}, 1)^T$$

式 (A.37) , (A.38) より観測過程は

$$\begin{aligned} \tilde{\mathbf{x}}_t &= \begin{pmatrix} \frac{f}{Z_{t-1}} & 0 & -\frac{f}{Z_{t-1}^2} X_{t-1} & \frac{f}{Z_{t-1}} X_{t-1} \\ 0 & \frac{f}{Z_{t-1}} & -\frac{f}{Z_{t-1}^2} Y_{t-1} & \frac{f}{Z_{t-1}} Y_{t-1} \\ 0 & 0 & 0 & 1 \end{pmatrix} \left(\begin{array}{c|c} R_t(\boldsymbol{\theta}_t^{(rot)}) & \boldsymbol{\theta}_t^{(tns)} \\ \hline 0 & 1 \end{array} \right) \tilde{X}_0 + D_X \mathbf{v}_t \\ &= P_{\theta,t} \tilde{\boldsymbol{\theta}}_t + D_X \mathbf{v}_t \end{aligned} \quad (\text{A.39})$$

の形に書き直すことが出来る．但し，

$$P_{\theta,t} = \begin{pmatrix} p_{1\theta,t} & p_{2\theta,t} & p_{3\theta,t} & p_{4\theta,t} & p_{5\theta,t} & p_{6\theta,t} & p_{7\theta,t} \\ p_{8\theta,t} & p_{9\theta,t} & p_{10\theta,t} & p_{11\theta,t} & p_{12\theta,t} & p_{13\theta,t} & p_{14\theta,t} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

とし、各要素は以下で求められる。

$$\begin{aligned}
p_{1\theta,t} &= \frac{f}{Z_{t-1}} \left\{ X_0 \left(f_{1x,t} - \frac{X_{t-1}}{Z_{t-1}} f_{7x,t} \right) + Y_0 \left(f_{2x,t} - \frac{X_{t-1}}{Z_{t-1}} f_{8x,t} \right) + Z_0 \left(f_{3x,t} - \frac{X_{t-1}}{Z_{t-1}} f_{9x,t} \right) \right\} \\
p_{2\theta,t} &= \frac{f}{Z_{t-1}} \left\{ X_0 \left(f_{1y,t} - \frac{X_{t-1}}{Z_{t-1}} f_{7y,t} \right) + Y_0 \left(f_{2y,t} - \frac{X_{t-1}}{Z_{t-1}} f_{8y,t} \right) + Z_0 \left(f_{3y,t} - \frac{X_{t-1}}{Z_{t-1}} f_{9y,t} \right) \right\} \\
p_{3\theta,t} &= \frac{f}{Z_{t-1}} \left\{ X_0 \left(f_{1z,t} - \frac{X_{t-1}}{Z_{t-1}} f_{7z,t} \right) + Y_0 \left(f_{2z,t} - \frac{X_{t-1}}{Z_{t-1}} f_{8z,t} \right) + Z_0 \left(f_{3z,t} - \frac{X_{t-1}}{Z_{t-1}} f_{9z,t} \right) \right\} \\
p_{4\theta,t} &= \frac{f}{Z_{t-1}} \\
p_{5\theta,t} &= 0 \\
p_{6\theta,t} &= \frac{f}{Z_{t-1}} \left(-\frac{X_{t-1}}{Z_{t-1}} \right) \\
p_{7\theta,t} &= \frac{f}{Z_{t-1}} \left\{ X_0 \left(f_{1c,t} - \frac{X_{t-1}}{Z_{t-1}} f_{7c,t} \right) + Y_0 \left(f_{2c,t} - \frac{X_{t-1}}{Z_{t-1}} f_{8c,t} \right) + Z_0 \left(f_{3c,t} - \frac{X_{t-1}}{Z_{t-1}} f_{9c,t} \right) \right\} \\
p_{8\theta,t} &= \frac{f}{Z_{t-1}} \left\{ X_0 \left(f_{4x,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{7x,t} \right) + Y_0 \left(f_{5x,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{8x,t} \right) + Z_0 \left(f_{6x,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{9x,t} \right) \right\} \\
p_{9\theta,t} &= \frac{f}{Z_{t-1}} \left\{ X_0 \left(f_{4y,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{7y,t} \right) + Y_0 \left(f_{5y,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{8y,t} \right) + Z_0 \left(f_{6y,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{9y,t} \right) \right\} \\
p_{10\theta,t} &= \frac{f}{Z_{t-1}} \left\{ X_0 \left(f_{4z,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{7z,t} \right) + Y_0 \left(f_{5z,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{8z,t} \right) + Z_0 \left(f_{6z,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{9z,t} \right) \right\} \\
p_{11\theta,t} &= 0 \\
p_{12\theta,t} &= \frac{f}{Z_{t-1}} \\
p_{13\theta,t} &= \frac{f}{Z_{t-1}} \left(-\frac{Y_{t-1}}{Z_{t-1}} \right) \\
p_{14\theta,t} &= \frac{f}{Z_{t-1}} \left\{ X_0 \left(f_{4c,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{7c,t} \right) + Y_0 \left(f_{5c,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{8c,t} \right) + Z_0 \left(f_{6c,t} - \frac{Y_{t-1}}{Z_{t-1}} f_{9c,t} \right) \right\}
\end{aligned}$$

F. 変分ベイズを用いた因子分解法

F.1 VB-E ステップ

VB-E ステップでは自由エネルギー \mathcal{F} を $q_n(s^{(n)})$ に関して最大化する。未定乗数法を用いて $\int q_n(s^{(n)}) ds^{(n)} = 1$ の制約の下、上記自由エネルギーの極値問題を変分法に

より解けば,

$$q_n(\mathbf{s}^{(n)}) = C \exp\left(\left\langle \log p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta) \right\rangle_{q(\Theta)}\right)$$

が得られる．ここで， C は正規化のための定数である．今，

$$\begin{aligned} \left\langle \log p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta) \right\rangle_{q(\Theta)} &= \left\langle \log p(\mathbf{s}^{(n)}) \right\rangle_{q(\Theta)} + \left\langle \log p(\mathbf{x}^{(n)} | \mathbf{s}^{(n)}, \Theta) \right\rangle_{q(\Theta)} \\ &= \left\langle \log \mathcal{N}_{N_d}(\mathbf{s}^{(n)} | \mathbf{0}, I_{N_d}) \right\rangle_{q(\Theta)} + \left\langle \log \mathcal{N}_D(\mathbf{x}^{(n)} | \tilde{M}\tilde{\mathbf{s}}^{(n)}, \Sigma_x^{(n)}) \right\rangle_{q(\Theta)} \\ &= -\frac{N_d}{2} \log(2\pi) - \frac{1}{2} \mathbf{s}^{(n)T} \mathbf{s}^{(n)} - \frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_x^{(n)}| \\ &\quad - \frac{1}{2} \left\langle (\mathbf{x}^{(n)} - \tilde{M}\tilde{\mathbf{s}}^{(n)})^T \Sigma_x^{(n)-1} (\mathbf{x}^{(n)} - \tilde{M}\tilde{\mathbf{s}}^{(n)}) \right\rangle_{q(\Theta)} \end{aligned}$$

と書くことができる．ここで，

$$\begin{aligned} &\left\langle (\mathbf{x}^{(n)} - \tilde{M}\tilde{\mathbf{s}}^{(n)})^T \Sigma_x^{(n)-1} (\mathbf{x}^{(n)} - \tilde{M}\tilde{\mathbf{s}}^{(n)}) \right\rangle_{q(\Theta)} \\ &= \mathbf{x}^{(n)T} \Sigma_x^{(n)-1} \mathbf{x}^{(n)} - 2 \text{Tr} \left[\tilde{\mathbf{s}}^{(n)} \mathbf{x}^{(n)T} \Sigma_x^{(n)-1} \left\langle \tilde{M} \right\rangle_{q(\Theta)} \right] + \text{Tr} \left[\tilde{\mathbf{s}}^{(n)} \tilde{\mathbf{s}}^{(n)T} \left\langle \tilde{M}^T \Sigma_x^{(n)-1} \tilde{M} \right\rangle_{q(\Theta)} \right], \\ &\left\langle \tilde{M} \right\rangle_{q(\Theta)} \equiv \hat{\Theta} = (\hat{M}, \hat{\mathbf{m}}_{N_d}), \\ &\left\langle \tilde{M}^T \Sigma_x^{(n)-1} \tilde{M} \right\rangle_{q(\Theta)} = \left\langle (\Sigma_x^{(n)-\frac{1}{2}} \tilde{M})^T (\Sigma_x^{(n)-\frac{1}{2}} \tilde{M}) \right\rangle_{q(\Theta)} \\ &= \left\{ \left\langle (\Sigma_x^{(n)-\frac{1}{2}} \tilde{M}) \right\rangle_{q(\Theta)} \right\}^2 + \text{Var} \left\{ \left\langle (\Sigma_x^{(n)-\frac{1}{2}} \tilde{M}) \right\rangle_{q(\Theta)} \right\} \\ &= \hat{\Theta}^T \Sigma_x^{(n)-1} \hat{\Theta} + \hat{V}^{(n)-1} \end{aligned}$$

であり，これより

$$\begin{aligned} \left\langle \log p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta) \right\rangle_{q(\Theta)} &= -\frac{N_d}{2} \log(2\pi) - \frac{1}{2} \mathbf{s}^{(n)T} \mathbf{s}^{(n)} - \frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_x^{(n)}| \\ &\quad - \frac{1}{2} (\mathbf{x}^{(n)} - \hat{\Theta} \tilde{\mathbf{s}}^{(n)})^T \Sigma_x^{(n)-1} (\mathbf{x}^{(n)} - \hat{\Theta} \tilde{\mathbf{s}}^{(n)}) - \frac{1}{2} \tilde{\mathbf{s}}^{(n)T} \hat{V}^{(n)-1} \tilde{\mathbf{s}}^{(n)} \end{aligned}$$

が成り立ち，更に

$$\tilde{I}_{\tilde{N}} = \begin{pmatrix} I_N & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix}$$

とすれば, 上式は

$$\begin{aligned} \langle \log p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta) \rangle_{q(\Theta)} &= -\frac{N_d}{2} \log(2\pi) - \frac{1}{2} \tilde{\mathbf{s}}^{(n)T} (\tilde{\mathbf{I}}_{N_d} + \hat{\mathbf{V}}^{(n-1)}) \tilde{\mathbf{s}}^{(n)} \\ &\quad - \frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_x^{(n)}| \\ &\quad - \frac{1}{2} (\mathbf{x}^{(n)} - \hat{\Theta} \tilde{\mathbf{s}}^{(n)})^T \Sigma_x^{(n-1)} (\mathbf{x}^{(n)} - \hat{\Theta} \tilde{\mathbf{s}}^{(n)}) \end{aligned}$$

と書くことができる. ここで,

$$\Xi^{(n)} = (\tilde{\mathbf{I}}_{N_d} + \hat{\mathbf{V}}^{(n-1)})$$

である. 今, $\tilde{\mathbf{s}}^{(n)} = (\mathbf{s}^{(n)T}, 1)^T$ であることから, $\Xi^{(n)}$ を以下の様に分割することを考える.

$$\Xi^{(n)} = \left(\begin{array}{ccc|c} \Xi_{1,1}^{(n)} & \cdots & \Xi_{1,N_d}^{(n)} & \Xi_{1,N_d+1}^{(n)} \\ \vdots & \ddots & \vdots & \vdots \\ \Xi_{N_d,1}^{(n)} & \cdots & \Xi_{N_d,N_d}^{(n)} & \Xi_{N_d,N_d+1}^{(n)} \\ \hline \Xi_{N_d+1,1}^{(n)} & \cdots & \Xi_{N_d+1,N_d}^{(n)} & \Xi_{N_d+1,N_d+1}^{(n)} \end{array} \right) = \left(\begin{array}{c|c} \Xi_{ss}^{(n)} & \Xi_{s1}^{(n)} \\ \hline \Xi_{1s}^{(n)} & \Xi_{11}^{(n)} \end{array} \right)$$

この時

$$\begin{aligned} \tilde{\mathbf{s}}^{(n)T} \Xi^{(n)} \tilde{\mathbf{s}}^{(n)} &= \begin{pmatrix} \mathbf{s}^{(n)T} & 1 \end{pmatrix} \begin{pmatrix} \Xi_{ss}^{(n)} & \Xi_{s1}^{(n)} \\ \Xi_{1s}^{(n)} & \Xi_{11}^{(n)} \end{pmatrix} \begin{pmatrix} \mathbf{s}^{(n)} \\ 1 \end{pmatrix} \\ &= (\mathbf{s}^{(n)} - \bar{\mathbf{s}}^{(n)})^T \Sigma_s^{(n-1)} (\mathbf{s}^{(n)} - \bar{\mathbf{s}}^{(n)}) + (\Xi_{11}^{(n)} + \Xi_{s1}^{(n)T} \bar{\mathbf{s}}^{(n)}) \end{aligned}$$

が成り立つ. 但し, $\bar{\mathbf{s}}^{(n)} = \Xi_{ss}^{(n-1)} \Xi_{s1}^{(n)}$, $\Sigma_s^{(n)} = \Xi_{ss}^{(n-1)}$ とする. 更に, $\Xi_{ss}^{(n)}$ の余因子行列を $\Xi_{ss}^{(n)\dagger}$ で書けば

$$\begin{aligned} |\Xi^{(n)}| &= \Xi_{11}^{(n)} |\Xi_{ss}^{(n)}| - \Xi_{1s}^{(n)} \Xi_{ss}^{(n)\dagger} \Xi_{s1}^{(n)} \\ &= |\Xi_{ss}^{(n)}| (\Xi_{11}^{(n)} - \Xi_{1s}^{(n)} \Xi_{ss}^{(n-1)} \Xi_{s1}^{(n)}) \end{aligned}$$

が成り立つので

$$\log |\Xi^{(n)}| = \log |\Xi_{ss}^{(n)}| + \log (\Xi_{11}^{(n)} + \Xi_{s1}^{(n)T} \bar{\mathbf{s}}^{(n)})$$

であるから，これを用いて

$$\begin{aligned} \langle \log p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta) \rangle_{q(\Theta)} &= \log \mathcal{N}_{N_d}(\mathbf{s}^{(n)} | \bar{\mathbf{s}}^{(n)}, \bar{\Sigma}_s^{(n)}) + \log \mathcal{N}_D(\mathbf{x}^{(n)} | \hat{\Theta} \tilde{\mathbf{s}}^{(n)}, \Sigma_x^{(n)}) \\ &\quad + \frac{1}{2} \log |\Xi_{ss}^{(n)-1}| - \frac{1}{2} (\Xi_{11}^{(n)} + \Xi_{s1}^{(n)} \bar{\mathbf{s}}^{(n)}) + \frac{1}{2} \log (\Xi_{11}^{(n)} + \Xi_{s1}^{(n)} \bar{\mathbf{s}}^{(n)}) \end{aligned}$$

と書くことができる．また，上式は

$$\begin{aligned} \hat{\mathbf{x}}^{(n)} &\equiv \hat{\Theta} \begin{pmatrix} \bar{\mathbf{s}}^{(n)} \\ 1 \end{pmatrix} \\ \hat{\Sigma}_x^{(n)} &\equiv \Sigma_x^{(n)} + \hat{M} \bar{\Sigma}_s^{(n)} \hat{M}^T \end{aligned}$$

として

$$\begin{aligned} \hat{\mathbf{s}}^{(n)} &= \bar{\mathbf{s}}^{(n)} + K^{(n)}(\mathbf{x}^{(n)} - \hat{\mathbf{x}}^{(n)}) \\ \hat{\Sigma}_s^{(n)} &= (I_{N_d} - K^{(n)} \hat{M}) \bar{\Sigma}_s^{(n)} \\ K^{(n)} &= \bar{\Sigma}_s^{(n)} \hat{M} \hat{\Sigma}_x^{(n)-1} \end{aligned}$$

とすれば，

$$\begin{aligned} \langle \log p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta) \rangle_{q(\Theta)} &= \log \mathcal{N}_{N_d}(\mathbf{s}^{(n)} | \hat{\mathbf{s}}^{(n)}, \hat{\Sigma}_s^{(n)}) + \log \mathcal{N}_D(\mathbf{x}^{(n)} | \hat{\mathbf{x}}^{(n)}, \hat{\Sigma}_x^{(n)}) \\ &\quad + \frac{1}{2} \log |\Xi_{ss}^{(n)-1}| - \frac{1}{2} (\Xi_{11}^{(n)} + \Xi_{s1}^{(n)} \bar{\mathbf{s}}^{(n)}) + \frac{1}{2} \log (\Xi_{11}^{(n)} + \Xi_{s1}^{(n)} \bar{\mathbf{s}}^{(n)}) \end{aligned}$$

となるので，自由エネルギーを最大化する $q_n(\mathbf{s}^{(n)})$ として以下を得ることができる．

$$\begin{aligned} q_n(\mathbf{s}^{(n)}) &= \mathcal{N}_{N_d}(\mathbf{s}^{(n)} | \hat{\mathbf{s}}^{(n)}, \hat{\Sigma}_s^{(n)}) \\ \langle \tilde{\mathbf{s}}^{(n)} \tilde{\mathbf{s}}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} &= \begin{pmatrix} \hat{\mathbf{s}}^{(n)} \hat{\mathbf{s}}^{(n)T} + \Sigma_s^{(n)} & \hat{\mathbf{s}}^{(n)} \\ \hat{\mathbf{s}}^{(n)T} & 1 \end{pmatrix} \end{aligned}$$

F.2 VB-M ステップ

VB-M ステップでは自由エネルギー \mathcal{F} を $q(\Theta)$ に関して最大化する．VB-E ステップと同様に，自由エネルギーの極値問題を変分法により解けば，

$$q(\Theta) = C' p(\Theta) \exp \left(\langle \log p(P_x, S | \Theta) \rangle_{q(S)} \right) \quad (\text{A.40})$$

が得られる．ここで C' は正規化のための定数である．今，

$$\begin{aligned}
\langle \log p(P_x, S | \Theta) \rangle_{q(S)} + \log p(\Theta) &= \sum_{n=1}^{N_f} \langle \log p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta) \rangle_{q(\mathbf{s}^{(n)})} + \sum_{d=1}^D \log p(\boldsymbol{\theta}_d) \\
&= \sum_{n=1}^{N_f} \langle \log \mathcal{N}_{N_d}(\mathbf{s}^{(n)} | \mathbf{0}, I_{N_d}) + \log \mathcal{N}_D(\mathbf{x}^{(n)} | \tilde{W} \tilde{\mathbf{s}}^{(n)}, \Sigma_x^{(n)}) \rangle_{q_n(\mathbf{s}^{(n)})} \\
&\quad + \sum_{d=1}^D \log \mathcal{N}_{N_d}(\boldsymbol{\theta}_d | \mathbf{e}_d, \Sigma_\theta) \\
&= \sum_{n=1}^{N_f} \left\{ -\frac{N_d}{2} \log(2\pi) - \frac{1}{2} \langle \mathbf{s}^{(n)T} \mathbf{s}^{(n)} \rangle_{q_n \mathbf{s}^{(n)}} \right. \\
&\quad - \frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_x^{(n)}| \\
&\quad \left. - \frac{1}{2} \langle (\mathbf{x}^{(n)} - \tilde{W} \tilde{\mathbf{s}}^{(n)})^T \Sigma_x^{(n)-1} (\mathbf{x}^{(n)} - \tilde{W} \tilde{\mathbf{s}}^{(n)}) \rangle_{q_n(\mathbf{s}^{(n)})} \right\} \\
&\quad + \sum_{d=1}^D \left\{ -\frac{N_d}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_\theta| - \frac{1}{2} (\boldsymbol{\theta}_d - \mathbf{e}_d)^T \Sigma_\theta^{-1} (\boldsymbol{\theta}_d - \mathbf{e}_d) \right\}
\end{aligned}$$

である．ここで，

$$\begin{aligned}
&-\frac{1}{2} \sum_{n=1}^{N_f} \left\{ \langle (\mathbf{x}^{(n)} - \tilde{W} \tilde{\mathbf{s}}^{(n)})^T \Sigma_x^{(n)-1} (\mathbf{x}^{(n)} - \tilde{W} \tilde{\mathbf{s}}^{(n)}) \rangle_{q_n(\mathbf{s}^{(n)})} \right\} - \frac{1}{2} \sum_{d=1}^D \{ (\boldsymbol{\theta}_d - \mathbf{e}_d)^T \Sigma_\theta^{-1} (\boldsymbol{\theta}_d - \mathbf{e}_d) \} \\
&= -\frac{1}{2} \sum_{n=1}^{N_f} \left\{ \left\langle \sum_{d=1}^D (x_d^{(n)} - \boldsymbol{\theta}_d^T \tilde{\mathbf{s}}^{(n)})^T \sigma_{x,d}^{(n)-2} (x_d^{(n)} - \boldsymbol{\theta}_d^T \tilde{\mathbf{s}}^{(n)}) \right\rangle_{q_n(\mathbf{s}^{(n)})} \right\} \\
&\quad - \frac{1}{2} \sum_{d=1}^D \{ (\boldsymbol{\theta}_d - \mathbf{e}_d)^T \Sigma_\theta^{-1} (\boldsymbol{\theta}_d - \mathbf{e}_d) \} \\
&= \frac{1}{2} \sum_{d=1}^D \{ (\boldsymbol{\theta}_d - \hat{\boldsymbol{\theta}}_d)^T \hat{V}_d (\boldsymbol{\theta}_d - \hat{\boldsymbol{\theta}}_d) + C \}
\end{aligned}$$

と書ける．ここで，

$$\begin{aligned}
\hat{V}_d &= \Sigma_\theta^{-1} + \sum_{n=1}^{N_f} \sigma_{x,d}^{(n)} \langle \tilde{\mathbf{s}}^{(n)} \tilde{\mathbf{s}}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} \\
\hat{\boldsymbol{\theta}}_d^T &= \left(\mathbf{e}_d^T \Sigma_\theta^{-1} + \sum_{n=1}^{N_f} \sigma_{x,d}^{(n)-2} \langle x_d^{(n)} \tilde{\mathbf{s}}^{(n)T} \rangle_{q_n(\mathbf{s}^{(n)})} \right) \hat{V}_d^{-1}
\end{aligned}$$

であるから，自由エネルギーを最大とする $q(\Theta)$ は

$$q(\Theta) = \prod_{d=1}^D \mathcal{N}_{\tilde{N}_d}(\boldsymbol{\theta}_d | \hat{\boldsymbol{\theta}}_d, \hat{V}_d^{-1})$$

で得ることができる．

この時，

$$\begin{aligned} \langle \tilde{M} \rangle_{q(\Theta)} &= \hat{\Theta} = (\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_D)^T \\ \langle \langle \log p(P_x, S | \Theta) \rangle_{q(S)} \tilde{M}^T \tilde{M} \rangle_{q(\Theta)} &= \left\langle \sum_{d=1}^D \boldsymbol{\theta}_d \boldsymbol{\theta}_d^T \right\rangle_{q(\Theta)} = \sum_{d=1}^D \langle \boldsymbol{\theta}_d \boldsymbol{\theta}_d^T \rangle_{q(\Theta)} = \hat{\Theta}^T \hat{\Theta} + \sum_{d=1}^D \hat{V}_d^{-1} \end{aligned}$$

より，VB-E ステップで用いた $\hat{V}^{(n)-1}$ は

$$\hat{V}^{(n)-1} = \sum_{d=1}^D \hat{V}_d^{-1}$$

で得ることができる．

F.3 自由エネルギー

自由エネルギー \mathcal{F} は

$$\mathcal{F} = N_f \mathcal{L} - \mathcal{H}$$

で得ることができる．ここで， \mathcal{L} は

$$\mathcal{L} = \frac{1}{N_f} \sum_{n=1}^{N_f} \int q_n(s^{(n)}) q(\Theta) \log \left(\frac{p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta)}{q_n(s^{(n)})} \right) d\mathbf{s}^{(n)} d\Theta = \mathcal{L}_{xs} - \mathcal{H}_s$$

である。但し，ここで

$$\begin{aligned}
\mathcal{L}_{xs} &\equiv \frac{1}{N_f} \sum_{n=1}^{N_f} \int q_n(\mathbf{s}^{(n)}) q(\Theta) \log p(\mathbf{x}^{(n)}, \mathbf{s}^{(n)} | \Theta) d\mathbf{s}^{(n)} d\Theta \\
&= \frac{1}{N_f} \sum_{n=1}^{N_f} \left\{ \left\langle \log \mathcal{N}_{N_d}(\mathbf{s}^{(n)} | \mathbf{0}, I_{N_d}) \right\rangle_{q_n(\mathbf{s}^{(n)})} + \left\langle \log \mathcal{N}_{N_D}(\mathbf{x}^{(n)} | \tilde{W} \tilde{\mathbf{s}}^{(n)}, \Sigma_x^{(n)}) \right\rangle_{q_n(\mathbf{s}^{(n)}) q(\Theta)} \right\} \\
&= -\frac{N_d + D}{2} \log(2\pi) + \frac{1}{N_f} \sum_{n=1}^{N_f} \left\{ -\frac{1}{2} \left\langle \mathbf{s}^{(n)T} \mathbf{s}^{(n)} \right\rangle_{q_n(\mathbf{s}^{(n)})} - \frac{1}{2} \log |\Sigma_x^{(n)}| \right. \\
&\quad \left. - \frac{1}{2} \left\langle \mathbf{x}^{(n)T} \Sigma_x^{(n)-1} \mathbf{x}^{(n)} \right\rangle_{q_n(\mathbf{s}^{(n)})} + \text{Tr} \left[\left\langle \tilde{\mathbf{s}}^{(n)} \mathbf{x}^{(n)T} \right\rangle_{q_n(\mathbf{s}^{(n)})} \Sigma_x^{(n)-1} \left\langle \tilde{W} \right\rangle_{q(\Theta)} \right] \right. \\
&\quad \left. - \frac{1}{2} \text{Tr} \left[\left\langle \tilde{\mathbf{s}}^{(n)} \tilde{\mathbf{s}}^{(n)T} \right\rangle_{q_n(\mathbf{s}^{(n)})} \left\langle \tilde{W}^T \Sigma_x^{(n)-1} \tilde{W} \right\rangle_{q(\Theta)} \right] \right\}
\end{aligned}$$

$$\begin{aligned}
\mathcal{H}_s &= \frac{1}{N_f} \sum_{n=1}^{N_f} \int q_n(\mathbf{s}^{(n)}) \log q_n(\mathbf{s}^{(n)}) d\mathbf{s}^{(n)} \\
&= \frac{1}{N_f} \sum_{n=1}^{N_f} \left\langle \log \mathcal{N}_{N_d}(\mathbf{s}^{(n)} | \hat{\mathbf{s}}^{(n)}, \hat{\Sigma}_s^{(n)}) \right\rangle_{q_n(\mathbf{s}^{(n)})} \\
&= \frac{1}{N_f} \sum_{n=1}^{N_f} \left\{ -\frac{N_d}{2} \log(2\pi) - \frac{1}{2} \log |\hat{\Sigma}_s^{(n)}| - \frac{1}{2} \left\langle (\mathbf{s}^{(n)} - \hat{\mathbf{s}})^T \Sigma_s^{(n)-1} (\mathbf{s}^{(n)} - \hat{\mathbf{s}}) \right\rangle_{q_n(\mathbf{s}^{(n)})} \right\} \\
&= -\frac{N_d}{2} (1 + \log(2\pi)) - \frac{1}{N_f} \sum_{n=1}^{N_f} \frac{1}{2} \log |\hat{\Sigma}_s^{(n)}|
\end{aligned}$$

$$\begin{aligned}
\mathcal{H} &\equiv \int q(\Theta) \log \left(\frac{q(\Theta)}{p(\Theta)} \right) d\Theta \\
&= \left\langle \sum_{d=1}^D \log \mathcal{N}_{\tilde{N}_d}(\boldsymbol{\theta}_d | \hat{\boldsymbol{\theta}}_d, \hat{V}_d^{-1}) \right\rangle_{q(\Theta)} - \left\langle \sum_{d=1}^D \log \mathcal{N}_{\tilde{N}_d}(\boldsymbol{\theta}_d | \hat{\boldsymbol{\theta}}_d, \Sigma_\theta) \right\rangle_{q(\Theta)} \\
&= \sum_{d=1}^D \left\{ -\frac{\tilde{N}_d}{2} \log(2\pi) - \frac{1}{2} \log |V_d^{-1}| - \frac{1}{2} \left(\boldsymbol{\theta}_d^T \hat{V}_d \boldsymbol{\theta}_d - \boldsymbol{\theta}_d^T \hat{V}_d \hat{\boldsymbol{\theta}}_d - \hat{\boldsymbol{\theta}}_d^T \hat{V}_d \boldsymbol{\theta}_d + \hat{\boldsymbol{\theta}}_d^T \hat{V}_d \hat{\boldsymbol{\theta}}_d \right) \right\} \\
&\quad - \sum_{d=1}^D \left\{ -\frac{\tilde{N}_d}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_\theta| - \frac{1}{2} \left(\boldsymbol{\theta}_d^T \Sigma_\theta^{-1} \boldsymbol{\theta}_d - \boldsymbol{\theta}_d^T \Sigma_\theta^{-1} \mathbf{e}_d - \mathbf{e}_d^T \Sigma_\theta^{-1} \boldsymbol{\theta}_d + \mathbf{e}_d^T \Sigma_\theta^{-1} \mathbf{e}_d \right) \right\}
\end{aligned}$$

とする。

研究業績

論文

- Takashi Bando, Tomohiro Shibata, Kenji Doya and Shin Ishii,
”Switching particle filters for efficient visual tracking,”
Robotics and Autonomous Systems, **54**(10), pp. 873-884, (2006).
- 坂東誉司, 柴田智広, 清水幹郎, 石井信,
”適応的サンプリングによる階層モデル化された対象の効率的状態推定,”
システム制御情報学会論文誌, **19**(10), pp. 369-377, (2006).

解説

- 柴田智広, 坂東誉司,
”パーティクルフィルタの切り替えによる頑健な実時間物体追跡,”
映像情報インダストリアル, **38**(10), pp.29-33.

国際会議

- Takashi Bando, Tomohiro Shibata, Kenji Doya and Shin Ishii,
”Switching particle filters for efficient real-time visual tracking.”
International Conference on Pattern Recognition (ICPR2004),
2(3), pp. 720-723, (2004).
- Takashi Bando, Tomohiro Shibata, Kenji Doya and Shin Ishii,

” Hard / soft switching particle filters for efficient real-time visual tracking.”
International Symposium on Artificial Life and Robotics (AROB 10th '05),
GS15-5, (2005).

研究会/全国大会

- 坂東誉司, 柴田智広, 銅谷賢治, 石井信.
”効果的な実時間視覚追跡のためのサンプリング切り替え手法.”
日本神経回路学会第 14 回全国大会, pp.56-57, (2004).
- 坂東誉司, 柴田智広, 清水幹郎, 石井信.
”階層的パーティクルフィルタのための適応的サンプリング分布推定とその剛体の実時間姿勢推定への応用”
電子情報通信学会技術研究報告 *NC ニューロコンピューティング*, **105**(657),
NC2005-119(2006-3), pp. 79-84, (2006).
- 坂東誉司, 柴田智広, 清水幹郎, 石井信.
”適応的サンプリング分布を用いた階層的状態推定.”
日本神経回路学会第 16 回全国大会, pp.208-209, (2006).
- 坂東誉司, 柴田智広, 清水幹郎, 深谷直樹, 石井信.
”確率的逐次因子分解による単眼画像時系列からの三次元構造復元” 電子情報
通信学会技術研究報告 *NC ニューロコンピューティング*, **106**(279), NC2006-
53(2006-10), pp. 49-54, (2006).