

時空間情報伝搬スキームに基づく動画像からの
3次元形状復元のオンライン実現

On-line Reconstruction of 3-D Structure from Image Sequences
Based on Spatio-temporal Information Propagation

15500117

平成 15 年度～平成 16 年度 科学研究費補助金（基盤研究 (C)）研究成果報告書

平成 18 年 3 月

研究代表者 田川 憲男
東京都立大学大学院工学研究科

はしがき

本報告書は、平成 15 年度から平成 16 年度にかけて、当該研究グループが本助成金（基盤研究 (C)）によって行った「時空間情報伝搬スキームに基づく動画像からの 3 次元形状復元のオンライン実現」の最終成果をまとめたものである。研究組織、交付決定額、研究発表は以下のとおりである。

研究組織

研究代表者： 田川 憲男 （東京都立大学大学院工学研究科・助教授）
研究分担者： 皆川 明洋 （東京都立大学大学院工学研究科・助手）

交付決定額（配分額）

（金額単位：千円）

	直接経費	間接経費	合計
平成 15 年度	2,200	0	2,200
平成 16 年度	800	0	800
総計	3,000	0	3,000

研究発表

- (1) 川口純矢，長沼将一，田川憲男，“多重解像度処理に基づく動きからのベイズ的三次元形状復元，” 電子情報通信学会総合大会，no. ISS-P-59, p. 77, 2006 年 3 月
- (2) 長沼将一，須藤優，田川憲男，“情報伝搬を基礎とする動きからの 3 次元構造復元，” 映像情報メディア学会メディア工学研究会技術報告，2006 年 5 月（発表予定）

目次

第1章	はじめに	1
1.1	本研究の目的	1
1.2	本研究の概要	2
1.3	本研究の成果	3
第2章	多重解像度処理に基づくベイズ的オプティカルフロー推定	6
2.1	まえがき	6
2.2	勾配方程式に基づくオプティカルフロー検出の原理	7
2.2.1	投影モデルとオプティカルフロー	7
2.2.2	勾配方程式	8
2.3	多重解像度処理による復元の概要	8
2.3.1	解決すべき問題点	8
2.3.2	解決策の基本的考え方	9
2.4	多重解像度処理に基づく計算原理	10
2.4.1	画像の分解と確率モデル	10
2.4.2	フローの階層的推定	11
2.4.3	パラメータの決定	13
2.5	計算アルゴリズム	15
2.5.1	EM アルゴリズムの適用	15
2.5.2	計算手続き	16
第3章	時間統合処理に基づく動きからのベイズ的三次元構造推定	20
3.1	まえがき	20
3.2	計算原理	21
3.2.1	オプティカルフローからの構造復元	21
3.2.2	基本となる推定方式	21
3.3	情報伝搬による逐次復元	23
3.3.1	逐次復元の基本的考え方	23
3.3.2	具体的定式化	24

3.4	計算アルゴリズム	28
3.4.1	EM アルゴリズムの適用	28
3.4.2	計算手続き	28
第4章	数値評価	32
4.1	評価の目的	32
4.2	評価条件	32
4.3	評価結果	33
第5章	おわりに	49
	参考文献	50

第1章 はじめに

1.1 本研究の目的

ロボットの視覚システムの実現を目指す Computer Vision (CV) の研究は、生理学者、視覚心理学者、物理学者、数学者そして工学者といった様々な研究者によって精力的に進められている。生理学者や視覚心理学者の興味は、主として人間の視覚メカニズムの解明にあり、その成果を工学的に実現することによって視覚システムを構築できると考えている。一方で数理工学的立場からは、物理的、幾何光学的観点から妥当とみなされる計算原理の解明とそのアルゴリズム化が期待され、現在までに膨大な数の研究成果が得られている。ただし、その計算原理の妥当性は、純粋に物理的・幾何光学的に正しいことのみではなく、人間の視覚メカニズムと深く関係を持つことによっても評価されるものであり、その意味で上記の二つの立場は全く独立なものではない。むしろ、積極的な協力関係が重要である。

当該研究者はこれまで、主に後者の立場から、視覚システムの入力に相当する画像雑音にロバストな計算アルゴリズムの開発に従事してきた。その際、どのような計算原理を基盤とするかが重要な問題となる [1]。CV の原理的枠組みは、一般に Structure from X と記されることが多い。ここで X は、3次元構造復元使用される物理的・幾何学的現象や法則を意味し、Structure from shading, Structure from texture, Structure from Vanishing Point 等多数提案されている。当該研究者は、主に Structure from Motion (SFM) に焦点を当てて研究を進めてきた。これは、カメラと対象物体との間の相対的運動によって生じる画像上の見掛けの速度場 (オプティカルフロー [2]) に基づいて、物体の3次元構造を復元する方法論であり、環境が静的であれば Structure from Stereo (両眼立体視) をより一般化したものと考えることができる [3], [4], [5]。SFM は CV の中でも最も活発な研究が進められている枠組みの一つである。初期の研究は、数学的立場での復元原理、すなわち解の存在条件、解の個数等に関する議論が中心であった。その後、高精度な計算アルゴリズムの開発、統計的推論を基盤とする画像雑音に強い方法の議論がなされてきた [6]。そして近年は、複数フレームの情報を安定かつ効率的に利用する方法 (因子分解法 [7] がその代表である)、カメラキャリブレーションを行わずに実現可能な構造復元の性質や具体的復元手法とその応用方法等が盛んに研究されている。さらには、Image

Based Rendering (IBR) や Virtual Reality (VR) への CV の応用も注目されている。これらの研究動向は、CV の可能性、有用性を広めるという意味では重要であり、完全な CV アルゴリズムが未開発であっても、CV を社会に還元する工学的研究としては貴重である。しかしながら、一方で CV、特に SFM のより完成度の高いアルゴリズムを開発することは、上記と並行して進められるべきである。

SFM の本質的困難は、見掛けの 2 次元速度場を高精度かつ高密度で得ることが本来不良設定問題であること、及び時々刻々と見え方が変化する動画像の情報統合が、計算上極めて難しいことである。上述の因子分解法は、これらの問題が解決されている場合の効率的な計算方法に相当するため、結局本質的な解決にはなっていない。そこで本研究では、これらの二つの困難を克服するための手法の検討を目的とする。

1.2 本研究の概要

前節での二つの問題に対応し、本研究では計算処理を二つに分離して実現する。一つは各時刻における解像度方向の処理であり、これによって連続する 2 フレームに基づく安定・高精度で密なオプティカルフローの算出を可能とする。もう一つは時間方向の処理であり、見え方の時間連続性を重視してオプティカルフロー情報を適切に時間統合することで、高精度な物体形状（奥行き情報）の復元を可能とする。なお、本研究で共通に用いるカメラモデルを図 1.1 に示す。奥行き方向に対応するカメラ座標系の Z 座標をレンズの光軸に合わせ、中心投影の投影面（画像面）を $Z = 1$ の平面とする。また、カメラの並進、回転速度ベクトルをそれぞれ、 u, r で表す。

本研究では、関連する物理量を確率変数と解釈し、それらの間の関係をモデル化した上で、統計的推論、特にベイズ推定の枠組みで解を得る方策をとる。図 1.2 は、本研究で想定する確率モデルをグラフィカル表現したものである。グラフィカル表現とは、確率変数をノードに配置し、それらの間の確率的関係を辺でつないだものであり、今回は有向グラフ表現を用いている。この場合の有向辺は、確率変数間の直接的な因果関係を表している。

まず、 f_i^j は時刻 i の解像度 j の画像（画素輝度値の集合）を表す。各時刻の画像は複数解像度に分解されて使用される。オプティカルフロー v_i^j の添字も同様である。 d_i は時刻 i での奥行き逆数である。なお、本研究では密な形状復元を目指すことから、物体上の特定の特徴点の奥行きを求めるのではなく、各時刻において各画素に対応する奥行き（逆数）(iconic depth) を求める pixel-based なアプローチを採用する。

図 1.2 からわかるように、ある時刻 t の d_t をベイズ推論の枠組みで算出するためには、 $\{f_i^j\} (i \leq t, j = 1, 2, \dots, L)$ が持つ情報を $\{v_i^j\}, \{d_i\}$ を経由して適切に伝

搬させなければならない．この伝搬は，上記の二つの処理によって実現される．各処理の詳細はそれぞれ2章，3章で述べるが，その概要は以下のように整理できる．

1. 解像度処理によるオプティカルフロー検出

勾配方程式に基づくオプティカルフロー検出における原理的な困難として，窓問題とエイリアス現象が古くから議論されている．前者に関する標準的な対処法として局所最適化と大域的最適化が存在するものの，ともに分解能を画像解像度に比べて低く制限してしまうという欠点を有する．一方，エイリアスに関しては画像上の動き量に対して十分長い波長成分のみを輝度値パターンから抽出して処理に使用することが一般的である．しかし，この場合も結果的に，用いる輝度値パターンの解像度以上の分解能を期待することはできない．本研究では，画像を空間周波数帯域の異なる複数の画像に分割し，解像度方向に適切な情報伝搬を行うことで，分解能低下を抑えて上記二つの困難を解決する手法を提案する．

2. 時間統合処理による3次元形状復元

オプティカルフローに基づく3次元構造復元においては，各時刻の構造情報を適切に統合することで，復元結果の高精度化を図る必要がある．そのための従来法の多くでは，連続するフレーム間でオプティカルフローを用いた3次元復元を行った後，得られた各画素での奥行き情報をカルマンフィルタを利用して統合している．しかしながら，本来はオプティカルフローそのものを観測量とし，奥行きをMPM推定として逐次更新していくことが望ましい．その際，カルマンフィルタの直接的適用のためには，フレーム間でのカメラと物体の相対的3次元運動が既知でなければならない．そこで本研究では，フレーム間での3次元運動を最尤推定の枠組みで決定しながら，カルマンフィルタによって奥行き更新を行う新たな手法を提案する．

1.3 本研究の成果

本研究の成果は，以下のように整理できる．

1. 時空間情報統合のための確率グラフの検討

今回は，時空間情報統合の適切さと実際に行う推論の容易さを併せて評価し，図1.2に示す確率構造を得た．この結果はあくまでも一例であり，更なる拡張が可能である．例えば， d_i についても解像度の異なる変数を増やし，解像度方向の情報伝搬を v_i^j に加えて， d_i^j も経由することが考えられる．ただしその場合は，推論に要する計算量が増加することから，新たな近似の工夫が必要となる．

2. 時空間情報伝搬の枠組みの構築

今回採用した確率構造に対しては，解像度方向のカルマンフィルタと時間方向のカルマンフィルタに処理を分離することができる．この種の考え方は，他の多くの初期視覚システムの構築に役立つものとする．

3. パラメータ決定機能を有するカルマンフィルタの構築

図 1.2 の各ノードには，未知のパラメータ，具体的には入力雑音の分散，状態（オプティカルフローや奥行き逆数）の時間あるいは解像度方向への遷移に関わるパラメータ，フレーム間での 3 次元運動パラメータ等が存在する．通常のカルマンフィルタの枠組みでは，これらのパラメータは既知である必要がある．しかしながら，一般には未知であることが多く，今回の問題においても，それらパラメータも同時に決定しなければならない．本研究では，パラメータを最尤推定として決定しながら，逐次的に奥行き逆数を計算していくアルゴリズムを構築した．この方法論は，他の多くのカルマンフィルタリング問題に適用可能である．

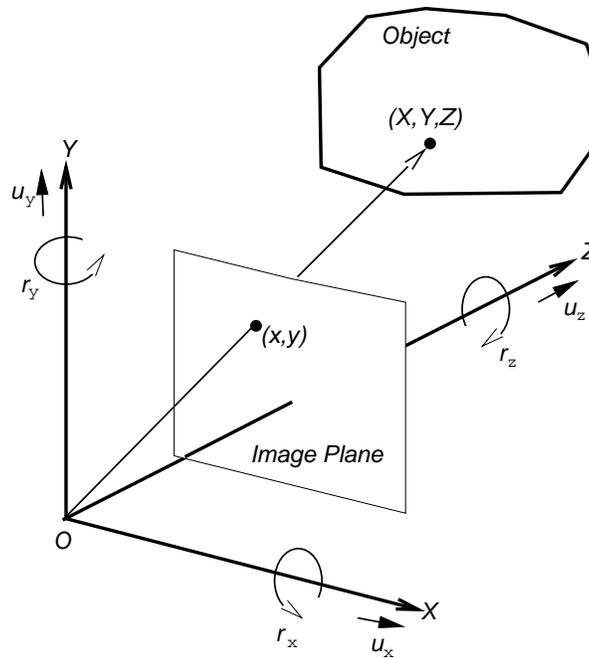


図 1.1: カメラ座標系

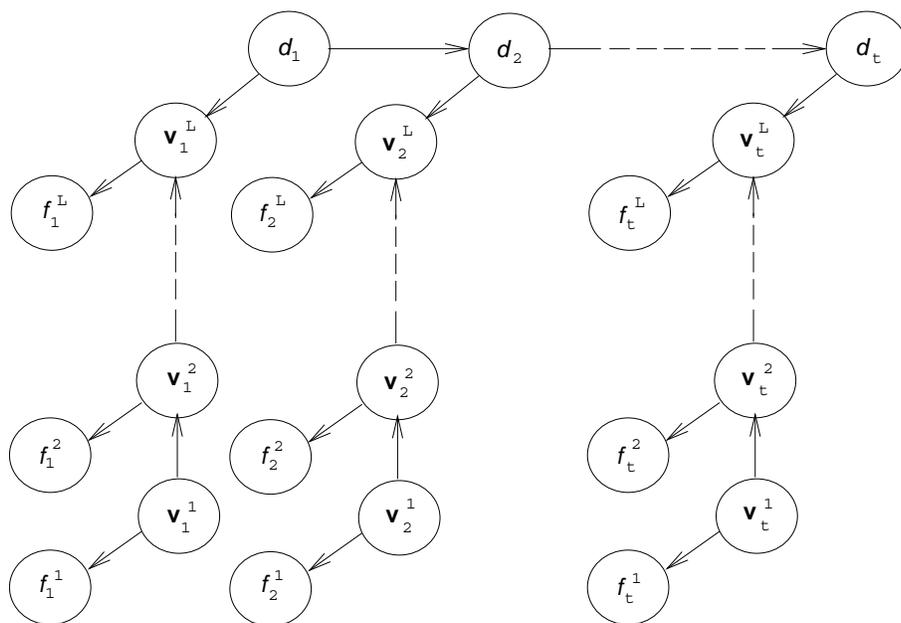


図 1.2: 諸量の確率的関係のグラフィカル表現

第2章 多重解像度処理に基づくベイズ的オプティカルフロー推定

2.1 まえがき

structure from motion (SFM) の枠組みにおける研究は、従来より computer vision (CV) の中核をなし、膨大な数の成果が報告されてきた [3], [4], [5], [6]. 研究初期には、数学的立場での復元原理に関する研究（解の存在条件、解の個数等）が脚光を浴び、その後は高精度な復元手法の開発が進められ、同時に CV における統計的推論の在り方や新しい考え方が議論されてきた．そして昨今は、因子分解法 [7] に代表される安定で高効率な手法、カメラキャリブレーションを要しない復元法とその利用法、そしてそれらの結果を用いた VR 技術の開発に注目が集まっている．しかし、要素技術研究の立場では SFM が本質的に有する困難の完全な解決は依然残されていると言わざるを得ない．その本質的な困難とは、画像上の密な変位・速度場の正確な検出、ならびに動画像処理としての適切な時間統合であり、上記の近年注目されている技術は画像上での粗な特徴点追跡が何らかの方法で行えていることを前提とすることで、これらの困難を形式的に回避しているに過ぎない．本稿では上述の困難のうちの前者に焦点を当て、それを解決する一手法を検討する．

勾配方程式に基づく二次元速度場（オプティカルフロー）の検出は本来不良設定問題であり、各画素におけるオプティカルフローの値は別の制約を付与しなければ求まらない．この現象は窓問題と呼ばれ、人間の視覚系においても錯視の要因となる基本的な困難である．従来は、オプティカルフローが局所的に一定であると仮定する局所最適化法 [8]、あるいはオプティカルフローが空間的に滑らかに変化すると仮定する大域的最適化法 [2] による解決がなされている．しかしこれらの仮定は、間接的に対象物の形状に制約を課しており、検出されるオプティカルフロー、したがってこれから得られる復元形状の分解能低下の原因となる．特に物体の輪郭等、奥行きが不連続に変化する領域に相当する画素において不正確な奥行きが検出されてしまう．最近、勾配方程式の指導原理である運動前後での輝度値不変性が成立しない場合を考慮して、幾何学的拘束以外に輝度値拘束も併せて利用する手法 [9], [10] が検討されており、この立場からの工夫も上記不良設定性を解決するために有効である可能性はあるが、未だ完全な解決には至っていない．

オプティカルフロー検出に関わるもう一つの困難として、エイリアス問題が挙げられる。すなわち、オプティカルフローの大きさに比べて短い波長の輝度値パターンをその決定に使用した場合は、得られるオプティカルフローは大きな誤差を含むことになる。これはオプティカルフローの大きさに対して相対的に時間サンプリングが粗いことによって生じ、エイリアシングと呼ばれる現象の一つである。アクティブビジョンの立場であれば、時間サンプリングも制御することによって、あるいはオプティカルフローが常にある一定の大きさ以内に収まるような制御を施すことでこのエイリアシングを回避することができる。しかし、通常のアプリケーションでは受動的に撮影された動画像に基づいて形状復元をすることが多く、信号処理に基づく回避法が望ましい。従来法の多くは、空間的に十分滑らかな輝度値パターンを low-pass フィルタによって取り出して使用しているため、この場合も結果的に空間分解能の低い形状が得られることになる。

以上の2点の解決を目指し、本研究では原画像を空間周波数帯域の異なる複数の解像度画像に分解した上で、低解像度画像から高解像度画像へと情報を伝搬させながら、安定かつ高分解能なオプティカルフロー検出を実現することを考える。なお、同様の考え方に基づくオプティカルフロー検出方法は既に存在するが、基本的にカルマンフィルタをそのまま適用していることから、いくつかのパラメータを事前に知っている必要があった [11]。それに対し、本研究では統計的推定に必要となるすべてのパラメータも同時に決定する方法を EM アルゴリズム [12] を利用して構築する。これにより、ヒューリスティックなパラメータ決定が不要となり、最適性の面でも有利となる。

2.2 勾配方程式に基づくオプティカルフロー検出の原理

2.2.1 投影モデルとオプティカルフロー

図 1.1 に示すように、奥行き値に対応するカメラ座標系の Z 座標をレンズの光軸に合わせ、中心投影の投影面（画像面）を $Z = 1$ の平面とする。3次元実空間内の物体上の点 (X, Y, Z) の投影点 $(x, y) \equiv (X/Z, Y/Z)$ におけるオプティカルフロー $\mathbf{v} \equiv [v_x, v_y]^\top$ は次式を満足する。

$$v_x = xy r_x - (1 + x^2) r_y + y r_z - (u_x - x u_z) d, \quad (2.1)$$

$$v_y = (1 + y^2) r_x - x y r_y - x r_z - (u_y - y u_z) d. \quad (2.2)$$

$\mathbf{u} = [u_x, u_y, u_z]^\top$, $\mathbf{r} = [r_x, r_y, r_z]^\top$ はカメラの並進及び回転速度ベクトルであり、物体がカメラ座標系に対して $-\mathbf{u}$, $-\mathbf{r}$ の速度で運動しているとも考えることもできる。ただし、輝度値の運動前後での不変性を厳密に考慮するならば、例えば物体の光反

射特性は完全な拡散反射であり，物体と光源が固定された状況でカメラが運動しているものとする必要がある． d は奥行き Z の逆数であり，画像上の各画素における未知数である．これに対し， (u, r) はすべての画素に共通する未知数となる．画像上で v が検出されれば，式 (2.1)，(2.2) に基づいて (u, r) とオプティカルフローが検出された画素に対応する d を求めることができる．したがって，高精度なオプティカルフローの検出は，高精度な奥行きを求めるために必須である．なお，以降ではオプティカルフローをフローと略記する．

2.2.2 勾配方程式

画像上の輝度値を $f(x, y, t)$ で表すと，運動前後での輝度値不変性の一次近似式（勾配方程式）は次のように与えられる．

$$f_t = -f_x v_x - f_y v_y = -\mathbf{f}_g^\top \mathbf{v}. \quad (2.3)$$

ここで， f_t, f_x, f_y はそれぞれ時空間の偏導関数であり， $\mathbf{f}_g \equiv [f_x, f_y]^\top$ である．本稿では後節で述べるように，画像を空間解像度（帯域）の異なる複数の画像に分解した後，各解像度に応じた微分操作を施すことで輝度値の空間微分に関する誤差を抑えることから，相対的に時間微分の誤差が顕著に現れる．そこで，式 (2.3) において左辺のみに誤差（雑音）が含まれるものと想定し，この観測方程式を基礎として， \mathbf{v} を決定する方法を検討する．

2.3 多重解像度処理による復元の概要

2.3.1 解決すべき問題点

勾配方程式に基づく高精度・高分解能な復元手法のためには，エイリアシング問題と窓問題を適切に回避する必要がある．まず，前者を簡単に説明する．各画素におけるフローを $[v_x, v_y]^\top$ とすると，空間周波数 (ω_x, ω_y) と時間周波数 ω_t の間には， $\omega_t = v_x \omega_x + v_y \omega_y$ の関係が成り立つことから，空間周波数毎に時間サンプリングのナイキスト周波数が異なることがわかる．しかし，通常の画像の時間サンプリングは空間周波数によらず一定であるため，高い空間周波数成分の時間サンプリングは低い空間周波数成分に比して標本化定理を満足しない可能性，すなわちエイリアシングが生じる確率が高いと言える．そこで多くの従来法では，原画像に適当な空間 low-pass フィルタを施して得られる低解像度画像のみを解析に用いている．この方法では高解像度情報が欠落し，その結果として復元形状の分解能が低下する．

一方窓問題は， $[v_x, v_y]^\top$ が等しいと仮定できる画像上の局所領域における複数の

式 (2.3) を用いても、各画素の空間勾配 (f_x, f_y) が等しい場合にはフローを一意に決定できないことを述べており、その根幹には単独画素で定義された式 (2.3) が不良設定性を有することが原因として存在する。この問題に対する従来の対処法は大別して、局所最適化法 [8] と大域最適化法 [2] がある。前者は時間あるいは空間について局所的にフローが一定であると仮定し、式 (2.3) を連立させて解く方法である。すなわち、最終的に未知形状に局所的な制限を加えるものであり、線形計算で解を得ることができる。これに対して後者は、空間的な滑らかさに代表される大域的な制限を仮定する方法であり、具体的計算は変分原理に基づいた微分方程式を解くことに帰着される。

以上、エイリアスに対しては使用する画像情報に制限を加える（低周波数成分を使用する）こと、窓問題に対しては形状に制限を加える（分解能すなわち自由度を下げる）ことがそれぞれ肝要である。従来の多くの方法では、この両工夫を併せて適用することで、双方の問題を回避している。しかし、これでは得られるフローの分解能が画像解像度に比べて低下してしまい、結果としてフローの検出精度も特に形状の不連続な部分において低くなる。

2.3.2 解決策の基本的考え方

本稿では、フローの分解能の低下を引き起こさずに上記二点の困難を同時に克服することを考える。基本的な考え方は、(1) 画像を空間周波数帯域の異なる複数の解像度画像に分割すること、(2) 各解像度の処理に際しては仮定するフロー分解能を解像度と適切に連動させること、及び (3) 処理を低解像度から高解像度へと再帰的に進めることである。

(1) と (3) の組み合わせはエイリアスへの対処を可能にする。エイリアスを避けるためには低解像度画像を使用することが望ましいが、分解能を低下させないためには高解像度情報を適切に使用することが重要となる。そこで、低解像度で得られる安定な（エイリアスの可能性が低い）情報を高解像度処理へと再帰的に伝搬させることで、エイリアスを避けた高解像度情報の利用を可能とする。この考え方は、文献 [11] の方法のそれと同様である。情報伝搬スキームにはベイズ推論方式を採用することで、低解像度でのフロー推定誤差を高解像度での処理で補償することが可能となる。

窓問題に関しては、(1) と (2) の組み合わせにより対処することができる。エイリアス対策として (1) を採用すれば、各解像度から得ることのできる形状情報は自然に階層構造をなすことになる。したがって、低解像度処理においては変数の分解能を下げ、逆に高解像度処理においては分解能を高めておくことが適切である。この

ような解像度に応じた分解能の制御は，大域最適化・局所最適化のいずれによっても可能であるが，今回は局所演算性を重視し，局所最適化の枠組みを採用する．つまり，低解像度ではフロー一定を仮定する領域サイズを大きくし，高解像度となるにつれてこのサイズを小さくする．

なお，情報伝搬の技術的側面としての本稿での寄与に，パラメータ（例えば観測雑音の分散）の値も観測量に基づいて系統的に決定することが挙げられる．フローと併せてパラメータ値も統計的推論の立場から決定することを目指し，EM アルゴリズム [12] を利用した方法を構築する．

2.4 多重解像度処理に基づく計算原理

2.4.1 画像の分解と確率モデル

まず，観測画像の複数解像度画像への分解について説明する．単純に解像度の異なる画像への分解でよければ，空間 Low-pass フィルタの遮断周波数を変えることで可能であるが，後の処理において輝度値の時間微分に加わる雑音の分解画像間での統計的独立性を利用することから，Band-pass フィルタによる分解を採用し，帯域の違いをインデックス $l = 1, \dots, L$ で表す． $l = 1$ が最低解像度画像（最下位画像）を指すものとする．この帯域すなわち輝度値の空間波長に連動させて，フロー分解能を定める．すなわち，帯域 l の画像においては局所領域サイズ N_l においてフローは一定と仮定し， $l_1 > l_2$ であれば， $N_{l_1} < N_{l_2}$ となるように定める．なお，情報の効率的扱いとしては Wavelet 変換のようにピラミッド画像構成を利用する方が望ましいが，ここでは雑音抑圧の効果を考慮し，すべての解像度画像のサンプリング，すなわち画像サイズは共通とする．

次に，確率モデルを定義する． v は上述のように，各解像度画像においてその帯域に応じた局所領域内では共通であり，局所領域間で独立と扱うことになる．さらに解像度画像間では条件付き独立な未知数とする．つまり， l から $l+1$ への線形補間演算子を $E^{(l)}$ とし，

$$v^{(l+1)} = E^{(l)} v^{(l)} + n_0^{(l)} \quad (2.4)$$

とする． $n_0^{(l)}$ は線形補間値からのずれを表す確率変数で，他の確率変数とは独立な平均 0，共分散 $\sigma_0^2 I$ （ I は 2×2 単位行列）の正規分布に従うとすると， $v^{(l+1)}$ の確率密度関数は次式で与えられる．

$$p(v^{(l+1)} | \sigma_0^2) = \frac{1}{\sqrt{(2\pi)^2 \det V^{(l+1)}}} \exp \left\{ -\frac{1}{2} (v^{(l+1)} - \bar{v}^{(l+1)})^\top V^{(l+1)-1} (v^{(l+1)} - \bar{v}^{(l+1)}) \right\} \quad (2.5)$$

ここで，平均 $\bar{v}^{(l+1)}$ 及び共分散 $V^{(l+1)}$ は，下位画像での平均，分散を用いて次のように再帰的に書き表される．

$$\bar{v}^{(l+1)} = E^{(l)}\bar{v}^{(l)}, \quad (2.6)$$

$$V^{(l+1)} = E^{(l)2}V^{(l)} + \sigma_0^2 I. \quad (2.7)$$

ただし，式 (2.7) は隣接する $v^{(l)}$ 間の共分散を無視した近似式であり，演算子 $E^{(l)2}$ は，演算子 $E^{(l)}$ における重み係数をそれぞれ二乗したものをその重み係数とする線形補間演算子である．

続いて観測量 f_t の確率モデルを考える．観測雑音を含む観測方程式を，式 (2.3) に基づいて次式で定義する．

$$\begin{aligned} f_t^{(l)} &= -f_x^{(l)}v_x^{(l)} - f_y^{(l)}v_y^{(l)} + n_1^{(l)} \\ &= -\mathbf{f}_g^{(l)\top} \mathbf{v}^\top + n_1^{(l)}. \end{aligned} \quad (2.8)$$

ここでは簡単のため， $n_1^{(l)}$ は他の確率変数と独立な平均 0，分散 σ_1^2 の正規分布に従う $f_t^{(l)}$ の観測雑音とする．このとき， $f_t^{(l)}$ の条件付き確率は

$$p(f_t^{(l)}|\mathbf{v}^{(l)}, \sigma_1^2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(f_t^{(l)} + \mathbf{f}_g^{(l)\top} \mathbf{v}^{(l)})^2}{2\sigma_1^2}\right\} \quad (2.9)$$

で与えられる．

2.4.2 フローの階層的推定

前節までの議論に基づき， $\{f_t^{(l+1)}\}, \{f_t^{(l)}\}, \dots, \{f_t^{(0)}\}$ を観測したときの最小分散推定量 $\hat{v}^{(l+1)}$ を求めることを考える．なお，ここでの記号 $\{a\}$ はすべての画素での a の集合を意味する．そのような $\hat{v}^{(l+1)}$ は想定するすべての観測を得た後の確率（事後確率）の平均値に相当する．そこで，以下ではその事後確率を求める．

$\{f_t^{(l)}\}_{\mathcal{N}}$ を $v^{(l)}$ を一定と仮定する局所領域 \mathcal{N} における $f_t^{(l)}$ の集合とすると，次が成り立つ．

$$p(\mathbf{v}^{(l+1)}|\{f_t^{(l+1)}\}, \{f_t^{(l)}\}, \dots) = p(\mathbf{v}^{(l+1)}|\{f_t^{(l+1)}\}_{\mathcal{N}}, \{f_t^{(l)}\}, \dots). \quad (2.10)$$

これは，ベイズの公式を適用すると次式となる．

$$p(\mathbf{v}^{(l+1)}|\{f_t^{(l+1)}\}_{\mathcal{N}}, \{f_t^{(l)}\}, \dots) = \frac{p(\{f_t^{(l+1)}\}_{\mathcal{N}}|\mathbf{v}^{(l+1)})p(\mathbf{v}^{(l+1)}|\{f_t^{(l)}\}, \dots)}{p(\{f_t^{(l+1)}\}_{\mathcal{N}}|\{f_t^{(l)}\}, \dots)}. \quad (2.11)$$

この事後確率の分子を具体的に計算してみる．まず，分子の二つめは次のように書き下すことができる．

$$p(\mathbf{v}^{(l+1)}|\{f_t^{(l)}\}, \dots) = \int p(\mathbf{v}^{(l+1)}|\{\mathbf{v}^{(l)}\}_{\mathcal{P}})p(\{\mathbf{v}^{(l)}\}_{\mathcal{P}}|\{f_t^{(l)}\}, \dots)d\{\mathbf{v}^{(l)}\}_{\mathcal{P}}, \quad (2.12)$$

$$p(\{\mathbf{v}^{(l)}\}_{\mathcal{P}}|\{f_t^{(l)}\}, \dots) = \prod_{\mathcal{P}} p(\mathbf{v}^{(l)}|\{f_t^{(l)}\}_{\mathcal{N}}, \dots). \quad (2.13)$$

ここで, $\{\cdot\}_{\mathcal{P}}$ は線形補間に使用する要素の集合を示す. 式 (2.4) 及び式 (2.8) より, 式 (2.12) も正規分布となり, 一つ下の階層での事後分布 $p(\mathbf{v}^{(l)}|\{f_t^{(l)}\}, \dots)$ の平均と共分散をそれぞれ $\tilde{\mathbf{v}}^{(l)}$, $\tilde{V}^{(l)}$ とすると, 式 (2.5) と同様に次式のように書ける.

$$\begin{aligned} p(\mathbf{v}^{(l+1)}|\{f_t^{(l)}\}, \dots) &= \frac{1}{\sqrt{(2\pi)^2 \det(E^{(l)2} \tilde{V}^{(l)} + \sigma_0^2 I)}} \\ &\times \exp \left\{ -\frac{1}{2} (\mathbf{v}^{(l+1)} - E^{(l)} \tilde{\mathbf{v}}^{(l)})^\top \{E^{(l)2} \tilde{V}^{(l)} + \sigma_0^2 I\}^{-1} (\mathbf{v}^{(l+1)} - E^{(l)} \tilde{\mathbf{v}}^{(l)}) \right\}. \end{aligned} \quad (2.14)$$

これより, 式 (2.11) の分子は以下のように表される.

$$\begin{aligned} p(\{f_t^{(l+1)}\}_{\mathcal{N}}|\mathbf{v}^{(l+1)}) p(\mathbf{v}^{(l+1)}|\{f_t^{(l)}\}, \dots) &= \left\{ \prod_{\mathcal{N}} p(f_t^{(l+1)}|\mathbf{v}^{(l+1)}) \right\} p(\mathbf{v}^{(l+1)}|\{f_t^{(l)}\}, \dots) \\ &= \frac{\exp\{-H\}}{Z}, \end{aligned} \quad (2.15)$$

$$Z = \left\{ \sqrt{2\pi\sigma_1^2} \right\}^N \sqrt{(2\pi)^2 \det(E^{(l)2} \tilde{V}^{(l)} + \sigma_0^2 I)}, \quad (2.16)$$

$$\begin{aligned} H &= \frac{\sum_{\mathcal{N}} (f_t^{(l+1)} + \mathbf{f}_g^{(l+1)\top} \mathbf{v}^{(l+1)})^2}{2\sigma_1^2} \\ &\quad + \frac{1}{2} (\mathbf{v}^{(l+1)} - E^{(l)} \tilde{\mathbf{v}}^{(l)})^\top (E^{(l)2} \tilde{V}^{(l)} + \sigma_0^2 I)^{-1} (\mathbf{v}^{(l+1)} - E^{(l)} \tilde{\mathbf{v}}^{(l)}) \\ &= \frac{1}{2} (\mathbf{v}^{(l+1)} - \tilde{\mathbf{v}}^{(l+1)})^\top \{\tilde{V}^{(l+1)}\}^{-1} (\mathbf{v}^{(l+1)} - \tilde{\mathbf{v}}^{(l+1)}) + \text{Const.}, \end{aligned} \quad (2.17)$$

$$\tilde{\mathbf{v}}^{(l+1)} = \tilde{V}^{(l+1)} \left\{ (E^{(l)2} \tilde{V}^{(l)} + \sigma_0^2 I)^{-1} E^{(l)} \tilde{\mathbf{v}}^{(l)} - \frac{\sum_{\mathcal{N}} f_t^{(l+1)} \mathbf{f}_g^{(l+1)}}{\sigma_1^2} \right\}, \quad (2.18)$$

$$\tilde{V}^{(l+1)} = \left\{ (E^{(l)2} \tilde{V}^{(l)} + \sigma_0^2 I)^{-1} + \frac{\sum_{\mathcal{N}} \mathbf{f}_g^{(l+1)} \mathbf{f}_g^{(l+1)\top}}{\sigma_1^2} \right\}^{-1}. \quad (2.19)$$

H を最小にする $\mathbf{v}^{(l+1)}$ の値が MAP 推定量 $\mathbf{v}_{MAP}^{(l+1)}$ であり, これは式 (2.11) の平均に一致する. ゆえに, $\mathbf{v}_{MAP}^{(l+1)} = \tilde{\mathbf{v}}^{(l+1)}$ である. なお, $\tilde{\mathbf{v}}^{(l+1)}$ の算出は, 上記のように下位階層からの情報を用いて逐次的に行なうことができ, これはカルマンフィルタに相当している. すなわち, 式 (2.4) が状態方程式, 式 (2.8) が観測方程式である.

カルマンフィルタは通常，動的に推移する状態のその時刻以前の観測に基づく周辺事後確率を計算するための効率的な手続きとして使用される．ここでも，求めたいものは $\{v^{(L)}\}$ であり，それより低解像度のフローそのものは不要であるため， $\{v^{(L)}\}$ 単独の MAP 推定 (MPM 推定) をカルマンフィルタによって実現する方策をとった．また，カルマンフィルタによって $\{v^{(l)}\}$ を逐次的に推定しなければならないもう一つのより本質的な理由として，エイリアス問題の回避がある． $f_t^{(l)}$ を連続フレーム間の差分として求めた場合，画像上での変位が $f^{(l)}$ の有する帯域に対して大きい場合にエイリアスが生じる．そこで，下位階層 (低解像度) 画像で得られた $\hat{v}^{(l-1)}$ を用いた次式により $f_t^{(l)}$ を定義する．

$$f_t^{(l)} = -f_x^{(l)} E^{(l-1)} \hat{v}_x^{(l-1)} - f_y^{(l)} E^{(l-1)} \hat{v}_y^{(l-1)} + \frac{\partial}{\partial t} \mathcal{W}(f^{(l)}, E^{(l-1)} \hat{v}^{(l-1)}). \quad (2.20)$$

ただし，画像のワープ処理を次のように定義する．

$$\mathcal{W}(f, v)(x, t + \delta t) = f(x - v\delta t, t + \delta t). \quad (2.21)$$

このように，エイリアスを避けて $f_t^{(l)}$ を得るためには，エイリアスの生じる危険性の少ない低解像度画像から得られるオプティカルフロー推定値を逐次利用しなければならず，このためにもカルマンフィルタの採用は必要不可欠である．

2.4.3 パラメータの決定

前節で示したフローの逐次的 MAP 推定のためには，いくつか既知でなければならないパラメータが存在する．各解像度に依存する $\Theta_{res} = \{\sigma_0^{2(l)}, \sigma_1^{2(l')}\}_{l=1, \dots, L-1; l'=1, \dots, L}$ と逐次処理の初期値となる $\Theta_{init} = \{\bar{v}_{init}, V_{init}\}$ である．

事前知識の特に存在しない未知量の統計的推論としては，最尤推定 (MLE) が一般的である．MLE とは，すべての観測の同時確率に実際の観測値を代入したものとして定義される尤度関数に基づき，これを最大とするパラメータを推定量とする推論方式である．本稿では，各層のパラメータの MLE を逐次的に計算することを考える．すなわち，ある層 l の観測を得たときに，その層でのパラメータ $\Theta_{res}^{(l)} \equiv \{\sigma_0^{2(l-1)}, \sigma_1^{2(l)}\}$ の MLE を，それ以前の層の全観測に基づくものとして逐次的に得る．

そのために，形式的に一様分布に従う事前確率 $\{p(\Theta_{res}^{(l)})\}$ を仮定する．すると，他のパラメータを周辺化した $\Theta_{res}^{(l)}$ の事後確率は， $l \geq 2$ に関して次式となる．

$$\begin{aligned} p(\Theta_{res}^{(l)} | \{f_t^{(l)}\}, \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}) &= \frac{p(\Theta_{res}^{(l)})}{P(\{f_t^{(l)}\} | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\})} \\ &\times \int p(\{f_t^{(l)}\} | \{v^{(l)}\}, \sigma_1^{2(l)}) p(\{v^{(l)}\} | \{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \sigma_0^{2(l-1)}) d\{v^{(l)}\}. \end{aligned} \quad (2.22)$$

ただし，被積分関数の第二因子は，該当する層のパラメータのみで規定されるフローの事後確率であり，次のように定式化できる．

$$\begin{aligned} & p(\{\mathbf{v}^{(l)}\}|\{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \sigma_0^{2(l-1)}) \\ &= \int p(\{\mathbf{v}^{(l)}\}|\{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \sigma_0^{2(l-1)}, \Theta_{res}^{(l-1)})p(\Theta_{res}^{(l-1)}|\{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\})d\Theta_{res}^{(l-1)}. \end{aligned} \quad (2.23)$$

ここで，上式の被積分関数の第二因子は，一つ下位の層のパラメータの周辺事後確率である．ゆえに，各層の式 (2.22) は，下位層の結果を用いて逐次的に計算できる．式 (2.22) を最大にする $\Theta_{res}^{(l)}$ は MPM 推定量であるが， $p(\Theta_{res}^{(l)})$ を一様分布と考えることにより，値としては MLE に一致する．なお，式 (2.23) 中の $\Theta_{res}^{(l-1)}$ に関する周辺化は一般に困難であり，適当な近似を用いる必要がある．本稿では最も簡単な近似として，鞍点近似を適用する．すなわち，既に求まっている下位層の MLE $\Theta_{res\ ML}^{(l-1)}$ を用い，次式で近似する．

$$p(\{\mathbf{v}^{(l)}\}|\{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \sigma_0^{2(l-1)}) \simeq p(\{\mathbf{v}^{(l)}\}|\{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \sigma_0^{2(l-1)}, \Theta_{res\ ML}^{(l-1)}). \quad (2.24)$$

この近似を採用すれば，式 (2.22) から $\Theta_{res}^{(l)}$ の推定に必要な部分を次のように取り出すことができる．

$$\begin{aligned} & p(\{f_t^{(l)}\}|\{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}, \Theta_{res}^{(l)}) \\ & \simeq \int p(\{f_t^{(l)}\}|\{\mathbf{v}^{(l)}\}, \sigma_1^{2(l)})p(\{\mathbf{v}^{(l)}\}|\{f_t^{(l-1)}\}, \dots, \sigma_0^{2(l-1)}, \Theta_{res\ ML}^{(l-1)})d\{\mathbf{v}^{(l)}\}. \end{aligned} \quad (2.25)$$

したがって，各層でこの式を条件付き尤度関数とみなして最大化を行うことにより，近似的に $\Theta_{res\ ML}^{(l)}$ を得ることができる．

$l = 1$ (一層目) の場合は， $p(\Theta_{init}, \sigma_1^{2(1)}|\{f_t^{(1)}\})$ に基づく MPM 推定を行うことになる．

$$\begin{aligned} & p(\Theta_{init}, \sigma_1^{2(1)}|\{f_t^{(1)}\}) \\ &= \frac{p(\Theta_{init})p(\sigma_1^{2(1)})}{p(\{f_t^{(1)}\})} \int p(\{f_t^{(1)}\}|\{\mathbf{v}^{(1)}\}, \sigma_1^{2(1)})p(\{\mathbf{v}^{(1)}\}|\Theta_{init})d\{\mathbf{v}^{(1)}\}. \end{aligned} \quad (2.26)$$

ただし，上述の鞍点近似を採用することから，この事後分布そのものは不要であり，パラメータの決定のためには次式の最大化を行えばよい．

$$\begin{aligned} & p(\{f_t^{(1)}\}|\Theta_{init}, \sigma_1^{2(1)}) \\ &= \int p(\{f_t^{(1)}\}|\{\mathbf{v}^{(1)}\}, \sigma_1^{2(1)})p(\{\mathbf{v}^{(1)}\}|\Theta_{init})d\{\mathbf{v}^{(1)}\}. \end{aligned} \quad (2.27)$$

なお，一層目における式 (2.15) に相当する同時確率，ならびにそれから得られる $\mathbf{v}^{(1)}$ の MAP 推定量は次式のように与えられる．ただし，推定精度を確保するために $V_{init} = \sigma_{init}^2 I$ と定義する．

$$p(\{f_t^{(1)}\}_{\mathcal{N}} | \mathbf{v}^{(1)}, \sigma_1^{2(1)}) p(\mathbf{v}^{(1)} | \bar{\mathbf{v}}_{init}, \sigma_{init}^2) = \frac{\exp\{-H_{init}\}}{Z_{init}}, \quad (2.28)$$

$$\begin{aligned} H_{init} &= \frac{\sum_{\mathcal{N}} (f_t^{(1)} + \mathbf{f}_g^{(1)\top} \mathbf{v}^{(1)})^2}{2\sigma_1^{2(1)}} + \frac{(\mathbf{v}^{(1)} - \bar{\mathbf{v}}_{init})^2}{2\sigma_{init}^2} \\ &= \frac{1}{2} (\mathbf{v}^{(1)} - \tilde{\mathbf{v}}^{(1)})^\top \tilde{V}^{(1)-1} (\mathbf{v}^{(1)} - \tilde{\mathbf{v}}^{(1)}) + \text{Const.}, \end{aligned} \quad (2.29)$$

$$Z_{init} = \left(\sqrt{2\pi\sigma_1^{2(1)}} \right)^N (2\pi\sigma_{init}^2), \quad (2.30)$$

$$\tilde{\mathbf{v}}^{(1)} = \tilde{V}^{(1)} \left\{ \frac{\bar{\mathbf{v}}_{init}}{\sigma_{init}^2} - \frac{\sum_{\mathcal{N}} f_t^{(1)} \mathbf{f}_g^{(1)}}{\sigma_1^{2(1)}} \right\}, \quad (2.31)$$

$$\tilde{V}^{(1)} = \left\{ \frac{I}{\sigma_{init}^2} + \frac{\sum_{\mathcal{N}} \mathbf{f}_g^{(1)} \mathbf{f}_g^{(1)\top}}{\sigma_1^{2(1)}} \right\}^{-1}. \quad (2.32)$$

式 (2.27) は，すべての領域にわたる式 (2.28) の積を $\{\mathbf{v}^{(1)}\}$ について周辺化したものであることに注意する．

2.5 計算アルゴリズム

2.5.1 EM アルゴリズムの適用

式 (2.25) の最大化によるパラメータ推定，ならびにその値に基づく式 (2.18) の計算を効率的に行うために，EM アルゴリズムの枠組みを採用することが可能である．EM アルゴリズムとは観測されていない未知確率変数（隠れ変数）を観測量に加えて想定したときに，両者（完全データ）が観測されれば問題が解きやすくなる場合に有効な推定スキームであり，その未知確率変数の事後分布とパラメータの ML 推定量を同時に得ることのできる反復手続きである．詳細は省略するが，各反復では次の 2 つのステップを実行する．

1. E-step：パラメータを以前の反復で得られている値としたときの隠れ変数の事後分布を求め，これを利用して Q 関数と呼ばれるパラメータ更新のための評価関数を算出する．

2. M-step : Q 関数を最大化するようにパラメータ値を更新する .

以降では , EM アルゴリズムを用いて各 l で Θ_{res} の ML 推定と , この値を固定したときの $\{\mathbf{v}^{(l)}\}$ の MAP 推定を行うための , 各層での具体的計算手続きを導出する .

2.5.2 計算手続き

まず , 解像度 $l(\geq 2)$ の画像に対して , それ以下の解像度で得られた情報に基づいて $\Theta_{res}^{(l)} = \{\sigma_0^{2(l-1)}, \sigma_1^{2(l)}\}$ の ML 推定と $\{\mathbf{v}^{(l)}\}$ の MAP 推定を行う計算式を導出する . E-step では次のように定義される $\Theta_{res}^{(l)}$ に関する関数 (Q 関数) を具体的に導出する .

$$Q(\Theta_{res}^{(l)}; \hat{\Theta}_{res}^{(l)}) \equiv E \left[\ln p(\{f_t^{(l)}\}, \{\mathbf{v}^{(l)}\} | \Theta_{res}^{(l)}) \middle| \{f_t^{(l)}\}, \hat{\Theta}_{res}^{(l)} \right]. \quad (2.33)$$

ただし , この式を含めて以降では , $\{\{f_t^{(l-1)}\}, \dots, \{f_t^{(1)}\}\}$ に関する条件付きであることを省略する . また , $\hat{\cdot}$ はその反復時に得られている値 , あるいはそれに依存する変数とする . $\{f_t^{(l)}\}$ と $\{\mathbf{v}^{(l)}\}$ の同時確率は

$$\begin{aligned} & p(\{f_t^{(l)}\} | \{\mathbf{v}^{(l)}\}, \sigma_1^{2(l)}) p(\{\mathbf{v}^{(l)}\} | \sigma_0^{2(l-1)}) \\ &= \prod_{i=1}^{M^{(l)}} \left\{ \prod_{\mathcal{N}_i^{(l)}} p(f_t^{(l)} | \mathbf{v}_i^{(l)}, \sigma_1^{2(l)}) \right\} \prod_{i=1}^{M^{(l)}} p(\mathbf{v}_i^{(l)} | \sigma_0^{2(l-1)}) \end{aligned} \quad (2.34)$$

であり , よって式 (2.33) において期待値計算を施される関数は次式となる . ただし , $\mathcal{N}_i^{(l)}$ に含まれる画素数は i によらず一定値 $N^{(l)}$ をとり , さらに各フレーム内の $\mathcal{N}_i^{(l)}$ の数を $M^{(l)}$ とする .

$$\begin{aligned} & -\frac{N^{(l)}M^{(l)}}{2} \ln \sigma_1^{2(l)} - \frac{1}{2} \sum_{i=1}^{M^{(l)}} \ln \det \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I \right) \\ & - \frac{\sum_{i=1}^{M^{(l)}} \sum_{\mathcal{N}_i^{(l)}} \left(f_t^{(l)} + \mathbf{f}_g^{(l)\top} \mathbf{v}_i^{(l)} \right)^2}{2\sigma_1^{2(l)}} \\ & - \frac{1}{2} \sum_{i=1}^{M^{(l)}} \left(\mathbf{v}_i^{(l)} - (E^{(l-1)} \tilde{\mathbf{v}}^{(l-1)})_i \right)^\top \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I \right)^{-1} \left(\mathbf{v}_i^{(l)} - (E^{(l-1)} \tilde{\mathbf{v}}^{(l-1)})_i \right) \\ & + \text{Const..} \end{aligned} \quad (2.35)$$

一方 , $\{f_t^{(l)}\}$ 観測後の $\mathbf{v}_i^{(l)}$ の各反復での事後分布 $p(\mathbf{v}_i^{(l)} | \{f_t^{(l)}\}_{\mathcal{N}_i}, \hat{\Theta}_{res}^{(l)})$ は , 式 (2.18) , (2.19) より , 平均及び分散がその反復での $\hat{\Theta}_{res}^{(l)}$ を用いて以下で表される正規分布と

なる．

$$\hat{\mathbf{v}}_i^{(l)} = \hat{V}_i^{(l)} \left\{ \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \hat{\sigma}_0^{2(l-1)} I \right)^{-1} (E^{(l-1)} \tilde{\mathbf{v}}^{(l-1)})_i - \frac{\sum_{\mathcal{N}_i^{(l)}} f_t^{(l)} \mathbf{f}_g^{(l)}}{\hat{\sigma}_1^{2(l)}} \right\}, \quad (2.36)$$

$$\hat{V}_i^{(l)} = \left\{ \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \hat{\sigma}_0^{2(l-1)} I \right)^{-1} + \frac{\sum_{\mathcal{N}_i^{(l)}} \mathbf{f}_g^{(l)} \mathbf{f}_g^{(l)\top}}{\hat{\sigma}_1^{2(l)}} \right\}^{-1}. \quad (2.37)$$

ゆえに，式 (2.35) を $p(\mathbf{v}_i^{(l)} | \{f_t^{(l)}\}_{\mathcal{N}_i}, \hat{\Theta}_{res}^{(l)})$ に関して期待値をとることで，式 (2.33) が具体的に得られる．M-step ではこの $Q(\Theta_{res}^{(l)}; \hat{\Theta}_{res}^{(l)})$ を $\Theta_{res}^{(l)}$ に関して最大化する．そこで， $Q(\Theta_{res}^{(l)}; \hat{\Theta}_{res}^{(l)})$ を -2 倍して定数を無視して，次の関数の最小化を行うものとする．

$$\begin{aligned} J(\Theta_{res}^{(l)}) &= N^{(l)} M^{(l)} \ln \sigma_1^{2(l)} + \sum_{i=1}^{M^{(l)}} \ln \det \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I \right) \\ &\quad + J_f(\sigma_1^{2(l)}) + J_v(\sigma_0^{2(l-1)}), \end{aligned} \quad (2.38)$$

$$\begin{aligned} J_f(\sigma_1^{2(l)}) &= \frac{1}{\sigma_1^{2(l)}} \left\{ \sum_{i=1}^{M^{(l)}} \sum_{\mathcal{N}_i^{(l)}} f_t^{(l)2} + 2 \sum_{i=1}^{M^{(l)}} \left(\sum_{\mathcal{N}_i^{(l)}} f_t^{(l)} \mathbf{f}_g^{(l)\top} \right) \hat{\mathbf{v}}_i^{(l)} \right. \\ &\quad \left. + \sum_{i=1}^{M^{(l)}} \text{tr} \left\{ \left(\hat{\mathbf{v}}_i^{(l)} \hat{\mathbf{v}}_i^{(l)\top} + \hat{V}_i^{(l)} \right) \sum_{\mathcal{N}_i^{(l)}} \mathbf{f}_g^{(l)} \mathbf{f}_g^{(l)\top} \right\} \right\}, \end{aligned} \quad (2.39)$$

$$\begin{aligned} J_v(\sigma_0^{2(l-1)}) &= \sum_{i=1}^{M^{(l)}} \text{tr} \left\{ \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I \right)^{-1} \right. \\ &\quad \left. \left(\hat{\mathbf{v}}_i^{(l)} \hat{\mathbf{v}}_i^{(l)\top} + \hat{V}_i^{(l)} - 2(E^{(l-1)} \tilde{\mathbf{v}}^{(l-1)})_i \hat{\mathbf{v}}_i^{(l)\top} + (E^{(l-1)} \tilde{\mathbf{v}}^{(l-1)})_i (E^{(l-1)} \tilde{\mathbf{v}}^{(l-1)})_i^\top \right) \right\}. \end{aligned} \quad (2.40)$$

$J(\Theta_{res}^{(l)})$ を $\sigma_0^{2(l-1)}$ ， $\sigma_1^{2(l)}$ それぞれに関して偏微分する．ただし，式 (2.39) を $J_f = \gamma/\sigma_1^{2(l)}$ ，式 (2.40) を $J_v = \sum_i \text{tr}\{((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I)^{-1} \Xi_i\}$ とおく．

$$\begin{aligned} \frac{\partial J(\Theta_{res}^{(l)})}{\partial \sigma_0^{2(l-1)}} &= \sum_{i=1}^{M^{(l)}} \text{tr} \left\{ \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I \right)^{-1} \right\} \\ &\quad - \sum_{i=1}^{M^{(l)}} \text{tr} \left\{ \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I \right)^{-2} \Xi_i \right\} \end{aligned}$$

$$= \text{tr} \left[\sum_{i=1}^{M^{(l)}} \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I \right)^{-1} \left\{ I - \left((E^{(l-1)2} \tilde{V}^{(l-1)})_i + \sigma_0^{2(l-1)} I \right)^{-1} \Xi_i \right\} \right], \quad (2.41)$$

$$\frac{\partial J(\Theta_{res}^{(l)})}{\partial \sigma_1^{2(l)}} = \frac{N^{(l)} M^{(l)}}{\sigma_1^{2(l)}} - \frac{\gamma}{\sigma_1^{4(l)}}. \quad (2.42)$$

なお，式 (2.41) の導出にあたっては，微分公式

$$\frac{d}{dt} (\ln \det A(t)) = \text{tr} \left\{ A(t)^{-1} \frac{d}{dt} A(t) \right\} \quad (2.43)$$

を用いた．

式 (2.42) をゼロとおくことにより，更新値 $\hat{\sigma}_1^{2(l)}$ が次式で得られる．

$$\hat{\sigma}_1^{2(l)_{new}} = \frac{\gamma}{N^{(l)} M^{(l)}}. \quad (2.44)$$

$\sigma_0^{2(l-1)}$ に関しては（狭義の）一般化 EM アルゴリズムの枠組みを採用し，式 (2.41) に基づく探索として $\sigma_0^{2(l-1)}$ を変化させてこの反復での更新値とする．EM アルゴリズムは，E ステップで隠れ変数に関するエネルギー最小化，M ステップでパラメータに関するエネルギー最小化を行う変分的方法と解釈することができる．一般化 EM アルゴリズムとは，両ステップにおいて，最小化の代わりに以前の反復値よりもエネルギーを小さくする値を更新値とする拡張版であり，収束先は保存される．したがって，上記は M ステップに関する一般化 EM アルゴリズムに相当する．

次に， $l = 1$ （一層目）の計算手続きを考える．この場合の式 (2.38) に相当する評価式は次のようになる．ただし， $\{\bar{\mathbf{v}}, \sigma_{init}^2\}$ は画像全体で共通とする．

$$J(\sigma_1^{2(1)}, \bar{\mathbf{v}}_{init}, \sigma_{init}^2) = N^{(1)} M^{(1)} \ln \sigma_1^{2(1)} + 2M^{(1)} \ln \sigma_{init}^2 + J_f^{init}(\sigma_1^{2(1)}) + J_v^{init}(\bar{\mathbf{v}}, \sigma_{init}^2), \quad (2.45)$$

$$J_f^{init}(\sigma_1^{2(1)}) = \frac{1}{\sigma_1^{2(1)}} \left\{ \sum_{i=1}^{M^{(1)}} \sum_{\mathcal{N}_i^{(1)}} f_t^{(1)2} + 2 \sum_{i=1}^{M^{(1)}} \left(\sum_{\mathcal{N}_i^{(1)}} f_t^{(1)} \mathbf{f}_g^{(1)\top} \right) \hat{\mathbf{v}}_i^{(1)} + \sum_{i=1}^{M^{(1)}} \text{tr} \left\{ \left(\hat{\mathbf{v}}_i^{(1)} \hat{\mathbf{v}}_i^{(1)\top} + \hat{V}_i^{(1)} \right) \sum_{\mathcal{N}_i^{(1)}} \mathbf{f}_g^{(1)} \mathbf{f}_g^{(1)\top} \right\} \right\}, \quad (2.46)$$

$$J_v^{init}(\bar{\mathbf{v}}_{init}, \sigma_{init}^2) = \frac{1}{\sigma_{init}^2} \left\{ \sum_{i=1}^{M^{(1)}} \text{tr} \left(\hat{\mathbf{v}}_i^{(1)} \hat{\mathbf{v}}_i^{(1)\top} + \hat{V}_i^{(1)} \right) - 2 \left(\sum_{i=1}^{M^{(1)}} \hat{\mathbf{v}}_i^{(1)} \right)^\top \bar{\mathbf{v}}_{init} + M^{(1)} \bar{\mathbf{v}}_{init}^\top \bar{\mathbf{v}}_{init} \right\}. \quad (2.47)$$

ただし，各反復での MAP 推定値は以下ようになる．

$$\hat{\mathbf{v}}_i^{(1)} = \hat{V}_i^{(1)} \left\{ \frac{\hat{\mathbf{v}}_{init}}{\sigma_{init}^2} - \frac{\sum_{\mathcal{N}_i^{(1)}} f_t^{(1)} \mathbf{f}_g^{(1)}}{\sigma_1^2} \right\}, \quad (2.48)$$

$$\hat{V}_i^{(1)} = \left\{ \frac{1}{\sigma_{init}^2} + \frac{\sum_{\mathcal{N}_i^{(1)}} \mathbf{f}_g^{(1)} \mathbf{f}_g^{(1)\top}}{\sigma_1^2} \right\}^{-1}. \quad (2.49)$$

以上より，まず σ_1^2 の更新値は，式 (2.44) と同様に次式となる．

$$\sigma_1^{2, new} = \frac{\gamma}{N^{(1)} M^{(1)}}. \quad (2.50)$$

ただし， γ は式 (2.44) と同様の定義である． $\bar{\mathbf{v}}_{init}$ の更新値は

$$\hat{\mathbf{v}}_{init}^{new} = \sum_{i=1}^{M^{(1)}} \hat{\mathbf{v}}_i^{(1)}, \quad (2.51)$$

で与えられ，この値を代入した式 (2.47) の分子を χ とおくと， σ_{init}^2 の更新値は次式となる．

$$\sigma_{init}^{2, new} = \frac{\chi}{2M^{(1)}}. \quad (2.52)$$

第3章 時間統合処理に基づく動きからのベイズ的三次元構造推定

3.1 まえがき

structure from motion (SFM) の枠組みにおける研究は、従来より computer vision (CV) の中核をなし、膨大な数の成果が報告されてきた [3], [4], [5], [6]. 研究初期には、数学的立場での復元原理に関する研究（解の存在条件、解の個数等）が脚光を浴び、その後は高精度な復元手法の開発が進められ、同時に CV における統計的推論の在り方や新しい考え方が議論されてきた．そして昨今は、因子分解法 [7] に代表される安定で高効率な手法、カメラキャリブレーションを要しない復元法とその利用法、そしてそれらの結果を用いた VR 技術の開発に注目が集まっている．しかし、要素技術研究の立場では SFM が本質的に有する困難の完全な解決は依然残されていると言わざるを得ない．その本質的な困難とは、画像上の密な変位・速度場の正確な検出、ならびに動画像処理としての適切な時間統合であり、上記の近年注目されている技術は画像上での粗な特徴点追跡が何らかの方法で行えていることを前提とすることで、これらの困難を形式的に回避しているに過ぎない．本稿では上述の困難のうちの後者に焦点を当て、それを解決する一手法を検討する．

本研究では、密な奥行き情報の獲得を重視し、連続フレーム間で勾配法に基づいて解析的に計算されるオプティカルフロー（画像上での見掛けの二次元速度場、以下フローと略記する）から三次元構造を復元する立場を採用する [3], [4], [5], [6]. 勾配法に基づく構造復元の特徴は、基本的には各画素に対応する奥行きが密に得られること、すなわち、pixel-based な方法であることである．しかし、このことは時間統合の立場では欠点となる．すなわち、対象物体上の点の複数フレーム間にわたっての対応関係が得られないため、観測フレーム数の増加に伴う奥行き精度の向上を図るのが困難となる．

この困難を解決する従来法の多くは、各画素に対応する奥行き値を状態とするカルマンフィルタを適用している [13], [14]. しかしながら、フレーム間の三次元運動や入力雑音の分散、状態遷移モデルに含まれるパラメータ等が既知でなければならず、実用上、あるいは精度上問題があった．特に、三次元運動の値は可能な限り正確でなければ、状態遷移モデルのずれが大きくなり、奥行き値の情報伝搬が逆効果

に成りかねない．現実には，ひとまず各フレーム間で独立に奥行きを計算することにより同時に得られる三次元運動の値を用いることになり，精度の面で問題がある．すなわち，奥行きの時間相関を考慮した解としての三次元運動の値を，カルマンフィルタにおけるパラメータ値として採用すべきである．

以上のことから本稿では，通常のカルマンフィルタを拡張し，パラメータである三次元運動も同時に推定する枠組みを検討する．それによれば，三次元運動を含むパラメータの値は最尤 (ML) 推定として，奥行き値は周辺化事後確率最大化 (MPM) 推定として実現できる．

3.2 計算原理

3.2.1 オプティカルフローからの構造復元

図 1.1 に示すように，奥行き値に対応するカメラ座標系の Z 座標をレンズの光軸に合わせ，中心投影の投影面 (画像面) を $Z = 1$ の平面とする．3次元実空間内の物体上の点 (X, Y, Z) の投影点 $(x, y) \equiv (X/Z, Y/Z)$ におけるフロー $\boldsymbol{v} \equiv [v_x, v_y]^\top$ は次式を満足する．

$$v_x = xyr_x - (1 + x^2)r_y + yr_z - (u_x - xu_z)d, \quad (3.1)$$

$$v_y = (1 + y^2)r_x - xyr_y - xr_z - (u_y - yu_z)d. \quad (3.2)$$

$\boldsymbol{u} = [u_x, u_y, u_z]^\top$, $\boldsymbol{r} = [r_x, r_y, r_z]^\top$ はカメラの並進及び回転速度ベクトルであり，物体がカメラ座標系に対して $-\boldsymbol{u}$, $-\boldsymbol{r}$ の速度で運動しているとも考えることもできる．ただし，輝度値の運動前後での不変性を厳密に考慮するならば，例えば物体の光反射特性は完全な拡散反射であり，物体と光源が固定された状況でカメラが運動しているものとも考える必要がある． d は奥行き Z の逆数であり，画像上の各画素における未知数である．これに対し， $(\boldsymbol{u}, \boldsymbol{r})$ はすべての画素に共通する未知数となる．画像上での \boldsymbol{v} を観測量とすれば，式 (3.1) , (3.2) に基づいて $(\boldsymbol{u}, \boldsymbol{r})$ とフローが検出された画素に対応する d を求めることができる．

3.2.2 基本となる推定方式

本稿では，画像上の各画素 i でのフロー \boldsymbol{v}_i を観測量とする．また，各画素に対応する奥行き逆数 d_i を未知確率変数として取り扱う．ここで， \boldsymbol{v}_i に含まれる観測雑音 \boldsymbol{n}_{vi} が $N(0, V_i)$ に従うと仮定する．

$$\boldsymbol{v}_i = \boldsymbol{v}_i^r + \boldsymbol{v}_i^u d_i + \boldsymbol{n}_{vi}, \quad (3.3)$$

$$\begin{aligned}
\mathbf{v}_i^r &\equiv \begin{bmatrix} x_i y_i r_x - (1 + x_i^2) r_y + y_i r_z \\ (1 + y_i^2) r_x - x_i y_i r_y - x_i r_z \end{bmatrix} \\
&= \begin{bmatrix} x_i y_i & -(1 + x_i^2) & y_i \\ 1 + y_i^2 & -x_i y_i & -x_i \end{bmatrix} \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} \\
&\equiv R_i \mathbf{r},
\end{aligned} \tag{3.4}$$

$$\begin{aligned}
\mathbf{v}_i^u &\equiv - \begin{bmatrix} u_x - x u_z \\ u_y - y u_z \end{bmatrix} \\
&= \begin{bmatrix} -1 & 0 & x_i \\ 0 & -1 & y_i \end{bmatrix} \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} \\
&\equiv U_i \mathbf{u}.
\end{aligned} \tag{3.5}$$

$\{d_i\}$ が与えられた条件のもとでの $\{\mathbf{v}_i\}$ の確率は次式で与えられる．ただし，全画素数を N ， $\Theta_m \equiv \{\mathbf{u}, \mathbf{r}\}$ とし， V_i は既知とする．

$$p(\{\mathbf{v}_i\}|\{d_i\}, \Theta_m) = \frac{1}{(2\pi)^N \sqrt{\prod_i \det V_i}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^N \delta \mathbf{v}_i^\top V_i^{-1} \delta \mathbf{v}_i \right\}, \tag{3.6}$$

$$\delta \mathbf{v}_i \equiv \mathbf{v}_i - \mathbf{v}_i^r - \mathbf{v}_i^u d_i. \tag{3.7}$$

続いて， $\{d_i\}$ に関する事前確率が以下の正規分布として与えられるとする．

$$p(\{d_i\}|\Theta_d) = \frac{1}{(2\pi)^{N/2} \alpha^N \sqrt{\prod_i (\sigma_{d_i}^2 + \sigma^2)}} \exp \left\{ -\frac{1}{2\alpha^2} \sum_{i=1}^N \frac{(d_i - \alpha d_{mi})^2}{\sigma_{d_i}^2 + \sigma^2} \right\}. \tag{3.8}$$

ここで， $\alpha (> 0)$ は事前知識のスケール不定性を表現するパラメータ， σ^2 は画素 i によらずに加わる正規性ランダム変量の分散であり， $\Theta_d \equiv \{\alpha, \sigma^2\}$ とおいた． $\{\mathbf{v}_i\}$ が得られたときの $\{d_i\}$ の事後確率は，ベイズの公式により，以下のように計算できる．

$$p(\{d_i\}|\{\mathbf{v}_i\}, \Theta_m, \Theta_d) = \frac{p(\{\mathbf{v}_i\}|\{d_i\}, \Theta_m) p(\{d_i\}|\Theta_d)}{p(\{\mathbf{v}_i\}|\Theta_m, \Theta_d)}. \tag{3.9}$$

上式の分母はその分子を $\{d_i\}$ に関して周辺化したものであり，パラメータ $\{\Theta_m, \Theta_d\}$ の最尤推定量 (MLE) $\{\Theta_{mML}, \Theta_{dML}\}$ は， $\{\mathbf{v}_i\}$ を観測量に固定した次式によって与えられる．ただし，本問題では \mathbf{u} と $\{d_i\}$ が積として現れるため，両者の大きさ

を同時に決定することはできない．そこで，本稿では制約条件 $\|\mathbf{u}\| = 1$ を課すこととする．

$$\{\Theta_{mML}, \Theta_{dML}\} = \arg_{\max} \{\ln p(\{\mathbf{v}_i\}|\Theta_m, \Theta_d)\}. \quad (3.10)$$

式 (3.9) は正規分布であり，その平均と分散は次のようになる．

$$\tilde{d}_i(\Theta_m, \Theta_d) = \tilde{\sigma}_{di}^2 \left\{ \mathbf{v}_i^{u\top} V_i^{-1} (\mathbf{v}_i - \mathbf{v}_i^r) + \frac{d_{mi}}{\alpha(\sigma_{di}^2 + \sigma^2)} \right\}, \quad (3.11)$$

$$\tilde{\sigma}_{di}^2(\Theta_m, \Theta_d) = \left\{ \mathbf{v}_i^{u\top} V_i^{-1} \mathbf{v}_i^u + \frac{1}{\alpha^2(\sigma_{di}^2 + \sigma^2)} \right\}^{-1}. \quad (3.12)$$

ここで $\tilde{\cdot}$ は，事後分布の母数であることを指す．したがって，上式に $\{\Theta_{mML}, \Theta_{dML}\}$ を代入することで， d_i の現実的な MAP 推定（ウィナーフィルタリング）を行うことができる．

3.3 情報伝搬による逐次復元

3.3.1 逐次復元の基本的考え方

以降では，時刻 t のフロー及び奥行き逆数を $\{\mathbf{v}_i^{(t)}\}$ ， $\{d_i^{(t)}\}$ と記述する．本稿では，カルマンフィルタの枠組みを採用し，観測量 $\{\mathbf{v}_i^{(j)}\}_{j=1, \dots, t}$ に基づいて各時刻の状態 $\{d_i^{(t)}\}$ の MPM 推定量を逐次的に求めることとする．それにあたり，パラメータ $\Theta \equiv \{\Theta^{(j)}\}_{j=1, \dots, t}$ も観測量から決定する方針をとる．ただし本稿では，パラメータは各時刻の観測を規定するもののみ存在し，全観測を共通に規定するパラメータは存在しないものとする．

各時刻のパラメータ $\Theta^{(t)}$ の，その時刻までの観測に基づく MLE は，形式的に一様分布に従う事前確率 $p(\Theta^{(t)})$ を仮定する MPM 推定量と等価である．したがって，計算すべき $\Theta^{(t)}$ の周辺事後確率は， $t \geq 2$ については以下ようになる．

$$\begin{aligned} p(\Theta^{(t)}|\{\mathbf{v}_i^{(t)}\}, \{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}) &= \frac{p(\Theta^{(t)})}{p(\{\mathbf{v}_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\})} \\ &\times \int p(\{\mathbf{v}_i^{(t)}\}|\{d_i^{(t)}\}, \Theta^{(t)}) p(\{d_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)}) d\{d_i^{(t)}\}. \end{aligned} \quad (3.13)$$

ただし，被積分関数の第二因子は，該当する時刻のパラメータのみで規定される奥行き事後確率であり，次のように定式化される．

$$p(\{d_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)})$$

$$= \int p(\{d_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)}, \Theta^{(t-1)})p(\Theta^{(t-1)}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\})d\Theta^{(t-1)}. \quad (3.14)$$

上式の被積分関数の第二因子は一つ前の時刻のパラメータの周辺事後確率である．したがって，式 (3.13) は過去の結果を用いて逐次的に計算可能であり，この式を最大にする値として $\Theta_{ML}^{(t)}$ を得ることができる．

式 (3.14) 中の $\Theta^{(t-1)}$ に関する周辺化は一般に困難であり，適当な近似を用いる必要がある．ここでは最も簡単な近似として，鞍点近似を採用する．すなわち，既に求まっている $\Theta_{ML}^{(t-1)}$ を用いた次の近似を用いる．

$$\begin{aligned} p(\{d_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \mathbf{v}_i^{(1)}, \Theta^{(t)}) \\ \approx p(\{d_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)}, \Theta_{ML}^{(t-1)}) \end{aligned} \quad (3.15)$$

この近似に従えば，式 (3.13) から $\Theta^{(t)}$ の推定に必要な部分を次のように取り出すことができる．

$$\begin{aligned} \int p(\{\mathbf{v}_i^{(t)}\}|\{d_i^{(t)}\}, \Theta^{(t)})p(\{d_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)}, \Theta_{ML}^{(t-1)})d\{d_i^{(t)}\} \\ \approx \int p(\{\mathbf{v}_i^{(t)}\}, \{d_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)})d\{d_i^{(t)}\} \\ \approx p(\{\mathbf{v}_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)}). \end{aligned} \quad (3.16)$$

$\{d_i^{(t)}\}$ の事後分布は，式 (3.16) 及びその左辺の被積分関数をそれぞれ分母，分子とする次式で与えられる．

$$\begin{aligned} p(\{d_i^{(t)}\}|\{\mathbf{v}_i^{(t)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)}) \\ = \frac{p(\{\mathbf{v}_i^{(t)}\}|\{d_i^{(t)}\}, \Theta^{(t)})p(\{d_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)}, \Theta_{ML}^{(t-1)})}{p(\{\mathbf{v}_i^{(t)}\}|\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta^{(t)})}. \end{aligned} \quad (3.17)$$

したがって， $\{d_i^{(t)}\}$ の MPM 推定量は，上式に $\Theta_{ML}^{(t)}$ を代入したものを最大にすることで得られる．以上の推定方式は，カルマンフィルタにパラメータの逐次推定機能を付加したものに相当する．

3.3.2 具体的定式化

式 (3.16) の被積分関数 ($\{\mathbf{v}_i^{(t)}, d_i^{(t)}\}$ の条件付き同時確率) を具体的に導出する．そのために，状態 $\{d_i^{(t)}\}$ の時間遷移モデルを次式で定義する．

$$d_i^{(t)} = \alpha^{(t)} \left(E_i^{(t-1)} d_i^{(t-1)} + n_{di}^{(t)} \right). \quad (3.18)$$

ここで、画素 i における $t-1$ から t への平滑化演算子を $E_i^{(t-1)}$ とする．具体的には、 $t-1$ で得られた奥行き逆数及び運動パラメータを用いて式 (3.1), (3.2) で計算されるフロー推定値に基づき、 t での画素 i に対応する $t-1$ フレーム中の位置 (サブピクセル) を計算し、その位置を中心とするガウス型重み関数を用いた $\{d_i^{(t-1)}\}$ の空間平均を i での奥行き逆数とする． $n_{di}^{(t)}$ はこの平滑化結果からのずれを表す確率変数で、他の確率変数とは独立な平均 0、共分散 $\sigma^{2(t)}$ の正規分布に従うとする．

ここで、この平滑化モデルの意味を説明する． $d_i^{(t)}$ の値は、 t における i に対応する $t-1$ での座標値によって指定される．この座標値 (η_i, ξ_i) は、 $t-1$ における i でのフロー $\tilde{v}_i^{(t-1)}$ を用い、 $(\eta_i, \xi_i) \equiv (x_i, y_i) - \tilde{v}_i^{(t-1)}$ として近似的に算出できる．ただしこの $\tilde{v}_i^{(t-1)}$ は、 $d_i^{(t-1)}$ 及び $\Theta_{mML}^{(t-1)}$ を式 (3.1), (3.2) に代入して得られる推定量とする．ここで、もしこの座標値に誤差がなければ、そこでの $t-1$ における奥行き逆数 (通常はサブピクセル値となるため、補間処理が必要) を $d_i^{(t)}$ とすることができる．しかし実際には、 (η_i, ξ_i) には無視し得ない誤差が含まれる．

まず考慮すべきは、本来は $t-1$ の画像上で i を終点とするフローの始点を (η_i, ξ_i) として求めなければならないことである．したがって、上記の近似は局所的にフローが一定であることを仮定していることになる．この誤差のモデルとして、本稿では i の近傍のフロー推定値を用い、分散が次式で定義され、平均が 0 となる二次元正規分布を採用する．

$$V_{f1_i}^{(t-1)} \equiv E \left[\frac{1}{N_i} \sum_{j \in \mathcal{N}_i} (\tilde{v}_j^{(t-1)} - \tilde{v}_i^{(t-1)}) (\tilde{v}_j^{(t-1)} - \tilde{v}_i^{(t-1)})^\top \mid \mathbf{v}_i^{(t-2)}, \dots, \mathbf{v}_i^{(1)} \right]. \quad (3.19)$$

ここで、 \mathcal{N}_i は i の近傍を表し、近傍の画素数を N_i とする．また、 $E[\cdot \mid \mathbf{v}_i^{(t-2)}, \dots, \mathbf{v}_i^{(1)}]$ は条件付き期待値操作を意味する．式 (3.19) は具体的に次のように書くことができる．ただし、 $\Theta_{mML}^{(t-1)}$ の推定誤差、および $d_i^{(t-1)}$ と $\{d_j^{(t-1)}\} (j \in \mathcal{N}_i)$ との間の相関は無視する．

$$\begin{aligned} V_{f1_i}^{(t-1)} &= \frac{1}{N_i} \sum_j \left\{ \tilde{v}_j^r \tilde{v}_j^{r\top} + (\tilde{v}_j^r \tilde{v}_j^{u\top} + \tilde{v}_j^u \tilde{v}_j^{r\top}) \tilde{d}_j^{(t-1)} + \tilde{v}_j^u \tilde{v}_j^{u\top} (\tilde{d}_j^{(t-1)2} + \sigma_{dj}^2)^{(t-1)} \right\} \\ &\quad - \frac{1}{N_i} \sum_j (\tilde{v}_j^r + \tilde{d}_j^{(t-1)} \tilde{v}_j^u) \tilde{v}_i^{r\top} - \frac{1}{N_i} \tilde{v}_i^r \sum_j (\tilde{v}_j^{r\top} + \tilde{d}_j^{(t-1)} \tilde{v}_j^{u\top}) \\ &\quad - \left(\frac{1}{N_i} \sum_j \tilde{v}_j^r \right) \tilde{v}_i^{u\top} \tilde{d}_i^{(t-1)} - \tilde{v}_i^u \left(\sum_j \tilde{v}_j^{r\top} \right) \tilde{d}_i^{(t-1)} \\ &\quad + (\tilde{v}_i^r \tilde{v}_i^{u\top} + \tilde{v}_i^u \tilde{v}_i^{r\top}) \tilde{d}_i^{(t-1)} + \tilde{v}_i^u \tilde{v}_i^{u\top} (\tilde{d}_i^{(t-1)2} + \sigma_{di}^2)^{(t-1)} + \tilde{v}_i^r \tilde{v}_i^{r\top}. \quad (3.20) \end{aligned}$$

$\tilde{d}_*^{(t-1)}$, σ_{d*}^2 は、 $p(d_*^{(t-1)} \mid \mathbf{v}_*^{(t-2)}, \dots, \mathbf{v}_*^{(1)}, \Theta_d^{(t-1)})$ の平均及び分散である．また、 \mathbf{v}_*

の添字 $(t-1)$ は省略した .

次に考慮すべき (η_i, ξ_i) の誤差として , $d_i^{(t-1)}$ の推定誤差に起因して生じる $\tilde{v}_i^{(t-1)}$ 自身の揺らぎがある . この揺らぎは , 上と同様に $\Theta_{mML}^{(t-1)}$ の誤差を無視すると , $\tilde{d}_i^{(t-1)}$ を用いて計算される $\tilde{v}_i^{(t-1)}$ に基づいて決まる (η_i, ξ_i) を平均 $\bar{x}_i^{(t-1)}$ とする一次元正規分布となり , その分散は次式で与えられる .

$$V_{f2_i}^{(t-1)} = \sigma_{di}^2 \tilde{v}_i^{u(t-1)} \tilde{v}_i^{u(t-1)\top}. \quad (3.21)$$

以上の二つの誤差が独立であると仮定すると , 平滑化モデルに用いるガウス型重み関数は , 次のように定量化できる . ただし , A は正規化定数である .

$$\begin{aligned} w_i^{(t-1)}(\mathbf{x}) &= A \exp \left\{ -\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_i^{(t-1)})^\top \left(V_{f1_i}^{(t-1)} + V_{f2_i}^{(t-1)} \right)^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i^{(t-1)}) \right\}. \end{aligned} \quad (3.22)$$

式 (3.18) の $E_i^{(t-1)}$ を式 (3.22) を用いた平滑化と定義すると , $t \geq 2$ に対する式 (3.16) 左辺の被積分関数の第二因子は , 式 (3.8) と同様に次式で与えられる .

$$\begin{aligned} p(\{d_i^{(t)}\} | \{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}, \Theta_d^{(t)}) &= \frac{1}{(2\pi)^{N/2} \alpha^{(t)N} \sqrt{\prod_i (E_i^{(t-1)})^2 \sigma_{di}^2 \tilde{d}_i^{(t-1)} + \sigma^2(t)}} \\ &\times \exp \left\{ -\frac{1}{2\alpha^{(t)2}} \sum_{i=1}^N \frac{\left(d_i - \alpha^{(t)} E_i^{(t-1)} \tilde{d}_i^{(t-1)} \right)^2}{E_i^{(t-1)2} \sigma_{di}^2 \tilde{d}_i^{(t-1)} + \sigma^2(t)} \right\}. \end{aligned} \quad (3.23)$$

なお , ここでも隣接する $d_i^{(t)}$ 間の共分散を無視している . 演算子 $E_i^{(t-1)2}$ は , 演算子 $E_i^{(t-1)}$ における重み係数をそれぞれ二乗したものをその重み係数とする平滑化演算子である .

続いて観測量 $\mathbf{v}_i(t)$ の確率モデルを考える . 観測雑音を含む観測方程式を , 式 (3.3) と同様に次式で定義する .

$$\mathbf{v}_i^{(t)} = \mathbf{v}_i^{r(t)} + \mathbf{v}_i^{u(t)} d_i + \mathbf{n}_{vi}^{(t)}. \quad (3.24)$$

したがって , 式 (3.16) 左辺の被積分関数の第一因子は , 式 (3.6) と同様に次式で与えられる .

$$p(\{\mathbf{v}_i^{(t)}\} | \{d_i^{(t)}\}, \Theta_m^{(t)}) = \frac{1}{(2\pi)^N \sqrt{\prod_i \det V_i^{(t)}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^N \delta \mathbf{v}_i^{(t)\top} V_i^{(t)-1} \delta \mathbf{v}_i^{(t)} \right\}, \quad (3.25)$$

$$\delta \mathbf{v}_i^{(t)} \equiv \mathbf{v}_i^{(t)} - \mathbf{v}_i^r(t) - \mathbf{v}_i^u(t) d_i^{(t)}. \quad (3.26)$$

$\{\mathbf{v}_i^{(t)}, d_i^{(t)}\}$ の条件付き同時確率は，式 (3.23) と式 (3.25) の積となり，よって正規分布に従うことがわかる．この同時確率は， $\{d_i^{(t)}\}$ の事後確率（式 (3.17)）の分子に相当し，ゆえにその事後確率の平均と分散は，各 $d_i^{(t)}$ に対して次式のように定まる．

$$\tilde{d}_i^{(t)}(\Theta_m^{(t)}, \Theta_d^{(t)}) = \sigma_{di}^2(t) \left\{ \mathbf{v}_i^{u(t)\top} V_i^{(t)-1} (\mathbf{v}_i^{(t)} - \mathbf{v}_i^r(t)) + \frac{E_i^{(t-1)} \tilde{d}_i^{(t-1)}}{\alpha^{(t)} \left(E_i^{(t-1)2} \tilde{\sigma}_{di}^2(t-1) + \sigma^2(t) \right)} \right\}, \quad (3.27)$$

$$\tilde{\sigma}_{di}^2(t)(\Theta_m^{(t)}, \Theta_d^{(t)}) = \left\{ \mathbf{v}_i^{u(t)\top} V_i^{(t)-1} \mathbf{v}_i^{u(t)} + \frac{1}{\alpha^{(t)2} \left(E_i^{(t-1)2} \tilde{\sigma}_{di}^2(t-1) + \sigma^2(t) \right)} \right\}^{-1}. \quad (3.28)$$

以上より， $t \geq 2$ に関しては式 (3.23) と式 (3.25) の積を $\{d_i^{(t)}\}$ に関して周辺化して得られる条件付き尤度関数（式 (3.16) を具体的に定式化したもの）より $\Theta_{ML}^{(t)}$ を求め，その結果を式 (3.27)，式 (3.28) に代入することで， $\{d_i^{(t)}\}$ の MPM 推定値が得られる．

$t = 1$ の場合は， $\{d_i\}$ に関する陽な事前知識がないことから，式 (3.8) に相当する事前確率として簡単に次式を用いることができる．

$$p(\{d_i^{(1)}\} | \Theta_{init}) = \frac{1}{(2\pi\sigma_{init}^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma_{init}^2} \sum_{i=1}^N (d_i^{(1)} - d_{init})^2 \right\}. \quad (3.29)$$

ここで， $\Theta_{init} \equiv \{d_{init}, \sigma_{init}^2\}$ である．観測方程式は式 (3.24) で与えられることから， $\{\mathbf{v}_i^{(1)}\}$ ， $\{d_i^{(1)}\}$ の同時確率は，式 (3.25) と式 (3.29) の積で与えられ，これを $\{d_i^{(1)}\}$ で周辺化して尤度関数とすることで， Θ_{initML} を得ることができる．また，その同時確率を規格化して得られる式 (3.9) に相当する $\{d_i^{(1)}\}$ の事後確率の平均及び分散は，次式で与えられる．

$$\tilde{d}_i^{(1)}(\Theta_{init}) = \sigma_{di}^2(1) \left\{ \mathbf{v}_i^{u(1)\top} V_i^{(1)-1} (\mathbf{v}_i^{(1)} - \mathbf{v}_i^r(1)) + \frac{d_{init}}{\sigma_{init}^2} \right\}, \quad (3.30)$$

$$\tilde{\sigma}_{di}^2(1)(\Theta_{init}) = \left\{ \mathbf{v}_i^{u(1)\top} V_i^{(1)-1} \mathbf{v}_i^{u(1)} + \frac{1}{\sigma_{init}^2} \right\}^{-1}. \quad (3.31)$$

3.4 計算アルゴリズム

3.4.1 EM アルゴリズムの適用

式 (3.16) の最大化によるパラメータ推定，ならびにその値に基づく式 (3.27) の計算を効率的に行うために，EM アルゴリズムの枠組みを採用することが可能である．EM アルゴリズムとは観測されていない未知確率変数（隠れ変数）を観測量に加えて想定したときに，両者（完全データ）が観測されれば問題が解きやすくなる場合に有効な推定スキームであり，その未知確率変数の事後分布とパラメータの ML 推定量を同時に得ることのできる反復手続きである [12]．詳細は省略するが，各反復では次の 2 つのステップを実行する．

1. E-step：パラメータを以前の反復で得られている値としたときの隠れ変数の事後分布を求め，これを利用して Q 関数と呼ばれるパラメータ更新のための評価関数を算出する．
2. M-step：Q 関数を最大化するようにパラメータ値を更新する．

以降では，EM アルゴリズムを用いて各 t で $\Theta^{(t)}$ の ML 推定と，この値を固定したときの $\{d_i^{(t)}\}$ の MPM 推定を行うための，各層での具体的計算手続きを導出する．

3.4.2 計算手続き

時刻 $t (\geq 2)$ の画像に対して，それ以前の情報に基づいて $\Theta^{(t)} = \{\mathbf{u}^{(t)}, \mathbf{r}^{(t)}, \alpha^{(t)}, \sigma^{2(t)}\}$ の ML 推定と， $\{d_i^{(t)}\}$ の MPM 推定を行う計算式を導出する．E-step では次のように定義される $\Theta^{(t)}$ に関する関数（Q 関数）を具体的に導出する．

$$Q(\Theta^{(t)}; \hat{\Theta}^{(t)}) \equiv \mathbb{E} \left[\ln p(\{\mathbf{v}_i^{(t)}\}, \{d_i^{(t)}\} | \Theta^{(t)}) \mid \{\mathbf{v}_i^{(t)}\}, \hat{\Theta}^{(t)} \right]. \quad (3.32)$$

ただし，この式を含めて以降では， $\{\{\mathbf{v}_i^{(t-1)}\}, \dots, \{\mathbf{v}_i^{(1)}\}\}$ に関する条件付きであることを省略する．また， $\hat{\cdot}$ はその反復時に得られている値，あるいはそれに依存する変数とする．式 (3.32) において期待値計算を施される関数は，式 (3.23)，式 (3.25) を用いて，次式で与えられる．

$$\begin{aligned} & -N \ln \alpha^{(t)} - \frac{1}{2} \sum_{i=1}^N \ln \left(E_i^{(t-1)2} \tilde{\sigma}_{di}^{2(t-1)} + \sigma^{2(t)} \right) \\ & - \frac{1}{2} \left(\mathbf{v}_i^{(t)} - \mathbf{v}_i^{r(t)} - \mathbf{v}_i^{u(t)} d_i^{(t)} \right)^\top V_i^{-1} \left(\mathbf{v}_i^{(t)} - \mathbf{v}_i^{r(t)} - \mathbf{v}_i^{u(t)} d_i^{(t)} \right) \\ & - \frac{1}{2\alpha^{(t)2}} \sum_{i=1}^N \frac{\left(d_i^{(t)} - \alpha^{(t)} E_i^{(t-1)} \tilde{d}_i^{(t-1)} \right)^2}{E_i^{(t-1)2} \tilde{\sigma}_{di}^{2(t-1)} + \sigma^{2(t)}} + \text{Const..} \end{aligned} \quad (3.33)$$

一方, $\{\mathbf{v}_i^{(t)}\}$ 観測後の $d_i^{(t)}$ の各反復での事後分布 $p(d_i^{(t)}|\mathbf{v}_i^{(t)}, \hat{\Theta}^{(t)})$ は, 式 (3.27), (3.28) より, 平均及び分散がその反復での $\hat{\Theta}^{(t)}$ を用いて以下で表される正規分布となる.

$$\begin{aligned} & \hat{d}_i^{(t)}(\hat{\Theta}_m^{(t)}, \hat{\Theta}_d^{(t)}) \\ &= \hat{\sigma}_{di}^{2(t)} \left\{ \mathbf{v}_i^{u(t)\top} V_i^{(t)-1} \left(\mathbf{v}_i^{(t)} - \mathbf{v}_i^{r(t)} \right) + \frac{E_i^{(t-1)} \tilde{d}_i^{(t-1)}}{\hat{\alpha}^{(t)} \left(E_i^{(t-1)2} \tilde{\sigma}_{di}^{2(t-1)} + \sigma^2 \right)} \right\}, \end{aligned} \quad (3.34)$$

$$\hat{\sigma}_{di}^{2(t)}(\hat{\Theta}_m^{(t)}, \hat{\Theta}_d^{(t)}) = \left\{ \mathbf{v}_i^{u(t)\top} V_i^{(t)-1} \mathbf{v}_i^{u(t)} + \frac{1}{\hat{\alpha}^{(t)2} \left(E_i^{(t-1)2} \tilde{\sigma}_{di}^{2(t-1)} + \sigma^2 \right)} \right\}^{-1}. \quad (3.35)$$

ゆえに, 式 (3.33) を $p(d_i^{(t)}|\mathbf{v}_i^{(t)}, \hat{\Theta}^{(t)})$ に関して期待値をとることで, 式 (3.32) が具体的に得られる. M-step ではこの $Q(\Theta^{(t)}; \hat{\Theta}^{(t)})$ を $\Theta^{(t)}$ に関して最大化する. そこで, $Q(\Theta^{(t)}; \hat{\Theta}^{(t)})$ を -2 倍して定数を無視し, 次の関数の最小化を行うものとする.

$$J(\Theta^{(t)}) = J_v(\Theta_m^{(t)}) + J_d(\Theta_d^{(t)}), \quad (3.36)$$

$$\begin{aligned} & J_v(\Theta_m^{(t)}) \\ &= \sum_{i=1}^N \left\{ \left(\mathbf{v}_i^{(t)} - \mathbf{v}_i^{r(t)} \right)^\top V_i^{-1} \left(\mathbf{v}_i^{(t)} - \mathbf{v}_i^{r(t)} \right) + 2 \left(\mathbf{v}_i^{(t)} - \mathbf{v}_i^{r(t)} \right)^\top V_i^{-1} \mathbf{v}_i^{u(t)} \hat{d}_i^{(t)} \right. \\ & \quad \left. + \left(\mathbf{v}_i^{u(t)\top} V_i^{-1} \mathbf{v}_i^{u(t)} \right) \left(\hat{d}_i^{(t)} + \hat{\sigma}_{di}^{2(t)} \right) \right\}, \end{aligned} \quad (3.37)$$

$$\begin{aligned} & J_d(\Theta_d^{(t)}) \\ &= 2N \ln \alpha^{(t)} + \sum_{i=1}^N \ln \left(E_i^{(t-1)2} \tilde{\sigma}_{di}^{2(t-1)} + \sigma^2 \right) + \frac{1}{\alpha^{(t)2}} \sum_{i=1}^N \frac{\hat{d}_i^{(t)2} + \hat{\sigma}_{di}^{2(t)}}{E_i^{(t-1)2} \tilde{\sigma}_{di}^{2(t-1)} + \sigma^2} \\ & \quad - \frac{2}{\alpha^{(t)}} \sum_{i=1}^N \frac{E_i^{(t-1)} \tilde{d}_i^{(t-1)} \hat{d}_i^{(t)}}{E_i^{(t-1)2} \tilde{\sigma}_{di}^{2(t-1)} + \sigma^2} + \sum_{i=1}^N \frac{\left(E_i^{(t-1)} \tilde{d}_i^{(t-1)} \right)^2}{E_i^{(t-1)2} \tilde{\sigma}_{di}^{2(t-1)} + \sigma^2}. \end{aligned} \quad (3.38)$$

まず, J_v を $\mathbf{r}^{(t)}$ で微分して次式が得られる.

$$\frac{\partial J_v}{\partial \mathbf{r}} = -2 \sum_{i=1}^N R_i^\top V_i^{-1} \mathbf{v}_i^{(t)} + 2 \left(\sum_{i=1}^N R_i^\top V_i^{-1} R_i \right) \mathbf{r}^{(t)} - 2 \left(\sum_{i=1}^N \hat{d}_i^{(t)} R_i^\top V_i^{-1} U_i \right) \mathbf{u}^{(t)}. \quad (3.39)$$

\mathbf{u} に関しては, $\|\mathbf{u}\| = 1$ の拘束を考慮し, $\mathbf{u} = [\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta]^\top$ とおいて, $\{\theta, \phi\}$ で微分する.

$$\begin{aligned} \frac{\partial J_v}{\partial \{\theta, \phi\}} &= 2 \frac{\partial \mathbf{u}}{\partial \{\theta, \phi\}}^\top \sum_{i=1}^N \hat{d}_i^{(t)} U_i^\top V_i^{-1} \mathbf{v}_i^{(t)} - 2 \frac{\partial \mathbf{u}}{\partial \{\theta, \phi\}}^\top \left(\sum_{i=1}^N \hat{d}_i^{(t)} U_i^\top V_i^{-1} R_i \right) \mathbf{r}^{(t)} \\ &\quad + 2 \frac{\partial \mathbf{u}}{\partial \{\theta, \phi\}}^\top \left(\sum_{i=1}^N (\hat{d}_i^{(t)^2} + \sigma_{d_i}^{2(t)}) U_i^\top V_i^{-1} U_i \right) \mathbf{u}^{(t)}. \end{aligned} \quad (3.40)$$

ただし,

$$\frac{\partial \mathbf{u}}{\partial \{\theta, \phi\}} = \begin{bmatrix} \cos \theta \cos \phi & -\sin \theta \sin \phi \\ \cos \theta \sin \phi & \sin \theta \cos \phi \\ -\sin \theta & 0 \end{bmatrix}. \quad (3.41)$$

したがって, 式 (3.39), (3.40) を用いて, 例えば最急降下法によって J_v の最小化を行うことにより, $\{\mathbf{u}^{(t)}, \mathbf{r}^{(t)}\}$ の更新値を得ることができる.

次に, 式 (3.38) を $J_d = 2N \ln \alpha^{(t)} + a/\alpha^{(t)^2} - 2b/\alpha^{(t)} + c$ とおくと, J_d を $\alpha^{(t)}$ で微分して 0 とおくことにより次式が得られる.

$$N\alpha^{(t)^2} + b\alpha^{(t)} - a = 0. \quad (3.42)$$

ゆえに, $\alpha > 0$ を用いると, $\alpha^{(t)}$ の更新値が次式で得られる.

$$\alpha_{new}^{(t)} = \frac{-b + \sqrt{b^2 + 4Na}}{2N}. \quad (3.43)$$

$\sigma^{2(t)}$ に関しては, 数値探索を実行する必要がある. ただし, 一般化 EM アルゴリズムの枠組みを採用すれば, J_d の完全な最小化を行う必要はなく, 前回の $\hat{\sigma}^{2(t)}$ よりも J_d を小さくする値に更新すればよい. したがって, $\Theta_d^{(t)}$ の更新にあたっては, まず数値探索によって $\sigma^{2(t)}$ の更新値を定め, この値を用いて式 (3.43) により $\alpha^{(t)}$ の更新を行えばよい.

次に, $l = 1$ (一層目) の計算手続きを考える. この場合の式 (3.36) に相当する評価式は次のようになる. ただし, $\Theta_{init} = \{d_{init}, \sigma_{init}^2\}$ は画像全体で共通とする.

$$J^{init}(\Theta_m^{(1)}, \Theta_{init}) = J_v(\Theta_m^{(1)}) + J_d^{init}(\Theta_{init}), \quad (3.44)$$

$$J_d^{init}(\Theta_{init}) = N \ln \sigma_{init}^2 + \frac{1}{\sigma_{init}^2} \left\{ \sum_{i=1}^N \left(\hat{d}_i^{(1)2} + \sigma_{di}^2 \right) - d_{init} \sum_{i=1}^N \hat{d}_i^{(1)} + N d_{init}^2 \right\}. \quad (3.45)$$

$J_v(\Theta_m^{(1)})$ は, 式 (3.37) において $t = 1$ としたものである. また, 各反復での $d_i^{(1)}$ の平均と分散は以下ようになる.

$$\hat{d}_i^{(1)}(\hat{\Theta}_m^{(1)}, \hat{\Theta}_{init}^{(1)}) = \sigma_{di}^2 \left\{ \hat{\mathbf{v}}_i^{u(1)\top} V_i^{(1)-1} \left(\mathbf{v}_i^{(1)} - \hat{\mathbf{v}}_i^{r(1)} \right) + \frac{d_{init}}{\sigma_{init}^2} \right\}, \quad (3.46)$$

$$\sigma_{di}^2(\hat{\Theta}_m^{(1)}, \hat{\Theta}_{init}^{(1)}) = \left\{ \hat{\mathbf{v}}_i^{u(1)\top} V_i^{(1)-1} \hat{\mathbf{v}}_i^{u(1)} + \frac{1}{\sigma_{init}^2} \right\}^{-1}. \quad (3.47)$$

以上より, まず J_d^{init} を d_{init} で微分して 0 とおくと, 次の更新式が得られる.

$$d_{init}^{new} = \frac{\sum_i \hat{d}_i^{(1)}}{N}. \quad (3.48)$$

この値を式 (3.45) に代入し, σ_{init}^2 で微分して 0 とおくことにより, 次式が得られる.

$$\sigma_{init}^{2, new} = \frac{\sum_i \left(\hat{d}_i^{(1)} - d_{init}^{new} \right)^2 + \sum_i \sigma_{di}^2}{N}. \quad (3.49)$$

$\Theta_m^{(1)} = \{\mathbf{u}^{(1)}, \mathbf{r}^{(1)}\}$ に関しては, 式 (3.39), (3.40) で $t = 1$ としたものを勾配とみて, 同様の数値探索により更新値を得ることができる.

第4章 数値評価

4.1 評価の目的

今回提案するパラメータ決定機能を有するカルマンフィルタの有効性を確認する目的で、人工画像を用いた数値評価を行った。ここでは特に、本アルゴリズムを解像度方向の情報伝搬に適用した場合の結果を示す。

4.2 評価条件

正弦波状のテクスチャを持つ平面を想定し、この平面の運動前後の画像を生成した。1枚目の画像を図4.1に示す。原理的な確認を行う主旨で、画像の時空間偏微分は理論的に計算し、 f_t について雑音を付加した。

多重解像度画像への分解には、ラプラシアンガウシアン演算子によるバンドパスフィルタを用いた。設定した運動の種類は以下の3通りである。

1. 平行移動 (x 方向の並進 u_x のみ)
2. 湧き出し (z 方向の並進 u_z のみ)
3. 一般 (回転・並進)

また、運動の大きさと雑音に関して、以下の4通りを考えた。

1. 運動が小さい場合 (雑音なし)
2. 運動が中程度の場合 (雑音なし)
3. 運動が大きい場合 (雑音なし)
4. 運動が大きい場合 (ダイナミックレンジの1%の雑音あり)

帯域分割数は以下の通りとした。

1. 1層 (全帯域を一括で使用: single)
2. 2層 (低域で1個のフローを推定して高域に伝搬: dual-1)

3. 2層（全帯域を2個に分割：dual-2）
4. 3層（全帯域を3個に分割：triad）
5. 4層（全帯域を4個に分割：quad）
6. 8層（全帯域を8個に分割：octad）

4.3 評価結果

オプティカルフローの設定値と推定結果を図4.2から図4.13に、また推定されたフローの真値からの自乗誤差を表4.1から表4.12に示す。

これらの結果において、情報伝搬を行わない場合はエイリアスによる大きな誤差が見られるのに対し、情報伝搬を行うことによりそのような誤差はほぼ取り除けることがわかる。また、今回程度の大きさの雑音には影響されないことも確認できる。

フローの真値と推定値との平均自乗誤差を見ると、帯域分割数が1から3にかけては誤差が小さくなり、それ以上に分割すると若干誤差が増大していることがわかる。これは、パラメータ σ_0^2, σ_1^2 についても同様であった。今回使用した画像は、低域成分が主要であり、高域成分ではS/Nが低下している。そのため、パラメータの推定精度が低下し、その結果としてフロー推定に大きな誤差が生じるものと解釈できる。この改善策としては、パラメータの推定精度も考慮した情報伝搬スキームの検討、あるいは帯域分割数を対象画像の特性に応じて適応的に決定する方法の開発等が考えられる。

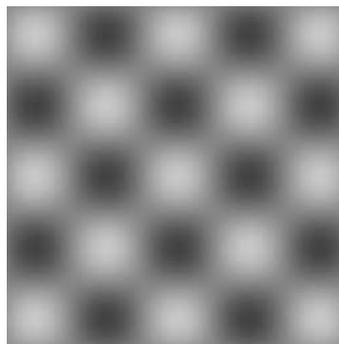


図 4.1: 入力画像の1フレーム目

表 4.1: フローの平均自乗誤差 (並行移動 (小))

帯域分割数	v_x	v_y
1(single)	1.01e-5	2.35e-6
2(dual-1)	9.03e-22	1.55e-21
2(dual-2)	9.12e-27	1.60e-26
3(triad)	6.78e-22	1.25e-21
4(quad)	2.03e-21	3.45e-21
8(octad)	3.88e-24	7.10e-24

表 4.2: フローの平均自乗誤差 (並行移動 (中))

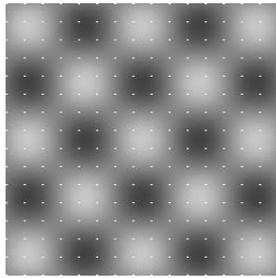
帯域分割数	v_x	v_y
1(single)	2.54e-4	5.88e-5
2(dual-1)	4.90e-26	9.32e-26
2(dual-2)	7.07e-25	1.24e-24
3(triad)	4.90e-26	9.32e-26
4(quad)	4.52e-25	7.94e-25
8(octad)	9.48e-26	1.71e-25

表 4.3: フローの平均自乗誤差 (並行移動 (大))

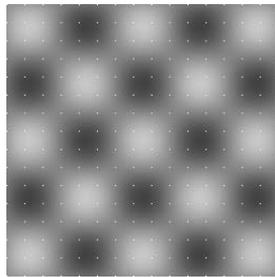
帯域分割数	v_x	v_y
1(single)	1.01e-3	2.35e-4
2(dual-1)	1.24e-26	2.03e-26
2(dual-2)	4.59e-25	9.72e-25
3(triad)	1.24e-26	2.03e-26
4(quad)	1.75e-24	3.13e-24
8(octad)	2.89e-15	4.73e-15

表 4.4: フローの平均自乗誤差 (並行移動 (大+雑音))

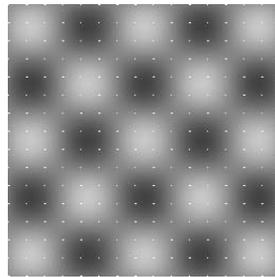
帯域分割数	v_x	v_y
1(single)	1.04e-3	2.57e-4
2(dual-1)	2.17e-5	2.18e-5
2(dual-2)	2.17e-5	2.18e-5
3(triad)	2.18e-5	2.20e-5
4(quad)	8.16e-5	1.15e-4
8(octad)	1.82e-3	1.80e-3



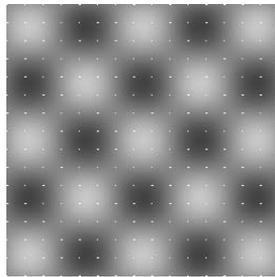
(a) 真のフロー



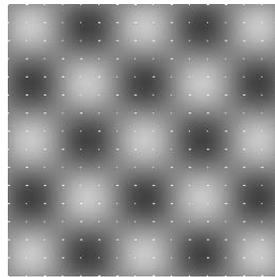
(b) 1(single)



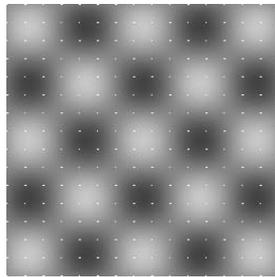
(c) 2(dual-1)



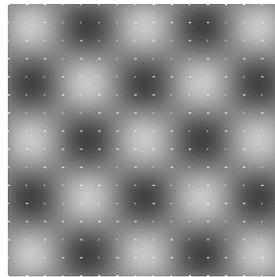
(d) 2(dual-2)



(e) 3(triad)

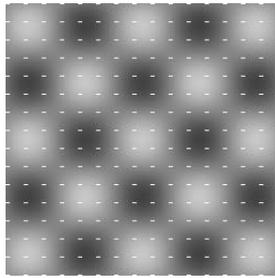


(f) 4(quad)

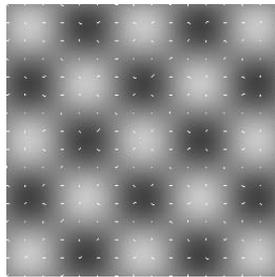


(g) 8(octad)

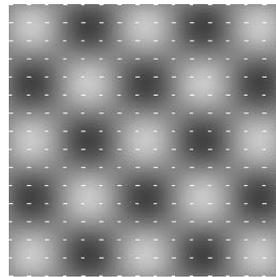
図 4.2: 並行移動 (小)



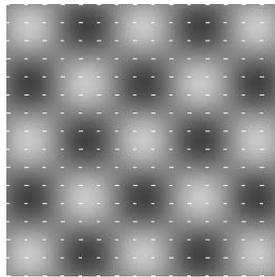
(a) 真のフロー



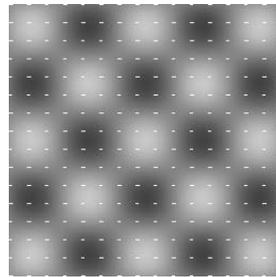
(b) 1(single)



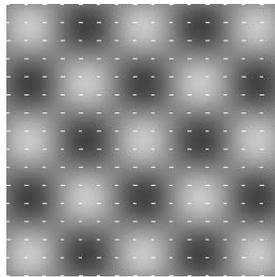
(c) 2(dual-1)



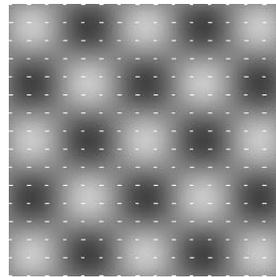
(d) 2(dual-2)



(e) 3(triad)

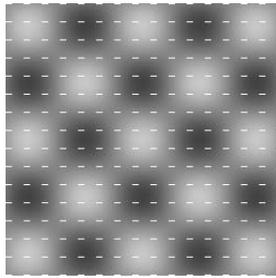


(f) 4(quad)

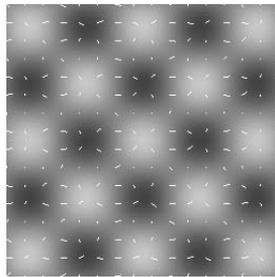


(g) 8(octad)

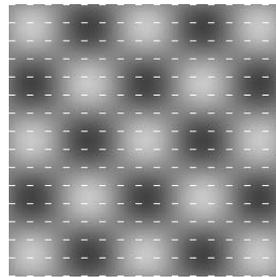
図 4.3: 並行移動 (中)



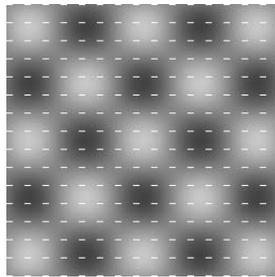
(a) 真のフロー



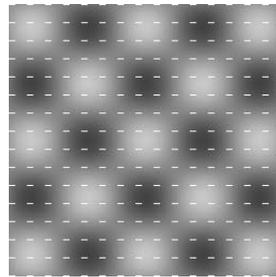
(b) 1(single)



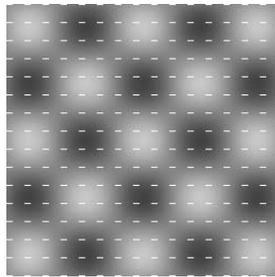
(c) 2(dual-1)



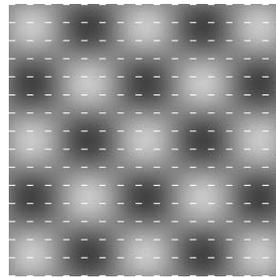
(d) 2(dual-2)



(e) 3(triad)

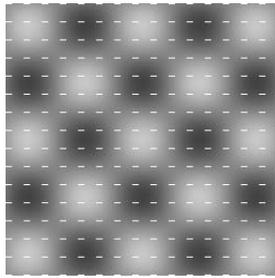


(f) 4(quad)

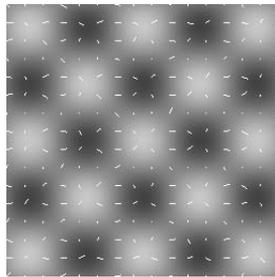


(g) 8(octad)

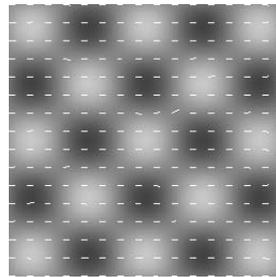
図 4.4: 並行移動 (大)



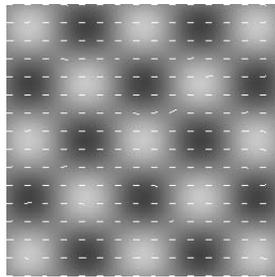
(a) 真のフロー



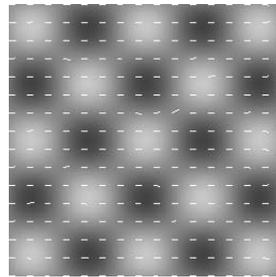
(b) 1(single)



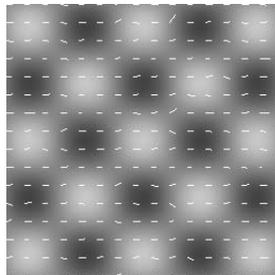
(c) 2(dual-1)



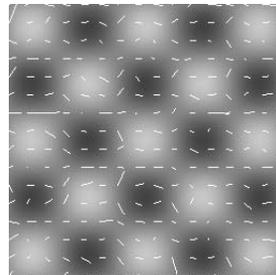
(d) 2(dual-2)



(e) 3(triad)

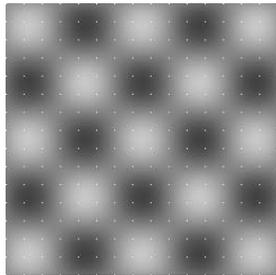


(f) 4(quad)

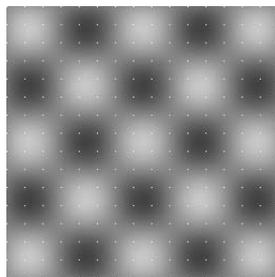


(g) 8(octad)

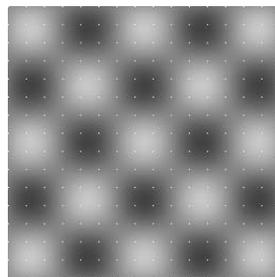
図 4.5: 並行移動 (大+雑音)



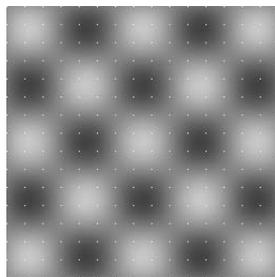
(a) 真のフロー



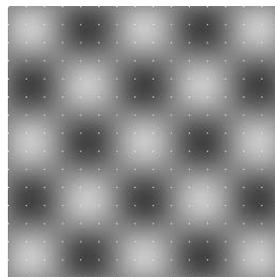
(b) 1(single)



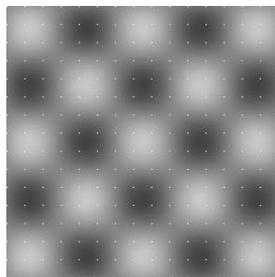
(c) 2(dual-1)



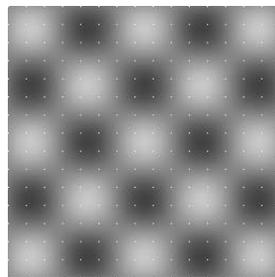
(d) 2(dual-2)



(e) 3(triad)

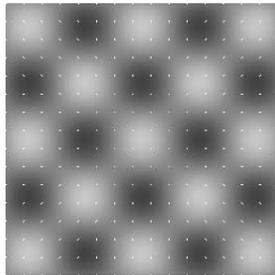


(f) 4(quad)

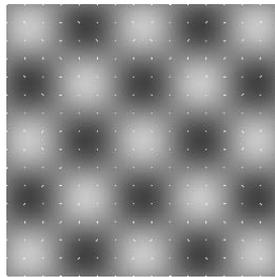


(g) 8(octad)

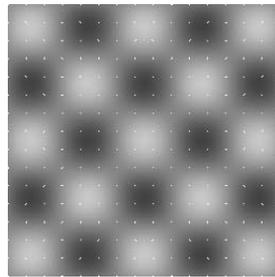
図 4.6: 湧き出し (小)



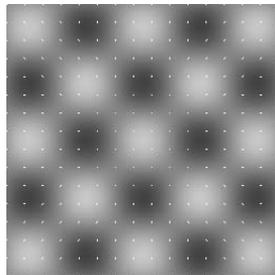
(a) 真のフロー



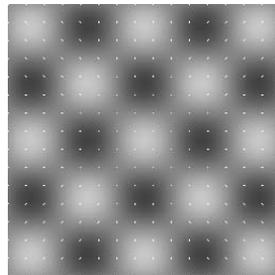
(b) 1(single)



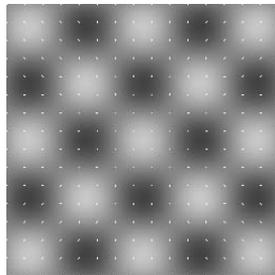
(c) 2(dual-1)



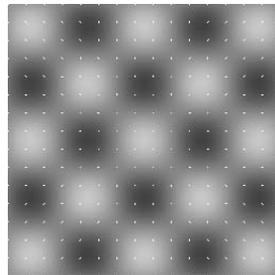
(d) 2(dual-2)



(e) 3(triad)

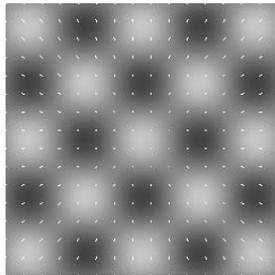


(f) 4(quad)

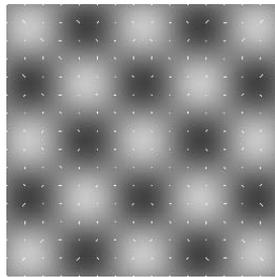


(g) 8(octad)

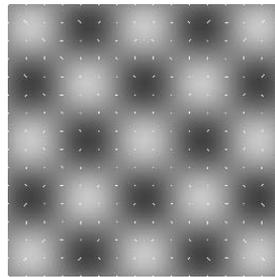
図 4.7: 湧き出し (中)



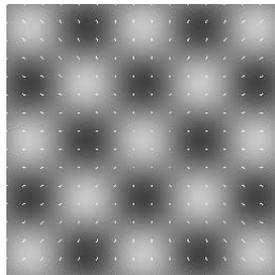
(a) 真のフロー



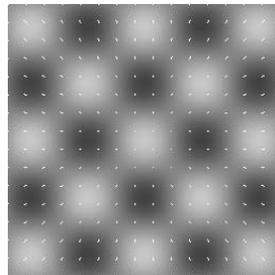
(b) 1(single)



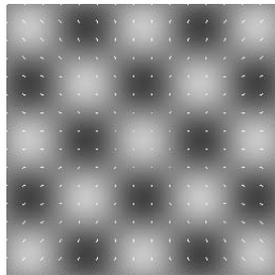
(c) 2(dual-1)



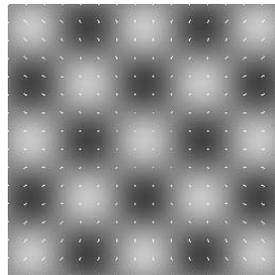
(d) 2(dual-2)



(e) 3(triad)

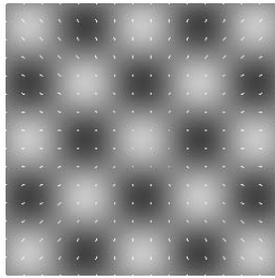


(f) 4(quad)

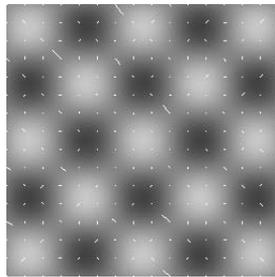


(g) 8(octad)

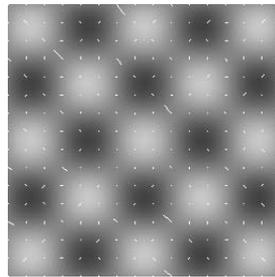
図 4.8: 湧き出し (大)



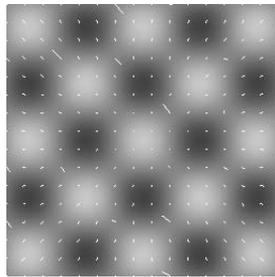
(a) 真のフロー



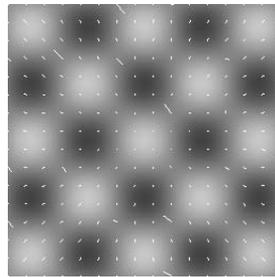
(b) 1(single)



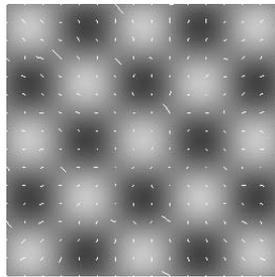
(c) 2(dual-1)



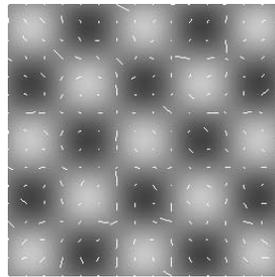
(d) 2(dual-2)



(e) 3(triad)

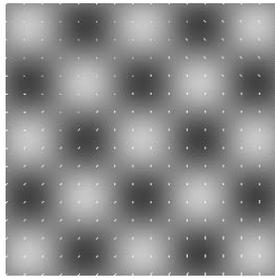


(f) 4(quad)

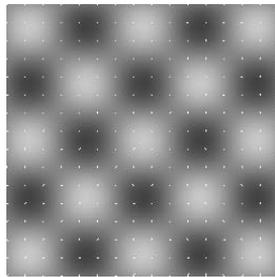


(g) 8(octad)

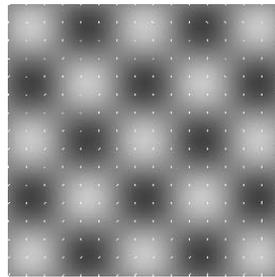
図 4.9: 湧き出し (大+雑音)



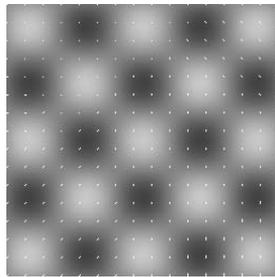
(a) 真のフロー



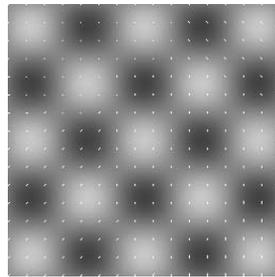
(b) 1(single)



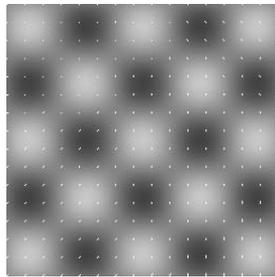
(c) 2(dual-1)



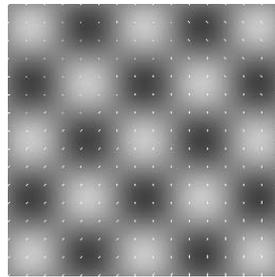
(d) 2(dual-2)



(e) 3(triad)

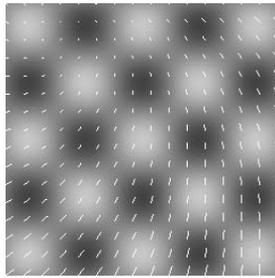


(f) 4(quad)

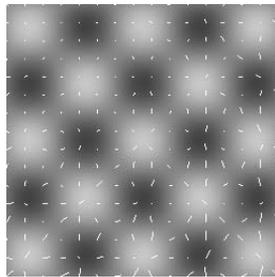


(g) 8(octad)

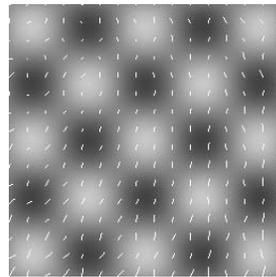
図 4.10: 一般 (小)



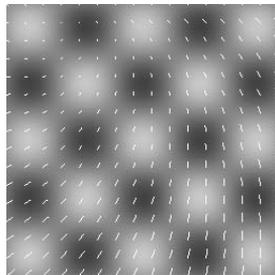
(a) 真のフロー



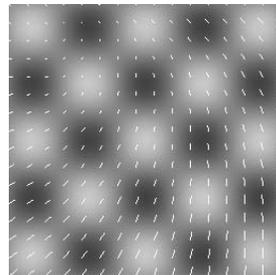
(b) 1(single)



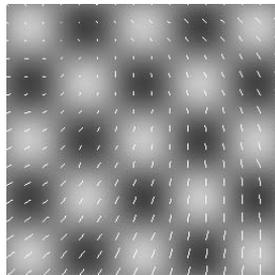
(c) 2(dual-1)



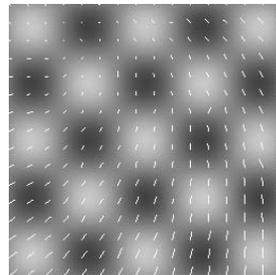
(d) 2(dual-2)



(e) 3(triad)

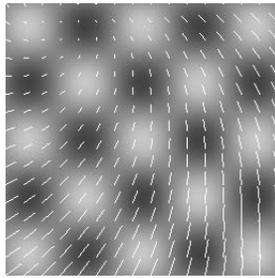


(f) 4(quad)

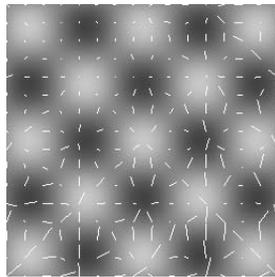


(g) 8(octad)

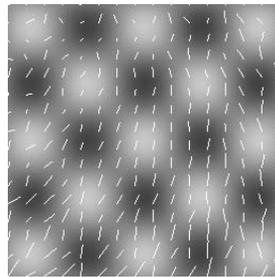
図 4.11: 一般 (中)



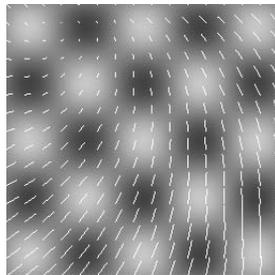
(a) 真のフロー



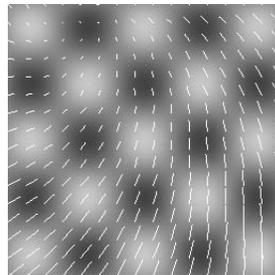
(b) 1(single)



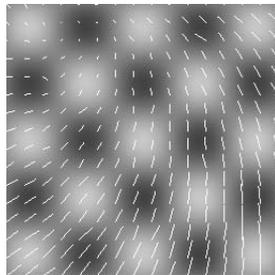
(c) 2(dual-1)



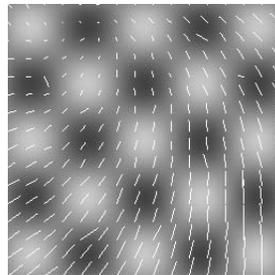
(d) 2(dual-2)



(e) 3(triad)

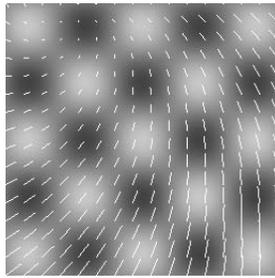


(f) 4(quad)

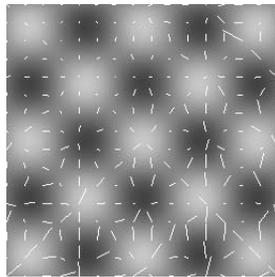


(g) 8(octad)

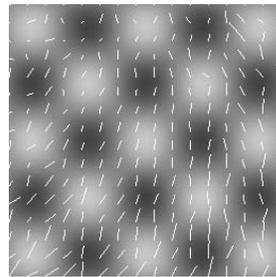
図 4.12: 一般 (大)



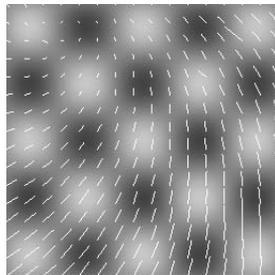
(a) 真のフロー



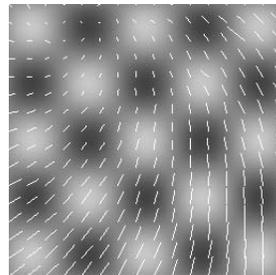
(b) 1(single)



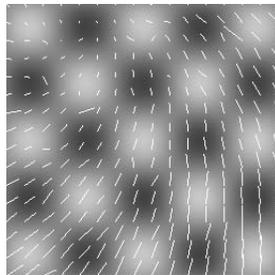
(c) 2(dual-1)



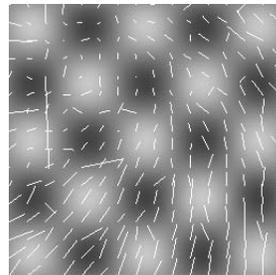
(d) 2(dual-2)



(e) 3(triad)



(f) 4(quad)



(g) 8(octad)

図 4.13: 一般 (大+雑音)

表 4.5: フローの平均自乗誤差 (湧き出し (小))

帯域分割数	v_x	v_y
1(single)	1.08e-6	1.08e-6
2(dual-1)	1.08e-6	1.08e-6
2(dual-2)	2.49e-9	2.49e-9
3(triad)	2.76e-9	2.76e-9
4(quad)	3.44e-8	3.44e-8
8(octad)	5.68e-8	5.68e-8

表 4.6: フローの平均自乗誤差 (湧き出し (中))

帯域分割数	v_x	v_y
1(single)	2.62e-5	2.62e-5
2(dual-1)	2.62e-5	2.62e-5
2(dual-2)	5.51e-8	5.51e-8
3(triad)	5.51e-8	5.51e-8
4(quad)	9.81e-7	9.81e-7
8(octad)	1.35e-6	1.35e-6

表 4.7: フローの平均自乗誤差 (湧き出し (大))

帯域分割数	v_x	v_y
1(single)	1.00e-4	1.00e-4
2(dual-1)	1.00e-4	1.00e-4
2(dual-2)	1.67e-7	1.67e-7
3(triad)	2.75e-7	2.75e-7
4(quad)	3.60e-6	3.60e-6
8(octad)	6.38e-6	6.38e-6

表 4.8: フローの平均自乗誤差 (湧き出し (大+雑音))

帯域分割数	v_x	v_y
1(single)	1.44e-4	1.43e-4
2(dual-1)	1.46e-4	1.44e-4
2(dual-2)	4.55e-5	4.47e-5
3(triad)	4.57e-5	4.49e-5
4(quad)	1.28e-4	1.12e-4
8(octad)	1.72e-3	1.79e-3

表 4.9: フローの平均自乗誤差 (一般 (小))

帯域分割数	v_x	v_y
1(single)	1.82e-5	3.85e-5
2(dual-1)	1.08e-5	1.08e-5
2(dual-2)	5.28e-8	5.31e-8
3(triad)	1.12e-7	1.12e-7
4(quad)	4.87e-7	4.86e-7
8(octad)	7.02e-7	7.30e-7

表 4.10: フローの平均自乗誤差 (一般 (中))

帯域分割数	v_x	v_y
1(single)	4.53e-4	9.71e-4
2(dual-1)	2.73e-4	2.73e-4
2(dual-2)	1.93e-6	1.95e-6
3(triad)	3.30e-6	3.30e-6
4(quad)	1.47e-5	1.48e-5
8(octad)	2.19e-5	2.17e-5

表 4.11: フローの平均自乗誤差 (一般 (大))

帯域分割数	v_x	v_y
1(single)	1.79e-3	3.93e-3
2(dual-1)	1.09e-3	1.09e-3
2(dual-2)	9.74e-6	9.77e-6
3(triad)	1.74e-5	1.65e-5
4(quad)	7.66e-5	7.72e-5
8(octad)	1.13e-4	1.15e-4

表 4.12: フローの平均自乗誤差 (一般 (大+雑音))

帯域分割数	v_x	v_y
1(single)	1.83e-3	4.00e-3
2(dual-1)	1.14e-3	1.19e-3
2(dual-2)	6.86e-5	1.14e-4
3(triad)	7.63e-5	1.21e-4
4(quad)	2.39e-4	3.07e-4
8(octad)	2.69e-3	2.88e-3

第5章 おわりに

本研究では，特定の特徴点の奥行き値ではなく，*iconic depth* としての密な形状情報を求める立場から，動画像系列の時空間情報伝搬スキームに基づく SFM の高精度化を検討した．成果として得られた確率構造や推論方式には，他の復元キー，例えば陰影情報やステレオ視における視差情報等を組み入れることが可能であり，初期視覚システムの全体像をこの方法論によって定式化することができる．今後はこの方法に，本研究の成果を活用していく予定である．

最後に，本研究を進めるにあたり，ご指導とご配慮を頂いた東京都立大学大学院工学研究科の守屋正教授，ならびにご協力頂いた本学電子通信工学講座の学生諸氏，特にプログラム作成に貢献して頂いた長沼将一氏に感謝致します．

参考文献

- [1] D.Marr, *Vision: a computational investigation into the human representation and processing of visual information*, W.H. Freeman and Co., 1982. (乾敏朗, 安藤広志 (訳), 産業図書, 1987.)
- [2] B.K.P. Horn and B.G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.
- [3] G. Adiv, “Determining three-dimensional motion and structure from optical flow generated by several moving objects,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, no. 4, pp. 384–401, 1985.
- [4] M. Subbarao and A.M.Waxman, “Closed form solution to image flow equation for planar surface in motion,” *Comput. Vision, Graphics, Image Process.*, vol. 36, pp. 208–223, 1986.
- [5] K. Kanatani, “Structure and motion from optical flow under perspective projection,” *Comput. Vision, Graphics, Image Process.*, vol. 38, pp. 122–146, 1987.
- [6] N. Tagawa, T. Toriu, T. Endoh, “3-D motion estimation from optical flow with low computational cost and small variance,” *IEICE Trans. Inf. & Syst.*, vol. E79-D, pp. 230–241, 1996.
- [7] 金出武雄, コンラッド ポールマン, 森田俊彦, “因子分解法による物体の形状とカメラ運動の復元,” *電子情報通信学会論文誌 D-II*, vol. J76-D-II, no. 8, pp. 1497–1505, 1993.
- [8] J.K. Kearney and W.B. Thompson, “Optical flow estimation: An error analysis of gradient-based methods with local optimization,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 2, pp. 229–243, 1987.
- [9] 武川 直樹, “動画像解析における光学モデル,” *電子情報通信学会論文誌 D-II*, vol. J77-D-II, no. 4, pp. 711–718.

- [10] 牧 淳人, 渡邊 睦, チャールズ ワイルズ, “Geotensity 拘束による 3 次元形状獲得,” 電子情報通信学会論文誌 D-II, vol. J83-D-II, no. 8, pp. 1741–1752.
- [11] E.P. Simoncelli, “Bayesian Multi-scale differential optical flow,” *Handbook of Computer Vision and Applications*, eds. B. Jahne, H. Haussecker, P. Geissler, Academic Press, 1999.
- [12] A.P. Dempster, N.M. Laird, and D.B. Rubin, “Maximum likelihood from incomplete data,” *J. Roy. Statist. Soc. B*, vol. 39, pp. 1-38, 1977.
- [13] L. Matthies and T. Kanade, “Kalman filter-based algorithms for estimating depth from image sequences,” *Int. Journal Comput. Vision*, no. 3, pp. 209–236, 1989.
- [14] Y.S. Hung and H.T. Ho, “A Kalman filter approach to direct depth estimation incorporating surface structure,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 6, pp. 570–575, 1999.