

複合現実感におけるカメラ位置姿勢推定の安定化とその考察

遠藤 隆明 佐藤 清秀 内山 晋二 山本 裕之

キヤノン株式会社 知覚システム開発センター

あらまし MR システムの実現に必要な観察カメラの位置姿勢を推定する手法が多く検討されている。多くの研究の主眼は、新たな画像特徴の利用方法や精度の向上を目指すものであり、その安定性に対する取り組みは単発的かつ副次的に取り扱われることが多い。それに対して、MR システムの実利用化において、安定性は極めて重要視される要件である。そこで本稿では、カメラ位置姿勢推定を不安定にする複合的な要因を分析・考察し、安定化させるための方策を様々な観点から総合的に議論する。さらに、その考察を踏まえて、我々が実現したカメラ位置姿勢推定安定化手法について述べ、実験を通してその安定性を示す。

Implementation and Discussion for Stabilizing Camera Pose Estimation

Takaaki Endo, Kiyohide Satoh, Shinji Uchiyama, and Hiroyuki Yamamoto

Human Machine Perception Laboratory, Canon Inc.

Abstract – Stable estimation of camera pose is one of the most important keys to put a Mixed Reality system into practical use. In this paper, we carefully analyze multiple factors that are destabilizing camera pose estimation, then comprehensively discuss and develop strategy how to stabilize it. According to the strategy, we propose two methods for estimating the stable camera pose; one is reference constrain-based method and the other is error distribution-based method. In addition to theoretical discussions, we demonstrate the effectiveness of our methods by experiments in comparison with conventional methods.

1. はじめに

MR (Mixed Reality) システムの実現に必要な現実空間と仮想空間の位置合わせ手法、すなわち現実空間を移動しながら観察するカメラの位置姿勢をリアルタイムで推定する手法が多く検討されている。実践的な MR システムを構築する上で、位置合わせの安定性は重要な課題である。しかし、位置合わせの安定化に向けた従来の取り組みは、単発的かつ副次的なものであり、不安定性の要因分析や対策に総合的に取り組んだものはなかった。例えば、画像情報(画像から検出したシーン中の指標の座標)に基づくカメラの位置姿勢の推定(以下ではビジョン推定と呼ぶ)において外れ値が多く含まれていても破綻しないようにするため、M-推定などのロバスト推定が一般に用いられる。しかし、ロバスト推定を適用するだけでは、位置姿勢推定結果の時間的な揺れ(ジッタ)や不連続変化(飛び)の発生を回避できない。

ここで、位置合わせが不安定であるとは、本来は現実空間中に固定的に表示されるべき CG が、小刻みに振動したり、ある瞬間に不連続に移動したりして表示される状態のことを表す。この不安定さが、現実空間中にあたかも仮想物体が存在するかのように感じさせる MR 体験を大きく障害する。カメラの位置姿勢推定が安定すれば、CG も表示位置に関係なく安定的に表示されることになる。そこで本論文では、カメラ位置姿勢推定

の安定化に焦点を当てて議論する。まず、位置合わせを不安定にする複合的な要因の分析に基づいて、これまで単発的に行われてきた位置合わせの安定化に対する取り組みを系統的に分類する。そしてこれらを考慮した上で、総合的に位置合わせを安定化させる手法を提案する。

位置合わせの安定性のみを求めるのであれば、非常に強いローパスフィルタをかければよい。しかし、この場合には自由なカメラ移動に対する位置合わせの追従性が悪くなり、幾何学的・時間的整合性が取れなくなる。つまり、実践的な MR システムでは、位置合わせの安定性と精度(幾何学的・時間的整合性)を適切にバランスさせることが求められる。

本論文では、図1に示すビデオシースルー方式の画像情報と、位置や姿勢を計測する物理センサ(観察者カメラ以外のセンサ)を併用したシステム構成を想定して、位置合わせの安定性について議論する。例えば、

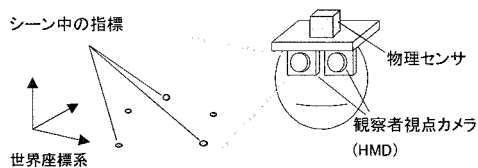


図1 想定システム構成

磁気センサのような6自由度の位置姿勢センサによって計測した観察者視点カメラの位置姿勢を画像情報に基づいて補正するシステムや、ジャイロセンサや加速度センサによって構成される3自由度の姿勢センサの計測値と画像情報とに基づいてカメラの位置姿勢を推定するシステムなどを考察の対象とする。

以下、2章で位置合わせの不安定さの要因を分析し、3章で位置合わせを安定化させるための従来手法について整理する。4章では3章までの分析・考察に基づいて位置合わせを安定化させる手法を提案し、5章で実験を通してその有効性を示す。そして6章でまとめと今後の課題を述べる。

2. 位置合わせの不安定さの要因分析

本論文では、次の1～3のような現象が発生する場合に、位置合わせが不安定であると考える。

【現象1】位置姿勢推定結果の時間的な揺れ(ジッタ)

【現象2】位置姿勢推定結果の不連続変化(飛び)

【現象3】位置合わせの破綻

ジッタとは、カメラの位置姿勢推定値が時間軸に沿って小刻みに振動する現象のことを表す。ジッタはビジョン推定値や物理センサ計測値のランダム変動によって発生する。2次元画像中で検出された指標の座標値には統計的誤差(多数回の計測を行うと平均値が0のガウス分布に従う誤差)が少なからず含まれている。その誤差が画像座標系からカメラ座標系へと伝播して、ビジョン推定値のランダム変動として観測される。ビジョン推定値の誤差の大きさ(ランダム変動の幅)は、指標検出の統計的誤差の画像座標系からカメラ座標系への伝播を表す次式によって近似される [Hoff00]。

$$\mathbf{C}_{cam} = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \begin{bmatrix} \mathbf{C}_{u1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{C}_{un} \end{bmatrix} ((\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T)^T \quad (1)$$

ここで、 \mathbf{C}_{cam} はビジョン推定値の誤差共分散行列、 \mathbf{C}_{u_i} は指標検出の統計的誤差の分散を対角成分に持つ、指標検出の誤差共分散行列を表す。また、 \mathbf{J} は画像中の特徴量の微小変化と位置姿勢の微小変化との関係を表す行列であり、画像ヤコビアンと呼ばれる。

また、飛びとは、カメラの位置姿勢推定値がある瞬間に不連続に切り替わる現象のことを表す。飛びは主にビジョン推定値の不連続変化によって発生する。2次元画像中で検出された指標の座標値に含まれる系統誤差(真値からの偏りを表す、多数回の計測を行っても同じように現れる誤差)は、画像座標系からカメラ座標系へと伝播する。ここで、検出される指標がフレーム間で切り替わると、カメラ座標系へ伝播する誤差の大きさが不連続に変化するため、ビジョン推定値も不連続に変

化することになる。指標検出の系統誤差は、指標の配置情報の校正誤差や、カメラ内部パラメータの校正誤差によって引き起こされる。また、画像からの指標検出処理自体に系統的な誤差が含まれる場合もある。

そして、破綻とは、カメラの位置姿勢推定値が発散したり局所解に収束したりする現象のことを表す。破綻は主にカメラ位置姿勢の推定計算に用いる行列が指標の偏在によりランク落ちすることによって発生する。

3. 位置合わせを安定化させるための方策

本章では、位置合わせの不安定さの要因に対処して位置合わせを安定化させるための方策について、従来技術を位置付けながら議論する。

多数の指標を検出して利用できれば [Skrypnik04] [Klein07]、指標の偏在に起因するジッタや破綻を抑えることができる。しかし、これらの手法は比較的乱雑なシーンに限り有効である。

時間軸方向の連続性を適切に取り扱えば、ジッタを軽減させることができる。一般的には、拡張カルマンフィルタなどの、時間遅れのない時系列フィルタリングが利用される [Koller97][Park99]。同様な効果が得られる手法として、Vacchettiらは、前フレームと現フレームのカメラ位置姿勢を同時に最適化する手法を提案している [Vacchetti04]。また、Skrypnikらは、「前フレームの位置姿勢推定値から大きくかけ離れない」という拘束条件を非線形最適化手法に組み込みながら誤差最小化を行う手法を提案している [Skrypnik04]。

効果的にジッタを抑えるためには、物理センサの計測誤差を統計的誤差と系統誤差に分けて、それぞれの誤差を適切に見積もることが望ましい。張替らは、時定数が大きい慣性航法装置のドリフト誤差をカルマンフィルタで推定することにより、適正な補正量を算出している [張替00]。

物理センサ計測値に基づいて算出した指標の画像座標と、画像処理によって検出した指標の画像座標とを比較し、そのずれ(以下、再投影誤差と呼ぶ)を打ち消すように物理センサ計測値の一部のパラメータに修正を加える手法(以下、部分パラメータ固定法と呼ぶ)も提案されている [Bajura95][佐藤05]。部分パラメータ固定法によれば、画像と物理センサの情報を相補的に用いることで位置合わせの精度を向上させることができるだけでなく、ビジョン推定値のランダム変動とセンサ計測値の統計的誤差に総合的に対処してジッタも軽減させることができる。ただし、センサ計測値の一部のパラメータを完全に信じているため、画像情報を十分に活かさず、十分に高い精度とならない場合もある。これに対し、横地らは、画像上での再投影誤差とGPSの計測誤差を同時に最小化してカメラ位置姿勢を推定

する手法を提案している [横地 06]. これは、画像と物理センサの両者の情報を、部分パラメータ固定法よりも有効に活用した手法であると言える。一方、Hoff らは、誤差分布を適切に考慮して複数センサの計測値を統合することにより、最小分散推定を実現している [Hoff00]. ただし、Hoff らの手法は系統誤差を考慮していない。

飛びの原因となる指標の配置情報の校正誤差やカメラ内部パラメータの校正誤差は、実際のシステムでは多少なりとも発生することは避けられない。この問題に対し Thomas らは、新たに検出された指標の座標を、これまで検出されていた指標によって求めたカメラ位置姿勢と矛盾しないように一時的に補正し、そのあと徐々に正しい位置に戻すことで、不連続変化を軽減する手法を提案している [Thomas97]. ただし、この手法は検出指標数が減少する場合に対処できない。

破綻を防止するには、指標の偏在によるランク落ちに対処した計算を行えばよい。たとえば、画像ヤコビアンを用いる手法の場合には、画像ヤコビアンとその転置行列との積の正則化を適切に行うことにより、安定した解を得ることができる [佐野 90]. 具体的には、再投影誤差最小化の繰り返し計算の過程で、通常的一般化逆行列 $\mathbf{J}^{\dagger} = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T$ の代わりに正則化一般化逆行列 $\mathbf{J}^{\dagger} = (\mathbf{J}^T \mathbf{J} + \mu^2 \mathbf{I})^{-1} \mathbf{J}^T$ を利用すればよい。ここで μ は微小な正則化パラメータである。あるいは、画像ヤコビアンの微小特異値を 0 としてランクを低減させた上で画像ヤコビアンを再計算しても、同等の効果が得られる。

4. 提案手法

前章では、位置合わせを不安定にする様々な要因に対する対処方法について議論した。本章では、前章までの分析および考察を踏まえて、全体として矛盾しない方法で位置合わせを総合的に安定化させる手法を提案する。

まず、ジッタに対処するために、カメラや物理センサを高精度にするには限界がある。また、大量の指標の利用も場面によっては困難である。したがって、ビジョン推定と物理センサ計測の誤差特性の違いを相補的に利用する考え方の有用性が極めて高い。その際に、ビジョン推定と物理センサ計測の誤差を同時に最小化する横地らの手法 [横地 06] や、誤差分布を考慮してビジョン推定値と物理センサ計測値を統合する Hoff らの手法 [Hoff00] が有効である。また、時系列方向の連続性も考慮すべきである。それにはカルマンフィルタの適用や、前フレームの位置姿勢推定値への拘束を行う Skrypnik らの手法 [Skrypnik04] が有効である。さらに、物理センサの系統誤差も分離して扱うべきである。カルマンフィルタを使う場合には、張替らの手法 [張替

00] が有効である。

また、飛びに対処するためには、Thomas らの手法が有効である [Thomas97]. ただし、検出指標数の増減に関わらず適用できるように拡張する必要がある。また、検出指標数などの条件に関わらず単一の手法でカメラ位置姿勢を推定できることが要求される。

そして、破綻に対処するためには、大量の指標の利用が有効であるが、場面によっては困難である。少数の指標しか検出されない場面における破綻を回避するには、画像ヤコビアンとその転置行列との積の正則化 [佐野 90] が有効である。

本章では、以上の分析を踏まえて、画像と物理センサの情報を相補的に活かしつつ、時系列方向の連続性を考慮して、かつ物理センサの系統誤差を分離して扱うことができる、二種類の手法を提案する。まず一つ目は、物理センサ計測値、前フレームの位置姿勢推定値、系統誤差推定値などへの拘束をかけつつ、再投影誤差を極力小さくするような最適化を行う手法である (4.1 節: 参照値拘束法)。そして二つ目は、Hoff らの手法を用いて画像と物理センサの情報を統合した上で、カルマンフィルタによって時系列方向の連続性を考慮する手法である (4.2 節: 空間分散・時間分散最小化法)。さらに、飛びへの対処として、ビジョン推定値の系統誤差補正量を時系列的に平滑化する手法も提案する (4.3 節)。以下では、これらの手法について詳細に述べる。

4.1. 参照値拘束法

従来の部分パラメータ固定法では、物理センサ計測値の一部のパラメータを完全に信じていることが問題であった。これに対し、「カメラ位置姿勢推定値は物理センサ計測値から大きくかけ離れない」という拘束条件を加味しながら、再投影誤差を極力小さくするような最適化を行う方法が考えられる。また、Skrypnik らの手法と同様に、前フレームの位置姿勢推定値を、かけ離れないようにする参照値として用いることもできる。そこで本章では、これらの複数種類の参照値から大きくかけ離れないという拘束条件を加味しながら再投影誤差を極力小さくするように最適化する方法を提案する。本手法を以下では、参照値拘束法と呼ぶ。

4.1.1. 再投影誤差最小化

まず、参照値拘束法が基本原理として用いる、非線形最適化処理による再投影誤差最小化の一般的な枠組みの概略を説明する。

再投影誤差最小化は、画像上での誤差が最小となる場合のカメラ位置姿勢推定値が最良であると考え、そうなるようにカメラの位置姿勢パラメータを修正する操作を反復する処理である。誤差の評価式は、

$$e = \sum_i |u_i - h(s, X_i)| \quad (2)$$

であり、 e を最小にする s を求める。ここで、 u_i は画像処理によって検出された点 i の画像座標、 s は現在のカメラ位置姿勢推定値、 X_i は点 i の既知の3次元座標である、そして、 $h(\cdot)$ は観測方程式であり、現在のカメラ位置姿勢推定値から求められる点 i の計算上の画像座標を表す。

ここで、 n 点の指標の再投影誤差を並べた $2n$ 次元のベクトルを Ψ と置くと、 s の補正量は、

$$\Delta s = J^+ \Psi \quad (3)$$

として n 点の最小二乗として求めることができる。なお、 J^+ は J の一般化逆行列である。この結果から s を求め、再投影誤差を再計算し、さらに Δs を求める、という計算を収束するまで繰り返して、最終的な s を算出する。

なお、参照値拘束法では、 J^+ として 3 章で説明した正則化一般化逆行列を用いることにより、ランク落ちによる破綻に対処する。

4.1.2. 各種参照値への重み付き推定

参照値拘束法では、各種参照値から大きくかけ離れないという拘束条件を加えて、再投影誤差最小化の計算を行う。参照値としては、

- (1) 物理センサ計測値
- (2) 前フレームの位置姿勢推定値
- (3) 前フレームの物理センサの系統誤差推定値

の三つを同時に用いる。そうすることにより、画像と物理センサの情報を相補的に活かしつつ、時系列方向の連続性を考慮して、かつ物理センサの系統誤差を分離して扱うことを可能としている。

このとき、最適化の評価式は、

$$e = \sum_i |u_i - h(s, X_i)| + w_1 |s_{\text{sensor}} - s| + w_2 |s_{t-1} - s_t| + w_3 |\delta_{t-1} - \delta_t| \quad (4)$$

となる。ここで、 w_1 、 w_2 、 w_3 は重みパラメータ、 s_{sensor} は物理センサ計測値、 s_t は時刻 t におけるカメラ位置姿勢推定値、 δ_t は時刻 t における物理センサの系統誤差の推定値である。

式 4 を用いた場合、検出指標が多数ならば式 2 に近い推定結果となる。一方、検出指標が少数の場合には、位置姿勢推定値が物理センサ計測値や前フレームの推定結果から逸脱しなくなり、または物理センサの系統誤差が前フレームの推定結果から逸脱しなくなり、安定した解を得ることができる。

4.1.3. 指標の分布の考慮

画像中の検出指標の分布状況が悪い場合には、画像情報への依存度を弱めることが望ましい。そこで、ま

ず検出指標の重心から見た座標値を並べた行列とその転置行列との積を求め、その最小固有値の逆数を分布考慮項とする。そして、式 4 における重みパラメータ w_1 、 w_2 、 w_3 に分布考慮項を掛け合わせた値を新たな重みパラメータとすることにより、検出指標の偏在の影響を軽減させることができる。すなわち、検出指標が偏在している場合には、新たな重みパラメータが大きくなり、参照値から逸脱しない安定した解を得ることができる。

4.2. 空間分散・時間分散最小化法

4.1. で提案した参照値拘束法で設定する重みパラメータは、次元、単位、スケールなどが異なる評価空間における誤差の大きさを比較するものであり、パラメータに物理的意味がないという問題がある。この問題に対処するために、式 1 にしたがって指標検出の統計的誤差をカメラ座標系へと伝播させ、それと物理センサ計測値の統計的誤差と同じ3次元空間で評価する方法が考えられる。この方法は、同じ3次元評価空間で観測値と予測値との重み付き平均を取るカルマンフィルタとの親和性も高い。

そこで本章では、ビジョン推定値と物理センサ計測値とに基づいて3次元空間における最小分散推定を行った上で、カルマンフィルタによって時間軸方向の最小分散推定を行う方法を提案する。本手法を以下では、空間分散・時間分散最小化法と呼ぶ。

4.2.1. 処理手順

空間分散・時間分散最小化法の処理手順は以下に示す通りである。

- (step1) 初期化
 - (step2) カルマンフィルタの状態ベクトル予測
 - (step3) 再投影誤差最小化(参照値拘束なし)
 - (step4) センサフュージョン(物理センサ計測値から系統誤差の予測値を差し引いた値と step3 の結果との重み付き平均)
 - (step5) カルマンフィルタの状態ベクトル更新(step2 の結果と step4 の結果との重み付き平均)
- step3 の再投影誤差最小化の計算過程において、3 章で説明した正則化一般化逆行列を用いることによって、ランク落ちや方程式不足に対処する。

以下では、センサフュージョンとカルマンフィルタの処理の内容についてさらに詳しく述べる。

4.2.2. センサフュージョンによる空間分散最小化

ビジョン推定値と物理センサ計測値のセンサフュージョンには、Hoff らが提案した枠組みを用いる [Hoff00]。空間分散・時間分散最小化法では、再投影誤差最小化の計算結果と物理センサ計測値との重み付き平均を取ることによって、統計的に最適なカメラ位

位置姿勢を推定する。ここで重みは誤差の分散の逆数となる。

再投影誤差最小化の計算結果を s_{cam} 、物理センサ計測値から系統誤差予測値を差し引いた値を s_{sensor} とすると、カメラ位置姿勢推定値は、

$$\begin{aligned} s_{fuse} = & C_{sensor} (C_{cam} + C_{sensor})^{-1} s_{cam} \\ & + C_{cam} (C_{cam} + C_{sensor})^{-1} s_{sensor} \end{aligned} \quad (5)$$

となる。ここで、 C_{sensor} は物理センサ計測値から系統誤差予測値を差し引いた値の誤差共分散行列であり、統計的誤差の分散と系統誤差の予測誤差の分散を足し合わせた値を対角成分に持つ。また、 C_{cam} は再投影誤差最小化の計算結果の誤差共分散行列であり、式1によって表される。

センサフュージョン結果の誤差共分散行列は、

$$C_{fuse} = C_{cam} (C_{cam} + C_{sensor})^{-1} C_{sensor} \quad (6)$$

となり、図2のように表される。

4.2.3. カルマンフィルタによる時間分散最小化

カメラの運動モデルとセンサフュージョン結果とに基づいて統計的に最適なカメラ位置姿勢を推定するために、カルマンフィルタを用いる。空間分散・時間分散最小化法では、張替らの手法 [張替 00] と同様に統計的誤差と系統誤差を分離して扱うことにより、統計的誤差が過大に見積もられることを防ぐ。以下では、物理センサとして位置姿勢センサを用いる場合を例にして説明する。

カルマンフィルタの状態空間モデルは、次の状態の予測を行う状態遷移方程式と、状態と観測の関係を表す観測方程式からなる。

状態遷移方程式は、

$$x_k = Fx_{k-1} + w \quad (7)$$

である。ここで、 x_k は状態ベクトル、 F は状態遷移行列、 w はシステムノイズを表す。具体的には、状態ベクトルは、

$$x = [t \ w \ \tau \ \omega \ \delta t \ \delta w]^T \quad (8)$$

とする。ここで、 t は位置、 w は姿勢、 τ は速度、 ω は角速度、 δt は位置の系統誤差、 δw は姿勢の系統誤差を表す。このとき、運動モデルを表す状態遷移行列は、

$$F = \begin{bmatrix} I & 0 & \Delta T \cdot I & 0 & 0 & 0 \\ 0 & I & 0 & \Delta T \cdot I & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & I \end{bmatrix} \quad (9)$$

となる。ただし、 ΔT はフレーム間の経過時間を表す。これは、等速度・等加速度運動モデルに加えて、傾斜角

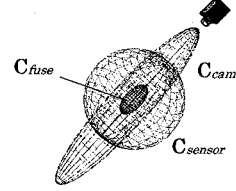


図2 センサフュージョン結果の誤差共分散行列

の等系統誤差モデルを表している。そして、システムノイズの共分散行列は対角行列であり、

$$C_w = \text{diag}[C_t, C_w, C_\tau, C_\omega, C_\alpha, C_{\delta w}] \quad (10)$$

となる。ここで、 C_t 、 C_w 、 C_τ 、 C_ω 、 C_α 、 $C_{\delta w}$ は、それぞれ、位置、姿勢、速度、角速度、位置の系統誤差、姿勢の系統誤差の予測に関する誤差共分散行列を表す。

一方、観測方程式は、

$$y_k = Hx_k + v \quad (11)$$

である。ここで、 y_k は観測ベクトル、 H は観測行列、 v は観測ノイズを表す。具体的には、観測ベクトルは、

$$y = [t_{fuse} \ w_{fuse} \ \delta t_{observe} \ \delta w_{observe}]^T \quad (12)$$

とする。ここで、 t_{fuse} 、 w_{fuse} はそれぞれセンサフュージョン結果の位置および姿勢を表す。そして、 $\delta t_{observe}$ は位置の系統誤差を表し、物理センサ計測値からセンサフュージョン結果の位置成分を差し引いた値となる。同様に、 $\delta w_{observe}$ は姿勢の系統誤差を表し、物理センサ計測値からセンサフュージョン結果の姿勢成分を差し引いた値となる。このことから、観測ノイズの共分散行列は

$$C_v = \begin{bmatrix} C_{fuse} & 0 \\ 0 & C_{sensor_noise} + C_{fuse} \end{bmatrix} \quad (13)$$

となる。ここで、 C_{fuse} はセンサフュージョン結果の誤差共分散行列(式6)、 C_{sensor_noise} は物理センサの統計的誤差の共分散行列を表す。

なお、物理センサとして地磁気センサのような方位ドリフトが発生しない姿勢センサを用いる場合には、状態遷移方程式および観測方程式から位置の系統誤差に関連する項目を取り除けばよい。また、方位ドリフトが発生する姿勢センサを用いる場合には、状態ベクトルなどに方位ドリフトに関する項目を追加し、状態遷移行列で等ドリフトモデルを表現すればよい。

4.3. 系統誤差補正量の時系列平滑化

以上で提案した二つの手法はいずれも、検出指標の切り替わりによって生じるビジョン推定値の不連続変化を十分には解消できない。指標検出に系統誤差が生じる原因を取り除くことは困難であるので、指標の切り替わりに特化した平滑化によって、系統誤差の影響を軽減することを考える。

本手法では、指標の切り替わり(増加・減少)に応じて以下の処理を実行する。あるフレームで指標が新たに追加された場合には、現フレームにおいて、前フレームとの共通指標のみを使ったカメラ位置姿勢推定と、全ての検出指標を使ったカメラ位置姿勢推定のそれぞれを行う。そして、これらの比較によって、指標が増加したことによる推定値の変動を得る。一方、あるフレームで指標を見失った場合には、現フレームとの共通指標のみを使ったカメラ位置姿勢推定を前フレームに対して行う。そして、これを前フレームで採用した位置姿勢推定値と比較することで、指標が減少したことによる推定値の変動を求める。それらの変動を以降の推定値に徐々に反映させることで、系統誤差に起因する飛びを緩和させる。これは、単純にローパスフィルタを施す場合と異なり、情報の変動分のみを推定値に反映させることから、カメラの位置姿勢推定に遅れが生じることはない。本手法によれば、ビジョン推定値の系統誤差成分の補正量が時間軸方向に平滑化される。指標が増加した場合には Thomas らの手法 [Thomas97] と同等の効果が得られ、指標の減少にも対処可能である。

4.4. 考察

4.1.で提案した参照値拘束法と 4.2.で提案した空間分散・時間分散最小化法は共に、系統誤差を考慮しつつ、画像と物理センサの情報を相補的に利用し、かつ時系列方向の連続性も考慮する、という点で共通している。二つの手法の違いは以下の点である。

参照値拘束法で設定する重みパラメータは、次元、単位、スケールなどが異なる評価空間における誤差の大きさを比較するものであり、パラメータに物理的意味がない。ただし、パラメータが適切に設定できれば、4.1.2 で述べたような実践的な効果を得ることができる。一方、空間分散・時間分散最小化法では、次元、単位、スケールなどが同じ評価空間で誤差を評価することができるため、物理センサの誤差特性などに応じて物理的に意味のあるパラメータを適切に設定することができる。ただし、本来は非線形な誤差の伝播を式 1 で線形近似するため、非線形性が強く現れる場面では精度が悪くなる。たとえば、平面上に配置された 3~4 個程度の指標を正面付近から観察する場合に、フレーム毎に複数の局所解のいずれかへの収束を繰り返す挙動を示すことがあるが、そのような現象を線形近似することはできず、画像上での位置ずれが発生する。これに対し、参照値拘束法は再投影誤差最小化に基づいているため、大きな位置ずれは発生しない。したがって、二つの提案手法のどちらが位置合わせの安定性および精度に優れているかは一概には言えない。

5. 実験

提案手法の安定性を従来手法と比較する実験を行う。比較対象となる従来手法としては、4.1.1.で説明した再投影誤差最小化法と、3 章で説明した部分パラメータ固定法を用いる。

5.1. 実験条件

シーンを撮影するカメラとしては、33 万画素の CCD を有する竹中システム機器社製の FC300M を用いる。また、レンズとしては、水平画角 61 度を有するフジノン社製の TF4DA-8 を用いる。焦点距離、主点位置、レンズ歪みパラメータなどのカメラの内部パラメータは、予め校正された値を用いる。物理センサとしては、6自由度磁気センサ (Polhemus 社 FASTRAK) をカメラに取り付けて利用する。そして、図 3 に示すように、世界座標系の原点および $Z=300\text{mm}$ の位置に一辺 60mm の正方形マーカを配置し、その 4 隅の座標を用いてカメラの位置姿勢推定を行う。画像中に描画された座標軸は、推定したカメラ位置姿勢で描画した世界座標系の座標軸である。位置合わせが正確であれば、画面左側マーカの中心位置に座標軸が描画される。

参照値拘束法では、位置合わせの安定性と精度が適切なバランスとなるように、発見的に重みパラメータを設定する。本実験で用いた重みパラメータの設定値

表 1 参照値拘束法における設定値

パラメータ種	設定値
センサ計測値(位置)への拘束の重み	0.001
センサ計測値(姿勢)への拘束の重み	0.005
前フレーム推定値(位置)への拘束の重み	0.03
前フレーム推定値(姿勢)への拘束の重み	0.02
前フレーム系統誤差(位置)への拘束の重み	0.00001
前フレーム系統誤差(姿勢)への拘束の重み	0.00001

表 2 空間分散・時間分散最小化法における設定値

パラメータ種	設定値
指標検出のランダム誤差(pixel)	0.02
センサ計測値(位置)のランダム誤差(mm)	2.0
センサ計測値(姿勢)のランダム誤差(rad)	0.02
位置予測の誤差(mm)	5.0
姿勢予測の誤差(rad)	0.003
速度予測の誤差(mm/frame)	300.0
角速度予測の誤差(rad/frame)	0.2
系統誤差(位置)予測の誤差(mm)	1.0
系統誤差(姿勢)予測の誤差(rad)	0.1

を表 1 に示す。一方、空間分散・時間分散最小化法では、観測ノイズとシステムノイズの標準偏差を、物理センサの誤差特性やカメラの運動モデルに基づいて設定する。本実験で用いた設定値を表 2 に示す。なお、指標検出の統計的誤差の標準偏差は、位置合わせ精度を高めるために実際よりも小さめに設定している。これは非線形な誤差伝播の線形近似を補正していることに相当すると考えられる。

5.2. ジッタの評価

初めに、分布の悪い指標を観察した時に生じるジッタを評価する実験を行う。本実験では、図 3 の画面左側マーカのみを利用する。ジッタの影響のみを評価するために、マーカから約 500mm 離れた位置にカメラを固定し、その位置姿勢を推定する。推定結果の時間変動が少ないほど、ジッタを抑制できる安定性に優れた手法であるということを意味する。本実験では、二つの提案手法を、再投影誤差最小化法および部分パラメータ固定法と比較する。部分パラメータ固定法では、位置パラメータを固定し、画像情報を用いて姿勢パラメータを補正する。

実験の結果を図 4 に示す。グラフの縦軸は、カメラ位置姿勢推定値の Y 座標の値から平均値を差し引いた値、横軸はフレーム数を表している。グラフより、提案手法はいずれも再投影誤差最小化法よりも大幅にジッタが少ないことが分かる。また、提案手法では、物理センサによる計測位置をそのまま利用する部分パラメータ固定法と比較して、推定値が滑らかに遷移していることが分かる。

5.3. 系統誤差補正量の時系列平滑化の評価

次に、4.3. で提案した系統誤差補正量の時系列平滑化の効果を評価する実験を行う。本実験では、再投影誤差最小化法をベースとして、平滑化を行う場合と行わない場合の位置姿勢推定値の推移を比較する。

本実験では、図 3 の画面右側マーカの配置情報に 3mm の系統誤差を付加する。そして、平滑化を行う場合の系統誤差補正量が 0 から最大値まで遷移する時間を 30 フレームに設定する。

図 5 に、画面右側マーカが 5 秒毎に見え隠れする場合のカメラ位置姿勢推定値の推移を示す。グラフの縦軸はカメラ位置姿勢推定値の Z 座標の値、横軸はフレーム数を表している。150 フレーム付近で画面右側マーカが検出され、300 フレーム付近で検出されなくなり、450 フレーム付近で再び検出された様子を表す。グラフより、系統誤差補正量の時系列平滑化を行うと、検出マーカが切り替わる際の飛びが抑えられて、カメラ位置姿勢推定値の推移が滑らかになることが分かる。

5.4. 位置合わせ精度の評価

提案手法によって位置合わせの安定性が向上したとしても、精度が悪化するのでは、実用には耐えられない。そこで、提案手法の位置合わせ精度を評価する実

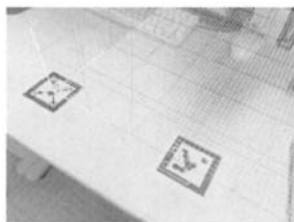
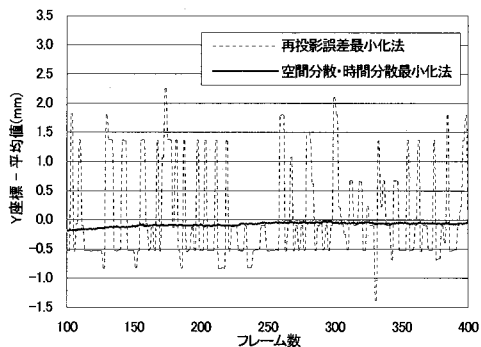
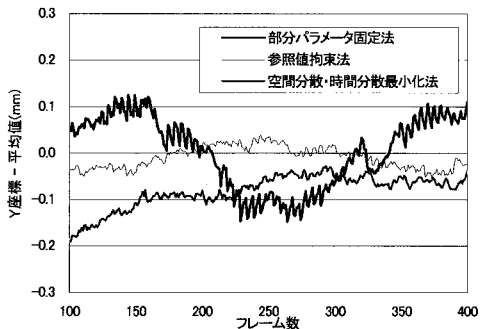


図 3 実験シーン



(a) 再投影誤差最小化法との比較



(b) 部分パラメータ固定法との比較

図 4 実験結果(ジッタの評価)

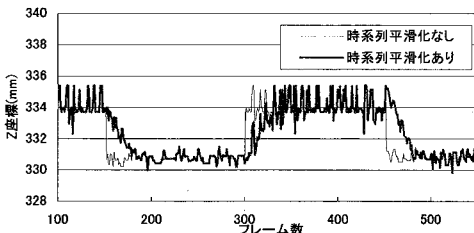


図 5 実験結果(系統誤差補正量の平滑化の評価)

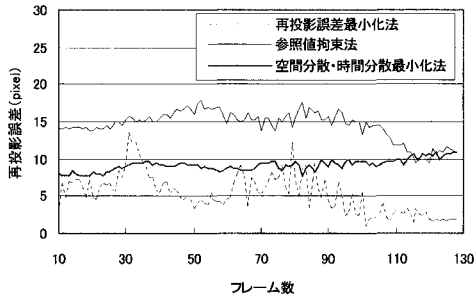


図6 実験結果(精度の評価)

験を行う。具体的には、位置合わせに利用する指標とは別に、評価用の指標を用意し、その指標の画像上での再投影誤差によってずれを評価する。このずれは、実際のMRアプリケーションにおける現実世界と仮想世界とのずれに相当するものである。そのため、提案手法の位置合わせの精度を評価する指標となる。本実験では、図3の画面左側マーカのみを位置合わせに利用し、画面右側マーカを評価用として用いる。

図6に、カメラを回り込むように動かしながらマーカを観察した時の再投影誤差の推移を示す。グラフの縦軸は再投影誤差、横軸はフレーム数を表している。グラフより、どちらの提案手法も、再投影誤差最小化法と比較して、安定性の向上と引き換えに精度が若干悪くなっていることが確認される。しかし、安定性と精度のバランスは、主観的には再投影誤差最小化法よりも大幅に改善されており、精度の悪化も許容可能なレベルであると考えられる。

6. おわりに

本論文では、複合現実感において位置合わせが不安定となる場合の現象であるジッタ、飛び、破綻について、その要因の分析を行い、それらの要因への対処方法について整理・考察した。そして、考察に基づいて、画像と物理センサの情報を相補的に活かしつつ、時系列方向の連続性を考慮して、かつ物理センサの系統誤差を分離して扱うことができる、二種類の手法を提案した。さらに、ビジョン推定値の系統誤差補正量を時系列的に平滑化する手法を提案した。

実用性を考えた場合、位置合わせの精度は安定性と同等に重要な要素である。提案手法を用いれば、位置合わせの安定性と精度をバランスさせることができる。

なお、提案手法では、指標の同定問題に関しては深く議論しなかった。今後は、指標の検出精度や同定の信頼性なども加味して、人為的なマーカとシーン中に元来存在する自然特徴を併用するような手法を検討する必要があると考える。

参考文献

- [Bajura 95] M. Bajura and U. Neumann, "Dynamic registration correction in augmented reality systems," *Proc. VRAIS '95*, pp.189-196 (1995).
- [張替 00] 張替, 辻井, 村田, 小野, 稲垣, 富田, "DGPS/INS 複合航法システムの理論精度解析とその飛行実証", 航空宇宙技術研究所報告 TR-1398 (2000).
- [Hoff00] W. Hoff and T. Vincent, "Analysis of head pose accuracy in augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol.6, no.4, pp.319-334 (2000).
- [Klein07] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," *Proc. ISMAR '07*, pp.225-234 (2007).
- [Koller 97] D. Koller, G. Klinker, E. Rose, D. Green, R. Whitaker, and M. Tuceryan, "Real-time vision-based camera tracking for augmented reality applications," *Proc. VRST '97*, pp.87-94 (1997).
- [Park 99] J. Park, B. Jiang, and U. Neumann, "Vision-based pose computation: robust and accurate augmented reality tracking," *Proc. IWAR '99*, pp.3-12 (1999).
- [佐野 90] 佐野, "帯域制限信号の外挿問題における特異値分解の最適化", 電子情報通信学会論文誌 A, vol.J73-A, no.6, pp.1139-1149 (1990).
- [佐藤 05] 佐藤, 内山, 山本, "UG+B 法:主観及び客観視点カメラと姿勢センサを用いた位置合わせ手法", 日本バーチャルリアリティ学会論文誌, vol.10, no.3, pp.391-400 (2005).
- [Skrypnik04] I. Skrypnik and D. G. Lowe, "Scene modeling, recognition and tracking with invariant image features," *Proc. ISMAR '04*, pp.110-119 (2004).
- [Thomas97] G. A. Thomas, J. Jin, T. Niblett, and C. Urquhart, "A versatile camera position measurement system for virtual reality TV production," *Proc. Int'l Broadcasting Convention '97*, pp.284-289 (1997).
- [内山 03] 内山, 山本, 田村, "複合現実感のためのハイブリッド位置合わせ手法—6自由度センサとビジョン手法の併用—", 日本バーチャルリアリティ学会論文誌, vol.8, no.1, pp.119-125 (2003).
- [Vacchetti04] L. Vacchetti, V. Lepetit, and P. Fua, "Combining edge and texture information for real-time accurate 3D camera tracking," *Proc. ISMAR '04*, pp.48-57 (2004).
- [横地 06] 横地, 池田, 佐藤, 横矢, "特徴点追跡とGPS 測位に基づくカメラ外部パラメータの推定", 情報処理学会論文誌, vol.47, no.SIG5(CVIM13), pp.69-79 (2006).