

ボトムアップ・トップダウン統合アプローチとアピアランス 自動登録によるマーカを用いない3次元物体追跡手法

A Fiducial-less 3-D Object Tracking Method Using Integration of Top-down and Bottom-up Approaches
and Automatic Appearance Registration

大隈隆史 蔵田武志 坂上勝彦

Takashi OKUMA, Takeshi KURATA, and Katsuhiko SAKAUE

産業技術総合研究所 知能システム研究部門

(〒305-8568 つくば市梅園 1-1-1 中央第2, {takashi-okuma, t.kurata, k.sakaue}@aist.go.jp)

Abstract: In this paper, a fiducial-less 3-D object tracking method are proposed. The proposed method acquires a position and an orientation of an object using a bottom-up approach which calculates an analytic solution with detected known points in an image frame. When a reliability of the acquired solution is low, the method uses a top-down approach to obtain the parameters. The top-down approach represents probability density of the parameters by samples and weights, and tracks the parameters by iteration of sampling probability density and evaluating each sample using known model. The method expands its tracking area by automatically registering image frames and estimated reliable parameters to a database.

Key Words: augmented reality, fiducial-less tracking, vision based tracking.

1. はじめに

拡張現実感技術において現実物体と仮想物体の位置合わせ問題は最も重要な研究課題の一つであり、これまでさまざまな取り組みがなされている[1]。位置合わせに用いるセンサは各応用システム構築時の条件に応じて適切に選択する必要があるが、ビデオスルー型の拡張現実環境では、現実環境提示用の画像を用いて画像処理により視点の位置姿勢を求める手法がよく採用される。この種の位置合わせ手法は原理的に環境側に計測用デバイスを配置する必要が無いことから、特にウェアラブルシステムでの利用が期待される[2-4]。画像処理による位置合わせは既知のモデルを持つ対象を撮影した画像の特徴から、対象に対するカメラの位置姿勢を推定する問題となる。これまでに提案されているカメラの位置姿勢推定手法は、

1. 画像を解析した結果と既知のモデルの特徴を対応付け、対応結果から位置姿勢を計算により求めるボトムアップのアプローチをとる手法[2][5-8]、
2. 推定される位置姿勢候補をサンプリングし、モデルから得られるパラメータと画像から得られるパラメータを比較し、評価することで解を決定するトップダウンのアプローチをとる手法[3][9]、

の2種類に分類できる。拡張現実感研究においては、主にボトムアップアプローチが採用されてきた。ボトムアップアプローチは

1. 基準点検出処理：外見上の特徴（アピアランス）と3次元座標が既知である位置合わせの基準点を、検出・追跡して画像

平面上での2次元座標値を取得する処理、

2. 位置姿勢計算処理：基準点の2次元座標値と3次元座標からカメラの位置姿勢を計算する処理、

で実現される。実環境画像を扱う場合、ノイズや視点位置の変化に伴い基準点のアピアランスが変化し、基準点検出処理は困難となる。そこで実時間で安定した推定を実現する必要から、環境中に配置したマーカを利用する工夫がなされてきた[5-8]。しかし、実環境中いたるところにマーカが配置される状況は不自然で、応用を限定しかねないため、マーカを用いない位置合わせ手法の開発が試みられている[2-4][8]。この場合、マーカで改善していた問題に別のアプローチから対応する必要がある。

加藤ら[8]は位置合わせの基準点となる自然特徴点をテンプレートマッチングで追跡する際、前フレームのカメラパラメータ推定結果によるテンプレート変形で自然特徴点のアピアランス変化に対応し、安定した追跡を実現している。また、多重解像度のテクスチャ情報を事前に準備しておくことで観察距離の変化にも対応している。ただし、基準点はテクスチャ情報を事前に取得できる平面上の点に限定される。また、テンプレートマッチングの回数を削減するために、位置姿勢計算に用いる基準点を画像上の位置によって必要最小限数順番に選択していく手法を取っているため、選ばれた基準点が誤追跡された場合には正しい解を計算することはできない。一般に必要な最小数の基準点のみを追跡して解を求めるアプローチの場合、基準点の誤追跡による位置合わせの失敗は避けられない。

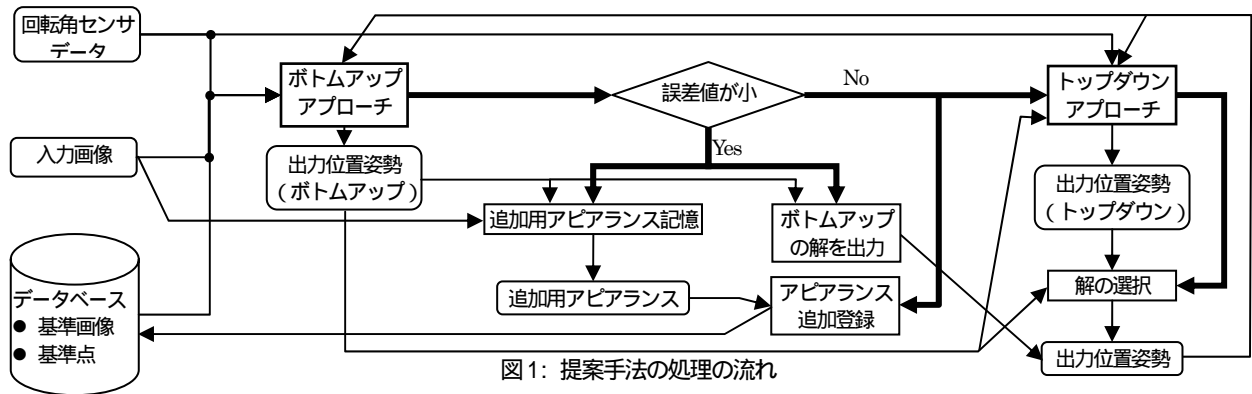


図1: 提案手法の処理の流れ

一方、トップダウンアプローチでは各フレームにおける位置姿勢の確率密度をサンプルの集合と各サンプルの評価値の組による離散表現で保持し、直前時刻の確率密度と運動予測に従ったサンプル生成と、既知のモデルと画像を用いての各サンプルの評価を繰り返すことで対象を追跡する。各フレームの出力は確率密度から位置姿勢の代表値を計算して用いる。トップダウンアプローチは確率密度の離散表現で位置姿勢を保持しているため、ボトムアップアプローチに比べて特徴点の誤検出の影響を受けにくい。しかし実時間処理での位置合わせを実現するためにはサンプル生成の範囲や密度などに工夫を要する。

以上の背景から本研究では、ボトムアップとトップダウンのアプローチを相補的に用いる安定した追跡手法を提案する。

2. 提案手法

2.1 手法の概要

処理の流れを図1の太矢印、データの流れを細矢印に示す。提案手法では多数の解候補をボトムアップアプローチにより計算し、ロバスト統計量により解を決定する。求めた解の誤差(後述)が小さい場合はこれを出力とし、誤差が大きい場合にはトップダウンアプローチからも解を求めて、より誤差の小さい解を出力する。サンプリング範囲を効果的に狭めるために、ウェアラブルシステムでも用いることのできる回転角センサをカメラに取り付けて利用する。また、解の誤差が小さい状態での入力画像と対象の位置姿勢を必要に応じてデータベースに自動登録することで、基準点のアピアランス変化に対応し、追跡範囲を拡大する。

提案手法では追跡対象とする3次元物体について、以下の項目は事前知識として利用可能とする。

- カメラ内部パラメータ。
- 基準画像：追跡対象を撮影した画像。

基準点：位置合わせの基準となる追跡対象上の点。Lucas-Kanade法[10]で基準画像から選択した追跡対象上の自然特徴点を基準点とし、基準画像上の2次元座標と実環境での3次元座標を計測しておく。

2.2 ボトムアップアプローチ

提案手法で用いるボトムアップアプローチの位置姿勢推定ステップでは複数の基準画像の利用と基準画像の光軸周りの回転処理によって基準点検出が可能な領域を拡大する。また、ロバスト統計量の導入によって位置姿勢計算処理の特徴点の誤追跡に対するロバスト化を図る。以下、提案手法で導入したそれぞれの工夫について説明する。

2.2.1 複数基準画像の利用と基準画像の回転

提案手法では画像局所領域の勾配情報を用いて繰り返し演算で基準点を追跡するLucas-Kanade法を用いる。繰り返し演算の初期値となる基準点の2次元座標は前フレームの位置姿勢の推定結果を用いて3次元座標を画像平面上に投影した位置とする。Lucas-Kanade法は特徴点の平行移動のみを仮定しているため、アピアランス変化に強いとはいえない。前述の加藤らの手法のように前フレームの位置姿勢を用いて基準画像を变形する方法も考えられるが、追跡対象物体が平面でない場合には見た目が变形結果と大きく異なる可能性もある。そこで提案手法では登録された複数の基準画像を用いる。登録基準画像の増加に伴う計算量の増加を最小限にするため、前フレームの推定結果の計算に使用された基準画像と、前フレームの位置姿勢と一番近い位置姿勢が登録されている基準画像のふたつ(一致する場合はひとつ)を選択して利用する。また、対象の形状の影響を受けない画像の光軸まわりの回転を考慮し、前フレームの推定結果と回転角センサによって得られた前フレームから現在までのカメラの回転量を用いて選択された基準画像を回転し、特徴点追跡に用いる。

2.2.2 LMedS基準を用いた位置姿勢計算処理

n 個の追跡された基準点の2次元座標と3次元座標からPnP問題の解法によりカメラ位置姿勢の解析解を求めることができる。しかしながら n 個全てが正確に追跡されることはまれであり、これを前提として解を求めると正確な解を得られない。そこで提案手法ではロバスト統計量として代表的なLMedS基準を用いて以下の手順で解を求める。

手順1 n 個の特徴点からランダムに3点を選択し、P3P問題の解法を用いて解候補を計算する。

手順2 次式で定義される解候補の誤差値を計算する。

$$err^2 = \text{med} \left((x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2 \right) \quad (1)$$

ただし、 (x_i, y_i) は i 番目の基準点の画像上での観測座標値、 $(\tilde{x}_i, \tilde{y}_i)$ は i 番目の基準点を解候補で画像平面上へ投影したときの座標値、 $\text{med}(f(i))$ は全ての i における $f(i)$ の中央値を表すものとする。

手順3 手順1.と2.を m 回繰り返す。割合 r の特徴点が正確に追跡された場合に確率 p より高い確率で正しい解を求めるための繰り返し回数 m は次式により計算される。

$$p < 1 - (1 - r^3)^m \quad (2)$$

手順4 誤差値最小となる解候補を解として出力する。

2.3 トップダウンアプローチ

2.3.1 処理の概要

提案手法ではボトムアップアプローチによって信頼できる解が得られなかった場合、トップダウンアプローチによる位置姿勢推定を行う。本研究ではトップダウンアプローチとして ConDensation アルゴリズム[9]を用い、以下の手順を毎フレーム繰り返すことで位置姿勢を推定する。

手順1 サンプル生成処理: 前フレームの状態に応じてサンプル集合を生成する処理。詳細は2.3.2節で述べる。

手順2 観測処理: 入力画像を観測して各サンプルの評価値を計算する。各サンプルの評価値の計算にも式(1)を用いるためには基準点3次元座標を各サンプルで画像平面上に投影した座標 $(\tilde{x}_i, \tilde{y}_i)$ を初期値とする特徴点追跡により対応する基準点の入力画像中の座標 (x_i, y_i) を求めなければならず、計算量が膨大になる。そこで、提案手法では事前知識の準備で用いた Lucas-Kanade 法で入力画像中の自然特徴点を検出し、各々の $(\tilde{x}_i, \tilde{y}_i)$ に最も近い検出特徴点 (x'_i, y'_i) で (x_i, y_i) を代用して式(1)の誤差値を計算し評価する。

手順3 代表値決定処理: 推定された解の確率密度からの代表値として誤差値最小のサンプルを出力する。

2.3.2 サンプル生成処理

提案手法では効果的なサンプルを重点的に生成するために、ボトムアップアプローチの計算結果と回転角センサ計測値を利用する。回転角センサでは3自由度しか計測できないため、ユーザの運動モデルを設定して3自由度の計測値を6自由度パラメータのサンプル生成処理で利用する。想定した対象観察動作は以下の3通りである。

- 観察対象を視野中で動かす観察動作: 回転角センサの値には影響を与えない。
- 観察対象を中心として頭部を動かす観察動作 (物体中心回転): 追跡対象は見かけ上、追跡対象を中心として回転角センサの計測結果と逆に回転する。
- 頭部を中心に大きな対象を見回す観察動作 (頭部中心回転): 追跡対象は見かけ上、頭部を中心として回転角センサで計測結果分回転する。

回転角センサのドリフトや誤差の蓄積を防ぐために、前フレームからの計測値の変化量のみを利用することとした。実際には以下の手順で時刻 t のサンプル集合 $\{s_t^{(1)}, s_t^{(2)}, \dots, s_t^{(N)}\}$ を生成する。ただし、 M は全サンプル数、 s_t は時刻 t のサンプルで位置姿勢を表す。

まず、全体の1/4のサンプル $\{s_t^{(1)}, s_t^{(2)}, \dots, s_t^{(N/4)}\}$ を、時刻 t のボトムアップアプローチの計算結果を分布中心とするランダムサンプリングによって生成する。残りの3/4のサンプルについては時刻 $t-1$ の出力がボトムアップ、トップダウンのどちらのアプローチにより計算されたかで生成方法を変える。

時刻 $t-1$ の出力がボトムアップアプローチで計算された場合:

B-1 時刻 $t-1$ の出力を分布中心とするランダムサンプリングによって $\{s_t^{(N/4+1)}, s_t^{(N/4+2)}, \dots, s_t^{(N/2)}\}$ を生成。

B-2 頭部中心回転運動を想定して時刻 $t-1$ の出力を回転角センサの値を用いて回転変換した位置姿勢を中心とするランダムサン

プリングによって $\{s_t^{(N/2+1)}, s_t^{(N/2+2)}, \dots, s_t^{(3N/4)}\}$ を生成。

B-3 物体中心回転運動を想定して時刻 $t-1$ の出力を回転角センサの値を用いて回転変換した位置姿勢を中心とするランダムサンプリングによって $\{s_t^{(3N/4+1)}, s_t^{(3N/4+2)}, \dots, s_t^{(N)}\}$ を生成。

時刻 $t-1$ の出力がトップダウンアプローチで計算された場合:

T-1 ConDensation アルゴリズムに従ったサンプル選択処理: 時刻 $t-1$ のサンプル集合 $\{s_{t-1}^{(1)}, s_{t-1}^{(2)}, \dots, s_{t-1}^{(N/4)}\}$ から各サンプルの評価値の比に従って全サンプル数の3/4を選択する。選択されたサンプル集合を $\{s_t^{(1)}, s_t^{(2)}, \dots, s_t^{(3N/4)}\}$ とする。

T-2 選択されたサンプルのうち $\{s_t^{(1)}, s_t^{(2)}, \dots, s_t^{(N/4)}\}$ をランダムウォークで更新して $\{s_t^{(N/4+1)}, s_t^{(N/4+2)}, \dots, s_t^{(N/2)}\}$ を作成。

T-3 選択されたサンプルのうち $\{s_t^{(N/4+1)}, s_t^{(N/4+2)}, \dots, s_t^{(N/2)}\}$ を頭部中心回転で変換した位置からのランダムウォークで $\{s_t^{(N/2+1)}, s_t^{(N/2+2)}, \dots, s_t^{(3N/4)}\}$ を作成。

T-4 選択されたサンプルのうち $\{s_t^{(N/2+1)}, s_t^{(N/2+2)}, \dots, s_t^{(3N/4)}\}$ を物体中心回転で変換した位置からのランダムウォークで $\{s_t^{(3N/4+1)}, s_t^{(3N/4+2)}, \dots, s_t^{(N)}\}$ を作成。

2.4 アピアランスの自動追加登録

提案手法では視点移動に伴う基準点のアピアランス変化のため、複数視点から撮影した基準画像と基準点を登録しているが、安定して追跡するために必要な画像の撮影位置は追跡対象の形状やテクスチャに依存する。このため、事前に必要な全情報を準備するためには追跡対象毎に必要な撮影位置を全て探し出し、その位置からの基準画像を登録するという煩雑な手順が必要となる。そこで提案手法では、事前に全てのアピアランスを準備するのではなく、追跡処理中に必要に応じてアピアランスを追加する方針をとる。すなわち、ボトムアップアプローチの誤差が大きくなり、トップダウンアプローチが開始されるときに新しいアピアランスが必要であると判断して、直前のフレームの入力画像、位置姿勢、追跡に成功した基準点の2次元座標を新しいアピアランスとしてデータベースに追加登録する。

3. 実験と考察

トップダウン・ボトムアップ統合処理と、アピアランス追加登録処理が追跡に与える影響を調べるための比較実験を行った。同じ画像系列を

手法1: 提案手法 (トップダウン・ボトムアップ統合処理 + アピアランス追加登録)

手法2: トップダウン・ボトムアップ統合処理のみ

手法3: ボトムアップ処理 + アピアランス追加登録

手法4: ボトムアップ処理のみ

の4通りの追跡方法で処理した場合の出力画像と誤差値を比較した。事前知識として登録される3次元座標の獲得にはレンジファインダ(MINOLTA VIVID700)を用いた。実験に用いた画像サイズは320x240、トップダウン計算処理のサンプル数は512サンプル、Intel Xeon 1.7MHz Dual CPUのPCで実験を行い、各フレームの処理時間はボトムアップのみの場合で平均約70ms、トップダウンまで実行された場合は約100msであった。ボトムアップの繰り返し回数 m は52回に設定した。図2に実験結果を示す。

使用した350フレームの入力シーケンスでは、第90フレーム

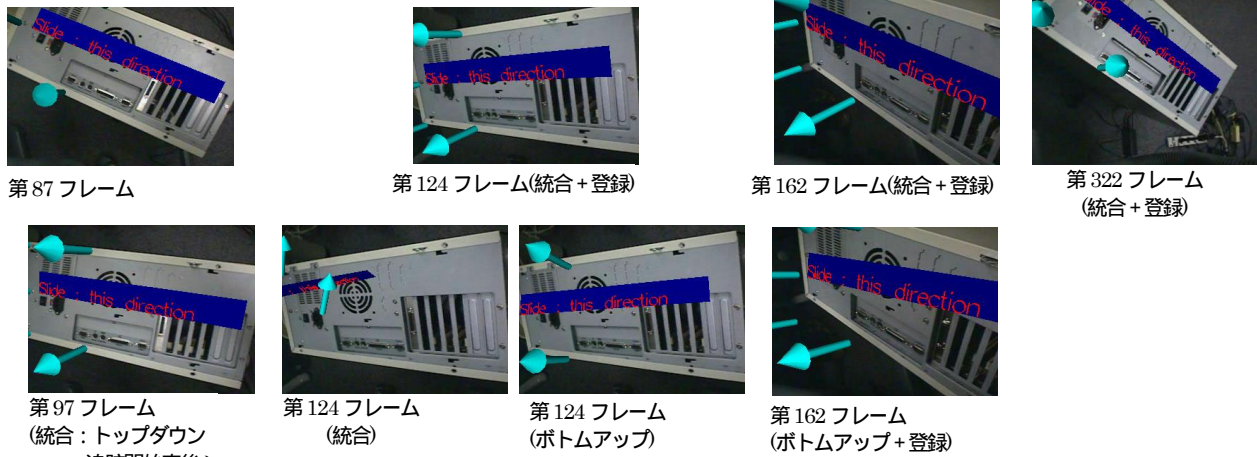
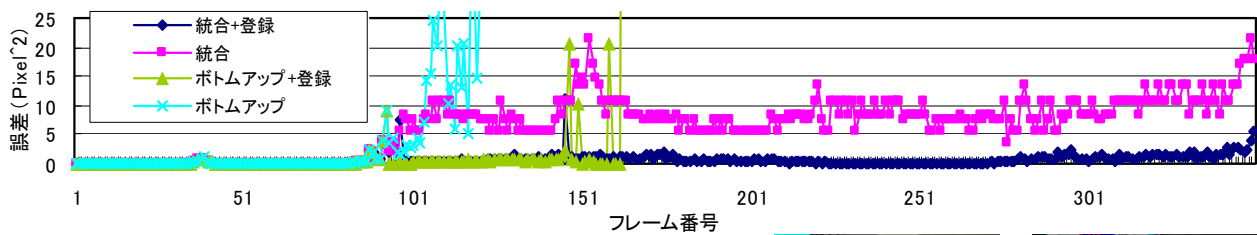


図 2: 追跡実験結果

前後までは主に平行移動と光軸方向の回転のみであったため、どの手法でも追跡ができた。その後姿勢を変えながら物体に近づくため、手法 4 での追跡が困難になる。手法 4 で追跡した場合、第 125 フレーム以降追跡できなかった。手法 1、手法 2 では第 96 フレームからトップダウン計算処理の出力が使用された。しかし、手法 2 では、しばらくたつと誤った位置姿勢で追跡し始めた(図中第 124 フレーム参照)。これはトップダウンアプローチで採用した評価法では特徴点を正確に対応付けていないため、特徴点が集中する部分で対応のずれが起きてしまったためと考えられる。手法 3 では視点の移動量が大きい第 162 フレーム目で正しい解が得られなかったため、第 163 フレーム以降の追跡に失敗した。ボトムアップアプローチのみを用いているので、多くの基準点が追跡に失敗して正しい解を計算できず、次のフレームでの処理が破綻していると考えられる。これに対して提案手法は短期間の失敗であれば基準点追跡結果を利用しないトップダウン計算処理により復帰することができる。また、アピアランス追加登録によりボトムアップ計算処理の追跡範囲を動的に広げている。この実験結果より提案手法で採用したボトムアップ・トップダウン統合処理とアピアランス追加処理の有効性を確認することができた。

4. まとめ

本稿ではマーカーを用いない 3 次元物体の位置姿勢追跡手法として、LMedS 基準とアピアランスの自動追加を用いたボトムアップアプローチとトップダウンアプローチの統合により安定した位置姿勢追跡を実現する手法について述べた。今後は基準点の自動追加によるオンラインでのデータベース生成を実現するシステムを構築する予定である。

参考文献

[1] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and

MacIntyre, B., "Recent Advances in Augmented Reality", IEEE Computer Graphics and Applications, Vol. 21, No. 6, pp. 34-47, 2001.

[2] Okuma, T., Kurata, T., and Sakaue, K., "Real-Time Camera Parameter Estimation for 3-D Annotation on a Wearable Vision System," IEICE Trans. Inf. Syst., Vol.E84-E, No. 12, pp.1668-1675, 2001.

[3] Okuma, T., Kurata, T., and Sakaue, K., "VizWear-3D: A Wearable 3-D Annotation System Based on 3-D Object Tracking using a Condensation Algorithm," In Proc. IEEE Virtual Reality 2002, pp.295-296, 2002.

[4] VizWear: <http://www.is.aist.go.jp/vizwear/>

[5] Neumann, U., and Cho, Y., "A self-tracking augmented reality system," In Proc. VRST 96, pp. 109-115, 1996.

[6] Rekimoto, J., "Matrix: A Realtime Object Identification and Registration Method for Augmented Reality," APCHI'98, 1998.

[7] Kato, H., and Billinghurst, M., "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System," In Proc. the 2nd IEEE and ACM International Workshop on Augmented Reality '99, pp.85-94, 1999.

[8] 加藤博一, 汐崎徳男, 橋啓八郎, " テクスチャー画像からオンライン生成されたテンプレートのマッチングに基づく拡張現実感のための位置合わせ手法 ", 日本バーチャルリアリティ学会論文誌, Vol.7, No.2, 2002.

[9] Isard, M. A., "Visual Motion Analysis by Probabilistic Propagation of Conditional Density," Ph.D. thesis, Department of Engineering Science, University of Oxford, 1998.

[10] Lucas, B.D. and Kanade, T., "An Iterative Image Registration Technique with an Application to Stereo Vision," Proc. DARPA Image Understanding Workshop, pp.121-130, 1981