

頂点追跡に基づく AR のためのマーカレス位置合わせ手法

廣瀬 亮[†] 斎藤 英雄[†]

[†] 慶應義塾大学理工学部情報工学科

現実世界をベースにして仮想空間を自然に融合する AR(Augmented Reality) 技術において、現実世界と仮想空間の位置合わせ問題が最も重要な課題となっている。この位置合わせ問題を解決する手法として、本稿では、実験環境中に配置された物体の 3 次元モデルを利用し、画像から物体頂点を抽出・追跡することによって、マーカレスな位置合わせを行う手法を提案する。また、本手法では 6 自由度センサである磁気センサを併用することで、素早い動きに対するロバスト性を高めることができ、対象物体が視野に入っていない場合においても、カメラ位置姿勢の推定を安定して行うことが可能となる。

Marker-less Registration Method for Augmented Reality Based on Tracking Vertex of Model

Ryo Hirose[†], Hideo Saito[†]

[†] Department of Information and Computer Science, Keio University

The most important issue in AR (Augmented Reality) is accurate registration between real world and virtual world. We proposed a marker-less registration method. We use the vertices of 3D model of the target object. We estimate 6DOF camera position by detecting and tracking the vertices every frame. Furthermore we use the magnetic 6DOF sensor. By combining vision base method and sensor based method, the robustness of rapid camera movement and accuracy of registration can be improved. We can also estimate camera position even if the object doesn't appear the field of view.

1 はじめに

現実空間とコンピュータによって生成された仮想空間 (VR:Virtual Reality) を融合してユーザに提示する方法として、複合現実感 (MR:Mixed Reality) という技術の研究が近年盛んに行われている [1][2]。この MR 技術のうち、現実空間をベースに仮想空間の情報を附加させる技術のことは、特に拡張現実感 (AR:Augmented Reality) と呼ばれている。AR/MR では現実世界と仮想空間を同時に扱うため、それぞれの空間の情報を補完・増強しあうことができ、単独では味わうことができなかつたような新しい世界を生み出すことが可能となる。そして、この技術を応用することで、医療・福祉、インテリアデザイン・都市計画、ナビゲーションシステム、作業支援などさまざまな分野に対して実用化させることができ期待されている [3]。

これらの AR/MR を自然な形で実現させるためには、幾何学的整合性、時間的整合性、光学的整合性といった課題を解決する必要があり、さまざまな手法が提案されている [4]。本稿では AR/MR において最も重要な課題である幾何学的整合性、すなわち現実空間と仮想空間の位置合わせ問題を解決する手法について述べる。

位置合わせ問題を解決する手法は大きく以下のように分類することができる。

1. 物理センサを用いる手法
2. ビジョンベースによる手法
 - 人工マーカを用いる手法
 - 自然特徴を用いる手法
3. 物理センサとビジョンを融合した手法

物理センサを用いる手法の利点としては、視点や環境の変化にロバストであり、ユーザやカメラの位置姿勢を簡単に計測することができるため、ユーザの負担が少ないとすることが挙げられる。欠点としては、センサなどの高価で特別なハードウェアが必要なこと、移動範囲がセンサの有効範囲に限られてしまうなどが挙げられる。

次に、ビジョンベースによる手法は入力画像中の情報を用いることによって位置姿勢を推定する手法で、現実空間中に人工マーカを配置しておくもの[5]と、現実空間中の自然特徴のみ用いる手法[6]、マーカと自然特徴を併用した手法[7]などがある。

人工マーカを使用すると画像中からマーカを抽出することが容易となり、安定かつ高速に位置姿勢の推定を行うことができる。しかしマーカの配置やメンテナンスに手間がかかること、マーカの3次元位置が既知であること、個別に識別できるマーカを用いる必要があること、マーカの存在する範囲でしか実現できないため屋外での使用が困難、オリジナルの景観を壊してしまうなどさまざまな欠点がある。

このような問題を解決する方法として、空間中に存在するエッジやコーナー、平面などの自然特徴を利用する手法[6]が提案された。自然特徴を利用するということは、位置合わせ可能な領域に制限がない上、オリジナルの景観を損ねるといった問題がなくなる。しかし、人工マーカを使用した場合と比べ、画像中から特徴を抽出することが困難であるため、誤対応・誤追跡により精度が落ちてしまう。また、計算時間が長くなるという欠点も存在する。

最近では、形状が既知である物体や環境などの3次元モデルを作成し、それを人工マーカの代わりとして使用する手法[8][9]が注目されている。大江らはあらかじめ撮影の対象とする環境を全方位カメラで撮影し、それを3次元復元とともに、ランドマークとして使用する特徴点の3次元位置とその周囲のテンプレート画像を取得し、入力画像とマッチングを取ることでカメラ位置姿勢を推定している。また、Lepetitらは、対象物体の3次元形状をCADモデルとして作成し、特徴点の周囲のテンプレートを取得し、それに少し変形を加えたイメージパッチを作成する。そして入力画像とそれらのイメージパッチを固有空間上で比較することで高速なマッチングを行い、カメラ位置姿勢の推定を行う。どちらの手法においても、対象とする環境

や物体のモデルの作成に複雑な処理が必要であり。入力画像がテンプレート画像を学習したときと大きく異なる軌跡を通った場合にはこれらの手法は適用できないといった問題点があった。そのため、簡単に作成できる3次元モデルを利用することや、視点の位置姿勢に依存するテンプレートマッチング法を用いない位置合わせ手法が望ましいといえる。

そこで、本手法では、実験環境に配置された対象物体の3次元モデルと、カメラ視点の位置姿勢に依存しない物体頂点を利用して正確な位置合わせを行う手法を提案する。また、ロバスト性を向上させるために磁気センサを併用する。

2 手法の概要

入力画像撮影用のカメラとして、磁気センサを内蔵したHMD(Head Mounted Display)を使用する。撮影環境中に3次元形状が既知の物体を配置し、HMDのカメラによって撮影された入力画像列と、磁気センサの出力である外部パラメータを入力とする。

続いて3次元モデルを磁気センサの出力によって画像上に投影した点と、画像から抽出した特徴点を対応付けることにより、3次元モデル頂点と画像の2次元座標の対応付けを行う。その際、初期フレームと2フレーム目以降では異なる処理方法をとる。初期フレームでは、磁気センサの出力値が信頼できない場合にも対処するために、カメラ位置姿勢の初期化として大まかなセンサの補正処理を行うことを目的とした対応付け(大局的位置合わせ)を行う。2フレーム目以降では、正確なカメラ位置姿勢の推定を行うことを目的とした密な対応付け(局所的位置合わせ)を行う。

現フレームにおけるカメラ位置姿勢は、前フレームで推定されたカメラ位置姿勢に、現フレームと前フレームの磁気センサの出力値の変化量を加えることによって推定することできる。このように磁気センサの変化量を足し合わせることは、カメラが大きく移動した場合にも対応できるようになり、ロバスト性を向上させることができる。次に、このカメラ位置姿勢の推定値 R_0, t_0 を入力として3次元モデル頂点を画像に投影し、その投影点の周囲で物体頂点と思われる特徴点を探して対応付けを行う。

最後に最適化処理によってモデル頂点の投影点と特徴点との投影誤差が最小となるようなカメラ位置姿勢を推定する。また、カメラの内部パラメータと磁気セン

サのキャリブレーションは事前に行われているため、推定された外部パラメータ R , t と内部パラメータ A を用いることで、仮想物体を投影表示することができる。

2.1 3次元モデルの投影

3次元モデルの頂点座標と画像から抽出した特徴点を対応付けるには、3次元モデルの頂点座標を画像に投影し、対応付けする必要がある。

磁気センサから得られる情報は、3次元の並進ベクトル $t = [t_x \ t_y \ t_z]^\top$ と、回転軸ベクトル $r = [r_x \ r_y \ r_z]^\top$ ($r_x^2 + r_y^2 + r_z^2 = 1$) と回転軸に対する回転角 r_a である。ここで回転行列 R は、回転軸ベクトル r と回転角 r_a を用いることにより、式(1)のように表すことができる。

$$R = \begin{bmatrix} r_x^2(1 - \cos r_a) + \cos r_a & r_x r_y(1 - \cos r_a) - r_z \sin r_a & r_x r_z(1 - \cos r_a) + r_y \sin r_a \\ r_x r_y(1 - \cos r_a) + r_z \sin r_a & r_y^2(1 - \cos r_a) + \cos r_a & r_y r_z(1 - \cos r_a) - r_x \sin r_a \\ r_x r_z(1 - \cos r_a) - r_y \sin r_a & r_y r_z(1 - \cos r_a) + r_x \sin r_a & r_z^2(1 - \cos r_a) + \cos r_a \end{bmatrix} \quad (1)$$

ここで、磁気センサから得られる外部パラメータはカメラ座標系原点に対する世界座標系の位置成分と姿勢成分であることから、3次元の並進ベクトル t と式(1)によって得られた 3×3 の回転行列 R を用いることで、式(2)のように、カメラ座標系上の3次元座標 X_c を、世界座標系上の3次元座標 X_w に変換することができる。

$$X_w = RX_c + t \quad (2)$$

したがって、世界座標系上の3次元点 \tilde{X}_w をカメラ座標系上の3次元点 X_c に変換するには、

$$X_c = [R^\top | -R^\top t] \tilde{X}_w \quad (3)$$

として計算する必要がある。

ところで、カメラの透視投影変換はカメラの内部パラメータ A を用いることで、

$$\lambda \tilde{x} = A X_c \quad (4)$$

と表すことができるので、式(3)と式(4)より、3次元モデル上の i 番目の頂点 \tilde{X}_{wi} は、式(5)によって画像上の点 \tilde{x}_i に投影することができる。

$$\lambda \tilde{x}_i = A [R^\top | -R^\top t] \tilde{X}_{wi} \quad (5)$$

2.2 初期フレームにおける対応付け処理

磁気センサは近くに磁場を乱すような金属類がないような広い空間で使用する場合には理想的な出力が期待できる。しかし、一般にそのような環境を作成することは容易ではないため、磁気センサの出力には磁場の乱れによる誤差が含まれてしまう場合が多い。そのような場合にも対処できるように、初期フレームではカメラ位置姿勢の初期化を目的とした大まかな補正処理を行うための対応付けを行う。このときの処理の流れを図1に示す。

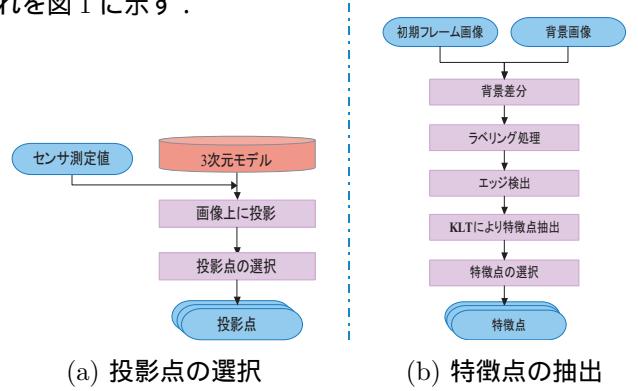


図1: 初期フレームにおける投影点・特徴点抽出処理

初めに、図1(a)に示すように対応付けさせる投影点を選択する。磁気センサの出力 R_0 と t_0 を用いて3次元モデルを画像上に投影する。初期フレームでは大まかな対応付けが求められているため、図2のように画像上で最も手前に見えている面を構成する4頂点を、画像から抽出する特徴点と対応付けする。同時に投影点の縦エッジの長さを算出し、画像処理によって検出された特徴点のうち対応点としてふさわしいものを選び出す過程において利用する。

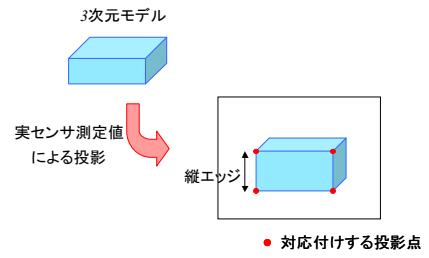


図2: 対応付けする投影点の選択

続いて図1(b)に示すように、特徴点の抽出処理を行う。本手法では、あらかじめ図3(a)のような対象とする環境の背景画像を撮影しておく必要がある。この背景画像と図3(b)の入力画像との差分をとることに

より、図 3(c) のように対象物体を抽出する。これには対象物体の影による影響などが含まれているため、ラベリング処理を行うことで、誤差による影響を取り除き、図 3(d) のように物体の正面のみを取り出す。

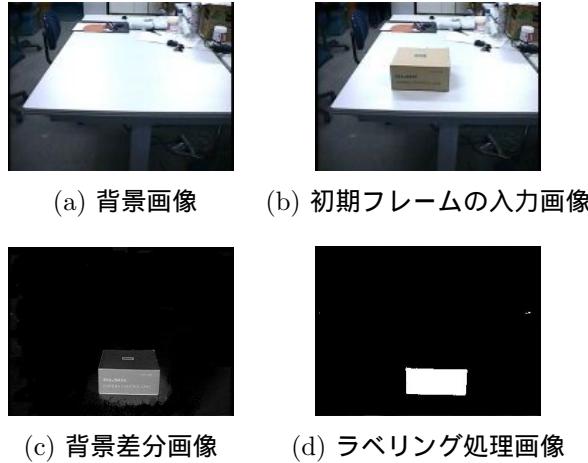


図 3: 背景差分とラベリング処理

ラベリング処理した画像から図 4(a) のようにエッジ検出をし、図 4(b) のように x, y 軸方向に対してそれぞれプロジェクションをとる。プロジェクションが極大となる 2 点を選び出すことによって、物体正面のエッジ位置を決定することができ、図 4(c) のように特徴点の探索範囲を限定できる。最後にその探索範囲内で物体頂点を検出するために、KLT Tracker によって特徴点検出を行い、検出された特徴点の中から先ほど求めた縦エッジの長さとほぼ同じ長さになるような特徴点対を見つけることで、図 4(d) のように両サイドからそれぞれ 2 点ずつ物体頂点を検出する。

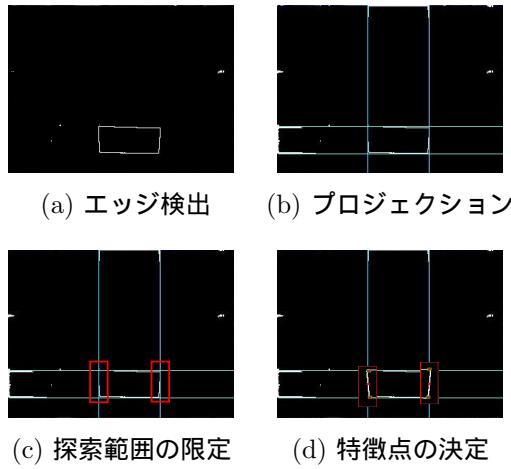
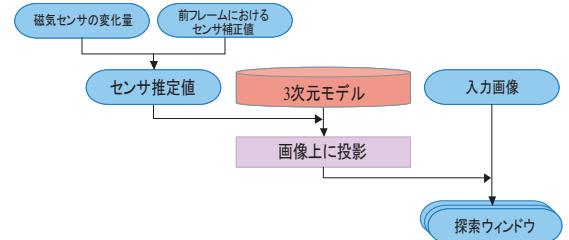


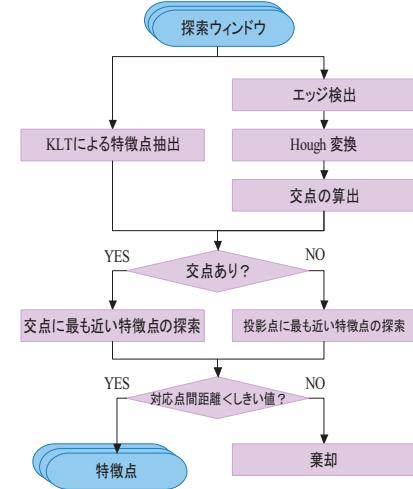
図 4: 特徴点の検出方法

2.3 2 フレーム目以降における対応付処理

2 フレーム目以降における 3 次元モデルの頂点と画像から抽出した特徴点の対応付けの処理を図 5 に示す。



(a) 探索ウィンドウの設定



(b) 探索ウィンドウ内での特徴点抽出と対応付け

図 5: 2 フレーム目以降における対応付け処理

2 フレーム目以降では、まず現フレームにおけるカメラ位置姿勢を推定する。前フレームと現フレームでの磁気センサの出力の変化分がカメラの移動量であるため、前のフレームで推定されたカメラ位置姿勢にこの変化量を加えることにより、現フレームでの初期カメラ位置姿勢 R_0, t_0 を推定できる。

推定された R_0, t_0 を用いて 3 次元モデルを画像に投影し、その周囲に図 6 のような小さい探索領域（探索ウィンドウ）を設定する。そして、全ての投影点に対して探索ウィンドウを設け、それぞれの探索ウィンドウの中から最も適切であると思われる特徴点を検出する。

続いて、図 5(b) で示した、各探索ウィンドウ内での特徴点抽出処理と対応付け処理について述べる。まず図 7 のような探索ウィンドウ内で KLT Tracker によ

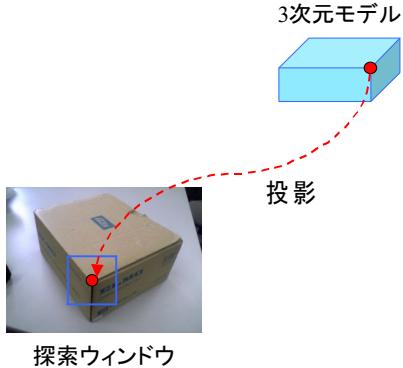


図 6: 探索ウィンドウの設定

る特徴点の抽出処理を行う。ここで検出される特徴点は図 8 のように、物体の頂点位置を検出することができているが、物体表面のテクスチャや、物体周囲に存在する物体などからも特徴点が検出されてしまうため、この方法だけで対応付けを行なうことはできない。



図 7: 探索ウィンドウ

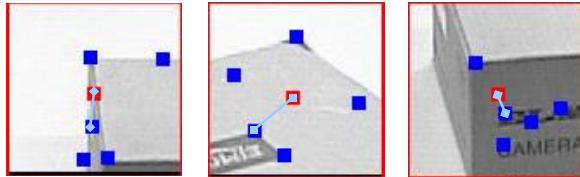


図 8: KLT Tracker による特徴点抽出

そこで探索ウィンドウ内で Hough 変換を行い、エッジと思われる直線を最大で 2 本検出する。画像によっては直線が 1 本しか検出されない場合や 1 本も検出されない場合も存在するため、2 直線検出された場合に限り交点座標を算出する。

そして図 10 のように、Hough 変換によって算出された直線の交点座標に最も近い KLT Tracker によって検出された特徴点を探索する。一方、Hough 変換によって交点が算出されなかった場合には投影点に最も近い KLT 特徴点を探索する。このとき、2 点間の距離が近い場合には対応点対として受理し、離れている場合には信頼できるものではないとして棄却する。

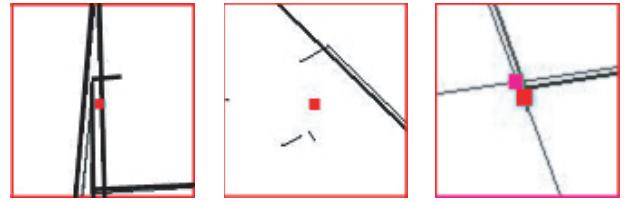


図 9: Hough 変換による交点検出

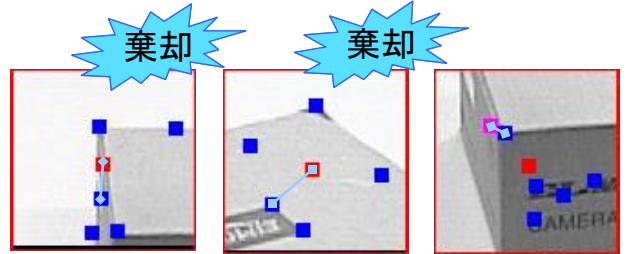


図 10: 特徴点の決定の仕方

2.4 最適化処理

3 次元モデル頂点と画像から抽出した特徴点との対応付けが行われた場合、それらを用いて最適な外部パラメータである R, t を推定する最適化処理を行う。

本手法において最適化させるパラメータは、並進成分 t と回転成分の回転軸ベクトル r 、回転角 r_a を合わせた 7 自由度であり、ベクトル s として表せる。

$$s = \begin{bmatrix} t_x & t_y & t_z & r_x & r_y & r_z & r_a \end{bmatrix}^\top \quad (6)$$

パラメータ s を最適化させる繰り返し演算の手法として最急降下法を用いる。そこで、式 (5) で表される $x = (x, y)$ を偏微分するヤコビ行列 J_{us} を求める。

$$J_{us} = \frac{\partial \mathbf{x}}{\partial \mathbf{s}} = \left[\begin{array}{ccccccc} \frac{\partial x}{\partial r_x} & \frac{\partial x}{\partial r_y} & \frac{\partial x}{\partial r_z} & \frac{\partial x}{\partial r_z} & \frac{\partial x}{\partial t_x} & \frac{\partial x}{\partial t_y} & \frac{\partial x}{\partial t_z} \\ \frac{\partial y}{\partial r_x} & \frac{\partial y}{\partial r_y} & \frac{\partial y}{\partial r_z} & \frac{\partial y}{\partial t_z} & \frac{\partial y}{\partial t_x} & \frac{\partial y}{\partial t_y} & \frac{\partial y}{\partial t_z} \end{array} \right] \quad (7)$$

ヤコビ行列が求まるとき、 s の微小な変化量 Δs に対する x の微小な変化量を計算することができる。

対応点対が N 組 ($4 \leq N \leq 8$) あるとき、画像から抽出された特徴点 $v_i = (u_i, v_i)^\top$ と、3 次元モデル頂点の投影点 $x_i = (x_i, y_i)^\top$ との二乗誤差の和は

$$f(s) = \sum_{i=0}^N \{(u_i - x_i)^2 + (v_i - y_i)^2\} \quad (8)$$

として計算できる。この $f(s)$ を 0 に近づけるために s を繰り返し更新することで、最適なカメラ位置姿勢を算出する。

3 実験

本手法の有用性を実証するために、図 11 で示す 3 次元形状が既知な物体を実験環境中に配置し、ビデオシースルーハードウェア（Canon VH-2002 Video See-through HMD）を使用して 3 つの実験を行った。



(a) 対象物体 A (b) 対象物体 B (c) 対象物体 C
図 11: 対象物体

3.1 実験 1

推定されたカメラ位置姿勢を用いて 3 次元モデルを画像上に投影し、比較することで本手法の有用性を示す。図 12～14 に処理の様子と補正前後の外部パラメータによって投影した結果を示す。いずれの場合も、投影した 3 次元モデルと対象物体が画像上で殆ど一致しており、本手法により安定に位置合わせが行われていることがわかる。また、図 12 では、対象物体が視野外に一旦出てしまった後でも、安定に追跡できていることがわかる。

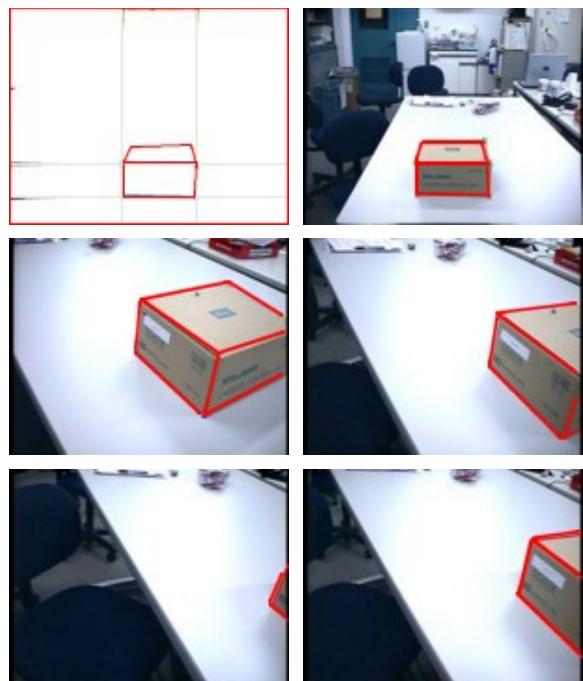


図 12: 対象物体 A の場合の結果

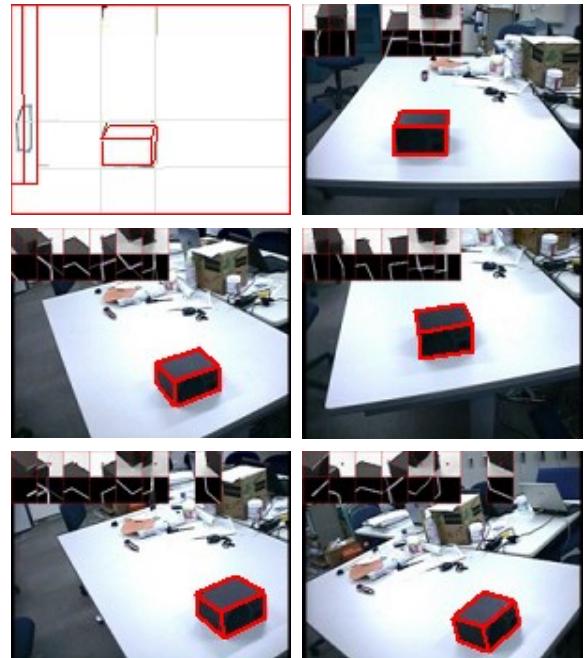


図 13: 対象物体 B の場合の結果

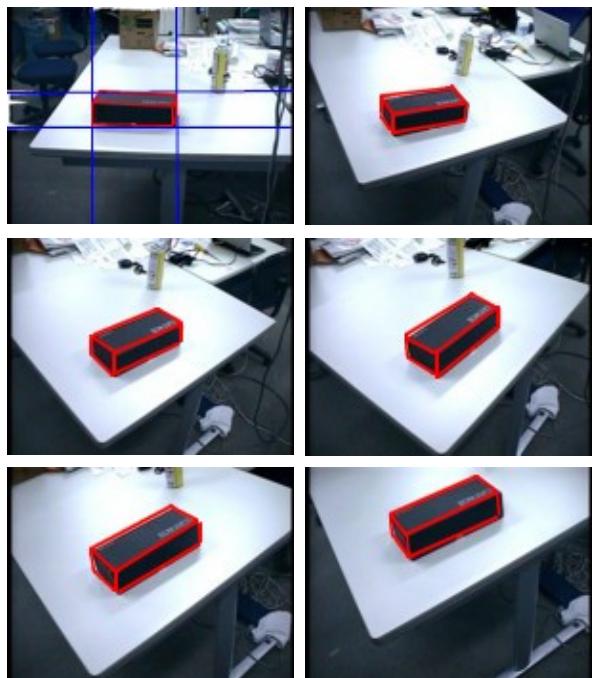


図 14: 対象物体 C の場合の結果

3.2 実験 2

本手法におけるカメラ位置姿勢の推定の精度が実際の AR に適用可能であることを示すため，OpenGL によってポットを実験環境中に重畳表示させる。

対象物体を図 11(a) としてカメラ位置姿勢を推定し，その結果を用いてポットを重畠させた。図 15 では対象物体である箱の上に重畠し，図 16 では，対象物体の箱と，奥に置いてある箱の上にポットを重畠した様子を示す。

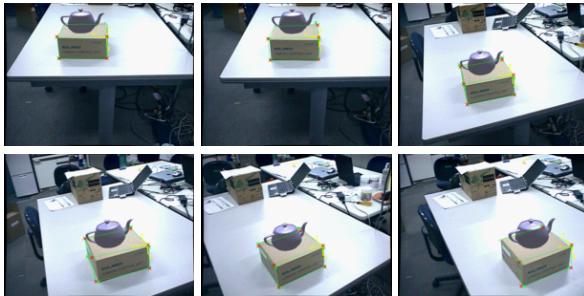


図 15: 実験 2 の結果 1



図 16: 実験 2 の結果 2

3.3 実験 3

対象物体を図 11(a) の箱として本手法の処理時間を測定する。初期フレームにおける処理速度の結果を表 1 に，また，2 フレーム目以降においては，ウィンドウサイズをいろいろと変更し処理時間を計測した。その結果を表 2 に示す。なお，本論文で使用した PC の構成は次のとおりである。

« PC 構成 »

- Intel Pentium4 Processor 2.8 GHz
- 1024 MB RAM
- 64 MB NVIDIA GeForce4 MX420 graphics card
- IO-DATA GV-BCTV/PCI capture card
- Redhat Linux 7.3

表 1: 初期フレームにおけるフレームレート

最小	最大
3.57 fps	3.73 fps

表 2: 2 フレーム目以降におけるウィンドウサイズとフレームレートの比較

ウィンドウサイズ	最小	最大
21 × 21 ピクセル	12.61 fps	12.76 fps
31 × 31 ピクセル	11.21 fps	12.32 fps
41 × 41 ピクセル	10.26 fps	10.26 fps
51 × 51 ピクセル	9.13 fps	9.13 fps
61 × 61 ピクセル	8.17 fps	8.17 fps

4 考察

4.1 従来手法との精度比較

従来のモデルベースにおける手法で一般的に用いられるテンプレートマッチング法では，カメラ位置姿勢を推定するのに複数のテンプレートが必要である。1 つの物体頂点について考えた場合でも，図 17 に示すように複数視点から撮影されたテンプレートが必要となる。また，これら全てとマッチングをとり，最も類似性の高いものを探し，位置姿勢を推定するため処理時間がかかる上，テンプレートとは大きく異なった角度から撮影された場合には正しい位置推定が行えない。

しかし，本手法では，カメラの視点の位置姿勢に不变である物体頂点を検出し，それを対応点として用いているため，常に安定した追跡が可能になる。図 17 に対して本手法に適用した場合でも，図 18 のように正しく物体頂点が検出されている。

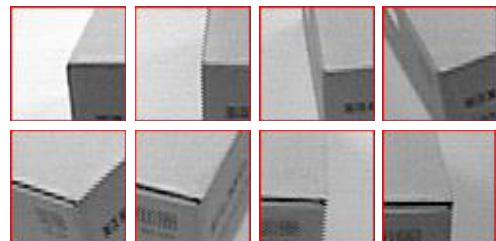


図 17: テンプレート

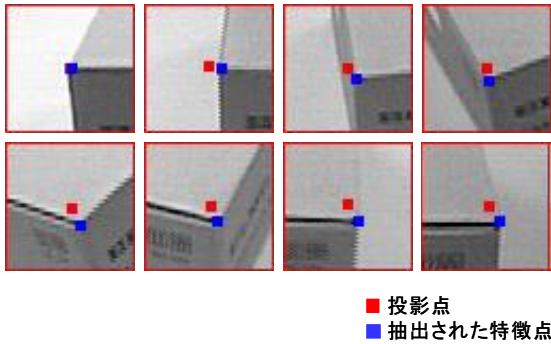


図 18: 図 17 に対して特徴点を検出した結果

この結果、図 12～14 に示したように、対象物体に対するカメラの視線方向が大きく変化し、見え方が大きく変わっても、安定したカメラ位置姿勢の推定が可能になっているわけである。

4.2 本手法における誤差の比較

手動で算出した実際の物体頂点座標（正解値とする）と、画像から抽出された物体頂点座標・本手法によって推定された外部パラメータを用いて投影した投影点との誤差をそれぞれ算出した。

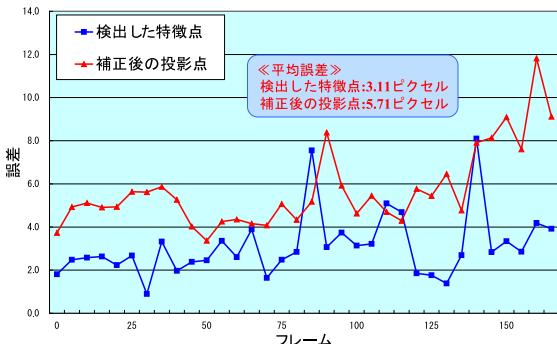


図 19: 誤差の比較

図 19 に示したように、正解値と画像処理によって抽出された物体頂点位置との誤差は平均 3.11 ピクセルと精度よく物体頂点が抽出できている様子が分かる。また、正解値と投影点の誤差は、最適化処理前では平均 15.16 ピクセルであったのに対し、最適化処理後は平均 5.71 ピクセルと外部パラメータが補正された様子が分かる。

5 結論

本論文では、物体の 3 次元モデルを用いて、物体の頂点を追跡することにより正確な位置合わせを行う手法を提案した。視点の位置姿勢の変化によって形状が変化しない物体頂点を利用することで、カメラの移動範囲に関する制約を取り除くことができ、磁気センサを併用することによってロバスト性を向上させることができた。また、実験により本手法の有用性を示すと共に、AR にも十分適用できるくらいの位置合わせの精度が得られていることを確認した。

今後は、物体頂点の検出精度の向上や、最適化処理の精度の向上を図るとともに、複数の物体を空間中に配置することでより広範囲に適用できるように改良することが重要である。

参考文献

- [1] Azuma, Ronald T “A Survey of Augmented Reality”. In Presence: Teleoperators and Virtual Environments, Vol.6, No.4, pp.355-385, 1997.
- [2] H.Tamura et al. “Mixed reality : Feature dreams seen at the border between real and virtual worlds”. IEEE Computer Graphics and Applications, Vol21, No.6, pp.64-70, 2001.
- [3] Hiroyuki Yamamoto “Case studies of producing mixed reality worlds”. Proc. IEEE SMC’99, vol.6, pp.42-47, 1999.
- [4] 佐藤清秀，内山晋二，田村秀行 “複合現実感における位置合わせ手法”. 日本バーチャルリアリティ学会論文誌, vol.8 no.2, pp.171-180, 2003.
- [5] 加藤博一，Mark Billinghurst, Ivan Poupyrev, 鉄谷信二，橋啓八郎 “拡張現実感技術を用いたタンジブルインターフェース”. 芸術科学会論文誌 , Vol.1, No.2, pp.97-104, 2002.
- [6] Gilles Simon and Marie-Odile Berger “Pose Estimation for Planar Structures”. IEEE Computer Graphics and Applications 2002.
- [7] Y.Genc, S. Riedel, F. Souvannavong, C. Akinlar, and N. Navab “Markerless tracking for augmented reality: A learning-based approach”. In Proc. International Symposium on Mixed Reality 2002.
- [8] 大江純子，佐藤智和，横屋直和 “画像特徴点によるランドマークデータベースを用いた動画像からのカメラ位置・姿勢推定”. PRMU Vol.104 No.572 pp.19-24 2004.
- [9] Vincent Lepetit, Luca Vacchetti, Daniel Thalmann, Pascal Fua “Fully Automated and Stable Registration for Augmented Reality Applications”. IEEE and ACM International Symposium on Mixed Reality and Augmented Reality (ISMAR2003)