

画像処理を用いた屋外 AR システムのための高精度定位

石川高志[†] 全へい東[‡]

[†]千葉大学大学院自然科学研究科 [‡]千葉大学総合メディア基盤センター

仮想環境と現実環境を継ぎ目なく融合する技術体系である拡張現実感技術 (AR: Augmented Reality) は、各方面で積極的な研究開発が行われている。その応用範囲は作業支援, 実環境への注釈提示, エデュテインメント (教育・娯楽) などと広範である。この研究では屋外 AR システムにおける現実環境と仮想環境の幾何学的ずれ (位置ずれ) の解消を目的としている。ずれの解消には画像センサを用いることが有効だが、屋外では照明条件等の変化により必ずしも安定した性能を得ることができない。そこでこの研究では利用者の位置・姿勢の検出に GPS, ジャイロセンサとビジョンを融合した手法を提案する。ビジョンを応用した画像センシングは精密な位置合わせには適しているが、探索空間が広いと計算時間がかかる。そこで GPS, ジャイロによって得た位置・姿勢データを探索の初期値として利用する。また画像特徴量には、屋外でも比較的安定して検出できる空と建物の境界線を用いた。事前に用意した建物の 3 次元モデル (3 次元地図) から推定した境界線と、実画像から得た境界線は DP マッチングから類似度を求め、正確な位置・姿勢を探索する。実験は学内を対象とした実環境で行い、手法の性能評価を行った。

Accurate Estimation of Pose and Location using Image Processing for Mobile Outdoor AR System

Takashi Ishikawa[†] Heitoh Zen[‡]

[†] Graduate School of Science and Technology, Chiba University

[‡] Institute of Media and Information Technology, Chiba University

Augmented Reality (AR) is a form of virtual reality that registers images of a virtual environment with the real world. In this paper, a hybrid sensing system for outdoor AR system is proposed, that enables accurate positioning and the pose estimation of the viewer. GPS and gyro sensing are used as the initial estimation. Then the vision (image sensing) is used for estimate the accurate position and pose. The boundary edges between the buildings and the sky is chosen as the image derived feature, for the fine search. Matching is done between the image derived boundary edges and the synthesized one from 3D maps. The similarity is defined as DP matching score. The evaluation of the proposed method is done through the experiments performed in the campus of Chiba University.

1. はじめに

各種センサの小型化が進み、携帯型情報端末上に組み込むことが可能になった。カメラ、GPS、ジャイロはその代表であり、これらのセンサから得たデータによって新たな機能やサービスの実現が一層進むことが予想される。

我々の研究グループでは、このような背景を念頭にして、建物などへの注釈提示を行う屋外拡張現実感技術を携帯型情報端末上で実現することを目指している。

拡張現実感技術 (AR: Augmented Reality) は、仮想環境と現実環境を視覚的に継ぎ目無く融合する技術体系であり、各方面で積極的な研究開発が行われている。その応用範囲は作業支援、実環境への注釈提示、エデュテインメント (教育・娯楽) など広範囲に及ぶ [1,2]。AR システムの課題は実環境と仮想環境の「ずれ」の解消である。「ずれ」には幾何学的 (位置) ずれ、光学的ずれ、時間的ずれ、の3種類が知られており、これらのずれを解消するための研究・開発は各方面で精力的に進められている。この研究で扱っているのは幾何学的ずれの解消であり、具体的には利用者の視点位置・姿勢を正確に計測することを目標としている。

従来の現実環境と仮想環境の位置合わせの手法には、磁気センサや超音波などを利用した手法があるが、屋外のように広い範囲で位置・姿勢を測定する目的では不適當である。またビジョンセンサを用いた多くの手法では人工的なマーク (fiducial) を事前に環境に設置しているが、この方法もまた屋外では不適當である。したがって屋外 AR を実現するためには「環境への作りこみ」が不要なビジョン (画像センシング) が望ましい [3] [4]。

この研究では、携帯端末に組み込み可能なセンサを利用することを前提に、位置検出に GPS を、また姿勢検出にジャイロセンサを用いる。いずれのセンサも小型化、省電力化が実現されており、携帯電話程度の機器用にすでに実用に供されている。GPS とジャイロセンサから得られる計測結果は計測原理に由来する誤

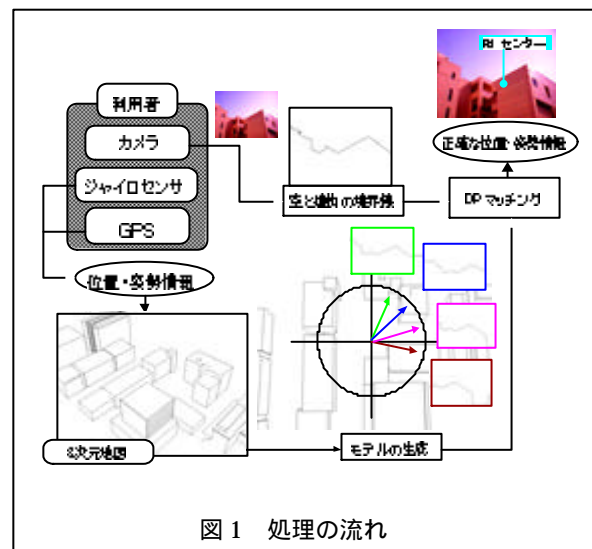
差が含まれており、建物などへの注釈に十分な精度が得られるとは限らない。

そこでこの研究では、GPS とジャイロから得た位置・姿勢を初期値とし、その後の詳細な推定には、画像から得られる特徴を用いる。画像特徴には屋外でも比較的安定して得られる建物と空の境界線 (建物のシルエット) を用いる。一方対象とする地域の 3 次元地図を事前に用意し、地図から生成した境界線と画像から得た境界線との間で DP マッチングを行う。初期値から開始したマッチングは DP によるスコアを類似度にした探索を行う。探索範囲は GPS やジャイロの精度により定義する。

千葉大学の構内対象にした実環境での実験を行い、手法の性能評価を行った。以下 2 で提案手法の詳細について述べ、3 で実験結果とその評価について記した。

2. 提案手法

図 1 に本手法の処理の全体像を示す。その概要を以下に述べる。



1. カメラより利用者の視点から見た、空を含む建物シーンの画像を取得する。同時に GPS およびジャイロセンサより、利用者の位置・姿勢情報を計測する。

2. 撮影画像から画像特徴量を抽出する。また、計測した位置・姿勢情報を基に、予め用意した 3 次元地図データから探索領域である一定の範囲・角度の視点位置・視線方向でのモデルを生成する。
3. DP マッチングによる撮影画像と探索領域内で生成した各モデルの相違度を算出し、相違度が最小となるモデルの視点位置・姿勢の値を、現実環境における利用者の正確な視点位置・姿勢情報と決定する。

2.1. 画像特徴量の抽出

屋外でビジョンセンサの利用に関して、安定して得られる画像特徴量を選定する必要がある。特殊な色や形状をしたマーカを人工的に配置してそれを追跡することも考えられるが、景観を損なうなど一般の環境では問題が多い。そこで環境中に既に存在する自然特徴を利用する方法が考えられる。屋外の画像シーン中に存在する自然特徴として、目立つ看板などをランドマークとして利用することも考えられるが、そのランドマークと光源・カメラの相対的な位置関係が変化することで画像シーン中での見え方が変化し、常に一定であるとは限らない。また建物表面の色やテクスチャなどの利用も、天候や日照の照明条件の変化、影の影響による見え方の変化を、事前のモデル化や解析で対応することは困難である。

そこで本研究では、画像シーン中の自然特徴として、空と建物の境界線に注目した。まず、利用者が屋外に位置しているのならば常に見えるもので、画像シーン中の存在領域は制限されない。また、照明条件の変化に際して、空と建物の境界は画像シーンの中でも比較的濃淡変化の差が大きく現れ、抽出の画像処理も容易に行える。そして、カメラとランドマークの位置関係の変化による見え方の変動にも、空と建物の境界線の幾何的特徴はその視点位置・視線方向により一意に決まるものであり、モデルとの照合によりその位置・視線方向の検出に有効的である。その抽出手法は以下の通りである。最初に、入力画像に対し Canny のエッ

ジ検出法を適用する。画像座標系における $y=0$ から y 軸の正方向に向かって走査して、画素値が初めて 0 (または 1) になる画素を見つけ、その次の画素からは値を 1 (または 0) と置き換える。すべての x について行った後、 y 軸方向の画素のギャップに結合処理を施した結果を、空領域との境界線とする。図 2 に求めた空と建物の境界線を示す。



図 2 空と建物の境界線

2.2. DP マッチングによる撮影画像とモデルの相違度算出

DP マッチングは、従来から音声処理、特に単語認識の分野で広く用いられている手法であり、2 つのパターンが与えられた場合、それらのパターンを非線形に伸縮し、それらのパターンの最適なマッチングを行う手法である。本手法では画像処理によって得られる、撮影画像の空と建物の境界線とモデルを対象として、DP マッチングによる相違度を算出する。

2.2.1. 線図形のリスト化

DP マッチング処理を行うにあたり、2.1 節で得られた境界線 (線図形) をリストによって表現する。その手法として区分的直線近似法を用いた。これは線図形を表す点列を再帰的に分割する次のような方法である。まず、点列の両端を結ぶ直線を求める。次にこの直線から最も離れた点を見つけ、その点で点列を 2 つに分割する。さらに、分割した各々の点列に対して、同様の操作を再帰的に繰り返し両端を結ぶ直線と点列までの最大距離がある決められた誤差以下になるまで分割を繰り返す。この方法により、任意の誤差範囲内で線図形を近似する線分列 (区分的直線) を得ることができる。

2.2.2 DP マッチング

撮影画像，推定モデルから得た線図形を，DP マッチングによって相違度を算出する．近似処理によって作成した線分列を，線図形のパターンとみなして図形間の対応付けを行う．即ち，作成した線分列を画像座標系 x 軸方向に非線形に伸縮し，線分列間の最適な対応付けを求めることにより，撮影画像の x 軸方向の任意の場所にある線分が，モデルでのどの場所に対応するかを求める．以下，DP マッチングのアルゴリズムについて述べる．

今，対応付けを行う二つの線分列を A, B とする．但し， $A = \{a_1, a_2, \dots, a_i\}$ ， $B = \{b_1, b_2, \dots, b_j\}$ である． I, J はそれぞれの線図形を直線近似して求められた線分の総数である．線分の始点の座標をそれぞれ (x_i, y_i) ， (x_j, y_j) として 2 つの線分の始点位置の距離，線分の長さの差，線分の角度の差を $d_1(a_i, b_j)$ ， $d_2(a_i, b_j)$ ， $d_3(a_i, b_j)$ とすると 2 つの線分間の相違度 $d(a_i, b_j)$ を次のように計算する．

$$d(a_i, b_j) = w_1 \cdot d_1(a_i, b_j) + w_2 \cdot d_2(a_i, b_j) + w_3 \cdot d_3(a_i, b_j) \quad (1)$$

$$d_1(a_i, b_j) = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (2)$$

$$d_2(a_i, b_j) = \sqrt{(x_{j+1} - x_j)^2 + (y_{j+1} - y_j)^2} - \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (3)$$

$$d_3(a_i, b_j) = \left| \tan^{-1} \left(\frac{y_{j+1} - y_j}{x_{j+1} - x_j} \right) - \tan^{-1} \left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right) \right| \quad (4)$$

ここで， w_1, w_2, w_3 はその重み係数である．そして，対応付けを行う 2 つの線分間で，前後に位置する 1 つの線分の対応を許すように伸縮をさせた場合，

$$D(a_i, b_j) = \min \left\{ \begin{array}{l} D(a_{i-1}, b_{j-1}) + 2d(a_i, b_j) \\ D(a_{i-1}, b_j) + d(a_i, b_j) \\ D(a_i, b_{j-1}) + d(a_i, b_j) \end{array} \right\} \quad (5)$$

となる．これは漸化式となるので初期条件を以下のよう設定する．

$$D(0,0) = 0$$

$$D(0,l) = \infty \quad (l > 0)$$

$$D(k,0) = \infty \quad (k > 0)$$

$$D(k,l) = \infty \quad (k < 0 \text{ もしくは } l < 0)$$

$$D(k,l) = \infty \quad (k = 0 \text{ もしくは } l = 0)$$

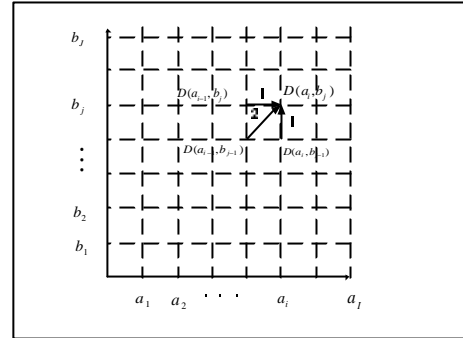


図 3 DP マッチングによる線分の対応付け

このように，順次 $D(a_i, b_j)$ を計算し， $D(a_i, b_j)$ を求めることで 2 つの線分列の相違度相違度を算出することができる．

2.3 視点位置・視線方向に基づく多次元空間内の探索

GPS やジャイロセンサによる視点位置・姿勢計測は何らかの誤差を含むもので，その計測値は必ずしも正確ではない．本研究では，GPS，ジャイロセンサによって得られる視点位置・姿勢計測を大域的な検出結果として捉え，その情報を基に探索空間を設定し，精密な視点位置・姿勢を推定する．その探索空間について，基本的に利用者の視点位置・姿勢は 6 つの自由度（3 次元位置座標値 (x,y,z) と各軸周りの回転のパン角，チルト角，ロール角）で決まる．いま，本手法での視点の高さはヒトの高さを想定しているので，その高さの変化は高々 1[m] 未満である．この視点位置の高さの変化によって，遠方に位置する建物を注視した場合シーン中に大きな変化は生まれない．そこで，3 次元位置座標の高さ方向の変化に関しては本研究では考慮しないことにする．また，姿勢のロール角についても建物シーンを撮影する際の大きな角度変化はないものとし

て、探索空間として本研究は、視点位置の2次元水平面での移動、姿勢のパン、チルト角を取り扱うものとする。探索空間内で視点位置・姿勢を変化させながら逐次生成したモデルと撮影画像に対してマッチング処理を施し、最小となる相違度の視点位置・姿勢を求める。そこで、その探索空間の領域設定、また、その領域内での視点位置・姿勢を如何に変化させるかが問題となる。例えば、姿勢変化をパン角+30[deg] ~ -30[deg]、チルト角+10[deg] ~ -10[deg]まで1[deg]間隔で、視点位置の変化を探索領域の半径10[m]、1[m]間隔の移動を仮定したときの、マッチング処理の計算量は、406077となる。実際に計算機（CPU：3.4[GHz]）で処理してみた結果、8498[sec]（2[h]）であった。このように全探索の手法は最も確実だが、計算コストにかかる負担が大きい。そこで、視点位置の2次元水平面での座標値を x, y 、姿勢のパン角を θ 、チルト角を ϕ とした変数として、撮影画像と推定モデルの相違度を求める評価関数に対して、多次元変数の最適化問題として捉え、最適化手法の一つである Nelder-Mead の滑降シンプレックスアルゴリズムを適用して解を求める。

2.3.1 Nelder-Mead の滑降シンプレックス法

最適化すべき変数の数を N とすると、 N 次元空間の極値を求める問題となる。シンプレックスと呼ばれる、 N 次元で $N+1$ 個の頂点とそれらを結ぶ辺、面からなる多面体の基本図形を考え、その頂点の関数値をもとにシンプレックスを移動、変形させながら最適値を求める手法である。この手法は導関数を用いずに、関数値だけで評価する。そのシンプレックスの更新ステップを以下に示す。

0. まず $N+1$ 個の初期値を与える。
1. まずシンプレックスの頂点についてループし、関数値の最悪、最悪の次、最良の点を調べる。関数値の最大の比を求めこれが十分に小さいのなら終了。そうでなければ2へ進む。
2. 最悪の点を残りの点の重心の反対側に対象移動す

る。現在の最良の点より良くなったならば、更に2倍対称移動させ、1に戻る。最良の点より良くならなければ3へ進む。

3. 対称移動した点が最悪のままならば1次元の収縮を試みる。点の関数値が最良の点よりも良くなれば、1へ戻る。それでも良くならなければ4へ進む。
4. 全体の点を最良点に向かって収縮させる。1に戻る。

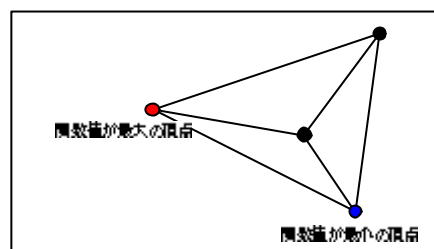


図4 シンプレックス ($N=3$ の例)

2.3.2 複数の撮影画像による拘束条件を利用した探索

また、場合として線図形が1本の直線で現れるときなど、探索結果として視点位置・姿勢が異なるにもかかわらず複数地点で相違度に差が表れず、推定値が一意に決められないことがある。そのような曖昧な推定結果にも対応する必要がある。

実際の正しい視点位置・姿勢からそのシーンを見ているのならば、その視点位置から別方向におけるシーンを見ても、探索において動いた姿勢成分のみ変化させれば同じ視点位置で正しいマッチングが得られるはずである。そこで、マッチングの拘束条件として、別方向の撮影画像を複数枚利用し、曖昧なマッチング結果に対応する。

撮影画像の画像処理によって得られる線図形の線分数が、ある閾値より小さい場合、同じ位置の別の姿勢で撮影した画像を複数枚用意する。最初の撮影画像に対して推定位置の探索処理を行った後、そこから求められた推定位置を、次に使用する異なる姿勢での撮影画像の探索における出発位置とすると同時に、GPS から得た位置からの移動量を2次元平面上の探索領域の

半径として、探索処理を行う。その同様の手順を撮影画像枚数分繰り返すことで、視点の推定結果の精度を向上させる。

3. 実験

提案手法である GPS、ジャイロセンサとビジョンセンサを組み合わせたハイブリッド手法による利用者の視点位置・姿勢推定の有効性の確認とその評価実験を本学内において行った。以下、3.1 節では機器の構成について、3.2 節では提案手法による位置合わせ結果について述べる。

3.1 機器の構成

まず、撮影画像とその姿勢情報を獲得するために使用した機器についての詳細を述べる。

カメラ (IBM : Ultra Port)

フレームレート : 30[fps]

解像度 : 320×240[pixels]

利用者の視点近傍から現実環境の建物が写りこんだシーンを撮影する。カメラのレンズをジャイロセンサの中心に、また、光軸をロール軸に一致するように設置する。

ジャイロセンサ (TOKIMEC : TMOS2000F)

ヨーレート : 0.03[deg/sec]

ピッチ角の誤差 : 0.05[deg]

データ更新間隔 : 10[ms]

カメラを上部に設置し、その視線方向(パン、チルト、ロール)を計測する。初期状態において、環境中の任意の位置に配置したマーカにカメラの光軸を向け、初期角度を決定しておく。

PC (SOLIDRAY)

CPU 周波数 : 3.4[GHz]

搭載メモリ : 2[GB]

カメラ、ジャイロセンサの各センサからそれぞれ情報を取得し、内部に保持した 3 次元地図情報よりモデルを生成し、撮影画像とのマッチング処理を行う。

本実験では、GPS の使用は想定としている。そこで探索範囲としては、都市部に含まれる測位の誤差を考慮する必要がある。そこで、都市部でも利用可能な測位精度として、gpsOne の測位精度を例として、GPS から得られる視点位置の最大誤差を半径 50[m]と想定し[5,6,7]、実験を行うものとする。

3.2 位置合わせ実験

提案手法を用いて、利用者が任意の位置から撮影した画像から、その視点位置・姿勢に等しい推定モデルが取得可能かを検証する実験を行った。図 5 は実環境中での撮影位置を原点として、2 次元平面上の探索範囲内でランダムに選び出した位置を出発点としてその位置から探索を行い、最終結果として得られた推定値と、そのときの推定モデルを示す。また、そのときの撮影画像と推定値の誤差を表 1 に、2 次元平面上での探索した視点位置の分布を図 6 に示す。いずれも GPS の誤差を想定した開始点 (図中 $(x, y) = (-20, -27)$) から、真値へ向けて、高い精度で位置合わせ(推定値 $(x, y) = (1.044, 0.512)$)ができていることが確認できる。実際にその撮影画像に対して推定モデルのワイヤフレームを重畳表示したものが図 7 である。また、探索の開始点を真値から一定の範囲 (0 ~ 10[m] , 10 ~ 20[m] , 20 ~ 30[m]) で開始点を複数与えた場合の、そこからの各範囲での推定結果の分布を図 8 に示す。

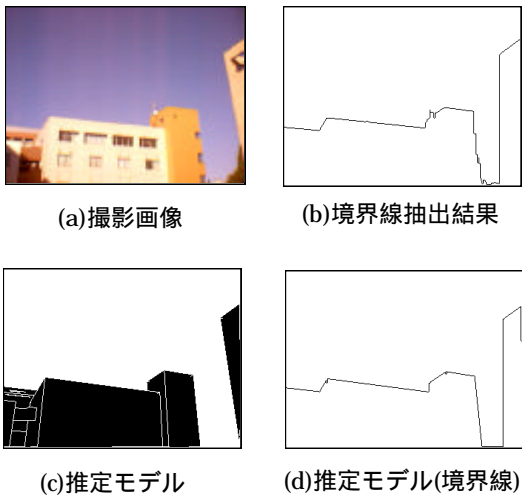


図5 撮影画像と推定モデル

表1 撮影画像と推定モデルの誤差

	視点位置 (x[m])	視点位置 (y[m])	パン角 ([deg])	チルト角 ([deg])
真値	0	0	22.80	23.73
推定値	-1.044	0.512	18.120	23.633
誤差	1.044	0.512	4.68	0.097

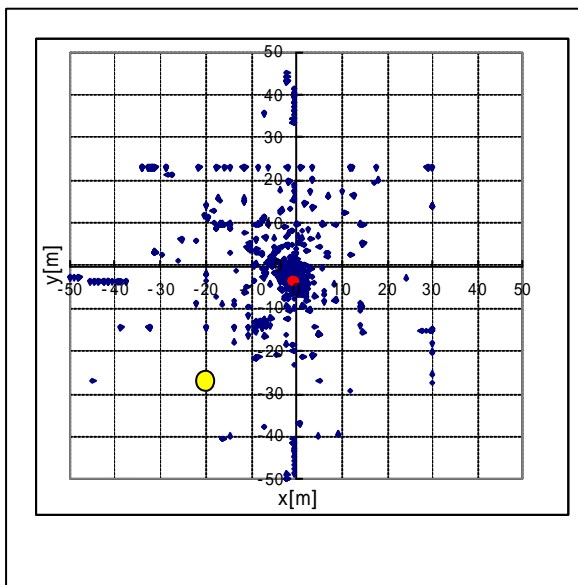


図6 探索した視点位置の分布



図7 ワイヤフレームによる重畳表示

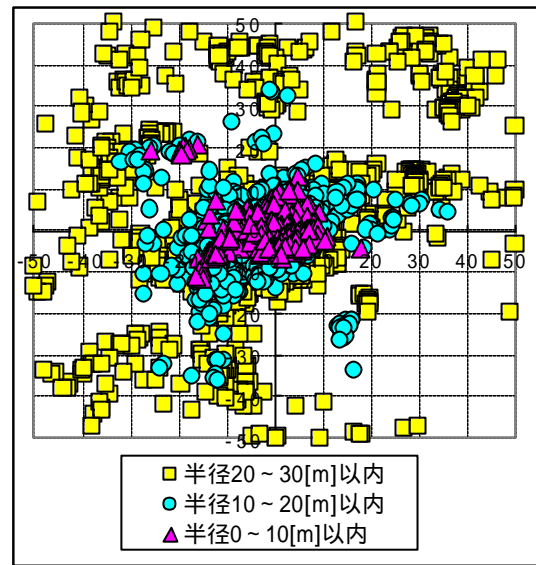


図8 開始点の各範囲からの推定結果

次に、曖昧な推定結果を持つ撮影画像に対する対応として、複数の異なる姿勢で撮影した画像による拘束条件を利用して、正しい位置を推定する手法について有効性を検証する実験を行った。まず、曖昧な推定結果の例を図9に示す。この結果は直線近似により生成された線図形が1本の線分しか現れない。撮影画像と推定モデルの間には当然低い相違度となるが、視点位置が変化しても同様な画像シーンは複数存在するため、推定値に曖昧さが含まれる結果である。そのときの相違度の分布を図10に示す。同じ視点位置から異なる姿勢で撮影した画像3枚に対して、撮影画像毎に得られる推定値を次に用いる撮影画像の探索の出発点として、推定値のずれの量を探索範囲に適用することにより位置を推定した。そのときの推定位置の軌跡を図

に示す．図 11 では上から探索に使用した画像の順序と，それぞれの推定値の軌跡を示す．その撮影画像の異なる姿勢による撮影画像の使用順序の入れ替えによる最終的な推定値の変化は，順序の違いに関わらず正しい視点位置に向かって約半径 3[m]の範囲に近づいたことが確認できた．

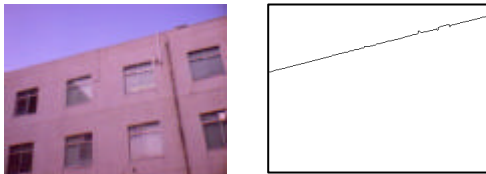


図 9(a) 撮影画像 図 9(b) 境界線

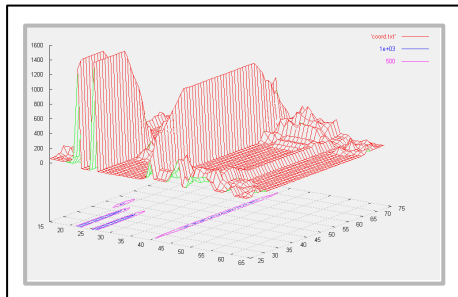


図 10 相違度の分布

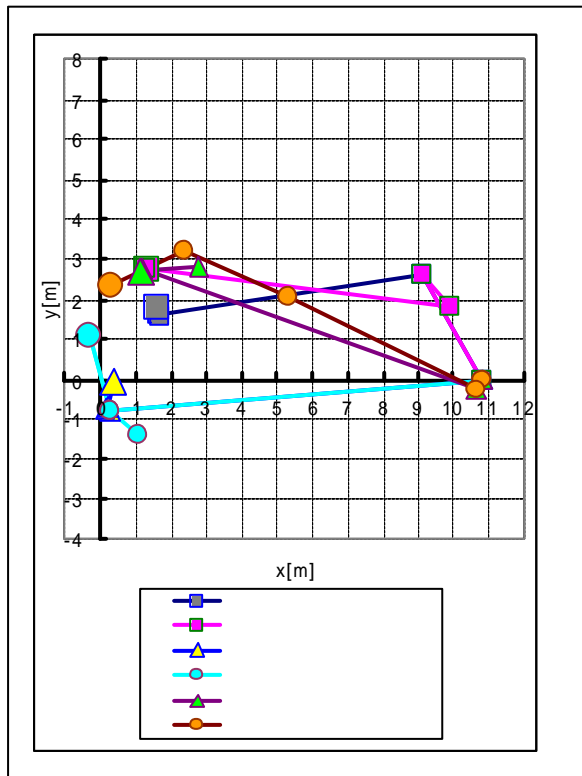


図 11 推定結果の軌跡

4. まとめ

ビジョンセンサと位置姿勢センサのハイブリッドな位置合わせ手法を提案した。GPS，ジャイロセンサで得られた計測結果である大域的な利用者の視点位置・姿勢情報を，撮影画像とその探索領域内で生成される推定モデルとの DP マッチングによる相違度を算出することによって物体の注釈提示に可能な視点位置・姿勢情報を正確に推定した．実環境での実験を通じ，有効性を評価した。屋外 AR システムとしてのアノテーションへの適用可能性を確認した。今後は実験規模の拡大と，マッチング処理の高速化を検討する予定である．

参考文献

- [1] 田村，大田：“複合現実感”，映像情報メディア学会誌，Vol.52，No.3，pp.266-272，1998.
- [2] R. Azuma：“A survey of augmented reality”，Presence，Vol.6，No.4，pp.355-385，1997.
- [3] D.Koller et al.：“Real-time Vision-Based Camera Tracking for Augmented Reality Applications”，Proc. VRST '97，pp.87-94，1997.
- [4] G.Simon et al.：“Markerless Tracking using Planar Structures in the Scene”，Proc. ISMAR2000，pp.120-128，2000.
- [5] N. F. Krasner: Position determination using hybrid gps/cellphone ranging. Inst. of Navigation Conf. GPS 2002，2002.
- [6] 安田：“GPS技術の展望”，信学技報 B Vol. J84-B，No.12，pp.2082-2091，November，2001.
- [7] 天野，柳原，初本：“さまざまな次世代GPS測位方式”，IP SJ Magazine，Vol.43，No.8，pp827-835，August，20

この研究は平成16年度科研費・基盤研究(B)「センサ統合による高精度定位装置の開発」(課題番号15300053)によった。