2 基礎 2：位置合わせ技術

植松裕子
慶應義塾大学

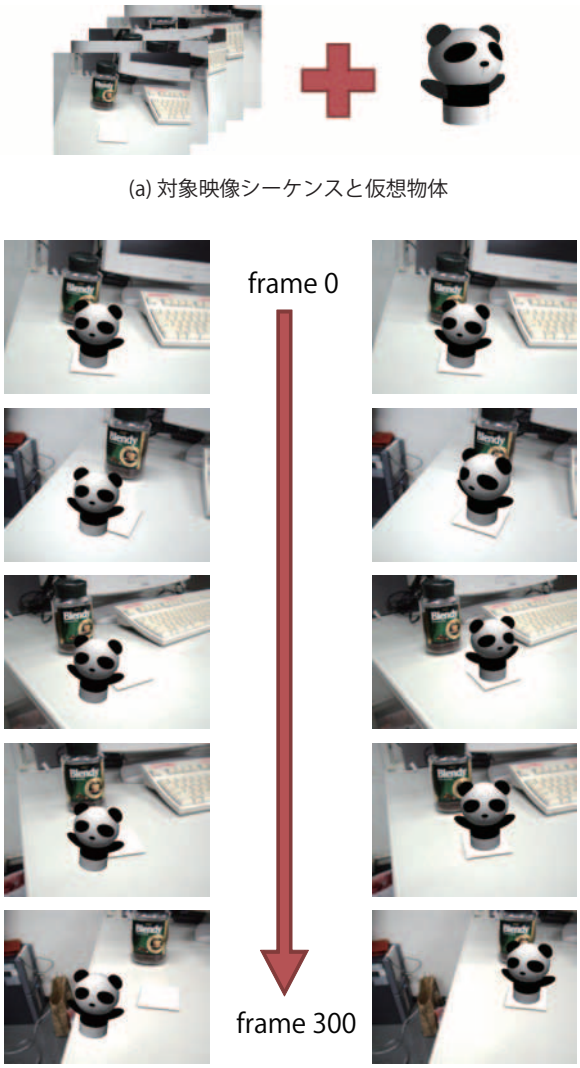
AR における幾何学的位置合わせ

AR（Augmented Reality）技術を利用すると，現実世界の中に仮想世界が溶け込んだような，従来のバーチャルリアリティとは異なった新しい世界を表現することができる．このような表現は，現実世界をカメラで撮影した映像の上に，CG などの仮想物体を重ねて表示することで実現する．

AR では，携帯デバイスのような手持ちのカメラを用いたり，HMD（Head Mounted Display）などに搭載されたカメラを頭や目の位置に装着して使用することが想定されるため，基本的にカメラが動いていることが前提である．そのため，動いているカメラで撮影された映像に対して仮想物体を重畳表示するには，映像中のどの位置へ仮想物体を配置すればよいかを考える必要がある．つまり，映像中の各フレームにおいて，そのときのカメラの動き（位置姿勢）に合わせて仮想物体を重畳する位置を決定しなくてはならない．この問題は，AR における“幾何学的位置合わせ問題”と呼ばれている．

たとえば，図 -1（a）のような映像シーケンスに対して，パンダの CG 仮想物体を重畳表示するとき，カメラの動きをまったく考慮せずに重ね合わせた場合と，カメラの動きに応じて位置合わせを行った場合の結果を図 -1（b）に示す．左列では，カメラが動いて視点が変化しても，CG の仮想物体は映像中の同じ位置姿勢に 2 次元的に表示されているだけである．そのため，仮想物体が現実物体と同様には見えず，現実世界と仮想世界が融合されているように見せることはできない．一方で右列では，カメラの動きに従って仮想物体の位置姿勢が 3 次元的に変化しており，実際に机の上にパンダが存在しているかのように見える．したがって，仮想物体を違和感なく現実世界に溶け込ませるためには，幾何学的な位置合わせが必要である．

幾何学的位置合わせ問題を解決するためには，カメラの位置と姿勢を毎フレーム取得・追跡することが必要になる．ここでのカメラの位置姿勢とは，現実世界に対して任意の場所に 3 次元座標系を設定したときに，その座標系に対してカメラがどんな位置姿勢にあるかという相対的な位置関係を表している．一般に，カメラの位置姿勢は，図 -2 に示すように，姿勢にあたる回転行列 $\boldsymbol{R}$ と位置にあたる並進ベクトル（物体の並進運動の大きさ・方向を表現するベクトル）$\boldsymbol{t}$ で表され，回転に 3 自由度，



(a) 対象映像シーケンスと仮想物体

(b) 左列：位置合わせなしの重畳表示結果
右列：位置合わせをした重畳表示結果

図 -1　カメラ動きの考慮（位置合わせ問題）の有無による仮想物体の重畳表示結果

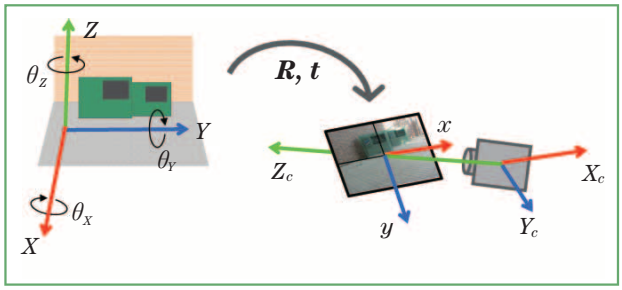


図 -2　カメラ位置姿勢を表す座標系の設定

並進に 3 自由度で，合計 6 自由度のパラメータを求めることが課題となる．

数学的に表してみると，以下のようになる．

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \simeq \boldsymbol{A}[\,\boldsymbol{R} \mid \boldsymbol{t}\,] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

$$\boldsymbol{R} = \begin{bmatrix} \cos\theta_x \cos\theta_y & \sin\theta_x \cos\theta_y & -\sin\theta_y \\ \begin{matrix}\cos\theta_x \sin\theta_y \sin\theta_z \\ -\sin\theta_x \cos\theta_z\end{matrix} & \begin{matrix}\sin\theta_x \sin\theta_y \sin\theta_z \\ +\cos\theta_x \cos\theta_z\end{matrix} & \cos\theta_y \sin\theta_z \\ \begin{matrix}\cos\theta_x \sin\theta_y \cos\theta_z \\ +\sin\theta_x \sin\theta_z\end{matrix} & \begin{matrix}\sin\theta_x \sin\theta_y \cos\theta_z \\ -\cos\theta_x \sin\theta_z\end{matrix} & \cos\theta_y \cos\theta_z \end{bmatrix} \tag{2}$$

$$\boldsymbol{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{3}$$

この式 (1) は，カメラが 3 次元を 2 次元に投影するという考え方から，現実世界の座標系（X, Y, Z）が画像の座標系 (x, y) へと投影されていることを表している．このとき，$\boldsymbol{A}$ は焦点距離やレンズの特徴を表すパラメータであり，$\boldsymbol{R}$ と $\boldsymbol{t}$ がカメラの姿勢（回転）と位置（並進）を表している．回転については，計算の便宜上 3 × 3 の行列 $\boldsymbol{R}$ で表されているが，内部の自由度は θ_x, θ_y, θ_z の 3 つであり，並進 $\boldsymbol{t}$ の t_x, t_y, t_z と合わせて自由度 6 となる．

カメラの 6 自由度を求める手法についてはこれまでにも非常に多くの研究がなされているが，それらを大きく分類すると，カメラや端末などにセンサを取り付けて，その動きを直接取得する“センサベースの手法”と，撮影された映像の中に映っている点や線などの特徴物の動きからカメラの動きを取得する“ビジョンベースの手法”に分けられる．

これらのアプローチは，使用環境や目的などによって使い分けられるが，特に高い位置合わせ精度を要求されるような状況では，画像を利用したビジョンベースの手法が多く用いられる．なぜなら，センサを用いた AR は照明の変化や未知の環境にも対応しやすく処理速度も高速である反面，仮想世界と現実世界の座標系をぴったりと合わせるのに足るような正確さを得ることは難しいからである．しかしながら，夜間や天候の変化などでカメラからの入力が適切に得られないような場面などでは，センサが効果的に用いられている．

本稿では，ビジョンベースの位置合わせ手法を軸として，AR における位置合わせ手法について解説していく．ビジョンベースの手法では，さらにアプローチごとに細かく分けて述べていく．また，センサベースの手法に関しては，各種センサによるアプローチについてや，近年注目されている携帯電話端末を利用したアプリケーションの仕組み，ビジョンとセンサの両者の利点を組み合わせることで精度と安定性を両立した“ハイブリッドな手法”について解説する．

ビジョンベース位置合わせ手法

前述の通り，AR における位置合わせでは，カメラから撮影される画像に対して仮想物体を重畳表示する際に，カメラの視点の動きに合わせて 3 次元的な位置と姿勢を求める．そして，そのパラメータに合わせて仮想物体の位置姿勢を変化させて重畳表示する．

ビジョンベースの手法では，撮影された画像の中に写っている画像特徴を利用して，カメラの位置姿勢を求める．一般に用いられる特徴には，点や線，平面構造などがある．計算でそのパラメータを求めるためには，点や線などの現実世界における 3 次元座標と，撮影された画像における 2 次元座標との対応関係を得ることが必要である．つまり，式 (1) で示した (X, Y, Z) と (x, y) の対応を何組か取得することで，式 (2)，(3) に示した回転と並進のパラメータを算出する．なお，このパラメータはカメラの動きに従って変化するため，固定カメラを用いない限りは，入力となる画像列のそれぞれのフレームにおいて算出する．そのため，毎フレーム安定に画像から特徴を抽出する必要があり，どのような特徴を用いるかによって精度や適した使用環境が決まる．また，ビジョンベースの特徴では，常に利用する特徴がカメラから見えている必要があるが，1 台のカメラを用いるだけではその動く範囲が限られてしまったり，他の物体などによって特徴が隠れてしまったりする状況も起こるため，カメラを複数台組み合わせて用いることで，隠れに対応したり精度を向上させるアプローチも存在する．

利用する特徴ごとにビジョンベースの手法を大別すると，人工的に特徴を作り出してその場に配置する“マーカベースの手法”，対象物体の 3 次元モデルを参照デー

タとして用いる“モデルベースの手法”，および，現実世界に自然に存在している特徴を用いる“自然特徴ベースの手法”のようになる．マーカとは，画像から点や線などの特徴が検出しやすいような記号やテクスチャパターンのことを主に指す．マーカを現実世界に配置すると，何も利用しない場合に比べて，入力画像中からその特徴を安定に検出できる確率が高く，さまざまなアプリケーションに簡単に応用しやすい．しかし，その都度マーカを配置する必要があったり，マーカそのものが現実環境に違和感を与えることもある．

モデルベースの手法では，対象とする物体の 3 次元モデル（CAD データなど）をあらかじめ保持しておくことで，撮影された画像中での見え方とモデルとのマッチングを行う．主に直線成分などで構成されている単純な形状の物体では，エッジ同士のマッチングなどによって高速に計算をすることも可能である．しかし，一般のユーザが対象物体の正確な 3 次元モデルをあらかじめ用意できるという前提は難しいため，特定の製品や環境に対して構築されたシステムなどに適していると言える．

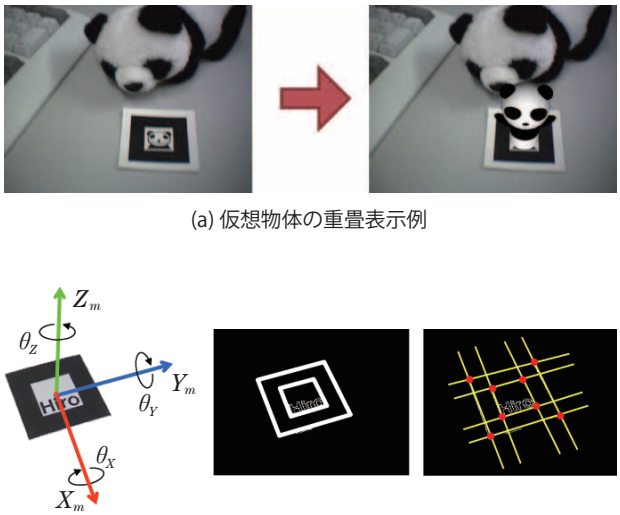
現実環境に手を加えることのない自然特徴を用いた手法では，実際に存在している物体の特徴を画像から抽出して利用するだけなので，視覚的にもより自然な AR を実現することができる．その分，自然特徴を安定に抽出・追跡するための技術が必要不可欠になり，照明変化などにも柔軟に対応することが求められる．画像処理の分野ではすでにさまざまな自然特徴抽出・追跡手法が提案されているが，それぞれに処理速度やカメラ移動範囲の制限などがあるため，目的に応じて選択する必要もある．

このように，それぞれ利用する特徴によって適した状況・環境があるため，次節以降では，各アプローチについて代表的な研究例を挙げながら解説する．

--- マーカベース位置合わせ手法 ---

AR におけるマーカベースの位置合わせ手法で最も有名な手法は，ARToolkit である[1]．図 -3（a）のような 2 値（白黒）の矩形マーカを用意し，現実世界の任意の場所に配置するだけで，簡単に 3 次元の仮想物体を重畳表示することが可能になる．

位置合わせの計算の際には，図 -3（b）のようにマーカ平面に対して 3 次元座標系を定義し，その座標系に対するカメラの位置と姿勢を表すパラメータを求める．ARToolkit では，マーカの矩形サイズおよび内部のパターンが既知であるため，撮影された画像から直線を検出し，それらの交点が矩形の頂点になると考える．そして，マーカが平面であることを利用して，回転および並進のパラメータを算出する．このような白黒のコントラストが強いマーカを使用することで，画像中から点や線の特徴を安定に検出できるので，多少の照明変化やカメラの急激な移動にも対応しやすい．そのため，非常に多くのアプリケーションが開発されており，応用の幅が広がっている．



(a) 仮想物体の重畳表示例

(b) マーカ上に定義される 3 次元座標系とマーカ検出の概念

図 -3　ARToolkit における座標系と仮想物体重畳の様子

一方で，このような白黒のマーカではなく，もっと現実世界に溶け込んだスタイルで見栄えや美観を重視したマーカを作ろうという研究も行われている．立命館大学のグループでは，図 -4 のような対象領域と同様の色相を持ったツートンカラー方式というマーカを開発し，現実環境に溶け込んで美観を損ねないマーカベースの位置合わせ手法を提案している[2]．また，目立ちにくい再帰性反射材を利用してマーカを作成し，赤外線 LED 付きの赤外線カメラで撮影して認識する手法も中里らによって提案されている[3]．

さらに，一見するとマーカとは分からないようなパターンをマーカ代わりに利用するアプローチもある．天目らは，自然なパターンとしてポスターを選び，ある一定のルールを設けて作成されたポスターをマーカ代わりに利用することで，従来の白黒マーカよりも景観を重視した手法を提案した[4]．また小野らは，任意の画像の高周波領域にマーカとなる信号を埋め込み，撮影された画像を 2 次元フーリエ変換することでマーカ情報を抽出する手法を提案し，人目に付きにくいマーカを開発している[5]．一般的な画像では，高周波成分が人間の目に知覚されにくいという特性を利用しており，事前に画像のテクスチャを必要としないため，利便性も高い．

これらのマーカも ARToolkit 同様に，特徴がすべて 1 枚の平面上に存在するという条件を利用している．これは，式（1）で示した 3 次元と 2 次元の対応関係からパ

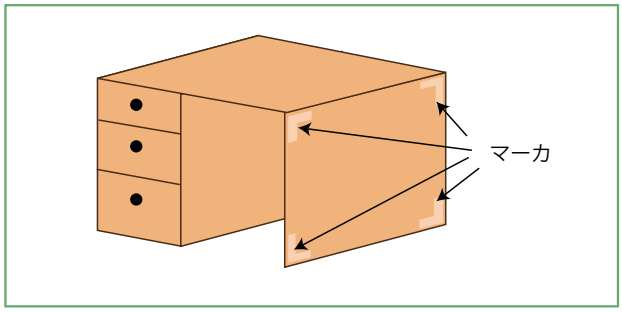


図 -4　美観を考慮したマーカの例

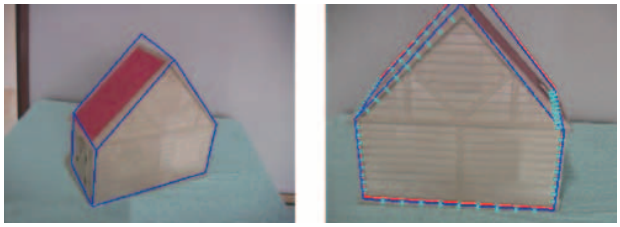


図 -5　モデルベースによる位置合わせの様子

ラメータを求める際に，それらの対応点がすべて平面上に存在しているという拘束条件を利用すると，計算式を単純に解くことができるからである．また，一般のユーザが使用する際にも，“紙に印刷するだけ”といった手軽さは非常に重要な要素であるため，平面形状のマーカが多数用いられている．

--- モデルベース位置合わせ手法 ---

モデルベースの手法では，マーカのように新たな特徴を付加するのではなく，位置合わせに利用したい物体の 3 次元形状を用意しておき，その 3 次元モデルと実際に入力画像に写った対象物体の見た目が一致するようにパラメータを推定する．つまり，3 次元モデルが分かっていると，“このパラメータに相当するカメラ位置姿勢で対象物体を撮影すると，このように画像に写るはずである”ということが分かるため，その予測されたモデルの位置姿勢と，入力画像中の物体の見た目（エッジなど）が一致するように，パラメータを最適化していく．図 -5 では，あらかじめ家の 3 次元モデルが分かっているため，予測したパラメータの位置姿勢にカメラがあった場合に，家のモデルが画像に写るであろう場所に青のワイヤフレームモデルを投影している．この青の直線と，画像中に写った家との差分を小さくするように，パラメータを最適化することになる．このアプローチでは，あらかじめデータベース中にモデル情報を保持しておくのだが，その際に対象物体の周囲にもともと存在している点や線などの情報も同時に保存しておくと，予測と入力画像との差分を最適化する際に，より多くの情報を用いることができるようになる．

Lepetit らの手法では，対象の 3 次元 CAD モデルを利用し，オフラインの学習フェーズにおいて対象物体周辺の自然特徴点をデータベースに登録する[6]．その後，オンラインで撮影された画像から抽出した自然特徴点とデータベース中のモデルおよび特徴点とのマッチングを取ることによって，対象物体に対するカメラの位置姿勢を推定し，正確な位置合わせを行う．また Kotake らは，プリンタの 3 次元モデルを利用して，そのモデルと撮影画像との位置合わせを初期化からすべて自動かつオンラインで行う手法を提案した[7]．この手法は，後述するセンサベースの手法とビジョンベースの手法とを併用するハイブリッドなアプローチであり，先にセンサから角度情報を得ることで，抽出した線分と 3 次元モデル上の線分との対応を効率的に取ることができるため，リアルタイムの AR 表示が可能である．

このように，対象とする物体および目的が特定されていて，その 3 次元モデルなどが手に入りやすい状況では，このアプローチは非常に完成度の高いアプリケーションになり得る．

--- 自然特徴ベース位置合わせ手法 ---

自然特徴とは，現実世界に存在する物体がカメラで撮影されたときに，画像中から抽出できる点や線などのことを主に指す．マーカなどのように現実世界に手を加える必要がなく，現実世界と仮想世界をより自然に融合することが可能になる．何の変哲もない現実空間に突然 CG の仮想物体が現れるようなデモンストレーションは，特に一般ユーザに対するインパクトが高い．

画像中から自然特徴を抽出・追跡するための研究も，以前より多数行われている．自然特徴点抽出のための Harris オペレータや，追跡のための KLT（Kanade Lucas Tomasi）Tracker などは非常に有名なアルゴリズムであり，開発から 10 年以上経った現在でも頻繁に用いられている．しかしながら KLT の追跡では，画像中の局所領域において微小時間経過後の特徴点の動きが一定であるという条件があり，動画像中の前後フレーム間で照明などの変化によって輝度変化が激しい場合や，カメラの動きに回転やスケール変化がある場合には，特徴点追跡がうまくいかない．

そこで近年注目されているのは，Lowe によって提案された SIFT（Scale Invariant Feature Transform）アルゴリズム[8]や，それを拡張した SURF（Speeded Up Robust Features）アルゴリズムのように，フレーム間での回転やスケール変化に対しても安定な特徴点追跡手法を用いたカメラ追跡である．従来のアルゴリズムと比較しても，

図 -6　SIFT アルゴリズムによる自然特徴点抽出・追跡

オクルージョン（隠れ）や環境の変化にも強い傾向にあり，図 -6 のようにカメラ位置姿勢が大きく変化したような場合でも特徴点を追跡できるため，自由にカメラを動かしたい AR の要求にも適している．SIFT はスケール変化に対応するための計算コストが高いため，リアルタイム処理には向かず，それを高速化した SURF は SIFT と比較するとマッチング精度は落ちるといった問題もあるが，これらの基本的なアルゴリズムを拡張した AR のための位置合わせ手法が多く提案されている．

自然特徴を利用した AR で近年最も注目を集めているのは，Klein らによって開発された PTAM（Parallel Tracking And Mapping）である [9]．これまでにロボットの自動ナビゲーションなどの分野で多く研究されてきた，カメラの自己位置推定と環境の 3 次元構造認識を同時に行う “SLAM（Simultaneous Localization And Mapping）” と呼ばれる技術を応用したものである．カメラ入力から特徴点を取得・追跡することで，対象空間の 3 次元マップを構築するのと同時に，現在のカメラ位置姿勢の算出をリアルタイムで行う．事前にマーカを配置したり，特徴点をデータベース化しておくなどの準備は必要なく，始めにカメラを数 cm 平行に動かして，基準となる平面を検出すると，すぐさま仮想物体の重畳表示が始まる．つまり，未知のシーンを対象としても，その空間の 3 次元モデルをリアルタイムで作ることができ，位置合わせに利用できる．実装面でも，カメラの追跡部と 3 次元マップの構築部とを並列に走らせることで高速に動作可能であり，数千個の特徴点をリアルタイムで扱えるように最適化されているため，精度も安定性も抜群である．現在では，特徴点だけでなくエッジも考慮することでカメラの高速な動きに対応させている．また，さらに処理を最適化して iPhone へも実装済みであり，これからの AR の実用化に大きな影響を与えるといえる．

センサベース位置合わせ手法

ビジョンベースの手法が，カメラからの入力画像をもとにカメラ位置姿勢を計算するのに対して，センサベースの手法は，カメラそのものや対象環境中にセンサを搭載し，直接その動きや構造を計測するアプローチである．対象とする環境によって，磁気センサやジャイロセンサ，赤外線センサ，GPS などのセンサを使い分けることが重要であり，またビジョンベースの手法との併用も効果的である．

たとえば，屋外などの照明環境が変化しやすい環境であったり，非常に広い範囲での AR を想定すると，ビジョンベースの手法では特徴となる点や線を安定して取得することが難しいため，センサベースの手法が適している．最近話題となっているセカイカメラも，屋外でのナビゲーションという非常に広い範囲が目的であるため，複数のセンサを利用した AR が行われている．ここでは，画像処理技術は使わずに，GPS によって現在地を特定し，端末に搭載されたセンサ（電子コンパス）によってカメラの向きを認識して，その方向に存在する建物などのタグを付加する．このように，画面上での数 pixel のずれや揺らぎが問題にならないようなアプリケーションでは，処理速度や動作の安定性からもセンサのみによる AR が適しており，今後もカメラ付き携帯電話のアプリケーションでの発展が大いに期待できる．

また，センサベースの利点であるロバスト性と，ビジョンベースの利点である精度を両立するために，両者を併用した “ハイブリッドな手法” も提案されている．画像のみでは安定に動作しにくい屋外であったり，センサのみの精度では不十分なテーブルトップシステムなどに多く用いられている．

キヤノンが開発した MR Platform システムは，ビデオシースルー型の HMD と磁気センサから構成されており，センサからカメラの 3 次元的な位置姿勢を取得して位置合わせを行う．このとき，センサからの情報だけではなく，マーカや自然特徴によるビジョンベースの手法とを併用することで，フレーム間の揺らぎや，仮想物体の細かな位置合わせずれを表すジッタなど位置合わせ精度の不安定さを解消することができる．

また，Schall らのシステムでは，屋外のあらゆる未知な環境に対応するために，複数のセンサとビジョンベースのトラッキングを融合している．GPS，ジャイロ，磁気，加速度計といった多数のセンサからの情報をカルマンフィルタリングによって統合し，さらにビジョンベースのトラッキング結果と融合することで，安定した精度を維持している．さらに，UMPC のような小型でパワーの少ない PC であってもリアルタイムに動作可能であり，屋

外でも精度の良い AR を行うのに最適である．

今後の課題

このように，AR のための位置合わせ問題はすでにビジョン・センサともにさまざまなアプローチが提案されてきた．しかし，これまでは主に学術的な研究としての側面が強く，デスクトップ PC を使った据え置き型のシステムや，それなりの計算機能を持ったモバイル PC を“背負う”タイプのウェアラブルシステムなどを想定した手法が中心となっていた．そこで，今後の課題としては，これらのアルゴリズム・手法をいかに実用的なものにしていくかということがある．ハード面に関しては，最近の iPhone などに代表される高機能携帯電話の普及に伴い，携帯端末を利用した AR が大いに注目されてきており，GPS などのセンサの精度・整備も格段に進んできた．よって今後は，どのようなサービスを提供すればよいかといったコンテンツ自身が重要となり，その特徴に適した位置合わせへと進化させていくことが必要である．

参考文献

1）加藤博一，Billinghurst, M.，浅野浩一，橘啓八郎：マーカー追跡に基づく拡張現実感システムとそのキャリブレーション，日本バーチャルリアリティ学会論文誌，Vol.4, No.4, pp.607-616（1999）．
2）吉田友祐，天目隆平，柴田史久，木村朝子，田村秀行：半人為的幾何位置合わせマーカの研究（第 1 報），電子情報通信学会技術研究報告 PRMU2006-195，Vol.106, No.470, pp.7-12（2007）．
3）中里祐介，神原誠之，横矢直和：ウェアラブル拡張現実感のための不可視マーカと赤外線カメラを用いた位置・姿勢推定，日本バーチャルリアリティ学会論文誌，Vol.10, No.3, pp.295-304（2005）．
4）天目隆平，西上彰人，柴田史久，木村朝子，田村秀行：ポスタを利用した複合現実感用幾何的位置合わせ，日本バーチャルリアリティ学会論文誌，Vol.14, No.3, pp.351-360（2009）．
5）小野友也，岩井儀雄，石黒　浩：二次元フーリエ変換を利用した物体の姿勢推定，情報処理学会研究報告，Vol.2010-CVIM-170, No.9.
6）Lepetit, V., Vacchetti, L., Thalmann, D. and Fua, P. : Fully Automated and Stable Registration for Augmented Reality Applications, *Proc. 2nd IEEE/ACM Int. Symp. on Mixed and Augmented Reality*, pp.93-102（2003）．
7）Kotake, D., Satoh, K., Uchiyama, S. and Yamamoto,H. : A Fast Initialization Method forEdge-based Registration Using an Inclination Constraint, Proc. 6th IEEE/ACM Int. Symp.on Mixed and Augmented Reality, pp.1-10（2007）．
8）Lowe, D. G. : Distinctive Image Features from Scale-invariant Keypoints, *Journal of Computer Vision*, Vol.60, No.2, pp.91-110（2004）．
9）Klein, G. and Murray, D. : Parallel Trackingand Mapping for Small AR Workspaces, *Proc. 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*（*ISMAR'07*），Nara，Japan（2007）．

（平成 22 年 2 月 22 日受付）

植松裕子
yu-ko@hvrl.ics.keio.ac.jp

2004 年慶應義塾大学理工学部情報工学科卒業．2006 年同大学院前期博士課程修了，2009 年同大学院後期博士課程修了．同年より同大理工学部助教．複合現実感やコンピュータビジョン等の研究に従事．博士（工学）．