

展望 2: ARのためのセンシング

亀田能成

筑波大学

AR 技術の技術基盤を成すのは、「基礎 1：拡張現実感 (Augmented Reality : AR) 概論」でも解説されているように、現実世界と仮想世界とを幾何的・光学的・時間的に一致させる技術である。その要求精度はアプリケーションによって異なるが、一般に AR はリアルタイムシステムであることが多いため、まず時間的一致が大前提となる。すなわち、何をやるにしても低遅延かつ高スループットであることが望まれる。また、人間の視覚心理上、幾何的に一致してからでないと光学的 (色彩的) 整合性の評価に入らない傾向があるので、AR では長く幾何的一致をリアルタイムに実現する方法、すなわちトラッキングについて研究が重ねられてきた。本稿では、幾何的一致のためのセンシングに関する話題を中心に取り上げる。

トラッキング

トラッキングとは、世界に対する対象物体の移動量と回転量を表す 6 自由度の変数をリアルタイムに求めることである。実は、環境にかかわらず常にこの目標を達成することは、現在の研究レベルをもってしてもまだ容易ではない。そのため、さまざまな条件付けが行われ、そのもとでトラッキングのためのセンシングが試みられている。

--- ユーザと注視対象との関係 ---

AR には、大きく分けて 2 通りの形態があり、それに応じてセンシングの前提条件が変わる。1 つが Inside-out 方式であり、もう 1 つが Outside-In 方式である (「基礎 1：拡張現実感 (Augmented Reality : AR) 概論」も参照)。

Inside-out ではユーザの周りのシーンすべてが「現実世界」であり、そのシーンの上に仮想物体が重畳される。この場合、ワークエリアはそのシーン全体と考えられ、その中をユーザが動き回っていることになる。Inside-out で重要なことは、広く動き回るユーザの視点、あるいはそのユーザの視点を代行するユーザ視点カメラを精度よくトラッキングすることである。

これに対して、Outside-in では、注視対象が 1 つの物

体ないし比較的狭いワークエリアであり、ユーザはそれを周りから観察する。あるいは、ユーザの視点の位置・姿勢情報をシステムが外部から観測する。この場合も重要なことはユーザの視点情報であるが、外部からの観測系を準備するとき、その観測系として多数のカメラを設置したり、特殊なライティングとマーキングを行ったりするなどさまざまな工夫を凝らすことができるので、視点情報推定のための工学的解法を見つけやすいという利点がある。システムが大規模になる難点もあるが、これまで商業化されているトラッキングシステムの多くは Outside-in を前提としている。

--- 必要精度 ---

具体的なカメラトラッキング技術について次章以降で述べる前に、AR に必要な精度について考えてみよう。

一般的なカメラベース AR システムを利用するときの代表的な指標としては、そのカメラ画像上での見かけのずれが挙げられる。これは再投影誤差と呼ばれ、単位は画素ないし画像に対して相対化された量である。再投影誤差が画素レベルまで小さくなれば、見かけ上はそれ以上の精度を追求する必要はない。

別の代表的な指標としては、利用者の視力に基づく指標が挙げられる。AR とは人間に見せるための技術であるので、人間の視点を基準に取ってその人の視力を考えると、到達すべき精度の上限が分かる (図-1「AR における視力の問題」も参照)。

単純化した例として、光学シースルー HMD による AR システムを考えてみよう。ユーザの視力を 1.0 とすると、ISO の視力の定義によればこれはランドルト環の C の形の切れ目部分が、1/60 度 (1 分) のものを区別できる視力に相当する。眼の 1m 先で AR を行う場合、この角度は 0.29mm に相当する。もしくはその光学シースルー HMD が水平画角として 45 度確保できるとすると、必要な水平画素数は 2700 画素必要となる。一方、ビデオシースルー HMD の場合は、その HMD の解像度以上の推定は無意味である。この場合、たとえば水平画角 60 度、水平画素数が 1280 画素の HMD を考えると、視

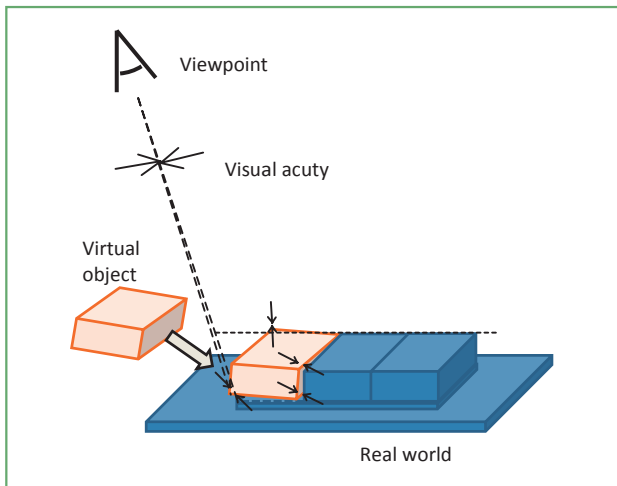


図-1 ARにおける視力の問題

力は0.36程度に相当する。表-1「視力と必要画素数との関係」にその関係を示す。

なおここでは簡単化のために静止視力のみ取り上げたが、実際には、ARでは動体視力など静止視力以外の要素も問われることに注意されたい。

また、携帯端末やディスプレイ上でARを行う場合は、通常はユーザがそのディスプレイの1画素が判別できるほどの近距離にいると考えられるので、ディスプレイ上での1画素が精度の上限となる。

画像以外のセンサによるカメラトラッキング

古典的には、ARのためのカメラトラッキングは、そのカメラの画像を用いることなく実現されてきた例がいくつも挙げられる¹⁾。これは、画像処理が時間のかかる処理ゆえに遅延が大きかったことや、画像処理時の頑健性に問題があることなどが原因である。

--- エンコーダ ---

カメラ位置を計測する最も古典的な方法の1つが、カメラやHMDを自在アームに載せ、自在アームの移動量や関節の角度をエンコーダから読み出す方法である。研究ではもう取り上げられることはないが、圧倒的な精度と堅牢性によって、商業的に最も成功した方法として分類できる。

映画撮影におけるVFXでは、モーションコントロールカメラを用いてマッチムーブすることでカメラトラッキングを実現していたが、あくまでポストプロセスであった。これに対して現在よく行われている手法は、主にカメラを三脚に載せ、そのカメラヘッドが向けられた方向をロータリーエンコーダで読み出すタイプである。バーチャルスタジオやスポーツ中継の現場では普及が進み、

視力	0.3	1.0	2.0	
分解能 [分]	3.3	1.0	0.5	
水平視野角 45°	必要画素数 [pixel]	810	2700	5400
水平視野角 60°	必要画素数 [pixel]	1080	3600	7200

※必要画素数：透視投影を仮定すると正面と端では必要解像度が異なるがここでは最も解像度が必要な正面部分の解像度で換算

表-1 視力と必要画素数との関係

特にテレビのニュース番組では一般的に用いられるようになった。

エンコーダを用いる方式の難点は、アームないしは三脚にカメラを取り付けなくてはいけない制約から、どうしても装置が大掛かりになる点である。そのため、番組制作現場のようにカメラが大きい場合か、カメラを三脚に載せても構わないような状況でしか用いることができない。また、移動に対するエンコーダを用意するのは機械的に大掛かりになることが多いため、カメラの方向は自由でもその位置は固定になることが多い。

一方で、利点は推定精度のよさとその堅牢性である。通常、ロータリーエンコーダはカメラのフレームレートより圧倒的に高速な読み出しが可能であり、かつ100,000C/T以上の角度分解能も珍しくない。誤推定も通常は発生しない。100,000C/Tあれば、1カウントあたりの角度分解能は0.216分に達し、視力換算で2.3となる。これは、人間の通常の視野角と同程度にカメラの撮影画角を設定すれば、その誤差を人の目が認識することが難しいことを意味する。実際にはズーム撮影が行われることもあるため、用いられるズームレンジに応じて角度分解能がさらに数倍高いロータリーエンコーダが用いられる。

--- ポイントマーカによる空間定位 ---

エンコーダベースのセンシングには可用性に制約が多い。それに代わるものとして、何らかの手段で特徴づけたある点またはある小物体を空間中に投入し、周辺に設置した観測装置でその位置を計測する方式が昔から提案されている。

• 計測手段：

可視光、赤外光、超音波

• アクティブマーカ：

LED (可視光・赤外光)

超音波発振機

• パッシブマーカ：

着色球や着色小片

2次元マーカ

赤外線反射球

光を計測手段にする場合は、観測装置側はそのマーカの

みを効率よく検出できるよう特殊加工されたカメラが複数台組み合わせられ、三角測量の原理で空間定位を行う。超音波の場合は、複数のマイクロフォンを配し、それらの間での音の到達時間差から空間定位を行う。

なお、ポイントマーカが1点だけでは、対象となるカメラの位置は得られても方位が得られないので、マーカを2点以上載せるか、後述するジャイロセンサを組み合わせることが一般的である。

ポイントマーカによる空間定位は、ポイントマーカと観測装置の両方に工夫を重ねることで、高精度かつ頑健に結果を出力することができる。また、観測側には実質的に専用システムが用意されるため、実行速度も確保できる場合がほとんどである。一方、欠点としては、専用システムが必要であること、ワークエリアの大きさに比して観測装置が大掛かりになることが挙げられ、いずれも導入コスト高に結びつく。また、ポイントマーカが観測装置から見えなくなるようなオクルージョンが発生する状況にも不適である。

--- 磁界センサ ---

ポイントマーカによる空間定位によるシステムと同程度の古い歴史を持つセンシング方法として、ワークエリアに強制的に磁界を発生させ、レーザーで電磁誘導を計測することでそのレーザーの位置と方位を同時に計測するシステムが提案・販売されている。

この方法は、6自由度を一度に推定でき、オクルージョンの心配もなく、かつカメラのフレームレートよりも高速に計測できるという点で優れている。1994年と比較的古くから完成されたシステムが販売されていたこともあり、VR・ARの研究開発に用いられることが多かった。一方で、この手法の欠点として環境の磁界変化に弱いという問題点がある。電磁波を発生する電子デバイスを組み合わせざるを得ないARシステムで利用するには、慎重なデバイスデザインが求められる。また、磁界の発生範囲内でしか計測できないため、原理的にワークエリアを大きくしづらいという点も制約の1つである。専用システムであるため、導入費用も安価とは言い難い。

--- ジャイロセンサ ---

ポイントマーカによる空間定位と相互補完的に用いたり、後述する画像ベースのカメトラッキングの補完をしたりする手段として、ジャイロセンサが用いられる。ARでは、主に振動型と光ファイバ型がその大きさや扱いやすさからよく用いられる。

振動型のジャイロセンサはMEMS技術の向上に伴い、チップレベルでユニット化され低廉化が進んだことから、急速に普及しつつある。

利点としては、軽量小型かつ高い頑健性が挙げられる。性能についても現在は3軸を1ユニット化したものが利用できるようになっている。一方で、ジャイロセンサは計測中にドリフトが発生することと、その性質から静止時には一切姿勢情報が入手できないことが欠点である。これらの欠点を解消するため、最近では地磁気センサ(電子コンパス)や重力を利用した傾斜角センサを統合したユニットが提供されている。

もう1つの方式の光ファイバ型ジャイロセンサは、リング状の光ファイバに光を投入し、光ファイバ経路が運動したときに発生する光の位相差を計測することで回転角度を求めるものである。精度がよくドリフトも小さいという利点があるが、振動型ジャイロに比べるとその機構上、小型化しにくく高価である。

--- GPS ---

屋外での位置合わせにおいて、現在最も普及が進んでいるのがGPSであろう。しかしながら、GPSによる位置測定精度は数mから数十mであり、対象物がこの測定精度に比して十分に遠距離にあるという状況でないといふARに直接用いることは難しい。

測定精度に影響を与える要因はいくつも挙げられるが、そのうち特に、電離層および対流圏での信号の乱れを補正する方法として、Differential GPS (D-GPS) と Real-Time Kinematic GPS (RTK-GPS) が挙げられる。

D-GPS方式では、計測対象のGPS受信機とは別に、受信機の近くにGPS受信局を用意し、その位置は厳密に既知であるとする。受信局での位置測定結果と受信局が存在しているはずの位置とのずれが誤差要因であるとして、受信局はそのずれ情報をGPS受信機に送る。GPS受信機ではずれ分だけ補正して測定精度を向上させる。これにより、精度は数m程度になると言われている。受信局とずれ情報の受け取り方法によってD-GPSシステムにはさまざまなバリエーションが存在するが、現在日本で最も汎用的に利用できるD-GPSは静止軌道人工衛星MTSATを利用した、Satellite Based Augmentation System (SBAS)の日本版であるMulti-Functional Satellite Augmentation System (MSAS)であり、2007年から本格運用が開始されている。SBASの場合、現在市販のGPSでも対応しているものがあるため、特に追加の費用の必要なく導入することができる。

これに対して、RTK-GPSはGPS衛星からの信号電波の位相差を数えてD-GPSよりさらに受信機での誤差を低減させる方法であり、その測定誤差を数cmにまで小さくできる。ただし、RTK-GPSを実施するためには受信局を受信機から数km内に用意する必要があるため、実質的には利用時には常に受信局と受信機の両方を用意

	分解能		ワークエリアの 一辺 [m]	更新レート
	角度 [分]	位置 [mm]		
エンコーダ	1 以下※ a	1 以下※ a	装置依存	◎
ポイントマーカ	(球・点)	N/A	1 ~ 10 ※ b	○
ポイントマーカ	(2次元マーカ)	N/A	0.1 以下	○
ポイントマーカ	(hybrid)	6	2 ~ 5	◎
ポイントマーカ	(hybrid)	5 ~ 15	2	◎
磁界センサ	10	1	1.5m	◎
ジャイロ	振動式	15	N/A	◎
ジャイロ	光ファイバ式	1 ~ 6	N/A	○
GPS	N/A	~ 10m	屋外	1fps
D-GPS/A-GPS	N/A	~数 m	屋外	1fps
RTK-GPS	N/A	数 cm ~数十 cm	屋外	1fps
無線 LAN アクセスポイント	N/A	5 ~ 50m	屋内外	約 1fps

※ a: 構造によっては桁違いに向上可能

※ b: 使用するカメラとその台数およびワークエリアに依存するがここでは一辺 1k 画素程度のカメラの利用で 1 立法 m 程度のワークエリアを仮定
数値はいずれも既存のいくつかのシステムの公表値に筆者の経験を加味した目安であることを断わっておく。

表-2 非画像センサの特徴

する必要があり、かつ信号電波の位相差まで計測できる GPS 装置はいまだ高価である。加えて、その性質上、RTK-GPS はマルチパスに非常に弱く、安定した計測のためには受信局・受信機ともマルチパスがほぼ存在しない環境であることが必須である。

--- 携帯電話基地局・無線 LAN アクセスポイントの利用 ---

GPS がマルチパスの影響により特に都市部で位置推定精度が保てないため、それに代わる手段として、都心部で高密度で見られるようになった無線電波を用いる手法が提案されている。具体的には、携帯電話基地局や無線 LAN アクセスポイントを利用する。携帯電話基地局の位置情報は携帯電話上のサービスとしては利用できるが、研究開発の基盤としては外部に公開されていないため、AR への利用は限定的である。それに対して、無線 LAN アクセスポイントは電波到達距離が短いこと、一般にその電波の参照は禁止されていないことから、これを利用した定位方法の普及が進められている。海外では Skyhook wireless 社が米国を中心にサービス展開しており、日本ではクウジツ社が PlaceEngine というサービス名で普及を進めている。

現在の推定精度は D-GPS とそれほど変わらないが、GPS と異なり天頂が開いてなくともよいこと、AR で用いる多くの機器は無線 LAN 受信機構を備えているため設備投資が不要であることから可用性に大きな利点があり、今後も普及が進むと思われる。

--- 他の屋内広域定位法 ---

現在、上記以外にもさまざまな研究が進められているが、利用者側の設備投資を最小限に抑えることを重視する取り組みとしては、国内では GPS の概念を屋内に拡張した Indoor Messaging System (通称屋内 GPS) と、Bluetooth を利用したユビキタスコミュニケーター (UC) が挙げられる。いずれも社会的な実証実験による検証まで研究が進んでいる状態である。

以上の非画像センサの特徴を表-2「非画像センサの特徴」に示す。筆者の主観的評価が入っているため、数値は目安程度の参考とされたい。

画像処理によるカメラトラッキング

コンピュータビジョンの進歩と計算機環境の向上に伴って、従来は困難とされていた、カメラ画像をトラッキングに直接用いる手法が現実的になりつつある。

カメラベースの AR を行う場合、画像処理によって現実世界と仮想物体との位置ずれが画像上でサブピクセルレベルにまで到達すれば、利用者から見れば幾何的には完全な AR が実現されたことになる。本章では、AR のためのリアルタイムカメラレジストレーション技術について概観する。なお、本特集の「基礎 2：位置合わせ技術」、「基礎 3：開発用ツール」に技術的詳細が述べられているので併せて参照されたい。

画像処理に基づくカメラトラッキング手法は、システムハードウェアとしてはカメラ 1 台が必要なだけであるため、導入コストを最小に抑えられるという利点がある。

--- カメラ特性 ---

AR システムでは、透視投影に従うカメラがもっぱら使われる。このとき、画角・焦点距離・分解能は固定値として扱われることが多い。これは、AR システムの可搬性を考えるとき、調整可能な大きなレンズをカメラに取り付けることは避けたいからである。

カメラレジストレーションのためには、使用するカメラの画角は広角であるほうがよい。これは、カメラの回転運動が発生したときに、広角カメラであるほど平行移動と回転運動の分離が行いやすくなるからである。この意味では、魚眼カメラや全方位カメラを用いれば一般的なカメラよりさらに平行移動と回転運動の分離が行いやすくなるが、レンズ系が大型化し可搬性を損なうこと、AR として画像をユーザに提示する際に透視投影の像になるように変換する必要があること等が問題となるため、一般的には利用されていない。

広角レンズの場合は歪差が表れることが多いので、透視投影を前提とする手法を適用する場合、事前に歪差を取り除く処理が必須となる。

--- マーカベーストラッキング ---

AR において、人為的マーカの利用が許されるとき、最も可用性に富むのが、板状のマーカを用意して追跡する手法であろう。この方法は ARToolKit をはじめとするさまざまな公開ライブラリと相まって現在に至るまで広く利用されている。

ARToolKit に限らず、板状マーカを用いる方法では、マーカのパターンによって個体 ID を割り付けることで、マーカに対するカメラの位置推定だけでなく、付加的なサービスを実現していることが多い。

板状マーカの亜種としては、近赤外線反射材でマーカを作成して赤外ライトで照射してレジストレーションを行う手法や、白黒の無機質なパターンではなくポスタなど自然なレイアウトの中にマーカとしての情報を巧妙に埋め込む方法なども提案されている。

ポイントマーカによる空間定位のうち、2次元マーカに基づく手法はこの分類に含めることができる。

なお、2次元マーカを単独で用いる場合、マーカ平面に垂直な方向のレジストレーション精度は、マーカ平面内の平行移動に対する精度に比べて格段に悪くなる。そのため、ワークエリア座標系のどの直交軸に対しても精度を確保するためには、3枚の2次元マーカを互いに直角になるように配置するなどの工夫が必要である。

--- ランドマーク・CAD ベーストラッキング ---

マーカベーストラッキングは可搬性が高く応用範囲が広いが、その性質上、検出されやすいパターン

がワークエリア中に必須である。屋外の広いエリアでの作業や、審美的観点が要求される場合等、このようなマーカが利用できない状況での解決法として、ランドマークおよび CAD データに基づくオンラインレジストレーション法が提案されている。

原理的にはマーカベースと変わらないが、ランドマークや CAD データから予想される画像特徴量はマーカと比べて安定して発見することが難しく、実現には工夫が必要になる^{2), 3)}。

--- 自然特徴によるトラッキング ---

Inside-out 環境における究極の解法は、周囲にたまたま存在している、ユーザが意図して設置したわけではないという意味での自然特徴を利用したオンラインレジストレーションである。カメラがいくら運動しても、シーン中で静止している自然特徴点は3次的に一意にしか存在できないというよく知られた当然の拘束条件を用いるわけであるが、この問題のリアルタイム解法が具体的に示され始めたのはここ数年のことである。

最近ではより低レベルのハードウェア性能でもリアルタイムトラッキングを成功させるための取り組みが実を結びつつあり、携帯電話クラスのハードウェアでも 10fps 以上のトラッキングが報告されている^{4), 5)}。

センサフュージョン

前述のように、トラッキングにおける個々のセンシング手段としては非画像ベースと画像ベースとに大別できるが、両者を統合して利用することが最近の AR システムでは一般的な傾向である。

画像ベースのセンシングや非画像ベースのセンシング技術はいずれも前提条件・方法・精度がさまざまであるため扱いが難しいが、それでもそれらを統合することで、単体のセンシングではトラッキング不可能な状況に対応し、かつ高い精度をする研究が進められている⁶⁾。ただ、これらは個別のセンサ技術の近年の成果に比べると、まだ改良の余地が多く残されていると思われるので、日本から研究発信できるよいチャンスであろう。

今後の展開

本稿では、AR システムの中核を成すカメラトラッキングを中心に、AR のためのセンシング技術の現状と動向を概観した。

トラッキングについては、画像ベースの技術を中心に、この5年で急速に実用化に近づきつつあるという印象である。現状ではエンコーダによる姿勢推定しか限界精度



図-2 Tracking Competitionの様子 (ISMAR 2009)

に到達しているとは言えない状況であり、まだ改良の余地は多々残されてはいるが、筆者は、冒頭で述べたようなARのために必要な限界精度にこれらの技術が到達する日はそう遠くないと確信している。特に画像ベースの技術の進展は今後も目を離さないであろう。

また、トラッキングの研究が成功を収めるためには、研究開発環境の整備のほかに、評価基準の確立が重要である。研究開発環境はOpenCVが1999年のリリース以降デファクトスタンダードの位置を急速に固めつつあるが、評価環境はInternational Symposium on Mixed and Augmented Reality国際会議においてTracking Competition (図-2「Tracking Competitionの様子」という場が設けられている以外は、頼れる基準がないのが現状である。この状況を改善すべく、現在、日本のARコミュニティにおいて、Trakmark⁷⁾という名でカメラトラッキングの評価用データベースの整備を進めているところである。日本発の技術貢献ということで、関連のあ

る方々にはご協力いただければ幸いである。

なお、今回は紙面の都合で割愛したが、ARのためのセンシングとしては、対象物体・シーンデータのリアルタイム構築も長年研究が進められている。ARでの取り組みの特徴は、カメラのレジストレーションと並行して3次元データの獲得を行うため、データが常に最新の状態にあり、即座にARサービスに組み込むことである。こちらの研究も長足の進歩を遂げているので、今後も注意していく必要があろう。

参考文献

- 1) Azuma, R. T. : A Survey of Augmented Reality, Presence, Vol.6, No.4, pp.355-385 (1997).
- 2) 大江統子, 佐藤智和, 横矢直和: 幾何学的位置合わせのための自然特徴点ランドマークデータベースを用いたカメラ位置・姿勢推定, VRSJ, Vol.10, No. 3, pp.285-294 (2005).
- 3) Kotake, D., Satoh, K., Uchiyama, S. and Yamamoto, H. : A Fast Initialization Method for Edge-based Registration Using an Inclination, ISMAR, 10 pages (2007).
- 4) Wagner, D., Schmalstieg, D. and Bischof, H. : Multiple Target Detection and Tracking with Guaranteed Framerates on Mobile Phones, ISMAR, pp.57-64 (2009).
- 5) Arth, C., Wagner, D., Klopschitz, M., Irschara, A. and Schmalstieg, D. : Wide Area Localization on Mobile Phones, ISMAR, pp.73-82 (2009).
- 6) Schall, G., Wagner, D., Reitmayr, G., Taichmann, E., Wieser, M., Schmalstieg, D. and Hofmann-Wellenhof, B. : Global Pose Estimation using Multi-Sensor Fusion for Outdoor Augmented Reality, ISMAR, pp.153-162 (2009).
- 7) TRAKMARK : Benchmark Test Schemes for AR/MR Geometric Registration and Tracking Methods, <http://www.trakmark.net/> (平成22年2月18日受付)

亀田能成 (正会員)

kameda@iit.tsukuba.ac.jp

1991年京大・工・情報卒業。1996年同大学院博士後期課程認定退学。同年同大学助手。2001～02年米国MIT客員研究員(併任)。2003年筑波大学講師。2004年同大学院助教授、現准教授。複合現実感やマルチメディア処理等の研究に従事。博士(工学)。

