

1

基礎 1：拡張現実感 (Augmented Reality : AR) 概論

神原誠之

奈良先端科学技術大学院大学

拡張現実感とは

拡張現実感 (Augmented Reality : AR) とは、ユーザが見ている現実のシーンにコンピュータグラフィックス (CG) によって描かれた仮想物体を重畳表示することで、ユーザがいる場所に応じた情報を直感的に提示する技術である。この技術は、現実環境と仮想環境を融合する技術である複合現実感 (Mixed Reality : MR) の一分野に含まれる。図-1 に示す複合現実感分野では、現実環境と仮想環境の融合方式として拡張現実感と仮想化現実という2つの分野が示されているが、この現実と仮想の融合方式に基づく分類ではその境界が明確に定義できないとされており、現実環境と仮想環境は連続であると言われている。

AR は MR の中で現実環境に近いところに位置づけられる。AR はユーザが現実環境で見ているシーン上に CG で描かれた仮想物体が重畳表示される。図-1 の例では、ユーザが特殊な眼鏡をかけると現実の何もなかったところにかつて建造されていた大極殿があたかもそこにあるように見ることができる。このように大部分が現実で仮想物体が部分的に融合し、現実を拡張するという意味で拡張現実感と言われている。一方、計算機内に構築される仮想環境に現実環境の情報を取り込むことで仮想物体のクオリティを向上させる技術が仮想化現実 (Augmented Virtuality) である。これは、もともとバーチャルリアリティ (VR) において CG で描画されていた仮想物体の写実

性を向上させ、より臨場感の高い仮想環境を構築することが可能である。図-1 の例では、現実環境で撮影した映像から、現実世界の建造物の形状とテクスチャを用いて 3D モデルを自動作成した例である。モデリングソフトのみで作成された 3D モデルに比べ写実的なモデルを生成することができる。この仮想化現実では、ただ単に現実環境の情報を仮想化するだけでなく、それをいかに人間に臨場感高く提示するかも技術課題に含まれる。

AR を実現するための基礎技術

AR は、VR の発展形として位置づけられており、AR を実現するための基盤技術の多くは VR 技術と共通しているため、それらを利用することが可能である。しかし VR で提示されるシーンはすべてが仮想環境であるが、AR の場合、背景となる実シーンに仮想物体が融合されるため、現実と仮想の正確な融合を実現するための技術課題が存在する。

その技術課題の1つは、人間が見ているシーンにどのように、仮想物体のレンダリング結果を重畳表示させるかというディスプレイの問題である。VR の場合、もともとすべてが CG で描かれたシーンのみを見るため、HMD や CAVE などの没入型ディスプレイを利用することが一般的であるが、これらを AR のためのディスプレイとしてそのまま利用することはできない。AR で用いられるディスプレイは、現実環境と仮想環境を融合して



図-1 複合現実感の技術体系

見せるためシースルー方式と呼ばれるディスプレイが利用される。

その他 AR では、現実環境と仮想環境の整合性に関する技術課題が挙げられる。これは現実環境と仮想物体のレンダリング結果を合成する際、いかに矛盾なくそれらを合成できるかという問題である。一般に AR で取り扱われる現実環境と仮想環境の間の整合性は以下の 3 つが挙げられる。

- 幾何学的整合性
- 光学的整合性
- 時間的整合性

幾何学的整合性とは現実環境と仮想環境の 3 次元的な位置合わせを意味しており、現実環境の正しい位置に仮想物体が存在するような映像を作り出すことである。2 つ目の光学的整合性は、現実物体と仮想物体の陰影や画質の整合性を取り扱った問題である。最後の時間的整合性は、現実環境と仮想環境の時間的な整合性であり、現実環境と仮想環境を合成した際、遅延や同期ずれがないことを意味する。一般的に AR の場合、現実環境でユーザが動的に移動するような環境が想定されるため、3 つ目の時間的整合性を保ちつつ、いかに幾何学的・光学的整合性を実時間で解決するかが課題となる。

AR のためのシースルーディスプレイ

AR で用いるシースルーディスプレイには、現実のシーンと仮想物体の合成方法の違いから、光学式シースルー方式とビデオシースルー方式の 2 つに分類できる。

〔光学式シースルー方式〕

図-2(a) に示すように、ハーフミラーなどを用いて、映り込む仮想環境と透過して見える現実環境を同時にユーザに提示する方式である。現実環境が時間遅れなく提示できる反面、仮想環境を提示する際の計算時間などは、現実環境と仮想環境の同期ずれとなり、これが位置ずれとしてユーザに知覚される。ハーフミラーを用いているため現実環境が若干暗く見え、仮想物体は半透明に表示される。そのため、現実物体と仮想物体の前後関係を隠蔽により表現するのは一般的に困難である。また、仮想物体を置くことによる実物体への影などの表現も難しい。ただし、奥行き隠蔽関係に関しては、近年、光学合成方式において、特殊な表示デバイスを用いて隠蔽関係を表現する研究も行われている。

〔ビデオシースルー方式〕

図-2(b) に示すように、カメラによって撮影された現実環境の画像上に、仮想物体を描画する方式である。ユーザの視線方向とカメラの光軸方向を一致させて撮影した現実環境の映像上に、仮想環境を合成して提示するこ

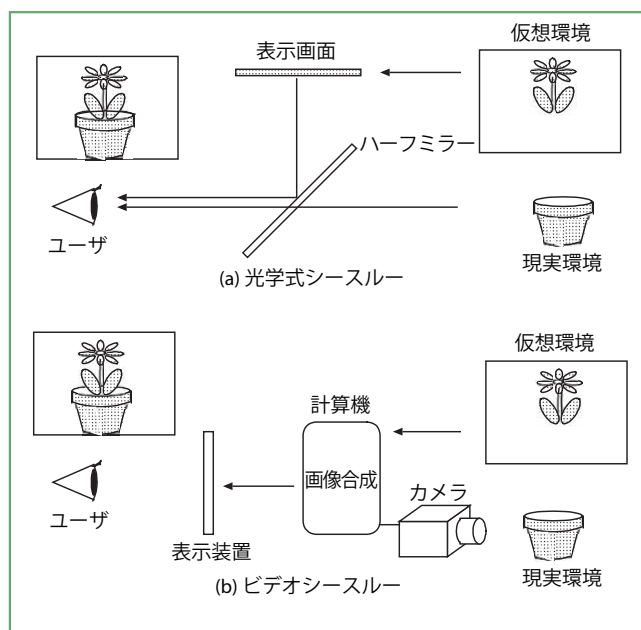


図-2 シースルー方式

とで実現される。現実環境と仮想環境の同期をとって提示できるため、同期ずれによる両者の位置ずれは生じない。本方式では、現実物体と仮想物体の前後関係が分かれば、お互いを隠蔽でき、現実物体との前後関係を表現できる。ただし、提示される AR 環境全体が、仮想環境を重畳合成する際の計算時間などによって遅れて提示される。

AR における幾何学的整合性

--- 現実世界と仮想世界の幾何学的位置合わせ ---

AR における幾何学的整合性は、現実環境と仮想環境の位置ずれのない合成画像を生成することを意味しており、ユーザの違和感の解消に最も大きな影響を与える整合性である。一般的な AR のシーンにおける座標系の関係を図-3 に示す。現実環境に基準となる世界座標系が設定され、その座標系中に合成したい仮想物体とユーザの視点が存在する。位置合わせ問題は、一般に仮想物体とユーザの視点位置の関係（図-3 中 C）を推定することで解決できる。現実世界（世界座標系）と仮想物体の関係（図-3 中 B）はアプリケーションの管理者によりあらかじめ設定されるため、世界座標系におけるユーザの視点位置・姿勢（図-3 中 A）の推定が必要となる。そのため一般的には、幾何学的整合性問題は、世界座標系におけるユーザ視点の位置・姿勢を推定する問題に帰着する。

一般にユーザ視点の位置・姿勢は、世界座標系の基準を示すためのインフラを用いることにより実現される。この問題は VR と共通の課題であったため、AR 初期では VR で利用されていた方法を利用することが多かった。

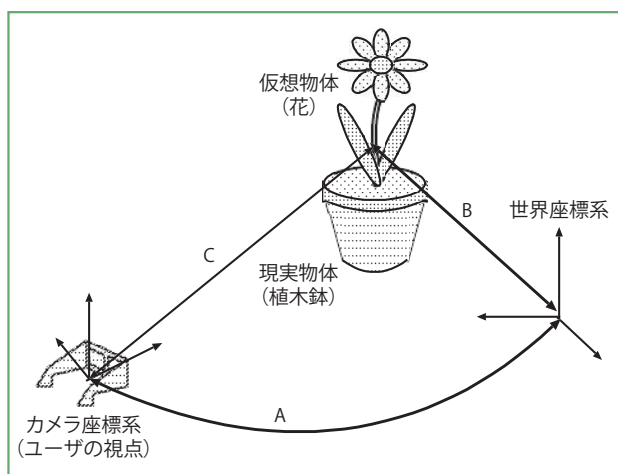


図-3 視点と世界座標系の関係

最も有名なものに、Polhemus 社 FastRak に代表される 3 次元磁気センサがある。これはトランスミッタと呼ばれる磁場発生装置を世界座標系における位置が既知の場所にインフラとして設置し、ユーザの頭部に取り付けられたレシーバで磁場を受け取ることで、トランスミッタとの 3 次元位置関係を計測するものである。これにより、世界座標系におけるユーザの視点位置を計測することで、位置合わせを行う。これに限らず VR で利用されていた光学式や超音波式の 3 次元センサも流用された。これらの方式は世界座標系の基準となるセンサをインフラとして実世界に設置し、ユーザの頭部に取り付けた機器などから何らかの情報を得ることで相対的な位置関係を計測し、世界座標系におけるユーザの視点位置を計測する。

1990 年代後半から、VR ではなかったカメラを用いた位置合わせ手法が見られるようになってきた。これは、ユーザの視点位置付近に取り付けられたカメラで撮影した映像を利用するものである。この頃から通常のデスクトップ PC を用いて実時間でカメラ映像のキャプチャ・画像処理が行えるようになってきたことが普及した 1 つの要因である。それに加え、この方法は、前述のビデオスルー方式の AR と非常に相性が良いという特徴がある。ビデオスルー方式の AR の場合、ユーザの視点付近に光軸と一致するように取り付けられたカメラで撮影された映像を AR 合成の背景として用いる。そのため、その映像をユーザ視点の位置・姿勢推定にも利用することで、カメラ以外の他のセンサが不要となる利点がある。さらに、近年 USB などで接続できるカメラに代表されるような安価なカメラが容易に手に入るようになり、簡単に AR システムを構築できるようになったため、カメラを利用した幾何学的位置合わせ手法は、AR 分野における位置合わせ手法の主流となっている。

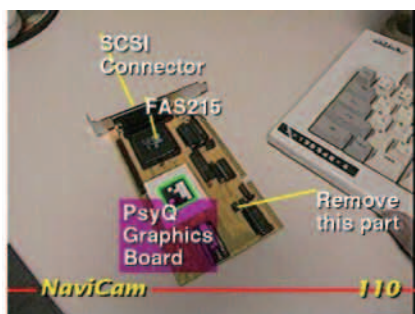
ビデオスルー方式の場合、カメラと視点の位置関

係は固定されるため、世界座標系におけるカメラの位置・姿勢を推定することが、ユーザ視点の位置姿勢を推定することと同義となる。この場合は、世界座標系における位置姿勢が既知の矩形パターンなどの画像マーカをインフラとして配置し、それらをカメラで撮影し、画像中にマーカがどのように映るかを解析することで、マーカに対するカメラの 3 次元位置・姿勢関係を推定することが一般的である。マーカが撮影された画像から、カメラとマーカの位置関係を推定する方法として、PnP (Perspective n-Points) 問題が利用されることが多い。PnP 問題では、位置関係が既知の平面上 4 点か平面上にない 6 点が、透視投影モデルに基づいて撮影された画像上でそれぞれどこに撮影されているかを対応付けることで、その点群とカメラの位置関係を推定することが可能になる。AR では一般に正方マーカの 4 頂点を位置関係が既知の平面上の 4 点として、それらの画像上の座標を検出することで位置合わせを実現している。

--- カメラを用いたさまざまな幾何学的位置合わせ手法 ---

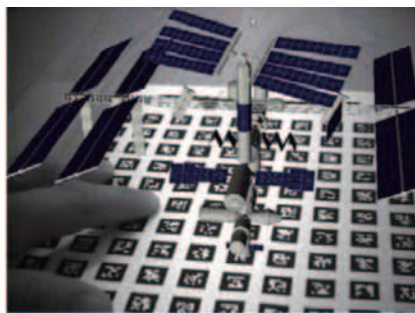
カメラで撮影した映像を用いた幾何学的位置合わせを行った先駆的な研究として、暦本の研究が挙げられる¹⁾。この研究では図-4(a) に示すように、正方形の黒い枠とその中の白黒のドットで表現される ID 部で構成された 2 次元バーコードをカメラで撮影した映像から検出することで、マーカとカメラの幾何学的な位置・姿勢関係を推定するものである。内部の ID 部でマーカの向きや種類を認識し、外部の正方マーカの 4 頂点を用いて位置合わせを行っている。

AR においてカメラを用いた位置合わせ手法を世に広めたのは、暦本の手法を応用して開発された ARToolKit の影響が大きい。ARToolKit とは、PC に接続されたカメラでマーカを撮影した画像から、カメラとマーカの位置関係を実時間で推定し、仮想物体をマーカ上に合成する公開ソフトウェアである。この公開により、誰もが AR 環境を簡単に構築することが可能となった。また、このことから AR 研究に取り組む研究者が増えたことなどから、ARToolKit の公開は当該分野の研究が飛躍的に加速するきっかけとなったと考えられる。その後、この正方マーカをベースとしてさまざまな研究がすすめられた。その代表的なものに、図-4(b) に示す多数のマーカを利用して位置合わせを行う ARTag がある。これは、非常に多くのマーカを利用することで、マーカの隠蔽による位置合わせ失敗を回避し、またその冗長性から安定性を向上させたものである。一方、マーカを利用しない手法として、図-4(c) に示す仮想立体絵本などがある。これは明示的にマーカを利用するのではなく、テクスチャを



(a) Matrix

1つのマーカを利用したAR



(b) ARTag

複数のマーカを利用したAR



(c) 仮想立体絵本

マーカを用いないAR

図-4 カメラを利用したさまざまな AR

マーカの代わりに利用することで位置合わせを実現している。

さらに、近年の計算機の高速化とともに増えてきたのが人工的に作成されたパターンを用いず、もともと現実環境中にある自然特徴点を利用する方法である。代表的なものは、PTAM (Parallel Tracking and Mapping for Small AR Workspaces)²⁾ という、画像の特徴点をフレーム間で追跡し、カメラの位置を推定するものである。これもソフトウェアが公開されているため、発表後これを利用した研究が見られるようになってきた。しかし、PTAMは初期の数フレーム間で現実環境中の代表的な平面を検出し、その平面上に仮想物体を合成するにとどまっており、実際のアプリケーションを想定した場合には、仮想物体の合成位置の基準となる世界座標系との位置関係を知る手段が必要となる。あらかじめ、現実環境のシーンの自然特徴点を収集して作成したランドマークデータベースを準備しておき、現在のユーザ視点から撮影した画像から検出されたランドマークと対応させることで、世界座標系における位置推定を実現する手法も提案されている³⁾。

これまで紹介したカメラを用いたARのための位置合わせ手法は、ユーザの視点付近に取り付けられたカメラを想定したものであったが、この方式はユーザから外を見るという意味でInside-out方式と言われる。ビデオシースルーディスプレイで背景として利用される映像を位置合わせに用いることができ、機器構成を簡単化できるといった利点があるため、カメラを利用した位置合わせのほとんどはこの方式である。一方、外界からユーザを撮影するOutside-in方式で位置合わせを行う手法もある。この手法は、マーカがインフラとなる前者とは異なり現実環境に固定されたカメラがインフラとなり、そこに映るユーザの視点の位置・姿勢を推定することになる。

--- もう1つの幾何学的整合性：隠蔽関係の実現 ---

ARにおいて幾何学的整合性というとは一般にはユーザ視点の位置・姿勢を推定する幾何学的な位置合わせ問題が主に挙げられるが、もう1つの課題として現実物体と仮想物体の隠蔽表現がある。ARにおいてシースルーディスプレイでは、原則仮想物体は実シーンに上書き合成される。そのため、たとえ仮想物体と視点の間に実物体があったとしてもその上から仮想物体は上書きされるため、正確な隠蔽関係は表現できない。これを解決するには、現実環境の3次元形状が必要となるが、一般に動的な環境を想定しているARでは、実時間でそれを推定することが要求される。従来、ビデオシースルーの左右のカメラ映像を利用してステレオ視によりシーンの形状を推定し正確な隠蔽表現を推定した手法が提案されている。

また、仮想物体と現実物体の干渉なども同様に課題として挙げられる。具体的には、仮想物体と現実物体の箱があった場合に仮想物体が実物体に衝突したような場合である。衝突判定そのものは、現実環境の形状を計測すれば可能であるが、仮想物体に押された場合、実物体がそれに応じて移動することが望ましいが、現段階ではこれを解決するのは困難である。またその逆に、仮想物体が実物体に衝突した際の反力の再現なども解決されていない。

ARにおける光学的整合性

光学的整合性は、実物体と仮想物体の陰影や画質を一致させることを意味している。先に述べた光学的整合性を含む3つの整合性はARを実現する際の一般的な課題とされている。しかし、光学的整合性は仮想物体の写実性を向上することが目的となるため、ARナビゲーションのための矢印や注釈情報を合成するようなアプリケーションの場合、必ずしも実現が必要な整合性ではない。仮想物体があたかも現実シーン中に本物の物体として存



図-5 幾何学的整合性と光学的整合性を実現した AR

在するかのように見せたい場合や、写実性の高い仮想物体を合成したい場合に特に重要となる整合性である。

光学的整合性は、一般に幾何学的整合性が実現された上で必要となる整合性であるため、光学的整合性を AR において実現する試みは、ARToolKit に代表される幾何学的位置合わせ手法が確立されてきた 2000 年ごろから本格的に議論されるようになった。まず、現実環境と仮想物体の陰影表現の一致に関する手法が提案され、その後 2005 年ごろから現実環境と仮想環境の画質の一致を試みる研究が見られるようになってきた。以下に、光学的整合性の中で重要な陰影表現と画質の一致に関してそれぞれ述べる。

--- 光学的整合性：陰影表現の一致 ---

光学的整合性の中でも陰影の整合性は、あたかも物体がそこにあるように見せたい場合に、解決すべき重要な課題である。図-5 に現実の机の上に仮想物体としてティーポットを合成した例を示す。机の上に置かれたマーカーにより位置合わせを行った中央の図では、幾何学的な位置合わせは正確に行われているが、机の上にティーポットが置いてあるように見えない。これは、横にある現実物体のコップと仮想物体の陰影に違いがあることから分かるように、仮想物体の正確な陰影が表現されていないからである。右図のように陰影を付加することにより机の上ののっているように見ることが出来る。

現実物体と同様の陰影を仮想物体で表現するには、現実環境の光源環境を推定し、その情報をもとに仮想物体をレンダリングすることで実現できる。AR において現実環境の光源環境を推定する手法として、カメラで光源を撮影する手法がよく利用される。最もシンプルな方法では、魚眼レンズを取り付けた広角なカメラを上向きに設置し光源環境を推定するものである。この手法は簡単に上空の光源環境が推定可能であるが、仮想物体への映り込み表現が要求される場合などには、上空だけでなくよりさまざまな方向に存在する光源の情報を獲得する必要がある。それを実現するために、鏡面球に映り込んだ

画像を利用することで、より広範囲な光源情報を獲得する手法がある。しかしこれらの方法は、ビデオシースルーディスプレイで利用されるカメラとは別に光源環境を獲得するためのカメラが必要となり、システムが煩雑になるという欠点がある。そこで、幾何学的整合性のマーカーに鏡面球を取り付けることにより、ビデオシースルーディスプレイで利用するカメラから光源環境を推定する手法が提案されている。近年では、広いダイナミックレンジを持ったハイダイナミックレンジ (HDR) 画像を利用して、より高精度な光源環境の推定を行う試みも見られるようになってきた。

--- 光学的整合性：画質の一致 ---

ビデオシースルーディスプレイで AR を構築した場合、ユーザが見る現実環境は一旦カメラで撮影された映像であるため人間が直接見る現実環境より劣化する。一方、仮想物体はコンピュータグラフィックスでレンダリングされるため劣化がない。そのため、それぞれの映像をそのまま合成すると画質の差が発生し、合成画像に違和感が生じる。この問題を解決するためには、カメラによって撮影される際、どのように映像が劣化するかを推定し、それに合わせて仮想物体のレンダリング結果を劣化させることで画質の整合性を実現することが可能となる。カメラで撮影された画像がどのように劣化しているかを、画像処理技術のボケ推定手法を用いて推定することで実現している⁴⁾。

AR の応用例：ウェアラブル／モバイル AR

近年、計算機の小型化・高性能化に伴い、いつでもどこでも利用可能であるという特徴を持つウェアラブルコンピュータや携帯電話などのモバイル機器が急激に進化してきた。これら機器上で、ユーザの見ているシーン中に情報を付加することで情報提示が可能な AR を実現すると、あらゆる場所でその場所に応じた情報を直感的にユーザに提供することが可能になる。そのため、ナビゲ

ーションなどの応用を想定した AR システムが数多く提案されるようになってきた。

ウェアラブル／モバイル型の AR を実現する際、最も重要な課題となるのは、幾何学的整合性の現実世界と仮想世界の位置合わせである。限られた場所で動作するこれまでの AR システムとは異なり、これら機器による AR はいかに広範囲でユーザの位置・姿勢を推定し続けられるかが最も重要な技術課題となる。一方、これら AR ナビゲーションシステムでは、矢印や注釈など写実性が必要ない仮想物体を合成することが多く、光学的整合性を考慮した例はあまり存在しない。

ユーザが移動するあらゆる場所でユーザ視点の位置・姿勢を 1 つの手法で推定し続けることは現時点では不可能である。そのため場所に依拠していくつかの手法を組み合わせて利用することが想定される。屋外環境では、位置を GPS で姿勢をジャイロなどの姿勢センサで推定するのが一般的である。ユーザが携帯できる小型 GPS では、AR ナビゲーションを実現するためには精度が不十分であるが、位置推定可能範囲が広いことや計測の安定性、価格の観点から考えると代替手段はなく、GPS を利用することが主流となっている。

屋内環境では AR 実現に必要な精度でユーザ位置姿勢を推定可能な決め手となる手法はなく、さまざまな位置推定手法が提案されている。アプローチとしては、通常の幾何学的整合性の位置合わせ問題の解決法と同様である。最も簡単な方法では磁気センサなどの 3 次元センサをインフラとして利用する手法が考えられるが、もともと限られた範囲で利用するものであるため広範囲に拡張するにはコストが問題となる。そこで、安価に広範囲に拡張する方法として、コストが低い画像マーカを利用する方法が考えられる。天井に多くのマーカを設置しそれをカメラで撮影する方法などがあるが、景観を損なうという問題が発生する。そのため、壁紙と同色の再帰性反射材と赤外線カメラを用いることで広範囲において位置推定を行うシステムも開発されている⁵⁾。

一方、広域をユーザが動き回るため上述のユーザの絶対位置を推定する手法のみではなく、相対移動量を推定する手法と組み合わせることで効率的に広範囲における位置推定を行うアプローチがある。このアプローチは、

ユーザの相対移動量を推定することで、絶対位置を計測するためのインフラの設置密度を下げ、コストを軽減することが可能である。ユーザの相対移動量を推定する手法では、ユーザが加速度計や磁気センサなどを装着することで、歩行動作計測を行う手法⁶⁾が代表的である。

AR の今後の展望

AR は当初 VR の延長として研究されていたが、AR 特有のさまざまな研究課題があり、現在では幅広く研究が行われている。また、実時間処理が要求される AR では、技術課題を実時間で処理する必要があるが、すべての問題を同時に解決することは現時点では不可能である。しかしこれまでも、計算機の高速化とともに技術の開発が進みさまざまな問題が解決されてきた。今後も、今まで着手されていない問題の解決に取り組み、さらに多くの研究課題が解決されることを期待したい。

参考文献

- 1) Rekimoto, J.: Matrix: A Realtime Object Identification and Registration Method for Augmented Reality, Proc. of Asia Pacific Computer Human Interaction (APCHI '98) (1998).
- 2) Klein, G. and Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces, In Proc IEEE/ACM International Symposium on Mixed and Augmented Reality (2007).
- 3) 武富貴史, 佐藤智和, 横矢直和: 拡張現実感のための優先度情報を付加した自然特徴点ランドマークデータベースを用いた実時間カメラ位置・姿勢推定, 電子情報通信学会論文誌, Vol. J92-D, No. 8, pp. 1440-1451 (Aug. 2009).
- 4) Okumura, B., Kanbara, M. and Yokoya, N.: Augmented Reality based on Estimation of Defocusing and Motion Blurring from Captured Images, In Proc IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 219-225 (Oct. 2006).
- 5) 中里祐介, 神原誠之, 横矢直和: 不可視マーカを用いた位置・姿勢推定のための環境構築とユーザ位置・姿勢推定システム, 日本バーチャルリアリティ学会論文誌, Vol. 13, No. 2, pp. 257-266 (2008).
- 6) 神原誠之, 濱口明宏, 山中一樹, 横矢直和: 装着した 3 次元磁気センサを用いた歩き・走り状態に対応したユーザ位置の自律計測, 日本バーチャルリアリティ学会論文誌, Vol. 13, No. 4, pp. 439-449 (2008). (平成 22 年 2 月 16 日受付)

神原誠之 (正会員)

kanbara@is.naist.jp

2002 年奈良先端科学技術大学院大学博士後期課程修了。同年同大情報科学研究科助手, 2007 年同大情報科学研究科助教, 現在に至る。複合現実感の研究に従事。博士 (工学)。2002 年電子情報通信学会学術奨励賞受賞, FIT2005 論文賞受賞。電子情報通信学会, 日本 VR 学会, IEEE 各会員。