

VR/MRのための画像計測

佐藤 智和* 横矢 直和*

* 奈良先端科学技術大学院大学, 奈良県生駒市高山町 8916-5
 * Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, Japan
 * E-mail: tomoka-s@is.naist.jp

キーワード：画像計測, バーチャルリアリティ, 複合現実感
 JL 002/02/4202-0086 ©2002 SIC E

1. はじめに

バーチャルリアリティ (VR: Virtual Reality) とは, コンピュータグラフィクス (CG) などを用いて実際には存在しない仮想的な世界を構築する技術の総称であり, ゲーム, 映画などの分野を中心に広く活用され, 一般に広く認知されている. 近年, このような VR の発展・進化形として, 従来から VR で用いられてきた CG などによって構築される仮想世界の映像と, ビデオカメラなどで撮影される現実世界の映像を, 融合してユーザに提示する複合現実感 (MR: Mixed Reality) と呼ばれる技術に注目が集まっている. 図 1 に示すように, MR によって構築される世界は, 現実世界と仮想世界の間には存在する連続的な概念であるが, 技術的には仮想世界の情報を用いて実世界を補強・増強する拡張現実感 (AR: Augmented Reality) と, 実世界の情報を用いて仮想世界を構築する仮想化現実 (AV: Augmented Virtuality) に大別できる. 本解説では, このような現実世界と仮想空間の融合による MR 環境の構築に欠かすことのできない, 現実世界の三次元計測手法について, 画像計測による手法を中心に紹介する.

2. 拡張現実感 (AR) のための画像計測

拡張現実感 (AR) は, ユーザが見ている現実世界の視覚情報に対して, CG などによって表現される仮想物体や注釈情報をリアルタイムで合成することで, 様々な付加情報を提示する技術である. 図 2 は AR を観光案内に利用したシーンの一例を表したイメージ図である¹⁾. ユーザはカメラが設置されたヘッドマウントディスプレイ (HMD) を装着しており, カメラはユーザの視点位置からの画像を撮影している. ここでは, 平城宮跡にいるユーザの視点から撮影された現

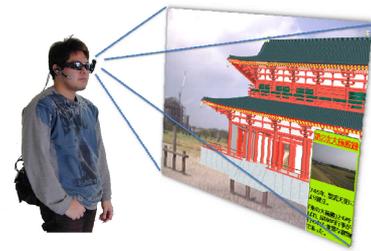


図 2 拡張現実感の利用イメージ (観光案内の例)

実環境の画像と, 現実には存在していない大極殿の CG が合成され, HMD を通してユーザに提示されている. CG はユーザの視点位置に応じてリアルタイムで更新され, あたかもそこに実際の大極殿があるかのように提示される. また, 提示画像中には大極殿に関する注釈情報が付加されている. AR は, このような観光案内や景観シミュレーションの他に, 映画・ゲームなどのエンタテインメント分野や医療分野でも用いられつつある. これらの分野で用いられる AR において, 現実感の高い仮想空間をユーザに提示するためには, 実画像と仮想物体を双方の継ぎ目が分からないように違和感なく合成する必要があり, そのためには特に, 現実環境と仮想環境の間の幾何学的整合性と光学的整合性を確保する必要がある. 以下では, 画像計測を用いたこれらの整合性問題の解決手法を紹介する.

2.1 幾何学的整合性

AR における幾何学的整合性とは, 仮想物体と現実物体の位置関係や隠蔽関係のような, 物体間の幾何学的な関係に関する整合性を指す. このような幾何学的整合性問題を解決するためには, (1) 現実世界と仮想世界の幾何学的な位置合わせ, (2) 現実物体と仮想物体の奥行き関係の再現, が必要である. 以下では, これら二つの課題に対するいくつかのアプローチを紹介する.

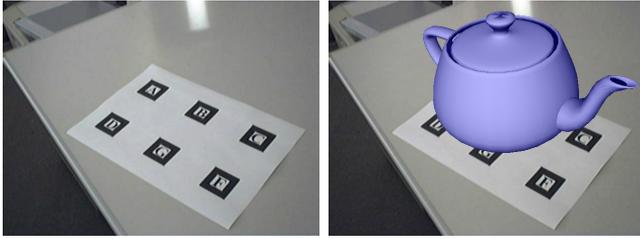
(1) 幾何学的位置合わせ

AR では, ユーザの視点位置がリアルタイムで変化するため, ユーザの視点位置・視線方向に応じた仮想世界 (CG) の描画を行う必要がある. このためには, 現実世界と仮想世界の座標系を一致させた上で CG を描画することが必要であり, この問題は, 現実環境に対するカメラの撮影パラメータを推定する問題に帰着する.

ただし, AR の利用シーンにおいては, カメラの内部パ



図 1 複合現実感が作り出す世界



(a) 実画像 (b) AR 画像

図3 拡張現実感に用いられるマーカと画像合成例

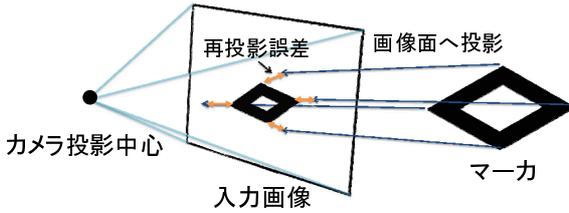
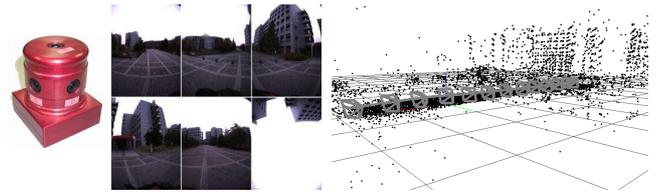


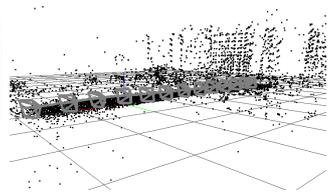
図4 再投影誤差の最小化によるカメラ位置・姿勢の推定

ラメータ(焦点距離, アスペクト比, レンズ歪みなど)は撮影中変化しないことが多いため, 内部パラメータはあらかじめオフラインで計測しておき, 外部パラメータ(カメラの位置・姿勢)のみをリアルタイムで推定することが一般的である. 比較的初期のAR研究では, カメラ位置・姿勢推定のために, 屋内では磁気センサを, 屋外ではGPSと姿勢センサを用いることが一般的であった. しかし, センサによって利用可能な範囲が限定されることや, 装置が複雑になるという問題があり, 近年は画像計測によるアプローチが多数提案されている.

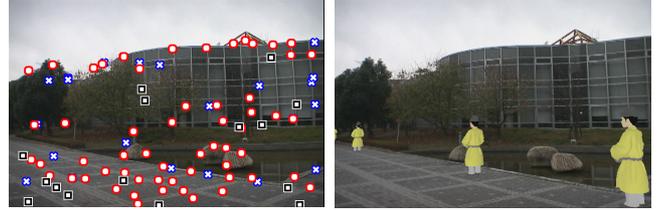
画像計測に基づくカメラ位置・姿勢推定において, 現在最も広く用いられているのは, 形状が既知の人工的なマーカを用いる手法であり, なかでもAR ToolKitと呼ばれる矩形マーカを用いる手法²⁾が世界的に有名である. これら形状が既知のマーカを用いる手法では, 図3(a)に示すようなマーカを環境中にあらかじめ配置しておき, これを画像中から抽出することで, マーカによって張られる座標系に対するカメラの位置・姿勢を推定する. 基本的には, マーカの四隅の点に対する既知の三次元位置と対応する画像座標の組を複数用いて, 図4に示すような再投影誤差の総和を最小とするカメラ位置・姿勢を求め, マーカ座標系とカメラ座標系との幾何変換パラメータを算出する. 推定されたパラメータに基づいてCGをレンダリングすることで, 図3(b)のような合成画像を生成できる. このようなマーカを用いた手法は, カメラ位置・姿勢推定のロバスト性が高く, 計算処理コストも低い扱いやすいという利点があり, また, 仮想物体を描画したい位置にマーカを置くだけで簡単に扱えるという分かりやすさも手伝って広く利用されている. しかし, 屋外などの広域な環境での利用を前提とした場合には, マーカの設置と計測に多大な人的コスト



(a) 全方位カメラと撮影画像



(b) ランドマークの三次元位置



(c) 検出されたランドマーク



(d) 幾何学的位置合わせの結果

図5 自然特徴点ランドマークデータベースを用いた幾何学的位置合わせ

がかかり, また景観を損ねるという問題がある.

これに対して, マーカの代わりに環境中の自然特徴をランドマークとして利用する手法が提案されている. 図5は, 環境の事前知識として自然特徴点の三次元位置と画像テンプレートをデータベース化しておき, これを用いて幾何学的位置合わせを行った例である³⁾. この手法では, 図5(a)に示す全方位カメラを用いて環境内を移動しながら撮影した上で, 動画像中の自然特徴点を自動追跡し, 後述するstructure from motion法によって同図(b)に示すような自然特徴点の三次元位置と全方位カメラの移動パラメータを推定する. これにより得られる自然特徴点をランドマークとし, その三次元位置と画像テンプレートをデータベース化する. 実際の利用シーンでは, ユーザが装着したカメラから撮影される画像上の自然特徴点をデータベース内のランドマークと照合し(同図(c)), マーカを用いる手法と同様に, ランドマークの三次元位置と自然特徴点の画像座標を用いて定義される再投影誤差を最小化することでカメラの位置・姿勢を推定する(同図(d)).

(2) 奥行き隠蔽関係の再現

現実物体と仮想物体の前後関係を再現することも幾何学的整合性問題の重要なテーマである. 図6(a)に示すように, 本来仮想物体を遮蔽するはずの指先が仮想物体により遮蔽されてしまうと, これらの前後関係が正しく再現されず, 違和感のある画像になってしまう. このような問題の解決法としてはいくつかの手法が提案されているが, 撮影対象を静的シーンに限れば, あらかじめ環境の三次元モデルを作成しておき, これを用いて仮想物体の描画をマスクすることで隠蔽関係を再現できる. また動的シーンにおいても, 実時間でのステレオ画像解析により奥行き画像を推定し, 仮想物体と実環境の前後関係を画素ごとに判定する手法が提案されている⁴⁾. シーンを限定すれば比較的容易

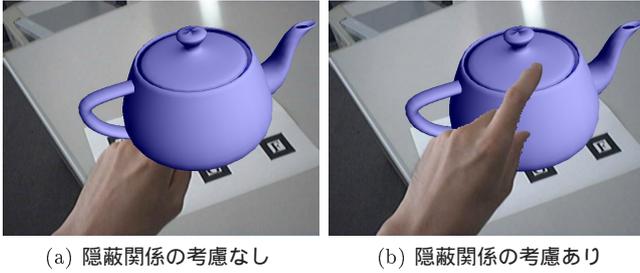
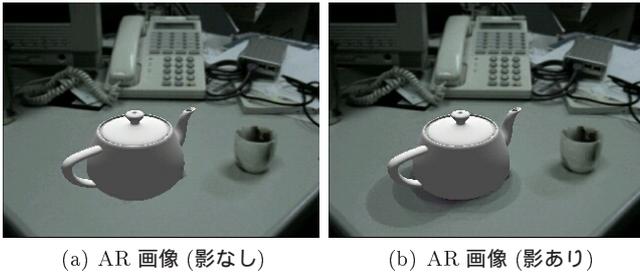
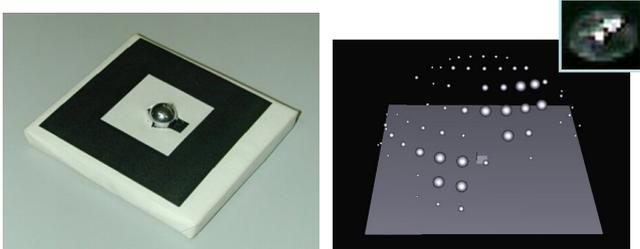


図6 奥行き隠蔽関係の再現



(a) AR 画像 (影なし) (b) AR 画像 (影あり)



(c) 鏡面球内蔵マーカ (d) 光源環境マップ

図7 鏡面球を用いた光源環境の推定と影の再現

に問題を解決できる場合もある。すなわち、図 6(b) では、手は常に CG の手前に存在するという仮定をおき、肌色領域の検出によって抽出された手領域をマスクして仮想物体を描画することで、隠蔽関係を再現している。

2.2 光学的整合性

AR において、描画する CG の画質が現実環境と一致しない場合には、画像上での画質の不連続から現実感が著しく損なわれることがある。例えば、図 7(a) に示すように、仮想物体と現実環境の幾何学的な整合性が保たれていなくても、仮想物体が現実環境に落とす影が存在しないだけで、仮想物体が浮き上がったように見え、違和感が生じる。このような照明条件の不一致などに起因する画質の不連続を解消するためには、現実環境の光源環境を測定し、仮想物体の描画時にそれを再現する必要がある。特に、AR では、リアルタイムで仮想物体と実物体の位置関係が変化するため、光源環境の測定もリアルタイム性が要求される場合が多い。

図 7 は、鏡面球を用いることで光源環境を計測し、陰影を再現した例である⁵⁾。この手法では、幾何学的整合性の解決に用いられるマーカを中心に鏡面球を埋め込み(図 7(c))、鏡面球に写りこんだ画像から環境の光源マップを推定して

いる(図 7(d))。光源マップを用いて仮想物体を描画することで、同図 (b) に示すように、物体表面の陰影や現実環境に落とす影を再現することができる。物体の影を正しく再現するためには、現実環境の三次元的な形状情報が必要となるが、この例では仮想物体が平面上に設置されていることを想定した影の描画となっている。

光学的整合性として、仮想物体の陰影や影を再現するだけでは不十分な場合もあり、更に写実性を高めるためには、仮想物体と実物体の間の相互の写りこみや影付けを考慮する必要がある。また、この場合には、光源環境の測定や現実環境の三次元計測に加えて現実環境の表面反射特性の推定が必要となる。

3. 仮想化現実 (AV) のための画像計測

仮想化現実 (AV) は、現実環境の情報を用いて仮想環境を構築することで、従来用いられてきた CG のみで表現される仮想空間に対して、より写実的な仮想空間を構築する技術である。現在ゲームなどで用いられている仮想空間や映画で利用される CG モデルには、物体表面のテクスチャとして実写画像を用いることで写実性を高める工夫が行われているが、これは最も簡単な部類の AV であるといえる。AV による仮想環境の表現手法は、幾何モデルを用いてシーンを描画するモデルベースドレンダリング (MBR) と、幾何モデルを直接用いずにシーンを描画するイメージベースドレンダリング (IBR) に大別できる。以下ではこれらのレンダリング手法と、それを実現するための画像計測手法について解説する。

3.1 モデルベースドレンダリング (MBR)

MBR は、現実世界の三次元形状や光学的特性をモデル化し、これに基づいて仮想世界の描画を行う手法である。MBR では、現実環境の三次元形状を用いるため、通常の CG と同様に任意の視点によるモデルの描画が容易であり、また光学的特性をモデル化した場合には、任意の照明条件による描画も可能となる。

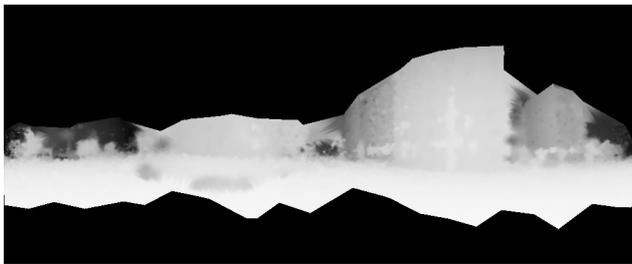
現実環境の三次元形状をモデル化する手法は、レーザレンジファインダなどを用いる能動的な手法と、画像による計測を行う受動的な手法に大別できる。本解説では、画像計測に基づく後者の手法を主に紹介するが、画像計測に基づく手法には、比較的簡単な機器構成で計測を行うことができ、また多くのデータを短時間(少ない人的コスト)で収集できるという特長がある。以下では、代表的な手法について、計測原理を紹介する。

(1) ステレオ法

画像に基づく三次元計測手法として最も基本的な手法である。ステレオ法では、異なる位置に設置された複数台のカメラで現実環境を撮影し、画像間における対応点を探索することで、三角測量の原理により注目画素のカメラからの奥行きを推定する。多数の画像を同時に用いた奥行き推



(a) 入力画像をパノラマ展開した画像



(b) 推定された奥行き画像をパノラマ展開した画像

図8 マルチベースラインステレオ法による奥行き画像推定

定法としては、マルチベースラインステレオ法⁶⁾が有名である。ステレオ法では、三角測量を行うためにカメラの位置・姿勢の関係をあらかじめ測定(校正)しておく必要があるが、カメラの校正には形状が既知のマーカなどが用いられることが多い。しかし、屋外環境のような広域環境では、マーカを用いてカメラの校正を行うことは容易ではないため、現実環境を動画像として移動撮影し、structure from motion(SFM)法によってカメラの位置関係を校正する手法が用いられる。図8は、図5(a)に示した全方位カメラを用いて現実環境を移動撮影し、SFM法によってカメラの校正を行った後に、得られた画像列からマルチベースラインステレオ法で奥行き画像を生成した例である⁷⁾。同図(b)は、推定された奥行き値を輝度値に変換した画像である。移動撮影によって入力画像を取得する手法は、現実環境が静的シーンであることを前提とするが、基本的には奥行き画像を多数の視点から作成し、それらを統合することで環境の三次元モデルを作成することができる。

(2) 視体積交差法

視体積交差法は、物体のシルエットから得られる幾何学的な制約を用いて撮影対象の三次元形状を推定する手法である。すなわち、図9に示すように、撮影位置の異なる多数の画像上で対象物体のシルエット形状を抽出し、各画像上で得られたシルエットを空間中に投影した視体積の交差部分(Visual Hull)を抽出することで、対象形状を推定する。この手法では、シルエットを正確に切りだすために単色の背景が用いられることが多く、また様々な方向からの撮影画像が必要となるため、撮影対象は小物体に限定されることが多い。図10は視体積交差法を用いて三次元モデルを推定した例である⁸⁾。ここでは、同図(a)に示す対象物

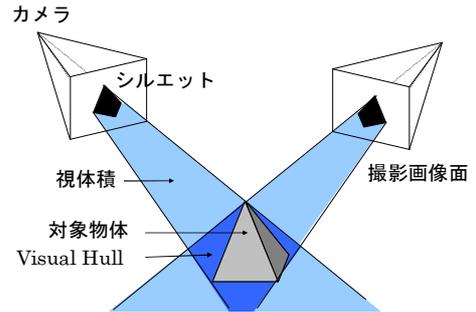
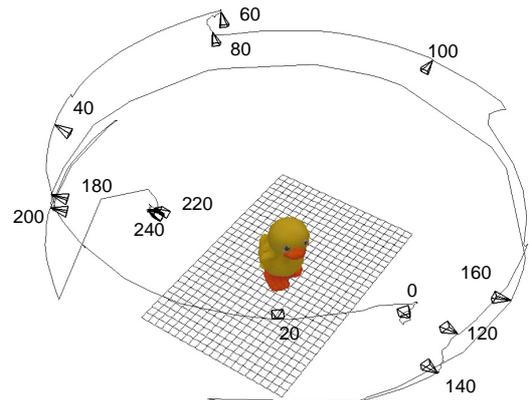


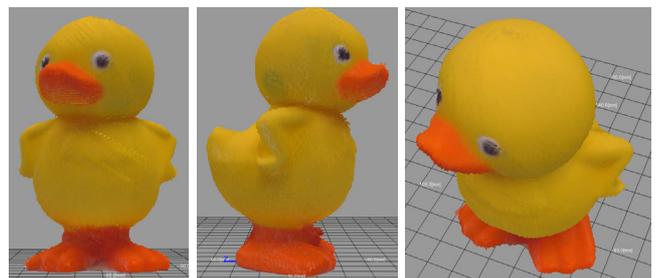
図9 視体積交差法で用いられるシルエット制約



(a) 対象物体



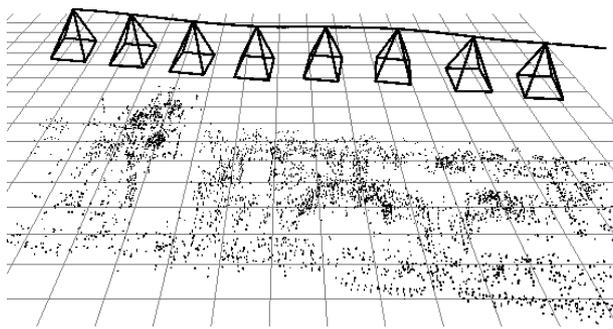
(b) 対象撮影時のカメラ位置・姿勢



(c) 推定された三次元モデル

図10 視体積交差法による三次元モデリング

体を形状が既知のマーカの上に置き、マーカを用いて2.1で紹介した手法によってカメラ位置・姿勢を推定している。図10(b)は、推定されたカメラ位置・姿勢に基づき、カメラの移動経路を図示したものである。また、同図(c)は、抽出されたVisual Hullに対して、各撮影画像から得られる



(a) 推定されたカメラパスと特徴点の三次元位置



(b) 生成された三次元モデル

図 11 Structure from Motion と動的輪郭法による屋外環境の三次元モデル化

テクスチャを張り付けたものである。視体積交差法は、それ単体では原理的に凹物体を復元できないため、復元される三次元物体上のテクスチャと画像上の輝度値の整合性を用いたスペースカーピング法と呼ばれる形状推定手法を組み合わせることで、凹物体にも対応可能な手法が提案されている⁹⁾。

(3) Structure from Motion(SfM) 法

SfM 法は、現実環境を移動撮影することで得られる複数の画像に対して、特徴点の動き情報を抽出することで、特徴点の三次元位置とカメラパラメータを推定する手法である。入力としては、自動で対応点探索を実現することが比較的容易な動画像が用いられることが多いが、複数の静止画像を用いた場合でも対応点が与えられれば SfM 法による復元が可能である。SfM 法によって得られる三次元情報は比較的疎であるが、後処理としてマルチベースラインステレオ法を用いることで、より密な三次元形状を得ることが可能である。また、計測時のカメラ位置・姿勢が同時に推定されるという特徴から、様々な画像計測手法の前処理として用いられる。図 11(a) は、空撮動画像に対して SfM 法を適用し、カメラパラメータと特徴点の三次元位置を推定した結果である。同図 (b) は、空撮データから推定された特徴点の三次元位置に対して、地上からレンジファインダを用いて計測した三次元点群データを位置合わせし、動

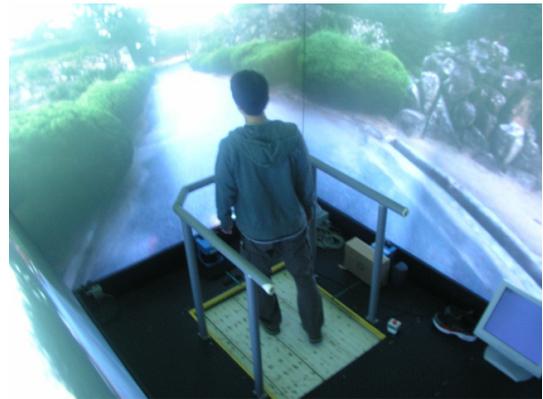


図 12 全方位映像と歩行装置による仮想化現実空間

的輪郭法を用いることで面を構成した屋外環境の三次元復元結果である¹⁰⁾。

3.2 イメージベースドレンダリング (IBR)

MBR では、幾何形状が計測されれば任意の視点での描画が可能となる一方、複雑なシーンに対して正しく幾何形状を得ることは容易ではなく、写実的な画像を生成することが難しいという問題がある。イメージベースドレンダリングは、このような幾何形状を用いず、現実環境を撮影した画像から直接仮想環境を描画する手法であり、写実性の高い仮想空間を構築できるという特徴を持つ。以下では、AV に用いられる IBR の手法について紹介する。

(1) 全方位動画像による画像提示

写実的な仮想環境を構築するための最も簡単な方法は、動画像として撮影された現実環境の映像を加工なしにそのままユーザに提示することである。しかし、単に一般的なビデオカメラで撮影された映像をディスプレイ上に再生するだけでは、視点・視線をユーザが制御できず、仮想世界を体験しているとは言えない。これに対して、全方位カメラを用いて環境の映像を移動撮影しておき、撮影経路上における視点位置(観測フレーム)と視線方向をユーザの要求に応じて変更し提示すれば、視点位置に対して 1 自由度、視線方向に対して 3 自由度の仮想環境を構築することができる。ユーザに提示される画像は、現実空間を撮影した画像そのものであるため、当然写実性は高い。図 12 の例では、観測者を中心として 3 面のスクリーンが設置されており、スクリーンには全方位カメラで撮影された画像が提示されている¹¹⁾。このシステムでは、ユーザは任意の視線方向を観察ことができ、またユーザの足元に設置されている歩行装置によってユーザの歩行を計測することで、ユーザの要求に応じた視点位置の切り替えを実現している。

(2) 視点補間

上記の手法では、視点位置の変更が 1 自由度に限られてしまうという問題がある。これを解決するためには、画像取得時には撮影されなかった新たな視点位置での画像を生成する必要がある。基礎的な手法として従来から用いられ

るのは、離散的な視点位置で撮影された画像に対して画像間をモーフィング処理で補間する手法である。この手法では、2枚の画像A,B間で多数の対応点を与え、生成する画像の視点位置に応じて、対応点の位置をAからBの間で移動させる。同時に対応点周辺の画像をAとBの間で適切に変形・合成することで中間視点の画像を生成する。これにより、視点位置の変更の自由度を高めることができるが、離散的な撮影位置で取得された画像群に対して、自動で対応点探索を行うことは難しく、人手で対応点を与える場合には人的コストが大きという問題がある。また、対応点間で二次元的な補間を行う単純な手法では、中間視点で生成される画像が歪みやすいという問題があり、この問題を回避するためには画像を密に取得しておく必要がある。

(3) 光線空間法

光線空間法は、画像を光線情報の集合として捉え、光線を観測位置と方向によってモデル化することで、任意視点・視線の画像を再現する手法である¹²⁾。この手法では、新たな画像を生成する視点位置に仮想的なカメラを設定し、仮想カメラから投影される光線に最も近い光線を観測データから収集することで画像を生成する。必要な光線情報を画像としてあらかじめ収集しておくことで、任意視点・視線方向の画像を生成することができるが、生成される画像の品質はあらかじめ収集された光線の量に依存する。従って、写実性の高い画像を生成するためには、あらかじめ多数の視点位置において画像を撮影しておく必要がある。

4. むすび

本解説では、VRの進化・発展形である複合現実感(MR)について、技術的要素を解説するとともに、画像計測の応用例を紹介した。MRでは、現実環境と仮想環境の継ぎ目を無くすために現実世界を計測することが必須であり、AR/AV実現のための画像計測の重要性は非常に高い。MRの研究分野は、他の分野に比べて研究の歴史が浅く、各手法ともにまだまだ改良の余地が残されている。このため、現在でも研究が精力的に行われており、今後の発展が期待される。また、カメラ付き携帯電話のような身近な機器を用いたMR研究も行われ始めており、MRは今後より身近な存在になると考えられる。

(一年・月・日受付)

参考文献

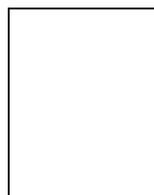
- 1) 天目 隆平, 神原 誠之, 横矢 直和: “「平城宮跡ナビ」マルチメディアコンテンツを利用したモバイル型観光案内システム”, 第1回デジタルコンテンツシンポジウム講演予稿集, No. S3-6, 2005.
- 2) H. Kato and H. Billinghurst: “Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System,” Proc. IEEE/ACM Int. Workshop on Augmented Reality, pp. 85-94, 1999.
- 3) 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”, 日本バーチャルリアリティ学会論文誌, Vol. 10, No. 3, pp. 285-249, 2005.

- 4) 神原 誠之, 大隈 隆史, 竹村 治雄, 横矢 直和: “ビデオスルー型拡張現実感のための実時間ステレオ画像合成”, 電子情報通信学会論文誌(D-II), Vol. J82-D-II, No. 10, pp. 1775-1783, 1999.
- 5) 神原 誠之, 横矢 直和: “現実環境の照明条件と奥行きの実時間推定による仮想物体の陰影表現が可能な拡張現実感”, 画像の認識・理解シンポジウム(MIRU2004)講演論文集, Vol. II, pp. 247-252, 2004.
- 6) M. Okutomi and T. Kanade: “A Multiple-baseline Stereo,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.15, No.4, pp. 353-363, 1993.
- 7) 佐藤 智和, 横矢 直和: “画像特徴点の数え上げによるマルチベースラインステレオ法”, 画像の認識・理解シンポジウム(MIRU2005)講演論文集, pp. 189-196, 2005.
- 8) 不殿 健治, 佐藤 智和, 横矢 直和: “ハンドヘルドビデオカメラを用いた撮影支援インタフェースを有するインタラクティブ三次元モデリングシステム”, 日本バーチャルリアリティ学会論文誌, Vol. 10, No. 4, pp. 599-608, 2005.
- 9) Y. Furukawa and J. Ponce: “Carved Visual Hulls for Image-Based Modeling,” Proc. European Conf. on Computer Vision (ECCV), pp. 564-577, 2006.
- 10) 北市 泰寛, 佐藤 智和, 横矢 直和: “屋外環境の異種計測データを対象とした動的輪郭法による三次元モデル生成”, 計測自動制御学会 第72回パターン計測部会研究会, pp. 21-28, 2007.
- 11) 和田 浩明, 池田 聖, 佐藤 智和, 横矢 直和: “複数経路上を歩行移動可能なテレプレゼンスシステムにおける違和感の低減に関する検討”, 日本バーチャルリアリティ学会 サイバースペースと仮想都市研究会研究報告, Vol. 11, No. 1, pp. 29-34, 2006.
- 12) 苗村 健: “実写とCGの融合”, 映像情報メディア学会誌, Vol. 53, No. 7, pp. 923-926, 1999.

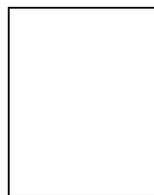
[著者紹介]

佐藤 智和



1999年阪府大・工・情報工卒。2003年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。現在、同大情報科学研究科助教。動画像からの三次元復元、奥行き画像推定、ビデオモザイクの研究に従事。博士(工)。電子情報通信学会、情報処理学会、IEEE各会員。

横矢 直和



1974年阪大・基礎工・情報工卒。1979年同大学院博士後期課程了。同年電子技術総合研究所入所。以来、画像処理ソフトウェア、画像データベース、コンピュータビジョンの研究に従事。1986-87年マッギル大・知能機械研究センター客員教授。1992年奈良先端科学技術大学院大学・情報科学センター教授。現在、同大情報科学研究科教授。電子情報通信学会・情報処理学会フェロー。工博。電子情報通信学会、情報処理学会、日本バーチャルリアリティ学会、人工知能学会、日本認知科学学会、映像情報メディア学会、IEEE、ACM SIGGRAPH各会員。