

[活動報告] TrakMark:

AR/MR カメラトラッキングのベンチマーク

牧田孝嗣*1 林将之*2 蔵田武志*1*2 亀田能成*2

佐藤智和*3 武富貴史*3 柴田史久*4

Abstract --- In the development of Mixed and Augmented Reality (MAR) applications, one of the most important technologies is geometric registration and tracking methods, especially vision-based methods. Since the studies on registration and tracking based on computer vision are getting flourishing and many algorithms have been proposed every year, TrakMark WG was established in 2009 to standardize benchmark schemes that permit objective evaluation and comparison of diverse registration and tracking methods. This paper reports on several topics that have been discussed in the WG such as dataset format, benchmark indicators, and the benchmark processes.

Keywords: geometric registration and tracking method, computer vision, benchmark, standardization, mixed and augmented reality

1 はじめに

近年、現実世界をCGや文字などによって電子的に増強する拡張現実感 (Augmented Reality; AR) や現実世界と仮想世界を長時間で融合する複合現実感 (Mixed Reality; MR) の分野へ注目が集まりつつある。AR/MR 分野の研究課題は数多く存在するが、中でも現実世界と仮想世界の幾何学的整合性を図る方法、すなわち位置合わせ手法 (カメラトラッキングとも呼ばれる) は最重要課題の1つであり、世界中の研究グループが新たな手法を毎年のように発表するなど百花繚乱の様相を呈している[1-3]。

このような状況の下では、提案される様々な手法の利点と欠点を客観的に評価する仕組みが必要となる。客観的評価の仕組みが確立することによって、例えば、AR/MR 技術を利用したアプリケーションを作成する場合には、そのアプリケーションに適した位置合わせ手法が選択でき、新たな位置合わせ手法を研究・開発する場合にも、従来手法との違いを明確化できるようになる。

客観的な評価の仕組みを確立するには、評価する項目を整理し、評価のための基準を定め、評価に用いるテストデータを整備する必要がある。これはすなわち、AR/MR カメラトラッキングのためのベンチマークの策定であり、日本 VR 学会の複合現実感

研究委員会 (SIG-MR) では、その下部組織として、「AR/MR トラッキング手法の評価方法&テストベッド策定に関するワーキンググループ」(通称、TrakMark WG) を2009年5月に設置し、ベンチマークの策定に取り組んできた[4-5]。

TrakMark WG の目標は、AR/MR における位置合わせ及びトラッキング手法に関して、基準となる評価方法を策定し、評価に用いるテストデータを準備した上で、その評価結果を広く国内外に公開することである。TrakMark WG は、SIG-MR 委員とその所属研究室のメンバーを中心に組織化し、国外からも位置合わせ手法を研究する著名な研究者をアドバイザーとして迎えて活動を進めてきた。

本稿では、これまでのWG活動の成果について報告する。以降、2節では、ベンチマーク策定に関連する研究について概観する。3節では、WGで準備したデータセット及び策定した評価指標について述べる。4節では、ベンチマーク環境の事例であるCasperカートリッジについて紹介する。また、5節では、TrakMark WGで策定したベンチマークをISOにおいて標準化する取り組みについて紹介する。

2 関連研究

AR/MR の分野においては、様々なCG位置合わせのためのカメラトラッキング手法が開発されてきたが、標準的に使われるデータセットが存在せず、評価は個々の研究者が作成したデータセットに対して個別に行われてきた。これに対して、近年、カメ

*1 産業技術総合研究所

*2 筑波大学

*3 奈良先端科学技術大学院大学

*4 立命館大学

ラトラッキング手法を同一の枠組みで定量的評価するベンチマークデータセットに関する提案がなされ始めている。

Lieberknecht らはカメラトラッキングにおいてしばしば用いられる平面上のテクスチャトラッキングの性能評価用データセットとして、入力画像系列、カメラ内部パラメータ、追跡対象となるテクスチャ画像、初期化用のフレームのカメラ位置・姿勢情報を公開している[6]。このデータセットには、光源環境の違い、テクスチャの密度の違い、繰り返しパターンの有無など異なる特徴を持つ複数のデータが含まれている。ユーザはこれらの情報を用いて評価対象となるアルゴリズムを実行し、推定された追跡対象の二次元位置情報をメールで送付することで、推定精度の定量的な評価結果を取得することができる。定量評価で用いる真値は、高精度な位置決めが可能なロボットアームの先端にカメラを固定することで取得されている。同文献では、代表的なアルゴリズムとして SIFT, SURF, FERNS, ESM を用いた場合の評価結果が公開されている。

Kurz らは ISMAR 2013 において、屋外環境でのカメラトラッキング手法の評価用データセットを公開した[7]。このデータセットは、スマートフォンなどのセンサ内蔵型のカメラ付き端末を用いたカメラトラッキング手法の定量評価を想定しており、評価用データには、入力画像系列に加えて、各フレームにおける重力方向、コンパス情報、GPS 測位情報が含まれる。また、9 地点で計測されたレンジデータとパノラマカラー画像を用いて作成された三次元モデルも提供されている。定量評価に用いるカメラ位置・姿勢の真値は、カメラ位置の計測と姿勢情報の取得の二段階により作成されている。具体的には、まず地面を格子状に分割し、トータルステーションを用いて各格子点の三次元座標を高精度に計測する。次に、重りを付けた一定の長さの紐をモバイル端末に付け、この紐を用いて計測された格子点の三次元位置からの高さを調整することで端末の三次元位置を取得している。最後に、端末の三次元位置を決定した後、レーザ計測によって作成された環境の三次元エッジモデルに基づくトラッキング処理によって端末の姿勢を推定している。本データセットでは、このようにして得られた端末位置及び姿勢の 6 自由度を真値として扱っている。このデータセットを用いれば、三次元モデルやテクスチャを用いた画像ベースのトラッキング手法だけでなく、センサ情報を融合した推定手法の定量的評価が可能となる。ただし、現状のデータセットでは、シーケンス中にカメラの移動があるような場合には対応できないという制約がある。

Sturm らは Kinect などの RGBD カメラを用いたトラッキング手法の精度評価を目的としたデータセットを公開している[8]。このデータセットには、RGBD カメラで取得されたカラー画像と深度画像及び各フレームに対する RGBD カメラのカメラ位置・姿勢の真値が含まれている。真値はモーションキャプチャシステムを用いて RGBD カメラに取り付けたマーカを追跡することで取得されている。また、文献[8]では、定量評価基準として、局所的な誤差の尺度である相対位置・姿勢推定誤差(Relative Pose Error: PRE)と全体的な誤差の尺度である絶対位置・姿勢推定誤差(Absolute Trajectory Error: ATE)を用いることを提案している。

Martull らは、ステレオ画像評価用データセットとして広く用いられている Tsukuba Stereo Dataset を三次元 CG モデルとして再構築し、様々なカメラパス、照明条件下で生成された CG 映像を公開している[9]。本データセットはシミュレーション環境下で構築されるため、カメラ位置・姿勢だけでなく、各フレームに対する奥行き画像なども利用可能である。

一方、コンピュータビジョンのアルゴリズム評価を目的として作成されたベンチマークデータセットとして、マルチビューステレオによる三次元形状推定結果を評価対象とした Middlebury Multi-View Stereo データセット[10]、車両走行環境を対象とした KITTI Vision Benchmark Suite[11]などが公開されている。これらのプロジェクトの web ページ上では、データセットを公開するだけでなく、様々な研究者により投稿されたアルゴリズムの評価結果を web ページ上で一覧・比較することが可能である。特に KITTI Vision Benchmark Suite は、車両走行環境下に限定されているものの、ステレオ、オプティカルフロー、ビジュアルオドメトリ、物体検出、動物体追跡、道路領域推定など、様々な目的に対応するアルゴリズム評価が可能であり、ビジュアルオドメトリについては車載機器を対象とした AR/MR のアルゴリズム評価にも利用できる場合があると考えられる。ただし、カメラ位置の真値は GPS の精度に依存しており、画素単位での位置合わせが要求されるアルゴリズムの評価には不十分である可能性もある。

3 データセット及び評価指標

ここでは、カメラトラッキング手法の評価に利用するために、WG で準備したデータセットを紹介する。また、AR/MR のためのカメラトラッキングを評価する指標として策定を進めている、仮想点の投影誤差について述べ、ベンチマーク結果の例を示す。



図 1: 映画のセットを用いたデータセット (作成: 立命館大学). 右画像は CG の合成例.



図 2: 屋外環境のデータセット (作成: 奈良先端科学技術大学院大学). 動物体の写り込む画像を含む.



図 3: 仮想化現実モデルを用いたデータセット (作成: 産業技術総合研究所). 左画像は国際会議の会場, 右画像は和食レストラン, にてそれぞれ作成.

[データセットの紹介]

これまで TrakMark WG では, 実画像を用いたデータセット (図 1, 2) と, 仮想化現実モデル (実環境で撮影した画像をテクスチャとして用いて作成された, 写實的 3 次元モデル) の画像を用いたデータセット (図 3) を作成・提供してきた. 以下, それぞれの特性を述べる.

実画像を用いたデータセットは, 実環境においてカメラ画像を撮影して作成し, カメラ位置・姿勢の真値 (参照値) は別途作成している. 真値 (参照値) の具体的な作成方法としては, 人手の作業による画像間の対応点座標データ作成や, 位置センサ・ロータリーエンコーダ等の外付けセンサを利用した計測を用いているため, 様々な環境におけるデータセットの量産は難しい.

一方で, 仮想化現実モデルの画像を用いたデータセットは, 実画像ではないために, ベンチマーキングの信頼性がやや劣る可能性があるものの, 先にカメラ位置・姿勢の真値を作成し, その後にデータセット用の画像を生成するという手順のため, 事後の真値作成が不要である. そのため, 仮想化現実モデルさえ作成しておけば, 様々なデータセットが効率的に量産可能である. また現在, 実画像を用いたデータセットに特性を近づけるために, カメラパスに歩行モーションを導入する機能や, 画像生成に画像

ぼけ (モーションブラー・デフォーカス) を導入する機能を持つツールを開発し, 新たなデータセット作成の準備を進めている.

[評価指標の策定]

カメラトラッキング手法の精度を評価するための一般的指標として, 3 次元空間におけるカメラの位置と姿勢の誤差 (参照値と推定値との差) がある. しかし, AR/MR アプリケーションにおいては, カメラ位置姿勢そのものよりも, 仮想物体をそのカメラに投影した時に生じる表示位置の誤差が重要である.

表示位置の誤差を知るための指標として, 特徴点の再投影誤差がある. これは, 画像中の特徴点の 3 次元位置を推定し, それを画像上へ再投影した際の 2 次元座標系における誤差であり, カメラトラッキング手法では, これを最小化する方法がよく用いられている. しかし, この方法は仮想物体が特徴点の 3 次元位置付近にある場合に有効であり, それ以外の位置にある仮想物体の投影誤差を推定することが難しい. そこで現在, 任意の位置に配置した仮想点の投影誤差をカメラトラッキングの精度評価の指標として導入し, 1 つの指標として策定を進めている [12-13]. 以下, これを投影誤差と呼び, 仮想点の数, 配置方法, 指標の算出方法について述べる.

A: 仮想点の数と配置方法

仮想点の配置方法には, 参照カメラ (真値または参照値の位置・姿勢が与えられたカメラ) からの相対的位置を固定する相対的配置と, 3 次元空間中における仮想点の位置を固定する絶対的配置がある. カメラから一定の距離に仮想物体を置いた場合の投影誤差を評価するには, 相対的配置が適している. そこで, 図 4 に示すように相対的配置を用い, 参照カメラの撮像面に平行で一定の距離にある仮想平面上に格子点を配置し, それを推定カメラに投影することで推定カメラの精度評価を行う.

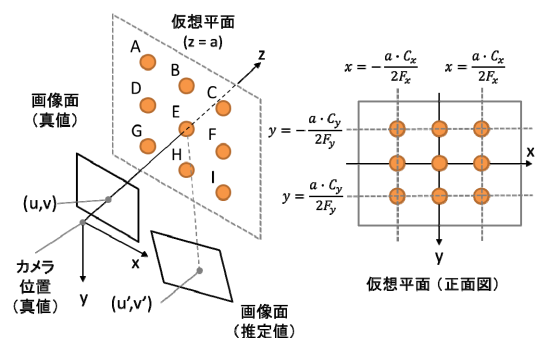


図 4: 仮想点の相対的配置の例 (a はカメラ位置から仮想平面までの距離, C_x , C_y , F_x , F_y はそれぞれ, Zhang のモデル [14] によるカメラ内部パラメータを示す).

B: 指標の算出方法

指標の算出について、以下の2段階が考えられる。

- (1) 可視性評価: 仮想点は可視であり、投影位置が推定カメラのフレーム内にあるか?
- (2) 誤差評価: (1)が真ならば、仮想点を参照カメラへ投影した時の位置と仮想カメラへ投影した時の2次元位置の差を算出し、その大きさを評価する。

誤差評価の指標値は、①画像上のユークリッド距離、②正規化画像座標系(焦点距離 $f = 1$ の平面上)のユークリッド距離、または③光軸との角度差、を用いて計算できる。①の場合、画像の画素数に依存するが、画像の大きさに対する誤差の割合がわかりやすい。図5に、誤差評価の指標値の概要を示す。

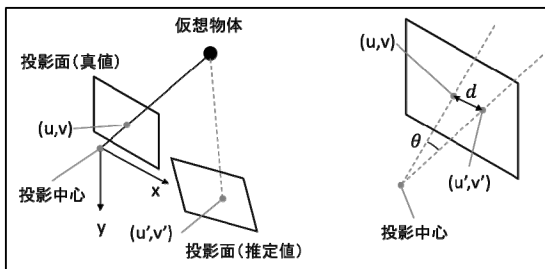


図5: 誤差評価の指標値の概要. d は画像上における距離を示す。一方、 θ は角度を用いた誤差で、大きさが画像の解像度には依存しない。

[ベンチマーク結果の例]

仮想化現実モデルを用いて生成したデータセットを入力として、特徴点追跡を用いたカメラトラッキングを行い、カメラから仮想平面までの距離が $3000[\text{mm}]$ の場合における投影誤差(指標値は画像上のユークリッド距離を用いて計算)を算出した結果を図6に示す。このような、仮想平面上の点の投影誤差の評価を行うことで、ARやMRを想定したカメラトラッキング手法のベンチマーキングは直観的、かつ定量的に実施できる。なお本結果では、同一仮想平面上において、点の位置の違いによる投影誤差のばらつきが見られた。実際、第250フレーム付近以降では、画像の左側(A, D, G)で比較的影響が増大しており、ばらつきがあることが確認できる。この結果から、同一のカメラ位置姿勢の推定値であっても、仮想物体の位置によってその評価にはばらつきがある場合があることが分かる。特徴点を利用したカメラトラッキング手法においては、このような投影誤差のばらつきは、2次元画像上における追跡点数の差によって発生することが多いと考えられる。そこで図7に示すように、画像面を 3×3 の9エリアに分割し、カメラトラッキング時に追跡された点の個数を調査した結果、画像の左側では、第250フレーム以降では追跡点数が極端に減少していることが分かった。このことから、仮想物体の位

置を事前設定できるアプリケーションを対象とした場合、仮想物体の位置に基づいてベンチマーキングを行うことで、複数のトラッキング手法から最適なトラッキング手法の選択することや、トラッキング手法の開発、及びパラメータ等の調整、などが可能であると考えられる。

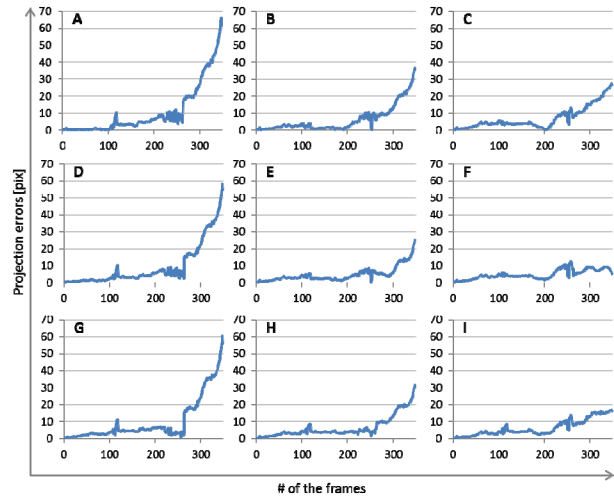


図6: 仮想物体の投影誤差($z=3000[\text{mm}]$)。第250フレーム付近以降では、画像の左側(A, D, G)で比較的影響が増大しており、ばらつきがあることが確認できる。

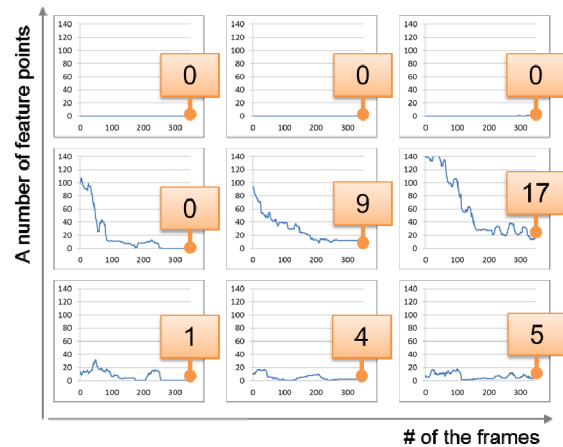
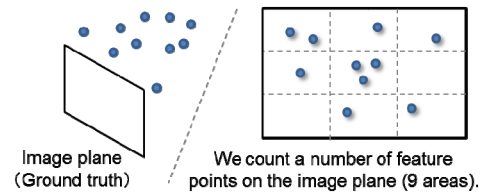


図7: 画像面を 3×3 の9エリアに分割し、カメラトラッキング時に追跡された点の個数を調査した結果(四角形内の数値は、最終フレームにおける点の個数を示す)。

4 ベンチマーク環境の紹介

前述のとおりAR用途での性能評価に適した指標[12-13]が提案されているが、今後、他の指標も提案されてくると考えられる。その指標を如何にして算出するかという実際の問題が研究者の頭を悩ませる

こととなる。一つの解決策は、KITTI Vision Benchmark Suite のように、算出をある研究グループが担う方法であるが、これはそのグループがその任を継続しなくなった途端に利用不能になるという難点がある。

そこで、AR 向けのポータブル開発・評価環境である Casper Cartridge を紹介する[15-17]。Casper Cartridge は Ubuntu Linux Desktop 環境を USB ブートで実現する。この環境内には、TrakMark データセットの利用プログラム例が用意されている。また、それによって自ら求めた外部パラメータと TrakMark が用意している真値とを比較するプログラム例も用意されている。Casper Cartridge は Linux OS ごと無償で公開されていてダウンロード可能なので、利用者は各自の手でそれらのプログラムを実行し、性能評価や検証を行うことができる。なお、本節での説明は、全て Casper Cartridge の Ver.550(CC.Ver.550)を想定している。

【1】TrakMark データセットを読み込めるカメラトラッキングプログラム例(PTAM-TrakMark)

自然特徴を利用したカメラトラッキング手法の著名な実装例として、PTAM[18]がある。その発展版[19]のプログラムコードを、(1)OpenCV ライブラリによる画像入力に対応させ、(2)TrakMark プロジェクトが提供するカメラ内部パラメータを利用できるようにし、かつ(3)PTAM の実行に必要な初期値をパラメータとして実行時に指定、できるようにした改良版が CC-Ver.550 上に用意されている。TrakMark で配布している BD-DVD があればすぐに検証を開始できる。そうでなくとも TrakMark データセット利用へのオンライン登録を済ませていれば、TrakMark 映像をインターネットから当該プログラムに読み込んで実行させることが可能である。図 8 はその実行の様子である。左側が(3)の初期化中の様子、右側がトラッキング中のスナップショットを示している。



図 8: TrakMark 映像での実行例

【2】性能評価指標の算出

3 節 B①で挙げたカメラトラッキングの精度評価の指標 [12-13] を求める参照実装を Casper Cartridge 上で提供している。評価結果は csv 形式で出力される。図 9 での実行で得られたカメラトラ

ッキング推定結果に対して評価指標を算出したものが図 9 である。横軸はフレーム番号、縦軸は pixel、画像は 640×480[pixel]、焦点距離は 415.7[pixel]、設定平面は 1000,2000,3000[mm]の距離にしている。40 フレーム目までは PTAM の初期ベースライン設定をしているので誤差が 0 となっている。

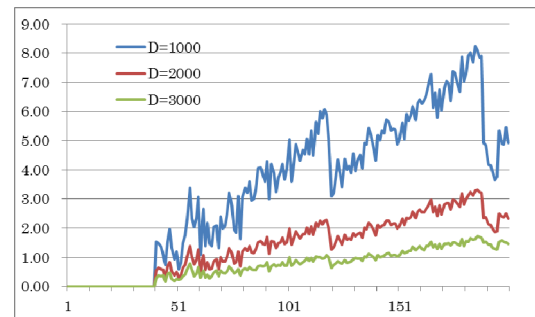


図 9: 投影誤差の計算例

Casper Cartridge では他にも ARToolkit 等の AR プログラム例が、いずれもソースコードからコンパイルして実行できる形式で用意されている。また、AR プログラム開発に有用な OpenCV ライブラリについても同様である。その中には多数のサンプルプログラム(c, c++, python など)、ライブラリリファレンス、チュートリアルといった資料一式が揃っていて、本環境単独で AR に関する研究開発を進めることが可能である。

5 ISO 標準化について

MR や AR は国内外の産学で盛り上がりを見せており、ISO/IEC JTC 1 においてその標準化が開始されている。2011 年 11 月の San Diego での JTC 1 会議では、AR の参照モデル(Reference Model)を中心とした規格に関して、SC 24 がミッションとして行うことが満場一致で可決されたことの報告があった(決議 42)。その結果、WG 9 が発足し、翌年の Rapid City 会議にも参加していたソウル大学の Gerald Kim 氏がコンビーナに就任した。SC 24 は歴史的に CG や VR(VRML)、画像処理に関する標準化に携わっており、現在は X3D の規格策定等が進められている。実環境とバーチャル環境の連続体(R-V 連続体)の概念から捉えると、SC 24 で AR/MR の標準化を進めていくことは妥当であると思われる。

現在、WG9 では、主に、(1)参照モデル(MAR RM: Mixed and Augmented Reality Reference Model)、(2)物理センサのための MAR 参照モジュール、(3)実人物表現のための MAR 参照モジュール 3 つの新規作業項目提案(NP)についての議論が活発に行われている。なお、SC 29/WG 11 も AR の参照モデルを技術報告書(TR)として別途準備していたため、

2012年11月の済州でのJCT1会議の決議34によって、JAHG(SC24/SC29 Joint AdHoc Group)を設置して、ARの参照モデルの提案内容の統一を推し進めることとなった[20].

TrakMarkに関するNPに関しては、その趣旨について、2012年のブリュッセル及び2013年のシドニーのSC24総会にて発表が行われ[21]、2013年12月にSC24国内委員会に提出がなされた。提案は主に、TrakMark WG及び2013年4月に設立されたWG9の国内小委員会での議論に基づいたものとなっており、スコープは以下のようになっている。

データセット

- ・内容：画像シーケンス、カメラの内部/外部パラメータの真値、オプション(特徴点の対応付けの真値、対象や環境の3Dモデルデータ、奥行、他のセンサデータなど)
- ・メタデータ：シナリオ(屋内/屋外ナビ、テーブルトップ、産業AR)、動きのタイプ(平行移動、回転、歩行、ハンドヘルド、車載など)、画質(解像度、焦点ぼけ、被写体ぶれなど)
- ・フォーマット：画像(JPEG, PNG, or Raw)、3Dモデル(X3D, Colladaなど)、真値とメタデータ(XML, X3D, JSONなど)

評価指標

- ・信頼性指標：仮想点の投影誤差、画像特徴の再投影誤差、位置・姿勢の誤差
- ・時間指標：遅延、更新レート(計算コスト)
- ・多様性指標：ベンチマークに使用したデータセットの数やメタデータ(プロパティ)の多様さ

ベンチマークプロセス

- ・評価指標の求め方、ベンチマーク結果の共有方法

6 おわりに

本稿ではTrakMarkに関する状況について概観した。今後は、AR/MRカメラトラッキングのベンチマークの実事例が増え、より効率的な開発、より公平な評価がなされていくことが期待される。

参考文献

- [1] 佐藤, 内山, 田村: “複合現実感における位置合わせ手法”, 日本VR学会論文誌, Vol. 8, No. 2, pp.171-180, 2003.
- [2] 加藤: “AR/MRにおける幾何位置合わせ手法の現状—人工マーカを使った手法の研究動向と課題”, 日本VR学会第14回大会論文集, 2009.
- [3] 佐藤, 横矢: “AR/MRにおける幾何位置合わせ手法の現状—自然特徴を使った手法の研究動向と課題”, 日本VR学会第14回大会論文集, 2009.
- [4] H. Tamura and H. Kato, "Proposal of International Voluntary Activities on Establishing Benchmark Test Schemes for AR/MR Geometric Registration and

- Tracking Methods," ISMAR, pp. 233-236, 2009.
- [5] TrakMark, <http://www.trakmark.net>.
- [6] S. Lieberknecht, S. Benhimane, P. Meier, N. Navab, "A Dataset and Evaluation Methodology for Template-based Tracking Algorithms," ISMAR, pp. 145-151, 2009.
- [7] D. Kurz, P. G. Meier, A. Plopski, G. Klinker, "An Outdoor Ground Truth Evaluation Dataset for Sensor-Aided Visual Handheld Camera Localization," ISMAR, pp. 263-264, 2013.
- [8] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers, "A Benchmark for the Evaluation of RGB-D SLAM Systems," IROS, pp. 573-580, 2012.
- [9] S. Martull, M. P. Martorell, K. Fukui, "Realistic CG Stereo Image Dataset with Ground Truth Disparity Maps," International Workshop on TrakMark (in conjunction with ICPR2012), pp. 40-43, 2012.
- [10] S. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," CVPR, vol. 1, pp. 519-526, 2006.
- [11] A. Geiger, P. Lenz, R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," CVPR, 8pages, 2012.
- [12] K. Makita, T. Okuma, T. Ishikawa, L. Nigay, T. Kurata, "Virtualized reality model-based benchmarking of AR/MR camera tracking methods in TrakMark," ISMAR 2012 Workshop on Tracking Methods and Applications (TMA), 4 pages, 2012.
- [13] 林, 北原, 亀田, 大田, "シーン中の3次元点群の投影誤差によるAR/MR向けカメラキャリブレーションの精度評価", 電子情報通信学会技術研究報告MVE, vol.112, no.221, pp.43-48, 2012.
- [14] Z. Zhang, "A flexible new technique for camera calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), pp.1330-1334, 2000.
- [15] 林, 北原, 亀田, 大田, "AR/MRデモンストレーションの再現性を保証するソフトウェア環境の構築—Casper Cartridge プロジェクター", 電子情報通信学会技術研究報告MVE, vol.111, no.235, pp.103-108, 2011.
- [16] Y. Kameda, I. Kitahara, and Y. Ohta, "Uniform Software Environment for AR Performance Evaluation based on USB Boot Linux," International Workshop on TrakMark (in conjunction with ICPR2012), 4 pages, 2012.
- [17] Casper, <http://www.kameda-lab.org/casper/>
- [18] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," ISMAR, pp. 225-234, 2007.
- [19] R. Castle, G. Klein and D. Murray, "Video-rate localization in multiple maps for wearable augmented Reality," ISWC, pp.15-22, 2008.
- [20] 蔵田: “拡張現実の参照モデルとトラッキングのベンチマーク”, 標準活動トピックス, pp.2-5, No.97, 情報技術標準 Newsletter, 2013.
- [21] T. Kurata, "Update Report: Benchmarking of Geometric Registration and Tracking Methods for Mixed and Augmented Reality," ISO/IEC JTC 1/SC 24 N 3543, 2013.