

[招待講演] カメラ位置推定技術を中心とした 複合現実感における基本課題への取り組み

池田 聖[†]

[†] 立命館大学情報理工学部 〒525-8577 滋賀県草津市野路東 1-1-1

E-mail: [†] ikeda.sei.jp@ieee.org

あらまし 本稿では、著者らの過去約5年間の複合現実感に関する研究を紹介する。およそ6年前に今後の複合現実感の研究としては、カメラ位置推定技術そのものの重要性は相対的に小さくなり、むしろ利用者の意図を読み取るようなセンシングや解析こそが重要になると若手研究者どうしていい加減に予言したが、利用者の意図を読み取るうえで最も重要な視線の検出機能を備え、デバイスの位置姿勢推定技術は概ね完ぺきではないかとも研究者の間でも評される Microsoft HoloLens が企業の製品として登場したのをみると、あながち間違っていない予言ともいえる。このいい加減な予言を信じて実施した、「十分な性能のカメラ位置推定技術が前提の応用研究」と「依然として残るカメラ位置推定技術の課題」について説明する。

キーワード カメラ位置推定, 複合現実感, 拡張現実感, テンプレート

1. はじめに

本稿では、著者がこれまで取り組んできた複合現実感(mixed reality, MR)に関する研究を紹介する。特に明確な大方針もなく場当たりに続けてきた研究であるが、強いて何かしらの一貫した考え方があるとすれば、MR において最も基礎的なカメラ位置推定という技術を中心に検討してきたことと、この MR は本来どうあるべきなのかという問いかけを意識していたといふことにあるかもしれない。その切掛けとなったのが 2010 年 9 月日本バーチャルリアリティ学会年次大会の複合現実感研究会のオーガナイズドセッションであった。

このセッションにおいて著者を含む当時“0x20 歳未満の若手”5 名がパネル討論形式で「AR¹ブームの次を見据え、研究社会は何に挑戦すべきか?」というテーマを討論した[1]。時間の制約上ディスカッションの筋書きは予め打ち合わせたものだったが、準備段階において著者らは自分なりにそもそも MR とは何なのか? MR に本当に必要なものは何なのか? 今の MR に足りないものは何なのか? 自分たちは MR この分野の何に貢献できるのか? ということを大真面目に議論した。

著者らの見解では、そもそも MR とは、日常の生活で我々人間が環境内にある様々な物体を見て情報を得るのと同じように、複雑な計算機内の情報を実物体のように見せて利用者をさりげなく支援する技術である。本質的な意味で、この「さりげなく」という条件を満たすためには、情報の表現方法である仮想物体を「実物体のように配置する」ことに注力するのみではなく、利用者の意図を読み取り、利用者の作業を妨げることのない情報提示により支援することだと当時理解した。この理解に基づけば、情報端末と実世界の間の問題で

あるカメラの位置姿勢推定技術は今後相対的に重要性が低くなり、利用者が実世界にどのように働きかけようとしているのかをくみ取るためのセンシングこそ著者らの分野では重要性を増すと予想した。このような考えに至る背景には、洗練されたカメラ位置推定手法である PTAM[2]のソースコードが公開されて、実際にその性能を目の当たりにしたことがある。会場のベテラン研究者らからは、PTAM とて理想的な MR を実現するための性能としては不十分で、そういった考えはまだまだ早いとご批判を受けたが、それでもなお著者ら当時の若手 5 名は、企業が興味をもつ程度の性能に達したことで研究のスピードが加速すると考えた。今、Google Project Tango や Microsoft HoloLens に搭載されるカメラ位置推定技術の性能をみると当時の予測の一部は正しかったと言えなくもない。また、同時期に複合現実感に関する国際会議 ISMAR2010 (International Symposium on Mixed and Augmented Reality)で、複合現実感にどのような研究課題が残されているかという観点で、認知科学系の内容を中心にまとめられたサーベイ論文[3]が発表されたのも上記考えに符合する。

2. 複合現実感におけるカメラ位置推定技術

カメラ位置推定は、複合現実感で最も基礎的な技術である[4]。利用者の眼前の風景(実シーン)に Computer Graphics (CG)によって表現される計算機内の情報を仮想物体としてあたかもそこにあるかのように描くためには、仮想物体を配置しようとする座標系に対するカメラの位置と姿勢を求める必要がある。仮想物体と実シーンとの照明状態を一致させる光学的整合法など他の技術も、このカメラの位置姿勢が既知であることが前提であることが多い。

PTAM などの visual SLAM と呼ばれる技術は、動画像列からカメラ位置とシーン中の点の三次元位置を同

¹ AR (augmented reality)の直訳は拡張現実感で、正確な定義は複合現実感よりも狭いが、MR より AR の方が知られているため、ここではあえて AR という単が用いられた。

時かつリアルタイムに算出する技術である。visual SLAM の原理は、三次元位置が既知のシーン点と画像上の点とが対応付けられるとカメラ位置姿勢が決まる PnP(Perspective-n-points)問題と、複数のフレームのカメラの位置姿勢が既知の時に画像間の点に対応付けられるとその三次元位置が決まる三角測量問題の組み合わせである[5]。PTAM では、初めてキーフレームという概念が導入され、最も計算量の大きい処理であるカメラの位置とシーン点の三次元位置の最適化処理がこのキーフレームのみで行われるため精度を損なわずリアルタイムで動作することが可能になった[6]。

著者らの予想のようにこのカメラ位置推定技術が成熟し、安定して動作することが前提となった時に必要となるはずの技術を考えて研究し、この技術が成熟してもなお必要となるはずのカメラ位置推定技術に関する研究課題を紹介する。

3. カメラ位置推定技術の応用研究課題

3.1. タブレット型 MR の幾何学的整合性

タブレットを用いた MR が実用化もされ、研究としても多数発表されているなかで、著者らが気づいた問題はタブレットのスクリーン上に表示される実シーンとタブレットを介さずに直接利用者が見る実シーンとの間の不整合である(図 1 左)。環境に対するタブレットの位置姿勢とタブレットに対する利用者視点の位置が分かればこの不整合を解決することが可能になる。

著者らは、タブレットの位置姿勢を推定するために広く利用されつつあった visual SLAM が推定する特徴点と視点を結び直線状にスクリーン上の特徴点が存在するように homography 変換する手法を提案した[7-9]。

提案手法は、ビデオレート以上で動作し、単純な仕組みにも関わらず様々なシーンで動作した²ため、今でも自由なタブレットの移動のもと視認性を評価するユーザスタディを実施した唯一のシステムである。

3.2. 個人 AR ディスプレイにおける焦点ボケ軽減

タブレット型 MR の幾何学的整合性の開発が成功し、スクリーン上の実シーンと利用者が直接みる実シーンの幾何学的整合性が解消された提示像(図 1 右)を見て初めて気づいたことは、図 1 のように写真上でみるよりも両者が連続して見えないということであった。前節で説明したシステムでは利用者視点の位置は効き目の位置に設定していたが、スクリーンと実シーンとの間に奥行の違いがあると両眼視と焦点ボケの影響で両者が脳内の像として整合しないことが原因として考えられた。



図 1: 通常が表示(左)と視点依存表示(右)

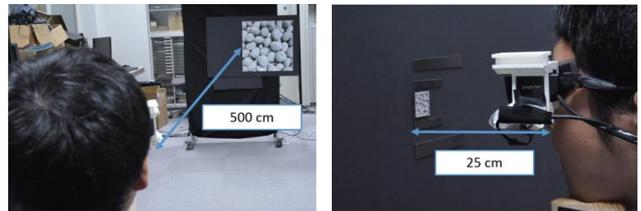


図 2: HMD による提示像と実シーンの焦点ボケの確認実験

両眼視による不整合については裸眼立体視用のディスプレイを用いることで解決できると考えて、焦点ボケによる不整合について、提示像の事前鮮鋭化による解決を試みた。著者らが提案した手法では視線検出器により注視点を検出し、注視点の二次元位置と visual SLAM の三次点群から注視点の奥行きを算出し、加えて視線検出器により瞳孔径をリアルタイム計測し、注視点の奥行きと瞳孔径のサイズに応じて PSF をスクーリングした上で、その PSF (point spread function) を用いて提示像を鮮鋭化し、スクリーンに提示する。

当初タブレットを用いて実験していたが、異なる奥行を同時に視認すること自体が難しかったので、図 2 に示すような光学シースルー型のヘッドマウントディスプレイ(HMD)上に像を提示し、実シーンと比較する実験を行った[10]。その結果注視する奥行と利用者が感じる最適な PSF のサイズの関係が予想通りであったことが確認された。

4. カメラ位置推定に関する研究課題

画像処理によるカメラ位置推定技術が成熟しても、画像特徴の抽出が難しい環境³においては、慣性センサに頼るしかない。この場合精度低下や蓄積誤差を回避できないため、マーカの研究や画像処理のみによるカメラの位置推定精度の評価はなお重要な課題のままであると予想される。

² タブレット型 MR の幾何学的整合性動画

- https://youtu.be/J-M59Fn56_8
 - <https://youtu.be/p0Gc20CZJgw>
 - <https://youtu.be/YUOwGj7EyfE>

³ テクスチャや幾何学的特徴が少ない場合、周辺の物体形状や反射特性が複雑すぎる場合、繰り返しパターンが存在する場合などが該当する。



図 3：小片マーカ（左：点灯なし，中央：点灯あり，右：重畳表示結果）

4.1. 目立たなさと頑健性を両立する小片マーカ

従来のマーカは，カメラ位置推定に目的を絞れば十分に成熟していると言え近年目新しいものは発表されていない．しかし，実用上は複数の問題が残されている．ID 識別と画像特徴の検出の両方に都合の良いパターンを備えていることが多く，この場合パターンを表現するための一定のサイズが必要となるため，個々のマーカの制作精度が要求されるし，コントラストの高いパターンが環境の景観を損なうという問題がある．また，デスクトップ以上のある程度の広さの MR 環境を構築するためにマーカを多数配置するには，各マーカの位置・姿勢を正確に計測し，配置するマーカの ID を管理する必要があった．

そこで著者らは，目立たず，各マーカの位置計測や ID 管理が不要で，マーカの作成自体も容易である実用性の高いマーカを提案している[11]．図 3 に示すように提案マーカはシート状のシールをハサミやカッターで切り刻み，三角形や四角形の薄片にしたものを環境中のランダムな位置に貼りつける．その配置パターン自体を識別用のパターンとして用い[12]，各マーカの三次元位置は特徴点ベースの visual SLAM[13]と同様の原理により推定する．シールを再帰性反射材にし，カメラ側にライトを加えれば頑健な検出が可能であるにも関わらず，他の観察者には目立たず，inpainting 技術と併用することで利用者自身にも目立たないマーカを実現することができる．

4.2. AR/MR 用カメラ位置推定法の性能比較

カメラの位置と姿勢の評価には次の 2 つの難しさがあある．1 つは，各フレームの位置と姿勢の誤差を剛体変換に不変な距離指標で表現することが不可能であること[14]．もう一つは，現実的なシーンの実画像列と十分な精度の真値の組み合わせを準備することが難しいことである．

1 つ目の問題に対しては，複数のフレームを束ねてカメラ位置のみを評価することで対処可能である[15]が，各フレーム単独の誤差は分かりにくくなる．著者らは TrakMark という MR 用途のカメラ位置評価基準の策定に関するプロジェクトでは，仮想物体の画像上の位置ずれを指標としてカメラの位置と姿勢を同時に

評価する指標も提案している．

2 つ目の問題に対して TrakMark やその他の研究グループ[16]でも実写および CG 画像の評価用動画を提供しているが，実写の動画像に真値の精度がセンチメートルレベルのものか，実シーン程複雑な形状の物体が登場しない CG の動画像で真値の誤差は殆ど無いものかのどちらかである．

2 つ目の問題に対しては別な取り組みも行っており，それがコンペ形式のカメラ位置推定精度の評価である．福岡で開催された国際会議 ISMAR2015 では，会議場でカメラ位置推定機能を有するデバイスを現場に持ち寄り，その精度を競いあう Tracking Competition を著者らのグループが開催した．ルールは極めて単純で，スタート地点で座標系を統一した後，複数のチェックポイントであらかじめ与えられた三次元位置にアノテーションなどの仮想物体を表示し，その点をポイントする作業により，カメラの位置・姿勢推定精度を評価するものである．事前取得した環境モデルや環境中に配置するセンサなどを用いずに自律的に動作させることと，デバイスをすべて競技者一人で携帯可能であることが条件である．

こうした評価方法の利点は，複雑な形状の物体を含み，ある程度の大きな環境における実シーンによる評価が可能であること，利用者がある程度システムの性能の限界を踏まえて操作するというより実利用時に近い状況で評価が可能なことである．欠点は，どうしても競技者がデバイスを不自然に動かすことにつながり，カメラ位置の推定精度のみが純粋に評価できないということである．とはいえ，会場での競技は実際には難しく論文等で好成績を収める手法でも，案外安定して動作しないということが露呈するという良さもある．Google Project Tango や Microsoft HoloLens に高性能なカメラ位置推定技術が搭載されていると話題であるが，本年度 9 月 14 日から 16 日にかけて同じく著者らのグループにより企画されている日本バーチャルリアリティ学会年次大会のトラッキングコンペティション 2016 では，こうした最新技術の実際の性能を直感的に体験できるのではないかと期待している．

5. まとめ

本稿では、著者らがここ5年間程度で取り組んできた複合現実感の基礎課題について紹介した。いずれも、カメラ位置推定技術を強く意識したものであったが、2010年から10年後を見据えて設定した研究課題であった。利用者の意図をくみ取るためのセンシングの一つとして利用者の視点位置、注視点位置、瞳孔径を用いたが、さらに高度なことを行うためには調節自体も必要になると考えられる。6年前と同様に、さらに10年後を見据えた研究を考えていきたい。

謝 辞

本研究の一部は、JSPS 科研費基盤研究(B)「符号化開口法を用いたHMDの被写界深度拡大研究課題」(15K12084)および挑戦的萌芽研究「注視センシングに基づく透過型ディスプレイの奥行き知覚整合技術」(15H02737)の助成を受けた。

文 献

- [1] 内山英昭, 一刈良介, 石川智也, 池田聖, 酒田信親: "ARブームの次を見据え, 研究社会は何に挑戦すべきか?", 第15回日本バーチャルリアリティ学会大会論文集, pp. 242-245, 2010.
- [2] Georg Klein and David Murray: "Parallel tracking and mapping for small AR workspaces," Proc. 6th IEEE/ACM Int. Symp. on Mixed and Augmented Reality (ISMAR2007), pp. 225-234, 2007.
- [3] Ernst Kruijff, J. Edward Swan II and Steven Feiner: "Perceptual issues in augmented reality revisited," Proc. 9th IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR2010), pp. 3-12, 2010.
- [4] 蔵田武志, 清川清(監), 大隈隆史(編): "AR(拡張現実)技術の基礎・発展・実践", 科学情報出版, 2015.
- [5] Sei Ikeda, Takafumi Taketomi, Bunyo Okumura, Tomokazu Sato, Masayuki Kanbara, Naokazu Yokoya, and Kunihiko Chihara: "Real-time outdoor pre-visualization method for videographers — Real-time geometric registration using point-based model—," Proc. IEEE Int. Conf. on Multimedia and Expo (ICME2008), pp. 949-952, 2008.
- [6] 内山英昭, 武富貴史, 池田聖: "[特別講演] 拡張現実感における Visual SLAM の技術動向", 電子情報通信学会技術報告, No. PRMU2015-114, pp. 143-148, 2015.
- [7] 富岡誠, 池田聖, 佐藤宏介: "カメラ内蔵タブレット型拡張現実感における実画像幾何補正", 電子情報通信学会技術研究報告, Vol. 112, No. 386, pp. 347-352, 2013.
- [8] Makoto Tomioka, Sei Ikeda and Kosuke Sato: "Approximated user-perspective rendering in tablet-based augmented reality," Proc. IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR2013), pp. 21-28, 2013.
- [9] Makoto Tomioka, Sei Ikeda and Kosuke Sato: "Pseudo-transparent tablet based on 3D feature tracking," Proc. 5th Augmented Human Int. Conf. (AH2014), Article No. 52, 2014.
- [10] Kohei Oshima, Kenneth R. Moser, Damien C. Rompapas, Edward J. Swan II, Sei Ikeda, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor and Hirokazu Kato: "SharpView: Improved clarity of defocused content on optical see-through head-mounted displays," Proc. IEEE Symp. on 3D User Interfaces (3DUI2016), pp. 173-181, 2016.
- [11] Sei Ikeda, Anh Nguyen Trung, Takumi Komae, Fumihisa Shibata and Asako Kimura: "Randomly distributed small chip makers", Proc. 15th IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR2016), 2016 (in press).
- [12] Hideyuki Uchiyama and Hideo Saito: "Random dot markers", Proc. IEEE Virtual Reality (VR) Conf., pp. 35-38, 2011.
- [13] Hideyuki Uchiyama, T. Taketomi S. Ikeda and Joao Paulo Silva do Monte Lima: "Abecedary Tracking and Mapping: A Toolkit for Tracking Competitions", Proc. 14th IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR2015), pp. 198-199, 2015.
- [14] F. C. Park: "Distance Metrics on the Rigid-Body Motions with Applications to Mechanism Design", Jour. Mechanical Design, Vol. 117, No. 1, pp. 48-54, 1995.
- [15] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers: "A Benchmark for the Evaluation of RGB-D SLAM Systems", Proc. IEEE/RSJ Int. Conf. on Intelligent Robot Systems (IROS), pp. 573-580, 2012.
- [16] A. Geiger, P. Lenz, R. Urtasun: "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite", Proc. Conf. on Computer Vision and Pattern Recognition (CVPR), 2012.