

# AR Replay を通じた作業理解の評価

李 云<sup>†</sup> 亀田 能成<sup>‡</sup> 大田 友一<sup>‡</sup>

<sup>†</sup>筑波大学 大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1  
E-mail: <sup>†</sup> s1320898@u.tsukuba.ac.jp, <sup>‡</sup> { kameda, ohta } @iit.tsukuba.ac.jp

**あらまし** 本研究では、実際の作業環境に合わせてチュートリアルビデオ中で教示者の様子を三次元的に再生する AR replay の評価を報告する。我々が提案している AR replay は、まず準備として、教示者の作業がよく見えるように、一台の RGB-D カメラを適宜動かしながら撮影し、記録する。次に、同じ作業環境中で、ユーザは RGB-D カメラを手で構えながら、記録の再生を行う。この時、記録時に構築した作業環境の三次元形状情報を利用することで、現在のカメラ映像の中に、教示者の作業記録を三次元的に正確に重ね合わせる。本稿では、プリンタに対する作業を例として、AR replay の評価実験を行い、作業理解に対する有効性の確認を行った。

**キーワード** RGB-D カメラ, 作業理解, KinectFusion, 三次元ビデオ, ユーザインターフェース, 形状復元

## The Assessment of Efficacy on Using AR Replay to Learn Tasks

Yun LI<sup>†</sup> Yoshinari KAMEDA<sup>‡</sup> and Yuichi OHTA<sup>‡</sup>

<sup>†</sup> Graduate School of SIE, University of Tsukuba Tennoudai 1-1-1 Tsukuba, Ibaraki, 305-8573 Japan  
E-mail: <sup>†</sup> s1320898@u.tsukuba.ac.jp, <sup>‡</sup> { kameda, ohta } @iit.tsukuba.ac.jp

**Abstract** We report the evaluation of our proposed method “AR replay”, which is a framework to record a tutor’s action in a working scene, and then replay the action inside a learner’s view that is taken by an RGB-D camera with his/her hand. On preparation, a tutor’s action is recorded by an RGB-D camera so as to frame it well. Then, on replay, a learner moves an RGB-D camera so that he/she can see the recorded action in the same working scene. The recorded action is precisely aligned to the scene by utilizing the 3D shape of the working scene built at recording. We conducted an evaluation experiment of AR replay on a printer maintaining task. Our advantage for better understanding of tutor’s action is confirmed through the experiment.

**Keyword** RGB-D camera, Learn tasks, KinectFusion, 3D video, User interface, 3D shape reconstruction

### 1. はじめに

機器に対する操作を学習する上で、チュートリアルビデオは効果的である。しかし、ビデオでは撮影された視点からしか操作法を眺められない。そのため、学習者が少し違う視点からその内容を確認しようとすることは不可能である。これに対して、近年に取り組み始めている AR 技術を用いて、実際の作業環境に合わせてチュートリアルビデオ中の教示者の様子を三次元的に再生できれば、操作方法の学習により有効である。

カメラを設置できるスペースが限られている実際の作業環境においては、多数のカメラを配置して様々な方向から作業の様子を記録し、その三次元情報を全て復元する方法は現実的でないことがある。教示者の作業理解ということを考えれば、作業に合わせて一台の手持ちカメラを移動させながら作業の様子を記録しておき、再生時にはそこから多少の視点移動を認めるようにできれば、作業理解に対して十分に効果的であると我々は考える。

そこで我々は、可動式の RGB-D カメラ一台のみを用いた作業現場における教示者の作業の様子の獲得と AR によるその再表示システムである “AR replay” を提案している [1][2]。図 1 は OA 機器の操作の学習をイ



図 1: AR replay のコンセプト

メージしたコンセプトを示している。AR replay では、一台の RGB-D カメラを用いて事前に作業シーンにおける教示者の作業の様子を記録しておくことで、教示者が作業現場にいなくとも、学習者は同じ作業現場に重ね合わせて AR 再表示される教示者の三次元的な作業の様子を見ることで効果的に作業を理解できる。

本稿では、AR replay の有効性を検証するため、実験によりユーザの評価を得た結果を報告する。実験を通じて、AR replay を用いた作業理解が有効であることを示す。また、実験を通じて、現システムの問題点と今後のアプリケーション化に向けた考察を行う。

## 2. 関連研究

AR は作業教示や理解に有用であるため、様々な取り組みがこれまでも成されてきた。例えば、作業環境を作り込むことができるならば、操作対象の仕組みまでよくわかるような提示方法[3][4]等が考えられるが、準備には労力がかかるため、このような手法を広く適用することは難しい。

教示者による作業を理解する上で、教示者の作業様子と作業環境の両方を即時的かつ三次元的に記録し再生できれば、教示現場に学習者があたかもいたかのような効果が得られることになる。しかし、この実現は容易ではない。

シーンの三次元形状の獲得については、完全に自動で取得するアプローチ[5]や手動的なアプローチ[6]が提案されている。しかし、これらの手法では、静的な作業環境の記録を目的としており、教示者の作業様子のような動的なシーンの記録には向いてない。一方で、動的な物体を対象にし、三次元形状の時間的変化を得る研究も取り組まれてきている[7][8]が、この方式を適用するとOA 機器などの形状が変化しない物体の精度がよくないという副作用が見られる。

RGB-D カメラを複数利用し、データを統合することにより、正面の視点と少し異なる視点で三次元映像を見ることが可能となる手法[9]が提案されているが、複数の RGB-D のカメラを利用するため、RGB-D カメラは固定であり、作業環境の範囲も限られている。小範囲の動きには対応できるが、作業者の動きに応じた撮影視点の変更は難しいため、作業者の作業様子の記録には制限がある。

KinectFusion [10]では一台の RGB-D カメラを環境中で動かしていくことで得られる時間軸上のデータを一つの volume に統合することにより、高精度で広い範囲の環境の復元を可能にしている。この研究では動的な物体の検知の可能性も示しているが、静的な物体や環境の三次元形状の復元に焦点においている。我々の手法は KinectFusion の方法をベースにしている。

## 3. AR replay の仕組み

本稿で用いる AR replay のシステムは、[1][2]で報告した方法に基づくものである。本節ではその概要を示す。AR replay は KinectFusion をベースとし、一台の RGB-D カメラのみで、作業シーンの記録と再生の両方を可能にする。

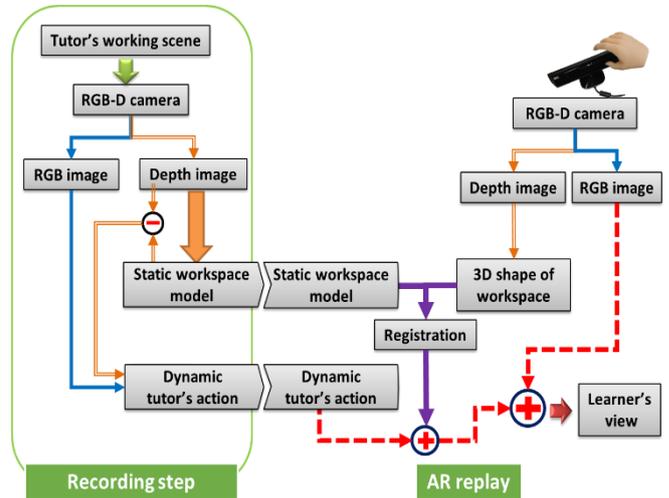


図 2: “AR replay”のシステム構成[2].

図 2 は AR replay のシステム構成である。提案システムは作業シーンの記録と AR replay の 2 つのステップに分かれる。

最初の作業シーンの記録のステップにおいては、まず、RGB-D カメラを用いて入力データを取得する。次に静的な作業環境を表す作業環境モデルとの形状比較により、幾何形状の整合性が取れた静的な要素と、整合性が取れなかった動的な要素に分割する。静的な要素を作業環境モデルに統合していくことで、蓄積されていく静的な作業環境モデルは高精度かつ広範囲に渡るようになる。この部分までは KinectFusion と同じアルゴリズムを利用している。

一方、各フレームの動的な要素を、教示者の作業様子を表す点群とする。

AR replay のステップにおいては、記録時に得られた静的な作業環境モデルを手掛かりに、KinectFusion のアルゴリズムを用いて、環境に対する学習者の手持ちカメラの現在の位置姿勢を求める。その上で、ビデオシーンスルーディスプレイ上に、教示者の作業様子の点群を実際の作業環境に合わせて重畳表示する。

この仕組みによって、AR replay する際には、学習者は、記録時のカメラ位置姿勢から外れてより良い眺めから作業様子を観察できるようになる。

## 4. 評価実験

AR replay の有効性を検証するため、実験によりユー

ザの評価を得る。本実験では、AR replay システムを一般ユーザに体験してもらい、AR replay が作業理解に有効であるかどうかを確認すること及び、通常のチュートリアルビデオと AR replay の間にどのような差異があるかを調べることを目的とする。

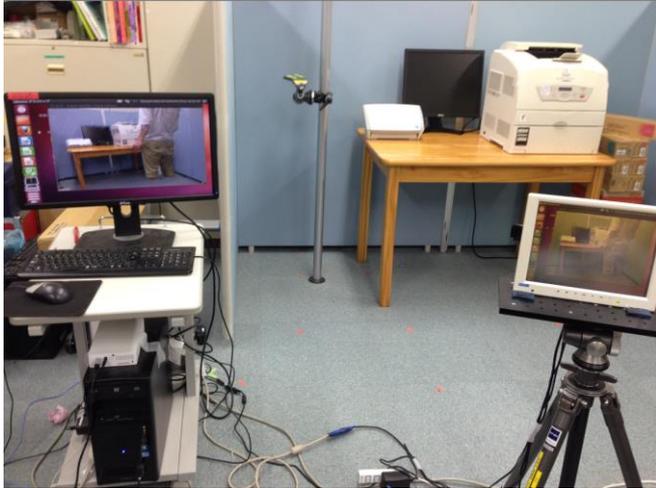


図 3: 実験環境

#### 4.1. システムハードウェア

AR replay システムと実験環境は図 3 のように用意した。右手奥の机の辺りが作業空間で、右側のプリンタが作業対象である。右手前の 10.4 インチディスプレイの裏側には RGB-D カメラが固定されていて、ディスプレイ付きの手持ちカメラとして利用できるようなになっている。左側のモニタと PC が AR replay システムを実行している。

本実験では記録と再生に同じ機材を使用した。RGB-D カメラは Microsoft 社製の Kinect for Xbox 360 である。システム実行用 PC には FRONTIER FRGX916/D (Intel Core i7-3770, 3.40[GHz], 4GB RAM) を使用した。GPU には NVIDIA の GTX660Ti (GDDR5, 2048MB RAM) を利用した。プログラムの開発と実行は Ubuntu Linux [11] 上で行われている。

#### 4.2. 実証システム

##### 4.2.1. 作業シーン記録の実証

まず、実験により、試作システムが静的な作業環境と動的な教示者の作業様子を同時に獲得できることを確認した。実装にあたっては KinectFusion をベースとして、高精度で広範囲に渡る静的な作業環境モデルを得た。結果の一例を図 4 に示す。灰色で描画されている部分が構築された作業環境モデルに相当する。物体の縁にある領域 (強調表示: エッジ) は、RGB-D カメラの映像フレームでデプス情報として得られたものの静的な作業環境モデルには統合されない部分である。

高精度で広範囲に渡る作業環境モデルは、動的な作業様子の点群の取り出しに有利なだけでなく、AR replay する際の作業環境の参照とトラッキングの安定にも有効である。512<sup>3</sup> の分解能で作業環境モデルの volume を構築した場合、フレーム当たり平均 45ms を要した。このとき volume には約 512MB のメモリ領域が必要となる。記録にあたっては、圧縮を施すことにより、およそ 5MB~10MB (モデルの表面形状により異なる) のサイズとなる。

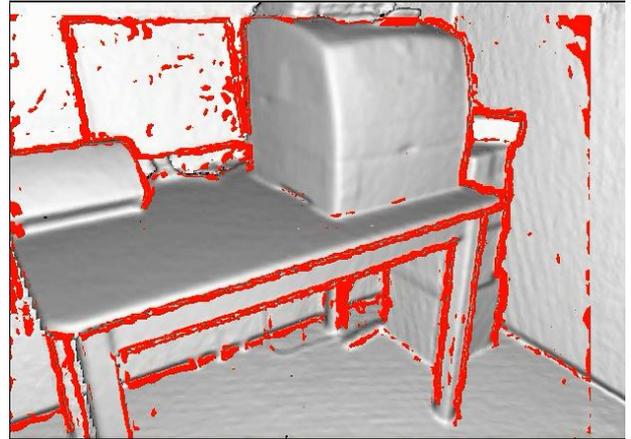


図 4: 獲得された静的な作業環境モデル



図 5: 教示者の作業様子を表す点群  
(上)撮影時の視点 (下)少し移動させた視点

一方、教示者の作業様子を表す点群は連続的に出力され、点群のストリームとして圧縮保存される。動的な作業様子を点群データにより三次元的に表現しているため、撮影時の視点と異なる視点から点群を見ることができる。図5上は記録時にあるフレームで得られた教示者の作業様子を表す点群を、撮影時と同じ視点で見た様子である。図5下はそれから少し移動させた視点から見た結果を示している。図5からもわかる通り、学習者独自の視点で教示者の作業様子を観察することが可能である。

提案手法の制約として、作業の様子を表す点群データを一台の RGB-D カメラにより得ているため、学習者が視点を大きく動かし、記録時のカメラ位置姿勢から視点が大きく離れると、点群データの見かけに欠損が多く含まれてしまうようになることが挙げられる。この問題については、元の記録時において、チュートリアルビデオとしてわかりやすいようにカメラが動かされていくことを前提にしているため、学習者は大きく視点を変える必要を感じないと我々は予想している。

#### 4.2.2. AR replay の実証

蓄積された静的な作業環境モデルは volume として保存される。教示者の作業様子を表す点群は点群ストリームとして保存される。同じ作業環境で教示者の作業様子を再表示する際には、保存された静的な作業環境モデルを読み込み、学習者が構えるカメラを作業環境へレジストレーションする。その上で、教示者の作業様子を現在の作業環境の上に重畳表示する。図6はプリンタのトナーカートリッジ交換を例とし、3つの基本操作における AR replay の結果を示している。図6の上から順次に、ボタンを押す操作、部品を取り出す操作と部品を取り換える操作の結果を表している。図6から教示者の作業様子は正しい位置に再表示され、融合した映像が十分な品質を保持していると考えられる。また、カメラの位置姿勢が常に推定されているため、学習者は適宜カメラを動かし、好みの観察視点にしてより良い眺めの映像を見ることが出来る。

### 4.3. 作業理解に対する有効性の評価実験

#### 4.3.1. 実験手順

提案した AR replay システムを作業理解に対する有効性を検討するため、15名(21~28歳の男性)の実験参加者に対して主観評価実験を行った。図3の作業環境内で、実験参加者は同じ作業内容の2種類の映像を見た後に、アンケートを記入する。作業内容は図6に示すようなプリンタのトナーカートリッジ交換であり、実験参加者はいずれもその作業方法について知識がほぼない状態である。映像の長さは約90秒である。



(上) プリンタのボタンを押す操作



(中) プリンタの部品を取り出す操作



(下) プリンタの部品を取り換える操作

図6: AR replay の実行結果

実験で見せた映像の一つ目は、RGB-D カメラの RGB 部で撮影されたビデオ映像である。実験参加者は映像をディスプレイ上でただ閲覧する。以後、これをチュートリアル映像 (A) と呼ぶ。二つ目は提案した AR replay を利用した映像表示である。実験参加者は Kinect を取り付けたディスプレイを把持し、視点と場所を変えながら、作業様子を閲覧することができる。以降、これを AR replay 映像 (B) と呼ぶ。

### 4.3.2. 評価実験結果と考察

各映像を見終わった直後に行った質問を表 1 に示す。Q1 から Q3 までは、映像 (A) (B) 両方についてそれぞれ評価させることで、それぞれの映像が趣旨に対して十分であるかどうかを確認する。Q4 から Q8 までは、映像 (A) に対して提案方式である映像 (B) が優位と思えるかどうかを尋ねる質問である。

実験参加者が各質問に対して 5 段階評価で回答した結果を図 7 に示す。

Q1 の作業内容の伝わりやすさについては、チュートリアル映像 (A) では、伝わりやすく感じた実験参加者数と伝わりにくいと感じた実験参加者数はほぼ同数であった。AR replay 映像 (B) では伝わりやすくと感じた実験参加者が増えたが、大きな変動とまではいえなかった。実験後の自由意見において、伝わりやすくないと思う原因の大きい要素として、プリンタが構造的に大きく開いて形状が変化したときに手持ちカメラのレジストレーション精度が悪化していたことが挙げられた。この問題については、物体形状が大きく変わるときは静的な作業環境モデルを分割しながら記録する手法を導入するということが対策として考えられる。

Q2 の作業の詳細まで見られるかについての設問の結果からは、映像 (A) ではほとんどの実験参加者が不満に感じていることが伺える。これは、教示者の作業記録時の視点が限定されているため、各学習者が見たいと思う作業の詳細を必ずしも見やすく撮影できていないことが原因と考えられる。このことは、AR replay 映像 (B) では詳細まで見えると思う人が多くなっていることから推測できる。一方、AR replay 映像 (B) でもまだ見にくいと思う実験参加者が 40% いる。この原因について実験後の自由記述で聞いたところ、視点が作業様子の点群に近づくにつれて、AR 部分の映像が薄くなり、結果的に細かい動きがはっきり見えなかったという意見があった。現在の実証システムでは、作業様子の点群が一定の大きさの点のまま表示されているため、学習者の手持ちカメラが対象に近づくにつれて、点群が離散的に見えてしまい、全体像がわかりにくくなる。これに対する改善策としては、カメラの点群に対する距離に合わせて、点群の各点の表示サイズを動的に変更することが考えられる。

表 1 : 質問一覧

映像A: KinectのRGB映像		映像B: AR Replay で再生した点群映像
各質問項目に対して“悪い評価”～“良い評価”5段階で回答		
共同質問	Q1 A & B	今の映像中で作業の内容は、伝わりやすいですか？伝わりにくいですか？(“伝わりにくい”～“伝わりやすい”)
	Q2 A & B	今の映像中で作業の細かいところまで、見られましたか？見られませんでしたか？
	Q3 A & B	今の作業者の動きと機械の位置関係は、見やすかったですか？見づらかったですか？
映像Bのみの質問	Q4 B	映像Aに比べて、今の映像では作業がわかりやすくなったと思いますか？思いませんか？
	Q5 B	使用したインターフェースは、操作はしやすいですか？使いにくいですか？
	Q6 B	このインターフェースの操作は、楽しいですか？楽しくないですか？
	Q7 B	このインターフェースを日常生活で、使いたいですか？使いたくないですか？
	Q8 B	作業理解のためのインターフェースとして、有効だと思いますか？思いませんか？

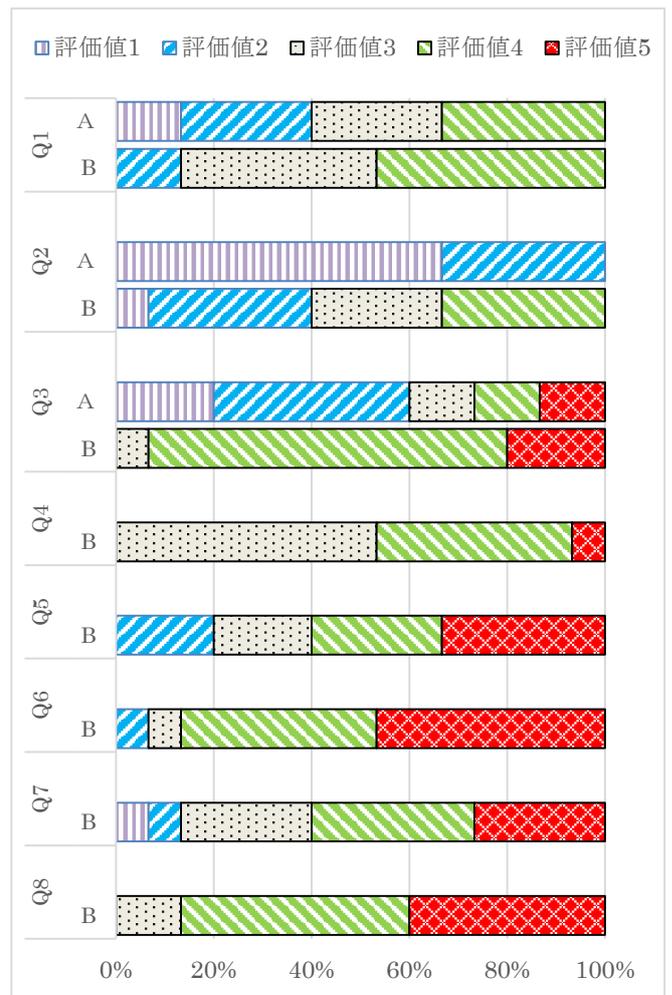


図 7 : 実験参加者による評価

## 謝 辞

本研究の一部は JSPS 科研費 23300064 の助成を受けて行われた。

## 文 献

Q3 の教示者と操作対象であるプリンタの位置関係の見やすさについてはチュートリアル映像 (A) では 60% の実験参加者が見にくいと感じたのに対して, AR replay 映像 (B) では 93% の実験参加者が見やすいと感じた。このことは, 学習者が視点操作を通じて, 教示者と操作対象との位置関係を好みの視点から確認できたためと言える。

これらの Q1~Q3 の評価値を通じて, AR replay 映像 (B) は通常のチュートリアル映像 (A) より作業理解に有効と言える。

Q4 の分かりやすさの向上については, 53% が二つ映像差はなく, 47% は AR replay のほうが見やすいと答えた。この結果は Q1 で映像 (A) (B) の差がそれほど大きくなかったこととも一致する。Q4 で否定的な評価値を選んだ実験参加者がいないことにも注目されたい。

Q5 から Q8 までは, AR replay というインタフェース自体への評価を想定しての質問である。Q5 で操作しやすいと思う実験参加者が 60%, Q6 で操作が楽しいと思う実験参加者が 87%, Q7 で使いたいと思った実験参加者が 60%, Q8 で作業理解に有効だと思った実験参加者が 87% を占めた。

以上により, 全体的には通常のチュートリアルビデオより, AR replay を用いたほうが教示者の作業内容の理解に有効だという傾向が確認できた。

現在の AR replay 実証システムについて, 自由記述の回答では, 「AR replay 映像のノイズがさらに除去できれば, より分かりやすくなる」, 「機械の内部まで変化する際は, 実際の映像と混乱し, 逆に見にくかった」, 「近づける際は, 映像が薄くなるのが気になる」, 「AR 映像の再生フレームレートがもっと欲しい」等の意見が挙げられた。

## 5. おわりに

教示者の作業様子をより良く見ることが出来る AR replay について, 実証システムによる評価実験を行い, その有効性を確認した。

AR replay の実証システムでは, 記録時には静的な作業環境モデルの構築と, 教示者の作業様子を表す点群ストリームとが実用的な速度で獲得できることを示した。また, 記録した情報に基づいて AR replay を実際に学習者が体験できることを確認した。

実証システム上で, プリンタに対する作業の教示を例に評価実験を行い, 通常のチュートリアルビデオよりも AR replay を利用した方が作業理解に有効であるという傾向を確認した。自由記述等から, 実証システムの現段階での改善点も明らかになった。今後は映像の品質の向上とアプリケーションの開発に取り組みたいと考えている。

- [1] Y. Li, Y. Kameda, Y. Ohta, "AR Replay in a Small Workspace," 23th International Conference on Artificial Reality and Telexistence (ICAT), pp.97-101, 2013.
- [2] 李 云, 亀田 能成, 大田 友一, "RGB-D カメラを用いた教示者の作業の AR 再表示," 電子情報通信学会 技術研究報告 MVE, vol.113, no.470, pp.223-227, 2014.
- [3] J. Quarles, S. Lamptang, I. Fischler, P. Fishwick, B. Lok, "Collocated AAR: Augmenting After Action Review with Mixed Reality," International Symposium on Mixed and Augmented Reality (ISMAR), pp. 107-116. 2008.
- [4] A. Kotranza, D. S. Lind, C. M. Pugh, B. Lok, "Real-Time In-Situ Visual Feedback of Task Performance in Mixed Environments for Learning Joint Psychomotor-Cognitive Tasks," International Symposium on Mixed and Augmented Reality (ISMAR), pp. 125-134, 2009.
- [5] Y. Furukawa, B. Curless, S. M. Seitz, R. Szeliski, "Reconstructing Building Interiors from Images," International Conference on Computer Vision (ICCV), pp. 80-87, 2009.
- [6] T. Ishikawa, T. Kalaivani, M. Kourogi, A.P. Gee, W. Mayol, K. Jung, T. Kurata. "In-Situ 3D Indoor Modeler with a Camera and Self-Contained Sensors," Virtual and Mixed Reality, Springer LNCS 5622, pp. 454-464, 2009.
- [7] T. Kanade, P. Rander, P. Narayanan. "Virtualized reality: Constructing virtual worlds from real scenes," IEEE Multimedia, Immersive Telepresence, vol.4, no.1, pp. 34-47, 1997.
- [8] T. Matsuyama, X. jun Wu, T. Takai, S. Nobuhara, "Real-Time 3D Shape Reconstruction, Dynamic 3D Mesh Deformation, and High Fidelity Visualization for 3D Video," Computer Vision and Image Understanding, vol. 96, no.3, pp. 393-434, 2004.
- [9] A. Maimone, H. Fuchs. "Real-time Volumetric 3D Capture of Room-Sized Scenes for Telepresence," 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1-4, 2012.
- [10] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, A. Fitzgibbon, "KinectFusion: Real-time Dense Surface Mapping and Tracking," International Symposium on Mixed and Augmented Reality (ISMAR), pp. 127-136. 2011.
- [11] 亀田 能成, "USB ブータブルな AR/MR プログラム開発・検証環境(Casper Cartridge Project)," 電子情報通信学会会誌, vol.96, no.7, pp.512-515, 2013.