一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report

RGB-D カメラを用いた教示者の作業の AR 再表示

李 云† 亀田 能成‡ 大田 友一‡

†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: †s1320898@u.tsukuba.ac.jp, ‡{ kameda, ohta} @iit.tsukuba.ac.jp

あらまし 教示者がいない作業現場において，作業を学習する時，学習者はチュートリアルビデオを見てビデオ内の教示者の作業と実際の作業環境の対応付けを目視で確認する．本研究では 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 3 次元的に再生することを提案する．本方法では，教示者の作業記録時に一台の RGB-D カメラを手で構えて撮影する．提案手法は KinectFusion をベースとし，それにより記録した作業環境と現在の作業環境のレジストレーション結果を利用することで，学習者に対して同じ作業環境上に教示者の 3 次元的な作業を重ね合わせて AR 再表示を実現する．

キーワード AR 再表示，RGB-D カメラ，教示者の作業，KinectFusion，3 次元ビデオ，形状復元

AR Replay of Tutor's Action by Using Single RGB-D Camera

Yun LI† Yoshinari KAMEDA‡ and Yuichi OHTA‡

† ‡ Graduate School of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan

E-mail: †s1320898@u.tsukuba.ac.jp, ‡{ kameda, ohta} @iit.tsukuba.ac.jp

Abstract We propose “AR replay,” a framework to record a working scene with a tutor’s action, and then replay the tutor’s action in front of a learner’s view in an AR fashion. This framework uses single RGB-D camera for recording and replaying.

On learning a task in a small workspace, when a tutor cannot be in the workspace, it is useful for a learner to check the action of the tutor by a video which was taken in advance in the same workspace. If the video can be replayed in an AR fashion, it will be more useful. Our new “AR replay” method exploits single RGB-D camera. In our “AR replay”, the action of tutor is aligned to the right place and the learner can check the action from various viewpoints. The action is shown as 3D dynamic shape with color and it is aligned to the workspace by the static geometric clues in the workspace.

Keyword AR replay, RGB-D camera, Tutor’s action, KinectFusion, 3D video, 3D shape reconstruction

1. はじめに

教示者がいない作業現場において，作業を学習する時，学習者がチュートリアルビデオを見ることは有用である．この方法では，ビデオ内の教示者の作業と実際の作業環境の対応付けを学習者自身が目視で確認することになる．近年，現実世界にバーチャル情報を重ね合わせて表示する AR 技術の発展につれて，AR 技術を用いた作業支援が取り組まれ始めている．例えば，チュートリアルビデオを用いて教示者の手の動きを AR で再現する研究[1]が提案されており，作業の進行に対する有効性を示している．一方，この取り組みではチュートリアルビデオは 2 次元的に表現され，卓上の手の操作に限定されていた．

卓上の平面に限らず，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 3 次元的に再生できれば，より有効であると我々は考えている．また，既存の機械や設備によりカメラを設置できるスペ

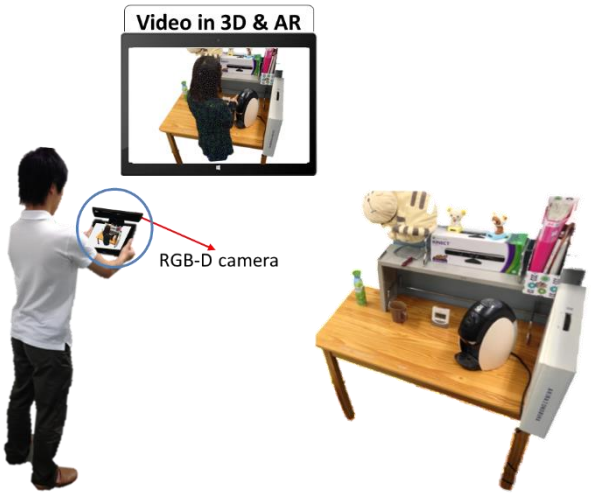

Figure 1: Concept of "AR replay".

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するよりも，1 台のカメラを手にしておき，作業に合わせて注目視点を換えながら作業の様子を記録しておくほうがより効果的であると考えられる．その上で，学習者に多少の視点移動を認める．

そこで，Figure 1 のような，RGB-D カメラ 1 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 による再表示システムを提案する．このシステムは，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き，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教示者の 3 次元的な作業の様子を重ね合わせて AR 再表示できる．

2. 関連研究

作業シーンの記録において，教示者の作業の様子と作業環境の両方を完全な 3 D で記録することが望ましい．しかし，それは容易には実現できない．まず，作業シーンの環境の 3 D 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[2]や手動的なアプローチ[3]が提案されている．しかし，これらの手法では，静的な作業環境を記録することは可能であるが，教示者の作業の様子のように動的なシーンの記録には向いていない．

作業環境と作業者の作業の様子の両方を獲得する研究として，自由視点映像生成技術がある．ビルボードを用いてサッカーのような広い領域を記録する方法[4]や，視体積法[5]を用いたステージ上の 3 D ビデオの獲得する方法がある．これらは空間に設置した複数カメラのデータの同期と統合により，環境と動きの 3 D 形状の同時獲得が可能となっている．しかし，この手法では設備の設置の複雑さ，カメラのキャリブレーションの難しさなどにより，作業現場での設置と記録には向いてない．

デプスセンサを持った Kinect を複数利用し，データの統合することにより，正面の視点と異なる視点で 3 次元映像を見ることを可能とした手法[6，7]は必要とするカメラ数も削減されているが，複数の RGB-D カメラが固定設置され，作業環境の範囲も限られている．極小範囲の動きに適用できるが，作業者の動きに応じた撮影視点の変更ができないため，作業者の作業の様子の記録には制限がある．KinectFusion[8，9]では一台のみの RGB-D カメラを用いて，時間軸上のデータを一つの volume に統合することにより，高精度でより広い範囲の環境の復元が可能となる．また，カメラ一台のみ使用することにより，複雑な設置を必要とせず，撮影の視点も自由度がある．これらの研究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 3 次元形状の復元に焦点においている．

3. 提案手法

提案手法は KinectFusion をベースとし，一台の RGB-D カメラのみ用いて，作業シーンの記録と AR 再生を行う．

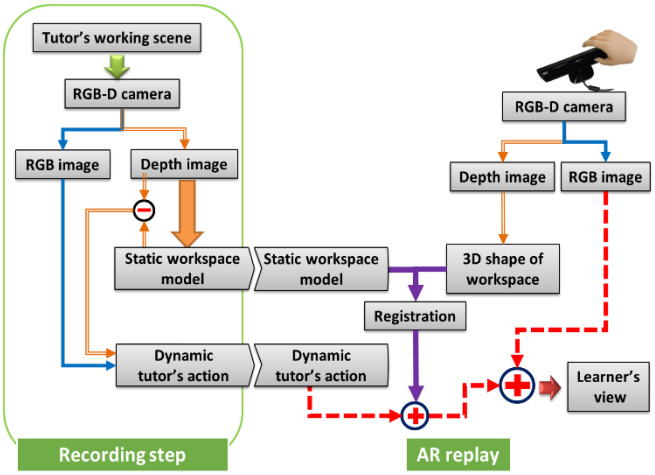

Figure 2: Block diagram of "AR replay".

Figure 2 に提案システムの構成を示す．提案したシステムは大きく作業シーンの記録と AR 表示の 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，蓄積された静的な作業環境モデルを更新する．それと同時に各フレームの動的な要素を、教示者の作業の様子を表す点群とする．

AR 再表示のステップにおいては，まず，記録時に獲得した静的な環境のモデルと実際の作業環境の参照により，カメラの位置姿勢を算出する． その上で，ディスプレイシースルービデオ上で獲得した教示者の作業の様子の点群を実際の作業環境に合わせて重畳表示する．

AR 再表示する際には，学習者の視点の操作により，元のカメラ位置からより良い眺めからその作業の様子及び相互作用を観察し，理解できるようになる．

4. 作業シーンの記録

作業シーンの記録においては，全体的な作業シーンから教示者の作業の様子を分割する必要がある．本研究は Kinect Fusion の手法をベースとして，作業シーン中の静的な作業環境を獲得しながら，シーン中の動的な作業の様子を分割して獲得する．獲得した静的な作

業環境は AR 再表示する際の作業環境との位置合わせにも用いられる．ここで，作業環境は volume データとして取り扱う．一方，動的な作業の様子である教示者の動きは連続的に表示する必要があるため，点群データで取り扱う．

本節では，作業シーンにおける静的な作業環境を表す volume の獲得と動的な教示者の作業の様子の点群の獲得ついて述べる．

作業シーンの記録の際には，先に教示者のいない状態の作業環境を数フレーム記録する．その後，作業する教示者の動きに合わせてカメラを動かしながら，教示者の動作様子を獲得する．この時，KinectFusion より同時に静的な作業環境を表す volume を得ていく．このステップは，具体的に Figure 3 に示す 3 つの Sub-step に分けられる．

Sub-step 1: 入力の獲得：

RGB-D カメラにより RGB 画像と Depth マップを同時に取得する．

Sub-step 2: 静的な作業環境の volume の更新：

Depth マップから得られていた現時点の表面形状を volume の表面形状と比較し，ICP アルゴリズム[10]を用いた反復計算によりカメラの位置姿勢を推定する．そして，推定したカメラの位置姿勢に基づいて，入力マップの各頂点と volume の整合性を求め，距離と角度が閾値以内の頂点を幾何的整合性が取れた ICP Inliers としてマークする．一方，閾値以上の頂点を幾何的整合性が取れなかった ICP Outliers として分離する．ここで前者はフレーム間において形状が変化しなかったものと考えられるため，静的な作業環境の一部と考えられる．ICP Inliers 頂点を，静的な作業環境の volume に統合し，更新することでより広く高精度な作業環境の記述を獲得する．また，関連した色の属性も加え，volume データとして保存する．

Sub-step 3: 動的な作業の様子の点群の獲得：

Depth マップのうち幾何的な整合性がとれない ICP Outliers としてマークされた頂点は，シーン中において形状変化が生じたものと考えられるので，教示者の作業の様子とみなす．教示者の動きにより，作業の様子が常に変化するため．点群データとして扱い，色情報と合わせて，ストリーミングデータとして獲得し，保存する．

以上の処理を各フレームで行うことにより，静的な作業環境の 3 D 形状データを表す volume と，教示者の作業の様子を表す点群のストリームを同時に獲得する．

5. 教示者の作業の様子の AR 再表示

AR 再表示においては，学習者が RGB-D カメラを手で構えて作業環境を見る．同じ作業環境に重ね合わせて教示者の作業の様子を表示するため，再表示の作業環境と予め獲得した教示者の作業の様子の位置対応付けを明確にすることが必要である．そのために，Figure4 に示すように，再度 KinectFusion をベースにして，再表示する際の作業環境の形状と，記録した静的な作業環境の volume を比較し，レジストレーション

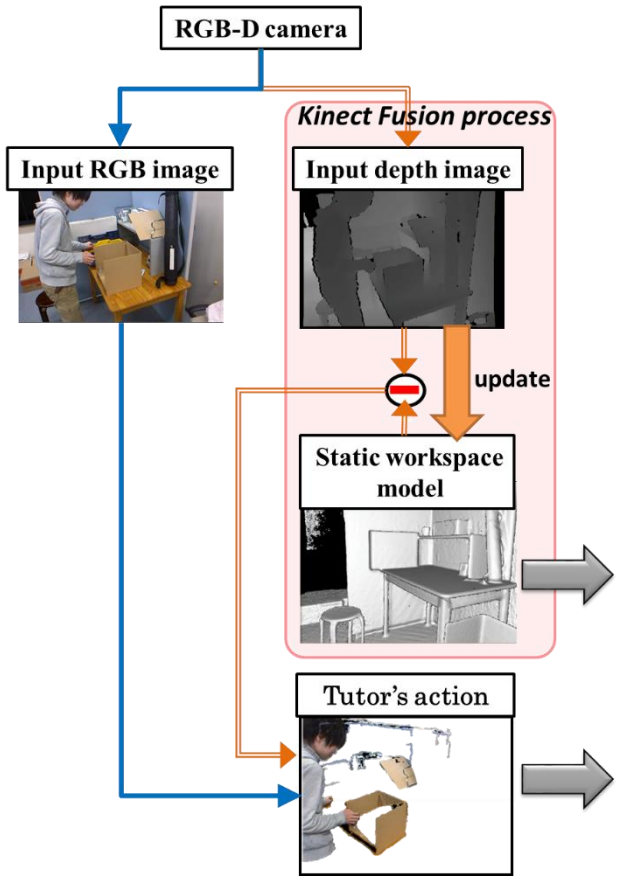

Figure 3: Data flow at recording a working-scene.

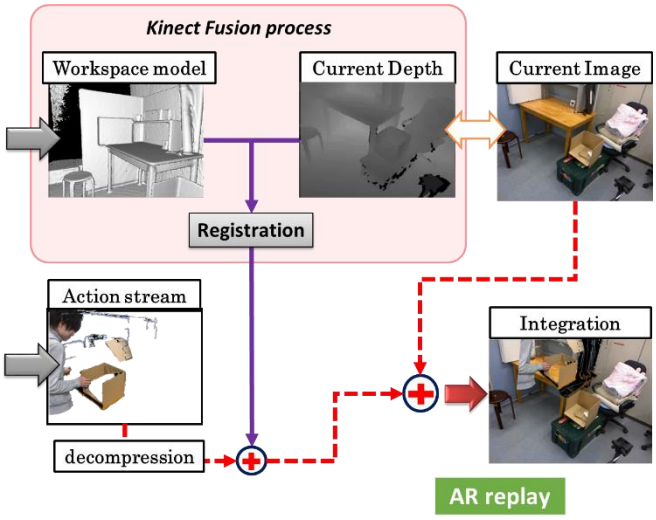

Figure 4: Data flow at AR replay.

を行う.そして,レジストレーションの結果を利用し,獲得した教示者の作業の様子の点群ストリームを受け取り,同じ作業環境の上に教示者の 3 次元的な作業を重ね合わせて AR 再表示する.それにより,学習者は RGB-D カメラを用いながら,ディスプレイシースルーによって記録された教示者の作業の様子を参照することが可能となる.

なお,現時点のシステムでは作業位置を合わせやすいように,学習者は再表示開始時には記録時のカメラ位置と同一地点から作業の様子の映像閲覧を開始するものとする.

6. 実験と考察

4節と5節で紹介した,事前の作業シーンの記録と,AR による教示者の作業の様子の再表示について,各過程で得られている結果とその考察について述べる.

6.1. 記録と再生時の実験環境

本実験では,記録時と再生時に同じ機材を使用する.RGB-D カメラとして Microsoft 社製の Kinect for Xbox 360 を使用した.また,処理に用いた PC の CPU は,Core i7-3770,3.40[GHz]で,メモリは 4GB RAM である.GPU は NVIDIA 製の GTX660Ti (GDDR5 2048MB RAM) を搭載しており,1344 個の CUDA コアをサポートする.プログラム開発は Ubuntu Linux Casper で行った.

6.2. 作業シーンの記録の結果

実験では静的な作業環境と動的な教示者の作業の様子を同時に獲得することに成功した.本実験では KinectFusion をベースとして,広く高精度な静的作業環境の volume (Figure 5) を得た.この処理は平均 1 フレームに約 45ms を要した.512^3 サイズの volume には約 512MB のメモリ領域が必要となる.出力する際には,作成した 圧縮保存関数に渡すことにより,およそ 5MB～10MB (モデルの表面形状により異なる)で出力される.

Figure 5 の左図が記録開始すぐの時点での volume,右図は蓄積獲得したモデルを表すものである.Figure 5 により,シーン中の静的要素の蓄積につれて,静的な作業環境のモデルがより広く,高精度に復元されることがわかる.より広く高精度な静的作業環境を表す volume の獲得は動的な作業の様子の点群の取り出しに有利なだけではなく,AR 再表示する際の作業環境の参照とトラッキングの安定にも有利である.

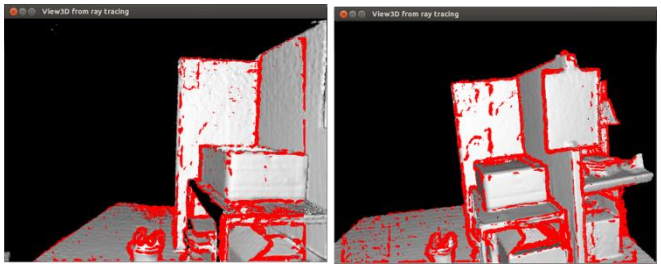

Figure 5: Result of static workspace environment.
(left) initial model, (right) final model (accmulated).

一方,動的な作業の様子を表す点群 (Figure 6) 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.動的な作業の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みることができる.Figure 6 の左図は撮影時と同じ視点で見た点群データ.Figure 6 の右図は撮影時の視点と少し異なる視点から観察する点群を表している.Figure 6 により,学習者独自の観察視点で教示者の作業の様子を観察できることがわかる.

提案手法の制約としては,作業の様子を表す点群データは一つの 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移動すると,点群データの見かけが破綻してしまうことが挙げられる.この問題については,元の記録時からカメラはチュートリアルビデオがわかりやすいように構えられているので,学習者は大きく視点を変える必要がないと我々は考えている.

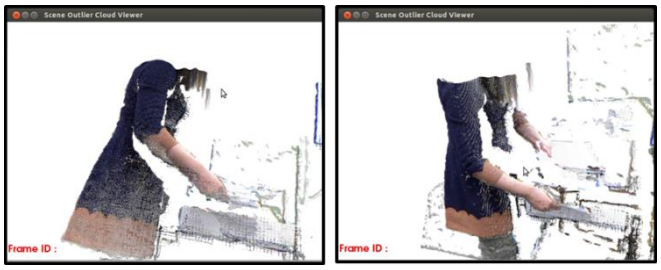

Figure 6: Result of tutor's action.
(left) original viewpoint, (right) different viewpoint.

尚,AR 再表示ではなく,ディスプレイ上のみで作業の様子を閲覧する際には,色付きの静的な作業環境と動的な作業の様子を統合することもできる.統合した映像は 3 次元点群で表現されているので,異なる視点から映像閲覧が可能である.静的な作業環境が継続的に蓄積されるため,元のカメラの位置からは撮影できない広いシーンを表現できる.これは,学習者にとって,教示者の作業の様子と作業環境の相互作用をより良く眺めることや,ズームインアウトすることがここでは可能となり,学習者にとってこれも有意義なものである.

6.3. AR 再表示の結果

蓄積された静的な作業環境は volume として保存される.教示者の作業の様子は点群ストリームとして保存される.同じ作業環境でその教示者の作業の様子を

再表示する際には，保存された静的な作業環境を読み込み，学習者が構えるカメラに対する，作業環境へのレジストレーションを実施し，現在の作業環境の volume と整列させるためのマッチングを行う．その上で，教示者の作業の様子を現在の作業環境の上に重畳表示する．Figure 7(左図が実際の環境，右図が AR 再表示結果)は AR 再表示の結果を示している．Figure 7 に示すように，AR 再表示する際，獲得した教示者の作業の様子は位置が正しく再表示され，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ことから，学習者はカメラを動かし，学習者の観察視点に合わせたより良い眺めの映像を見られる．

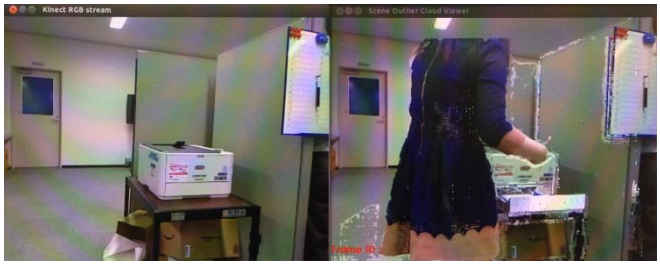

Figure 7: Result of AR replay.
(left) current view, (right) AR view.

7. まとめと展望

教示者の作業の様子の獲得及び AR による同じ作業環境の再現を一体化とした AR 再表示システムを提案した．このシステムは，教示者がいない作業現場における学習者の作業支援を目的としている．事前の作業シーンの記録の段階と AR 再表示の段階には，1 つの RGB-D カメラが利用される．

作業シーンの記録の段階において，我々は KinectFusion アルゴリズムに基づいて動的な教示者の作業の様子を点群ストリームとして，及び静的な蓄積した作業環境を volume として同時に獲得する方法を提案し，実験により獲得結果を確認した．

そして，AR 再表示の段階において，保存した静的な作業環境の volume のロードと，カメラによるトラッキングの安定性を確認した．その後に，現在の作業環境と視点に合わせた作業の様子の合成を確認した．これにより，幾何整合性が取れた教示者の作業の様子を AR 再表示できることを確認した．今後は，AR 再表示する際の品質の向上に取り組むことを考えている．

この手法は通常のビデオ撮影とほぼ同じように手動で 1 台の RGB-D カメラを構えて撮影するだけで，教示者の作業の様子の 3 次元的なデータを捉えられるので，便利なシステムとなりうる．作業支援に限らず，様々な展開の可能性があると考えている．Figure 8 はその一つ例である．このシステムを介して，環境中の物体に合わせるだけで，簡単に二人の自分（現実の自分と AR の自分）の合成動画を生成できる．面白い CG 映像の再生にも活用できるではないかと考えている．

Figure 8: One of the applications of AR replay.
(left) current view, (right) AR view.

謝　辞

本研究の一部は JSPS 科研費　23300064 の助成を受けた，ここで謝意を表す．

文　献

[1] M. Goto, Y. Uematsu et al. “Task support system by displaying instructional video onto AR workspace.” International Symposium on Mixed and Augmented Reality (ISMAR), pp. 83-90, 2010.

[2] Y. Furukawa, B. Curless, S. M. Seitz, R. Szeliski, “Reconstructing Building Interiors from Images.” International Conference on Computer Vision (ICCV), pp. 80-87, 2009.

[3] 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 (HCII2009), LNCS 5622, pp. 454-464, 2009.

[4] T.Koyama, I.Kitahara. Y.Ohta. “Live Mixed-Reality 3D Video in Soccer Stadium.” International Symposium on Mixed and Augmented Reality (ISMAR), pp. 178-186, 2003.

[5] A. Maimone, H. Fuchs. “Encumbrance-Free Telepresence System with Real-Time 3D Capture and Display using Commodity Depth Cameras.” International Symposium on Mixed and Augmented Reality (ISMAR), pp. 137-146, 2011.

[6] A. Maimone, H. Fuchs. “Real-time volumetric 3D capture of room-sized scenes for telepresence.” The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1-4, 2012.

[7] S. Rusinkiewicz,M. Levoy. “Efficient Variants of the ICP Algorithm.” Third International Conference on 3D Digital Imaging and Modeling, pp. 145-152, 2001.

[8] R. A. Newcombe, S. Izadi et al. “KinectFusion: Real-time Dense Surface Mapping and Tracking.” International Symposium on Mixed and Augmented Reality (ISMAR), pp. 127-136. 2011.

[9] S. Izadi, D. Kim, O. Hilliges, R. Newcombe, A. Fitzgibbon, et al. “KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera.” The Symposium on User Interface Software and Technology (UIST), pp. 559-568, 2011.

[10] P. Besl, N. McKay. “A Method for Registration of 3D Shapes.” IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 14(2), pp. 239-256, 1992.