

3D LIVE MAKER: 好きなキャラクタに命を吹き込むまで

岩本 尚也

佐藤 浩之

Bo Zheng

Huawei Technologies Japan K.K.

E-mail: {iwamoto.naoya, sato.hiroyuki1, bozheng.jp}@huawei.com



図 1: 3D LIVE MAKER を使ってぬいぐるみのスキャンをしている様子

1 はじめに

スマートフォン端末の発展及びソーシャルネットワークサービスの台頭によって、ユーザが日常的に様々なコンテンツを目にする機会が増えている。加えて、多種多様なスマートフォンアプリを通じて、ユーザは写真や動画といったコンテンツを能動的に加工、編集することも容易になってきた。近年では、開発者向けに搭載されたフェイシャルやスケルトンのトラッキング、仮想オブジェクトの Augmented Reality(AR) 表示といった機能を用いた様々な AR アプリケーションが開発されており、今後より多くの AR コンテンツに触れる機会が増えていくことが期待される。

AR コンテンツの中でも特にキャラクタをアニメーションさせるようなコンテンツを制作する場合には、形状のモデリングやテクスチャリング、リギング、アニメーションといった作業が必要となるが、動画像の編集とは大きく異なり、3DCG ツールに対する高度なスキルや経験が必要となる。そのため、AR コンテンツを能動的に制作することは未だにユーザにとって敷居が高く、現状フェイシャルやスケルトンのトラッキング機能は事前にアーティストによって設計されたキャラクタへの利用に限られている。

そこで我々は、ユーザが能動的に AR コンテンツを制作できるようにすることを目的とし、キャラクタのモデリング

からテクスチャリング、リギング、アニメーションを包括的に可能にするスマートフォンアプリ「3D LIVE MAKER」を開発した。本アプリケーションは、スマートフォン端末に搭載されたカラーカメラ及びデプスカメラを用いてぬいぐるみのような小物体を三次元スキャンすることでモデリングとテクスチャリングのプロセスを代替し、自動リギングとモーションリターゲッティングを通じてスキャンしたモデルをアニメーションすることができる（図 1）。

本論文では、まず本アプリケーションに関連する研究や既存技術について触れた後、全体的なパイプライン、具体的な開発内容とその結果について述べる。最後に製品発表時の反響や今後の課題について述べる。

2 関連事例

キャラクタを一から作り、動かすまでには、モデリング、テクスチャリング、リギング、アニメーションといった作業が必要となる。近年では Blender をはじめ多くの 3DCG ソフトウェアが手軽に使用できるようになったものの、未だに各作業にはスキルや経験が必要となる。そこで、より直感的にキャラクタをモデリングするための方法として、二次元上に描いたイラストを三次元化する技術が提案されている [1]。特に、Wonder Painter[2] や Rakugaki AR[3] は、

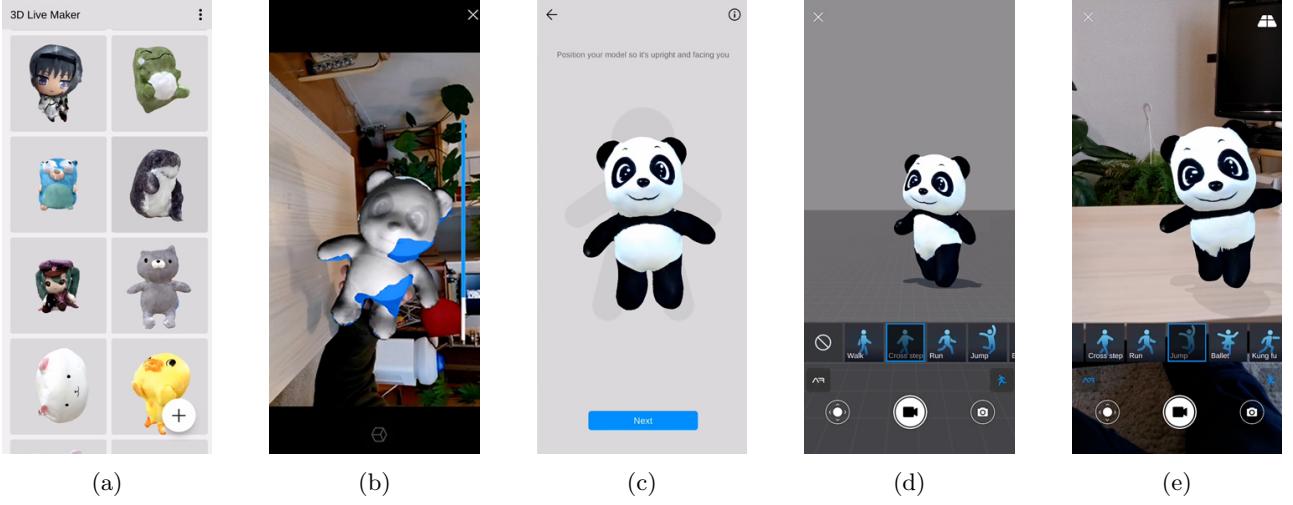


図 2: 3D LIVE MAKER の各処理におけるスクリーンショットを示す. (a) ギャラリー画面. 過去にスキャンしたぬいぐるみの形状を確認可能. 右下の+ボタンでスキャン画面へ遷移. (b) スキャン画面. 詳細は図 3 にて. (c) リギング画面. モデルを回転し, 直立の向きに修正する. Next ボタンを押すことによりリギング開始. (d) バーチャル空間内のアニメーション画面. 各モーションボタンを選択し, アニメーションを確認可能. 左下のコントローラで歩行動作を生成可能. 左下の AR ボタンにて AR 画面へ遷移. (e) AR 画面. 平面を認識後, 同様にアニメーションが可能. 右下のカメラボタンで写真撮影も可能.

スマートフォン端末のカメラで撮影した二次元上のイラストを 3DCG 化し, 自動リギングすることによって, 誰でも簡単に好きなイラストのキャラクタを AR でアニメーションして楽しむことができるようとした. また Magnenat らは, ユーザが色を塗ったイラストをカメラにかざすと, その色が反映された三次元キャラクタを表示するといった体験を提案した [4]. 上記に示した直感的なモデリングやテクスチャリング技術によって, 好みのキャラクタを手軽に生み出すことができるようになりつつある.

二次元のイラストを三次元化する方法以外にも, 実物のオブジェクトを三次元スキャンすることでモデリング及びテクスチャリングを代替する方法も提案されている. 2017 年に発表された 3D Creator[5] は, ユーザがスマートフォン端末のカメラを使用して小物体や人の顔形状をスキャンすることを可能にした. しかしながら, 対象物体を静止させる必要があるため, 物体全周のスキャンが難しいといった課題がある. また近年では, 写真一枚から三次元の形状及びテクスチャを推定する技術も提案されている [6]. これらの手法では, 形状およびテクスチャの推定のために事前に大量のデータを用意する必要があり, 今回のスキャン対象となるぬいぐるみを始めとする多種多様の形状やテクスチャが存在するモデルへの対応は大変困難となる. よって静止物体のスキャンニング及び形状の推定手法は未だに課題が多い.

アーティストが手作業で制作したモデルやスキャンしたモデルをアニメーション可能な状態にするには, リギングと呼ばれるスケルトンの埋め込みとスキンウェイトの計算処理が必要となる. 近年では, Mixamo[7] 等の自動リギン

グサービスを使用することで, 多くのユーザが手軽にリギングからアニメーションを行うことができるようになった. しかしながら Mixamo は手作業による関節の指定が必要になる上にウェブサービスという観点からリギング計算の自動化やパイプライン化が難しいといった課題がある.

我々の開発した 3D LIVE MAKER では, モデリングからアニメーションまでの高度な処理を民主化するために, 上記に示した実物のオブジェクトを三次元スキャンする方法を採用しつつ, 従来課題だった物体全周のスキャンや多種多様な形状, テクスチャのモデルへの対応, 自動リギングのパイプライン化といった問題を解決し, さらにすべての処理をスマートフォン端末上で行うといった過去に前例のないアプリケーションとなっている.

3 3D LIVE MAKER の概要

3D LIVE MAKER はモデリングからリギング, アニメーション, そして AR までのプロセスをスマートフォン端末上で実現することで, ユーザが能動的に AR コンテンツを制作して楽しむことができるようなアプリケーションを目指している. 本アプリケーションが搭載されているスマートフォン端末, Huawei Mate20 Pro[8] はインカメラに通常のカラーカメラに加えて IR プロジェクタと IR カメラからなる Structured Light 方式のデプスカメラが搭載されており, 色と奥行きの情報, RGB-D データをリアルタイムで取得することが可能となっている. 上記によるハードウェアとソフトウェアの組み合わせによってぬいぐるみの三次元スキャンを実現するだけでなく, 自動リギング及び

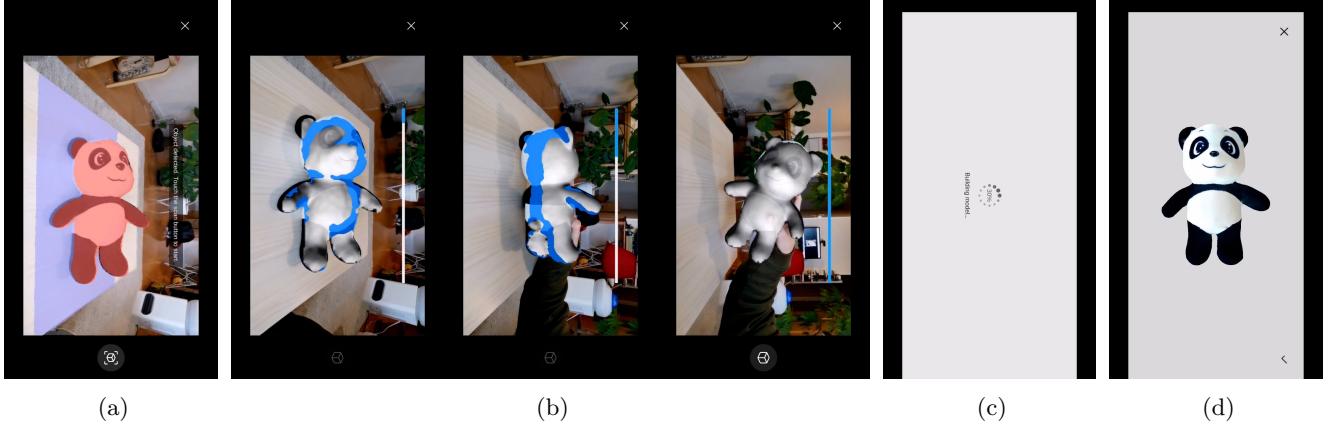


図 3: 3D LIVE MAKER のスキャン処理. スマートフォン端末を横向きにしてスキャンした際の本アプリケーションのスクリーンショットを示す. 紙面の都合で 90° 回転している. (a) 開始処理. 検出した平面が薄青, ぬいぐるみ領域内が薄赤で表示される. (b) オンライン処理. 左から順に開始直後, 処理中, 終了直前である. 推定中のジオメトリのうち十分に撮影されている領域は灰色, まだ撮影が必要な領域は青でオーバーレイされる. 画面上部のバーは進行度を表し一定以上になると終了ボタンが有効になる. (c) オフライン処理. 進行度が表示される. (d) テクスチャ付きメッシュ.

アニメーションリターゲットアルゴリズムを組み込むことで, スマートフォン端末上でスキャンからアニメーションまでを可能にした. 最後に Huawei AR Engine[9] を使用してスキャンしたキャラクタを実世界の平面上に AR 表示してアニメーションすることができるようになっている. 全体の流れを図 2 に示す. 以後, アプリケーションの開発環境, 各アルゴリズムの詳細について順に述べる.

3.1 開発環境

本アプリケーションはぬいぐるみのスキャン, リギング, アニメーション, AR といった機能から構成される. 各機能はそれぞれ高度なアルゴリズム開発を必要とした. アルゴリズム開発をスマートフォン端末上で行うことは非効率であるため, PC で実装やデバッグを行ってからスマートフォン端末で動作確認を行うという流れで開発を進めた. アルゴリズムは C++ で実装し, PC とスマートフォン端末で同じコードが動作するようにした. なおスマートフォン端末では Android NDK を経由し, UI 実装に用いられる Java から C++ のコードを呼ぶことができる. PC とスマートフォン端末で実装が大きく異なるのは画面表示である. PC では OpenGL, Android では OpenGL ES が用いられる. そこで環境に応じて OpenGL/GLES が自動で切り替わるレンダラを開発し, 開発効率を向上させた. 一部の機能はスマートフォン端末上でのみ動作するためスマートフォン端末上で直接実装とデバッグを行った. 例えば AR 表示機能は Huawei AR Engine を使用するためスマートフォン端末上でのみ使用可能となっている.

3.2 ぬいぐるみの三次元スキャン

三次元スキャンでは Mate20 Pro のインカメラから取得した RGB-D データを処理しテクスチャ付きメッシュを出力する. 処理は開始処理, オンライン処理, オフライン処理からなる. 各処理とテクスチャ付きメッシュを図 3 に示す.

3.2.1 開始処理

開始処理のスクリーンショットを図 3 (a) に示す. 撮影を始める際にぬいぐるみがテーブルなどの平面に置かれ, カメラの中央に見えていることを仮定する. デプス画像を変換した点群から平面を検出し, その平面より上部に存在する点群にクラスタリングを行うことでぬいぐるみが内包される三次元領域（以下, ぬいぐるみ領域）を検出する. 検出したぬいぐるみ領域が正しいことをユーザが確認するとその情報がぬいぐるみを基準とした世界座標に登録され, オンライン処理へ移行する.

3.2.2 オンライン処理

オンライン処理のスクリーンショットを図 3 (b) に示す. オンライン処理ではユーザが対象となるぬいぐるみを手で持つて動かすことで裏面を含むぬいぐるみ全周のシーケンシャルな RGB-D データを撮影する. 同時に KinectFusion[10] をベースとしたアルゴリズムを用いて各フレームのカメラポーズとジオメトリの推定をリアルタイムで行う. しかしユーザが手でぬいぐるみを持って動かすことで背景や手に対する位置関係が変化してしまう. このような撮影環境に生じる動的な変形を KinectFusion で扱うことはできない. そこでぬいぐるみ以外を無視し, 前述した世界座標における

る六自由度の三次元剛体変換をカメラポーズとしてトラッキングする。背景を無視するため、各フレームのRGB-Dデータを世界座標に変換し、ぐるみ領域に内包される部分に対してのみ処理を行う。またRGB-Dデータからカラー画像に基づいて検出した手領域を除去する。オンライン処理中に一部のフレームの情報をキーフレームとして保持する。UIに表示される進行度を確認しながら、ユーザはぐるみを手で動かし様々な方向から撮影する。現在のフレームにオーバーレイされた推定中のジオメトリを確認することで、撮影が正常に進行しているか失敗しているかをユーザは直感的に理解できる。スキャンがぐるみの全周をカバーするとオンライン処理は終了する。

本アプリケーションのオンライン処理のように手で物体を動かしながら行うin-hand方式の三次元スキャンは長年研究されているものの[11, 12]、スマートフォン端末に搭載された前例は我々の知る限りなかった。手の三次元トラッキングを活用する手法[13]やトポロジ変化に対応しながら動的な変形をトラッキングする手法[14]を用いることも考えられたが、計算量、ロバスト性、実装の複雑さ等を総合的に判断し、KinectFusionベースのアプローチを採用した。

3.2.3 オフライン処理

オフライン処理のスクリーンショットを図3(c)に示す。オフライン処理では数十秒から数分程度かけて全体最適化、ジオメトリ処理、テクスチャ生成を行う。ジオメトリ処理ではノイズ除去、穴埋め、スムージング、簡略化等を行い、クリーンでwatertightなメッシュを生成する。テクスチャ生成ではキーフレームを用いてメッシュにUV座標とテクスチャを付与する。最終出力であるテクスチャ付きメッシュを図3(d)に示す。

3.3 自動リギング・アニメーション

スキャンしたぐるみを自動でリギングするため、本アプリケーションではPinnochio[15]のアルゴリズムをベースにして開発を行った。スケルトンの推定は事前に用意するテンプレートのスケルトン構造に大きく依存することから、本アプリケーションで対象とするモデルは二足歩行の人型モデルであることを前提とした。本アルゴリズムは、球のパッキング及びメッシュの接続を考慮したウェイト計算を行うことから、穴の開いていないメッシュにのみ適用可能となっている。また、スキャンしたオリジナルのメッシュは頂点が非常に多く、リギングに1分程度の時間を要したため、スキャン後に頂点削減を行うことでリギング処理を5秒程度まで短縮した。

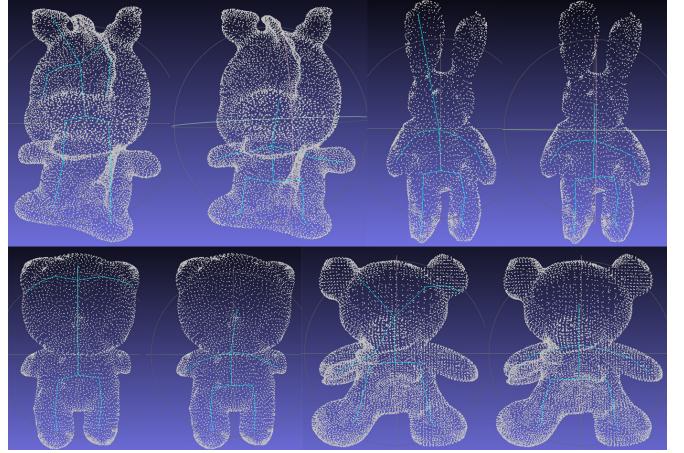


図4: ぐるみに特化した制約による効果の比較(左: 制約なし, 右: 制約あり)

3.3.1 姿勢の自動調整

Pinnochio[15]では、入力となるメッシュが地面に立っている状態が前提となっているが、我々のスキャンされたメッシュはモデルごとにそれぞれ異なる向きで保存されているため、リギングを行う際にはユーザが手動でモデルを回転し、正しい向きにする必要があった。しかし、ユーザの調整によるわずかな向きの違いによってもスケルトンの位置が大きく異なることから、三次元モデルの自動姿勢推定の開発を行った。その際、ぐるみの多くが左右対称であるという仮定のもと、モデルに対して方向付きバウンディングボックス(Oriented Bounding Box)を計算し、その軸の向きをY-upの向きになるように回転させることで自動調整を行えるようにした。本アルゴリズムは多くのぐるみで期待した効果が得られ、またユーザは少ない手順で向きの修正が行えるようになった。同時に端末上でのデバッグ作業の際にリギングの結果を同じ条件で比較できるようになった。

3.3.2 ぐるみに特化した制約の追加

スケルトンの位置推定の際、Pinnochio[15]では著者が事前に用意したキャラクタメッシュデータセットに対して最適化された重みや制約が使用されているが、我々の対象となるぐるみに同一のものを適用した場合、失敗するケースが多くみられた。特にぐるみは頭部が大きく手足が短いものが多いため、頭部の中に腕のスケルトンが入ってしまったり、胴体の大半を脚が占めてしまうようなケースがよく見られた。そのため我々は、ぐるみのようなキャラクタでもうまくいくよう、頭部や脚部に対してスケルトンが正しく推定されるよう新たな制約を加えた。その結果、図4に示すように、元論文では失敗したケースに対して、我々の手法では期待した結果が得られるようになった。



図 5: Huawei Mate20 Pro のリリース発表会の様子

3.3.3 アニメーションのリターゲット

リギング済みのモデルにモーションを適用する際、ぬいぐるみによってそれぞれ腕の長さや足の長さが異なることから、モーションのリターゲットを考える必要がある。本手法では、Pinnochio[15]でも採用されているリターゲット手法を同様に適用する [16]。なお、元となるモーションはそれぞれモーションキャプチャスタジオにて取得したモーションを使用した。

4 リリースと反響

本アプリケーションは、2018年10月英国ロンドンにて、スマートフォン端末、Huawei Mate20 Pro の発表会にて正式に発表された。当日は壇上での約8分間の長いプレゼンテーションデモに加え、直接端末に触れることのできるブースも用意されており、そこで来場したメディア関係者に対面でのデモを行った。三次元スキャンには経験が必要なため、スタッフがその場で来場者に対してデモを行い、最後にAR機能によるツーショット写真をプリンタで印刷してプレゼントするようなブースを設けた。この際、多くのメディア関係者によって記事として取り上げていただき、スマートフォン端末のARの文脈として大きな貢献を果たしたと考えている。発表会以後、本スマートフォン端末を購入したユーザが、本アプリケーションでスキャンしたぬいぐるみをアニメーションしている様子を各SNSで確認することができ、開発した技術が製品になって、一般のユーザの手元に届くまでの貴重な経験をすることができた。その後も、SIGGRAPH ASIA 2018 や VC2019, ICPC2019といったイベントで出展し、来場者のぬいぐるみをスキャン、リギングしたモデルをglTFとして提供するような取り組みも行い大変好評であった。

5まとめと今後の課題

本スマートフォンアプリケーション「3D LIVE MAKER」は、ユーザが自分の好きなぬいぐるみを自由にスキャンし、



図 6: ICPC2019 にて展示を行った際にスキャンしたぬいぐるみに対する結果

アニメーションして遊ぶことができるよう、従来課題であった小物体全周のスキャン、リギングの自動化といった課題を解決しつつ、すべての機能をスマートフォン端末上で動作可能にした。結果として、ユーザが能動的にARコンテンツ制作を制作して楽しむ未来への第一歩となったと考えている。なお本アプリケーションは現在、アウトカメラにToF方式のデプスカメラを搭載したスマートフォン端末Huawei P30 Pro[17]でも使用可能となっている。また本技術は米国をはじめとする各国で特許申請済みである [18]。

今後の課題として、ぬいぐるみの形状は多種多様故、スキャンに失敗するケースやリギングがうまくいかないケース、ライティング状況によってテクスチャが汚くなってしまうケースなどがあり、引き続きアルゴリズムの更新が必要であると考えている。またスキャンの手続きが煩雑であるため、今後はより直感的かつロバストなスキャン手法の提供を目指していく。

今後のアプリケーション応用として、スキャンしたぬいぐるみとユーザ間のコミュニケーションの実現を目指しており、そのためのぬいぐるみへの表情付与や発話機能、身体的なインタラクション設計について取り組んでいきたい。またCGキャラクタがAR空間になじむようにするための環境光の考慮、障害物によるオクルージョンの考慮、環境を理解したモーションプランニングなど、キャラクタの存在感をより感じることができるような設計が今後よりAR体験を豊かにしていくと考えている。こうした未来の実現のために、CG分野に限らず、コンピュータビジョンや人工知能技術、様々なハードウェアや高速通信技術との連携も積極的に進めていきたい。

6 謝辞

本アプリケーションが製品として世に出るまでご協力頂いた社内外の全ての人に感謝する。中でも東京研究所の3D Vision and Graphicグループのメンバーのサポートによってプロトタイプ作成及び特許業務を完遂することができた。本社研究部門のメンバーには製品搭載に向けたアルゴリズ

ム改良において、端末事業部のメンバーには製品への組み込みやマーケティングにおいて、多大な支援を受けた。また、データ収録等を手伝って下さった協力企業の皆様に謝意を表する。最後に、この度応用論文として発表する機会を下さった向井智彦先生、楽詠コウ先生の両名に深謝の意を表する。

参考文献

- [1] T. Igarashi, S. Matsuoka, and H. Tanaka, “Teddy: A sketching interface for 3D freeform design,” *Int'l Conf. on Computer Graphics and Interactive Techniques*, p. 409–416, 1999.
- [2] Beijing Xiaoxiaoniu Creative Technologies Ltd, “Wonder Painter,” <https://wonderpainter.net/>, released: 2018.
- [3] Whatever Inc., “Rakugaki AR,” <https://whatever.co/post/rakugakiar/>, released: 2020.
- [4] S. Magnenat, D. T. Ngo, F. Zünd, M. Ryffel, G. Noris, G. Rothlin, A. Marra, M. Nitti, P. Fua, M. Gross, and R. W. Sumner, “Live texturing of augmented reality characters from colored drawings,” *IEEE Trans. on Visualization and Computer Graphics*, vol. 21, no. 11, pp. 1201–1210, 2015.
- [5] Sony Corporation, “3D Creator,” <https://www.sonymobile.co.jp/xperia/function/3dcreator/>, released: 2017.
- [6] S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, and H. Li, “PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization,” *IEEE Int'l Conf. on Computer Vision*, pp. 2304–2314, 2019.
- [7] Adobe Inc., “Mixamo,” <https://www.mixamo.com/>.
- [8] Huawei Technologies Co., Ltd., “HUAWEI Mate20 Pro,” <https://consumer.huawei.com/en/support/phones/mate20-pro/>, accessed: 2020/11/11.
- [9] Huawei Technologies Co., Ltd., “HUAWEI AR Engine,” <https://developer.huawei.com/consumer/en/hms/huawei-arengine/>, accessed: 2020/11/11.
- [10] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. W. Fitzgibbon, “KinectFusion: Real-time dense surface mapping and tracking,” *Int'l Symposium on Mixed and Augmented Reality*, vol. 11, pp. 127–136, 2011.
- [11] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy, “Real-time 3D model acquisition,” *ACM Trans. on Graphics*, vol. 21, no. 3, p. 438–446, Jul. 2002.
- [12] T. Weise, B. Leibe, and L. Van Gool, “Accurate and robust registration for in-hand modeling,” *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [13] D. Tzionas and J. Gall, “3D object reconstruction from hand-object interactions,” *IEEE Int'l Conf. on Computer Vision*, pp. 729–737, 2015.
- [14] M. Dou, S. Khamis, Y. Degtyarev, P. Davidson, S. R. Fanello, A. Kowdle, S. O. Escolano, C. Rhemann, D. Kim, J. Taylor, P. Kohli, V. Tankovich, and S. Izadi, “Fusion4D: Real-time performance capture of challenging scenes,” *ACM Trans. on Graphics*, vol. 35, no. 4, Jul. 2016.
- [15] I. Baran and J. Popović, “Automatic rigging and animation of 3D characters,” *ACM Trans. on Graphics*, vol. 26, no. 3, p. 72–es, Jul. 2007.
- [16] K.-J. Choi and H.-S. Ko, “Online motion retargetting,” *The Journal of Visualization and Computer Animation*, vol. 11, no. 5, pp. 223–235, 2000.
- [17] Huawei Technologies Co., Ltd., “HUAWEI P30 Pro,” <https://consumer.huawei.com/en/phones/p30-pro/>, accessed: 2020/11/11.
- [18] N. Iwamoto, T. Wang, and C. Lei, “Object modeling and movement method and apparatus, and device,” *U.S. Patent Application*, no. 16/931,024, Nov. 2020.