

## 研究論文

# 深度カメラと姿勢推定モデルを用いた 3D キャラクターリアルタイムモーションシステムの開発

三浦彰人

2020年2月13日

## 概要

VRデバイスやモーションキャプチャ技術の普及に伴い、三次元コンピュータグラフィックスとその関連技術を表現手法として用いることが一般的となりつつある。しかし、これらの技術を用いたアプリケーション開発においては、開発環境の構築やカスタマイズで躓きがちである。また、要求されるハードウェア性能が高く、コストが大きくなりやすい。そこで本研究では、安価な深度カメラ RealSense と姿勢推定モデル PoseNet を用い、一般的なノート PC の性能でも利用可能な 3D キャラクターリアルタイムモーションシステムを開発した。

## 1 はじめに

2010年代に入り、三次元コンピュータグラフィックス(3DCG)とモーションキャプチャ技術のコモディティ化が進んでいる。更に2016年ごろから、Oculus Riftをはじめとした Virtual Reality (VR) デバイスが急激な普及や、スマートフォンアプリへの Argumented Reality (AR) 技術の取り入れ、広報・広告活動における 3DCG キャラクターの利用など、3DCG を用いたコンテンツの多様化と一般化が進んでいる。このため、次世代の情報伝達手法・表現手法として 3DCG はより身近で欠かせない存在となっていくと考えられる。

また、筆者の所属している東北公益文科大学が、文部科学省の私立大学研究ブランディング事業に採択され<sup>\*1</sup>、事業のテーマとして、「日本遺産を誇る山形県庄内地方を基盤とした地域文化と IT 技術の融合による伝承環境研究の展開」が掲げられている。その中で、モーションキャプチャや CG アニメーション、

---

<sup>\*1</sup> 文部科学省. 平成 29 年度「私立大学研究ブランディング事業」選定事業一覧. [http://www.mext.go.jp/a\\_menu/koutou/shinkou/07021403/002/002/1398494.htm](http://www.mext.go.jp/a_menu/koutou/shinkou/07021403/002/002/1398494.htm).

VRなどのIT技術を用いた地域文化の活用・デジタルアーカイブ化が推進されており、これらの技術に対する知見の集約が不可欠である。本事業においては、学生も交えた研究プロジェクトの推進が必要であるが、学生が3DCG技術に関する知識・技術を身に付けられる環境が乏しい。

そこで、学生の所有しているノートPCのような環境でも3DCGに関連する技術に触れ学ぶことのできるシステムが必要と考え、2018年度より、シンプルで軽量の3Dキャラクターリアルタイムモーションシステム[1]を開発している。

## 2 課題

2018年度に開発した3Dキャラクターリアルタイムモーションシステムは、Webカメラで撮影した画像フレーム内で顔が検出されたら、キャラクターがそれに応じた動きをするというものであった。そのため、腕や手、足などはあらかじめプログラムされた通りにしか動かせず、単調な動作となっていた。そこで本研究では、比較的安価な構成でリアルタイムに全身のモーションキャプチャによる動作が可能な3Dキャラクターモーションシステムを開発する。

## 3 システム要件

本システムの要件として、下記の4点を重点に置く。

- シンプルでカスタマイズが容易なシステムとすること
- 低コストで利用可能なシステムとすること
- 人の頭、首、肩、腕、胴体、腰、足の動きに追従すること
- リアルタイムで動作すること

シンプルでカスタマイズが容易であり、かつ低コストであることは、教育的なシステムである故に欠かせない要素である。また、ソフトウェアライセンスによるカスタマイズなどの制約を最小限に抑えるため、可能な限りオープンソースソフトウェアを活用し、開発したシステム自体もオープンソースソフトウェアとして公開する。更に、可能な限りリアルタイムで動作するようなシステムを目指すことで、応用の可能性を高める。性能と価格に関する要件は、下記の通り設定する。

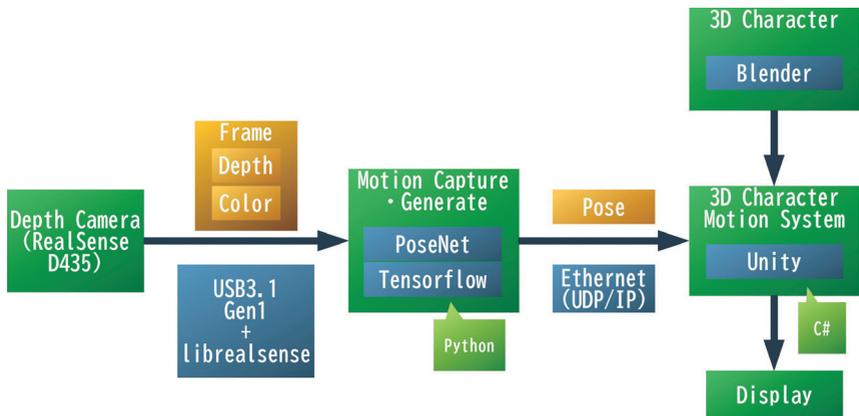


図1 システム全体の概要

- モーションキャプチャレート：30fps 以上
- キャラクターモーションのフレームレート：30fps 以上
- キャラクターのレンダリング解像度：1920 × 1080
- 2019年時点において、システム一式が15万円以内で導入可能であること

## 4 システム設計と実装

本システムは、3Dキャラクターモデル、モーションキャプチャシステム、キャラクターモーション生成システム、3Dキャラクターモーション表示システムの4つで構成される。4つの要素を疎結合するシステム構成とすることで各システムをシンプルな構成に保ち、カスタマイズが容易となるよう設計する。

### 4.1 3Dキャラクターモデル

3Dキャラクターモデルの作成には、旧システムと同様にBlender[2]を用いる。本システムでは、三頭身程度にデフォルメされたキャラクターモデルを制作する。キャラクターに設定するボーンは、モーションキャプチャシステムで取得できる身体パーツ座標を適用しやすい部分に設定する。今回制作したキャラクター(図2)には、頭、首、背骨、肩、上腕、前腕、掌、大腿、下腿、足のボーンを設定した。リアルタイムで動作させることを考慮し、頂点数は20000程度、ボーン数は23(うち動作対象は14)とした。

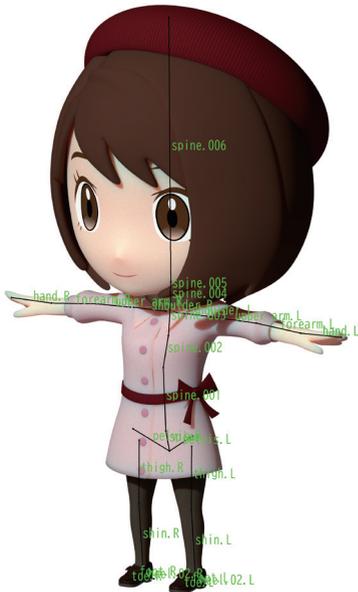


図2 キャラクターとボーン

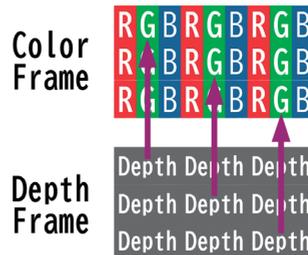


図3 RGBフレームとDepthフレームの構造

## 4.2 モーションキャプチャシステム

旧システムでは、Webカメラとdlib[3]を用いて顔のみのモーションキャプチャをおこなっていた。全身のモーションをキャプチャできる安価なデバイスにはPerception Neuron<sup>\*2</sup>などがあるが、システム一式で25万円以上のコストがかかる上、モーションキャプチャ時に各種センサー類を全身に装備しなければならず煩雑である。そこで本システムでは、安価な深度カメラであるIntel RealSenseD435[4]と、オープンソースの機械学習ライブラリTensorFlow[5]のリアルタイム姿勢推定モデルであるPoseNet[6]を用いて、顔、胴体、手、足の三次元座標を取得する。

RealSense D435では、カラー画像と深度画像をラスタ画像形式で取得できる。取得したカラー画像と深度画像は、図3のようにピクセル毎に座標が一致するように調整されている。モーションキャプチャの際は、まず取得したカラー

<sup>\*2</sup> Noitom. Perception Neuron  
<https://www.noitom.com/perception-neuron-2-0>

表1 モーションキャプチャ速度の評価

演算装置	種類	50	75	100	101
Ryzen 3600 + RTX2800	GPU	56.4	55.3	52.7	54.2
Ryzen 2700X	CPU	58.4	51.5	39.5	37.4
Core i5 8400	CPU	51.9	41.2	29.6	29.2
Ryzen 2600U(評価機)	CPU	27.6	20.0	18.3	17.3
Jetson Nano	GPU	15.6	14.9	10.0	8.4

画像を元に、PoseNetで人体の各パーツの二次元座標を取得する。この各パーツの二次元座標と深度画像の情報とを組み合わせることで、三次元座標を取得することができる。

モーションキャプチャ対象以外の人物が画像内に映り込んでいると、そちらの人物をメインのモーションキャプチャ対象として捉えてしまうことがある。イベント会場での展示の際など、背面に人が映り込みやすい環境において不具合の原因となるため、一定の距離を超える物体はカラー画像から消去する。また、カメラの視野内に全身が収まっていないと、モーションキャプチャの精度が著しく低下する。そこで、縦長の画像を取得できるようにカメラ本体を90度回転し、取得した画像も合わせて90度回転した上でポーズ認識をおこなう。画像のアスペクト比は、手を広げた際にも視野内に収まるよう4:3とし、画像解像度は320×240とする。

#### 4.3 キャラクターモーション生成システム

キャラクターのボーンを動作させる際は、各ボーンをひとつずつ回転させる手法(フォワードキネマティクス, FK)と、あるボーンの移動に関連するボーンを追従させる手法(インバースキネマティクス, IK)がある。本システムでは、FKを用いた動作をおこなう。FKによる動作をおこなうためには、モーションキャプチャシステムを用いて取得した三次元座標を、各ボーンの回転に変換する必要がある。また、モーションキャプチャシステムによって取得した座標には揺らぎがあり、そのままキャラクターに反映すると異常な振動をする。これを低減するため、移動平均法を用いて動作の平滑化をおこなう。変換し生成した回転情報を3Dキャラクターモーションシステムに送信する際は、リアル

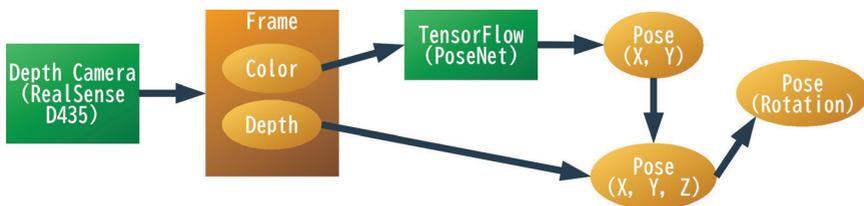


図4 モーションキャプチャシステムの概要

タイム性と扱いやすさを考慮し、JSON形式でシリアルライズした上でUDPを用いて送信する。モーションキャプチャシステムとキャラクターモーション生成システムによりモーションが生成される流れは、図4のようになる。

#### 4.4 3Dキャラクターモーション表示システム

3Dキャラクターのレンダリングと動作にはゲームエンジンであるUnityを用いる。Unity内でキャラクターモーション生成システムから受け取った回転情報をキャラクターに反映しレンダリングする。ボーン回転情報の更新とデータの受信を同期させるとレンダリングが遅延し、動作のカクつきが生じる。このため、ボーン回転情報の更新とデータ受信は別スレッドにて非同期でおこない、ボーン回転情報更新の際はその時点で受信済みの最新のモーションデータを参照する。

## 5 評価

先述の通り、本システムは学生が所有しているノートPCなどでも動作することを要件としている。そこで、2019年時点で10万円程度で購入可能なノートPCで実際に稼働させ、フレームレートに問題がないか評価する。システム評価用ノートPCの構成は下記の通りである。

- CPU : AMD Ryzen 5 2600U
- GPU : AMD Radeon Vega8
- メモリ : DDR4-24008GB
- ストレージ : NVMeSSD128GB
- 画面解像度 : 1920 × 1080

## 5.1 モーションキャプチャ・モーション生成システムの評価

モーションキャプチャシステムに用いる TensorFlow と PoseNet は、演算装置により処理速度が大きく変化する。また、PoseNet の動作検出品質設定は、50、75、100、101 の4段階でおこなうことができ、品質を高めると処理に掛かる時間が増大する。そこで、モーションキャプチャ・モーション生成システムによる、画像取得からモーション生成までの一連の処理をどの程度のフレームレートで実現可能かを評価した。また評価の際、計算機の性能によりどの程度影響があるかを確認するため、性能の異なる複数の環境を用意し、それぞれの環境で同様の実験をおこなった。(表1)

動作検出品質を50に設定した場合、主な環境として想定しているノートパソコン用プロセッサにおいて、30fpsに近いフレームレートを実現することができた。また、最高品質の101にて高フレームレートを実現したい場合は、RTX2080のようなGPUを用いる必要があることが確認できた。

## 5.2 キャラクターモーション表示システムの評価

モーションキャプチャ・モーション生成システムを稼働させた状態で、キャラクターモーション表示システムをどの程度のフレームレートでおこなえるか、システム総合評価用ノートPCを用いて評価をおこなった。結果として、モーションデータを受け取って動作しているキャラクターのレンダリングについては、概ね70~80fps程度でおこなうことができた。よって、評価機場では現状においてシステム全体を30fps程度で動作させることができ、モーションキャプチャ・モーション生成の高速化が可能となれば、より滑らかな動きを追求できることが分かった。

一方、現状のモーションデータでは三次元座標からボーンの回転への変換に問題があるケースが多く、想定しない方向にボーンが曲がる症状が散見されるため、改善が必要である。

# 6 本システムの応用

## 6.1 Looking Glassによる3Dホログラムシステム

本システムの応用例として、3Dホログラフィックディスプレイである Looking Glass[7]への表示をおこなった。Looking Glassに表示する場合も

Unityを用いるが、HoloPlay Unity SDKを用いてLooking Glassに対応した表示領域の設定をおこなう必要がある。

## 6.2 三次元空間記録システムとの連携

私立大学研究ブランディング事業の別プロジェクトとの連携として、三次元空間記録システムにて撮影した三次元空間記録データを、本システムの背景映像として用いる機能を追加した。これにより、三次元空間記録システムのデータを容易に他の3DCGを用いたシステムやコンテンツに応用できることを示すことができた。

## 6.3 高校生を対象とした3DCG教育

筆者は、2019年4月より酒田光陵高校情報科の三年次生の課題研究において、3Dキャラクターモデルの制作と、それを用いたシステムの開発指導を担当している。そこで、高校生が作成した3Dキャラクターを実際に本システムに組み込み稼働させる演習をおこなった。結果として概ね想定通り動作したものの、一部のモデルにおいて首が180度回転してしまうなどといった不具合が発生し、モデル製作時点での考慮が必要な場合が残されていることが判明した。これに対しては、モデル制作時に考慮できるよう、一般的な3Dキャラクターの制作手順の他に本システムの利用を想定したモデルカスタマイズ手順の体系化が必要である。本課題研究の成果は、2020年2月8日に開催された酒田光陵高校IT-ACE発表会にて発表・展示をおこなった。

## 7 おわりに

安価な深度カメラとPoseSenseによるリアルタイムモーションキャプチャシステムと、それを用いた3Dキャラクターリアルタイムモーションシステムを作成し、一般的なノートPCで動作させることができた。これにより、学生などが所有するノートPCのような環境であっても、比較的安価で手軽に3Dキャラクターリアルタイムモーションシステムを利用できるようになった。また、本システムの応用として、三次元コンピュータグラフィックスを対象とした教育や、地域資源の活用などを実現できた。一方、モーションキャプチャにおいては、モーション変換精度の向上や指先モーションの取得などといった課題が残っている。

## 参考文献

- [1] 三浦彰人. 3D キャラクターリアルタイムモーションシステムの開発. 東北公益文科大学 平成 30 年度ブランディング事業関連論文集, pp.46–49,2019.
- [2] The Blender Foundation. Blender. <https://www.blender.org/>.
- [3] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, Vol. 10, pp.1755–1758,2009.
- [4] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, and Achintya Bhowmik. Intel re-alsense stereoscopic depth cameras. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*,2017.
- [5] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, DandelionMané, RajatMonga, SherryMoore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, FernandaViégas, OriolVinyals, PeteWarden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous systems, 2015. Software available from [tensorflow.org](http://tensorflow.org).
- [6] Tensorflow- pose estimation. [https://www.tensorflow.org/lite/models/pose\\_estimation/overview](https://www.tensorflow.org/lite/models/pose_estimation/overview).
- [7] Looking Glass Factory Inc. The world's leading holographic display. <https://lookingglassfactory.com/>.