

一般的な単眼カメラによる表情・骨格の認識と アバターへの反映に関する研究

松本 拓未[†] 佐々木 茂[†] 田中 誠一[‡]

帝京大学理工学部情報電子工学科[†] 文星芸術大学美術学部美術学科マンガ専攻[‡]

1.はじめに

本研究では、一般的な単眼カメラ(Web カメラやスマートフォンに内蔵されているカメラ等)を用いて顔の表情と身体の認識を行い、アバターに反映するシステムを開発する。既存のアプリケーション等の中で、Web カメラのみで手軽に利用できる環境では、顔の表情のみ、もしくは身体全体の大まかな動きしかキャプチャできない。一方、表情と身体を両方キャプチャする場合は、Web カメラで表情を認識し、別のデバイスを複数組み合わせることで身体の動きを認識している。以上を踏まえて、特別なデバイスを使わず、Web カメラのみで表情と上半身の動きを自分のアバターへ同期するアプリケーションとして実装し、配信ソフト等と連携して使用することを目的とする。

2. 関連研究および事例

顔の表情と身体の動きを認識し、アバターに同期させることができるアプリケーションとして、たとえば次の2つが挙げられる。

- 3tene Studio [1]
- Hitogata [2]

上記2例はどちらも Web カメラで表情と顔の傾きを認識し、身体の動きは Perception Neuron を使用してトラッキングしている。本研究で目指している、Web カメラのみで顔の表情と身体の動きをトラッキングする例は見られなかった。

3.方法

3.1 開発環境とプログラミング言語

開発環境は Unity を使用し、プログラミング言語は Unity でスクリプト作成に用いられる C# を使用する。

3.2 3D モデルについて

3D モデルのアバターは VRM 形式 [3] のものを使用する。プログラム上では、VRM モデルを Unity で扱うことができるライブラリ UniVRM を用いて実装する。

3.3 機械学習ライブラリの活用

表情と身体の動きを認識する際には機械学習を活用する。これに伴い、機械学習ライブラリである MediaPipe [4] と BodyPix [5] を使用する。MediaPipe には虹彩や手、物体など様々なものに対する学習モデル及び API が付属している。このうち本研究では Face Mesh(顔の目以外の部分)、Iris(目)、Hands(手)を使用する。

BodyPix は TensorFlow [6] をベースとして開発されており、人体の各部位の領域が検出できる学習モデルである。

これらを Unity から使用するため、Unity パッケージである HandPoseBarracuda、FaceMeshBarracuda、BodyPixBarracuda を利用する。

3.4 プログラムの全体構成について

図1にプログラムの全体的な流れを示す。各プログラムは以下のようになっている。

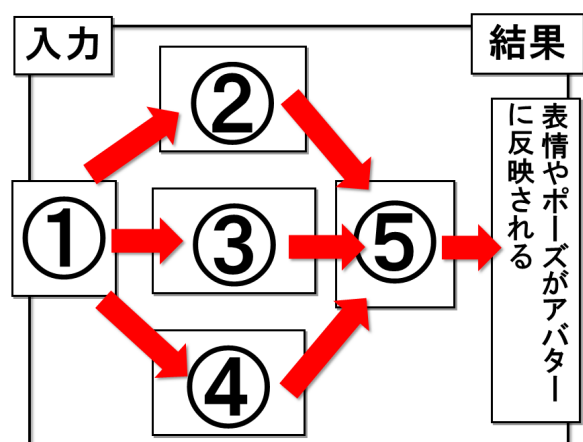


図1 プログラムの流れ図

Research on facial expression and skeleton recognition with a general monocular camera for avatar manipulation

[†]Matsumoto Takumi, Sasaki Shigeru • Teikyo University

[‡]Tanaka Seiichi • Bunsei University of Art

図1中の①～⑤の具体的な内容は次のとおりである。

- ① ウェブカメラの映像を得る。
- ② ウェブカメラからの画像を MediaPipe の Face Mesh に与える。(顔の各点の推定座標が出力される)
- ③ ウェブカメラからの画像を MediaPipe の Hands に与える。(手の各点の推定座標が出力される)
- ④ ウェブカメラからの画像を BodyPix に与える。(身体各点の推定座標が出力される)
- ⑤ ②～④の出力結果を適切に変換・分析し、VRMモデルのシェイプキーとボーンを動かす。

4. 結果

これまでのところ図1の③の手のポーズの認識を除いてUnityアプリケーションとして実装できている。アプリケーションの実行結果を図2に示す。カメラの画像から図1の②の顔の各点の認識と④の上半身の各点の認識結果をVRM形式のアバターの表情およびポーズに反映できている。

目および口の開き具合については、②において得られた顔の各点の推定座標のうち、上下左右端の座標から求めた縦横比を用いて設定している。視線の方向(黒目の向き)については、黒目の中心の座標を用い、左右の平均を求めて設定している。

また頭のポーズ(前後左右の傾きと首の回転)については、顔の上下左右端(左右は頬のあたり)の座標により設定する。顔の各点の座標は3次元座標で推定されるため、前後方向や首周りの回転についての設定が可能である。

④の上半身の認識においては、奥行き方向の座標が推定されないため、カメラ画像に投影された肩、肘、手首の座標から上腕および前腕の長さを求めて、奥行き方向の座標を推定している。

また、両肩を結ぶ線の傾きから上半身の傾きを求めて、VRMの背骨(spine)の角度を設定している。首の付け根より上の頭の座標は②の顔認識により求められるため、顔認識から得られた座標から首および頭の角度を設定している。手首より先のポーズについては、現在のプログラムでは認識していないため、手首の角度も含めて設定していない。

5. 考察

現在のプログラムでは手首より先のポーズの認識を行っていないが、これは MediaPipe の Hands で認識できる手のポーズにおいて1度に複

数の手を認識できないことによる。これについては、左右の手を含む画像の一部を切り出して、それぞれを認識したのち、全体のポーズに反映させることで実装できないかを検討している。同時に、2つの手を一度に認識できる機械学習ライブラリがないかについての調査も行う予定である。

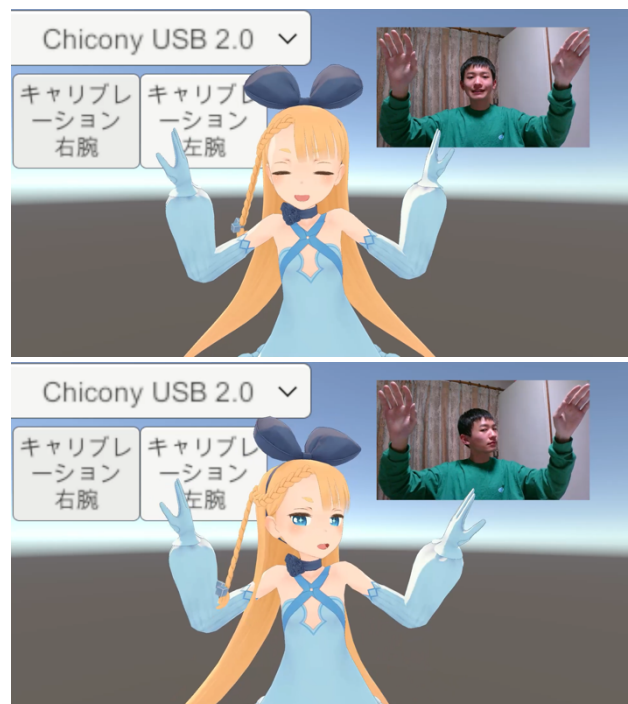


図2 表情および上半身の動きの同期の様子

参考文献

- [1] 3tene : <https://3tene.com/> , 閲覧日:2021/7/7
- [2] Hitogata : <https://sites.google.com/site/vhitogata/> , 閲覧日:2021/7/7
- [3] VRMコンソーシアム : <https://vr-consortium.org/> , 閲覧日:2021/11/19
- [4] MediaPipe : <https://google.github.io/mediapipe/> , 閲覧日:2021/11/19
- [5] Google Developers BodyPixの概要:ブラウザとTensorFlow.jsによるリアルタイム人セグメンテーション : <https://developers-jp.googleblog.com/2019/04/bodypix-tensorflowjs.html> , 閲覧日:2021/11/19
- [6] TensorFlow : <https://www.tensorflow.org/?hl=ja> , 閲覧日:2021/11/19