

1. 担当 PM

田中 邦裕（さくらインターネット株式会社 代表取締役社長）

2. クリエータ氏名

原 拓也（東京大学 大学院 学際情報学府）

田丸 裕己（東京大学 大学院 学際情報学府）

平城 裕隆（東京大学 大学院 学際情報学府）

3. 委託金支払額

2,736,000 円

4. テーマ名

全身の姿勢推定が可能なイヤークセサリの開発

5. 関連 Web サイト

e-Pose デモ動画：<https://youtu.be/Nz-2SIjE0Cs>

6. テーマ概要

デジタルトランスフォーメーション（DX）が注目されるなか、スポーツやヘルスケア、作業中の動作の把握などにおける DX において、姿勢推定が必要となるケースが増えてきた。

姿勢推定とは、ある人間の姿勢、すなわち、肩や腰、手や足など複数の関節の 2 次元または 3 次元位置を推定することであり、私たち人間の「動き」をコンピュータへの自然な入力インタフェースとして利用することができる。

姿勢推定を行う方法としては、固定の単眼カメラもしくは深度カメラによる映像ベースによるもの、複数台の高性能カメラと光学マーカ付きのスーツを用いた光学式モーションキャプチャによるもの、慣性センサを内蔵したスーツによる慣性式モーションキャプチャによるものの 3 つがあるが、場所の制約やスーツの着脱など、ハードルが高く手軽な方法とは言い難い。

そのため、本プロジェクトでは、日常的に利用可能なヘッドホン型デバイスで、場所や動作の制約のない全身の姿勢推定を実現するためのデバイスとソフトウェア群の開発を行う。

なお、本デバイスの左右のハウジング部分にはそれぞれ、ユーザの全身映像を撮影する魚眼カメラが下向きについており、その映像をもとに深層学習によって全身の3次元骨格座標を推定した。

7. 採択理由

全身の姿勢推定は有用な情報ソースであり、スポーツにおけるフォームのチェックなどのスポーツDXや、リハビリ支援や高齢者の見守りなどのヘルスケアの分野のほか、さまざまな分野で利用されている。そのため、近年姿勢推定技術は大きく進化してきたが、外部カメラを使ったり、全身に多くのセンサをつけたという手法は、手軽な姿勢推定からは程遠く、日常的に計測するにはハードルが高かった。

今回の提案では、イヤークセサリに内蔵されたカメラやジャイロなどのセンサを用いて、全身の姿勢推定を実現し、それを元にしたさまざまなアプリケーションを開発することになっている。応用範囲は非常に広く、これまで大きな設備や、多くのセンサを使って実現していた状況から、劇的なUX改善を進められる可能性を秘めており、技術的に解決すべきことは多いものの、大変有益な取り組みであると評価している。

8. 開発目標

本プロジェクトでは、場所や動作の制約のないウェアラブルな姿勢推定を実現するために、日常的に身につけられる、ヘッドホン型のデバイスに取り付けた魚眼カメラの映像をもとに、装着者の全身の姿勢を推定し、その姿勢推定結果を自由視点で閲覧可能にすることを目標とする。

事前に魚眼カメラを装着可能な体の部位を複数検討し、耳以外にも、時計を装着する習慣があり親和性が高い手首や、上向きにカメラを装着すると装着者の全身が映り込むつま先などが有力な候補として上がっていた。

しかし、手首では腕の位置によって映り込む体の部位が様々なことや、つま先ではスカート着用時に使用が難しいことや性犯罪等への悪用の懸念があることが問題となり、魚眼カメラの装着位置として両耳が最適であると考えた。

9. 進捗概要

本プロジェクトにおいては、魚眼レンズからの映像をもとに深層学習を用いた姿勢推定を行うためのソフトウェアの開発と、推定を行うためのデータセットの作成、そしてヘッドホン型デバイスの製作を行った。

姿勢推定を行うためには、画像上の各関節の2次元位置と正解の3次元関節位置を用いて教師あり学習を行い、そのデータセットをもとに画像から関節位置を導き出す(図1)。

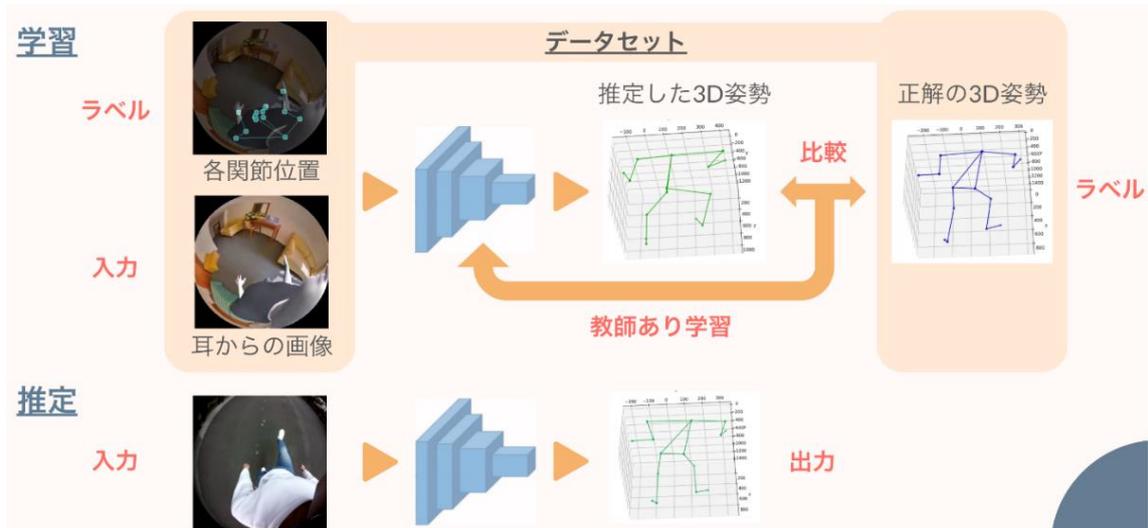


図 1：深層学習モデルの学習と推定

なお、左右の両耳からの画像があるため、それぞれ左耳からの魚眼カメラ画像に対して全身の姿勢推定を行うモデルと、右耳からの魚眼カメラ画像に対して全身の姿勢推定を行うモデルの2つがある。

そのため、ある瞬間の左右の魚眼カメラ画像に対して、2つの3次元姿勢が出力されることから、左右の魚眼カメラの位置や特性に応じて適切に合成することで、装着者の最終的な3次元姿勢を得る。

具体的には、まず左右それぞれの魚眼カメラ画像による3次元関節位置を首からの相対位置に変換する。次に、各部位に関して、①左右の3次元関節位置の平均をとる、もしくは②右半身の部位であれば右、左半身の部位であれば左の魚眼カメラ画像による推定結果を採用する、のいずれかによって最終的な3次元関節位置を決定する

次に、これらを実現するためのデータセットであるが、既存の単眼カメラのデータセットを魚眼カメラでの姿勢推定に利用できず自作する必要があった。

ただ、自作するにあたっては数十万から数百万枚の画像データが必要となり、教師データを用意するハードルが高かったため、Unity上でCGをもとに作成することとした。

モーションデータセットをもとにしたCGをUnity上に作成し、その空間内に現実の魚眼レンズと特性の合うモデルを作成して、耳の位置から撮影した画像を生成し、この画像とモーションデータを元に教師あり学習を行い、深層学習に用いるデータセットを作成した(図2)。



図 2 : CG による合成データセット

最後に、ヘッドホン型デバイスについては、カメラをマウントした Raspberry Pi Zero を左右に装着し、学習済みデータセットをもとに姿勢推定を行うこととした (図 3)。

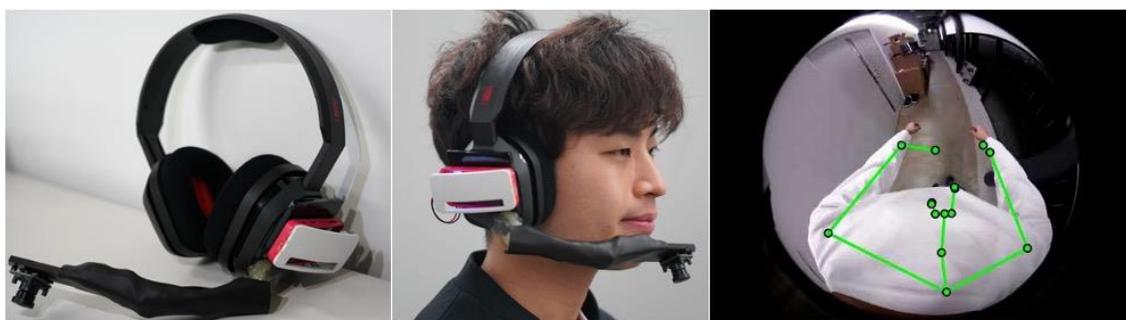


図 3 : ヘッドセット型デバイスと 2 次元姿勢推定結果

プロジェクトの成果として、日常的に着用できるヘッドホン型のデバイスで、装着者の姿勢推定を実現した。既存手法のように全く新しいウェアラブルデバイスを提案するのではなく、魚眼カメラとマイコン、バッテリーを外付けするだけのシンプルなハードウェア構成で、既製品のヘッドホンを拡張したデバイスを提案した。

10. プロジェクト評価

当初想定されていたイヤークセサリからヘッドセット型に変えるなどの変化はあったものの、Unity でデータセットを作るなどさまざまな工夫も行き、ある程度の動きは推定することができた。

外部からカメラで撮影したり、センサをつけて計測したりする手段に比べると、センシングの精度には課題があるが、歩く、手を振るなどの大まかな動きは推定できており、今後の改善を通じて実用レベルになるのではないかと考えている。

11. 今後の課題

今後の課題としては、スポーツや介護などの現場で実際に使用できるレベルの精度で、関節位置を連続的に判定できるようにすることである。

プロジェクト期間において、大まかな身体の動きは推定できるようになったが、手を振っているとか、足が動いているとかといった粒度であり、この精度であれば腕時計型デバイスなどで推計することもできる。

データセット作成などにはしっかり取り組めていたので、さらに教師データを蓄えて精度を高めるとともに、より正しい関節位置を推計できるようにアルゴリズムを改善し続けてほしい。