

全身の姿勢推定が可能なイヤークセサリの開発

—e-Pose: 身近なヘッドホン型デバイスでいつでもどこでも姿勢推定を実現—

1. 背景

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る。姿勢推定により、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ーフェースとして利用することができる。しかし、外部カメラによる対象の人物の撮影や、全身への複数の加速度センサの装着を必要とする既存の姿勢推定技術は、屋内外を問わず日常的に利用するにはハードルが高かった。

2. 目的

私たちの理想は、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し、その姿勢推定結果を活用して人々の生活を豊かにすることである。この理想の実現のために、本プロジェクトでは、日常的に身につけられるヘッドホン型のデバイスで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した。これにより、将来姿勢推定結果を用いたアプリケーションを作成する際の基盤技術を開発することを目指した。

3. 開発の内容

本プロジェクトの概要を図1に示す。本プロジェクトでは、図1のように左右に小型の魚眼カメラを備えた、装着者の映像を用いて全身の姿勢推定が可能なヘッドホン型デバイス e-Pose を開発した。



図 1: 本プロジェクトの概要



図 2: 装着者の姿勢推定を実現するヘッドホン型デバイス

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いる。これを屋内外で使用可能にするためのハードウェアとサーバ通信の実装を行った。

ソフトウェア群の開発では、装着者の全身の姿勢を推定する深層学習モデルと、それを訓練する自作の合成データセット、さらに推定結果の描画や推定のリアルタイム化のための実装を行った。

まず、姿勢推定モデルについて説明する。図 2 に、耳から撮影した魚眼カメラ画像を用いて全身の 3 次元姿勢推定を行う深層学習モデルの学習と推定の様子を示す。最終的な目標は、開発したデバイスを用いて取得した耳からの画像を用いて全身の 3 次元姿勢を推定すること(図 2 下段)である。そのために、耳からの画像を入力とし、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う(図 2 上段)。

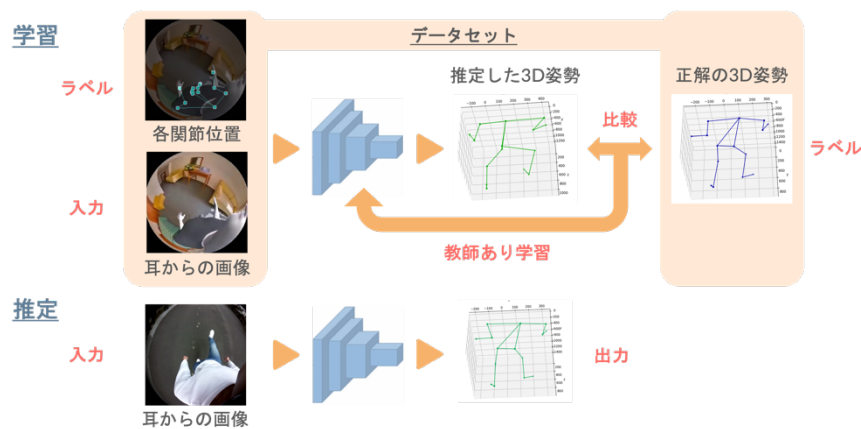


図 2: 姿勢推定モデルの学習と推定の流れ

図 3 に、具体的な姿勢推定モデルの構造を示す。両耳からの映像を入力とし、基礎的な特徴量抽出を行うバックボーンと、2 次元の関節位置のヒートマップを出力とする 2D モジュール、画像上の各関節位置までの深度を推定する深度モジュールの 3 つがある。最終的にヒートマップと深度から、魚眼カメラの歪みを考慮して 3 次元関節位置を出力する。

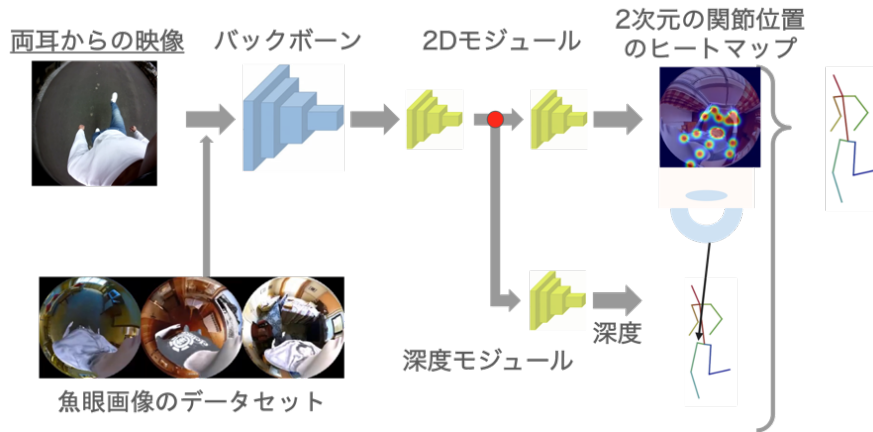


図 3: 姿勢推定モデルの構造

次に、自作の合成データセットの開発について述べる。大規模なデータセットを作成するために、デバイスをつけた人が実際に動く様子を Unity 内に再現した。具体的には、人物のモデル・服のテクスチャ・モーション・背景画像を、Unity 内のアバターに反映させて動かし、同時にアバターの耳元に魚眼カメラを配置して、アバターの全身の映像を録画する(図 4)。これにより耳元から撮影した魚眼カメラ映像とその時の関節位置を対応づけ、これを用いて上記の姿勢推定モデルを学習し、装着者の 2 次元・3 次元姿勢推定を行った(図 5)。



図 4: 自作データセットの作成

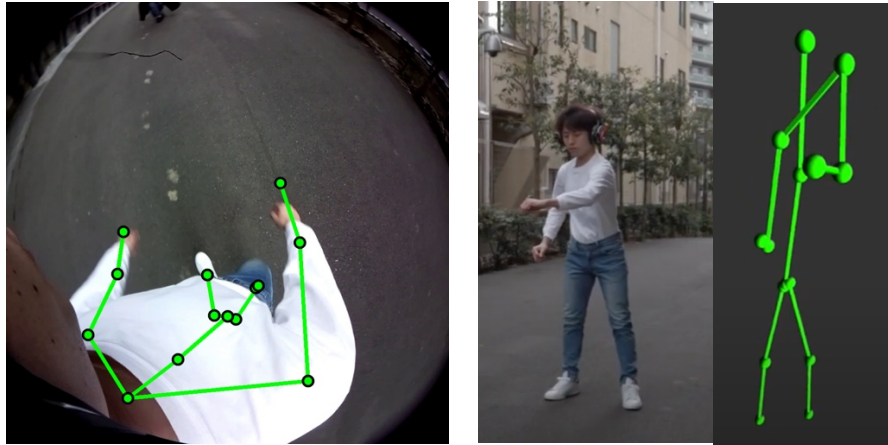


図 5:2 次元姿勢推定の例(左)と3次元姿勢推定の例(右)

4. 従来の技術(または機能)との相違

表 1 に、既存の 4 つの姿勢推定技術と比較した、e-Pose の立ち位置を示す。e-Pose は場所や動作の制約がないことと、日常的に利用できることの 2 点を両立した、今までにない姿勢推定技術である。

表 1: 既存の姿勢推定技術との相違

	推定精度	場所や動作の制約なし	日常利用
深層学習を用いた、 固定の単眼／深度カメラ映像ベース	○	×	○
光学式モーションキャプチャ	◎	×	×
慣性式モーションキャプチャ	○	○	△
深層学習を用いた、 頭部／胸に装着した魚眼カメラ映像ベース	○	○	×
e-Pose	○	○	◎

5. 期待される効果

課題の解決に加えて下記のようなことが可能になると考えている。

① ウェアラブルな姿勢推定のための多様なデバイスの開発

本プロジェクトでは、ユーザがデバイスを装着して動く様子を任意の魚眼カメラ視点で記録可能な、データセット作成環境を開発した。その上で、ヘッドホン型とヘッドセット型の 2 つのデバイスを開発し、そのデバイスのカメラ位置に合わせてデータセットを作成

した。ただし、開発したデバイスには、デザインの改善や小型化の余地がある。そのようなデバイスの改善に伴ってカメラ位置が変更になった場合でも、Unity 上のカメラ位置を変更するだけで対応できる。さらに、耳以外の他の位置にカメラを設置するようなデバイスを開発することも可能である。

② 周囲の環境の情報の取得によるアプリケーション

本デバイスは、場所の制約なく利用できることに加えて、魚眼カメラによって周囲の環境の情報とユーザの視線方向も取得できる。そのため、装着者の姿勢情報を反映した、長期的なライフログの取得が可能であると考えられる。さらに、AR グラスなどと併用することで、現実空間におけるユーザの行動に応じて、AR 空間で情報提示を行うなどのインタラクションも可能になる。

6. 普及(または活用)の見通し

まずは、AR・VR 向けのモーションインタフェースとしての活用が見込める。さらに、長距離ランナーなどのスポーツ選手のフォーム確認や、高齢者のライフログ取得などによるリハビリ支援等にも活用できる。

7. クリエータ名(所属)

- 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)

(参考)関連 URL

e-Pose デモ動画: <https://youtu.be/Nz-2SjE0Cs>