

## 3次元人体姿勢推定を実現するためのOpenPoseの拡張

大野祐汰<sup>†1</sup> プリマ オキ ディッキ アルディアンシャー<sup>†1</sup> 伊藤久祥<sup>†1</sup>

近年、単眼およびステレオカメラを利用した3次元の人体姿勢の推定手法が提案されている。これらの手法は、比較的広い範囲にいる多人数を同時に処理できる反面、精度などの面で制約を受ける。本研究では、2次元の人体姿勢の推定を実現するOpenPoseライブラリを拡張して、3次元の姿勢推定を可能にする手法を提案する。さらに、リアルタイム性と推定した姿勢の精度を分析し、動作解析などへの応用に向けた課題を論じる。

### Extending OpenPose to enable 3D human pose estimation

YUUTA OHNO<sup>†1</sup> OKY DICKY ARDIANSYAH PRIMA<sup>†1</sup> HISAYOSHI ITO<sup>†1</sup>

Recently, many studies have proposed methods to estimate 3-dimensional (3D) human pose from either single or stereo cameras. These methods allow to process the 3D human pose at the same time for a limit number of persons in target with an extent of accuracy. This study extends the library to estimate 2-dimensional human pose, OpenPose to enable the 3D human pose estimation for multiple persons, and discusses the accuracy and further application of the extended library to process real-time 3D human pose.

#### 1. はじめに

可視光カメラを利用した人体姿勢推定に関する研究は、数多く行われている。初期の研究では、三眼視に基づく手法<sup>1)</sup>や多視点画像から回帰分析に基づく手法<sup>2)</sup>、パーティクルフィルタに基づく手法<sup>3)</sup>などがある。しかしながら、近年、畳み込みニューラルネットワーク (CNN) による深層学習の発展に伴い、カメラ映像から人体関節の高精度な検出が可能になり、単眼カメラを利用してリアルタイムで複数人の人体関節を自動的に検出できるようになった。OpenPose<sup>4)</sup>は、単眼カメラの映像に対しCNNを利用して2次元人体関節と顔の特徴点を検出し、これらの特徴点をPAFs (Part Affinity Fields)で関連づけている。現時点において、OpenPoseは3次元人体姿勢の推定に対応していないが、今後の発展として3次元人体姿勢への対応が期待されている。

本研究では、OpenPoseの入力としてステレオカメラの映像を利用して、左右の映像から個別に人体姿勢を推定し、これらの視差情報をもとに重回帰式で3次元人体姿勢の推定を試みる。

#### 2. 提案する3次元人体姿勢推定手法

OpenPoseを拡張し、ステレオ映像から3次元人体姿勢を推定する概略図を図1に示す。ここで3次元人体姿勢を推定するためには、次のような手続きをとる。まず、ステレオカメラの撮影領域において、水平方向や垂直方向、奥行き方向に複数の基準点を設け、それらの基準点の座標  $\mathbf{P}(x_i, y_i, z_i)$  と左映像の座標  $(x_L, y_L)$ 、右映像の座標  $(x_R, y_R)$  との関係を高次多項式

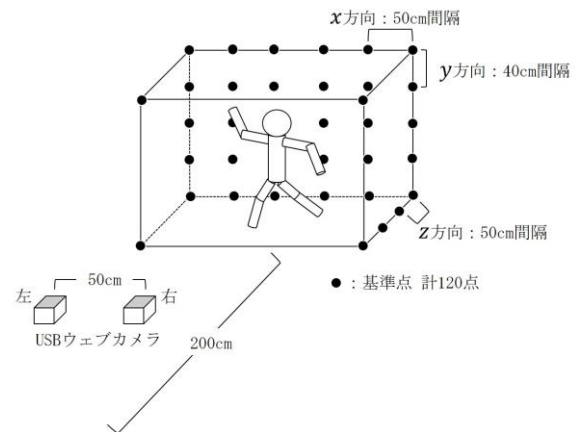


図1 提案する3次元人体姿勢推定手法の概略図  
Figure 1 Outline of the proposed 3D human pose estimation

$$\mathbf{P} = \mathbf{A} \cdot \mathbf{F}(x_L, y_L, x_R, y_R) + \boldsymbol{\xi} \quad (1)$$

で求める。ただし、 $\mathbf{A}$ は $m \times n$ の行列、 $\boldsymbol{\xi}$ は残差、 $\mathbf{F}$ は左映像の座標  $(x_L, y_L)$  と右映像の座標  $(x_R, y_R)$  から対応する基準点の3次元座標  $(x_i, y_i, z_i)$  を推定する関数である。 $\mathbf{A}$ は、次式より、最小自乗法で求める。

$$\operatorname{argmin} \sum_{i=1}^N \|\mathbf{A} \cdot \mathbf{F}(x_L, y_L, x_R, y_R) - \mathbf{P}\|^2 \quad (2)$$

次に、OpenPoseで求めたステレオ映像の2次元人体関節座標  $(x_L, y_L)$ 、 $(x_R, y_R)$  と、式(2)で求めた $\mathbf{A}$ を利用して、3次元人体姿勢

$$\text{Pose} = \mathbf{A} \cdot \mathbf{F}(x_L, y_L, x_R, y_R) \quad (3)$$

を求める。なお、ステレオ映像を取得するためのステレオ

<sup>†1</sup> 岩手県立大学ソフトウェア情報学部  
Faculty of Software and Information Science, Iwate Prefectural University

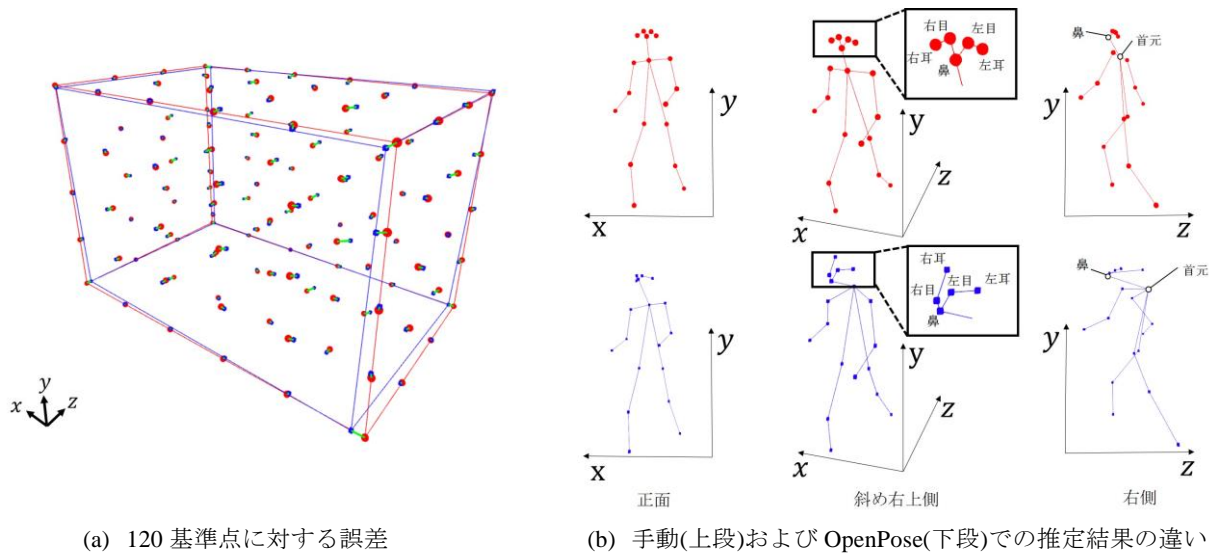


Figure 2 Regression errors and the estimated 3D human pose estimation

カメラに対して事前にステレオキャリブレーションを実施するものとする。

### 3. 実験

実験を通じて、提案する3次元人体姿勢推定手法の結果を検証する。まず、解像度1280px × 720pxの2台のUSBウェブカメラを50cmの間隔で配置し、ステレオキャリブレーションを行う。次に、水平方向に50cm、垂直方向に40cm、奥行き方向に50cmの間隔で複数の基準点を配置し、式(2)により行列Aを求める。最後に、OpenPoseを利用して各カメラから個別に2次元人体姿勢を求めたものと、比較のために手動による2次元人体姿勢のそれぞれについて3次元人体姿勢を推定する。

### 4. 結果

実験結果を図2に示す。まず、図2(a)のように、配置した基準点と重回帰式によって推定した基準点にズレが見られ、その平均自乗誤差は3.44cmである。次に、図2(b)にはOpenPoseをもとに推定した3次元人体姿勢と、手動による2次元人体姿勢をもとに推定した3次元人体姿勢をそれぞれ3つの視点で投影して比較する。顔の部位(目や鼻、耳に該当する部分)と、首元においてOpenPoseで正しく求められない場合がある。その結果、当該部位における3次元人体姿勢の推定も誤差が発生する。一方、その他の部位については、OpenPoseで各関節の位置を比較的精度良く求めることができ、3次元人体姿勢の推定結果も手動による2次元人体姿勢からの3次元人体姿勢の推定結果と類似していることが分かる。ただし、顔の部位について、広く利用されている顔パーツのトラッカー<sup>5)</sup>を利用すれば、顔の部位における3次元座標の推定結果を改善することができる。同様に、首元については推定した両肩の3次元

座標の midpoint として求めることで解決できると考える。

### 5. 結論

本研究では、ステレオカメラを利用してOpenPoseを拡張し、リアルタイムで3次元人体姿勢の推定を実現した。ステレオカメラと複数の3次元基準点を重回帰式で対応させることによって、高速に3次元人体姿勢を推定できるようになる。しかし、推定した3次元人体姿勢の精度は、OpenPoseの2次元人体姿勢の抽出精度に依存しており、今後、フレーム間の抽出結果を考慮し、安定した3次元人体姿勢の推定を目指したい。

### 参考文献

- 1) 高橋和彦, 坂口竜己, 大谷淳: 実時間非接触非装着型3次元人物姿勢推定法に関する一考察, 電子情報通信学会論文誌, D-II, Vol.J83-D-II, No.5, pp.1305-1314 (2000).
- 2) 大西克則, 滝口哲也, 有木康雄: HOG 特徴に基づく単眼画像からの人体3次元姿勢推定, 画像の認識・理解シンポジウム (MIRU2008), (2008).
- 3) 藤田祐輔, 広瀬真一, 松田一朗, 青森久, 伊東 晋: 単眼カメラによる人体の3次元姿勢推定, 2009年映像情報メディア学会年次大会, (2009).
- 4) Cao, Z., Simon, T., Wei, S.E., and Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields, Computer Vision and Pattern Recognition, arXiv:1611.08050, (2017).
- 5) Kazemi V. and Sullivan, J.: One millisecond face alignment with an ensemble of regression trees, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, (2014).