

シングルショット画像からの 3D モデルの生成とボーン座標推定

小野塚 昇大 安藤 崇央

昨今、深層学習や画像認識技術の急激な技術進化により、画像生成や 3D モデル生成など、生成 AI の研究が盛んに行われている。本研究では、シングルショット画像からの 3D モデル生成技術の拡張として、ボーンの 3 次元座標を 3 次元復元アルゴリズムなどを用いて推定し、繊細な動きを持つアニメーションなどにも応用できるような 3D モデルの生成に資するシステムを試作した。その結果、シングルショット画像から指定した人物の 3D モデルを生成したうえで、ボーンの 3D モデルを生成することができた。

1 はじめに

本研究では、複雑な姿勢や細部の調整などを可能とするために、指定した画像内の人物を検出し、対象人物のボーン情報を所持した 3D モデルを作成することを検討する。

まず、画像内から人物を検出し、検出した被写体をマスク加工する。次に、マスク加工した画像を入力とし、目的の被写体の 3D モデルを生成する。続いて、生成した 3D モデルを内部パラメータが同一で外部パラメータの異なるカメラで撮影し、各視点から被写体のキーポイントを検出する。さらに、検出した各撮影点のカメラ座標系キーポイントを 3 次元復元アルゴリズムを用いて、モデルの 3 次元座標をワールド座標系で表す。最後に、算出した 3 次元的情報から、モデルの 3D ボーン座標を決定する。

2 関連技術

本研究では、シングルショットの画像からボーン情報を所有した 3D モデルを生成する方法について検討

をした。ここでは、本研究において活用した技術に関連することについて述べる。

2.1 OpenPose

OpenPose [1] は、Caffe や TensorFlow などのディープラーニングフレームワークをベースに、人物の関節などのキーポイント情報をリアルタイムに抽出する姿勢推定ライブラリである。サンプルの静止画像を用いた姿勢推定の様子を、図 1 に示す。人物の姿勢推定方法には、トップダウン方式とボトムアップ方式があり、OpenPose はボトムアップ方式を採用している。トップダウンでは、まず人物の検出を行い、検出矩形内でキーポイントを探す。ボトムアップでは、画像中の全ピクセルに対して、それぞれが関節の一部であるか否かを判定し、各人物ごとに関節の一部であるピクセル同士を結合して、関節の位置を推定する。

同技術 [1] は、映像内人物の主要部位のキーポイントだけでなく、顔や手指に対するいっそう詳細なキーポイントを抽出することが可能である。また、加速度センサーといった特殊なセンサーを使用せずに、カメラ撮影による動画や画像のみで解析が可能であるという特徴がある。また、GPU を使用することで、動画や画像に複数の被写体がいる場合でもリアルタイムに解析が可能である。

Estimation for Born Coordinates of 3D Model Generated from Single-Shot Images

Shota Onotsuka, 群馬大学大学院情報学研究科, Graduate School of Informatics, Gunma University.

Takahiro Ando, 群馬大学大学院情報学研究科, Faculty of Informatics, Gunma University.



姿勢推定前

姿勢推定後

図 1 OpenPose による姿勢推定 [1]

2.2 DreamGaussian

DreamGaussian [2] は、シングルショットの画像から高速で高画質の 3D モデルを生成するための技術である。同技術 [2] は、従来のスコア蒸留サンプリング (Score Distillation Sampling) による最適化ベースの 3D モデル生成技術と比較し、モデルの生成開始から終了までの処理時間の大幅な短縮や、品質向上を実現した。これは主に、ガウススプラッティング (Gaussian Splatting) モデルの使用、UV 空間でのメッシュ抽出、およびテクスチャの精緻化によって実現されている。ガウススプラッティングとは、連続的な情報を離散的な画像に変換するための手法の 1 つであり、同技術 [2] では点群データを 3 次元のボリュームデータ (画像) に変換するために用いている。

3 関連研究

本研究では、画像中の前景検出や 3D モデル生成、特徴点追跡などの技術を活用する。ここでは、本研究に関連する研究について述べる。

3.1 大局的幾何制約を用いた特徴点追跡に基づく画像列からの 3 次元モデル生成

文献 [3] では、物体の全周囲を撮影した画像列から、大局的幾何制約により特徴点追跡を行い、再撮影誤差最小化原理に基づき安全に 3 次元モデルを生成する一連の手法を提案している。同文献 [3] では、まず 2 画像間の変形をアフィン行列で表せると仮定し、変形した画像における画素の近傍領域における特徴点の対応付けを限定する大局的幾何制約 (アフィン制約) により、初期的な特徴点追跡を行う。この追跡

結果をもとに、非線形最適化計算を行い、特徴点位置とカメラ位置・姿勢を推定する。得られた特徴点位置とカメラ位置・姿勢をもとに、特徴点の存在する奥行き範囲を限定する大局的幾何制約 (奥行き制約) を導入し、限定した範囲内を画像上で検索することで、より信頼性の高い特徴点追跡を行う。以上の処理を反復することで明らかな誤対応を確実に排除し、精度を向上させ、幾何学的に正しい 3 次元復元を行う。最終的に、信頼できる特徴点のみから構成されたテクスチャマッピングを行った 3 次元モデルを生成する。

3.2 Lifting 2d Human Pose to 3d : A Weakly Supervised Approach

文献 [4] は、人物の 2D ポーズから 3D ポーズを推定するための手法を提案するものである。一般的に、シングルビューから人物の 3D ポーズを推定することは、人物の姿勢の多様性と複雑性や、奥行きを復元する際の固有の曖昧さのため困難な問題である。そこで同文献 [4] では、3D ポーズの Ground Truth データと 2D ポーズの Ground Truth データの組み合わせに対して、弱教師あり学習のディープニューラルネットワークを用いて、人物の 2D ポーズから 3D ポーズを効果的に予測する方法を提案している。また、3D ポーズの Ground Truth データが存在しない場合の 3 次元姿勢推定の精度向上のため、人体解剖学に基づくより幾何学的な制約を導入している。

4 検討手法

この章では、指定画像内にいる人物の検出と対象人物の 3D モデル化システムの検討手法について述べる。まず 4.1 節で、開発ツールや使用した技術について述べる。その後 4.2 節で、検討したシステムの具体的な構成を提示する。

4.1 開発ツール

本研究で開発したシステムの主要なツールについて記述する。本研究では、Python 3.10.11 を使用して開発を行った。実装に際して利用した主要な外部ライブラリのリストを、表 1 に示す。

表 1 主に使用した外部ライブラリ

Python : 3.10.11

Package	Version	Package	Version
numpy	1.26.3	rembg	2.0.53
pandas	2.1.4	scikit-learn	1.4.0
opencv-python	4.9.0.80	matplotlib	3.8.2
opencv-contrib-python	4.9.0.80	openpose (GPU release)	1.7.0
pillow	10.2.0	dreamgaussian	

4.2 検討したシステムの構成

本研究で検討したシステムの概略図と、各機能の概要について述べる。

本システムにおける、指定した画像内にいる人物の検出と、対象人物の 3D モデルから、モデルのボーンに相当する 2D ポーズ情報の三次元復元を行うシステムのディレクトリ構造図を図 2 に、シーケンス図を図 3 に示す。また、本システムでは、図 3 に示したシーケンス上の 3D モデルの方向別撮影において、撮影方向をモデルの正面と、左右 45 度地点の計 3 方向に設定している。

4.2.1 extract.py

extract.py は、OpenPose で検出した全人物のキーポイント情報を用いて、画像内の各人物に対して、マスク画像を作成するためのスクリプトである。このスクリプトでは主に、OpenCV ライブラリ、Pillow ライブラリ、Rembg ライブラリを使用してマスク画像を生成する。

4.2.2 epipolar_geometry.py

epipolar_geometry.py は、DreamGaussian ライブラリで 3D モデルを方向別に 2 次元画像として撮影し、それらの画像を基に OpenPose ライブラリで検出した 2D ポーズキーポイントデータ群を用いて、3 次元復元を行うためのスクリプトである。3 次元復元を行う際には、透視投影変換アルゴリズムや RANSAC アルゴリズムを利用した。ここで、透視投影変換アルゴリズムとは、3 次元空間上の点を 2 次元平面上に投影する際、近方のオブジェクトは大きく、遠方のオブジェクトは小さく描写される座標に変換するアルゴリズムである。このスクリプトでは主に、OpenCV ライ

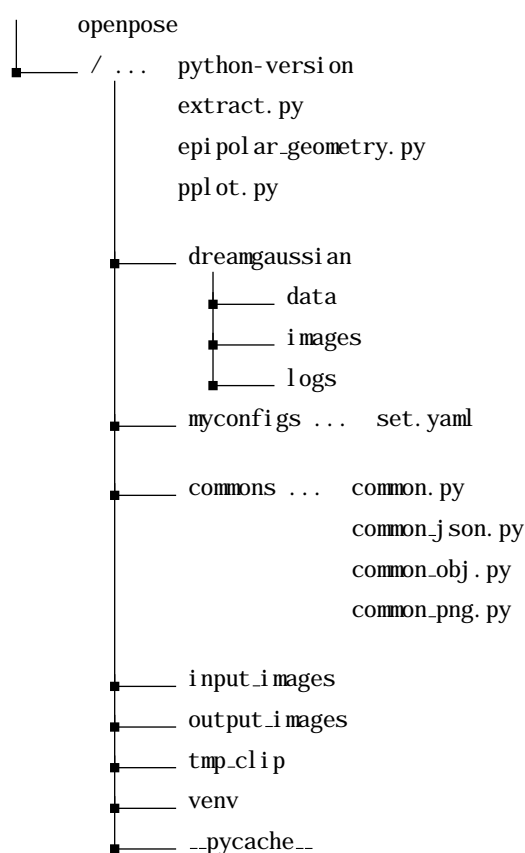


図 2 本システムのディレクトリ構造

イブラリ、scikit-learn ライブラリを使用して、3 次元復元を行う。

4.2.3 pplot.py

pplot.py は、epipolar_geometry.py で行った 3 次元復元の結果を可視化するためのスクリプトである。このスクリプトでは主に、Matplotlib ライブラ

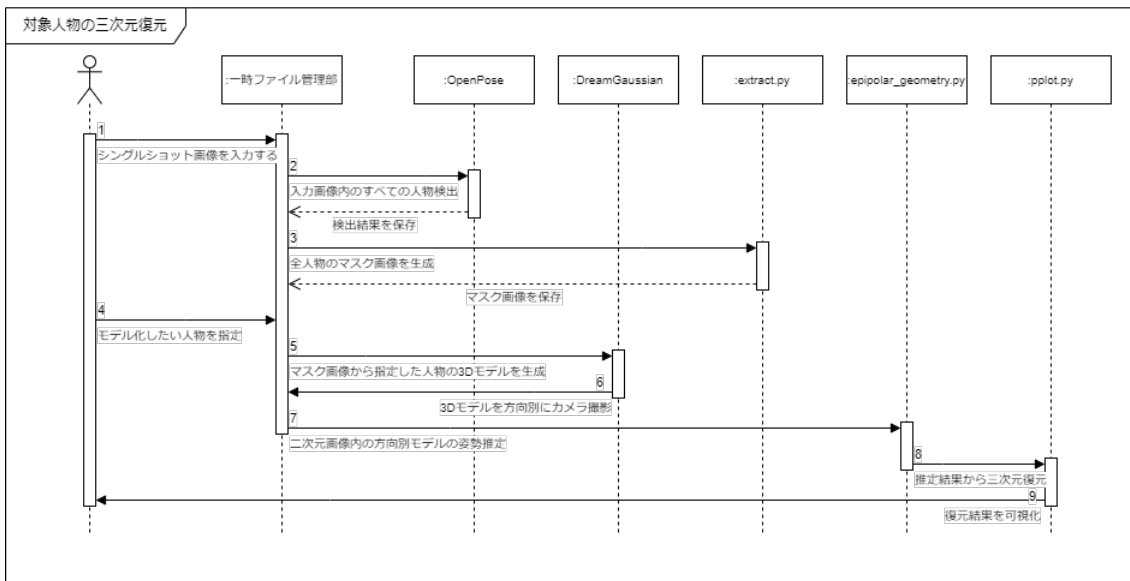


図 3 本システムのシーケンス図

りを使用して、3次元復元の結果を可視化する。

4.2.4 commons

commons は、ファイル操作やデータ整形、初期化処理などの共通処理を行うためのモジュールをまとめたディレクトリである。

5 本研究の成果

本研究では、複雑な姿勢や細部の調整などを可能とするために、指定した画像内の人物を検出し、対象人物のボーン情報を所持した 3D モデルを作成することを検討した。本章では、本研究で検討したシステムの成果について述べる。

5.1 動作結果

本システムの動作を確認するために、図 4 を入力するシングルショット画像として使用した。そして、画像入力のちにボーンの 3次元復元を行った結果、図 5、図 6 に示したような 3D ボーンが得られた。

5.2 動作過程

本システムの動作過程を確認する。入力画像と出力画像は、それぞれ前項の図 4 と、図 5、6 である。また、動作過程の確認において対象人物の選択は、図 8 の人物 A とした。

OpenPose による入力画像内のすべての人物の検出



図 4 入力したシングルショット画像

結果は、図 7 の通りである。各人物のマスク画像生成結果は、図 8 の通りである。選択した人物の 3D モデル生成結果は、図 9 の通りである。3D モデルを方向別に撮影し、2D ポーズ推定を行った結果は、図 10 の通りである。

5.3 評価

本システムの評価は、システムの検証をもとに行った。5.1 節、5.2 節の通り、指定した画像内にいる人物の検出と、対象人物の 3D モデルから、モデルのボーンに相当する 2D ポーズ情報の 3次元復元におおむね成功した。しかし、実際に復元された 3D ボーン 図 5、6 を見ると、外れ値とみられる座標が存在して

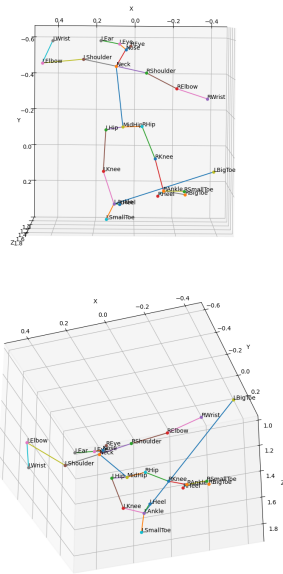


図 5 2D ポーズ座標の 3次元復元結果 1

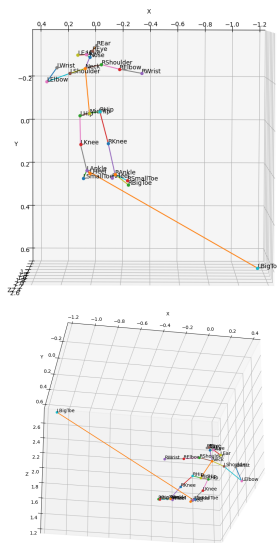


図 6 2D ポーズ座標の 3次元復元結果 2

しまっている。したがって、完全な 3D ボーンの復元を行うためには、ノイズ除去などの処理方法の再思考が必要である。



図 7 すべての人物の検出結果



人物 A



人物 A のマスク画像



人物 B



人物 B のマスク画像



人物 C



人物 C のマスク画像

図 8 マスク画像生成結果

6 まとめと今後の課題

本研究では、指定した画像内にいる人物の検出と対象人物の 3D モデル化システムの検討手法について述べた。検討したシステムでは、指定した画像内にいる人物の検出と、対象人物の 3D モデルから、モデルのボーンに相当する 2D ポーズ情報の 3次元復元を実現することができた。

