

Open3D + MediaPipe による 3Dモーションキャプチャ / アバター駆動

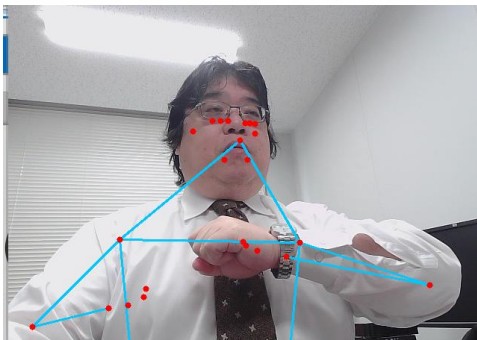
金子邦彦



人体姿勢の取得から3次元アバター駆動まで



カメラ画像
RGB映像・深度

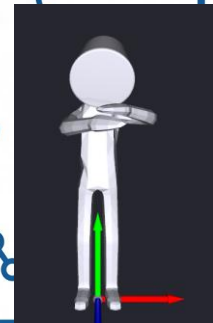
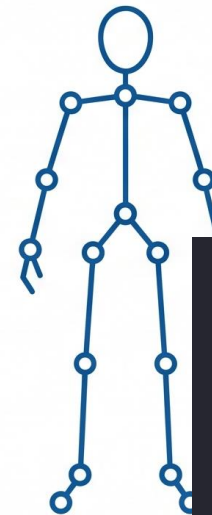


人体姿勢推定
各関節の位置・角度を
数値として抽出

座標系の整合
推定の座標を骨格の基準軸へ
変換・対応付け

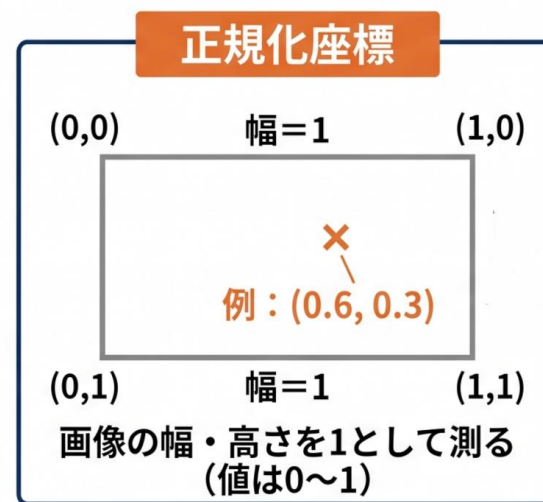
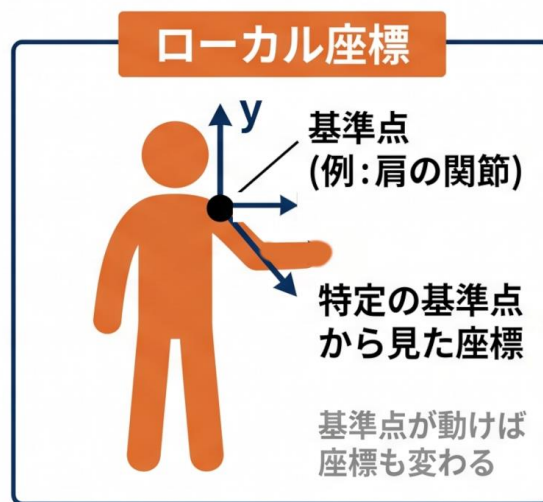
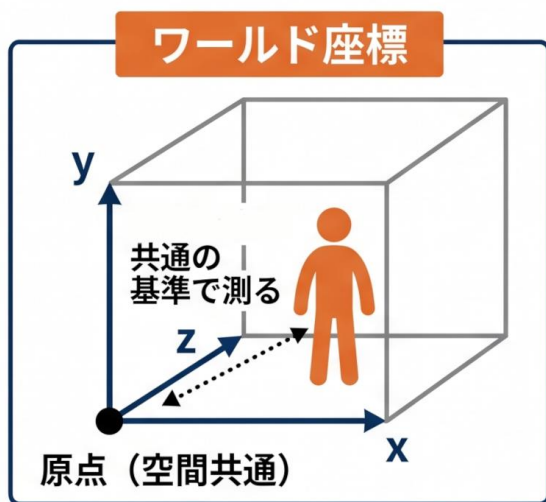
スケルタルアニメーション
骨格の変形で全身を駆動

3次元アバター
骨格・人物像として
可視化し駆動

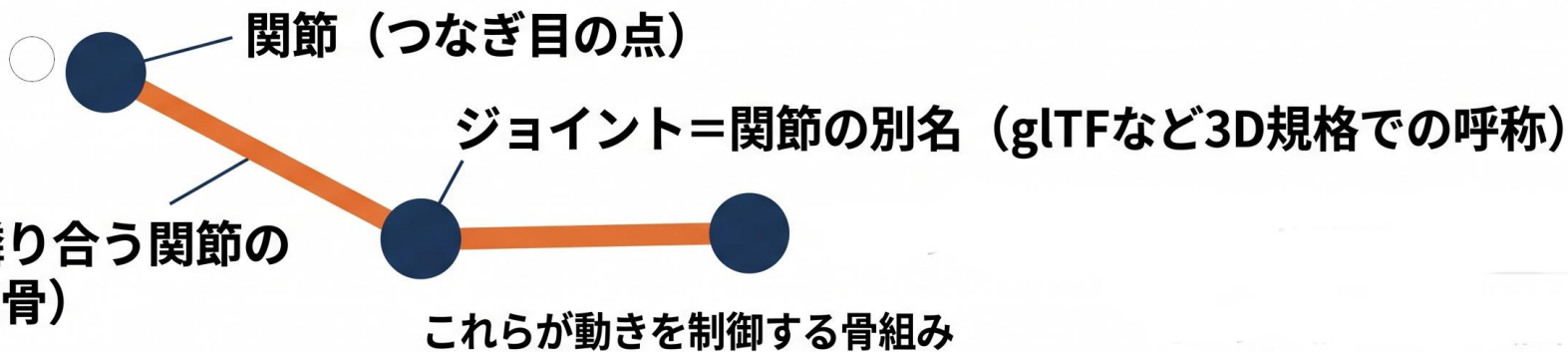


姿勢推定と骨格は別々の座標基準を持つため、両者をそろえる『座標系の整合』が成立条件となる。
整合が崩れると姿勢が正しく伝わらず、アバターが破綻する。

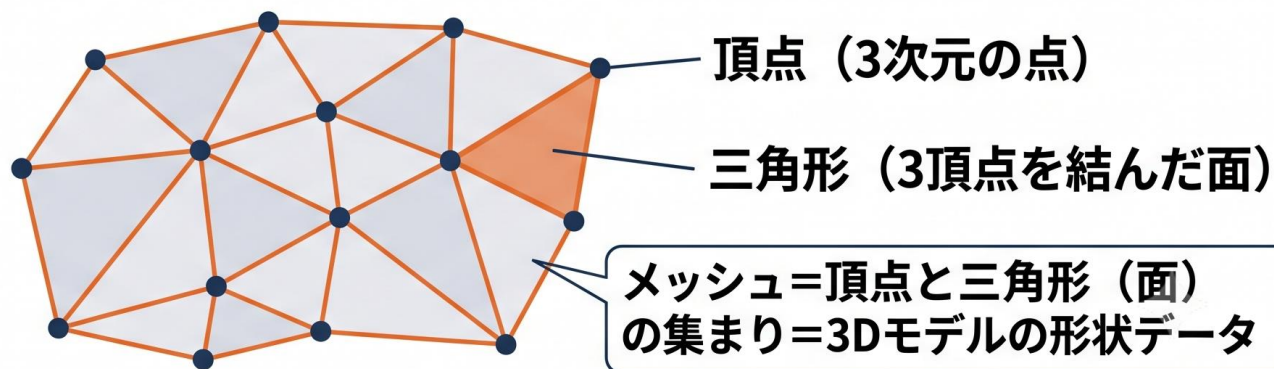
3次元の位置を測る基準



関節・ボーン・ジョイント・メッシュ



形そのものは下の“メッシュ”が担う



骨格 (関節・ボーン) で動かし、メッシュ (頂点と三角形) で形を表す

変形は『基準姿勢からの差分』として計算する



差分 (どれだけ動いたか)



この差分が変形量になる



基準姿勢 (バインド姿勢)

モデルが最初から持っている基準の姿勢

変形後の姿勢

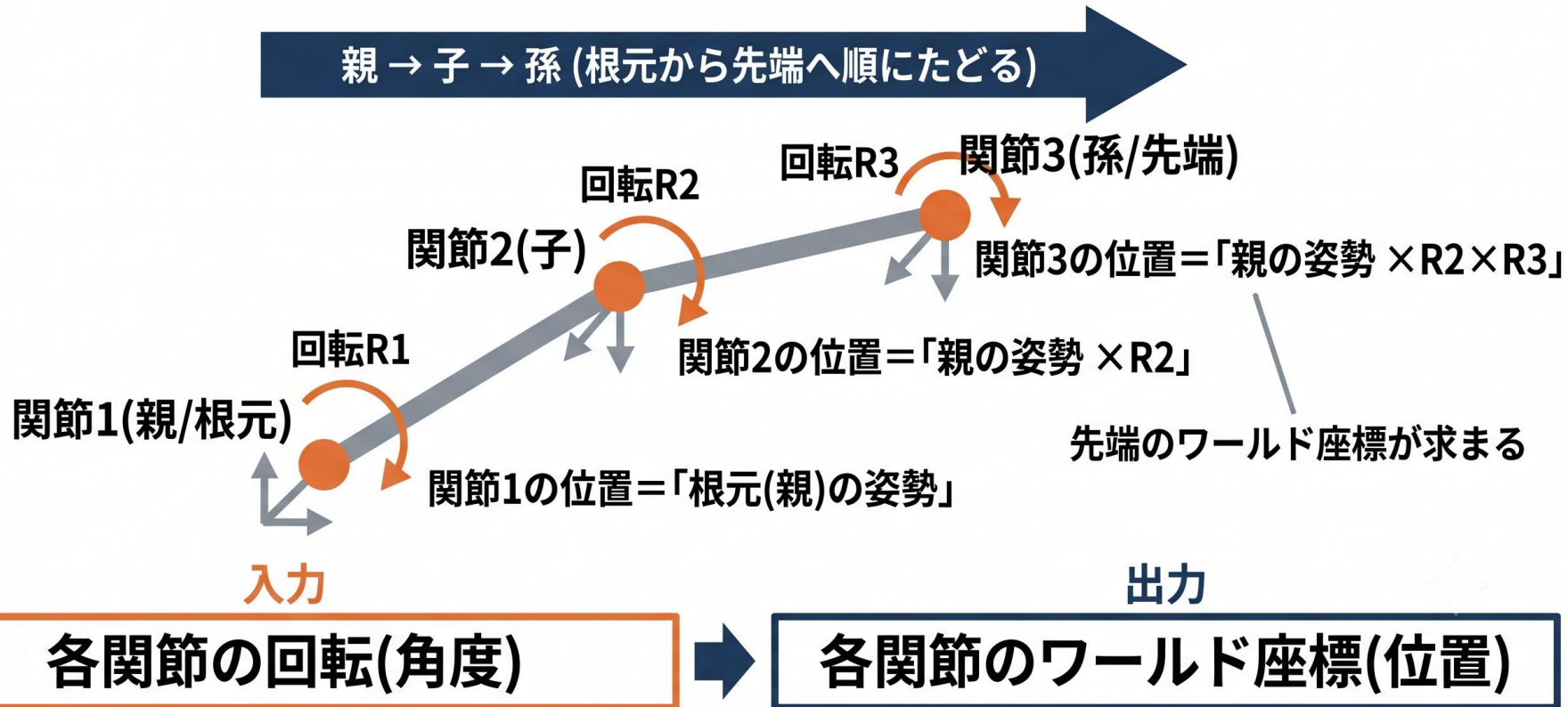
変形 = 変形後の姿勢 - 基準姿勢

基準姿勢を出発点とし、そこからの差分でモデルを駆動する

順運動学 (Forward Kinematics)



各関節の回転を親から子へ積み重ね、各関節のワールド座標を求める手法



リニアブレンドスキニング



各頂点を複数ジョイントから重みつきで変形し、メッシュを骨格の姿勢に合わせる手法

入力データ

頂点の初期位置
 v (バインド姿勢)

ジョイントの
ワールド変換 M_j

逆バインド行列 B_j^{-1}
(バインド姿勢→ジョイントの
ローカル座標へ移す行列)

※詳細はglTFの項を参照

頂点ごとの計算

$$v' = \sum_j w_j \cdot M_j \cdot B_j^{-1} \cdot v$$

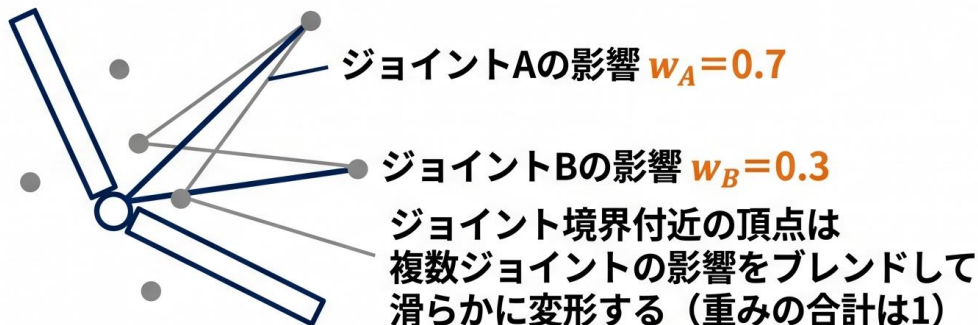
ジョイントjの重み (合計=1) ワールド変換 逆バインド行列 頂点の初期位置

各ジョイントが頂点を動かした結果を、
重み w_j で混ぜ合わせて最終位置を得る

変形後のメッシュ

最終頂点位置 v'

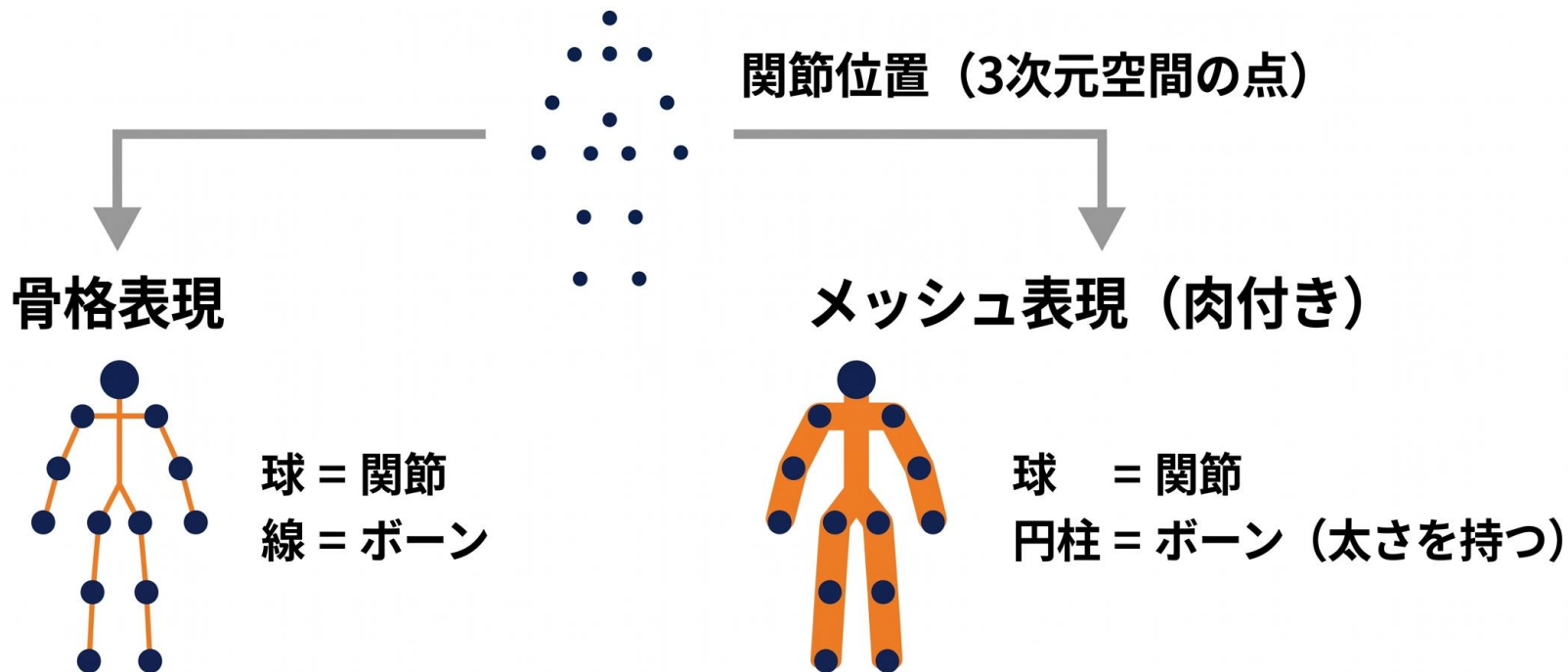
メッシュが骨格の姿勢に
追従して変形



骨格表現での描画、メッシュ表現での描画

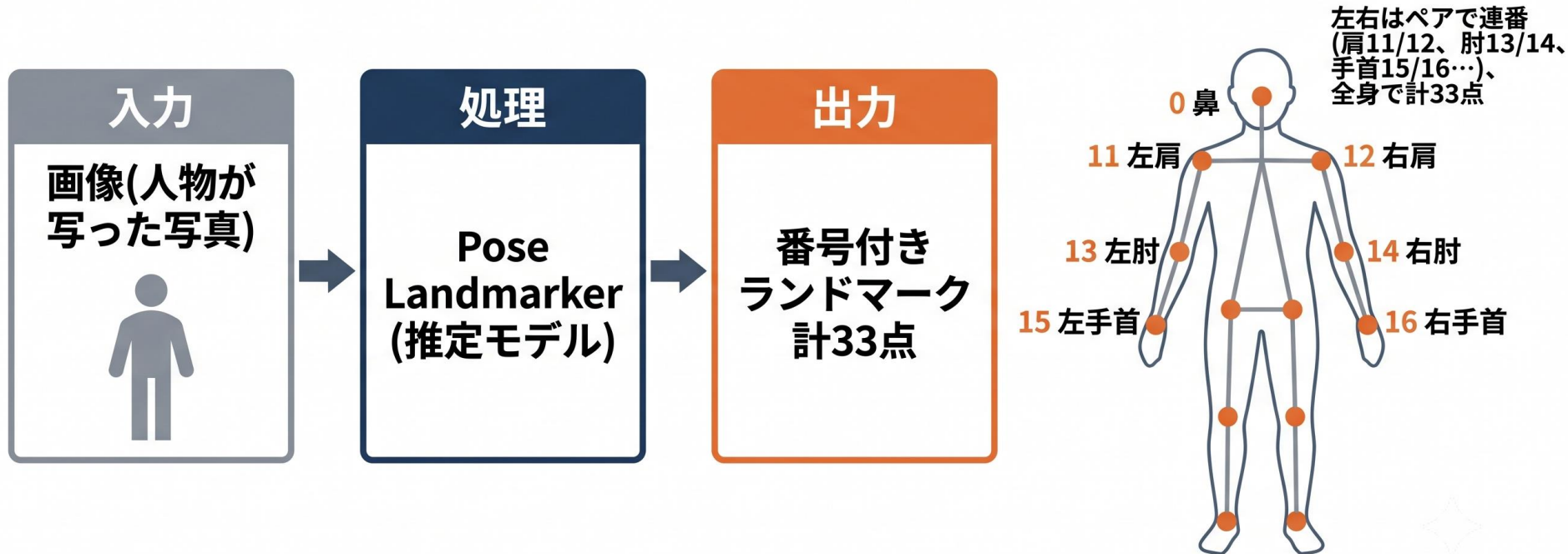


同じ関節位置を、描画する形の選び方で2通りに可視化できる



線を円柱に置き換えるだけで、骨格から肉付きのある表現へ変わる

入力画像から人体の関節位置(ランドマーク)を推定する仕組み



画像を入力すると、各関節が番号付きの座標(ランドマーク)として返る