

入力インターフェースとしての3D認識処理

亀田嘉人 土井滋貴 奈良工業高等専門学校

画像認識分野でのAI的な処理系の進歩は目覚ましいが、立体の認識についてもいくつかの手法が検討されている。今回は立体認識の手法の1つであるPointNetの実装を試み、さらに、PointNetのネットワークの途中から立体の角度情報を取り出し、入力インターフェースとして使用する方法を紹介する。

はじめに

近年、自動運転や農業用ロボットの需要の高まりと共にAIによる画像認識技術の向上が必要とされている。しかし現在使用されているYOLOなどの二次元ベースの画像認識では得られる情報が限られており、システム全体が複雑になる傾向にある。そこで本研究ではカメラから得た三次元情報に対して直接解析を行い、物体認識を実行することを目的とする。この手法は果樹収穫ロボット(図1)などへの応用が期待できる。また立体の傾き情報を1つの認識ネットワークで得られることは、ゲームのコントローラのインターフェース(図2)としても利用できる可能性がある。

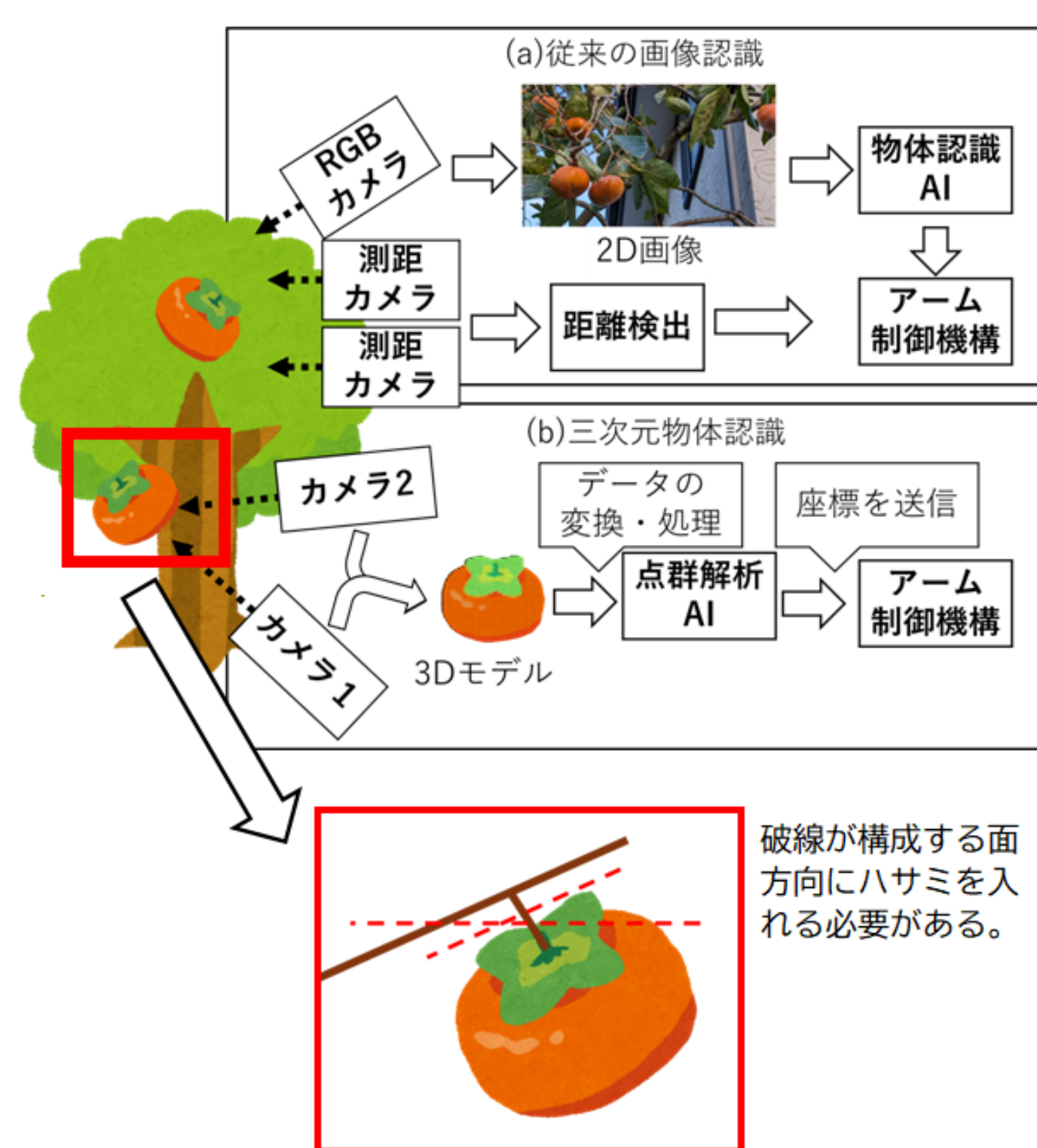


図1 3D認識系の果樹収穫ロボットへの導入



図2 3D認識系のゲームコントローラへの応用

3D認識系の実装

既にいくつかの立体情報の認識系は提案されているが、ここでは基本的な処理系であるPointNetを利用する。深度カメラのRealSenseとPointNetを組み合わせた実物体の認識系の処理例を示す(図3)。使用したコンピュータ環境はIntel Core i5 6500とNVIDIA GeForce RTX3050である。一般的なディープラーニングを用いた画像処理と比べるとやや高い性能が必要となる。

傾き情報の取り出し

PointNetではネットワークの前段と中段においてT-Netと呼ばれる補正のためのネットワークが用いられる(図4)。T-Netでは図形の正規化が行われる(図5)。T-Netで行われた正規化のパラメータは補正行列として取り出すことができ、この補正行列を用いれば物体の傾き情報を得ることができる。

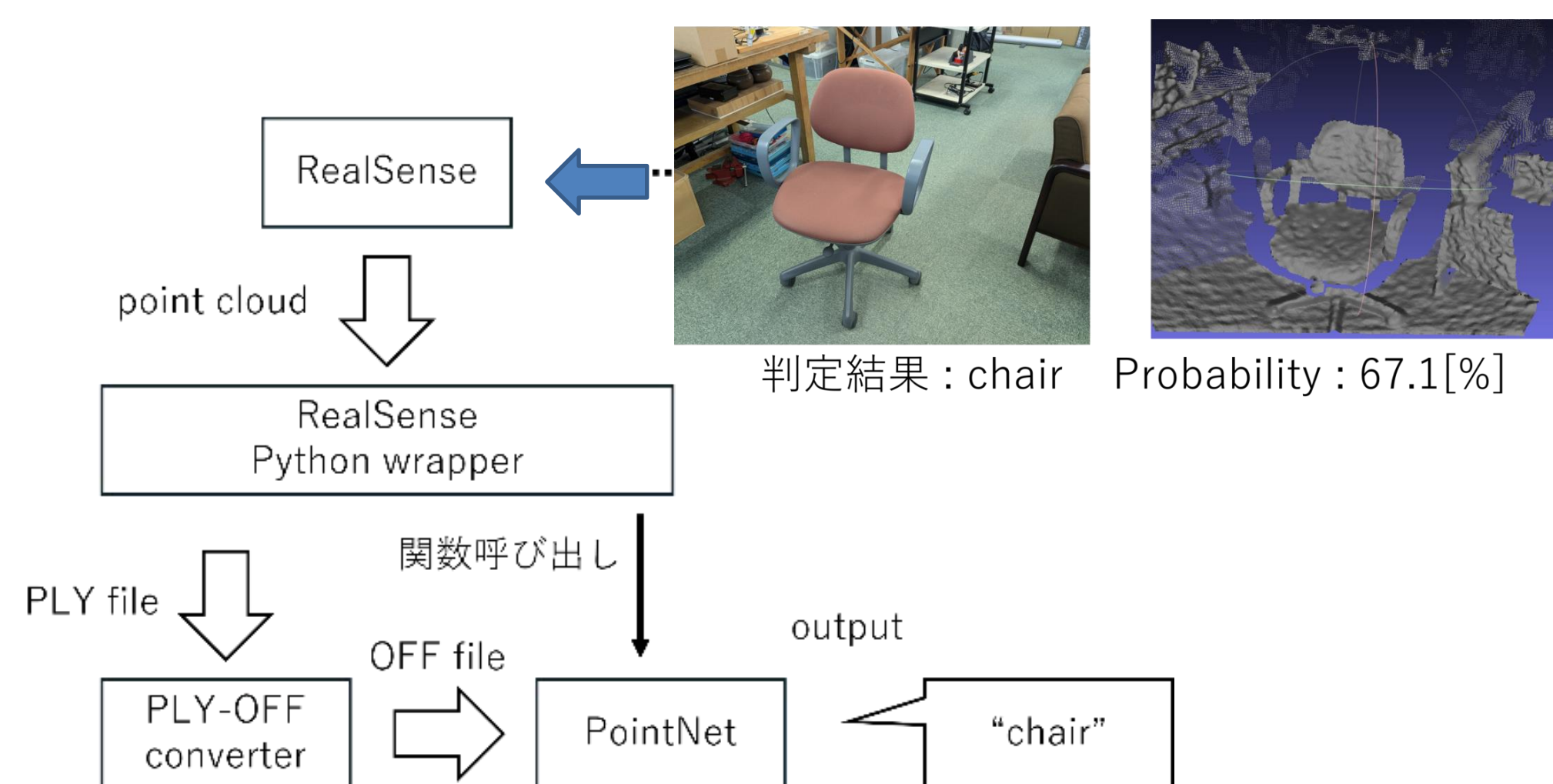


図3 深度カメラRealSenseから取得した椅子の3DデータをPointNetに入力し認識させる

球、円柱、円錐を学習させて、傾けた立体を認識させた場合の角度出力を示す(図6)。特異値分解(SVD)を適用した場合、ほぼ1に近づいていることが分かる。ModelNet10の中のモデル「ソファ」を傾けた場合の角度出力を示す(図7)。この場合もほぼ1に近づいていることが分かる。しかしながら、狭い範囲でしか比例しないことやほかの軸も変化してしまうなど、十分な精度がえられているとは言い難い。

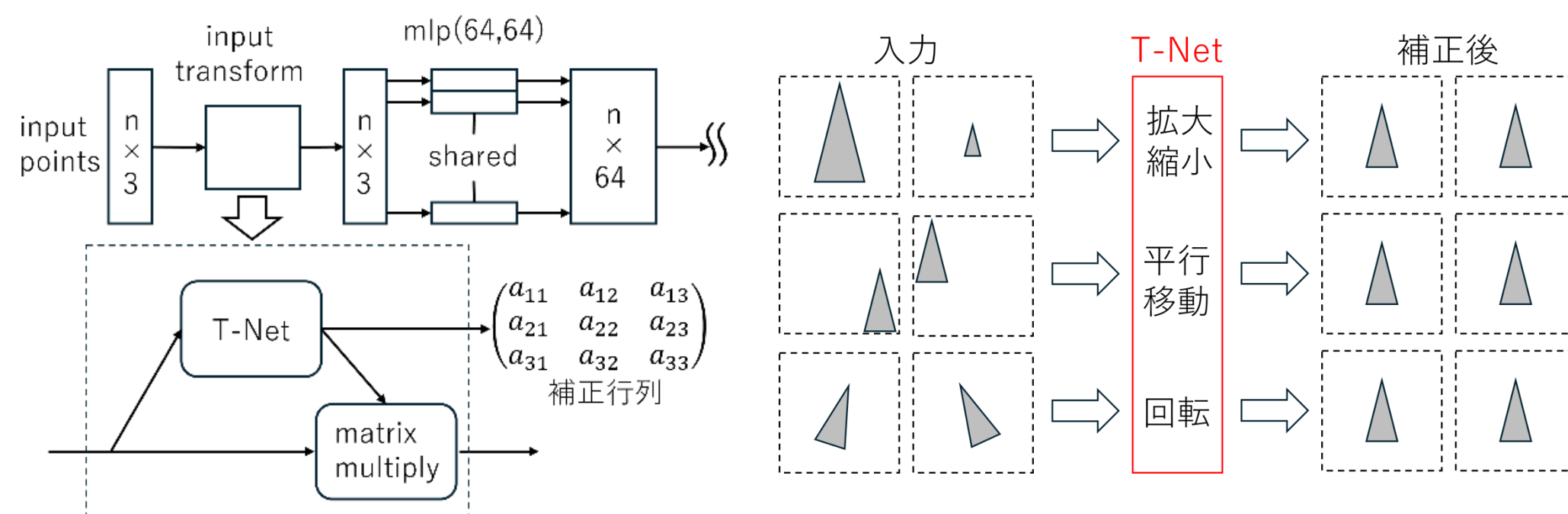


図4 PointNetの構造

図5 立体の正規化

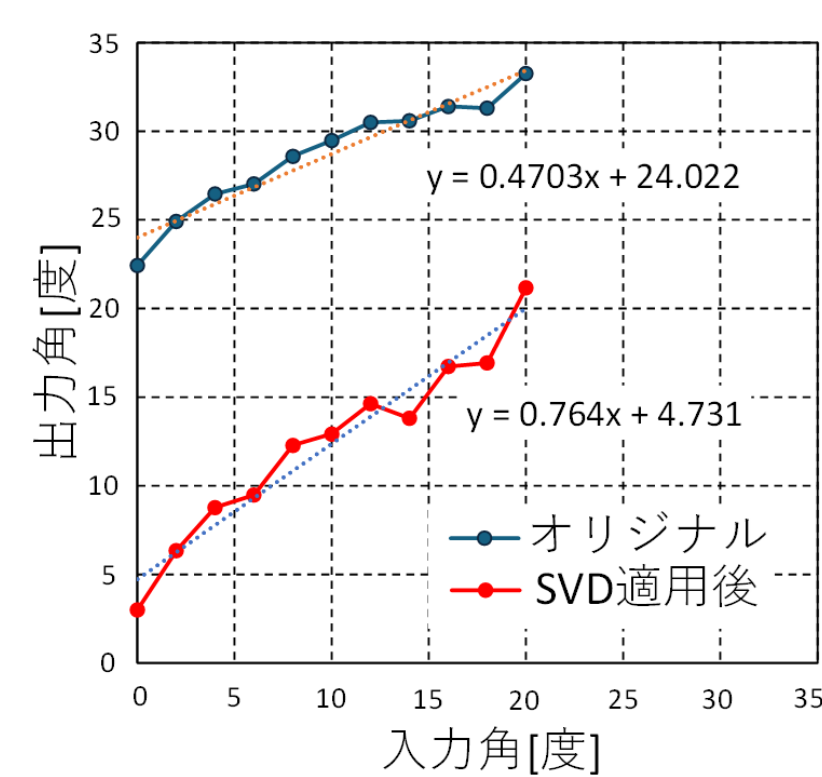
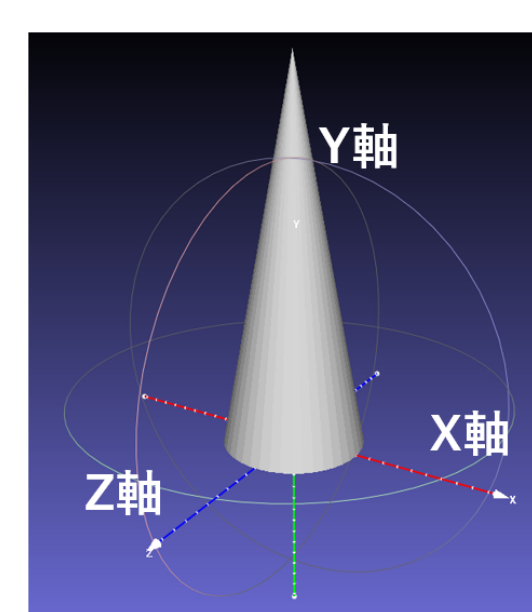


図6 円錐を傾けた時の取り出された傾き情報

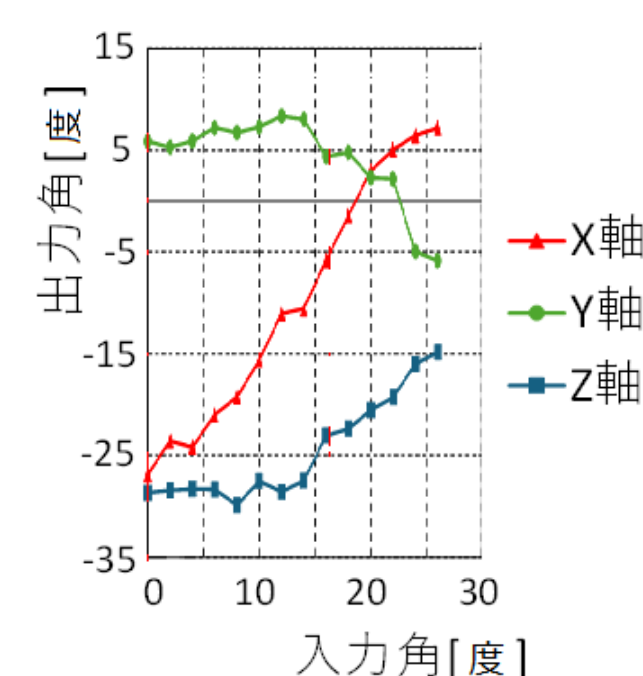
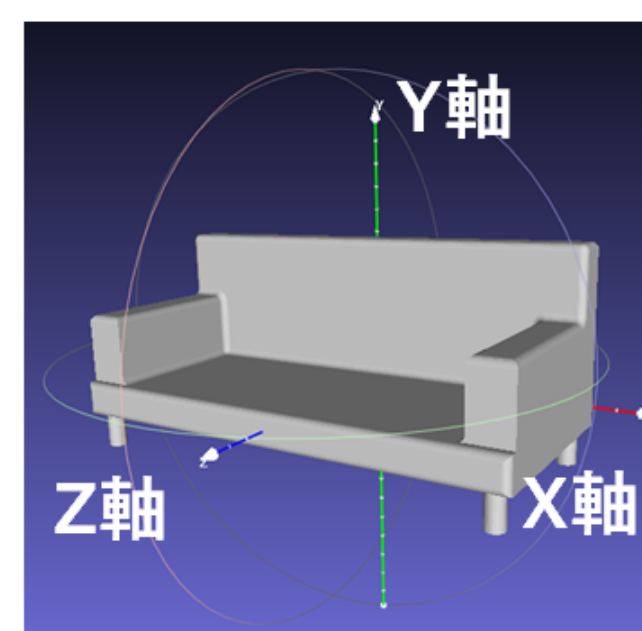


図7 「ソファ」を傾けた時の取り出された傾き情報

まとめ

「傾き情報の取り出し」で示したように、現状の処理系では十分な精度が得られているとは言い難い。しかしながら、立体の認識およびその傾き情報を直接的に取り出す手法としての有用である可能性が確かめられた。