

上方に設置した深度センサを用いた人物姿勢推定の検討

高木 和久[†] 橋本 敦史[†] 松富 卓哉^{††} 飯山 将晃[†] 美濃 導彦[†]
[†] 京都大学 ^{††} 奈良先端科学技術大学院大学

1. はじめに

深度センサによる姿勢推定の成功により、動作による認識やジェスチャによるインターフェイスを始めとした様々な応用が急速に発展している。一方で、現状の姿勢推定技術は人物を低い視点から水平方向に観測することを前提としており、街角の監視カメラのように高い位置からの姿勢推定の検討は進んでいない。本研究では、上方に設置したKinectなどの深度センサから得られる、カメラから視野内の物体までの距離を各画素に保持した深度画像を用いて、写っている人物の姿勢推定を行うことを目指す。上方にカメラを設置することで、下半身は見え難くなるものの、人同士や人と環境中の物体とのインタラクションに関わる上半身の姿勢については、人ごみの中でも観測できるという利点を得られ、商業施設などの環境で深度センサを役立てることができる。

2. 深度画像を用いた人物姿勢推定

センサを水平方向に設置した場合に、深度画像から人物姿勢推定を行う手法として、Shottonら[1]の手法が知られている。この手法は推定を開始するにあたって、決められたポーズをとる必要がなく、客や歩行者を対象とする応用が可能である。Shottonらの手法では、姿勢推定の前段階として、各画素が人体のどの部位に属するのかをその周辺2画素の間の深度の差を用いて分類している。また、対象画素と周辺2画素それぞれとの間隔を、対象画素の深度により正規化している。これらにより、センサが水平向きの場合には、視野内の人物位置変化に頑健な推定を行うことが出来る。

一方で、上方から撮影する場合には、図1のように透視投影の影響を受け、人物が同じ姿勢であっても立ち位置に依存して、周辺2画素間の深度の差が大きく変わる。このように、Shottonらの手法は、特徴が立ち位置に依存してしまう問題を考慮していないため、精度が低下することが予想される。

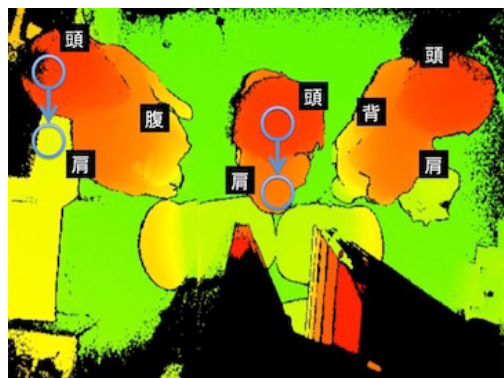


図1: 透視投影による見えの変化。
丸印は特徴量における2つの比較対象画素。

3. 予備実験

Shottonら[1]の手法における透視投影の問題を検証するために、地上から3mの高さに、斜め下方向45度の向きでセンサが設置してある状態を想定し、人物位置による部位分類結果の変化を確認する実験を行った。この状態を再現するために、3DCGソフトを用いて深度画像を生成した。また、学習には1000画像で20万画素を学習サンプルとして用い、学習サンプルにおいて、人物の位置は視野内の床の上を一様に動くこととした。図2に結果を示す。図内右下が図中央の人物の部位分類の正解例である。右手や右脚に着目すると、中央の人物の部位分類の方が、左端のものに比べて精度が高いことが確認された。

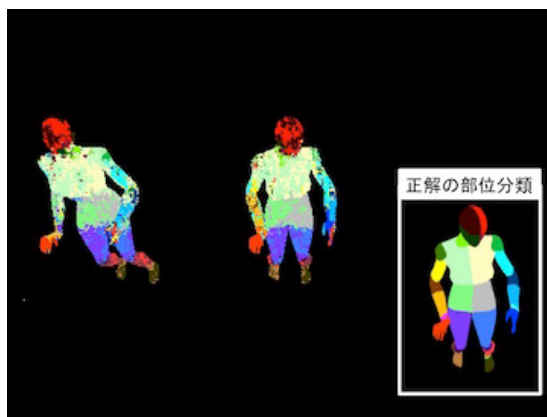


図2: 人物姿勢推定の結果。

4. 視点統一による人物姿勢推定

立ち位置依存性の問題を避けるには、視点統一を行うことが有効と考えられる。深度画像は、視野内に写っている物体のカメラ側の面の3次元位置の情報を保持しており、この情報を用いて3D点群データを生成することができる。どの向きのカメラから得られる3D点群データも、水平方向や鉛直方向など常に同じ視点からのデータとして扱うことにより、立ち位置に依存せずに同様の特徴が抽出できる。

ここで、このアプローチでは、視点変換により欠損画素と重複画素が発生する。欠損画素とは、視点変換を行った上で再び新たな視点からの深度画像として再投影した場合に値が不明となる画素のことであり、重複画素とは、その深度画像において2つ以上の値が入る画素のことである。欠損画素への対処法や、重複画素の値を出来るだけ上手く使う方法は、今後検討する予定である。

参考文献

[1] Ross Girshick, Jamie Shotton, Pushmeet Kohli, Antonio Criminisi, and Andrew Fitzgibbon. "Efficient regression of general-activity human poses from depth images." 2011 IEEE ICCV, pp. 415-422.