

現場安全性・生産性向上のための深層学習とコンピュータビジョンによる 人物・位置・動作の認識に関する研究

Research on recognition of person, position and movement for improving on-site safety and productivity by deep learning and computer vision

○上田 航平^{*1}, 本田 悠樹^{*1}, 加戸 啓太^{*2}, 平沢 岳人^{*3}
Kohei Ueda^{*1}, Yuki Honda^{*1}, Keita Kado^{*2} and Gakuto Hirasawa^{*3}

*1 千葉大学大学院 融合理工学府 博士前期課程

Graduate Student, Graduate School of Sci. and Eng., Chiba University.

*2 千葉大学大学院 工学研究院長 助教

Assistant Professor, Graduate School of Engineering, Chiba University.

*3 千葉大学大学院 工学研究院 教授

Professor, Graduate School of Engineering, Chiba University.

キーワード：深層学習；コンピュータビジョン；Azure Kinect DK；Openpose；Recap photo

Keywords: DNN; computer vision; Azure Kinect DK; Openpose; Recap photo.

1 はじめに

工事事務において、全産業の死亡者数に占める建設業の割合は最も高く 30%~40%で推移しており、死亡者は毎年減少傾向にあるものの依然として高い比率のままである¹⁾。死亡事故の主な発生原因として墜落・転落・交通事故・はさまれなどが挙げられるが、死亡事故の他にも軽微な事故やヒヤリハットも含めると膨大な量の例が発生している。施工現場や作業場ではこうした事故を未然に防ぐため、安全教育・訓練²⁾・様々な約束事³⁾が決められている。しかし、毎日変化する作業員の健康状態や作業内容に対する意識、天候、その周辺環境、工期、予算など⁴⁾いくつもの要因が組み合わさり依然として事故は発生してしまう。事故要因の一つとして、施工現場や作業場での作業の複雑性が挙げられる。例えば、現場では複数の作業員・複数の業者が何度も入れ替わり作業するので作業間の連携が難しく、現場の状況を漠然としか把握できないのでリアルタイムに監視・管理することが難しい。こういった事故に対し近年では、作業員にセンサ等のデバイスを装着し、心拍数や体温などによる健康状態の可視化・管理や定点カメラ・建機に取り付けたカメラを用いて、危険区域への進入はないかなどの管理・警告を行うなど、作業員の労働環境のモニタリングによって安全性を向上させる取り組みが行われている⁵⁾。このように施工現場や作業場において、作業員がどこにいてどの状態（健康状態や作業位置、実行中の動作）にあるのかという情報は事故を防ぐ上で重要な要素である。

また国土交通省では、ICTなどを施工現場に導入することで生産性向上を図る取り組みである i-Construction

を進めている。その一つとして建設ロボットなどの現場導入・人との協働が挙げられるが、安全性の観点から作業員がどこで何をしているかを把握することは重要な課題である。これらを把握することで、例えば無人搬送車などを走行中、近辺にいる作業員の作業内容や接近中の作業員がオペレータか否かによって軌道や速度を制御するなど作業効率の向上も期待される。

そこで本研究では、施工現場や作業場における安全性向上・生産性向上を目的に単眼カメラによる作業員とその位置、動作の推定方法を提案し、試行結果を報告する。

2 研究手法

近年では、機械学習を用いたパターン認識が耳目を集めており、なかでも単眼カメラから入力された映像を Convolutional Neural Network: CNN とコンピュータビジョンで処理し活用する手法は、センサや深度カメラより「安価で入手可能」、「キャプチャ範囲が広い」、「設置が簡易である」という特長から様々な分野で活用されている。そこで本研究ではこれを採用し、単眼カメラ・CNN・コンピュータビジョンで、作業員の人物・位置・動作を推定する手法の開発を試みる。図1は映像から人物・位置・動作を推定するまでのシステム図である。一般にこのような画像認識、本研究における人物や動作を推定する CNN は教師あり学習が用いられるが、教師データ（学習用データ）の収集が課題となる。本研究ではこの課題にも着目し上図のようなシステムを試行した。カメラの入力は Openpose で前処理し関節を抽出しておき、そこから顔付近の映像、足首位置、各関節のシーケンスを取得

し、人物・位置・動作を判別する。

Openpose とは、単眼カメラによる深層学習を用いた人物の関節推定アルゴリズムのことである。多人数でも推定可能で、各関節部位を推定してそれらを繋げるボトムアップのアプローチ方法を採用している (図2)。

カメラ座標系における計 18 の各関節位置とその認識精度を示す信頼度 (0.0~1.0) を取得可能である (図3)

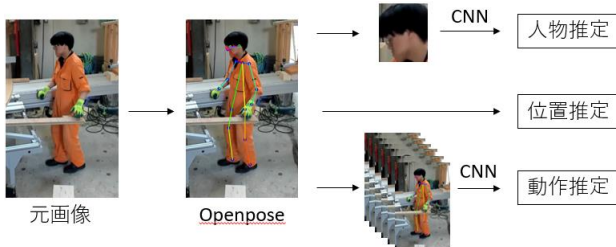


図1 システム図

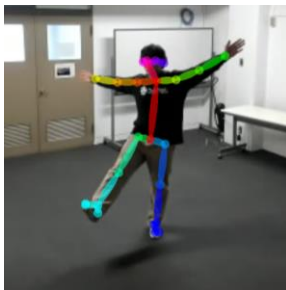


図2 関節推定の様子

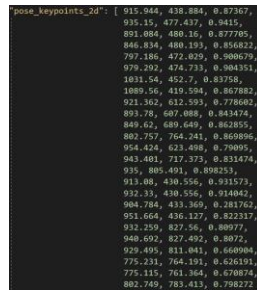


図3 各関節位置と信頼度

3 人物の推定

3.1 学習データ

作業員推定の学習データは下記の手順で作成する。

1. 作業員の顔の周囲を動画で撮影する (図4)。
2. 動画データから静止画を切り出し、Recap Photo を用いて作業員の頭部を 3D モデルとして再現する。
3. このモデルを多角度からキャプチャし、施工現場や作業場を想定した背景に合成する (図5)。キャプチャ角度は roll 角 (-10° ~10°)、pitch 角 (-30° ~40°)、yaw 角 (45° ~135°) をランダムに組み合わせたものとした。背景画像は当大学が保有する作業場内を撮影したものを使用した。
4. 運用時の環境・解像度・Openpose による切り出した

画像のブレを想定しぼかしを加え、ランダムに切り取り、ランダムに明度を変化させたものを学習データとした (図6)。

本稿では、当研究室の学生 7 人を対象に分類を行い、一人に対して 25,000 のデータを用意した。



図4 元データ



図5 3D モデル

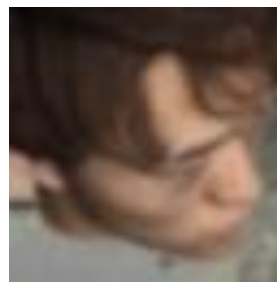


図6 学習データ

3.2 Recap photo

Recap photo とは、Autodesk 社がリリースしている Recap Pro に含まれるサービスであり、オブジェクトなどに対して複数撮影した写真を処理し、3D 表現を作成することができる (図7)。



図7 元画像 (左) と作成した 3D モデル (右)

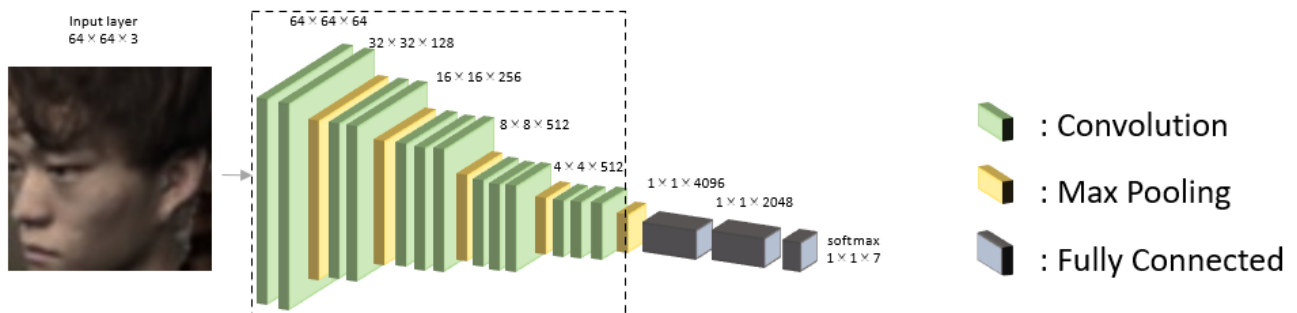


図8 分類モデル

3.3 分類モデル

VGG16の最下層の畳み込み層(図8点線内)までの重みデータを利用し、全結合層3層を追加して転移学習させた(図8)。

VGG16とはオックスフォード大学が2014年に提案した畳み込み13層・全結合3層の計16層からなるCNNモデルのことで、ImageNetと呼ばれる大規模画像データセットで学習されたものである。

4 位置の推定

現場における相対的な位置が既知の平面上4点を定義しカメラ座標と対応させ、射影変換することにより施工現場や作業場における作業員の位置を算出する。推定に用いる箇所はOpenposeから抽出した足首関節の中心とした(図9)。

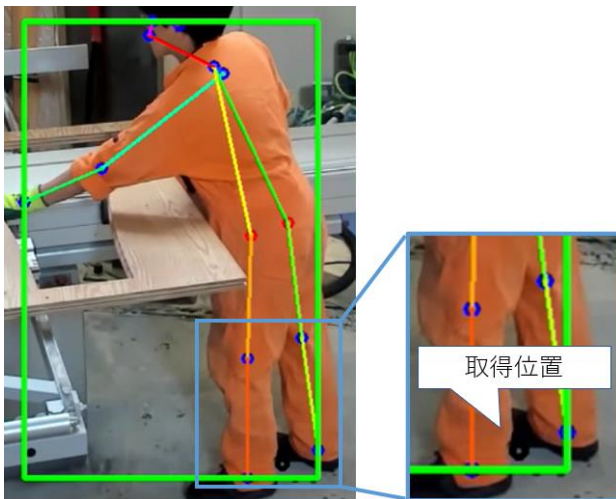


図9 位置の推定

5 動作の推定

5.1 学習データ

動作推定の学習データは下記の手順で作成する。

1. Azure Kinect DKを用いて、ある動作中の各関節の三次元座標を取得する(図10)。また、被写体とAzure Kinect DKとの距離は2mとし、高さ1mの位置から撮影。Openposeから抽出した各関節位置と対応するもののみを抽出し、取得フレームは1動作につき3600フレームとした。これをランダムな位置から30フレーム分切り出す。
2. 体形や身長による各関節位置の個人差を考慮し、三次元座標上で各関節の値にノイズを加える。
3. 運用時、Openposeから得られる関節位置にはジッターが含まれるので、これを想定し二次元座標へ透視変換後、各関節点に二次元座標上でノイズを加え、間引きを行う(図11)。また運用時、Openposeから抽出した各関節の信頼度が0.2未満のものは推定から除外する。

4. 各関節の最大値・最小値をもとに-1.0~1.0の範囲に正規化した(図12)。

本稿では、作業場内での諸作業を想定し、分類動作は「立ち止まっている・しゃがんでいる・かがんでいる・掃き掃除をしている・工具で材料を切っている(スライドソーを使用している)」とし、一動作に対して約7000のデータを用意した。

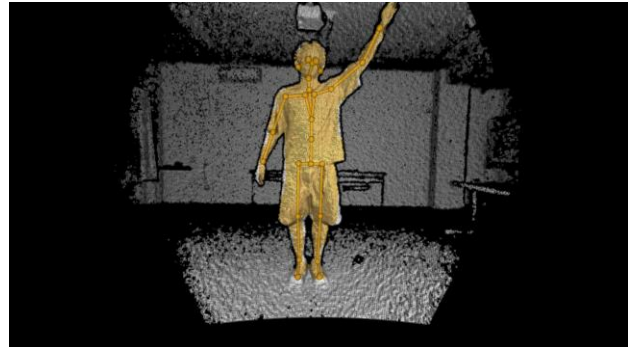


図10 Azure Kinect DKで関節位置を取得している様子

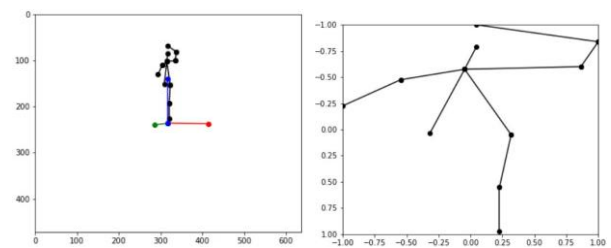


図11 三次元の学習データ 図12 正規化した学習データ

5.2 Azure Kinect DK

Azure Kinect DKとは、Microsoft社がリリースした開発者向けデバイスであり、深度センサ・マイク・RGBカメラ・IMUから構成されている(図13)。本研究では、深度センサ(モード:NFOV Unbinned、解像度:640×576、FOI:75°×65°)のみ利用した。

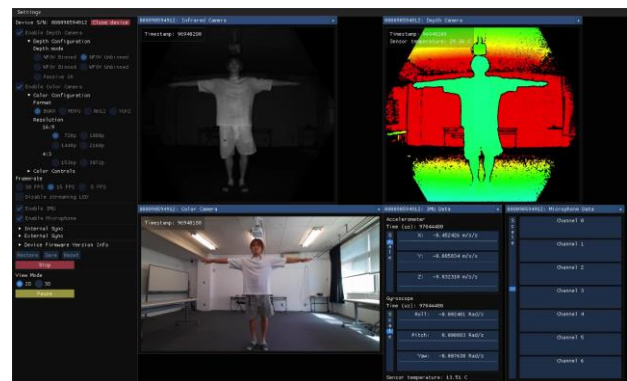


図13 Azure Kinect DKの動作の様子

5.3 分類モデル

三つの畳み込み層と二つのプーリング層、三つの全結合層からなるモデルを作成し学習させた(図14)。

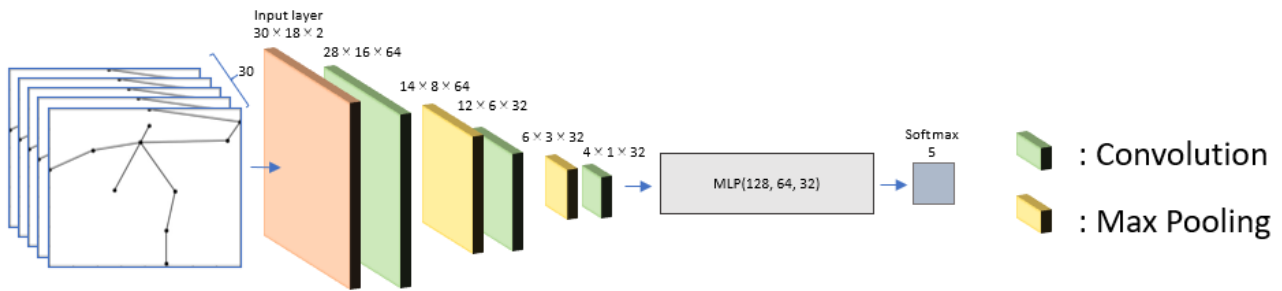


図 14 動作分類モデル

6 検証

当大学の作業場内で検証を行った（図 15）。

作業員の推定に関して、分類可能であったがカメラに背を向ける状況や被験者の頭部を Openpose で正しく切り抜きできない状況では、精度が著しく低下した。

位置の推定に関して、誤差±100mm 以内で推定可能であったが、足部が隠れてしまう状況では推定不能。

動作の推定に関しても分類可能であったが、Openpose の推定精度により動作の推定精度も左右された（図 16）。



図 15 工房での検証の様子



図 16 それぞれの推定の様子

7 考察

作業員の推定に関して、学習データ作成時の 3D モデルが頭部の前面部しか再現できなかったため、運用時、作業員がカメラに背を向けたような状況では Openpose から切り出した画像データと学習データとの間に乖離が発生し、低精度となったと考えられる。また、人物 1 人に対し 1 つのモデルでデータの水増しを行ったため、髪型・メガネの有無でも精度が左右されたと考えられる。

位置の推定に関して、誤差±100mm 以内は有効であると思われる。しかし作業場では、資材や工具等で足部が隠れることが多いので、頭部など別の部位で位置を推定することで、より有効となると考えられる。

動作の推定に関して、大きな工具を使用するとカメラ位置によって身体の大部分が隠れてしまうため、Openpose によって得られる関節点の推定精度が下がり、動作の推定精度も下がってしまった。

8 まとめ

施工現場や作業場では、どの作業員がどこで何をしているのかという情報は事故を未然に防ぐ上で重要な要素である。さらに、ロボットの現場導入の観点から作業効率を向上させる上でも重要である。そこで本研究では、単眼カメラを用いた作業員とその位置、その動作の推定手法を提案し、試行結果を報告した。

検証した結果、それぞれの推定は可能であったが、精度は Openpose による関節位置の推定精度に左右された。

【参考文献】

- 1) https://www.hrr.mlit.go.jp/gijyutu/seisan/H30zenki/seisan_H30_s5.pdf (2020/09/14 アクセス)
- 2) 広兼、他：安全教育における危険予知訓練について、土木学会論文集, 2010. 2
- 3) 山田、他：安全標識の認知規定要因に関する研究、日心第 74 回大会, 2010
- 4) 渡邊、他：建設現場施工における安全と生産性に関する一考察、建設マネジメント研究論文集 Vol. 5, 1997
- 5) M. Wang, P. Wong, H. Luo, S. Kumar, V. Delhi, J. Cheng, Predicting Safety Hazards Among Construction Workers and Equipment Using Computer Vision and Deep Learning Techniques, ISARC 2019