

部分欠損した骨格データにおける GAN を用いた 補間手法に関する検討

雲 河晨[†] 中村 悦郎[†] 景山 陽一[†] 石沢 千佳子[†]

加藤 伸彦[‡] 五十嵐 健[‡] 川本 健[‡]

秋田大学[†] エイデイケイ富士システム[‡]

1. 背景・目的

近年、日本における労働者の高齢化の進行に伴い、特に建設現場などにおける高齢労働者の安全管理を行うことが求められている¹⁾。また、高齢労働者に限らず、労働者の行動や身体状態を収集し、分析する安全管理システムは、作業の効率化および安全性向上を図り、安全かつ安心な労働環境の構築に寄与すると考える。

一方、安全管理システムにおける動作認識手法として、背景ノイズを抑制しつつ、動作認識において肢体位置(骨格データ)の時系列的变化を利用する手法が提案されている²⁾。しかしながら、作業環境の状況やカメラの撮影角度によって、カメラの位置から体の部位が物体に遮蔽されることに起因し、身体情報の検出および骨格データの推定が必ずしも良好に行われない場合があり、動作認識精度のばらつきが発生する課題が挙げられる。

本稿では、上記課題を解決するために、10種類の動作および3種類の遮蔽状態を対象として、部分欠損した骨格データの補間手法に関する検討を行った。具体的には、人物の部位および作業環境が遮蔽される場合を対象として想定し、行っている動作の様子から抽出した時系列的骨格特徴点を補間ネットワークに学習させ、部分欠損した骨格データの補間実験を行った。

2. 使用データ

一般的な蛍光灯(照度 500~900lx)の下、3台の単眼カメラ(FDR-AX60:ソニー社製, HC-VX2M:パナソニック社製×2台)を用いて、被験者10名が下記の10種類の動作(歩行, しゃがみ, 立ち上がり, 脚立上り, 脚立下り, 座り, 起立, ものを持ち上げる, 倒れる, 起きる)を行う様子を動画データ(1920×1080pixel, 60fps)として3回ずつ取得した(図1参照)。次に、取得した動画データを動画画像データ(30fps)に変換した。

歩行動作は、被験者正面に配置した1台のカメラに向かって歩く様子を、左・右方向、接近する方向、ならびに遠ざかる方向を撮影した。他の2台のカメラに対しても同様の撮影を行った。また、歩行以外の動作は、3台のカメラで正面、側面、ならびに背面からそれぞれ撮影した。なお、本研究は「秋田大

A Study on Interpolation Method using GAN in Partially Missing Skeleton Data
Hechen Yun[†], Etsuro Nakamura[†], Yoichi Kageyama[†],
Chikako Ishizawa[†], Nobuhiko Kato[‡], Ken Igarashi[‡], Ken Kawamoto[‡]

[†]Akita University, [‡]ADK Fuji System

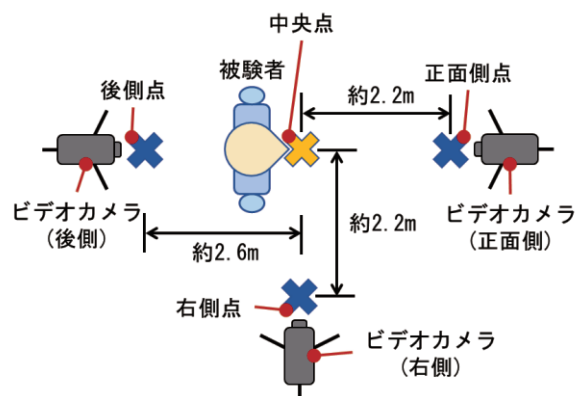


図1 データ取得環境

学手形地区における人を対象とした研究に関する倫理規程第6条第2項」に基づき、被験者の同意を得た上でデータを取得している。

3. 提案手法

3.1. 骨格特徴点の取得処理

動画画像から人間の骨格データを検出可能なオープンソースのOpenPose³⁾を用いて、取得した動画データの各フレームから15点の骨格特徴点および座標値を抽出しこれを実験に使用した。

3.2. GANを用いた補間ネットワークの構築

本研究では、GAIN⁴⁾というGAN(敵対的生成ネットワーク)に基づいたランダムな欠損データの補間を可能にするネットワークを用いて、欠損した骨格特徴点の座標値を推定する。人物の骨格構造による空間的特徴、および動作による部位の時系列変化を抽出しネットワークに学習させるために、補間ネットワークにおける生成モデルと判別モデルにResNetとLSTMを結合してそれぞれに使用した。また、生成モデルの学習を安定させるために、FCN(完全畳み込みネットワーク)を追加判別モデルとして用いて、元骨格データと生成骨格データを区別し、BinaryCrossEntropyを損失関数として算出した損失値を各モデルの損失関数と加算し、補間ネットワークの更新に使用する。補間ネットワークの全体的構造を図2に示す。

3.3. 動作データセットの作成処理

動作の姿勢や物体などの影響を受け、人体の部位が遮蔽されることに起因し、骨格特徴点の位置情報が検出できない場合が想定される。そこで本稿では、下記の3種類の遮蔽状態作成方法を実施し、欠損点

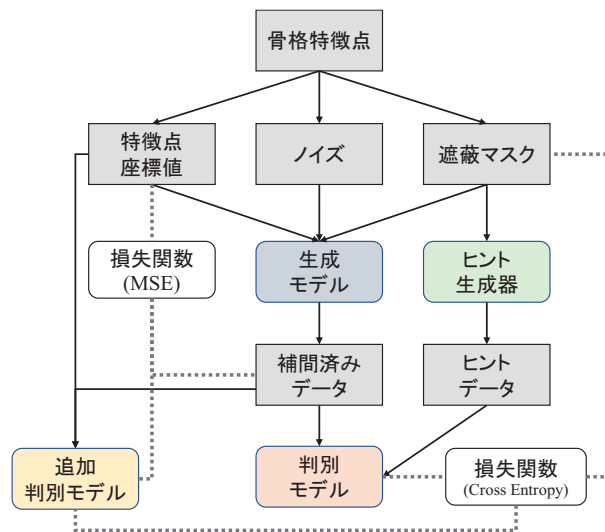


図2 補間ネットワークの構造

が含まれる部分欠損した骨格データを作成した。また、算出した体幹の中心点を含めている骨格特徴点計 16 点から TSSI^[5]という対象人物の身体部位における空間的關係が表れる特徴量を算出した。加えて、60 フレームのフィルタを時系列順に 1 フレームずつシフトし、動画像データにおける各区間の TSSI 特徴量の時系列変化を取得し、これを動作データセットとして使用した。

- (1) **MCAR** : ランダムに 3~12 ヶ所以内の特徴点を欠損させる状態である。さらに、動画像データにおけるフレームごとにもランダムな遮蔽状態を付加する。
- (2) **MAR-Near** : ランダムに 1 ヶ所の特徴点を選択し、その点と連続している 2 ヶ所の点を欠損させる状態である。
- (3) **MAR-Part** : ランダムに四肢部位、もしくは上下半身の位置の特徴点を欠損させる状態である。

3.4. 補間ネットワークの学習処理

構築した補間ネットワークを用いて、動作の種類に応じる 10 パターンの補間ネットワークを作成し、各動作に対して遮蔽状態を付加した骨格特徴点を使用し学習を行った。ネットワークのパターンごとに学習処理を 1~200 回まで 1 回刻みで変更し検討した。

4. 動作判別実験

4.1. 実験内容

提案手法の有用性を評価するために、被験者 10 名を対象として、骨格特徴点の補間実験を実施した。具体的には、被験者 10 名の遮蔽状態が付加している動作データセットを動作ごとに 10 パターンの補間ネットワークに学習させる。学習済みの補間ネットワークにおける生成モデルを用いて、動作データセットからランダムに各動作における「遮蔽状態あり」のサンプルに対して補間を行い、補間結果を可視化して主観評価を行う。

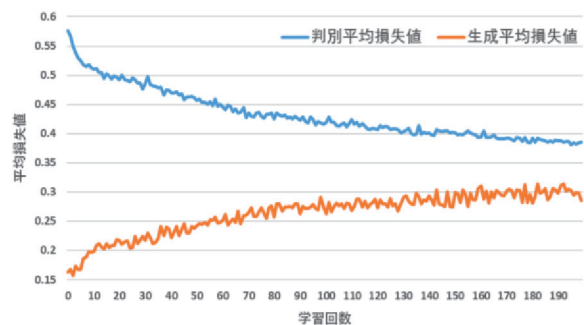


図3 学習処理における平均損失値の変化

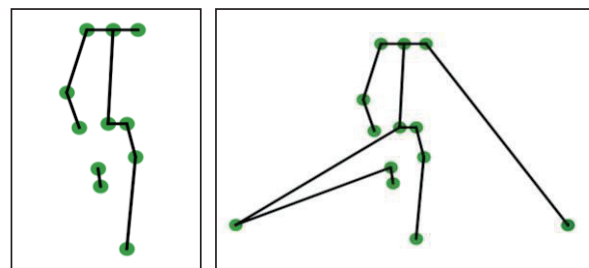


図4 しゃがみ動作中における MCAR の補間結果

図4 しゃがみ動作中における MCAR の補間結果

4.2. 実験結果および考察

提案手法における各パターンの学習処理に伴う損失値変化を図 3 に示す。学習回数に応じて変動し、生成損失値および判別損失値が収束する見込みが示されている。特に、しゃがみ動作に対して、ネットワークの更新が安定に進んでおり、各損失値のばらつきが小さいことがわかる。一方、図 4 に示すように、欠損していないデータから人体の空間的構造に基づいて、欠損している左手部・右足部の特徴点における大まかな座標値が推定可能であるものの、時系列変化の推定が困難であることが明らかになった。特に、フレームごとに遮蔽状態が変化している MCAR においては、TSSI 特徴量として時系列変化と人体の構造情報が同時に参考できないため、推定精度が低下すると考えられる。

今後は、対象人物の空間的な情報と時系列的な位置変化が学習可能なモデルに関して検討を行う予定である。

参考文献

- [1] 国土交通省: “建設業の働き方改革の現状と課題”, <https://www.kensetsu-kikin.or.jp/news/57a42379796b2a6c1d23286d40ea5b611f163364.pdf> (Accessed 2023/01/13)
- [2] P. Elias, J. Sedmidubsky, and P. Zezula: Understanding the limits of 2D skeletons for action recognition, *Multimedia Systems*, 27(3), pp. 547-561 (2021)
- [3] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei and Y. Sheikh: OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 1, pp. 172-186 (2021)
- [4] J. Yoon, J. Jordon, and M. Schaar: Gain: Missing data imputation using generative adversarial nets, *Inter. Conf. on machine learning*. PMLR (2018)
- [5] Z. Yang, Y. Li, et al.: Action Recognition with Spatio-Temporal Visual Attention on Skeleton Image Sequences, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 29, no. 8, pp. 2405-2415 (2019)