

局所的な動作姿勢に着目した作業行動認識手法の検討 Work Action Recognition Focusing on Local Motion Postures

小林 拓椰¹⁾ 新倉 雄大¹⁾ 大橋 洋輝¹⁾
Takuya Kobayashi Takehiro Niikura Hiroki Ohashi

1 はじめに

高齢化社会の到来により新たな労働力の確保は困難になってきており、ものづくりの現場においてはこの傾向が顕著である。そのため限られた労働資源で作業の付加価値や生産量を維持するために、作業員一人一人の生産効率を上げる作業生産性向上が不可欠である。限られた労働資源の中では作業に不慣れな作業員も労働力として早くから現場配属され、作業遅延や作業ミスが発生リスクは高まる。このような作業遅延や作業ミスによる作業効率低減は作業生産性向上に向けた課題となりえる。これを防ぐため、作業の様子を人手で監視する取り組みが実施されているが、余計な労働力を割く必要があり、一人一人の生産効率向上という観点において本末転倒である。そこで、追加の労働力を介さずに作業遅延や作業ミスを低減する手段として、センサを用いた作業行動認識技術の活用が注目されている。作業行動認識によって各工程の作業時間を把握できるため、作業全体のボトルネック発見や、人員配置や作業工程も自動で最適化できる。また、作業行動認識結果を現場の手順書と比較し、作業抜けや順番間違えが発生した際に瞬時に警告し、重大事故や不良品生成を未然に防ぐこともできる。

本研究では作業行動認識を現場適用する際の障壁となりえるセンシングデータ取得時の制約と、作業行動認識器の学習コストの両側面について改善し、より現場適用しやすい作業行動認識方法の実現を目標としている。

これまで作業行動をセンシングデータから認識する研究は数多くなされているが、多くの研究ではセンシングデータとして作業現場に固定されたカメラ映像が用いられている [1, 2, 3]。しかしカメラの映像からデータを取得する場合、カメラの撮影範囲による場所の制約が大きい。場所の制約が少ないデータ取得方法として、加速度計・角速度計を内蔵した IMU(Inertial Measurement



図 1: 衣服型センシングデバイス (各矩形位置に IMU センサが縫い付けられている)

Unit) センサを用いた研究も進められている [4, 5] が、IMU センサを用いた計測は体の複数個所にセンサをバンドで固定する必要があり、作業員にとって負担となる懸念があるため長時間の利用には適さない。そこで本研究では、IMU センサを直接衣服に縫い付けた衣服型センシングデバイス (図 1) を活用し、場所の制約もなく、かつ衣服型のため作業員にかかる負担を低減したセンシングデータ取得方法を提案する。

また、現場毎に異なる作業を認識可能とするための認識器の学習コストも無視できない。現場毎の学習コスト低減のため、少量の学習データで高い認識精度を実現するための少量教師下の行動認識手法に関する研究も行われている [6, 7] が、多くの研究では限られた学習データを有効的に活用する方法が論点となっている。本研究では体の局所的な動作や姿勢は一定の粒度において汎用的に定義できることに着目し、体の局所的な動作や姿勢を収集したデータを事前情報として作業行動認識器を事前学習し、事前情報を活用して少量学習データから効率よく作業行動の識別を学習する方法について提案する。作業現場を模擬した実験環境において 5 名の被験者により評価実験を行ったところ、体の局所的な動作や姿勢を事前知識として活用することで 7.6 % の識別率向上を確認できた。

1) 日立製作所 研究開発グループ

Research Development Group, Hitachi, Ltd.

2 関連研究

2.1 骨格ベースの行動認識

行動認識手法として様々な方法が提案されているが、本研究では作業者の骨格(の関節点座標) 情報を入力とした骨格ベースの行動認識手法に着目する。画像や音声を入力とする方法 [8, 9] と比較して、骨格情報は背景情報やノイズの影響を受けにくく、環境変化に対して頑健であることが特徴である。本報告では少量データでの学習・評価を行うが、少量データ条件では認識モデルが単純な作業動作以外の環境情報に影響されやすいため、そのような影響の少ない骨格ベースの手法が適している。骨格ベースの行動認識手法では、カメラ映像やセンサから取得した骨格の関節座標点データの時系列的遷移を作業認識器の入力とする方法が一般的である。また、人の骨格構造は一種のグラフ構造として捉えられるため、グラフのノード間の関係性も学習するグラフ畳み込みと相性が良い。時空間的なグラフ畳み込み構造(図 2) を特徴抽出器に採用した ST-GCN[10] をきっかけに、グラフ畳み込み構造を取り入れた行動認識手法は数多く提案されてきている [11, 12]。本研究においては、提案する局所的な動作姿勢情報の影響の大きさについて単純に比較するため、最も基本的なグラフ畳み込み構造を有している ST-GCN[10] をベースラインモデルとして用いることとする。

2.2 少量の学習データによる行動認識

新しい行動種類が追加される度に学習データを大量に用意するコストは、行動認識の実導入において大きな課題となる。この問題を解決するため、少ない学習サンプルで行動認識器の学習を可能とする few-shot[14] や zero-shot[15] の学習方法が提案されている。これらの手法では、学習時の工夫として表現学習が用いられており、行動種類毎に特徴空間上で距離が離れた特徴量が出力されるように学習を進めていく。ただしここで獲得される入力データの特徴表現はあくまで少量の学習データを基に決定されるものであるため、必ずしも行動を識別するために有用な表現を獲得できるとは限らない。それに対し、既知の情報を活用し、人間の理解に近い形で新しい行動を認識する手法も研究されている。[16] では人間の姿勢情報を認識できる認識器の情報を活用し、姿勢状態から行動の種類を予測する方法を提案している。本

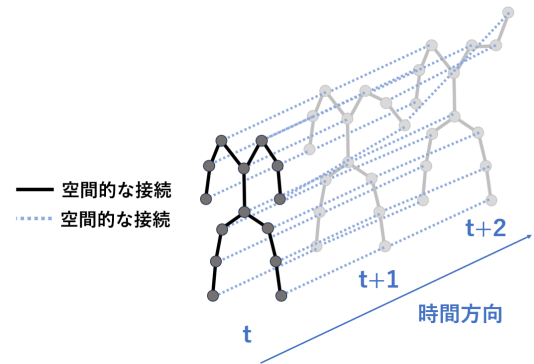


図 2: 時空間的なグラフ畳み込み構造

研究の動機はより時空間的な特徴を捉えるために、これを局所的な動作姿勢の状態に拡張したものである。未知の行動を人間が理解しようとするとき、全体としては未知な行動であっても、各体部位に着目すると、例えば腕であれば上向きに対して何かを捻る動作をしている等、既知の動き・姿勢に当てはめて理解することができる。そのため、各体部位の既知の動作姿勢の組み合わせとして、未知な新しい行動を理解していくアプローチがごく自然である。そこで本研究においては、各体部位毎に既知の動作姿勢を認識し、後段の行動認識に有用な特徴表現を獲得することを考える。

2.3 体の局所分解による行動認識

局所的に分解した各体部位の動作姿勢の組み合わせから行動認識を行うことは人間の理解に近いアプローチであり、従って関連の手法も提案されている [17, 18]。これらの行動認識手法ではアテンション機構を用い、各行動種類がどの体部位と強く結びついているかも同時に学習することで、特定の体部位に着目して最終的な行動を認識する。ただし、各行動クラスと関連性の高い体部位との結びつきや、そもそもの各体部位の最適な特徴表現を学習するには、学習データを大量に用意することが前提となっている。それに対し、本研究では各体部位ごとに汎用的な既知の動作姿勢状態を定義するため、既知クラスに基づいた最適な特徴表現を獲得し、後段の行動認識へ活用することができる。

3 提案手法

3.1 衣服型センシングデバイスによる作業データ取得

本研究では、図 1 に示した衣服型センシングデバイスによる作業データ取得を実施する。電子回路布と呼ばれ

る、伸縮性回路を布上に形成する方法により作業者が実際に使用する作業着に IMU センサが縫い付けてあるため、データ取得中も作業者は普段と変わらない状態で作業を行うことができる。作業者が衣服型センシングデバイスを着用した際、初期姿勢のキャリブレーションと、取得した加速度、回転、位置の時間変化情報を計算することで、世界座標系における作業者の 3 次元骨格座標データを各時刻で取得することができる。取得できる 3 次元の骨格座標データの関節点とその接続関係は図 3 の左側部に示した通りである。本研究ではこの骨格座標データを骨盤を原点とした 3 次元座標データに直し、体格差を補正するために胸椎-骨盤間の長さを参照して全体を正規化したのち、作業行動認識モデルの入力として用いる。骨格データの入力の際は、図 3 に示した関節点の接続関係の情報を含んでいるため、グラフ畳み込みでは各関節点(ノード)間の繋がりに基づいて学習が実行される。

3.2 局所的な動作姿勢情報の活用

ST-GCN[10] では各関節点(ノード)間の関連性を時空間的なグラフ畳み込みによって学習し、関節間の関係性を反映した上で、全ての関節点の特徴量を用いて作業行動予測結果を出力する。提案手法ではそれに加え、いくつかの関節の集まり毎に体を局所的に分割し、未知の作業行動クラスを既知の体部位の動作姿勢の組み合わせとして認識することを可能とするアプローチをとる。本研究における体部位の局所的な分割方法について図 3 に示した。体の局所的な分割方法としてはいくつか考えられるが、体の軸である体幹部に加え、それに付属した形で接続している付属肢骨格のうち、人の動作の大部分を決定する左腕部、右腕部、左足部、右足部を対象とする。作業現場の作業行動においても、ほとんどの作業が体幹部による姿勢保持と、左腕部、右腕部、左足部、右足部の動作姿勢によって成り立っている。左腕部を左前腕、左上腕と細分化していく方法も考えられるが、関節点をまたがない局所領域で表現できる動作の種類は限られるため、局所の動作姿勢情報を認識することによるメリットの恩恵を受けられない。さらに、実際の作業現場において左右の腕で別々の動作を行うことはあっても、左右の足でそれぞれ動作を行うケースは少ない。そのため、最終的には局所的な部位として体幹部、左腕部、右腕

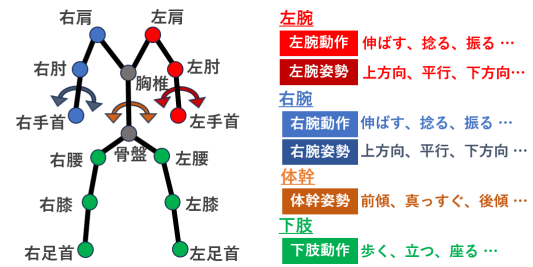


図 3: 取得した骨格情報と局所的な動作姿勢例

部、左足部と右足部をまとめた下肢部の 4 種類を本研究では用いる。それぞれの部位の該当関節点に関しては、図 3 の通りである。左腕部・右腕部に関してはそれぞれ動作情報と姿勢情報、体幹部に関しては姿勢情報、下肢に関しては動作情報を定義する。体幹部は主に姿勢保持のために使われ、骨の構造から動作の種類は限られること、また下肢の姿勢情報に関しては動作情報が既に姿勢情報も内包していることを理由に、今回は検討から除外している。図 3 の動作や姿勢の例からも分かるように、それぞれの部位の動作情報・姿勢情報はどのような作業現場においても共通で発生し得る動作姿勢によって構成されている。これらの動作姿勢情報については、それぞれの部位の動作・姿勢を教師ラベルとしたデータセットを事前に用意し、局所的な動作姿勢を認識する特徴抽出器・識別器を事前に学習させておく必要がある。

3.3 学習・推論方法

本研究の提案手法とベースライン手法の比較を図 4 に示す。ベースライン手法は ST-GCN[10] をそのまま用いており、作業行動の予測結果と教師ラベル間の交差エントロピーロスを損失として学習を実行する。提案手法では事前に局所的な動作姿勢認識モデルを学習する必要があるため、2 段階の学習フェーズによって構成される。なお、1 フェーズ目の局所的な動作姿勢認識の学習については初回のみ必要であり、新しい作業動作クラス群に置き換えて学習しなおす場合は 2 フェーズ目から開始してよい。1 フェーズ目では、それぞれの局所的な動作姿勢認識モデルに関して、ベースの特徴抽出器は重みを共有し、その後各部位の該当する関節点に恣意的に強い重みかけられたアテンション機構を導入する。アテンション機構によって重みかけられた特徴量はそれぞれの特徴抽出器・識別器を経由し、それぞれ教師ラベルとの間で交差エントロピーロスを損失として学習が進行

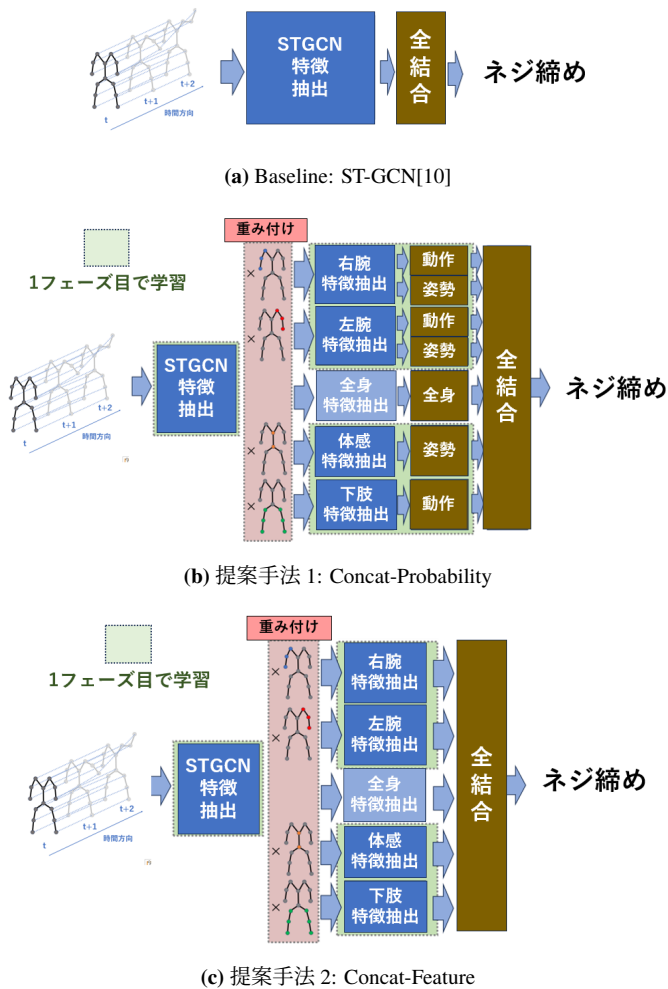


図 4: 手法比較

する。2 フェーズ目では、局所的な動作姿勢の認識結果について、softmax を計算する前の確率ベクトルを利用する方法と、特徴抽出器の出力を利用する方法の 2 パターンについて検討する。狙いとしては、局所的な動作姿勢情報がダイレクトに作業行動認識のための特徴として使われる場合 (提案手法 1: Concat-Probability) と、局所的な動作姿勢情報を認識するために必要な局所的な特徴量を作業行動認識のために用いる場合 (提案手法 2: Concat-Feature) について、影響の違いを比較するためである。今回の評価において比較する 3 手法について、以下にまとめる。

(a) ST-GCN(Baseline): ST-GCN[10] をそのまま用いており、学習は 1 段階。softmax の出力結果と正解作業行動間の交差エントロピー損失を用いて学習を実施する。

(b) Concat-Probability(提案手法 1): 局所的な動作姿勢の特徴抽出部と全身の特徴抽出部の出力結果を組み合わせ、最終的な作業行動を識別する。学習フェーズは 2 段階に分解され、1 フェーズ目では共通の ST-GCN[10] 特徴抽出器と、それぞれの局所的な特徴抽出器・動作姿勢識別器を学習させる。損失として、それぞれの局所的な動作姿勢の正解データと識別結果のクロスエントロピーロスを計算し、同時に全ての識別器の学習を進める。2 フェーズ目では、学習済みの共通の ST-GCN[10] 特徴抽出器とそれぞれの局所的な特徴抽出器・動作姿勢識別器の重みは固定し、全身の特徴抽出器と、最終的な作業行動識別器のみを学習させる。各体部位の動作姿勢の識別結果は softmax 適用前の確率ベクトルをそのままコンカットし、作業行動識別器の入力として活用する。

(c) Concat-Feature(提案手法 2): (b) 提案手法 1 と基本的な構造、学習方法は同一である。違いとして、提案手法 1 では各体部位の動作姿勢の識別結果は softmax 適用前の値を入力としていたが、こちらでは各局所特徴抽出器の出力時点でコンカットし、作業行動識別器の入力として活用する。

4 提案手法の評価

4.1 実験データ

20 代男性 4 人、20 代女性 1 人を含む身長体格も異なる被験者 5 人分のデータを、作業現場に見立てた実験室環境で収集した。各被験者は図 1 の衣服型センシングデバイス上下を着用し、表 1 にある 8 種類の作業行動 (待機、移動、保持、運搬、指差し、ネジ締め、叩きつけ、挿入) を各 10 回ずつ、計 20 分間程度実施し、衣服型センシングデバイスから作業時の骨格データを取得した。ただし、各作業行動における 10 回の作業は反復的な作業ではなく、作業行動種類は同一でも作業対象の工具や体勢、使用する腕を変えることで実際の現場で起こりうる様々な作業環境を想定して実施しており、少量データから認識するにあたって難易度の高い作業内容となっている。各作業行動の実施例を図 5 に示した。各データには、作業行動種類に関する教師ラベルだけでなく、局所

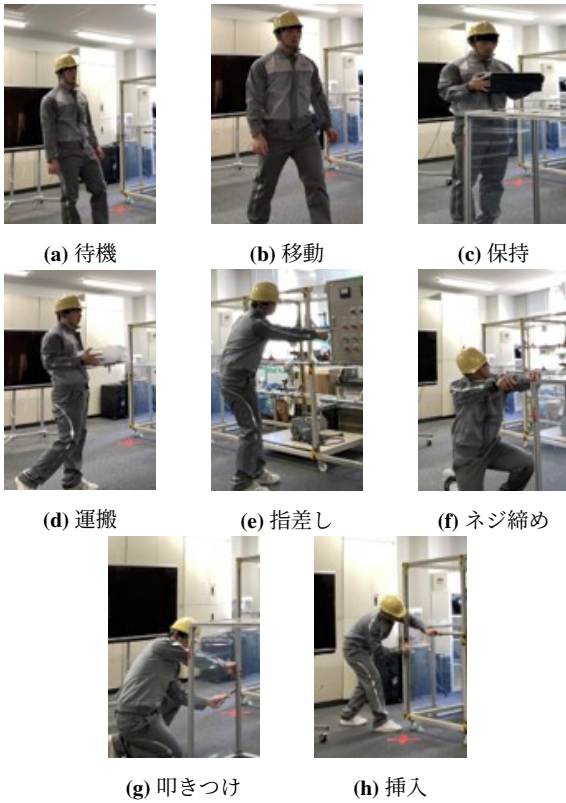


図 5: 作業動作の種類

表 1: 作業行動一覧

作業行動	割合 (%)	説明
待機	3.7	作業と作業の間の待機時間
移動	15.5	歩いて作業場所の移動
保持	10.9	対象物体を持つ (片手もしくは両手)
運搬	7.0	対象物体を運ぶ (片手もしくは両手)
指差し	11.2	指差し点検
ネジ締め	28.2	手でネジ締め
叩きつけ	16.0	ハンマーによるたたきつけ (模擬)
挿入	7.5	棒状物体の筒への挿入

的な動作姿勢モデルを学習させるために、左腕部の動作・姿勢、右腕部の動作・姿勢、下肢部の動作、体幹部の姿勢についても教師ラベルが付与されている。表 2 に各部位の認識対象の動作種類と全体に占めるデータ割合 (%) を示した。左腕部、右腕部はそれぞれ 6 種類 (待機、保持、伸ばす、捻る、振る、挿入する) の動作、下肢部は 5 種類 (立つ、歩く、座る、膝立ち、しゃがむ) の動作を定義している。表 3 には各部位の認識対象の姿勢種類と全体に占めるデータ割合 (%) を示した。左腕部、右腕部はそれぞれ 3 種類 (上方向、水平、下方向) の姿勢、体幹部は 3 種類 (前傾、真っすぐ、後傾) の動作を定義している。

表 2: 局所的な動作クラス

左腕		右腕		下肢	
動作種類	割合 (%)	動作種類	割合 (%)	動作種類	割合 (%)
待機	18.7	待機	13.7	立つ	42.8
保持	14.8	保持	12.2	歩く	14.8
伸ばす	34.2	伸ばす	28.2	座る	22.7
捻る	23.0	捻る	36.0	膝立ち	13.3
振る	6.3	振る	6.6	しゃがむ	6.4
挿入する	2.9	挿入する	3.0		

表 3: 局所的な姿勢クラス

左腕		右腕		体幹	
姿勢種類	割合 (%)	姿勢種類	割合 (%)	姿勢種類	割合 (%)
上方向	24.4	上方向	24.4	前傾	30.1
水平	49.8	水平	49.8	真っすぐ	54.5
下方向	26.5	下方向	26.5	後傾	16.3

4.2 実験設定

全ての実験・手法において、以下同様の実験設定のもと評価を行った。5 人分のデータセットを 4 人分の学習データ、1 人分の評価データに分割し、学習データを用いて学習させたモデルを評価データで評価し、評価結果を算出する。最終的な評価結果に関しては、5 通り全てのデータ分割方法による交差検証を行い、その平均値を算出する。

4 人分の学習データを更に分割し、そのうちの 1 割程度を検証データとして取り出す。学習時のエポック数を最大で 20 とし、各エポック後に検証データを用いて損失関数の計算を行う。4 エポック以上連続で検証データの損失関数が最小値を更新しなければ、それ以上は学習データに対する過学習であると判断して学習を終了し、4 エポック前のモデルを評価モデルとして採用する。途中で学習が終了しなかった場合は、20 エポック学習後のモデルが評価モデルとして使用される。

また表 1、表 2、表 3 からわかるように、各認識対象クラスはデータセットに含まれる割合に大きくばらつきがある。学習時にはデータセットに含まれる対象クラスの逆比で損失関数を重みづけし、クラス間のばらつきの影響をできるだけ抑える。

取得したデータの骨格構造は図 3 の通りであり、図 3 に記載の全ての関節点を入力として扱う。使用する骨格データのフレームレートは 100 であり、そのサンプルレートにおいて 60 フレームずつを 1 つの認識ウィンドウサイズとする。学習・評価データとして用いたシーケ

表 4: 動作クラスの認識精度

左腕		右腕		下肢	
動作種類	MoF	動作種類	MoF	動作種類	MoF
待機	99.8	待機	48.2	立つ	100.0
保持	78.6	保持	84.7	歩く	99.8
伸ばす	71.7	伸ばす	55.6	座る	55.3
捻る	25.2	捻る	49.3	膝立ち	37.2
振る	0.2	振る	0.4	しゃがむ	0.0
挿入する	0.0	挿入する	0.0		

ンスのウィンドウ数は 5 人分合わせて計 3122 個であり、各ウィンドウが 0.6 秒であることを考慮すると使用した総データ量は時間換算で約 31 分程度であった。

4.3 実験と結果

4.3.1 局所的な動作認識精度の評価

提案手法において、作業行動の認識時に局所的な動作姿勢情報を活用するため、事前に動作姿勢を学習させる必要がある。姿勢に関してはそれぞれ 3 種類の姿勢クラスかつ、識別内容も関節点の位置関係のみによって決定されるシンプルなものであり、100 %に近い識別精度となるため、ここでは局所的な動作認識に関する評価結果のみ評価を実施した。識別精度の指標として MoF (Mean of Frames) を用いた場合の局所的な動作認識精度の評価結果を表 4 に示す。MoF はフレーム単位での正解率であり、全ての認識フレームのうちの正解フレームの割合をパーセント表示した値と同値である。まず下肢部の動作に関しては、立つと歩くの 2 クラスに関しては 99 %以上の識別率で正しく認識できているが、座り、膝立ちの 2 クラスに関してはお互いに混同して認識してしまっているケースが目立ち、しゃがみに関しては 100 %誤認識していることを確認した。左腕部と右腕部の動作に関しては、それぞれ保持が 80 % 近くの高い水準で認識できていたのに対し、振る、挿入するの識別率がどちらも 1 % 以下と全く識別できていないことを確認した。

4.3.2 作業行動認識精度の評価

今回ベースラインの作業行動認識モデルとして用いた ST-GCN[10] に対して、提案した局所的な動作姿勢の導入による作業行動認識精度の変化に関する評価検証を行った。まず、識別精度の指標として MoF を用いた場合のベースライン手法と 2 種類の提案手法の比較結果を表 5 に示した。ベースラインである ST-GCN による作業行動認識方法に、局所的な動作姿勢情報を認識し利用す

表 5: 作業行動の認識結果

手法	動作姿勢情報	MoF
ST-GCN[10]	-	64.2
提案手法 1	Concat-Probability	71.6
提案手法 2	Concat-Feature	70.9

るブロックを導入することにより、最大 7.4 の MoF 向上が見られ、少量学習データ条件において局所的な動作姿勢情報を用いることの有用性を確認した。また提案手法内での比較では、局所的な動作姿勢情報をより直接反映した softmax 直前の段階でコンカットした方が、局所的な動作姿勢を認識するための特徴量の段階でコンカットするより 0.7 の MoF 向上が見られた。

4.4 考察

4.4.1 局所的な動作姿勢の認識

局所的な動作姿勢の認識結果について考察する。下肢部の動作について、座りと膝立ち、しゃがみという、骨格だけで見たときに姿勢の似た 3 つの動作クラスが混同して認識されていることを確認した。座っている椅子の高さが低い場合、骨格だけで見たときに座りはしゃがみに類似の体勢であり、また膝立ちとしゃがみの違いは片膝でも地面についているか否かだけであるため、これら 3 つの動作を区別することは難易度の高いタスクであると考えられる。これらの区別はできなかったものの、立ち、歩き、もしくは何かしらの体勢で地面に近い状態で作業している、の 3 種類の識別はできるということが確認できた。今回の検討では、何かしらの体勢で地面に近い状態を更に座り、膝立ち、しゃがみの 3 種類へと分割を試みたが、正しく認識できなかったことを踏まえ、3 種類への細分化の必要性についても含め、引き続き検討していく必要がある。左腕部の動作と右腕部の動作について見ると、振ると挿入の動作が全く識別できていない。この要因として 2 つ考えられる。1 つ目は表 2 にあるように、振ると挿入の全体に占めるデータ割合が極端に低く、十分な学習データが用意できていない点が挙げられる。学習時はこのデータ割合のばらつきを考慮して損失の重みづけは行っているものの、そのような対策だけでは不十分であったと考えられる。10 % 強のデータ割合である保持クラスが 80 % 以上で認識できていることを考慮すると、今回のデータ全体の 10 % に該当す

る 300 シーケンス (ウィンドウ) 数程度は最低限収集すべきデータ量であると考えられる。2つ目として、クラス間の動作の類似度が挙げられる。振るは右腕部に限ると、捻ると大きく混同してしまっていた。振るはハンマーを叩く際等の細かく工具を振る動きを示すが、振り上げの動きが小さいと、ネジ締めの際の捻る動きと同様に細かく揺れている動作として捉えることもできる。挿入に関しては、(腕を) 伸ばすと混同されていた。今回はウィンドウサイズが 0.6 秒と短く、動作の一部だけを切り取ると異なる動作クラスが同様の特徴として切り取られてしまったと考えられる。よりウィンドウサイズを拡張し、時系列的に長期的な動作特徴を踏まえた認識方法についても検討していく必要がある。

4.4.2 局所的な動作姿勢情報の効果

表 5 により、局所的な動作姿勢情報によって作業行動認識率が向上することを確認した。また、局所的な動作姿勢情報をより直接反映した softmax 直前の段階でコンカットした方が、局所的な動作姿勢を認識するための特徴量の段階でコンカットするより認識率の向上が見られた。図 6 に、最も認識精度が高かった提案手法 1 における具体的な識別結果を示す。縦軸の各作業行動種類に対して、横軸で各作業行動であると認識した割合を表している。図 6 から、移動、保持、運搬の作業行動が 90 % 以上で認識できていることがわかる。また、待機に関しては、必ずしも静止している作業行動クラスではなく、ストレッチしたり、リラックスしている動作状態も含まれているため難しいクラスであり、ベースラインでは認識率が 10 % 以下であったが、提案手法では認識率 50 % 以上であることを確認した。体全体の動きだけでとらえてしまうとこれらの動き情報から他の作業行動クラスに分類されてしまうが、局所的な既知の動作の組み合わせとして捉えたとき、他の作業動作として分類できないものが待機として認識できるようになったと考えられる。一方、ネジ締めに関しては、指差しと誤認識するケースが目立つ。これは動作認識モデルにおいて、左腕部や右腕部の動作認識モデルが捻るを伸ばすと誤認識していることに起因していると考えられる。左腕部や右腕部の動作認識モデルの認識精度を向上させれば、ネジ締めの作業行動クラスの認識率も向上すると考えられる。

同様に、表 4 と図 6 間の関係についてみていくと、そ

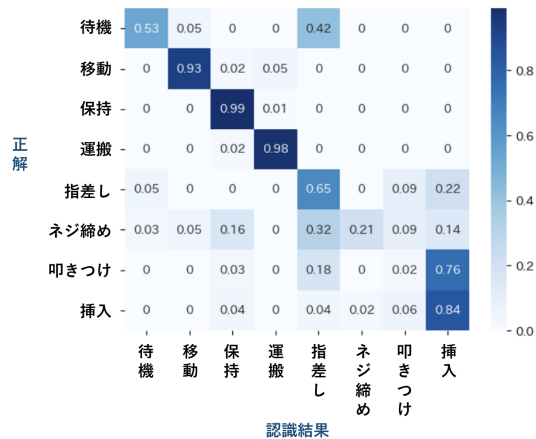


図 6: 作業行動認識結果: 提案手法 1

もその関連する局所的な動作認識精度が低い作業行動クラスは、その認識精度も悪いことが分かる。例えば、各腕で認識できていない振る動作を含む叩きつけ作業は、ベースライン・提案手法どちらにおいても極端に低い認識率である。同様に挿入動作に関して各腕の動作としては認識できておらず、作業行動の挿入作業に関しても一見正解率が高いが、適合率でみると低くなっている。このことから、今回定義した局所的な領域において動作を認識できることは作業行動全体の認識率に影響し、ひいては局所的な動作の認識率を上げていくことが作業行動の認識率向上につながると考えられる。

5 おわりに

本研究では作業最適化や作業ミス防止による作業生産性向上のために、作業者にかかる負担を低減した衣服型のセンシングデバイスを用いた作業行動認識手法に関して検討を行った。現場毎に必要な学習データを最小限に抑えた作業行動認識方法の実現のため、局所的な動作姿勢情報を事前情報として活用する方法について提案した。認識難易度の高い少量データ条件を想定し、5名の被験者により評価実験を行ったところ、提案した局所的動作姿勢特徴量の導入により、従来の骨格ベースの作業行動認識手法と比較して 7.4 % の作業行動認識率の向上を確認できた。

参考文献

- [1] Qianqian Xiong, Jianjing Zhang, Peng Wang, Dongdong Liu, and Robert X. Gao. Transferable two-stream convolutional neural network for human action recognition. *Journal of Manufacturing Systems*, Vol.56, pp.605–614, 2020.
- [2] Hanbin Luo, Chaohua Xiong, Weili Fang, Peter E.D. Love,

- Bowen Zhang, and Xi Ouyang. Convolutional neural networks: Computer vision-based workforce activity assessment in construction. *Automation in Construction* Vol.94, pp.282–289, 2018.
- [3] Chengjun Chen, Tiannuo Wang, Dongnian Li, and Jun Hong. Repetitive assembly action recognition based on object detection and pose estimation. *Journal of Manufacturing Systems*, Vol.55, p.325-333, 2020.
- [4] Mattia Pesenti, Giovanni Invernizzi, Julie Mazzella, Marco Bocciolone, Alessandra Pedrocchi, and Marta Gandolla. IMU-based human activity recognition and payload classification for low-back exoskeletons. *Scientific Reports* Vol.13 No.1184, 2023.
- [5] Ismael Espinoza Jaramillo, Jin Gyun Jeong, Patricio Rivera Lopez, Choong-Ho Lee, Do-Yeon Kang Do-Yeon Kang, Tae-Jun Ha, Ji-Heon Oh, Hwanseok Jung, Jin Hyuk Lee, Won Hee Lee, and Tae-Seong Kim. Real-Time Human Activity Recognition with IMU and Encoder Sensors in Wearable Exoskeleton Robot via Deep Learning Networks. *Sensors* Vol.22, 2022.
- [6] Mengyuan Liu¹, Fanyang Meng, Chen Chen, Songtao Wu. Novel Motion Patterns Matter for Practical Skeleton-Based Action Recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.
- [7] Joe Yue-Hei Ng, Jonghyun Choi, Jan Neumann, Larry S. Davis. ActionFlowNet: Learning Motion Representation for Action Recognition, In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision(WACV)*, 2018.
- [8] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [9] Fanyi Xiao, Yong Jae Lee, Kristen Grauman, Jitendra Malik, and Christoph Feichtenhofer. Audiovisual slowfast networks for video recognition. *arXiv*, 2020
- [10] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.
- [11] C. Song, Y. Lin, S. Guo, and H. Wan. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.
- [12] Maosen Li, Siheng Chen, Xu Chen, Ya Zhang, Yanfeng Wang, and Qi Tian. Actional-structural graph convolutional networks for skeleton-based action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [13] Ziyu Liu, Hongwen Zhang, Zhenghao Chen, Zhiyong Wang, and Wanli Ouyang. Disentangling and Unifying Graph Convolutions for Skeleton-Based Action Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [14] Yaqing Wang, Quanming Yao, James Kwok, and Lionel M. Ni. Generalizing from a few examples: A survey on few-shot learning. *CSUR*, pp.1-34, 2020.
- [15] Dipika Singhania, Rahul Rahaman, and Angela Yao. Iterative contrast-classify for semi-supervised temporal action segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.
- [16] Hiroki Ohashi, Mohammad Al-Naser, Sheraz Ahmed, Katsuyuki Nakamura, Takuto Sato, and Andreas Dengel. Attributes' importance for zero-shot pose-classification based on wearable sensors. *Sensors*, Vol. 18, 2018.
- [17] A. Zhu, Q. Ke, M. Gong, and J. Bailey. Adaptive local-componentaware graph convolutional network for one-shot skeleton-based action recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023.
- [18] Chenyang Si, Ya Jing, Wei Wang, Liang Wang, and Tieniu Tan. Skeleton-based action recognition with spatial reasoning and temporal stack learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.