

到達点系列情報と骨格情報を用いた動作予測

宮原 雅人^{1,a)} 佐藤 大祐¹ 福田 匡人¹ 松村 成宗¹ 瀬下 仁志¹

概要: 人とロボットが協調して効率的かつ安全に作業を進めるため、ロボット側で作業中の人の動作や意図を事前に予測してサポートを行うことが必要である。しかし、人の作業中の動きは多くの移動方向の変化が発生しており、将来の動きを予測することは困難である。本研究では過去の到達点の系列情報を用いて未来の到達点を予測し、到達点情報と骨格座標系列を用いた動作予測手法を提案する。到達点が切り替わることを考慮した予測モデルを用いて、将来の動作到達点を予測し、動作予測モデルの入力に加えることで、動作が切り替わる区間において現在の到達点情報のみを用いたモデルよりも改善させられることを示す。

キーワード: 動作予測, 到達点系列, 人とロボットの協調作業, 遠隔操作

Motion prediction using arrival point sequence information and skeletal information

1. はじめに

ロボットは、あらかじめ決められた作業を休みなく正確に繰り返し行うことができるという点で人の作業員より優れており、工場などでは産業用のロボットが広く導入されている。このような産業用ロボットは、外乱の入らない環境で作業を行うことを想定しており、安全性の観点から人が立ち入ることのできない空間において稼働させることが一般的である。多品種少量生産を行う場合など、作業内容が一定ではなく、時に柔軟な判断が求められる場面では、いまだロボットが人を代替するにはいたっていないものの、センシングやAI技術の進展に伴いロボットの能力が向上するにつれ、人と同じ空間で作業を支援することが可能となってきている。また、小型の掃除用ロボットなどはすでに各家庭内で稼働しており、今後より多くの場面で人と同じ空間で稼働するロボットが増えてくると予想される。

ロボットが人と同じ空間で稼働する上で、人に接触をしないことが重要である。安全性を重視し、人とロボットの距離に近づくとロボットを緊急停止させることで接触は防げるが、これでは頻りにロボットが動作を停止することとなり、大きく効率性を損ねる。そこで、人の動きや意図

を予測することで、接触の確率が高い場面を精度よく検知するというアプローチが検討されている。人の動きの予測は、センシングされた画像や点群情報から抽出された人の骨格情報を時系列データとして入力とし、未来時刻の骨格情報を予測することで行われる。Martinez らが Recurrent Neural Network(RNN) に基づく骨格予測手法 [1] を提案して以降、多くの深層学習を用いて骨格の時空間的な相関関係をモデリングする予測手法が提案されている。また近年では人が動作の終着点として到達するであろう到達点を明示的に推定し、従来の骨格予測手法に組み込む手法が提案されており、より高い精度で人の動きを予測することが可能となっている。一方で、人とロボットが協調して作業を行う際には、両者はより近接した状態が長く続くことになる。この場合においても、ロボットは人との接触を避けるだけではなく、人の動きを正確に予測し、人の動きに合わせて動作させることで様々な支援を行うことが可能となる。しかし、歩行する人を予測する場合とは異なり、作業中の人の動きはより複雑となる。人の動作を予測するこれまでの多くの取り組みでは、同じ方向に歩き続けるなどの単一の動作を対象としており、作業のように複数の動作を切り替えながら連続して行われる動きを対象としていない。そのため、従来の単一の動作を想定した予測モデルでは作業動作を精度よく予測することは難しい。特に動作が切り替わるタイミングでの予測は難しく、最大予測誤差が大き

¹ 日本電信電話株式会社 人間情報研究所
NTT Human Informatics Laboratories
1-1 Hikarinoka, Yokosuka-shi, Kanagawa-ken, 239-0847 Japan
^{a)} masato.miyahara.gc@hco.ntt.co.jp

なることで、安全な協調作業の妨げとなるため、システムが許容する予測誤差より抑える必要がある [2].

本研究では、人とロボットの近接した協調作業の実現を目指し、時系列骨格情報を用いた動作予測に取り組む。これまでの予測手法では、単一の動作を対象としており、複数の動作を連続して切り替える作業者の動きを精度よく予測することは困難であった。そこで、過去の作業者の到達点系列情報をもとに次の動作の到達点を予測し、予測された到達点を骨格予測に組み込むことで、動作が切り替わる区間においても誤差を抑えることが可能な動作予測手法を提案する。本稿では、人が室内の複数の場所に物を運ぶ作業を行う際の骨格データを含む Mogaze データセット [3] を用いて実験を行い、提案手法の有効性の検証を行った。

2. 関連研究

2.1 ロボットと人の協調作業

ロボティクス技術の発達に伴って、人と協調して作業を行うロボットが幅広い場面に普及しつつある。人とロボットの協調作業 (HRC, Human Robot Collaboration) を効率的に進める上で、人の動作や意図の予測をおこなってサポートをする手法が有効である。Zhang[4]らはロボットが、教示者である人のデモンストレーションから事前に作業計画を学習し、人の作業意図予測を活用した補助を行うことで、予測を行わない手法よりも短い待機時間で組み立て作業を実行可能にした。

2.2 動作予測技術

動作予測技術とは予測対象者の未来の体の動きを予測する技術であり、基礎研究だけでなく、ロボットと人の協調作業や VR(Virtual Reality) や AR(Augmented Reality)[5], 遠隔操作 [6] などの遅延のある環境下で操作者の動作の先読みと先行実行による低遅延化など、幅広い応用が行われている。基本的な問題設定では、入力として直前の予測対象の人の骨格位置座標を用いて、数 100msec から数 sec 先の骨格位置座標を予測を行うため、骨格予測とも呼ばれる。骨格予測技術の研究が近年活発に進められている理由として、センサーデバイスの発達と普及によって、大量の骨格データ収集が可能になった点が挙げられる。Microsoft Kinect[7] などの深度センサーを用いた非接触でのデータ収集や、2D の静止映像から 3 次元の骨格情報を推定する OpenPose[8] などの骨格推定技術を用いることで、容易に骨格情報の取得ができる。データ収集が容易となり、近年では Human3.6M[9] などの大規模データセットが公開されたことで、事前学習で大量の学習データセットを必要とする深層学習技術を適応した骨格予測の研究が活発に進められている。

Martinez らが Recurrent Neural Network(RNN) に基づく骨格予測手法 [1] を提案して以降、多くの深層学習による骨

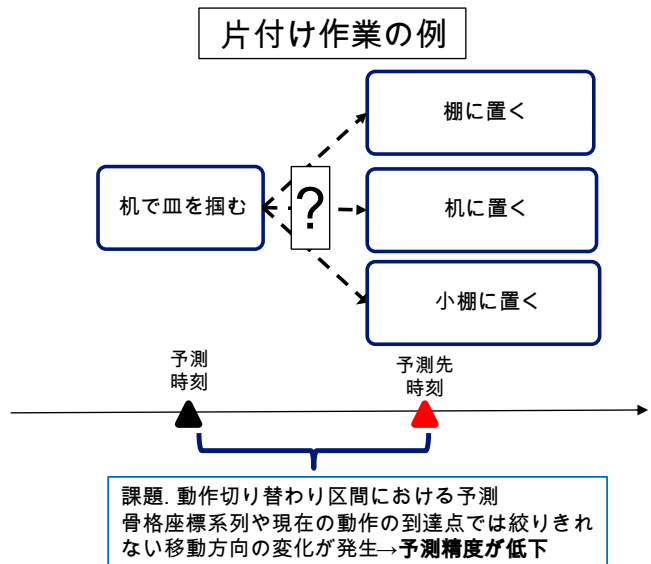


図1 作業中の動作予測の問題点

格予測手法が提案されている。Wu らは、Long Short Term Memory(LSTM) によって時間方向の動きの特徴量をとらえ、リアルタイムに 0.5 秒先の骨格予測が可能な Future Pose を提案した [10]。Chiu らは、人の全身運動において身体部位や関節毎に時間スケールが異なっている様子を捉えるために、階層的なモデルである Triangular-prism Recurrent Neural Network models(TP-RNN) を提案した [11]。Wang らは、従来の骨格予測技術が教師あり学習であるため、汎化性能が低い問題に着目した。未知の観測データに対する予測性能を高めるために、骨格予測を逐次的な意思決定問題として定式化した上で、逆強化学習の一種である Generative adversarial imitation learning(GAIL)[12] を適応した [13]。

近年、直前の骨格位置座標だけでなく、予測対象者がどこに向かって動いているかを表す到達点や何を行っているかを示す動作ラベルなど、人の意図に対応する情報を活用することで、より高精度に予測する手法が提案されている。Wang らは、既存の RNN に基づく動作予測モデルが、人の動作の慣性運動としての特徴量しか捉えられておらず、特に長期の動作予測において不十分であるという問題点を指摘し、動作ラベルを活用した動作予測モデルを提案した [14]。予測対象者が現在行っている歩く、座るなどの動作ラベルを推定し、予測に活用することで、160msec 以上のスケールでの予測精度を向上させた。Corona らは、人の動きが周囲の物体や環境との相互作用から生成されることに着目し、周囲のオブジェクトと予測対象者の関係をグラフ構造として活用する予測手法を提案した [15]。さらに Cao らは人の動きが、周囲の物体の空間的な配置などの環境情報の影響を受けるだけでなく、到達点を志向した goal-directed であることに着目した [16]。予測対象者である歩行者が映る第三者視点の映像から、到達点を推定し、移動経路、全身骨格の軌跡を順に予測する手法を提案した。

ここまで紹介した既存の動作予測技術は歩行や座る動作などの単一の運動に対する動作予測であったが、多様な動作から構成される作業中の人の動作予測にも同じ枠組みを適用することができる。例えば、基本的な作業タスクの一つとして、片付けや陳列業務など対象物を掴んで所定の場所に置く pick & place がある。人が物体を掴む動きをする際には「掴む位置」が、物体を置く際には「置く位置」が到達点となるため、pick & place を行う作業員に対して到達点情報を活用した予測は有効であると考えられる。[16]と同様に、これらの動きの到達点を事前に推定して、骨格情報の時系列データと併せて動作予測に活用することで、対象の動作区間内における予測精度を上げることが期待される。

既存の動作予測技術では、直前の動きから予測先時刻にかけて学習データと予測対象者が同一の到達点に向かって単一の動作をとることを仮定して予測を行う。具体的には、歩く、走るなどの特定の動作についての学習データから、その動作特有の速度変化等の特徴量やパターンを事前に学習し、既に観測された直前の映像フレームを入力として未来の映像フレームの動作の予測を行う。限られた動作パターンの軌跡を予測する場合は、学習データ量を増やすことで既存手法でも対応できる可能性があるが、多様な複数の動作が混在する作業においては、予測すべきパターン数が膨大になる。また、動作予測技術を手とロボットの協調作業などの他のシステムの一部として応用する際には、安定的に稼働させるため、システムが許容する予測誤差よりも最大予測誤差を抑えられるようなロバストな動作予測手法が必要である [2]。

3. 提案手法

多様な動作から構成される作業中においては、予測時刻と予測先時刻において、到達点の異なる動作が実行される可能性がある。基本的な作業タスクの一つである pick & place の例を図 1 に示す。予測時刻において、「皿を掴む」動作を行っていたとしても、予測先時刻において棚に置くか、机に置くかなど複数の到達点候補が存在する。起こりうる動作の候補が複数あるため、骨格位置座標などの直前の観測情報だけでは次の到達点を予測することが困難である。したがって、直前の観測情報以外の情報も活用しながら、予測時刻先の作業員の到達点を正確に予測する技術が必要となる。

そこで、本研究では、図 2 に示すように、骨格系列座標と過去の到達点の履歴情報から未来の到達点を予測する到達点予測器と、動作予測器の二つの予測器から構成される動作予測器を提案する。到達点予測器は予測時刻先の到達点を予測するモデルであり、骨格系列座標と過去の到達点の系列情報から到達点を予測する。予測した到達点と骨格座標系列を入力として、動作予測を行う。現在の動作の到

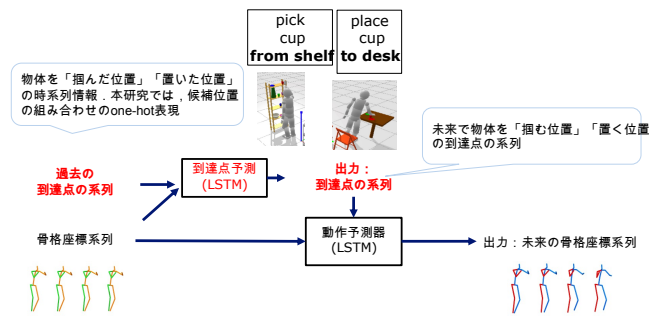


図 2 提案手法

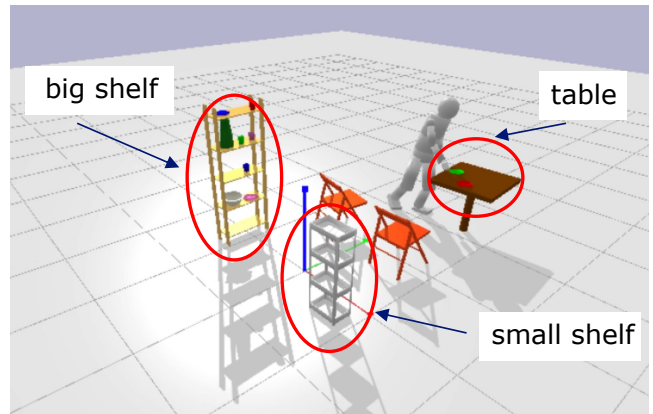


図 3 Mogaze データセット

達点ではなく、予測時刻における動作の到達点を動作予測に活用することで、従来手法では困難であった到達点の切り替わり区間における骨格予測精度を向上させる。

4. 実験

4.1 実験方法

過去の系列情報から予測した到達点情報が動作予測の精度へ与える影響を確認すべく、基本的な作業タスクである pick & place から構成される作業データセットである Mogaze を用いて実験を行った。到達点に対応する動作ラベルを手動で事前にアノテーションし、予測時刻先に予測時刻先に複数の動作候補がある到達点の切り替わり区間をデータセットから評価区間として用いた。評価区間内で骨格情報のみの動作予測モデル、入力に現在の動作の正解到達点と骨格情報を与えたモデル、過去の到達点系列と骨格系列から到達点を予測し、予測した到達点と骨格情報を用いて予測する提案手法の 3 つのモデルを用いて予測精度を比較した。

4.2 使用データセット

人の動きのゴールとなる到達点が急激に変化する作業データセットとして、Mogaze[3] データセットを使用した。Mogaze は、被験者が室内のオブジェクトを操作する際の全身骨格の位置座標、オブジェクトの位置情報、環境情報を記録したデータセットである。表 1 に Mogaze データ

表1 Mogaze データセットに含まれるデータ項目

データ項目	内容
作業者の全身骨格	7人分の21関節の角度情報, 位置情報
オブジェクト	10種類のオブジェクト (cup, bowl) の位置情報
環境情報	室内の家具 (大棚, 机, 小棚) の位置情報

セットに含まれるデータ項目を示す。各被験者は実験実施者からランダムに与えられる指示情報に従って、室内のオブジェクトを右手で掴み、所定の位置まで運び、配置する pick & place を行う。全体で7人の被験者、延べ10サンプルの情報を含む3時間長のサイズのデータになっており、10種類のオブジェクトに対する合計1627回分の pick & place が記録されている。作業者の動きは、全身21関節の3次元のオイラー角度と3次元の base position からなる66次元、120hz で記録されている。pick & place の操作対象となる cup, jug, plate などの10種類オブジェクトは、位置座標の時系列情報が120hz で記録されている。被験者は、pick & place 操作で、table, big shelf, small shelf の3つの場所の間でオブジェクトを移動させる。また、実験指示者が口頭で被験者に指示した作業内容と指示を開始した時刻の情報、操作者がオブジェクトを掴む時刻と置く時刻の情報が記録されている。

4.3 アノテーション

Mogaze データセットに対して、到達点情報に対応する正解情報をモデルに学習させて、評価に用いるため、次の5種類のラベルのアノテーションを行った(図(4))。

Upper half of the shelf 操作者が bigshelf の上段に向かって移動する運動

Lower half of the shelf 操作者が bigshelf の下段に向かって移動する運動

Table 操作者が table の下段に向かって移動する運動

Small shelf 操作者が small shelf に向かって移動する運動

Wait 操作者が実験指示者からの操作を聞く待機時間

ラベルについては、到達点情報に対応したラベルと操作者が実験指示者からの操作を聞く待機時間のラベルの2種類に分けられる。到達点情報は、作業者がオブジェクトを掴む場所と置く場所を4ヶ所 ((bigshelf の上段下段, table, small shelf) を候補点として用いる。各到達点情報ラベルは、あるオブジェクトを掴む動作 (pick), もしくは置く動作 (place) を開始した瞬間から終了するまでをラベル区間としている。pick 動作における到達点情報ラベル情報は、被験者がオブジェクトを掴んだ時刻を開始時間として、置いた時刻を終了時間とする。place 動作における到達点情報ラベル情報は、被験者がオブジェクトを掴んだ時刻を開始時刻として、置いた瞬間を終了時刻とする。wait ラベルは、Mogaze データセットに記録されている実験指示者が

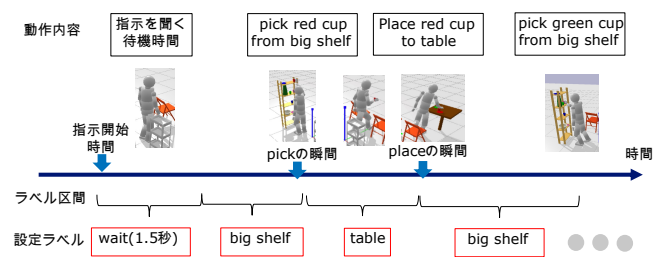


図4 ラベル設定

表2 モデル設定

項目	モデル	入力
比較手法1	動作予測	骨格情報
比較手法2	動作予測	骨格情報, 現在の動作到達点 (正解)
提案手法	到達点予測+動作予測	骨格情報, 到達点系列

被験者へ作業内容を指示する時刻を開始時刻として、1.5秒の長さをラベル区間としている。

4.4 評価方法

本研究では、予測時刻と予測時刻先で到達点が切り替わる区間を評価区間とし、その区間における予測精度を二つの指標を用いて評価を行った。動作予測を行う瞬間の予測時刻から予測対象の時刻である予測先時刻の間の時間の長さを予測時刻幅とする。評価区間は、図5のように、被験者が pick もしくは place を実施する瞬間を中心として、前後の予測時刻幅の区間に設定した。図5の例では、評価区間前半の区間から予測を実行すると、予測時刻では、被験者は table を到達点とした pick 動作をおこなっている、予測先時刻では、bigshelf を到達点とした place 動作をおこなっており、動作の切り替わりが発生している。

評価指標として、評価区間内の予測値と真値間の平均絶対誤差 (Mean Absolute Error, MAE) と最大誤差を使用した。MAE は動作予測の評価指標で広く用いられる指標であり、

$$MAE = \frac{1}{VT} \sum_{k=t}^{k=t+T} \sum_{v=1}^{v=V} |\hat{y}_{vk} - y_{vk}| \quad (1)$$

で表される。ただし、 y_{vk} , \hat{y}_{vk} はそれぞれ時刻 k の関節 v の真値, 予測値に対応し、 V は評価対象の関節数、 T は評価対象の時間幅である。また、特定のシステムの一部として動作予測技術を活用するためには、予測誤差の平均値だけでなく最大予測値を算出して、活用するシステム内での最大誤差許容量との比較を行う必要がある [2] ことから、最大予測誤差も指標として用いる。Mogaze データセット中の pick & place の動作が図6のように右腕で行われていることから、評価区間内における右肩, 右ひじ, 右手首の3つの関節角度を対象として評価を行った。

4.5 モデル

提案手法の動作予測モデル, 到達点予測モデルは共に

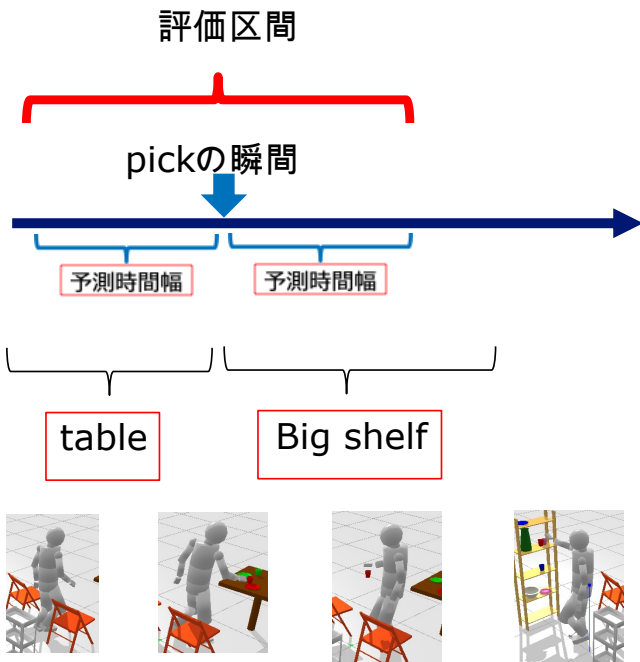


図5 評価区間

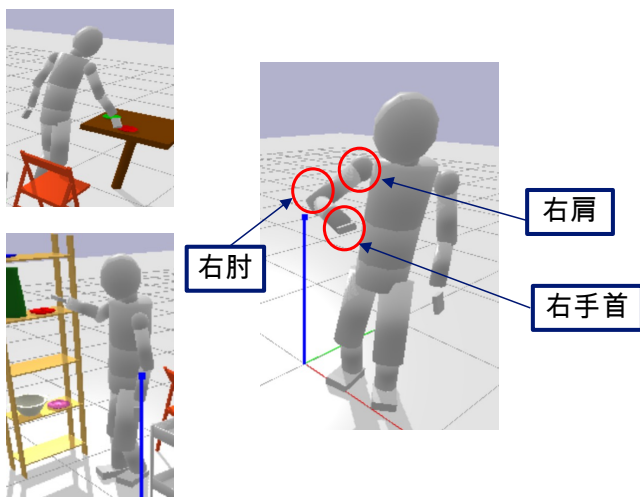


図6 右腕3関節による評価

LSTM(Long Short Term Memory)で実装した。到達点予測モデルは、過去5秒分の全身骨格の21関節63次元の角度情報と到達点情報系列を入力とする。到達点情報系列は入力のtime step分の図4に対応する5次元のone-hotベクトルを用いた。出力として、予測時間幅 $\Delta t = 300\text{msec}$ 先の到達点情報ラベルを予測する。動作予測モデルは、過去2秒分の全身骨格の21関節63次元の角度情報と到達点予測モデルで予測した、予測時刻先のone-hotベクトルで表される到達点ラベルを入力として、予測時間幅 $\Delta t = 300\text{msec}$ 先の全身21関節角度を出力する。線形層において、隠れ状態ベクトル h_t とconcatさせる。到達点情報の動作予測への効果の検証のため、提案手法と合わせて3つの動作予測モデルで実験を行う。使用した3つのモデルの概要とモデルへの入力情報を表2に示す。到達点情報ラベル自体の

表3 各動作切り替わり区間における右腕の評価値

項目	MAE	最大予測誤差
比較手法1	6.25	10.91
比較手法2	6.23	10.42
提案手法	6.16	10.34

動作予測への影響を確認するため、関節情報のみを入力として用いた動作予測モデルを比較手法1として用いる。さらに、到達点切り替わりに対応した本モデルの有効性を確認するため、関節情報と現在の動作の正解到達点ラベルを入力とする動作予測モデルを比較手法2として用いる。検証モデルでは関節情報に加えて、入力として予測時刻先の5次元の到達点情報ラベルをone-hotのベクトル与える。

実験時には、train/validデータを被験者の10人分を9:1の割合に分割し、testデータとして1人分の被験者のデータを用いた。データの前処理として、[0, 1]の値を取るようにMax-Min Normalizationと120hzから20hzのデータの間引きを行った。動作予測器の損失関数には、21関節角度の予測値と真値間の平均絶対誤差(Mean Absolute Error, MAE)を用いる。学習時のバッチサイズは512と設定し、OptimizerにはAdamを用いた。Validation Lossが10エポックで連続で下がらない場合はEarly Stoppingを実施した。

5. 結果

本実験は、提案手法の時系列の骨格情報と予測時刻先での到達点情報予測モデルを活用した動作予測手法が、動作切り替わり区間において有効か調査するために実施した。比較手法1、比較手法2、提案手法の3つのモデルについて、図5で示した評価区間における右腕のMAEと最大予測誤差を表3に示す。過去の骨格情報のみを入力とした比較手法1と、現在の動作の到達点と過去の骨格情報を入力とした比較手法2を比較すると、MAEについてはあまり改善が見られなかったが、現在の動作の到達点情報が最大動作予測誤差を抑えるのに有効であること確認できた。提案手法と比較手法2を比較すると、MAEと最大動作予測誤差の両方で改善が確認でき、動作切り替わり区間での動作予測において、予測時刻先の到達点を用いることの有効性を確認できた。

6. 考察

提案手法の動作予測モデルにおいて予測時刻先の到達点情報があることで、動作の到達点切り替わりによる移動方向の変化のタイミングの情報が伝わり、到達点の切り替わりで通り過ぎなかったことで、最大予測誤差が抑えられたと考えられる。一方で、現在の正解動作ラベルと骨格軌跡を与えたモデルが入力が骨格軌跡のみの予測モデルよりも良かったのは、通り過ぎてしまうことで発生する誤差より

も到達点情報によって同一の動作が安定して継続する効果の方が大きかったためと考えられる。今後、提案手法については、到達点情報と骨格軌跡の入力情報を3次元の位置情報に置き換えることで、モデルに対して、より効果的に人の goal-directed な動きを学習させることでがきると期待される。また、到達点情報の予測についても、視線情報を活用したり、作業者と環境のインタラクションを考慮したりすることで、時間方向、空間方向の両方でより精度高く予測することができるようになると考えられる。更に、到達点について、決定的な推定ではなく確率的に推定することで、人とロボットの協働作業において、推定の確信度が低い場合でも、衝突などのリスクを考慮したより安全な HRC システムが実現できると考えられる。

7. まとめ

本研究では、動作切り替わり区間でのロバストな動作予測技術の創出のため、時系列の骨格情報と予測時刻先での到達点情報予測モデルを活用した動作予測技術の提案と検証を行った。到達点が切り替わる区間において、従来手法のように現在の動作の到達点情報のみを用いるだけでは、予測時間と予測時刻先の間で到達点が切り替わる区間において動作の絞り込みができず、予測精度が低下する。そこで、本研究では、過去の到達点の遷移と骨格情報から、切り替わり先での動作到達点の遷移を予測し、予測時刻先での到達点の切り替わりに対応した。本研究では、pick & place 作業データセットである Mogaze を用いて検証を行った。作業データセットに対して、到達点情報に基づく動作ラベルのアノテーションを行い、骨格情報のみ、骨格情報と現在の到達点ラベルを用いた場合、提案手法である将来の到達点を予測した場合の3つの場合で比較を行い、評価区間における最大予測誤差と平均予測誤差の両方で提案手法が改善することを確認した。今後は、各個別の pick & place 動作シーンにおける切り替わりを考慮した到達点情報の有効性について検証していく。

参考文献

[1] Martinez, J., Black, M. J. and Romero, J.: On Human Motion Prediction Using Recurrent Neural Networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4674–4683 (2017).

[2] Itai, T., Kurai, T., Toide, Y., Makino, Y. and Shinoda, H.: Evaluation of Maximum Error in Jump Motion Prediction, *2022 61st Annual Conference of the Society of Instrument and Control Engineers (SICE)*, pp. 772–777 (2022).

[3] Kratzer, P., Bihlmaier, S., Midlagajni, N. B., Prakash, R., Toussaint, M. and Mainprice, J.: MoGaze: A Dataset of Full-Body Motions That Includes Workspace Geometry and Eye-Gaze, *IEEE Robotics and Automation Letters*, Vol. 6, No. 2, pp. 367–373 (2021).

[4] Zhang, Z., Peng, G., Wang, W., Chen, Y., Jia, Y. and Liu, S.: Prediction-Based Human-Robot Collaboration in Assembly Tasks Using a Learning from Demonstration Model, *Sensors*,

Vol. 22, No. 11 (2022).

[5] Hou, X., Zhang, J., Budagavi, M. and Dey, S.: Head and Body Motion Prediction to Enable Mobile VR Experiences with Low Latency, *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7 (2019).

[6] Luo, J., Huang, D., Li, Y. and Yang, C.: Trajectory Online Adaption Based on Human Motion Prediction for Teleoperation, *IEEE Trans. Autom. Sci. Eng.*, pp. 1–8 (2021).

[7] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A.: Real-Time Human Pose Recognition in Parts from Single Depth Images, *CVPR 2011*, pp. 1297–1304 (2011).

[8] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y.: OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 43, No. 1, p. 172–186 (online), DOI: 10.1109/TPAMI.2019.2929257 (2021).

[9] Ionescu, C., Papava, D., Olaru, V. and Sminchisescu, C.: Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 36, No. 7, pp. 1325–1339 (2014).

[10] Wu, E. and Koike, H.: FuturePose - Mixed Reality Martial Arts Training Using Real-Time 3D Human Pose Forecasting With a RGB Camera, *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1384–1392 (2019).

[11] Chiu, H.-K., Adeli, E., Wang, B., Huang, D.-A. and Niebles, J. C.: Action-Agnostic Human Pose Forecasting, *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1423–1432 (2019).

[12] Ho, J. and Ermon, S.: Generative Adversarial Imitation Learning, *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16*, Red Hook, NY, USA, Curran Associates Inc., p. 4572–4580 (2016).

[13] Wang, B., Adeli, E., Chiu, H., Huang, D. and Niebles, J. C.: Imitation Learning for Human Pose Prediction, *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7123–7132 (2019).

[14] Wang, J. and Su, X.: Enriching Intention of Human Motion Prediction, *Proceedings of 2020 the 6th International Conference on Computing and Data Engineering, ICCDE 2020*, New York, NY, USA, Association for Computing Machinery, pp. 173–178 (2020).

[15] Corona, E., Pumarola, A., Alenyà, G. and Moreno-Noguer, F.: Context-Aware Human Motion Prediction, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6990–6999 (2020).

[16] Cao, Z., Gao, H., Mangalam, K., Cai, Q.-Z., Vo, M. and Malik, J.: Long-Term Human Motion Prediction with Scene Context, *Computer Vision – ECCV 2020* (Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.-M., eds.), Lecture Notes in Computer Science, Cham, Springer International Publishing, pp. 387–404 (online), DOI: 10.1007/978-3-030-58452-8_23 (2020).