

認知リハビリテーションのための行動認識手法の検討

皆本光[†] 佐野睦夫[†]

概要：我々は在宅での認知リハビリテーションを行う方式の確立に取り組んできた。本稿ではそのために必要な行動認識手法について述べる。行動認識はまず未知の時系列骨格データから動作認識を行い、動作認識の結果から行動認識を推定する方式を提案した。動作認識は状態遷移モデルと CNN を用いた方式を提案し、行動認識については HMM と CNN を用いた手法について検討した。未知入力に対する動作認識の全体のエラー率は、CNN を用いた方式が 10% 以上低い結果となった。今後は動画から骨格データを推定するモデルと組み合わせて学習データ量の問題を改善し、実際の認知リハビリテーションのデータに対して認識を行っていく予定である。

A Study on Behavior Recognition Method for Cognitive Rehabilitation

HIKARU MINAMOTO[†] MUTSUO SANO[†]

1. はじめに

高次脳機能障害者や認知症患者・認知症予備軍などの認知障がい者は年々増加傾向にあり、リハビリテーションの専門家や介護福祉士の数は患者数に比べて圧倒的に少ない。その為我々は在宅での認知リハビリテーション手法を確立させようと試みてきた。生活行動に対する認知リハビリテーションに必要とされる行動サイクルを図 1 に示す。

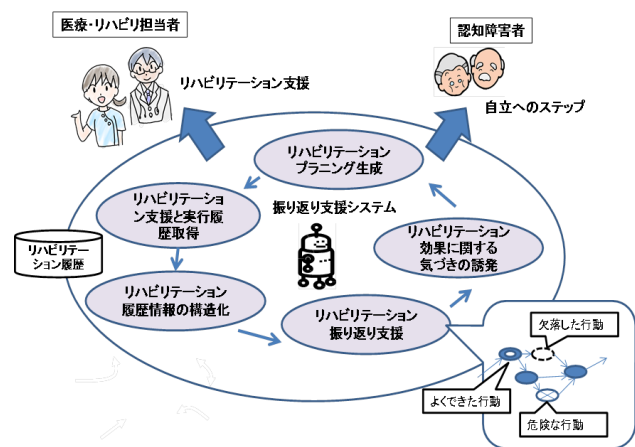


図 1 振り返り支援システムを介した
認知リハビリテーション効果の促進

Fig.1 Rehabilitation effect through review support system

振り返りシステムでは人物の行動を認識し、その行動のタイミングの適切性や内容等について自動で評価を行い、結果を提示することによってユーザに気付きを与えより高い認知リハビリテーションの効果を与えることを目的としている。その一つとして我々が開発を行ってきた生活行動のうち五感を用いて行うことで、より高いリハビリテーショ

ン効果が期待される調理に着目した認知リハビリテーションシステムがある。調理行動に適応させた振り返り支援システムの流れを図 2 に示す。



図 2 調理行動に着目した振り返り支援システム

Fig.2 Reflection support system focusing on cooking behavior

このような振り返り支援システムを実現する為には現在の認知機能を正しく評価できる必要がある。そのためにはロボットが人物の動作やその動作から構成される行動を精確に認識できなければならない。本稿ではこの人物の行動認識手法について述べる。複数の動作より構成される行動を推定するために、まず人物の動作認識を行いその結果を用いて行動を推定する。動作認識は加速度センサを用いての研究が多いが、認知障がい者の自立促すとするならばセンサを正しくつけることが負担であるなどの問題があり、今回のようなシステムでは利用しにくい。そこで RGB-D センサ 1 つだけを用いて動作認識を行う手法を検討する。そして行動認識については、本稿で定義する、複数の動作から構成されるとしたものを識別する研究は多くなされてはいない。そのため行動認識は図 1 のようなシステムで用いられることを想定した手法を提案する。本研究の目標は行動認識が可能となるシステムの構築および可能な限り実用

[†] 大阪工業大学
Osaka Institute of Technology University

レベルの認識精度を達成することである．動作認識では RGB-D センサを用い人物の骨格情報を取得し、骨格状態遷移モデル、畳み込みニューラルネットワーク (Convolutional Neural Network 以下 CNN)を用いた動作認識システムを構築し、システムに適する方式を実験によって決定した．動作認識システムが決定したのち行動認識システムを隠れマルコフモデル (Hidden Markov Model 以下 HMM) と CNN を用いた方式のシステムを構築し、実験によって目的に沿った方式を検討した．

2. 従来研究

生活行動のひとつである調理行動認識に関して、我々は一人称視点映像を用いて両手をトラッキングし、その移動ベクトルを学習し認識を行う方法の研究を行ってきた．また山肩らの研究では温度センサや調理器具に設置した加速度センサを用いて各工程を認識している[1]．今回提案する手法は、ユーザの骨格情報を取得し行動モデルおよび動作モデルの仮説を検証することにより、調理行動・動作の認識を行う．骨格情報を取得する RGB-D センサを基本的に最小数 1 個とする点が大きな相違点となる．また、我々が今まで行ってきた一人称視点映像解析による手法である池ヶ谷[2]らの研究では、調理を行っている手が前面にある物体によって見えなくなるオクルージョンが発生しにくいことや、使用器具の特定が可能であり、本提案手法と相補的な関係にある．映像からの動作認識は①動きを元に認識する方法、②物体やシーン認識を使って認識する方法、③人間の姿勢(ポーズ)を元に認識する方法、に大別されている．小林らの研究[3]では、テレビ映像から特定動作シーンの自動検出を行ったものがある．現時点では食べるシーンのみであるが、顔認識、物体認識、動作認識を組み合わせることによってシーンの判別を行おうとした．ここで行われている動作認識は dense trajectories という時空間特徴量を学習することで行っている．食べる動作の正例および負例を識別する 2 値分類であるが、90%前後の精度である．研究全体としては先の①、②の方法を用いて識別を行っている．本研究では特徴量を用いずにセンサデータを動作ごとにそのまま学習し、最尤推定する点で異なる．この方法により 2 値分類器を複数構成した場合の再学習の手間が大幅に軽減できると考えられる．自身が扱う骨格情報からの認識に近いもので③のアプローチがある．映像から姿勢推定の研究としては DeepPose という Toshev らの研究がある[4]．この研究は DeepNeuralNetworks(DNN) を使った姿勢推定を行っている．図 3 に示すように、映像より人物の姿勢を推定できている．自身の研究との大きな差は DeepPose と異なり 3 次元の骨格情報を扱い、その動作情報も学習する点である．これにより 2 次元の情報では表現できない奥行きを含めた情報が利用可能となり、動作の表現が幅広くなる．

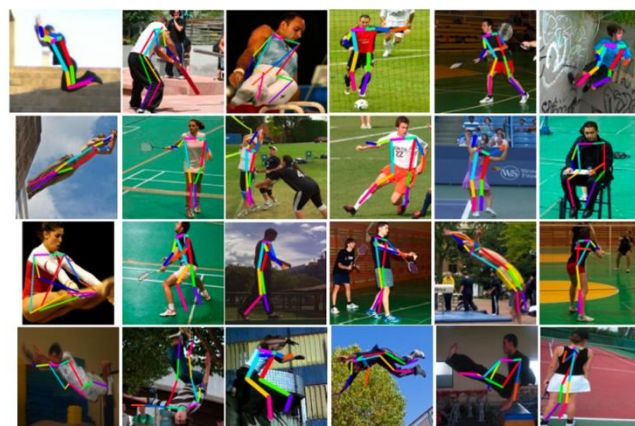


図 3 Deep pose の例

Fig.3 Deep pose example

3. 提案方式

3.1 行動認識システムの流れ

本研究では図 1 で表される認知リハビリテーションシステムにおけるリハビリテーション支援の実行履歴取得およびリハビリテーション履歴情報の構造化、リハビリテーション振り返り支援を行うために必要な動作認識と行動認識手法を検討する．現在、認知リハビリテーションで用いられる行動履歴の取得に用いられるセンサは 1 人称視点映像を取得するためのウェアラブルカメラと加速度センサ、3 人称視点映像および骨格情報を取得するための RGB-D センサが用いられている．本研究で用いるデータは RGB-D センサによって取得できる骨格データのみを扱う．システムの大まかな流れは図 4 のように、骨格データを用いて動作認識層にて動作認識を行い、動作認識層より出力される動作情報を元に行動認識層で入力に対して行動を識別する．

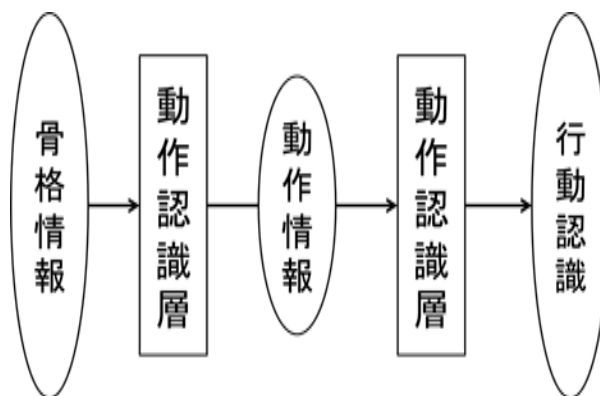


図 4 行動認識システムの流れ

Fig.4 Flow of Behavior Recognition System

3.2 提案システムで扱う骨格データ

提案システムでは、モーションキャプチャなどで取得できる各骨格点の 3 次元座標で表されている骨格データを扱う．骨格データの例を図 5 に示す．

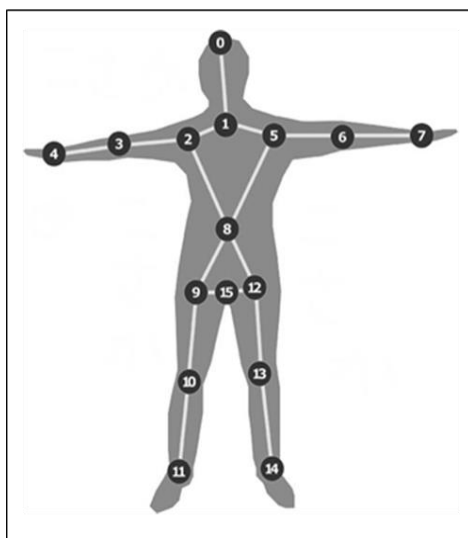


図5 OpenNI で取得できる各骨格点

Fig.5 skeleton points that can be acquired

図5で骨格点の数は16点であるが、学習および識別において同じセンサデータあれば骨格点はいくつでも構わない。

3.3 動作認識層

この節では骨格情報を元に動作認識する手法として、状態遷移に基づく手法とCNNを用いた手法の2種類の手法を提案する。以下の節でその2つの手法について述べる。

(1) 状態遷移モデルに基づく動作認識手法

人物の動作は骨格の動きの流れで表現することが可能だと考えられる。例えばピーラーで皮を剥くといった動作では、右手と左手が非常に近い骨格の状態①から一定の方向へ移動させ離れさせた骨格の状態②の繰り返しで表現できる。この2つの状態をあらかじめ定義しておけば、リアルタイムで取得した骨格状態の遷移が①→②→①→②の経路をたどったときにピーラーで皮を剥く動作を行ったと判定することができる。その他の多く動作についても骨格情報を用いて認識に必要な骨格状態を定義し、その骨格状態の遷移経路によって表現することが可能である。この表現を実現する為に各状態に遷移した際に記号を出力させる。そして遷移ごとに生成された記号列を端末から始点に向けて解析し、それがその行動特定の記号列であった場合、動作情報を出力する。このモデルでは同じ記号を連続で出力することはない。つまり違う骨格状態に遷移した場合のみ記号を出力するような制限をかけている。この制限がない場合、先に挙げた皮むき動作の定義は①→①→②→①→①→①→②や①→②→②→①→②、他にも非常に多くの経路を定義しなければならない。以下の図6に例として皮むき認識を行うための記号列を出力するモデルを示す。この図6表される手振りモデルに新たな行動を定義したい場合は、定義したい行動に必要な骨格状態をこのモデルにフルメッシュで接続するだけで可能である。

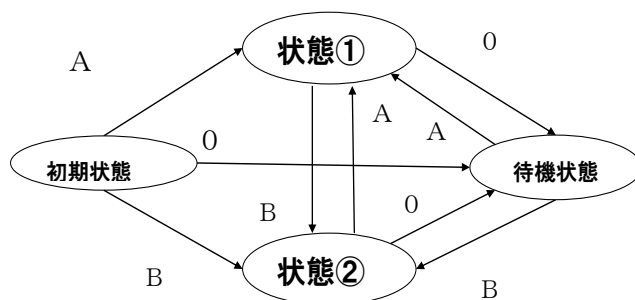


図6 皮むき動作認識モデル例

Fig.6 Example of Peeling Motion Recognition Model

このモデルを使うことによって動作を人がイメージしやすい形で1つ1つ定義でき、必要なものだけを組み合わせることもできる自由度の高い動作定義になることが期待できる。

(2) CNNを用いた動作認識手法

時系列骨格情報を先に述べたCNNに入力し動作認識結果を出力するよう学習する。学習方法はあらかじめ用意した動作の骨格時系列データを入力し、その出力結果がそれぞれの出力動作カテゴリユニットが最も発火するように教師あり学習を行う。入力に際して骨格情報をCNNへ入力できる形に変更しなければならない。用いる骨格情報は1フレームで15点XYZ座標値が取得できるとすると、データのある骨格点を基準としそれぞれの点を基準点からの相対座標に変換し、基準の点を除いた14×3のデータを1フレームのデータとする。この処理は骨格データがセンサの位置によって変化しない(人物の位置情報を持っていない)のであれば必要ない。このデータを任意のフレーム数(n)分並べ、42×nのグレースケール画像とすることによってCNNへの入力を行う。この動作認識手法のイメージを図7に示す。

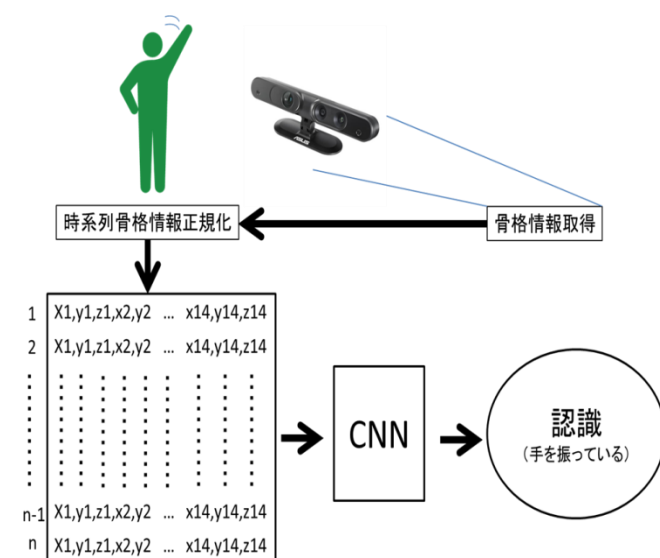


図7 CNNを用いた動作認識手法

Fig.7 Behavior recognition method using CNN

認識の際は未知の時系列骨格データに対し動作カテゴリユニットが最も発火したものを識別された動作とする。この手法の最大の利点は学習データを用意するだけで研究者の経験や知識に基づく特徴量設計や統計的解析を介さず識別を行える可能性があることである。仮に状態遷移モデルに基づく手法と同程度の識別精度であっても、この利点が実用時に大きく影響すると思われる。

3.4 行動認識層

この節では先の動作認識層で骨格情報から動作情報が出力されるので、この情報を元に行動認識を行う。この行動認識層はHMMを用いた手法と、動作認識層の構築でも利用したCNNを用いた手法の2種類で構築を行った。以下にそれぞれの手法について詳しく述べる。

(3) HMMを用いた行動認識手法

動作認識層より出力される動作系列をそれぞれの行動HMMにバウム・ウェルチアルゴリズムによってパラメータ推定させ、未知の動作系列に対し全ての動作HMMにビタビアルゴリズムによって観測系列を出力する最大確率を求める。ここで最も出力確率が高かった行動HMMを未知の動作列に対しての行動とする。

3.5 CNNを用いた行動認識手法

動作認識層より出力される動作情報を動作認識層と同様にCNNに入力し行動認識結果を出力するよう学習する。ただし最終的な行動の意味付けは動作認識とは異なり、未知の動作列に対し入力開始から終了まで入力を終了時間方向へスライドさせながら複数回の行動認識層への入力を行う。データの終端まで入力が終了した際に出てきた観測列において最も発火した回数が最大のユニットを入力に対しての行動とする。この行動認識手法の概略を図8に示す。

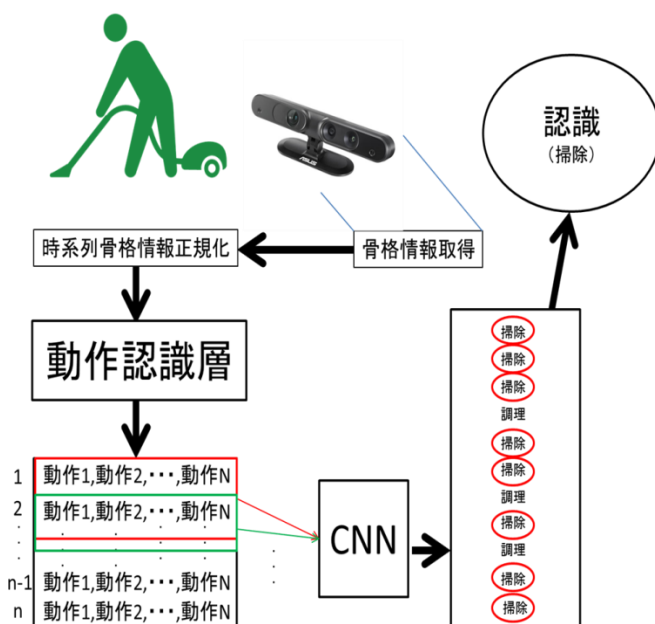


図8 CNNを用いた行動認識手法

Fig.8 Behavior recognition method using CNN

4. 動作認識層の構築・実験

ここでは実際の認知リハビリテーションシステムを想定し、提案手法によって構築した動作認識層の詳しい仕様とその精度の検証実験方法について述べる。

4.1 対象となる認知リハビリテーションシステム

実際に利用された認知リハビリテーションシステムで調理リハビリテーションシステムがある。図2のサイクルで認知リハビリテーションが行われる。そのなかでリハビリ目的と調理ナビゲーションシステムのレシピの決定がされ、そのレシピでの大きな項目が今回認識すべき行動の抽象化レベルとなる。例として図9に肉じゃがのレシピの調理ナビゲーションシステムを示す。



図9 肉じゃがの調理ナビゲーションシステム

Fig.9 Nikujyaga cooking navigation system

今回の行動認識では、この調理行動における下ごしらえおよび加熱調理を識別できることを検討し、動作認識はそれぞれの行動を構成する動作を識別できることを検討する。また現在我々は掃除・片づけで認知リハビリテーション方式の確立に取り組んでいる。現時点で具体的な認識すべき行動の抽象化レベルは判明していないが、調理と同様に、掃除を構成する動作および掃除行動を認識出来るかを確認する。

4.2 使用したRGB-D センサおよび骨格データ

本研究で使用したRGB-D センサは図10で示すASUSのXTION PRO LIVEである。このセンサのスペックは表1に示す。



図10 XTION PRO LIVEの外観

Fig.10 Appearance of XTION PRO LIVE

表 1 XTION PRO LIVE のスペック

Table.1 XTION PRO LIVE spec

センサ有効範囲	水平 58°、垂直 45°、対角 70°(0.8~3.5m)
センサ解像度	640×480ドット(VGA):30fps / 320×240ドット(QVGA):60fps
深度センサの有効距離	0.5~10m

このセンサと OpenNi というライブラリを利用することで、センサからの相対的位置となる 3 次元骨格情報を取得することが可能となる。

4.3 動作認識カテゴリ・学習データ決定

先の 4.1 節で最終的に識別すべき行動を決定した。これらを認識するために、動作認識のカテゴリは「切る」、「混ぜる」、「机を拭く」、「掃除機をかける」、「取る」、「しゃがむ」、「移動」、「手を振る」の 8 つとした。このカテゴリ数に決定したのは実際に状態遷移モデルに基づいて動作認識層を構築していた際、動作認識カテゴリを増やすと骨格状態の組み合わせが爆発的に増え、動作定義が困難となった為である。その為、今回は現実的な範囲で定義できるように、それぞれの動作について骨格状態を定義した。CNN を用いた手法ではそれぞれの動作について学習データを用意した。データについては、まずセンサに向かって正面からの動作についてのみ扱い、さらに骨格情報が実際の人物映像より大きく外れたものは扱っていない。また物を取る、しゃがむはデータ取得開始から 1 回だけ行ったものを 1 データとし、それ以外はデータ取得開始から終了まで(5 秒間)動作を行い続けたものを 1 データとした。1 データあたり 150 フレームあり、時間窓は 42 フレームとし、15 フレームずつスライドさせデータを入力するよう設定した。そのため 1 データあたり 7 回ずつ学習を行った。用意したカテゴリごとのデータは 10 個である。その際 CNN のパラメータは以下の通りである。

- ・入力 42×42×1 の骨格情報
- ・畳み込み層とプーリング層は 3 層ずつ交互に重ねた
- ・畳み込み層の出力にバッチ正規化をかけて学習効果を高めた
- ・第 1 層目の畳み込み層のフィルタサイズは 18×18
第 2, 3 のフィルタサイズは 6×6
- ・それぞれ畳み込み層の出力は 256 枚のフィルタ
- ・プーリングの方法は最大値プーリングでプーリング領域のサイズは 2
- ・活性化関数は tanh、ネットワークの最適化手法は確率的勾配効果法(SGD)
- ・出力に至る全結合層は 2 層で 10%のパーセプトロンの出力をドロップアウト(無効)にした
- ・ネットワークの出力は識別カテゴリ数である 8 ユニット

4.4 動作認識層の性能実験

以上のように作成した 2 種類の動作認識層の性能実験を行う。CNN の学習に用いた教師データ以外の骨格の時系列データを用意し、学習した CNN と状態遷移モデルによる動作認識層に入力して認識精度の比較を行う。学習データ同様、1 データあたり 7 回動作認識を行った。用意したカテゴリごとのデータは 10 個である。

4.5 動作認識層の実験結果

実験の結果は以下の表 2, 表 3, 図 11 のとおりである。表 2 は状態遷移モデルに基づく手法の認識結果。表 3 は CNN を用いた認識手法の結果である。図 11 は対比させた結果を示す。

表 2 状態遷移モデルに基づく手法の結果

Table.2 Results of method based on state transition model

	失敗数(回)	試行数(回)	エラー率
手を振る	6	70	8.6%
拭く	23	70	32.9%
掃除機をかける	58	70	82.9%
切る	8	70	11.4%
取る	65	70	92.9%
混ぜる	7	70	10.0%
しゃがむ	4	70	5.7%
歩く	14	70	20.0%
計	185	560	33.0%

表 3 CNN を用いた手法の結果

Table.3 Results of method based on CNN

	失敗数(回)	試行数(回)	エラー率
手を振る	0	70	0.0%
拭く	3	70	4.3%
掃除機をかける	2	70	2.9%
切る	0	70	0.0%
取る	27	70	38.6%
混ぜる	4	70	5.7%
しゃがむ	52	70	74.3%
歩く	6	70	8.6%
計	94	560	16.8%

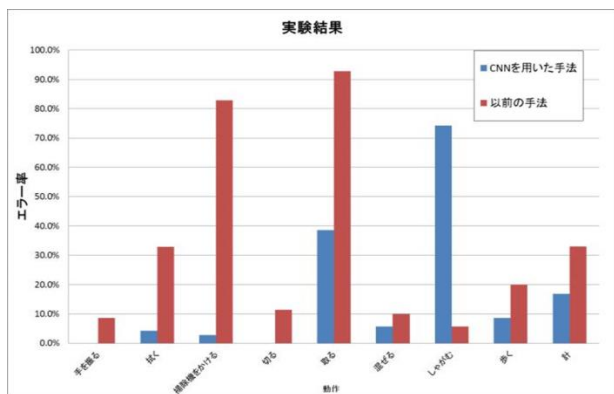


図 11 2 つの手法の比較

Fig.11 Comparison of the two methods

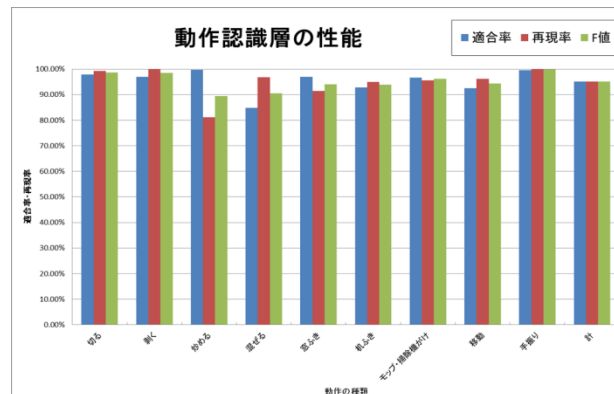


図 12 決定した動作認識層の検証結果グラフ

Fig.12 Verification result graph of action recognition layer

5. 行動認識層の構築・実験

ここでは先の動作認識層の性能実験の結果より、適する動作認識層を決定し、提案手法によって構築した行動認識層の詳しい仕様とその精度の検証実験方法について述べる。

5.1 動作認識層の決定

4.5の結果より全体的にCNNを用いた手法がより良かったため以下の行動認識層で用いるデータは、CNNを用いた手法の動作認識層より得られる動作データを用いる。また動作認識層にてエラー率が高かった取る、しゃがむ動作に関しては行動認識層の検証に大きく影響すると考えられ、学習データを正しく用意することも難しいと考えられた為、これらに代わり皮をむく、炒める、混ぜる動作を学習に加えて、さらに各データを12データ(約1900フレーム分)にし、学習時の窓の移動を1フレームずつに設定し、データを疑似的に増やした。また時間窓の設定は長期の時間的関係性を量み込んでくれることを期待して90フレームとした。その他出力ユニット数以外のパラメータは動作認識層と同様である。最終的に完成した動作認識層に対し、学習と同様にして、各動作1800データについての適合率、再現率、F値を検証した結果を以下の表4および図12に示す。

表 4 決定した動作認識層の検証結果

Table.4 Verification result of motion recognition layer

動作	テスト(回)	認識成功(回)	認識失敗(回)	誤認識(回)	適合率	再現率	F値
切る	1800	1786	14	38	97.92%	99.22%	98.57%
剥く	1800	1800	0	58	96.88%	100.00%	98.41%
炒める	1800	1460	340	4	99.73%	81.11%	89.46%
混ぜる	1800	1743	57	311	84.88%	96.83%	90.45%
窓ふき	1800	1645	155	52	96.94%	91.39%	94.08%
机ふき	1800	1710	90	134	92.73%	95.00%	93.85%
モップ・掃除機がけ	1800	1721	79	59	96.69%	95.61%	96.15%
移動	1800	1732	68	140	92.52%	96.22%	94.34%
手振り	1800	1800	0	7	99.61%	100.00%	99.81%
計	16200	15397	803	803	95.04%	95.04%	95.04%

全体の性能を表すF値が95%程度である為、以降の行動認識ではこの動作認識層の出力を用いる。この動作認識層の未知データに対する出力は一秒ごとに設定した。先の学習で表すならば、これは時間窓を30フレームごとにスライドさせるよう設定したことと同じである。

5.2 行動認識層の学習データ

先の動作認識層は、一回当たりの入力で各ユニットがどれだけ発火したかを表す出力が得られる。CNNを用いた行動認識手法ではこれを任意の数並べて入力すれば良い。今回は9つの出力(9秒分)を並べて入力した。HMMを用いた行動認識手法では、動作認識層の出力ユニットそれぞれに記号を割り当て、一回の入力に対し最も発火したユニットの記号をひとつ出力し、未知の時系列骨格データに対して記号列を出力するようにした。この記号列を用いて行動HMMをバウム・ウェルチアルゴリズムによってパラメータ推定し、ビタビアルゴリズムによってどの行動HMMの出力確率が高いかを求め行動を決定する。学習データはどちらの手法も共通であり、今回識別する下ごしらえ、温調理、掃除の3つの行動を約1分間とったデータを動作認識層に入力し得られたものを用いる。学習データは行動ごとに10個ずつ用意した。CNNのパラメータは入力と出力ユニット以外は動作認識層と同様であり、学習時の時間窓は1ずつスライドさせた。またもう一方の手法であるHMMは、パラメータ推定の際モデルの状態数を指定しなければならないが、これは出力記号の種類数と同じ9とした。

5.3 行動認識層の性能実験

以上のように作成した2種類の行動認識層の性能実験を行う。学習に用いたデータ以外の各行動データを用意し、それぞれの行動認識層に対して入力し、データを正しく識別できるか検証する。各行動10個のデータに対し、それぞれの手法の適合率、再現率、F値求め比較する。

5.4 行動認識層の性能実験結果

実験の結果は以下の表5、表6の通りである。表5はCNNを用いた認識手法、表6はHMMを用いた認識手法の結果である。

表 5 CNN を用いた行動認識手法の結果

Table.5 Results of behavior recognition method using CNN

行動	テスト(回)	認識成功(回)	認識失敗(回)	誤認識(回)	適合率	再現率	F値
下ごしらえ	10	10	0	0	100%	100%	100%
温調理	10	10	0	0	100%	100%	100%
掃除	10	10	0	0	100%	100%	100%
計	30	30	0	0	100%	100%	100%

表 6 HMM を用いた行動認識手法の結果

Table.6 Results of behavior recognition method using HMM

行動	テスト(回)	認識成功(回)	認識失敗(回)	誤認識(回)	適合率	再現率	F値
下ごしらえ	10	10	0	0	100%	100%	100%
温調理	10	10	0	0	100%	100%	100%
掃除	10	10	0	0	100%	100%	100%
計	30	30	0	0	100%	100%	100%

6. 考察

ここでは動作認識層および行動認識層の性能実験の結果について考察する。

6.1 動作認識層の考察

図 4.3 より、しゃがむ動作を除き今回提案した CNN を用いた手法がより、エラー率が低いという結果となった。また CNN を用いた手法では動作の定義を人の手でを行わなくてよい為、認識動作カテゴリを増やした際の動作定義の難しさという問題は無いと考えられる。以上より動作認識層は CNN を用いた手法がより良いと考えられた。CNN を用いた動作認識手法の中で特にエラー率の高かった、取る、しゃがむ動作であるが、これらは学習データ、テストデータともに 5 秒間の間に 1 度しか行動を行っておらず、検証では何もしていない時間でもしゃがむや取る動作が認識失敗とされるためであったと思われる。この問題についてはこれらのデータに対して骨格の時系列データのうちの時点がしゃがむ動作であるとし、教師データに定義する、あるいは学習データを短くし、学習データも増やすといった方法で改善できると考えられる。また学習データを疑似的に増やし、カテゴリ数も増やして作成した図 12 の動作認識層の実験結果であるが、全の F 値が 90% 近くを超えており高い性能が達成できていると考えられる。しかしながら学習に平常状態やこれらの動作以外を判定するようなカテゴリを学習していないため、この結果は実用した際の性能とは乖離していることが予測される。今後その他の動作を学習するためには大量のデータを用意する必要があると考えられ、現時点でも学習するために大量の時系列骨格データに対して動作の教師タグ付けを行うのは大変な労力が必要であると考えられる。今回は奥行き情報がなくなるとして利用しなかった DeepPose であるが、今後はこれを 3 次元座標で取得できるように拡張することが出来れば、動画か

らの学習が可能となるためこの問題は改善されることが考えられた。

6.2 行動認識層の考察

表 5 および表 6 より、どちらの手法も全 30 データに対し正しく識別することが出来た。しかしながらこれも動作認識層と同様に実用を想定した際には問題が生じる。そのためこの結果からは未知入力が少なくとも下ごしらえ、温調理、掃除の 3 つのどれかに絞られている場合に実用可能であるといえなく、実用レベルの認識精度という目標を達成したとは言えない。また 2 つの手法について結果に差異がないため、どちらが行動認識に優れているかは導けなかった。以上より今回作成した行動認識システムは動作認識を CNN で行い行動認識を HMM あるいは CNN を用いて行うことで、未知データが学習カテゴリである場合において高い精度で識別できるものと言える。

以上より掃除と調理といった大別されるような行動を識別する際は、同様のシステムおよび学習方法で識別できることが予想される。また動作認識に関して、炒めるや窓ふきといった継続的な動作は、時系列骨格情報に対し CNN を用いることで高い精度で認識できることが期待される。

7. おわりに

本研究では、認知リハビリテーションシステムにおける、認知能力を評価するために必要な人物の行動認識手法について検討を行った。RGB-D センサを用いて取得できる人物の骨格情報を学習 CNN で学習することで動作認識および行動認識を行う方法を提案した。また行動認識に関しても CNN あるいは HMM を用いて学習する手法を提案した。現時点では未知データが学習データのいずれかのカテゴリに属する仮定のもとで高い性能が得られるという結果となり。行動認識システム全体としては動作認識に CNN を用い、行動認識に HMM あるいは CNN を用いる構造が良いと考えられた。また動作認識に関しては継続的な動作について、時系列骨格情報に対し CNN を用いて学習することで高い認識精度が得られると考えられる。動画から骨格データを推定するモデルと組み合わせて学習データ量の問題を改善しシステム全体の改善を図っていく予定である。

謝辞 本研究の一部は文部科学省研究費補助金(基盤 C15K00368)の支援を受けた。

参考文献

- [1] 山肩洋子,角所考,美濃導彦:食材および調理操作の流れからなる調理行動認識に基づく動的なレシピ認識手法の提案;京都大学情報学研究科 京都大学学術情報メディアセンタ
- [2] 池ヶ谷剛,大井翔,佐野睦夫:認知リハビリテーションのための一人称視点による調理動作認識;PRMU IPSJ-CVIM IBISML 研究発表会(2016)
- [3] 小林隼人,柳井啓司:テレビ映像からの特定動作シーンの自動検出;DEIM Forum 2016
- [4] A. Toshev and C. Szegedy. DeepPose: Human pose estimation via deep neural networks. In Proc.of IEEE Computer Vision and Pattern Recognition, pp. 1653–1660, 2014.