深度画像認識手法を用いた老化に伴う運動機能低下の検知

槌道 慎也^{1,a)} 青木 工太^{1,b)} 槇原 靖^{1,c)} 中村 友哉^{1,d)} 八木 康史^{1,e)} 西川 博文

概要:高齢化が進む社会では介護を必要とする人の増加が問題となっており,その解決には健康寿命を延ば すことが重要である.高齢者が介護を必要とする大きな要因として骨折・転倒が存在しており,転倒事故は 高齢化による運動機能の低下に伴って発生しやすく,より重大なケガに繋がる.転倒事故防止には転倒リス クをいち早く本人が理解し,運動や生活環境の改善などの対策を行うことが必要である.そこで本研究では, 深度画像の動作認識に用いられる手法と画像処理を組み合わせることで転倒リスクの高さを推定する手法 を提案した.転倒リスクは TUG(Timed up & Go Test)を用いて定義し,使用するデータは TUG を真横か ら深度カメラで撮影することで取得した.得られた深度画像を任意に 1 秒間切り出し, 3DV(3D Dynamic Voxel)と PointNet++を用いて学習を行うことで転倒リスクの推定を自動で行う分類器の作成を行った. その結果, 1 秒間に撮影される深度映像から転倒リスクの推定は可能であることが分かり,提案手法の有効 性を確認した.

キーワード:高齢者見守りシステム,転倒リスク推定,深度画像,深層学習

Detection of physical ability decline associated with aging using depth image recognition method

Shinya Tsuchimichi $^{1,a)}$ Kota Aoki $^{1,b)}$ Yasushi Makihara $^{1,c)}$ Tomoya Nakamura $^{1,d)}$ Yasushi Yagi $^{1,e)}$ Hirohumi Nishikawa

Abstract: In an aging society, the increase in the number of people who require nursing care has become a problem, and extending healthy life expectancy is important to solve this problem. Fractures and falls are major causes of the need for nursing care among the elderly, and accidents involving falls tend to occur as physical ability decline due to aging, leading to more serious injuries. In order to prevent falling accidents, it is necessary for people to understand the risk of falling as soon as possible and to take measures such as exercise and improvement of their living environment. Therefore, in this study, we proposed a method for estimating the risk of falling by combining image processing with methods used for action recognition in depth images. The fall risk is defined using the TUG (Timed up & Go Test), and the data used are obtained by capturing the TUG from the side with a depth camera. The obtained depth images were arbitrarily cut out for one second and trained using 3DV (3D Dynamic Voxel) and PointNet++ to create a classifier that can automatically estimate the risk of falling. As a result, we found that it is possible to estimate the risk of falling from the depth image captured in one second, and confirmed the effectiveness of the proposed method.

Keywords: Elderly person watching system, Falls prediction, Depth image, Deep learning

¹ 大阪大学

- University of Osaka, Osaka, Japan
 2 三菱電機株式会社情報技術総合研究所 Information Technology R&D Center, Mitsubishi Electric Corp, Kanagawa, Japan
- ^{a)} tsuchimichi@am.sanken.osaka-u.ac.jp
- ^{b)} aoki.k@am.sanken.osaka-u.ac.jp
- c) makihara@am.sanken.osaka-u.ac.jp

1. はじめに

近年, 医療の発展により平均寿命が延び続けると共に, 健 康上の問題で日常生活が制限されることなく生活できる期

^{d)} nakamura@am.sanken.osaka-u.ac.jp

 $^{^{\}rm e)} \quad {\rm yagi}@{\rm am.sanken.osaka-u.ac.jp}$

間である健康寿命を延ばすことについて関心が高まってい る. 平均寿命と健康寿命の差を縮めることで介護を必要と する人の数を減らすことができ, 社会問題の1つである介 護の負担を減らすことにつながると考えられる. 高齢者が 介護を必要とするようになった主な原因のうち, 骨折・転 倒は認知症, 脳卒中, 高齢による衰弱に次いで4番目に多 い12.5%を占めている[1]. 転倒事故は年を重ねるにつれて 筋力が衰えることでケガが重症化しやすく, けがによる入 院生活や安静は筋力や身体機能のさらなる衰えを引き起こ し, 転倒事故の危険性を高めることにつながる. このような 悪循環を未然に防止するためには, いちはやく運動機能の 低下を検出し、対策を講じることが必要である.

高齢者の転倒事故を事前に予測するものとして転倒ア セスメントスコアシートが利用されている [2][3]. これは, 対象者の状態がスコアシートの各項目に当てはまるかを チェックすることでリスクの推定を行うものである. しか し, このスコアシートは対象者の状態をよく知る医療従事 者や介護者でなければ評価を行うことができず, 定期的な 評価の更新は業務の増加につながると考えられる.

こうした問題に対して,自動で転倒リスクを推定する研 究が行われている.人工知能を用いて電子カルテからリス クの推測を行う転倒転落予測システムのサービス [4] が提 供されており,このサービスの AI は医療従事者が診断し た場合と同程度の精度でスコアを算出できることから業務 の効率化や,定期的なスコアの更新を行うことが期待され る.他には運動機能を評価するテストとして定められた動 作を行い,それにかかる時間を計測する手法や [5],機械学 習を用いた手法として,身に着けた加速度センサからの情 報をもとに転倒リスクを推定する手法 [6] などが提案され ている.

しかし, 上記の手法は定期的に病院に通う必要があることと, 数十秒かかるテストやセンサの着脱を毎日継続することは生活の負担に繋がると考えられる.

そこで本研究では, 深度カメラを用いて日常生活を撮影 し, その映像から自動でリスクを推定するための手法を提 案する. カメラによる撮影は設置をすれば特別な操作は必 要なく, また深度カメラは映る物体までの距離を撮影する ことから RGB 画像と比較して被写体のプライバシーに 配慮されているといったメリットがある. 背景差分と深度 画像の動作認識で用いられる 3D Dynamic Voxel(3DV:以 下 3DV と表記する), 点群に対する機械学習の手法である PointNet++を組み合わせることで深度画像から転倒のリ スクが推定できることを示す.

2. 関連研究

2.1 転倒のリスクを予測する研究

転倒によるケガや転倒そのものを予防するために,リハ ビリによる運動機能の改善や周りの環境をより安全なもの

にするなどの対処を適切に行うには、転倒する危険性を予 測することが重要である. 運動機能を評価するテストとし て Timed Up and Go Test(TUG:以下 TUG と表記する)[5] が一般的に用いられている.しかし、このテストを実施する にはタイムを計測する人や、場合によってはテストを受け る人の介助を行う必要があり、このテストを用いて高齢者 全員の転倒リスクを予測することは, 労力やテストにかか る時間の問題から定期的に行うことは難しいと考えられる. 加えて、TUG で行う動作は日常生活で行うすべての動作を 反映したものではなく, 転倒リスクの予測には限界がある. そのため、より多くの要因を考慮したうえで転倒リスクを 即座に、また定量的に評価するための研究が数多く行われ ている. その1つに, 被験者に装着された加速度センサの 計測結果や被験者の過去の転倒や現在の健康問題などの聞 き取り調査,そしてバイタルサインから得られる情報それ ぞれに重み付けを行い、転倒リスクを3段階に分けて推定 する研究がある [7]. 機械学習を用いた研究の例として加速 度センサを装着した被験者が日常生活を行うことで得られ た情報を、時系列データに対して学習を行うことができる long short-term memory(LSTM) ネットワークに入力する ことで転倒リスクの推測を行う研究がある [6].

2.2 点群に対するクラス分け

点群とは3次元空間上にある点の集合によって構成され たもので、画像と比較して3次元物体の形をよりよく表現 している. 機械学習を用いた点群を分類する手法は数多く 考案されており、その1つが点群をボクセルに変換して3 次元の Convolutional Neural Network(CNN:以下 CNN と 表記する)[8][9] を適用する手法 [10][11] である. しかし, 点 群の密度によってボクセルの解像度が変化することや3次 元の CNN の計算は非常に時間がかかることから、3 次元の 点群から2次元の画像へのレンダリングを複数の視点から 行い,得られた2次元の画像に対して2次元のCNNを適 用する手法 [12] などが提案された. この手法は形状の分類 や検索タスクでは非常に高い精度を達成したが. シーンの 理解や点群の部分ごとの分類や形状補完などの3次元情報 を扱うタスクへの適用が難しい. この問題を解決する手法 として、3次元のデータの各点の座標をベクトルに変換し、 得られた3次元のベクトルに対して全結合層を用いて形状 分類する手法 [13] 及び, それを拡張した手法 [14] が提案さ れている.

2.3 3次元空間上の動作認識

深度カメラは撮影する物体までの距離を計測し, それを 画像として出力する. 深度画像の各ピクセルの値はカメラ に写った物体までの距離であるため, 3 次元空間上の対応し た座標に点を打つことで撮影した物体の形を構成すること が可能である.RGB 画像が 2 次元の平面の情報であるのに 対し, 点群は 3 次元の立体の情報である. このため機械学習 を用いて点群の特徴を捉えるためには RGB 画像で用いら れている手法とは異なるものが必要となる. 既存の 3 次元 の動作認識の手法は深度をベース [15][16] にしたものと骨 格をベース [17][18] にしたものがあり, 近年では RNN[17] や GCN [19] を用いた骨格ベースの手法が注目を集めてい る. しかし, RGB 画像や深度画像から正確に人間の骨格を 推定することは難しい場合がある. そのため, より実用的 な手法として深度ベースの手法が好ましいと考えられてお り, CNN を用いた手法 [20][21] は高い精度での動作認識を 実現している.

3. 提案手法

本研究では TUG の様子を撮影して得られた深度画像列 からテストの結果を分類することを行う.TUG は 11 秒以 上かかった場合,運動器不安定症の可能性が高いと診断さ れるため,11 秒を境界として深度画像列を2値分類できる かどうか実験する.動作認識に用いられる手法を利用して 深度画像列から11 秒以上の人が持つ特徴を抽出し,未知の データを分類することができるか否かを実験する.本章で は使用する手法について説明する.

3.1 3D Dynamic Voxel

深度画像に加えてカメラの焦点距離と画像の中心を意味 するパラメータを用いることで深度画像を点群に変換する ことができる. 深度画像列で表現される動作の特徴を抽出 するためにすべての点群を入力に使用することは学習の 実行速度やメモリの使用量の観点から見て非効率である. そのため,複数の深度画像の情報を圧縮しつつ,学習を行 うのに十分な情報を保持した表現手法である 3D Dynamic Voxel (3DV:以下 3DV と表記する) [22] を入力データとし て用いた.3DV は複数の点群を1つの3次元空間上に重ね 合わせて構成されるため,1つの点群で3次元空間上の点 群の動きを表現することができることから,深度をベース にした動作認識を高い精度で行うことができる [22].

3.1.1 Point cloud から Voxel への変換

点は体積がないため, 深度画像から得られる点群を重ね 合わせることはできない. そのため点を体積を持つボクセ ルに変換することで点群の重ね合わせを実現した. ボクセ ルの一辺は 35mm で各ボクセルは 0 か 1 の値を持ち, ボク セル内部に点が存在している場合は 1, それ以外の場合は 0 がボクセルの値となる. 図 1(b) は値が 1 のボクセルのみを 青のブロックで表現したものである.

3.1.2 Voxel から 3DV への変換

フレーム総数 N の深度画像列から 3DV を作成するこ とを考える.t フレーム目のボクセル V の座標 (x, y, z) を $V_t(x, y, z)$ とすると $V_t(x, y, z)$ の値は 1 か 0 であり, 作成さ れる 3DV を D, その座標 (x, y, z) を D(x, y, z)としたとき,



Fig. 1 Point clouds and the voxels created by transforming them

D の各座標の値は

$$D(x, y, z) = \frac{V_t(x, y, z) * t}{N} \tag{1}$$

となる.

撮影する角度や距離による 3DV の値や座標の不揃いは 機械学習を行うにあたってノイズとなる可能性がある. こ のため, D に対して以下のように座標と 3DV の値を正規化 する.

- **座標** y の最大値が 0.5,最小値が-0.5 になるように正規化 し, x と z はそれぞれ y に対して行われた正規化の比 率に応じてスケーリングする
- **3DV の値**最大値が 0.5,最小値が-0.5 になるように正規 化する

3DV を N フレーム全体から 1 つ作るだけでは動作の 途中で現れる細かな特徴が消えてしまう可能性がある. そ のため、細かな特徴を保持し続けるために m_G に加えて m_1, m_2, m_3, m_4 を N フレームの一部から 3DV を追加で作 成する (図 2). 本実験では $\frac{2N}{5}$ フレームずつスライド幅 $\frac{N}{5}$ で切り出して 4 つの 3DV(m_1, m_2, m_3, m_4) とした.





最終的にフレーム総数 N の深度画像列から作成される データ P の座標 D(x, y, z) が持つ値は

$$P_{D(x,y,z)} = (\overbrace{x',y',z'}^{Spatial}, \overbrace{m_G,m_1,m_2,m_3,m_4}^{Motion})$$
(2)

となる.Pに含まれるデータの内, -0.5 以外の 3DV の値を

持つ座標をランダムに 2048 ヶ所サンプリングし, 作成された 2048 × 8 のデータがネットワークに入力される.

3.2 PointNet++

ネットワークへの入力データは3.1節で説明した2048×8 のデータであり、出力は正例(TUGの所要時間11秒以上) と負例(TUGの所要時間11秒未満)のそれぞれに対する 確信度を表す2つの値である.正例に見られる特徴の抽出 には点群データに対してうまく機械学習を行えるPoint-Net[13]を拡張したPointNet++[14]を使用し、損失関数は cross-entropyを用いた.この節では実験に使用したネット ワークについて説明する.

3.2.1 PointNet++の概要

点群は近傍の点同士は強い関係性(局所性)を持つこと から,1つの点が持つ情報だけでなく複数の点同士の関係 性を学習することで,点群の分類やセグメンテーションを 高精度で行うことができると考えられる.また,点群を回転 移動や平行移動,拡大,縮小したものも同じ種類の物体を意 味する点群であり(移動不変性),点の生成順に関わらず 各点の座標が同じであれば同一の点群である.点群が有す るこれらの特徴を考慮しつつ学習を行うことができるネッ トワークとして PointNet が提案されている.しかし,点群 はある程度離れた点同士も弱い関係性を有していることか ら近傍の点同士の関係性を学習するだけでは不十分である 場合がある.このことから PointNet++は PointNet を再帰 的に用いることである程度離れた点群同士の関係性も学習 できる.

3.2.2 PointNet++のネットワーク構造

今回使用した PointNet++の実装の概要を示したものが 図 4 である.

このネットワークでは局所性を sampling & grouping, 順序不変性を Max Pooling を用いることで獲得した. 移動 不変性の平行移動や拡大,縮小はネットワークへ入力する 座標データの正規化を行うことで対応できるが,回転移動 に対しては対応していない. これは実験に使用したデータ がすべて同一の角度から撮影されたものであるため,今回 はネットワーク内で回転移動に対応する処理を実装せずに 学習を行った.

sampling & grouping では局所性を獲得するために近傍 の点の情報をまとめる処理を行う.入力された点の中から ランダムにサンプリングを行い,サンプリングした点の近 辺にある点の情報を収集する.その後,PointNet にて 収集 した複数の点の情報を CNN を用いて一つにまとめる.こ れを繰り返すことで学習に必要な情報を残しつつ情報量を 圧縮することで全結合層での学習を効率的に行っている.

4. 運動機能評価テストのデータ

本研究では被験者が TUG を行う様子を深度カメラで撮

影して得られた深度画像を用いて実験を行った. 本章では 実験で用いたデータについて説明する.

4.1 運動機能評価テストの概要

本研究では,運動機能が低下しているか否かを判断する ためのテストとして TUG を用いた.TUG は椅子に座った 姿勢から立ち上がり, 3m 先の目印で折り返して再度椅子に 座るまでの時間を測定する.TUG にかかった時間が 11 秒 以上の場合,歩行・移動能力の低下による転倒の危険性が 高い運動器不安定症の診断根拠の1つとなる [23].

4.2 撮影環境

今回の実験では Microsoft Kinect v2 を用いて深度画像 を解像度 512 × 424, 30fps で撮影を行った. 図 5 はカメラ と TUG を行う場所の位置関係を表したものでカメラは高 さ 1.6m の地点に設置されている. 図 6 は撮影された RGB 画像と深度画像である.

4.3 画像データの選別

TUG の映像は必要なテストの映像だけを撮影するので はなく、カメラを止めることなく撮影を続けて後から必要 な画像だけを選別することで映像を取得した. テストの開 始と終了の判断は同時に撮影された RGB 画像を確認する ことで行われ、開始は人が立ち上がるために動き始めた瞬 間,終了は人が座り終えて静止する瞬間とした. 開始から終 了までの時間に撮影された深度画像列を1回分の TUG の 映像として用いる. 被験者実験によって得られたデータの 年齢と TUG の時間の関係を図7に示す.

TUG が 11 秒以上である被験者 (Positive) とそうでない 被験者 (Negative) のデータ数に偏りがあり, 機械学習を行 う上でこの偏りは過学習につながる場合がある. そのため, 65 歳以上すべての被験者のデータを用いるのではなく, 一 部の被験者データのみを用いることにした. 加えて, 学習 ネットワークの汎化性能やデータセットの分析のために交 差検証 (cross-validation) を用いて実験を行う.

Group 1		Group 2		Gi	roup 3	Group 4				
ID	TUG	ID	TUG	ID	TUG	ID	TUG			
0	11.514	6	11.416	12	12.736	18	10.768			
1	9.522	7	11.133	13	8.276	19	13.803			
2	6.501	8	7.167	14	11.335	20	7.931			
3	12.624	9	9.033	15	9.234	21	12.100			
4	10.033	10	10.433	16	10.234	22	9.200			
5	8.100	11	8.868	17	7.800	23	8.567			
					-					

表 1:	谷グルー	70)	ID	2	TUGのスコア

 Table 1
 Subject ID and TUG scores for each group

表1は各グループの被験者の ID と TUG のスコアをま



図 3: PointNet++の概念図

 \boxtimes 4: Conceptual diagram of PointNet++



Fig. 5 Filming environment





(b) Depth image図 6: 撮影画像Fig. 6 Acquired images

とめたものである. 各グループは 11 秒以上の被験者が 2 人, 11 秒未満の被験者は 6 秒以上 8 秒未満, 8 以上から 9 秒未 満, 9 秒以上から 10 秒未満, 10 秒以上から 11 秒未満の被 験者がそれぞれ 1 人含まれており, 今回の実験はこの 24 名 の被験者のデータを用いて行った.

5. 実験

5.1 前処理

3DV の参考元の論文 [22] では深度カメラで撮影した画 像をそのまま用いて 3DV を作成していたが, 深度カメラ で撮影した被写体の行動に関係するのは画像内の一部の画 素のみである. そのため, 部屋の中にある被験者以外の物体





が学習に影響を及ぼさないよう,深度画像内の被験者の領 域のみを抽出することにした.この処理は被験者以外の人 やノイズの除去 8(b),そして被験者以外の静止物体の除去 8(c)の2つの処理に分けて行った.2つ目の処理では背景差 分を取ることで被験者のみの領域の抽出を行っているが, その際,被験者以外の動く物体が差分画像に含まれてしま うため1つ目の処理を行った.

図 8(b) は図 8(a) の画面下部分と, 一定以上の距離の画素 値を持つピクセルをすべてゼロ埋めした画像である. これ により深度画像内の TUG が行われる領域のみ抽出するこ とができる. 図 8(c) は OpenCV に実装されている MOG2 を用いて背景差分を取得することで人物領域のみの抽出を 行った.

5.2 実験の組み合わせ

今回の実験では、図 8(b) まで処理を行った画像と図 8(c) まで処理を行った画像の 2 種類の深度画像から 3DV を作 成して学習・テストを行った.また、実験は以下のような異 なる設定で行い、それぞれの結果を比較した.

- **実験1** 椅子や床を含んだ深度画像から作成した 3DV を 使用
- **実験 2** 人物のみが写った深度画像から作成した 3DV を 使用
- **実験3** 実験2のデータに加えて、それらを垂直方向の軸 を中心に±15度ずつ回転させたデータを学習用デー



(a) Acquired depth image







(c) Image of subject area only

図 8: 不要な情報の除去

Fig. 8 Removing unnecessary information タに使用

1	2	3	4
Group 3	Group 4	Group 1	Group 2
Group 4	Group 1	Group 2	Group 3
Group 2	Group 3	Group 4	Group 1
Group 1	Group 2	Group 3	Group 4
	1 Group 3 Group 4 Group 2 Group 1	1 2 Group 3 Group 4 Group 4 Group 1 Group 2 Group 3 Group 1 Group 2	1 2 3 Group 3 Group 4 Group 1 Group 2 Group 2 Group 3 Group 4 Group 4 Group 4 Group 3 Group 4 Group 4 Group 4 Group 3 Group 4 Group 4 Group 4 Group 3 Group 4 Group 4

表 2: 実験データの組み合わせ

 Table 2
 Combination of experimental data

また,本実験は交差検証を用いて行った.1 で示した 4 つ のグループの内, 2 つを Training データ, 1 つを Validation データ, 1 つを Test データとして使用し,表 2 のように交 差検証を行う.

5.3 データセット

3DV データは 3.1.2 で記述した通りに作成され,本実験 では N = 30 として実験を行った (図 9). フレーム総数 M の TUG 映像から切り出す 30F の範囲を 1F ずつずらす ことで,計 M-29 個の 3DV を作成した.24 名の被験者デー タに対してこの処理を行った結果,得られたデータの数を まとめたものが表 3 である.また,各 3DV データの正解 ラベルは TUG のスコアが 11 秒以上の被験者のデータを 1(Positive), 11 秒未満の被験者のデータを 0(Negative) と 定義した.



図 9: TUG の映像と 3DV の数の関係 (F:フレーム)

Fig. 9 Relationship between TUG picture and 3DV number (F:Frame)

label\Group	1	2	3	4			
Negative	808	953	913	868			
Positive	532	513	514	640			
合計	1340	1466	1427	1508			
表 3: 実験 1, 2 のグループ別のデータ数							

 Table 3
 Number of data by group for experiments 1, 2

表3は各グループの被験者データから作成されたラベル ごとのデータの数とその合計を表したものである. 実験3 では軸の回転によるデータ拡張によって表3の3倍の数の データが作成され,その中かから Negative のデータを1984 個, Positive のデータを992 個ランダムにサンプリングし て Training データとした. また, Validation データと Test データは実験2と同様のデータを使用した.

5.4 評価手法

実験結果に対し, ROC(Receiver Operating Characteristic) カーブと識別率を用いて精度評価を行った.ROC カー ブは縦軸に感度 (Sensitivity), 横軸に 1-特異度 (Specificity) の 2 つの指標をとる.ROC 曲線はグラフ左上の頂点に近 付く曲線ほど精度が高い結果である.入力された 3DV か ら実験で用いるネットワークは 2 つのスコア (Negative, Positive の確信度)を出力し,スコアの和は 1.0, 各スコア は 0.0 1.0 の値をとる.Positive の確信度がある閾値以上の 時に推定値を 1 とし, 閾値を 0.01(1%) 刻みで 0.0 1.0 の間 で変化させたときの精度をプロットすることで ROC カー ブを作成した.

6. 結果

実験は以下の3つを調査するために行った.

- 運動機能低下の検知に対する提案手法の有効性
- 床と椅子が精度に与える影響
- 3DV の角度の変更によるデータ拡張の有効性

表4は5.2で述べた各実験のグループごとのテストデータ に対して最も感度と特異度の和が高かった時の閾値の値 を示したものである.また,図10に表4の閾値毎の結果を 実験別にまとめた ROC カーブを示した.表5にその時の AUC,感度,特異度を示す.

表 6 は TUG で行われる動作を起立(①),歩行 1(②), ターン 1(③),歩行 2(④),ターン 2(⑤),着席(⑥)の 6 つ の動作に分割したときの動作別の精度を実験ごとにまとめ たものである.また,表 7 は TUG のタイム別の精度を実験 ごとにまとめたものである.

図 10 において, 実験 1 と実験 2 ではほとんど違いがな く, 実験 3 が最も精度が高いことが確認できる. また, 表 6, 7 から実験 3 においてある程度の精度で任意の 1 秒の映像 から提案手法を用いることで運動機能低下の推定が可能で あることと, 3DV の回転によるデータ拡張の有効性が確認 できる.

実験	グループ 1	グループ 2	グループ 3	グループ 4			
1	0.14	0.29	0.01	0.01			
2	0.02	0.05	0.78	0.04			
3	0.03	0.98	0.77	0.97			

Table 4 Thresholds for each experiment



図 10: 実験ごとの ROC カーブ **Fig. 10** ROC curves for each experiment

実験	AUC	感度	特異度
1	0.6469	0.5918	0.6469
2	0.6196	0.6104	0.5956
3	0.6874	0.6941	0.6430

表 5: 実験ごとのテスト結果

Table 5 Test results for each experiment

75.0%	60.8%
101070	00.070
69.0%	60.4%
72.3%	67.2%
	69.0% 72.3%

表 6: 実験ごとの動作別の精度

Table 6 Accuracy by behavior for each experiment

実験	6s & 7s	8s	9s	10s	11s	12s	13s
1	63.3%	74.7%	56.8%	45.2%	67.8%	58.3%	66.1%
2	67.4%	76.2%	62.5%	43.5%	57.8%	56.2%	73.0%
3	85.3%	82.3%	66.1%	47.3%	53.4%	73.4%	78.6%
表 7: 実験ごとの TUG のスコア別の精度							

Table 7 Accuracy by TUG score for each experiment

7. 考察

6章において実験3がほとんどの項目において精度が最 良の結果となった.実験1と2の比較から床や椅子の有無 は精度に大きく影響を及ぼさないことから,撮影環境の違 いによる静止物体の差異による影響を受けない人物領域の みの抽出が分類に適していると考えられる.また,実験2と 3の結果から3DVの回転によるデータの拡張は推定の精度 を大きく向上させることが分かった.

表4の通り実験1や2は閾値が大きく0(負例)に偏って おり,実験1は精度も正例と比較して負例の方が高い(表 5).これは学習に用いた正例と負例のデータの数に偏りが あり,負例の数が多いため推測値が0に偏ってしまったこ とが原因であると考えられる.そのため,正例と負例のデー タ数を均衡にするためにデータ拡張によって正例,負例共 にデータ数を増やし実験を行った.しかし,実験3の結果か ら分かる通り,データ拡張を行うと実験1や2とは逆に推 測値や閾値が1(正例)へ偏っており,負例の精度がほとん ど向上していないことが分かる.これは正例の被験者のバ リエーションが少ないことからデータ拡張によって正例の 特徴をより強く学習してしまい,負例の被験者の特徴の一 部を正例として分類してしまったことが原因であると考え られる.元の被験者と3DVデータの正解ラベルの比率を検 討することが重要であると考えられる.

動作別, TUG のスコア別の考察は最も精度が高かった実 験3のテスト結果を用いて行った.動作別の精度を比較す ると歩行1,ターン2,着席で精度が高く,ターン1と歩行 2で精度が低い.ターン1,2は人によって動き方が異なる 動作であり,運動機能低下の特徴ではなく右回り,左回りな どの動き方を学習して分類している可能性が高い.このた め,ターンではテストデータに対して高い精度で推測を行 えたとしても汎用性は低いと考えられる.また,歩行2は椅 子に座るための減速が入るため,一部の負例のデータを正 例と推定したため精度が低くなったと考えられる.このこ とから日常生活の映像から分類を行う場合,人によって動 作に違いが表れにくい動作が分類に適していると考えられ る. しかし, 歩行は部屋の広さによってはまっすぐ歩くこと が少ないことから, 起立や着席が最も分類に適した動作で ある.

表7から9秒未満と12秒以上の被験者のデータに対し て高い精度で推測が出来ている一方で,10秒代と11秒代 の被験者のデータに対しては精度が低い.また,実験3は 正例に推測が偏ったことから9秒台の被験者の精度は低く なっている.10秒代と11秒代の被験者データに対して精度 を向上させるためには被験者の数を増やすことに加えて, 正例と負例のデータの数のバランスを調整することが重要 であると考えられる.

8. 結論と課題

本研究では 3DV と PointNet++を用いることで深度映 像の任意の1秒から運動機能が低下している人とそうでな い人を分類できる可能性を示した.また,撮影環境が同じ場 合でも床や椅子が分類に与える影響はほとんどないことか ら,撮影環境が変化した時でも安定して推定が行えると考 えられる人物領域のみのデータが推定に最も適しているこ とがわかった.

今後の課題としては, 今回は TUG で行われるすべての 動作を1つの分類器に学習させたが, 動作毎にデータを分 類し学習させる手法や, 3DV データ自体をクラスタリング し, そのクラスの分類結果と 3DV のデータを用いて運動機 能の低下の検知をするなど, 行われている動作の情報を学 習に用いる手法の検討をすることが挙げられる.

謝辞

本研究の一部は三菱電機株式会社との共同研究で実施し たものである.

参考文献

- [1] 内閣府. 令和2年版高齡社会白書(全体版), 2020.
- [2] 久保和子大杉博美. アセス メントスコアを用いた効果 的な転倒転落防止への取り組み. Tokushima Red Cross Hospital Medical Journal 8, pp. 42–145, 2003.
- [3] 平井さよ子 賀沢弥貴 安西由美子森田恵美子. 転倒アセス メントスコアシートの改訂と看護師の評定者間一致性の 検討. 日看管会誌, Vol. 14, No. 1, 2010.
- [4] 株式会社 FRONTEO. 転倒転落予測 ai システム "coroban®". https://www.fronteo.com/.
- [5] D PODSIADLO. The timed "up & go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc, Vol. 39, pp. 142–148, 1991.
- [6] Deep learning to predict falls in older adults based on daily-life trunk accelerometry. *Pattern Recognition Letters*, 2018.
- [7] H. GholamHosseini, M. M. Baig, M. J. Connolly, and M. Lindén. A multifactorial falls risk prediction model for hospitalized older adults. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3484–3487, 2014.
- [8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In 2016 IEEE Conference on

Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

- [9] L. Zhang, Z. Shi, M. M. Cheng, Y. Liu, J. W. Bian, J. T. Zhou, G. Zheng, and Z. Zeng. Nonlinear regression via deep negative correlation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.
- [10] Zhirong Wu, S. Song, A. Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and J. Xiao. 3d shapenets: A deep representation for volumetric shapes. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1912–1920, 2015.
- [11] D. Maturana and S. Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 922–928, 2015.
- [12] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In 2015 IEEE International Conference on Computer Vision (ICCV), pp. 945–953, 2015.
- [13] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 77–85, 2017.
- [14] Hao Su Charles Ruizhongtai Qi, Li Yi and Leonidas J Guibas Tuytelaars. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. 2017.
- [15] Amir Shahroudy D. Xu Liu, Jun and G. Wang. Spatiotemporal lstm with trust gates for 3d human action recognition. 2016.
- [16] J. Liu, G. Wang, P. Hu, L. Duan, and A. C. Kot. Global context-aware attention lstm networks for 3d action recognition. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3671–3680, 2017.
- [17] O. Oreifej and Z. Liu. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 716–723, 2013.
- [18] X. Yang and Y. Tian. Super normal vector for activity recognition using depth sequences. In 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 804–811, 2014.
- [19] L. Shi, Y. Zhang, J. Cheng, and H. Lu. Two-stream adaptive graph convolutional networks for skeletonbased action recognition. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12018–12027, 2019.
- [20] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould. Dynamic image networks for action recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3034–3042, 2016.
- [21] H. Bilen, B. Fernando, E. Gavves, and A. Vedaldi. Action recognition with dynamic image networks. *IEEE Transactions on Pattern Analysis and Machine Intelli*gence, Vol. 40, No. 12, pp. 2799–2813, 2018.
- [22] Fu Xiong Wenxiang Jiang Zhiguo Cao Joey Tianyi Zhou Yancheng Wang, Yang Xiao and Junsong Yuan. 3dv: 3d dynamic voxel for action recognition in depth video. 2020.
- [23] 公益財団法人 日本整形外科学会. https://www.joa.or. jp/index.html.