Vol.2021-NL-248 No.17
Vol.2021-CVIM-226 No.17
2021/5/21

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. 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

[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
[d)] nakamura@am.sanken.osaka-u.ac.jp
[e)] yagi@am.sanken.osaka-u.ac.jp

Vol.2021-NL-248 No.17
Vol.2021-CVIM-226 No.17
2021/5/21

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

Vol.2021-NL-248 No.17
Vol.2021-CVIM-226 No.17
2021/5/21

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

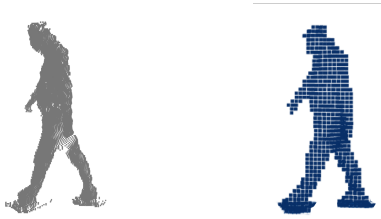
深度画像に加えてカメラの焦点距離と画像の中心を意味するパラメータを用いることで深度画像を点群に変換することができる. 深度画像列で表現される動作の特徴を抽出するためにすべての点群を入力に使用することは学習の実行速度やメモリの使用量の観点から見て非効率である. そのため, 複数の深度画像の情報を圧縮しつつ, 学習を行うのに十分な情報を保持した表現手法である 3D Dynamic Voxel（3DV:以下 3DV と表記する）[22] を入力データとして用いた.3DV は複数の点群を 1 つの 3 次元空間上に重ね合わせて構成されるため, 1 つの点群で 3 次元空間上の点群の動きを表現することが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



(a) Point cloud　　(b) Voxel

図 1: 点群とそれを変換したボクセル

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.

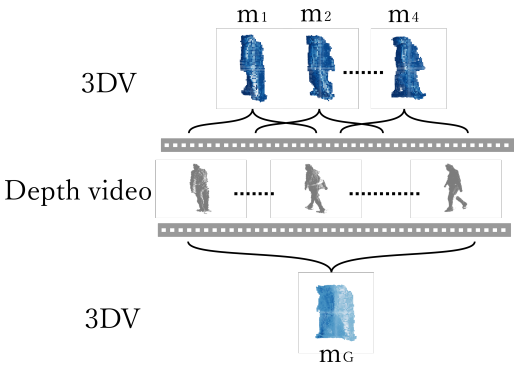


図 2: 時間で分割した映像からの 3DV の作成

Fig. 2 Temporal split for 3DV extraction

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

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

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 PointNet[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		Group 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: 各グループの被験者 ID と 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

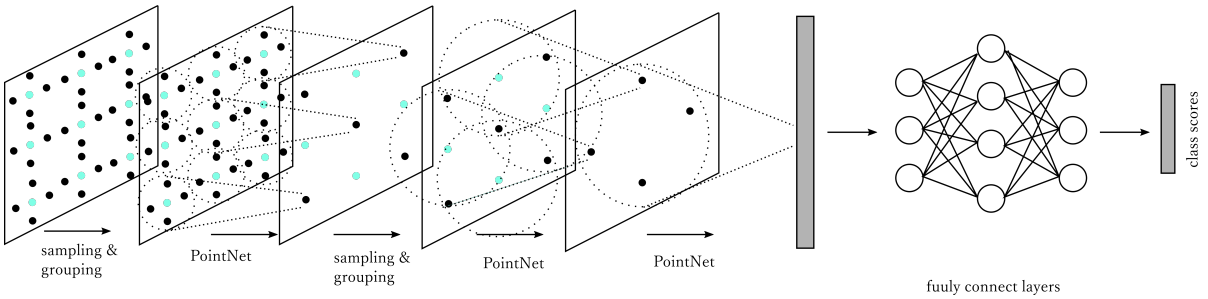


図 3: PointNet++の概念図

図 4: Conceptual diagram of PointNet++

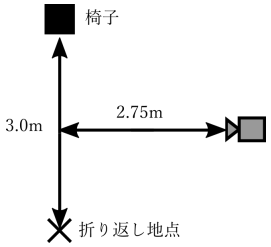
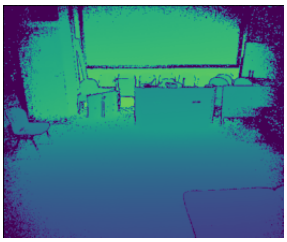


図 5: 撮影環境

Fig. 5 Filming environment

(a) RGB image



(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

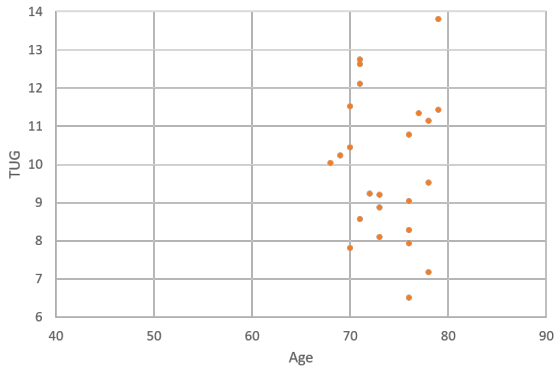


図 7: 年齢と TUG の時間の分布

Fig. 7 Age and TUG score distribution

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

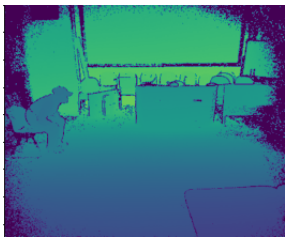
今回の実験では, 図 8(b) まで処理を行った画像と図 8(c) まで処理を行った画像の 2 種類の深度画像から 3DV を作成して学習・テストを行った. また, 実験は以下のような異なる設定で行い, それぞれの結果を比較した.

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

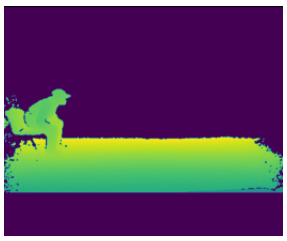
実験 2 人物のみが写った深度画像から作成した 3DV を使用

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

Vol.2021-NL-248 No.17
Vol.2021-CVIM-226 No.17
2021/5/21



(a) Acquired depth image



(b) Image of the floor and chair with the subject

(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

データの種類 \ 回数	1	2	3	4
Training	Group 3 Group 4	Group 4 Group 1	Group 1 Group 2	Group 2 Group 3
Validation	Group 2	Group 3	Group 4	Group 1
Test	Group 1	Group 2	Group 3	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.

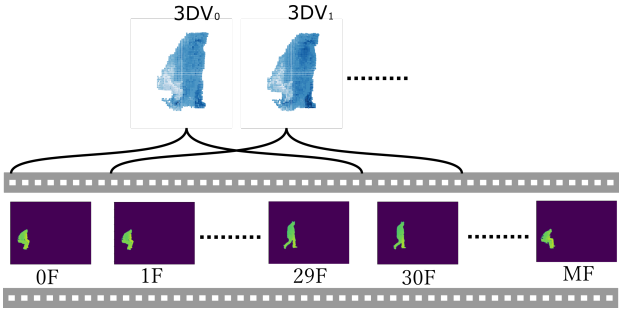


図 9: TUG の映像と 3DV の数の関係 (F:フレーム)

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

表 4: 実験ごとの閾値の値

Table 4 Thresholds for each experiment

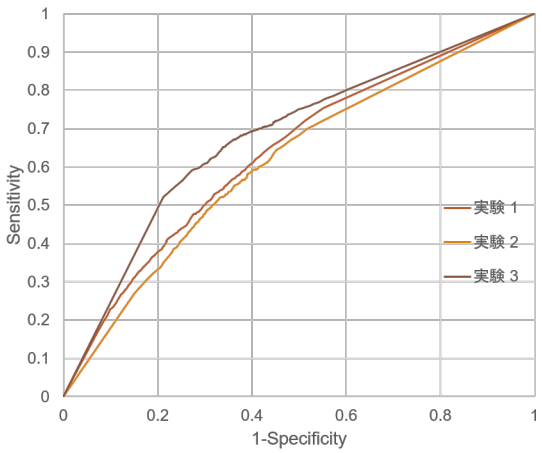


図 10: 実験ごとの ROC カーブ

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

実験	①	②	③	④	⑤	⑥	全体
1	53.5%	50.4%	63.9%	61.6%	68.8%	75.0%	60.8%
2	60.8%	59.7%	58.6%	56.4%	64.1%	69.0%	60.4%
3	67.1%	**69.7%**	64.0%	63.3%	69.2%	**72.3%**	**67.2%**

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 年版高齢社会白書（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. Spatio-temporal 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 skeleton-based 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 Intelligence*, 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`.