

## 〔論文〕

## 骨格検出技術を用いた知的障害児学習支援のための集中度推定法

春田 大河<sup>\*1</sup>・小田 まり子<sup>\*2\*3</sup>・石垣 大樹<sup>\*3\*4</sup>・新井 康平<sup>\*3</sup>

## Concentration Estimation Method Using Skeletal Detection Technology to Support Learning for Learners with Intellectual Disabilities

Taiga HARUTA<sup>\*1</sup>, Mariko ODA<sup>\*2\*3</sup>, Taiki ISHIGAKI<sup>\*3\*4</sup> and Kohei ARAI<sup>\*3</sup>

## Abstract

In individual learning (self-study) using e-learning in special-needs schools, learners tend to lose concentration when the instructor is not nearby. Thus, we are developing an e-learning system that alerts learners when the students lose concentration. In this study, we propose a method to support learning by detecting postural changes based on skeletal detection using real-time video images during e-learning and estimating and judging the learner's learning (focused state) and non-learning states (unfocused states). Using MediaPipe, a machine learning library that can detect the skeleton in real-time in a CPU environment, we could detect postures unsuitable for learning such as "standing up," "lying face down," and "turning to the side" as unfocused states based only on two-dimensional coordinate information at three points: the nose and both shoulders. Additionally, we confirmed that the unfocused states can be detected using MediaPipe, without using a special device or a high-performance computer.

**Key Words** :Artificial Intelligence, Skeletal Recognition, Emotion recognition, e-learning, mentoring

## 1. はじめに

内閣府の令和4年版障害者白書によれば、知的障害者の総数は109万4千人であり<sup>(1)</sup>、18歳未満の知的障害者の総数は22.5万人<sup>(1)</sup>に上る。近年、知的障害に対する認知度が高くなり、療育手帳取得者が増加したことから、2018年の知的障害者総数は2011年と比較して約34万人増加しており<sup>(2)</sup>、特別支援学校における教員不足が深刻化している。公立の特別支援学校教諭における特別支援学校教諭等免許状(知的障害教育)の保有状況を見ると、14.0%の教員が特別支援学校教諭等免許状(知的障害教育)を保有しておらず<sup>(3)</sup>、これを講師や助教諭に限定すると実に46.9%の教員が免許非保有<sup>(3)</sup>という状態である。しかしながら、本来、特別支援教育には通常学級とは異なる専門的スキルが必要であり、特別支援教育のあり方を専門に学んでいない教員にとって、障害児教育を担当することの心理的負担は大きい<sup>(4)</sup>。また、近年では身体障害と知的障害の重複障害児も増えている。特に、音声言語による表出が困難な児童生徒の場合はコミュニケーションの取り方にも専門性が求められるため、経験の浅い教員にとって、児童生徒の意思を汲み取ること、感情を把握することは容易でない。特別支援学校では、児童生徒の教育に加え、保護者対応のほか、体調の管理、食事の介助など介護の側面でもあらゆる役割をこなさねばならないため、教員の休憩時間の確保<sup>(5)</sup>も課題の一つとなっている。

<sup>\*1</sup> 電子情報システム工学専攻, <sup>\*2</sup> 情報ネットワーク工学科

<sup>\*3</sup> AI応用研究所 <sup>\*4</sup> 株式会社プロレド・パートナーズ  
令和4年12月8日受理

一方、コロナ禍において対面での教育活動が著しく制限されるなか、情報通信技術(ICT)を利用した学習である e ラーニングの普及が加速している。GIGA スクール構想<sup>(6)</sup>の実現により、特別支援学校においても一人一台端末の環境が整いつつある。そして、特別支援学校における新学習指導要領では、各教科の指導計画の作成に当たっての配慮事項として、障害種毎にコンピュータ等の ICT の活用に関する規定を示し、指導方法を工夫することや、指導の効果を高めることを求めている<sup>(7)</sup>。特別支援教育においても、ICT 機器の有効活用により、障害児特有の学びにくさやコミュニケーションの困難さを克服・支援し、児童生徒一人ひとりに適した教育ができれば、大きな学習効果が期待できる。しかし、e ラーニングによる個別学習(独習)の場合、学習者が常にモチベーションを高く保ちながら学習を継続し、自らの意志で学習目標を達成する必要がある、知的障害を持つ児童生徒にとって学習の継続は容易ではないと考えられる。実際に、特別支援学校での個別学習では、ICT 機器を用いて学ぶ児童生徒の横で教員が学習内容や ICT に関する教育支援を行うとともに、継続的で双方向のコミュニケーションにより学習者の情意面を支え、学習に集中するように支援する働きかけがなされている場合が多い。従って、ICT 機器を用いた e ラーニングを導入したとしても、知的障害児の学習においては教員の学習支援が必要であり、教員の負担軽減は期待できない。そして、教員の負担を増やさずに多種多様な子どもたちに向けた「誰一人取り残すことのない、公正に個別最適化された学び」<sup>(8)</sup>を実現させるためには、e ラーニングにおいても、障害児教育に慣れた指導教員(メンター)が行うような障害を考慮した学習支援機能を設ける必要があると考える。

我々は、指導教員の代わりに CG キャラクターである AI メンターが知的障害を持つ児童生徒の個別学習を支援する感情認識 AI メンタリング機能を有した e ラーニングシステムを開発している<sup>(8)</sup>。本システムでは、e ラーニングを用いた学習時における学習者の様子をカメラから取り組み、学習者の骨格検出や表情認識の結果から、学習者の集中度や感情を推定している。そして、AI メンターが学習者の集中力が途切れたタイミングで注意喚起の声かけや、学習者の感情を配慮した励ましの声かけを行うことにより、学習状態に導くとともに、e ラーニングの正答率、応答時間、学習者の集中度、感情状態から判断した最適な学習内容を選択し、学習者に提供する。

従来、初等教育学の研究では、授業中における児童生徒の集中度を調べるために様々な姿勢推定法が提案されてきたが、近年では人工知能の応用技術である骨格検出による姿勢推定が注目されている<sup>(9)(10)</sup>。この手法は、従来の児童生徒の身体にマーカーを着けて、加速度センサーなどの接触型センサーで集中度を推定する手法に比べ、児童生徒の負担が少なく、準備の手間と時間がかからない。例えば、OpenPose<sup>(11)</sup>はリアルタイムに複数人の関節を推定できるため、教育学研究における姿勢推定にも用いられている<sup>(10)</sup>が、OpenPose の場合、GPU の積まれた高性能な PC を用いなければリアルタイム動画の姿勢推定は難しい。

本稿では、特別支援学校での導入を視野に入れ、GPU を用いず CPU 環境において顔、手、姿勢のリアルタイム検出が行える機械学習ライブラリの MediaPipe<sup>(12)</sup>を用い、骨格検出によって得られた鼻と両肩の計 3 か所の座標のみから姿勢推定を行って、学習者の学習状態(集中状態)と非学習状態(非集中状態)を推定・判断し、学習を支援する手法について提案する。

## 2. 提案手法

### 2・1 AI による骨格検出

本研究では、e ラーニング学習者の集中度を推定する際に、PC に接続された Web カメラから取り込んだ学習時動画のリアルタイム骨格検出による姿勢推定結果を用いる。e ラーニングによる個別学習(独習)の際、指導者がそばに居なければ、緊張感が緩み、他に気を取られて学習が進まない場合がある。そこで、学習者の集中力が途切れたタイミングで注意喚起の声かけを行う目的で学習者の骨格検出を行う。学習時の行動を読み取るカメラは決まった位置に固定し、学習者の座席位置も固定することにすれば、カメラと正対して座った学習者の特定動作における座標や 2 点座標間の距離は、学習者によらず同様の値を取ると考えられる。従って、e ラーニングにおける応答時間が長い場合や不正解が続く場合に、学習者の身体、顔の各部位の相対座標を求め、学習者が「立ち上がる」、「横を向く」、「うつむく」、「上を向く」、「寝そべる」などの学習時にふさわしくない動作を行っていないかどうかを検出できる。そして、学習者の姿勢から集中度が低いと判断した場合には、学習者の動作に合わせて、しっかりと画面を見て学習するよう、注意を促すことができる。また、集中度が高いにも関わらず e ラーニングの正解率が低い場合には、提示する学習内容の難易度を変更したり、説明を加えたりなどし、集中度に変化がみられるかどうかを確認する。

本研究で用いる骨格検出モデルには Google 社が開発・提供している MediaPipe を用いる。MediaPipe は Yolo のように、骨格検出を行うための高性能 GPU を必要とせず、CPU のみで動かすことができる。MediaPipe では身体の両肩、首、腕、手首

等, 33 点の骨格点の二次元座標をリアルタイムに抽出できる. また, 骨格検出だけでなく, 目や鼻などの顔の特徴点(顔ランドマーク)もリアルタイムで検出できるので, 学習時における学習者の顔の表情を認識したり, 個人認証にも活用したりできる. 図 1 に MediaPipe を用いた骨格検出例を, 図 2 に顔ランドマーク検出例を示す.

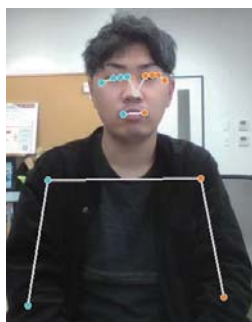


Fig.1 Example of Skeletal detection with MediaPipe



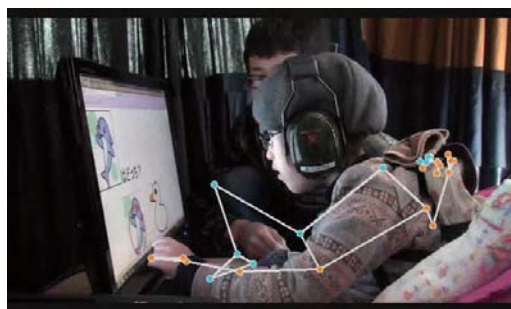
Fig.2 Example of facial landmark detection with MediaPipe

## 2・2 学習者の動作検出

学習時における児童生徒の動作・姿勢を把握するため, 以前久留米特別支援学校で ICT 機器による教育支援を行った際の学習効果確認用動画を OpenCV<sup>(13)</sup>に読み込ませ, MediaPipe で学習者の骨格検出を試みた. 具体的には対象動画を OpenCV でフレーム分割し, PNG 形式の連続した静止画として抽出した. その際, 全ての動画のフレームレートは 30 フレーム/秒(fps)だったので, 30 フレーム毎に骨格検出を行った. 図 3 に MediaPipe を用いた学習者の骨格検出の例を示す. 図 3(a)のように学習者を正面から撮影した学習時動画は, 集中度が推定できる精度で骨格検出ができた. しかし, 学習者を横方向から撮影した動画(図 3(b))や, 学習教材用タブレットを机の上において学習する様子を上から撮影した動画(図 3(c))では, 学習者の顔全体が映っておらず, 特徴点を検出できなかったり, 著しく骨格点の推定精度が悪かったりした. そのため, 過去に特別支援学校で撮影した動画像だけを用いて骨格推定し, 集中度・非集中度を分析・検出することは難しいと考えた.



(a) From the front



(b) From the side



(c) From above (with a tablet device)

Fig.3 Skeletal detection with student images during educational support in special needs schools

そこで, 特別支援学校で撮影した学習効果確認用動画を見て, 筆者と特別支援学校での教育支援経験者の計 4 名が合議の上で, 明らかに非集中状態と見受けられる動作として, 横(左右)を向く, うつ伏せになる, 立ち上がる, ものを投げるの 5 パターンの動作を選び, 研究を進めることにした. そして, 学習時の姿勢(骨格状態)を筆者が模倣し, その動作を顔や身体が映るように正面からカメラで撮影し, 骨格検出ならびに座標の抽出を試みた. 本研究では PC による e ラーニングを想定しており, 基本的に椅子に座り机の上の PC を操作する学習形態をとるので, 上半身のみの動きを推定できればよい. そのため本研究では鼻, 左肩, 右肩の 3 か所の座標のみに着目し, 骨格検出を行うことにした. 学習者は ICT 機器に正対した際, 正面を向いている状態を集中状態とし, 横(左, 右)を向く, うつ伏せになる, 立ち上がる, ものを投げる動作が見られた際に, 非集中状態として検出するために骨格検出を試みた. そして骨格の動きから図 4 の場合は集中状態, 図 5 の場合は非集中状態として検出した.





Fig.4 State of concentrating on studies



Fig.5 State of not concentrating on studies

### 2・3 座標データの検出

今回は座標データの出力を容易にするためにライブラリ化された MediaPipe を Anaconda<sup>(14)</sup>の仮想環境下にインポートし、Python で実行することにした。座標データは $x$ 座標、 $y$ 座標ともに、各々画像サイズの幅と高さによって0～1に正規化されて出力される<sup>(15)</sup>。そのため、実際の $x$ 座標、 $y$ 座標は MediaPipe で出力された座標 $x_m$ と $y_m$ 、カメラの解像度(単位: Pixel)  $p_x$ 、 $p_y$ を用いて、以下の式(1)、式(2)で算出することができる。

$$x = x_m \times p_x \quad (1)$$

$$y = y_m \times p_y \quad (2)$$

今回使用した PC の内蔵カメラの解像度は横 640 Pixel×縦 480 Pixel であるため、 $p_x=640$ 、 $p_y=480$  として検出を行った。また、リアルタイムでの検出において、フレームレートは 30 フレーム/秒であったため、撮影開始から 300 フレーム(= 10 秒)間で集中状態・非集中状態の再現を行い、鼻と両肩の座標推定を行った。なお、座標の(0, 0)は画像の左下(隅)の座標であることを意味し、数値が大きくなる毎に $x$ 座標なら右に、 $y$ 座標なら上に位置することを意味する。図 6 から図 11 に MediaPipe で骨格検出によって座標検出を試みた結果の例を示す。



Fig.6 Estimated in a concentrated state

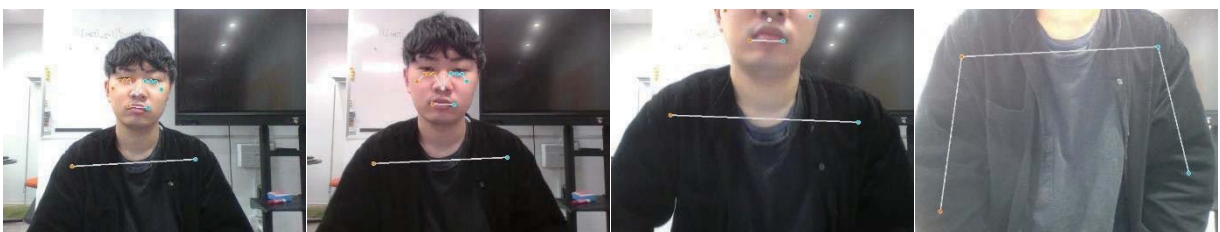


Fig.7 Estimated appearance in a standing position



Fig.8 Estimated in a prone lying position



Fig.9 Estimated state of throwing



Fig.10 Estimated view with left facing

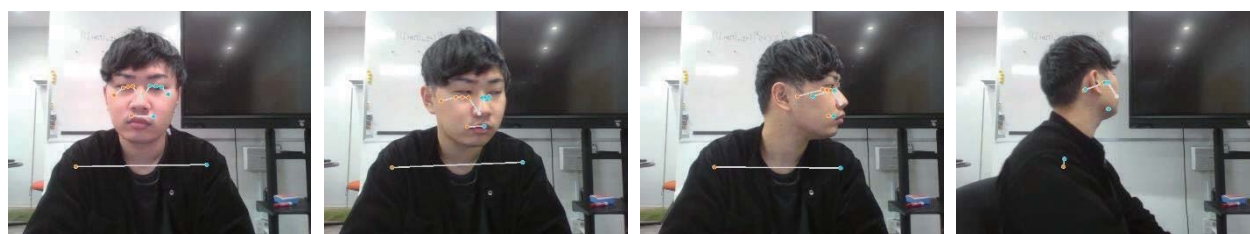


Fig.11 Estimated view with right facing

MediaPipe で検出認識した座標データの出力結果の保存については、Python プログラミングで鼻・両肩の座標データのみを抽出し、非学習状態の種類別に csv 形式で格納した。

### 3.骨格検出データによる集中度推定と考察

集中状態と非集中状態の種類毎に、鼻、両肩の $x$ 座標、 $y$ 座標がどのように変化するかを調べるために、ライブラリの Matplotlib を用いて、 $x$ 座標、 $y$ 座標のフレーム毎の変動を 3 次元時系列グラフにプロットした。

図 12 は集中状態(図 6)における鼻、両肩の座標の時系列変化を表したグラフである。集中状態では鼻と両肩はほとんど動かない(静止状態)ため、鼻、両肩ともに座標の時系列に大きな変化は見られない。

図 13 は、図 7 のように立ち上がった状態における鼻、両肩の座標の時系列変化を表したグラフである。図 13 のように、立ち上がった状態では左肩、右肩の $y$ 座標が座っていたときに比べ、最大で約 2.7 倍、両肩の座標位置が上昇し、鼻の $y$ 座標もフレーム No.94 を最後に推定できなくなった。学習者の身長には個人差があり、特に小学部の児童が立ち上がった場合、 $y$ 座標が大幅には変わらない可能性はあるものの、両肩の $y$ 座標位置の上昇ならびに、鼻の $y$ 座標が推定範囲から外れた場合は、立ち上がっており非集中状態だと判別できるであろう。

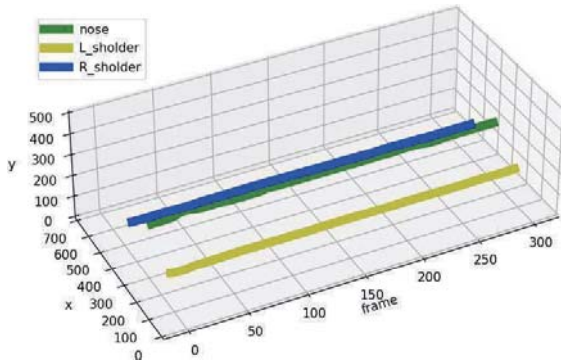


Fig.12 coordinate variation in a concentrated state

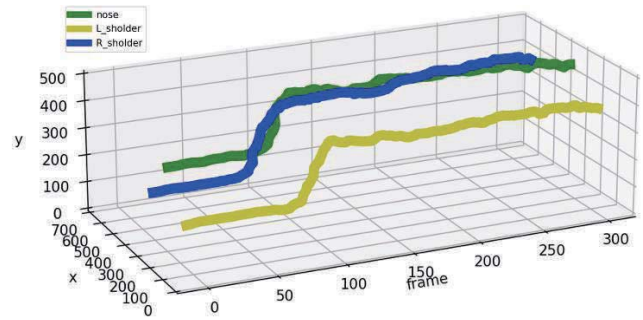


Fig.13 Coordinate variation in the standing state

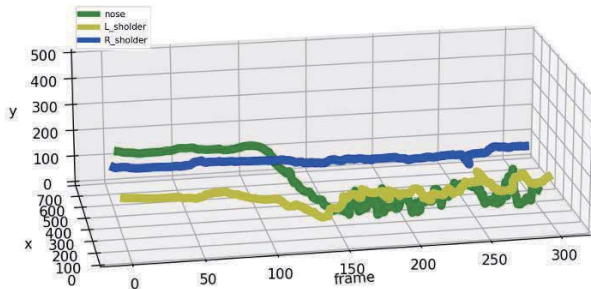


Fig.14 Coordinate variation in prone lying position

また図 14 は、机の上でうつ伏せになった状態における鼻、両肩の座標の時系列変化を表したグラフである。図 14 のように、うつ伏せの状態では頭部自体が下がっていくため鼻の $y$ 座標位置が下がり、フレーム No.138 の時点で鼻の $y$ 座標が両肩の $y$ 座標を下回った。表 1 にフレーム No.138 の各 $y$ 座標をまとめ、図 15 にフレーム No.138 の画像を示す。このような現象が長時間続いた場合、図 15 から学習者は PC の画面が見えないことが推定できるため、集中状態ではないと判断できる。

Table.1 Each  $y$ -coordinate of frame No.138

Frame Number	138
Nose_y	65.0873
Lsholder_y	71.8081
Rsholder_y	129.6702



Fig.15 Frame No.138 image

図 16 は、右手でものを投げようとしている状態における鼻、両肩の座標の時系列変化を表したグラフである。また、図 17 に、ものを投げる前の画像(フレーム No.1)と投げる瞬間の画像(フレーム No.120)の例を並べて示し、表 2 には、フレーム No.1 とフレーム No.120 の各座標の値を示す。

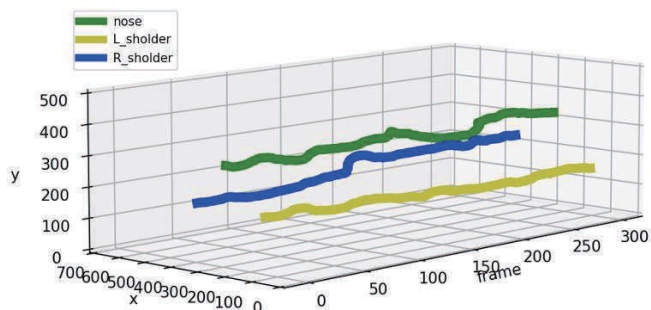


Fig.16 Coordinate variation in the state of throwing images



Fig.17 Comparison of frame No.1 and No.120



Table.2 Coordinate of frame No.1 and coordinate of the frame (frame No.120) at the moment of throwing

Frame_Number	1	120
Nose_x	287.7635	259.8120
Nose_y	309.6770	324.4851
Lshoulder_x	144.691	67.3369
Lshoulder_y	168.7424	168.8917
Rshoulder_x	395.1721	331.1885
Rshoulder_y	178.3609	217.0275

表2に示したフレームNo.1とフレームNo.120の座標値を比較すると、右肩のy座標(**Rshoulder\_y**)と左肩のy座標(**Lshoulder\_y**)の差、すなわちものを投げようとしている右肩のy座標と、逆である左肩のy座標の差が、フレームNo.1は約9.62であるのに対し、フレームNo.120では約48.14と約5倍の開きがあった。また、フレームNo.1は右肩のx座標(**Rshoulder\_x**)と鼻のx座標(**Nose\_x**)との差が約107.42であるのに対して、フレームNo.120では約71.38であり、ものを投げる動作の際、右肩のx座標と鼻のx座標との差が小さくなることが示された。これはものを投げる際、後ろに手を振りかぶる必要があるため、少なからず肩を通じて顔も投げる方に手引っ張られるからと考えられる。特別支援学校ではタッチペンを投げるような動作をする生徒が見られたが、このような非集中状態も両肩と鼻の座標位置の変化から検出できることが確認できた。

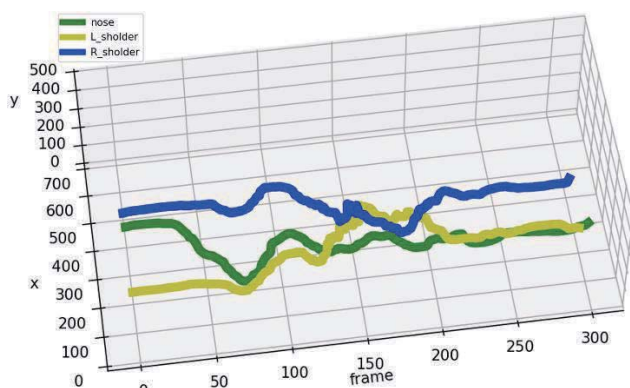


Fig.18 Coordinate variation with facing left

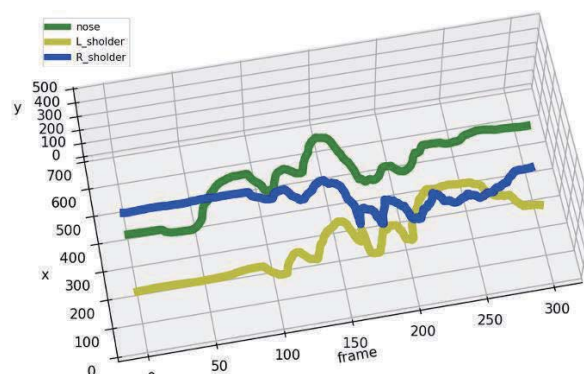


Fig.19 Coordinate variation with facing right

図18は顔が左に向いているときの、図19は顔が右に向いているときの、鼻、両肩座標の時系列変化を表したグラフである。図18、19ともに、最初に顔のみを横に向けると、鼻のx座標と横に向いた方の肩のx座標との差がだんだんと縮まっていき、やがて身体全体が横を向くと、鼻のx座標と横に向いた方の肩のx座標の位置関係が逆転している。図20に、顔を徐々に左に向けた時の画像例を示す。フレーム番号1の画像(図20(a))は撮影開始直後のものであり、フレーム番号50の画像(図20(b))は顔だけが左に向いたときの画像、そして、フレーム番号150の画像(図20(c))は身体全体が左を向いた状態の画像である。また、表3に、各フレームにおける鼻と両肩のx座標の値を示す。

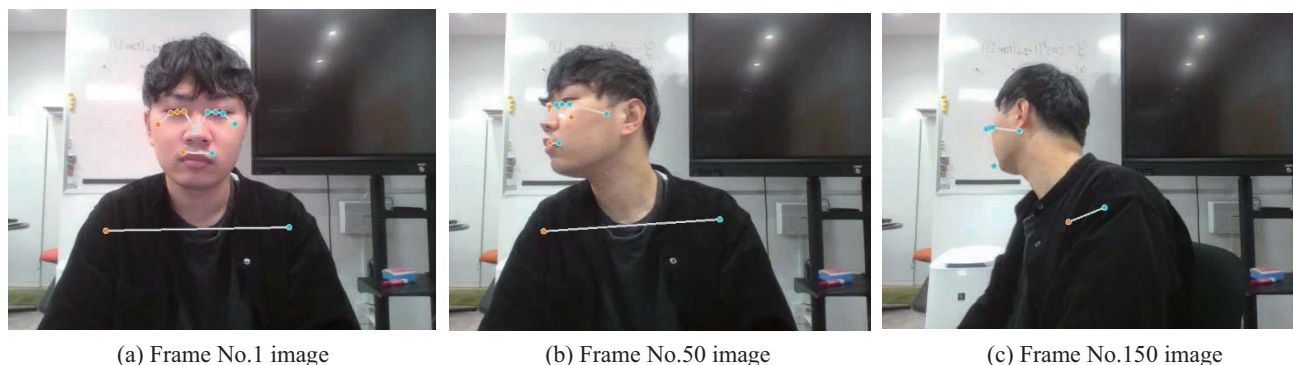


Fig.20 Comparison of frame No.1, No.50 and No.120 images(When the face is turned to the right)

Table.3 Coordinates of each frame with facing left

Frame_Number	1	50	150
Nose_x	281.2117	144.8510	146.8163
Lsholder_x	146.3724	144.8516	283.0860
Rsholder_x	420.6845	410.4799	338.0814

図 20 のフレーム 150 の画像では、左肩が完全に隠れてしまっており誤推定されてしまっている。しかし、正面を向いているときと比べ、鼻の $x$ 座標が左右に変動し鼻と肩の $x$ 座標の位置関係が長時間逆になっている場合には、非集中状態であると推定して差し支えないと考える。

最後に表 4 に以上の座標の位置関係をまとめて示す。

Table.4 Location of 2D coordinates for each skeletal state

骨格状態		座標の位置関係
集中状態		$Lsholder\_x < Nose\_x < Rsholder\_x$ $Lsholder\_y \approx Rsholder\_y < Nose\_y$
明らかな 非集中状態	立ち上がったとき	$Lsholder\_x < Nose\_x < Rsholder\_x$ $p_y < Nose\_y$
	うつ伏せになったとき	$Lsholder\_x < Nose\_x < Rsholder\_x$ $Nose\_y < Lsholder\_y \approx Rsholder\_y$
	ものを投げたとき	$ Lsholder\_x_0 - Rsholder\_x_0  <  Lsholder\_x - Rsholder\_x $ $L(R)sholder\_y < R(L)sholder\_y < Nose\_y$
	左を向いたとき	$Nose\_x \leq Lsholder\_x < Rsholder\_x$ $Lsholder\_y \approx Rsholder\_y < Nose\_y$
	右を向いたとき	$Lsholder\_x < Rsholder\_x \leq Nose\_x$ $Lsholder\_y \approx Rsholder\_y < Nose\_y$

ここで、 $p_y$ は 2-3 節と同じくカメラの縦解像度、つまり $y$ 座標がとりうる最大値を表しており、 $Lsholder\_x_0$ と $Rsholder\_x_0$ はそれぞれ左肩と右肩の初期 $x$ 座標を表している。

ただし、これらの式においてこの座標変動が微小な場合は集中状態か非集中状態かが不明瞭な可能性がある。そのため、座標の移動量に対する閾値を設け、学習状態か非学習状態かを判別する必要があるだろう。

#### 4. おわりに

本研究では、カメラから撮影した動画に対して、Google 社の提供する機械学習ライブラリーMediaPipe を用いた骨格検出(上半身・頭部動作、姿勢変化)を行い、得られた鼻や両肩の座標データに基づき、学習者の集中度をリアルタイムに推定する手法の提案を行った。MediaPipe を用いることで、特殊な装置や高性能なパソコンを用いなくても 3 か所の座標の情報のみで「立ち上がる」「うつ伏せになる」「横を向く」などの学習時にふさわしくない姿勢を検出し、非集中状態(非学習状態)として検知できることが確認できた。

今後は、学習者の集中度や感情に合わせた声かけを行う AI メンタリング機能を実現し、知的障害のある学習者の感情をより理解した助言や励ましができる e ラーニングシステムによる教育を実践していきたい。

#### 文 献

- (1) 内閣府，“令和 4 年度版障害者白書” (2022)，pp.215
- (2) 内閣府，“令和 4 年度版障害者白書” (2022)，pp.214



- (3) 文部科学省初等中等教育局特別支援教育課，“令和元年度特別支援学校教員の特別支援学校 教諭等免許状保有状況等調査結果の概要” (2022), [https://www.mext.go.jp/content/20220328-mxt\\_tokubetu01-000021585\\_02.pdf](https://www.mext.go.jp/content/20220328-mxt_tokubetu01-000021585_02.pdf)
- (4) 山口順也, 岩田吉生, “小中学校の特別支援学級の教員の精神健康度とストレス要因 –メンタルヘルスチェックの分析結果から–”, 愛知教育大学教職キャリアセンター紀要, Vol.2(2017), pp.33-40
- (5) 齋藤菜那, 江原寛昭, “特別支援学校の教員の教育的ニーズ”, 滋賀大学教育学部紀要, No.68(2019), pp.47-61
- (6) 文部科学省, “(リーフレット)GIGA スクール構想の実現へ”, [https://www.mext.go.jp/content/20200625-mxt\\_syoto01-000003278\\_1.pdf](https://www.mext.go.jp/content/20200625-mxt_syoto01-000003278_1.pdf)
- (7) 文部科学省, “特別支援教育における I C T の活用について”, [https://www.mext.go.jp/content/20200911-mxt\\_jogai01-000009772\\_18.pdf](https://www.mext.go.jp/content/20200911-mxt_jogai01-000009772_18.pdf)
- (8) 小田まり子, 呉 濟元, 八坂 亮祐, 河野 央, 高橋 雅仁, 新井 康平, “感情認識 AI メンタリングによる知的障害児の e ラーニングシステム”, 教育システム情報学会第 46 回全国大会(2021), pp.103-104
- (9) Shao Bochao, 金 尚泰, “深層学習を用いたユーザの生体情報認識に基づくリアルタイム集中度検出手法の検討”, 情報処理学会第 84 回全国大会(2022), No.6S47, pp.2\_451-452
- (10) 藤白 智也, 浅利 恭美, 宮田 真宏, 山田 徹志, 大森 隆司, “授業中の画像からの姿勢推定による集中度の評価”, 人工知能学会全国大会論文 集, 2020, JSAI2020 巻, 第 34 回 (2020), セッション ID 2C1-GS-12-03, p. 2C1GS1203, [https://doi.org/10.11517/pjsai.JSAI2020.0\\_2C1GS1203](https://doi.org/10.11517/pjsai.JSAI2020.0_2C1GS1203)
- (11) Z. Cao, T.Simon et al., "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", IEEE Transactions on Pattern Analysis and Machine Intelligence( 2019), pp.7291-7299
- (12) MediaPipe <https://mediapipe.dev/>
- (13) OpenCV <https://opencv.org/>
- (14) Anaconda <https://www.anaconda.com/>
- (15) MediaPipe on github <https://google.github.io/mediapipe/solutions/pose.html#output>