

# 授業中の児童の挙手行動の分類による学習意欲の推定

## Estimating Study Motivation by Classifying Hand Raising Behavior in Class

藤白 智也<sup>1</sup> 宮田 真宏<sup>2</sup> 山田 徹志<sup>3</sup> 大森 隆司<sup>2</sup>

Tomoya Fujishiro<sup>1</sup>, Masahiro Miyata<sup>2</sup>, Tetsuji Yamada<sup>3</sup>, and Takashi Omori<sup>2</sup>

<sup>1</sup> 玉川大学大学院工学研究科

<sup>1</sup> Graduate School of Engineering, Tamagawa University

<sup>2</sup> 玉川大学脳科学研究所

<sup>2</sup> Brain Science Institute, Tamagawa University

<sup>3</sup> 玉川大学教育学部

<sup>3</sup> College of Education, Tamagawa University

In order for the agent and the child to interact in the educational setting, it is necessary to estimate the child's willingness to learn. However, at present, such a method has not been established. To solve this problem, we have proposed a method for estimating children's interests from elementary school lesson videos. In this study, this know-how was applied to the raising hands of children in class. In this paper, we discuss a method for estimating the learning motivation and individuality of children for lessons from group and individual hand-raising information obtained from image recognition AI.

### 1. はじめに

教育現場では現在、文部科学省から個に応じた教育が推奨され、個々の子どもの個性を活かす教育の充実が求められている[1]。そこでは、子どもの評価指標として授業中の関心・意欲・態度を記録することが重要であるとされている。しかし、一人の教員が多数の子どもに向けて同時に授業する初等教育の現場では、授業に参加しているすべての子どもの状態を認識・支援することは困難である。この問題に対して、教育場面にエージェントを導入し、エージェントが個々の子どもの学びへの姿勢を推定し、それを教員にフィードバックすることで、教員の能力増強と負荷軽減することが考えられる。このようなエージェントの実現は質の高い教育の実現に有効であろう。しかし現時点で開発・販売されているエージェントはセンシング能力が低く、多数の子どもはおろか個人の学びへの姿勢の推定も難しく、教室単位での個々の子どもの把握は不可能である。

この問題に対して我々はロボット等のエージェント自身の対象人物のセンシングに加え、外部からの教室全体のセンシング・分析を組み合わせることが有効であると考え、これまで我々は、カメラを用いて授業を計測し、その映像に対して AI 技術を用

いて個々の子どもの姿勢を抽出し、その情報から個々の子どもの授業への参加状態を推定することで、個々の子どもの学習意欲の推定を目指してきた。これを踏まえて本研究では授業内の子どもの挙手行動に注目し、これを姿勢情報から推定して授業への参加状態を推定する可能性を検討する。そして、授業への参加状態を用いた個々の子どもの学習意欲の推定および教師との対話を含めて子どもの学びを支援するエージェントの可能性について議論する。

### 2. 教育現場の計測と子どもの姿勢の抽出

教育現場の広域かつ多人数での計測を実現するために本研究では、4K 解像度を持つカメラ 2 台からなるステレオカメラを作成し、8fps で同期計測が可能なセンシング装置を開発した[2]。これを玉川学園小学部の英語教室の前方と後方に各 1 台ずつ設置し、保護者から同意を得て 2020 年 11 月 27 日に実施された 1 年生の 2 クラス合計 20 名（以後、後半をクラス①、前半をクラス②）と 2021 年 6 月 29 日に実施された 2 年生の 1 クラス合計 10 名（クラス③）を対象に、授業中の子ども達の活動を映像記録した。この映像に対して人物の骨格推定アルゴリズムである OpenPose[3]を適用し、個々の子どもの骨格データを抽出した（図 1）。

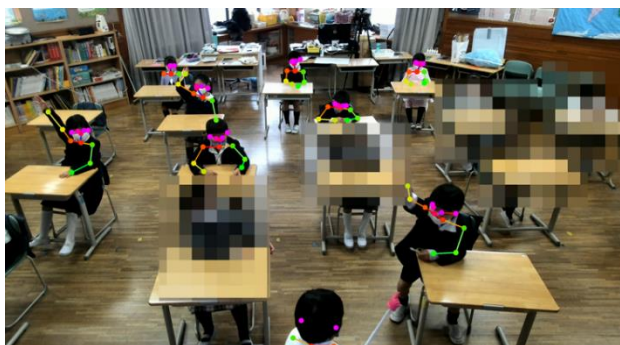


図 1. 教室での挙手場面の骨格抽出

### 3. 前処理：骨格情報の揺らぎの補正

OpenPose を用いた骨格検出には、人物と同等レベルの骨格推定精度が出る場面がある一方で、推定された骨格点の位置に揺らぎが生じるという問題もある。また、検出される骨格は各フレームで独立しており、個人の判別ができない。本研究ではこれに対して検出された骨格データのフレーム間での対応付けを行い信号処理手法のメジアンフィルタ (MF) [4]を用いて、骨格点の欠損の補間や突発的な外れ値などの揺らぎを除去した。

図 2 (a)は、補正前の骨格推定の位置を、図 2 (b)は補正後の骨格推定の位置をそれぞれ示している。図 2 (a)では挙手している子どもの手が検出できていないのに対して、図 2 (b)では肘、手首の位置など人が認識する位置と概ね一致する推定ができています。

### 4. 挙手行動による学びの状態の推定

#### 4-1. 骨格情報を用いた挙手場面の検出

本研究では、エージェントから教員へのフィードバック内容としては学習意欲が重要であると考えた。さらに、子どもたちの学習意欲は授業への参加状態として現れるとの仮説を立て、その参加行動に対応する動作として、授業中の挙手場面における姿勢に

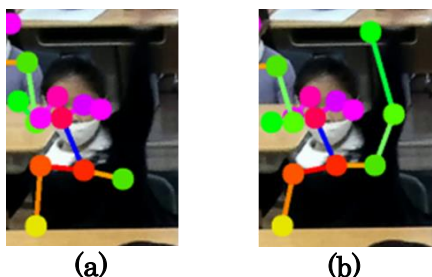


図 2. メジアンフィルタによる補正

着目した。本研究における挙手場面の定義は、教員が授業への習熟度を子どもたちに確認することを意図して挙手を指示した場面とした。骨格情報を用いた個々の子どもの挙手行動は、画像中の子どもの左右のいずれかの手首の骨格点の位置が左右の目の骨格点の位置よりも上にある場合とした。また集団としての挙手場面の検出は、挙手をしている子どもの人数が 3 人以上となっている範囲に対して±5 秒までの範囲とした。

これをクラス①～③のすべて範囲に適用し、挙手場面を検出した。クラス①の結果を図 3 に示す。図 3 の横軸は映像中の時刻 (分) を示し、縦軸は各時刻における挙手人数を示す。図中の赤枠は本研究手法により推定された挙手場面の内、本研究で定義した挙手場面の定義に当てはまる 6 場面を示している。図 3 の 5 分から 6 分半程度までの範囲に、最大で 8 人の子どもが挙手していると検出された場面があるが、この場面では教員が子どもたちの気を引くためにポンポンを使用している場面であり、本研究で考える教員が授業への習熟度を確認するための場面ではないとして分析対象場面から除外した。

抽出された挙手場面の例を図 4 に示す。図 4 の横軸は時刻 (秒) を、縦軸は各フレームで挙手と判断された子どもの人数を示す。この場面では、はじめの数秒の内に、3 名の子どもが挙手をし、その後 5 秒程度の間を空けた後にさらに 2 名程度の子どもが挙手をした場面であるということがわかる。実際の映像を確認したところ、教員より一人の子どもが指名されて前に出て他の子どもに曜日などを英語で質問する場面であった。はじめの 3 名は指名された子どもが教室の前に移動する段階から挙手しており、後の 2 名は質問が始まった段階で挙手していた。

自動抽出による場面の検出精度を確認するため、保育経験者 2 名による合議によりクラス①の授業内の挙手場面を選定し、どの子どもが挙手したかを 1 秒間隔で記述した (アノテーション)。ここで、自動抽出により分析した時間は 8fps であり、人が記述した 1 秒間隔とは時間のスケールが異なるが、本研究ではヒトの記述した瞬間の前後を含む 1 秒の間の 1 フレームでも検出されていれば挙手検出とした。

その後、クラス①での人手による記述で得られた 7 場面中の 6 場面と、本手法で検出された挙手場面が一致した 6 場面 (合計約 40 秒間) について混同行列を用いて評価した。本手法と一致しなかった場面は挙手時間が 2 秒程度と短く、また子どもが挙手をしていないと判断されたため、本手法では挙手場面として検出されなかった。本手法で画像から挙手と検出され、アノテーションでも挙手とされた場合を True Positive (TP)、本手法で挙手と検出され、ア

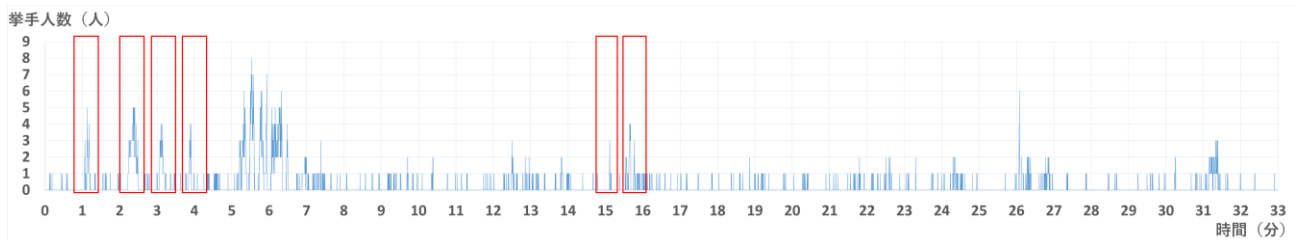


図 3. クラス①における挙手場面の推定例

ノテーションでは挙手ではないと判断された場合を False Positive (FP), 本手法で挙手の検出ができず, アノテーションでは挙手であるとされた場合を False Negative (FN), 本手法とアノテーションの双方で挙手ではないと判断された場合を True Negative (TN) とした.

対象とする 6 場面すべてを含めた場合, 自動検出のアノテーションに対する正解率((TP+TN)/全体)は 0.87, 適合率(TP/(TP+FP))は 0.84 であった. また, 6 場面のそれぞれについて求めると, 正解率, 適合率が最も高く得られた場面では, 正解率 0.90, 適合率 0.89, 逆に最も低い場面では, 正解率は 0.83, 適合率は 0 という結果が得られた. 表 1 に正解率, 適合率が高かった場面での検出の例を示す.

#### 4-2. 集団としての挙手場面の特徴抽出

挙手場面は, クラス①, クラス③ではそれぞれ 6 場面が検出され, クラス②では 7 場面が検出された. これらの授業では同じ教員が授業の導入・内容部分・終わり方など同一のパターンで授業をしており, 教員や子どもからの問いかけに対して子ども達が挙手する類似したパターンが多く見られた. それが今回検出された挙手場面である.

それらの挙手場面から, 各クラスの挙手傾向を分析するため, 各クラスの挙手場面の挙手の数を挙手が 3 人以上になったタイミングで重ねて平均を取った. 図 5 はクラス①, 図 6 はクラス②, 図 7 はクラス③の結果である. 平均を算出する際, 図 5 の 15 秒以降のように, 挙手の時間長の違いにより平均する場面が少なくなった箇所には 0 を入力した.

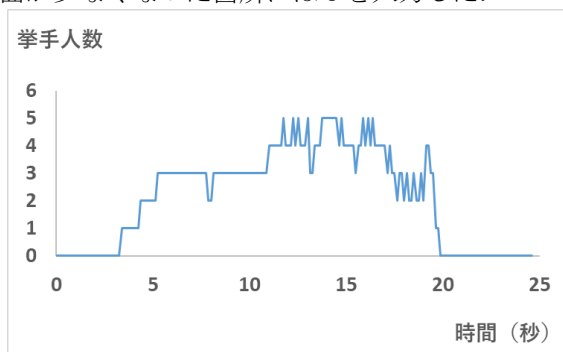


図 4. 挙手場面での挙手数の変化の例

#### 4-3. 個々人の挙手行動の抽出

続いてアノテーションにより得られた, クラス①の授業の前半にあたる 4 場面, 後半にあたる 3 場面を対象に, 個々の子どもの挙手について分析した. アノテーションにて集団の挙手場面の判断と同時に個々の子どもの挙手の状況についても記述し, その結果より得られた各場面での個々人の挙手時間の合計を図 8 に示す. 図 8 の縦軸に人物の ID, 横軸に挙手時間の合計を示す. 挙手場面全体における挙手時間の平均は約 10.3 秒であった. また授業の前半での挙手時間の平均は 8.2 秒, 授業の後半にあたる場面での平均挙手時間は 2.1 秒であった. この結果に対し, 本手法で検出された挙手場面全体における挙手時間の平均は 10.9 秒であった. 授業の前半での挙手場面における挙手時間の平均は 9.3 秒, 授業の後半にあたる場面の平均は 1.5 秒という結果が得られ, 挙手場面全体, 前半における挙手場面, 後半における挙手場面でのアノテーションで得られたものと本手法で得られた挙手時間の平均の誤差はそれぞれ, 0.6 秒, 1.1 秒, 0.6 秒という結果が得られた.

表 1. 正解率, 適合率が高かった挙手場面

時刻 (s)	子ども1	子ども3	子ども4	子ども5	子ども8	子ども9	子ども10	子ども13	子ども14	子ども15
0	TN	TP	TP	TN	TN	TP	FN	TN	TN	TN
1	TN	TP	TP	TN	TN	TP	FN	TN	TN	TN
2	TN	TP	TP	TN	TN	TP	TP	TN	TN	TN
3	TN	TP	TP	TN	TN	FN	TP	FP	TN	TN
4	TN	TP	TP	TN	TN	TP	TP	TN	TN	TN
5	TN	TP	TP	TN	TN	TP	TP	FP	TN	TN
6	TN	TP	TP	TN	TN	TP	FN	FP	TN	TN
7	TN	TP	TP	TN	TN	TP	TP	FP	TN	TN
8	TN	TP	TP	TN	TN	TP	FN	TN	TN	TN

#### 5. 考察

##### 5-1. 挙手場面の検出手法の妥当性

表 1 を見ると, 検出の精度は子ども毎に大きく異なっている. 映像で確認すると, 実際には挙手をしていた子どもでも, OpenPose で得られた骨格が実際の骨格位置と異なっていたことが原因であった. メジアンフィルタなどの事後の補正を用いたとしても, OpenPose で検出された骨格位置が大きく間違

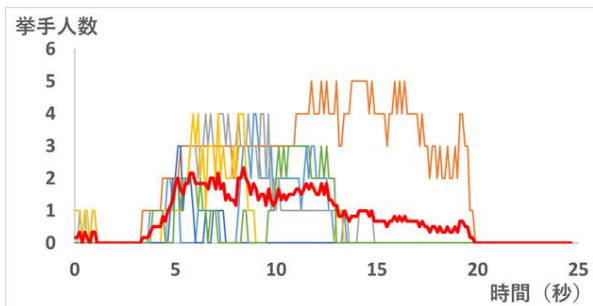


図 5. クラス①での挙手場面の平均

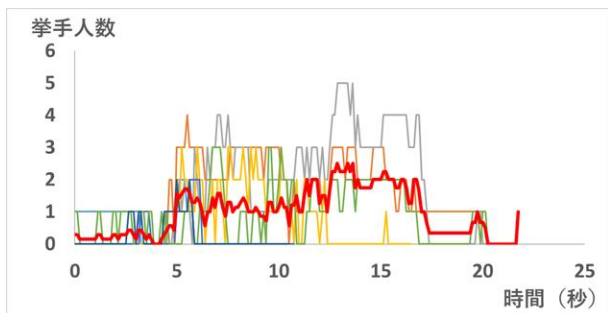


図 6. クラス②での挙手場面の平均

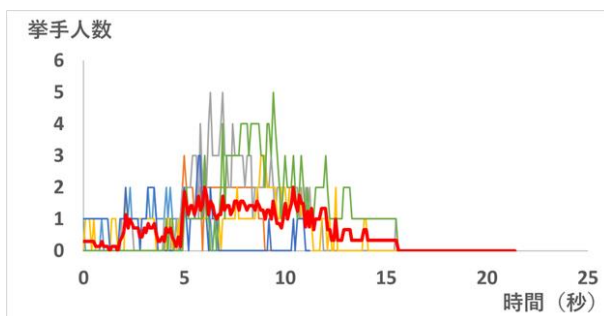


図 7. クラス③の挙手場面の平均

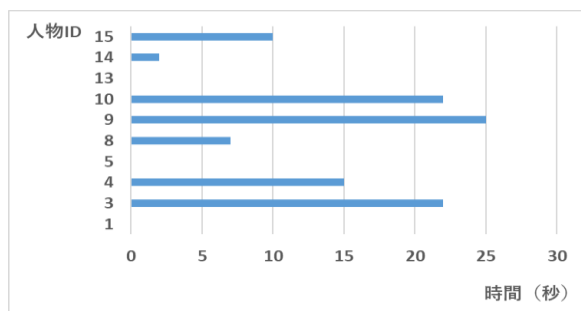


図 8. 挙手場面全体での個人の挙手時間

っている場合には、位置の完全な修正は難しいためである。OpenPose は複数の人物が重なっていたり背景が複雑な場合には、骨格の誤検出が比較的多い。しかし、手首と目の骨格情報のみを用いての挙手場面の検出は、検出された 6 場面の正答率・適合率を考えると、骨格検出の誤りの影響はあっても妥当であると考えられる。

## 5-2. 集団の挙手行動から判るクラス特性

11 月 27 日の 2 つのクラスを比較すると、クラス①では平均 2 人程度の子どもの挙手が続いているが (図 5), クラス②では 挙手の立ち上がり約 3 人の子どもが挙手をして、途中から挙手人数が下がるような傾向が得られた (図 6). 6 月 29 日のクラス③を見ると、挙手場面の開始から 5 秒付近までにすでに一部の生徒が挙手している、また、11 月 27 日の 2 クラスと比較しても、挙手場面の継続時間が短くなっている。これは、クラス①、②は 1 年生であったのに対して、クラス③は 2 年生であるため、学校の授業に慣れてきて質問の前に挙手するなどの無駄な行動は減少したためと考えられる。

また、クラス②、クラス③では、挙手場面の開始から数秒経過後に数人が挙手する場面や、挙手場面の中盤から挙手をする子どもや、挙手場面の後半に挙手をする子どもなどが見受けられた。実際に映像を確認すると、教員が質問をし終わるの待つように促す子どもがいる一方で、質問を言い終わる前から挙手をする子どもが確認されており、挙手数の平均は実際の子どもの行動を反映していると言える。

以上、全体の挙手場面の平均から、例えば以下のような仮説を考えることができる。クラス①では、平均的に約 2 名の子どもの挙手が続いていることから、序盤中盤、終盤のどのタイミングでも教員の質問の終了を待つことなく挙手をする傾向が推測できる。この 2 名が毎回同じ子どもであるなら、それはこの子ども達の何らかの特性が表われている可能性がある。一方のクラス②では、後半に挙手をする子どもの人数が多く、教員の質問を聞き終えてからの挙手が多い傾向と解釈できる。それに対して二年生のクラス③では、序盤に数人挙手をするが、途中から挙手する子どもが最も多く、先生の質問を最後まで聞いてから挙手をするという質問に対する標準的な答え方が身に付いた子どもが多いという仮説が生まれる。これらの仮説からはさらに、知識だけではない一年生と二年生の違い、すなわち授業における期待される行動の修得の程度の高さの違いである、という仮説が考えられる。

これらの仮説は実質 2 回だけの授業についてのケーススタディからの過剰な解釈である。これらの仮説を検証するにはより多くのデータが必要となるが、集団の挙手行動の定量性がこのような仮説の生成を可能にしたと考えられる。

### 5-3. 個々の子どもの授業参加と挙手行動

アノテーションによって得られたクラス①の授業内における挙手時間の平均は、前半は 8.2 秒、後半は 2.1 秒であった。また、本手法で得られた挙手場面では、アノテーションで得られた前半 4 場面は検出することができたが、後半では、アノテーションでは 3 場面に対し、2 場面しか検出できなかった。挙手時間の平均は、前半は 4 場面は 9.3 秒、後半 2 場面は 1.5 秒であった。すなわち、どちらも後半になるにつれ挙手場面の時間が短くなっていた。アノテーションでは検出されたが本手法では検出されなかった後半の 1 場面は後述する OpenPose による姿勢推定の限界が原因であった。一方で、アノテーションで得られた挙手をしている人数は前半、後半ともに 6 人と増減はなかったが、本手法では前半が 9 人、後半が 6 人と前半での挙手人数が多く検出されていた。これにより前半での挙手人数の平均の誤差に影響したと考えられる。実際の映像を確認すると、骨格の誤検出により、挙手をしていない子どもが挙手をしていたと判断されていた。アノテーションでの結果と本手法での結果を比較したところ、全体の傾向としては、どちらも個々人の挙手の時間は後半に短くなったことから、本手法の数値はクラスで起きている現象を反映していると考えられる。

一方で、図 8 中での 4 番の子どもは後半の 3 場面では挙手をせず、14 番の子どもは後半に挙手を始めていた。また、最も挙手時間が長かった 9 番の子どもは 7 場面の中で 6 回、挙手に参加していたなど、個々人の挙手行動の特性は異なっている。これらの行動特性と挙手場面の特性を組み合わせると、その行動の元となった個々人の学びに対する姿勢や学びの状況を推定できる可能性もあろう。

例えば、7 場面での子ども全員の挙手時間の平均に近かった 15 番の子どもは、実際には 7 場面中で参加は 2 回であった。この 2 場面は共に単語の意味を問う質問であった。これは、15 番の子どもは授業に興味が無くて不参加なのではなく、単語の意味を答える質問の時を選んで参加しているとも解釈できる。逆に、この子どもが参加しなかった別の例は、先生が子どもを一人指定し、その子が教室の前に出てクラスの子どもたちに英語で、その日の曜日を問いかける場面であった。指名された子どもは教室の前に出て英語で質問をするということもあり、この場面では子どもが目立つことになる。これより、15 番の子どもは単語の意味は答えられても、目立つような場面では挙手がしづらい性格、あるいは単語の知識はあっても文章の理解や英語での質問は苦手、という解釈もありうる。

これらはすべて限られたデータからの推測的な仮説である。これらを検証するには、一般には多くのデータが必要であり、クラスの継続的な計測により挙手行動の自動抽出が可能であれば、このような個々の子どもの個性の理解も実現できよう。一方で、これらの仮説は学習支援エージェントが個々の子どもと会話して反応を観察するための話題ともなりうる。この点については後述する。

### 5-4. OpenPose による姿勢推定の限界

OpenPose で得られた骨格情報を用いることで、クラスの挙手状況が定量化でき、さらに子どもの授業の参加傾向が抽出できた。しかし、子どもたちが映像内で重なった場合などでは、OpenPose による誤検出の発生は現状では避けられない。その補正のために用いたメジアンフィルタは、前後のフレームを参照しているため、複数のフレームをまたいで誤検出が起きた場合には、補完自体が行えない。この問題は現状の OpenPose を用いた姿勢推定の限界ともいえる。

OpenPose はまた、体を動かしたり、移動を行うようなアクティブな場面では映像にブレが生じるため適さない。そのため、幼稚園のような子どもが自由に動き回る環境では、本研究の手法はそのままでは適用が難しいと考えられる。OpenPose 以外の骨格情報を抽出する手法の検討や、OpenPose のアルゴリズムの改良が実現されることで、さらなる精度の向上および適用範囲の拡大が期待できよう。

また、挙手の検出についても現在は OpenPose で検出された両手首と両目の位置情報を利用している。そのため、OpenPose の誤検出の影響を直接的にうけてしまう。顔や目を画像から検出するアルゴリズムは多くあるが、手首を高精度で検出する方式は限られている。これも OpenPose の改良を期待される。

## 5- 教育におけるエージェントのあり方

本研究では授業中の子どもたちの骨格情報から挙手の姿勢を抽出し、授業中の挙手行動から挙手場面と個々人の挙手状況を検出した。これを自動である程度の精度で推定可能なエージェントが実現できるなら、教師の支援だけでなく、子どもの学習の質の向上にもつながることが期待できる。

現時点では以下のようなエージェントの動作が想定されよう。

1. センサシステムが各子どもの授業中の個々の活動への参加の程度などの情報を抽出する。
2. 授業終了後に、その日の各授業での挙手回数や

各科目での各種活動についての興味の度合いや、授業の内容の理解などの各個人の情報を、教師に報告する。

3. 教師は、エージェントからの報告と自身の認識をあわせて、各子どもの個性に合わせた評価および仮説を設定する。
4. エージェントは、教師に設定された仮説に基づき、子どもの授業理解、さらに子どもの状態や個性についての仮説を検証する対話を各子どもと行い、結果を教師に報告する。
5. 教師は、エージェントからの報告に基づき、各子どもの状態に応じた家庭学習課題の設定や、ホームルームでの指導、保護者への対応などを行っていく。また、自身の授業の方法について検討し、改善を行っていく。
6. エージェントは各子どもについての学びの履歴を、長期的な分析のためのデータベースに残す。

このように、本センシングシステムはエージェントの目として授業毎の子どもの状態の観察、教師への報告と振り返り支援、長期的な観察記録など、教室で教師の能力を拡張する機能を実現し、結果としてより質の高い教育を実現することを目指す。その鍵となるのが、センシングによる子どもの状態の推定である。

また、現在はカメラ2台からなる装置を用いているが、外装の工夫をすることで、子どもと接しやすくすることも必要であろう。これは、保護者が自身の子どもの機械により計測されると考え、それに嫌悪感を抱かれる方もいるという問題への対応として必要である。教育の現場のような、子どもとその保護者、教育者など様々な立場のステークホルダーが関わる場で人とエージェントがインタラクションをすることを想定するならば、できるだけ子どもたちや保護者の方に不安感を与えないエージェントの姿の検討が必要であろう。

## 6- まとめ

授業中の映像から子どもの骨格を抽出し、メディアンフィルタで骨格情報の平滑化および欠損値を補正し、授業場面の子どもの参加状態の1つとして、挙手行動の検出を行った。次いで、クラス全体の挙手の数から挙手場面の検出を試みた。クラスごとの挙手場面の平均を求めることで、クラスの傾向を検討し、さらに、個々の子どもの挙手行動から個々人の特性を推定する可能性を検討した。結果、本手法は教室で座って授業を受ける小学校の子どもの授業参加状態の定量化の可能性を示した。

本手法の特長は、従来は教員の主観によるしかなかった教室全体や個々の子どもの状態を客観的なデータに基づき定量化したことにある。定量化によりエージェントがそれを利用することが可能となり、エージェントによる教師の能力の拡張、さらに教師の指示による子どもとエージェントの対話のための仮説の生成などの可能性が示された。

ただ、今回の分析はごく限られたケーススタディに基づくものである。客観的な定量化の威力はデータの蓄積による仮説の検証に現れることを考えると、継続的な観察の必要性が高まってくる。

今後は、センシングシステムの安定稼働、計測の半自動化と分析ツールの充実による、負荷の低い形での継続的な観察の実現を行う必要がある。そして、現場の教師が必要とする情報の抽出手法の開発および、このシステムの現場での運用による実用性の評価が課題である。いずれは、5章で示したようなエージェントの実現を目指したい。

本研究は、科研費基盤研究 B19H01718 及びキヤノン財団からの研究支援により実施された。支援に感謝する。また本研究は、玉川大学研究倫理委員会及び保護者の承認のもと実施された。

## 参 考 文 献

- [1] 文部科学省 「4 「個に応じた指導」の一層の充実」：  
[https://www.mext.go.jp/b\\_menu/shingi/chukyo/c\\_hukyo0/toushin/attach/1399860.htm](https://www.mext.go.jp/b_menu/shingi/chukyo/c_hukyo0/toushin/attach/1399860.htm), 2018
- [2] 宮田 真宏, 山田 徹志, 大森 隆司: 複数台のカメラを用いた授業観察システムの提案, 教育工学会 2020 秋季全国大会, 2020
- [3] Z Cao et al: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, IEEE Trans. on PAMI, 2018.
- [4] 藤白智也 他: 授業画像からの姿勢推定による子どもの特性推定の試み, 日本教育工学会 2020 年秋季全国大会, 2020.