

顔情報を用いた子どもの関心対象の推定 ～教室の認知科学に向けたセンシング手法の検討～ Estimation of Child's Interest Target Using Face Information ～Trial of Field Sensing Toward Cognitive Science in Classroom～

宮田 真宏[†], 森下 雄介[‡], 山田 徹志[†],
高本 亮[‡], 今岡 仁^{‡†}, 大森 隆司[†],
Masahiro Miyata, Yusuke Morishita, Tetsuji Yamada,
Makoto Takamoto, Hitoshi Imaoka, Takashi Omori

[†]玉川大学, [‡]日本電気株式会社 バイオメトリクス研究所, ^{‡†}日本電気株式会社
Tamagawa University, NEC Biometrics Research Laboratories, NEC Corporation
m.miyata@lab.tamagawa.ac.jp

概要

心の状態の一つである関心を推定することは、学習者達の理解状態の評価指標となりうるため教員の重要な技能である。これまでに我々は人手アノテーションを用いて子どもの関心推定を試み、人の位置・向きから子どもの関心が推定可能となる結果を得てきた。今回、遠隔から人の視線を計測できたので、視線と関心・位置・顔向きなどとの関係を示すとともに、教室という広い空間を対象とした人の行動計測の現状について示し、教育実践フィールドにおける認知科学の可能性について議論する。

キーワード：関心推定, 視線検出, 教室の認知科学

1. はじめに

現在、教育現場では授業の経過や子どもたちの心身の成長状況を口頭や文書を用いて教員間で共有している。しかし、授業が連続して行われる教育現場において、授業毎に起こった多様な事象を担当教員が記憶・記録することは困難である。その為、各教員個が授業内容をビデオに録画するなどし、授業研究に用いていることが少なくない。しかし、この方法では日常の授業の計測は困難であり、適用の場が限られるという問題がある。

この問題は教育学研究においても同様である。教育学研究では、教育の質の向上が重要であるとされ、これに関する行動研究が進められている[1]。一方で、その多くは、教育場面をビデオカメラなどにより観察し、その内容を人が記述し教育の質を評価している。つまり、ケーススタディ主体の質的研究が有用な手法として採用されている。質的研究は、子どもの育ちを深く議論できる利点があるが、その記述には多大な時間や労力を要し、日々の授業で各子どもに起こった事象すべての記述は不可能であり、事例数に限界があるという教育現場と同じ問題が生じる。また、子どもの観察による心

の状態の読み取りは、教育者・研究者ごとに獲得してきた経験や感覚知に強く依存する。その為、読み取りの安定性・信頼性の担保は容易ではないことも課題である。

他方で近年、深層学習手法の確立にはじまり AI 技術は日々進歩している。特に、AI 技術を用いた画像認識は、人の能力に近い人物の情報推定を実現するものもある[2][3][4]。しかし、このような AI 技術を教育現場、または教育学研究に適用する方法は現時点で確立されていない[5][6]。

これらの状況を踏まえ、本研究ではこれまで、人の心の状態の一つである関心に着目し、関心を人の物理的行動量より推定する方法を検討してきた[7]。そこでは、教室のような広い空間でも関心は人の位置・向きを用いることで推定可能であることが示唆された。同時に関心推定の精度は身体(位置・向き)の情報よりも、頭の情報(顔、視線など)を用いることで、より高精度となることも報告した。しかし、従来の研究における人の位置・向き情報の取得は人手によるアノテーションに依存している。

物理量である人の位置・向き情報の取得については、AI 技術と組み合わせることで、推定可能とする研究もある[8][9]。これらの研究では、人の視線の検出は現在の AI 技術をもってしても容易ではないことから、人の身体または頭の向きに着目しており、人の視線を用いた分析はなされていない。

ところが、画像処理技術は日々飛躍的に進歩しており、遠隔からの視線検出もまた可能となってきた[10]。これにより、人の顔情報のうち画像中の顔の位置および顔・視線の向きの抽出が可能となった、この手法を従来から知られている人の頭の3次元的な位置を検出する方法と組み合わせるなら、これまで人手によるア

ノテーションに頼っていた人の関心とその対象の推定が自動化され、精度や記述頻度が格段に向上することが期待できる。

以上より本研究では、実際の教育現場（教室）での授業中の子どもたちの関心推定を題材に、ステレオ画像処理と遠隔視線検出を組み合わせた人物の位置・向き・視線検出のシステムを提案し、それが可能とする子ども達の関心・集中・発散などの教室の場の状態推定について説明する。さらに、教室の場の計測が可能とする教員の技能の「見える化」の可能性について検討する。そしてこれらを通して、AI技術による新しいセンシングが可能とする教室の認知科学に向けた本手法の在り方について議論する。

2. 位置・向きを用いた関心推定

我々は日々の生活において何かに関心を向ける際、その対象のある方向に身体や、顔、および視線を向け対象を確認する。さらに、その関心対象への関心の多寡に従い、対象物に向き続ける継続時間も変化すると考えられる。この仮説に基づき我々はこれまで、人の位置・向きを用いた関心推定手法を提案してきた[7] [11]。これらの研究ではまず、実際の幼稚園の教室にて撮像範囲の異なるカメラ4機を用いて子ども18名、大人4名が映る18分53秒の活動場面の映像を計測した。その後、この映像に対して保育経験が5年以上の保育経験者（アノテータ）3名の合議により、1秒ごとにそのフレームに写っているすべての子どもの位置と顔・体の向きを記述し、さらに、その子たちの関心対象とその位置を記述した。この結果より、対象とする子どもの位置と体あるいは顔の向きと関心対象への向きのなす角度が算出できる。この研究では、記録映像中のすべての事象・子どもを対象としてこの角度の確率分布を作成し、それに正規分布をフィッティングすることで、関心対象に関する確率分布を作成した（式1、角度は degree）。

$$f(x) = \left(\frac{1}{22.62}\right) e^{-\left(\frac{x^2}{2 \times 9.02^2}\right)} \quad (\text{式1})$$

式1に示す通り、子どもの顔の正面方向に対する関心対象の角度の標準偏差 σ は約9度であった。さらに、先に述べたように、人はある程度関心の強い対象は時間をかけて観察する。本研究ではこの関心の持続時間を考慮し、以下の式2で関心対象の尤度を計算した。

$$C_t = \operatorname{argmax}_i \left(\prod_{k=0}^3 f(i, t-k) \right) \quad (\text{式2})$$

式2では、式1にて推定された時刻 t の関心対象の確率をもとに時刻 $t-k$ から時刻 t までの間、対象 i に関心を向けている確率を計算し、その確率の最大となった対象の番号を算出している。

この式を用いて各子どもの関心を推定した結果を図1に示す。図1の横軸は映像中に登場する子どものIDであり、縦軸はその関心対象の推定精度を示している。ここでいう正答とはアノテータにより記述された関心対象を指す。各個人の青色のバーはその瞬間（1フレーム）のみを用いた際の推定精度を示し、同様に黄色、緑色、赤色のバーはそれぞれ1秒間連続から3秒間連続までの間の1秒ごとの関心推定の精度を示している。

また、図1の各バーの下側の色の薄い部分は、得られた関心対象の第一候補のみが正答と一致した場合の精度を示しており、上側のバーは関心対象の第三候補までに正答の関心対象が含まれていた場合の推定精度である。その結果、顔の向きを用いることで、第三候補までを用いることで瞬間的な推定精度は48.8%であったのに対し、3秒間連続で顔が関心対象に向いていた場合には推定精度が67%となった。現状、この精度は決して高いとは言えないが、教室で行われている活動などの他の文脈要因を考慮に加え、フィルタをかけることができるならばより高い精度となることが期待できる。ここで対象を子どもと限定したのは、一般的な成人は社会的な理由により、かならずしも顔を関心対象に向けるとは限らず、その場合にはこの手法では推定が困難であると判断したためである。

以上より、子どもの位置・向きを用いた関心推定は可能であると示唆された。さらに、子どもの位置と向きを自動的に推定することができるなら、少ない労力で長時間にわたり子ども全員の関心推定が可能になると期待できる。

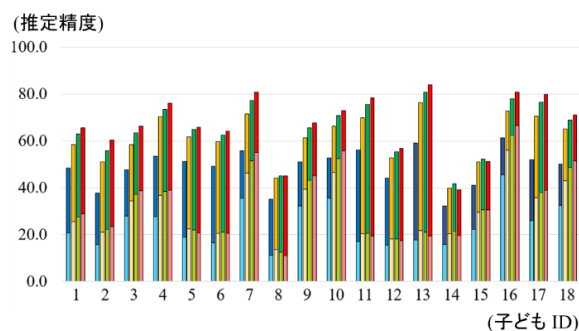


図1 子どもごとの関心推定の推定結果

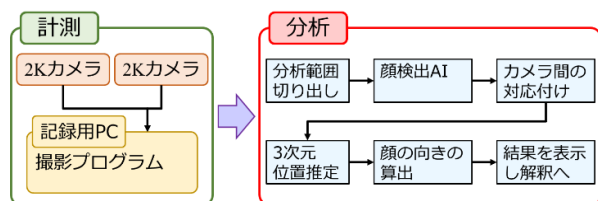


図2 子どもの半自動的な関心推定に向けた分析フロー



図3 子どもの顔認識の結果例

3. 行動計測の自動化の試み

3.1. センシングシステムの開発

人の位置・向きを推定するための手法として本研究では、ステレオカメラによる三角法を用いた。Kinect V2 に代表されるレーザー投射型の RGBD カメラを使用しなかった理由は、小学校の教室など奥行きが 10m を超える広範囲の計測にはレーザー投射は安定しないためである。一方、最近の AI 技術により画像中の同一人物の同一特徴点の位置が複数のカメラ間で同じ対象として検出可能となれば、ステレオ画像を用いた三角法により、広範囲かつ多人数の位置などが推定できる。本研究ではこれまでにこの方式のセンサ装置を試作し、2K の解像度のカメラ 2 台を視軸が平行になるよう左右に設置し、15fps で両カメラを同期して記録した (図 2)。

本稿で分析に用いた映像は、2019 年 10 月 10 日に玉川学園小学部にて行われた 1 年生 16 名が参加した英語の授業 (48 分 50 秒間) にて計測されたものである。教室内のカメラの設置位置は、教室正面にある黒板などの上部の授業時に教員の妨げにならない位置へ設置した。画角は着席時の子ども全員が正面方向から写る角度とした。なお、本計測データには個人情報が含まれるため、計測は事前に保護者に対して説明を行い、同意を得た上で実施している。

3.2. 子どもの位置・向きの推定

本研究ではこれまで、関心推定に用いる情報として、先行研究で身体の向きに比べ高い精度を示した顔の向きに着目し、分析に用いてきた[9]。その際、顔情報の認識には OpenFace[4]を用いてきた。OpenFace は人の顔の向き、および視線の向きを推定できることが特長である。一方、学習データが大人であるためか、対象を子どもにして視線の向きを推定した際に結果が不安定になる問題が生じており、分析に用いることはできなかった。この問題に対して本稿では、顔認識 AI として NEC の遠隔視線推定技術[10]を用いた。これを用い

ることで、これまでの顔情報に加えて子どもの視線を推定できるようになった (図 3)。図 3 中の青色、オレンジ色の縦横に同色の線が交わる位置は、それぞれ右目、左目の黒目の位置の推定結果を可視化したものである。黄緑色の線は左右の目の間隔を示す。図 3 中には図示していないが、実際には顔の向きや視線の向きも推定されている。さらに図 3 の子どもは図 4 上段の教室の最後列の一番左の席であり、図 4 中のカメラと子どもとの間の距離は約 7m 程度であるが、そのような条件下においても検出は概ね成功しており、さらに顔・視線の向きについても推定できている。

これらを踏まえ本研究の分析では、左右のカメラ映像の個々のフレームに対して本視線推定技術を用いて画像中の人物の顔を検出し、その特徴量を抽出した。しかし、左右のカメラ間の画角、誤認識などの影響により本視線推定技術より推定された人物は左右のカメラ間で異なる可能性がある。そのため、検出結果に対して左右のカメラ間で認識される位置のズレを考慮し、左右の画像間における人物の顔同士を対応付け、その後、三角法により対象の子どもの顔の 3 次元的位置を算出

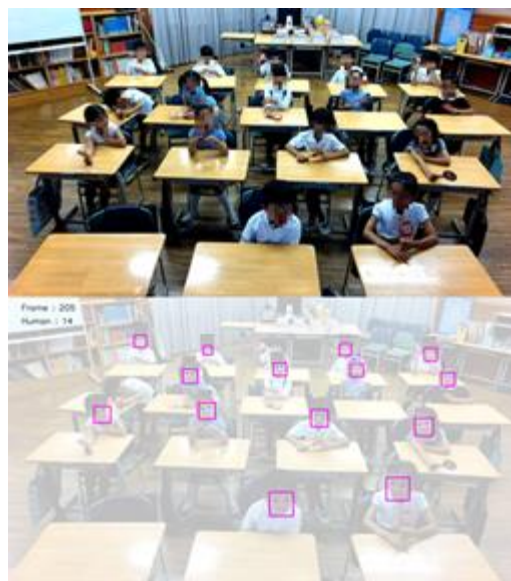
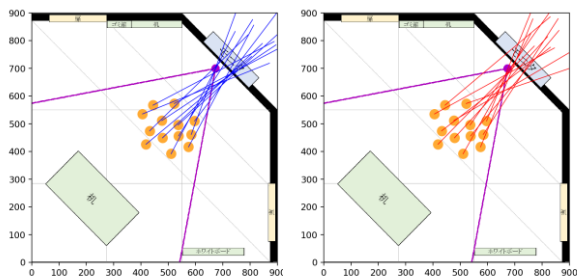
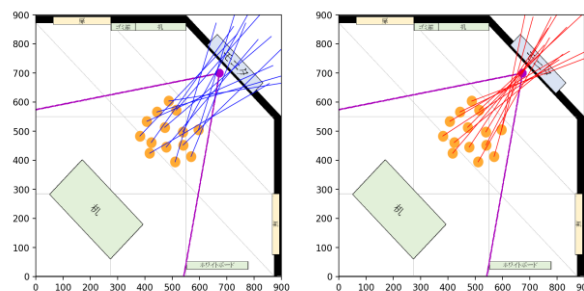


図4 計測場面と顔認識 AI の結果一例
上段：計測場面の例
下段：顔認識 AI により検出された顔、および目の位置の例



(a) (b)
図5 子どもたちの頭、視線の向きが一致した場面の例

- (a) 子どもたちの顔の向きが一点に集中
(b) 子どもたちの視線の向きが一点に集中



(a) (b)
図6 子どもたちの頭、視線の向きが不一致であった場面の例

- (a) 子どもたちの顔の向きが発散
(b) 子どもたちの視線の向きが一点に集中

した。そして、顔・視線の向きについては本視線推定技術にて抽出された顔・視線の角度を流用することで推定結果とした。本稿では、顔の向きに関しては3軸ある顔の向き検出情報の中でも対象人物の左右方向の回転角 (yaw 角) をその画像中の顔の向きとした。そして、視線の向きについても水平方向と垂直方向の角度が検出されるが、本稿では水平方向の角度を視線の向きとした。さらに、左右で対応する顔ではそれぞれに顔の向きが求まるが、本稿では両者の平均を対象の子どもの顔、および視線の向きとした。次節では、そのようにして求めた人の位置、顔の向き、および視線の向きの解釈について述べる。

4. 顔情報を用いた分析結果とその解釈

4.1. 顔情報の向きとその集合点との関係

図5は図4と同様の場面における各子どもの位置・向きの推定結果の一例である。図5のオレンジ色の丸は子どもの位置を示しており、紫色の丸は設置したカメラの位置を、紫色の線はカメラの画角 (yaw 角) をそれぞれ示している。そして、図5(a)(b)は同じ瞬間の子どもの顔の向き(a)、および視線の向き(b)を示している。また、図5(a)の青色の線は子どもの位置からの顔の向きを示しており、図5(b)の赤色の線は視線の向きを示している。

図4と図5(a)を比べると、推定された顔の位置は概ね正しく、相対的な位置関係は適切に示されている。図4中の左側の子どもの図5(a)での推定位置がずれているように見える。画像を確認したところ、これはレンズの歪みの影響と考えられ、その補正を行うことでさらに精度が高くなることが期待できる。

顔や視線の向きについては、この瞬間すべての子ど

もが映像中のやや左の方を見ているが、それが図5(b)でも検出できているなど、概ね正しく推定できていると考えられる。さらに、この顔の向きを示している青の線を伸ばした交点に着目すると、モニター的位置ではなく、その少し手前の位置に集中している。この場面は、教員が授業中に子ども達に次に行う活動を説明している瞬間であり、本分析では用いていないが、計測記録用の他のビデオカメラの映像を確認すると、この瞬間に教員はカメラのやや左にいた。これを考慮すると、子ども達はこの瞬間には教員の方を向いていると解釈できる。

また一方で、図5の場面から10秒程度時間が過ぎると顔と視線の向きが一致しない図6のような場面が現れた。図6(b)では、図5の結果と同様に視線の向きが一点に収束している一方で、図6(a)では図5の例とは異なり、顔の向きは概ね前正面を向いているが、何か特定の一点には収束せず、発散している。この瞬間は、教員から次の授業内容に関する説明が終わり、実際に説明された行動を行うまでの間の一瞬である。顔の向きの結果だけを確認すると多くの子どもたちはこの瞬間授業に集中していないという解釈となる。しかし、視線の結果では図5と同様に教員のいる方向に集中していた。この結果から、子どもたちが授業に参加しているかどうかの判断は顔の向きだけで判断することは難しいと言える。この結果より、同じ授業内においても集団活動において顔と視線の両方を教員の方向に向ける場面だけでなく、各自の作業内容に従った行動が必要になる場面も存在し、顔と視線の関係にもいくつかのパターンが存在すると思われる。この点に関してはより多様な場面の分析が必要であり、本稿ではこれ以上議論できない。

以上より、子ども達の顔・視線の向きを用いること

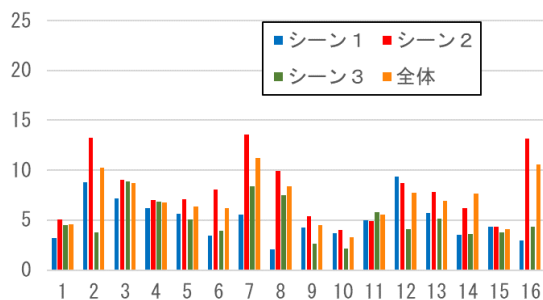


図7 各子どもの授業内参加時における活動毎の顔の向きの変化

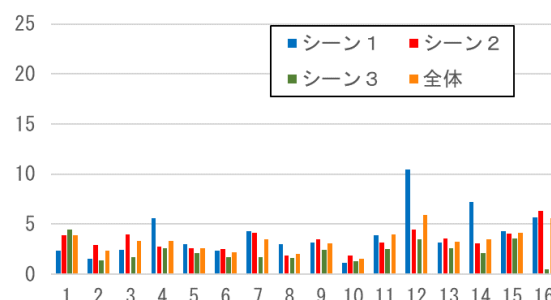


図8 各子どもの授業内参加時における活動毎の視線の向きの変化

で、集団活動における関心対象の推定ができる可能性がある。また本手法では、ある瞬間あるいは一定時間の間、教員を見てない子どもについても検出可能であり、その教育的な立場での解釈は今後の課題である。

4.2. 子ども達の顔・視線の向きのパターンと場面との関係

前節では集団の子どもの顔・視線の向きとその交点の収束・発散の情報を用いることで、授業の場で起こっている事象にはいくつかのパターンが存在することが示唆された。さらに一般的な座学の授業においてありうる授業中の子ども達の顔の向きとして、説明をしている先生や黒板などを見る、机に向かって各自の作業として教科書を読む、ノートを取る、何かを作成するなどといった、授業の活動として望まれる行動に対応した顔の向きのパターンが想定できる。そこで、これらの行動が授業内の活動とどのような関係にあるか、知る必要があると考えた。

そこで本稿ではまず、計測した英語の授業の内、子ども達の関心対象となり得る担当教員の移動が少なく、かつ複数の授業活動が含まれる1分10秒の対象場面を選定した。この場面には、(A) 教員が次に行う活動内容の説明をする場面、(B) 授業内の活動としてクイズをする場面、(C) クイズの内容を子ども達に思い出させる時間、の大きく3つの活動が含まれている。

分析に先立ち、この動画内の1秒ごとに、各子どもがその瞬間の活動に参加しているか否かを人手によりアノテーションした。また、分析には本分析場面において多くの場合、関心対象となり得る教員の位置を知り、そこからの差分の角度を用いて分析することが望まれる。しかし、今回の分析場面においてはステレオカメラを用いた計測では子どもたちを映すことを目的といたため、教員の位置のステレオ計測を用いた推定は行なわなかった。そこで、この場面においては多くの子ど

もが多くの時間、教員の方を向いていることから、各子どもの顔、および視線の方向の平均が関心対象(教員)となると仮定し、補正した。つまり、対象とする映像の時間区間中の各フレームで計測された個人の向き X_t に対して、子どもごとにその時間区間全体の平均を差し引き、さらに1秒ごとに人手でアノテーションしたフレームの時刻 F_T に頭の向きの計測値を対応させるため、 $F_T \pm 0.5$ 秒の範囲に含まれる X_t の平均 X_T を求めた。結果、各子どもについて、1秒ごとに個々の子どもが活動に参加/不参加を示すアノテーション $Y_T = \{\text{参加}, \text{不参加}\}$ と、その時刻に対応する頭の向き X_T が得られた。多くの子どもは活動内容に従い先生の方向を向いていると考えられるが、関心の対象や向き方によって X_T がゆらぐことが予想される。そこで、前述の映像区間中の各活動に対応する時間区間および映像区間全体についての X_T の標準偏差 σ_K ($K = \text{説明}, \text{クイズ}, \text{学習}, \text{全体}$) を顔・視線の向きに対してそれぞれ求めた。

図7に各活動区間 K の各子どもの顔の向き X_T の標準偏差 σ_K を示す。図7は横軸にこの英語クラスに参加していた子どものIDを、縦軸に顔の向きの標準偏差 (degree) を示す。子どもIDの1番や10番のように場面による顔の向きの変化が小さい子どもがいる一方、2番、7番、16番のように授業のシーンごとに顔の向きの変化が大きく変わる子どももいた。さらにアノテーションによりすべての活動に参加していると記述された子ども1番、5番、10番について見ると、1番、10番の子どもはすべての活動において顔の向きの動きが小さく、じっと先生の方向を向きながら授業に参加している。それに対して5番の子どもは顔の向きの標準偏差は子ども達全体を通してやや大きい傾向にある。このことから、授業に参加している子どもの行動パターンは必ずしも顔を教員や活動内容の方向に向けているわけではないことが判る。さらに、各活動への子ど

もの参加状況 Y_T と顔の向き標準偏差の関係を見ることで、各子どもの活動参加に関わる行動パターンを知ることができる可能性もあろう。

また、図8に図7と同様の各活動区間 K の各子どもの視線の向き X_T の標準偏差 σ_K を示す。それぞれの軸は図7同様に横軸にこの英語クラスに参加していた子どものIDを、縦軸に顔の向き標準偏差を示す。視線を用いた結果は12番の子どものシーン1では、視線にばらつきがあるよう結果となった。この時、12番の子どもは先生の話聞きながら自分の机を動かしており、その際の視線の変化が影響したのではないかと解釈できる。しかし、このような例外も若干存在するが、ほとんどの場合において、顔の向きとは異なり、授業への参加時の子どもの視線の向け方にはほとんど差がない結果となった。

以上より、授業に参加している際の顔の向きは授業内の活動により変化がみられるが、視線については今回分析した場面においては概ね一定であり、場面間の差異はないことを示唆する結果を得た。なお、今回示した結果は分析対象範囲が限定されており、視線の向きについて統計的な解釈をするには至っていない。この点については計測対象場面を増やすなどで対応が可能となるが、現時点では今後の課題とする。

5. フィールドの認知科学に向けて

前節までに高解像度のステレオ画像を用いることで、教室のような広い空間における人の位置、および顔・視線の向きを機械的な手段で定量化できることを示した。さらに、授業中の子どもの注視対象、ひいては関心対象の推定ができることが示唆された。これらより、本計測技術を実際の教育現場へ実装することで、子どもたちの行動及び教員の行動も併せて計測することが可能となろう。

教員の行動と子どもの行動と合わせて計測できることの意義は大きい。授業運営の方法は教育の質につながる重要な変数であり、教員経験の長い教員と短い教員（新人、若手など）とで計測・比較することができたならば、教員技能の定量的な見える化にもなり、さらに教員養成/研修においての利用も期待できる。また、実験室レベルではマルチモーダルな情報の取得が可能であるが[12]、実際のフィールドの広い空間では使える機器も限られ、映像からの情報を中心とした評価になる。従来、このような情報は人のアノテーションに依存

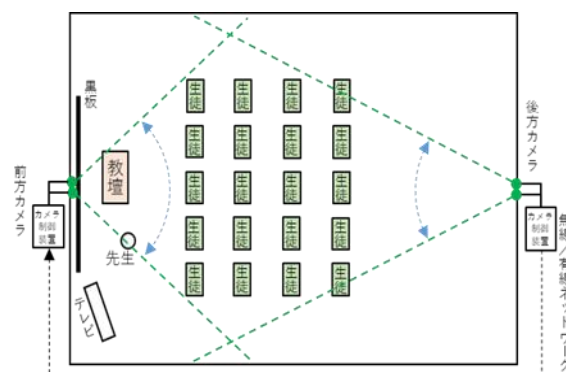


図9 教室の認知科学に向けた構想案

していたが、本研究のようなセンサシステムを用いることでその収集が自動化でき、多くのデータが少ない手間で得られるようになる。これは、教室のようなフィールドの認知科学を大きく変える可能性がある。

図9に、本研究において現在構想中の教室の認知科学に向けた計測システム案を示す。ここではステレオカメラのセットを教室内の前方・後方に一つずつ設置することを想定している。さらに、このカメラのセットは前後で同期して撮像から分析までが行われ、その結果を必要に応じて実時間で教員に返す、または記録し、後から詳細に分析することが想定される。本システム案は決して不可能なことではなく、実際に本研究にて用いた遠隔視線推定技術はリアルタイムで複数人の人の視線を検出できるとしている。そのため、教室の活動・状況として何を計測し、何を残すことが重要であるかが認知科学として解明されることで、実際にこのようなシステムの開発が進んでいくと期待する。

6. まとめ

本稿では、顔情報を用いた人の関心推定に向けて以下の3点についての結果を示した。

- ① 集団活動中の子ども達の顔および視線の向きを用いると、その瞬間にその教室で関心に向けておくべき対象を推定できる可能性がある。
- ② 授業における顔の向き・視線の向きの現れ方は場面ごとに異なる。仮に顔の向きがゆらいでも、視線の分析より子どもの授業への参加を評価できる可能性がある。
- ③ 授業への参加と顔・視線の向きとの関係は、顔の向きに関しては子どもごとに個性のようなものがあり、いくつかのパターンがあることが予想される。一方で、視線の向きについては比較的一

定である傾向がある。

なお本研究にて示した結果のうち、子どもの授業内活動シーンにおける参加・不参加時と顔の向きの標準偏差との関係については著者が自身の経験を基に解釈したものであり、実際の教育者との間では解釈が異なる可能性がある。しかし結果の数値自体は、センシングシステムから得られた情報と人手アノテーションとを突き合わせた際に得られた事実である。今後は、これらの結果を現場の教員にフィードバックして、教員からの評価を得る必要がある。さらに、これらの情報を日々の授業の振り返りの中に取り入れることで、授業中の子ども達の活動の変化を定量化できると期待される。

さらに本研究では教室の正面のカメラのみを用いて子どもを計測し、その行動について分析した。本手法では、子ども達の顔の向きから関心対象の位置を推定することはできるが、実際に何に関心を向けていたか、例えばそれが教員であった瞬間に教員が行っていた動作・活動との関係等は分析できない。そのため本稿では、教室の認知科学に向けた計測の一案として、教室後方にもカメラを設置し、教員の行動情報を子どもの行動情報とタイミングを合わせて計測するシステム案を示した。このような手法により教員の計測ができたならば、教員の位置だけでなく、例えば授業内で特定の子どもと向き合っている場面の分析など、実際に教員の取った教育活動の定量化もできよう。それができれば、例えば新人教員とベテランの教員、あるいは教員志望者の教育活動をそれぞれ計測し、それぞれの良い部分を定量化してフィードバックすることで教員研究や教員養成にも利用できると考える。

なお本研究では子どもの顔に関する情報のみに着目したが、身体の動きも人の心の状態を推定する有効な情報源であろう。これまで我々が身体の骨格情報の推定に用いていた OpenPose[2]は出力に揺らぎがあるため、人の3次元的な姿勢推定へ適用した際、その結果は不安定であった。しかし画像相関や信号処理手法を用いることでその安定化を可能にしようとする研究もある[13]。身体の向きについては顔の向きよりも関心推定の精度が低いとされている[7]が、これは人の身体の情報のうち向きのみを用いていることが原因であろう。人の身体の姿勢を安定的に推定できるならば、顔の向きのからの精度に相当する関心推定、さらには関心以外の多様な心の状態の推定が可能になると期待する。

また、センシングシステムについても課題は残る。本

稿にて提案したステレオセンシングシステムでは解像度が2Kのカメラを用いており、最大7m程度の距離での子どもの視線を検出可能であることを示した。しかし、教室のような広い空間での計測の場合、本研究で用いた解像度2Kのカメラでは7m程度離れた子どもの顔のサイズは30pixel程度とかなり小さく、人が見ても視線を推定することは容易ではなく、視線検出の精度に影響すると考えられる。そのため、本研究にて提案したセンシングシステムの内、カメラについては2Kだけでなく、さらに高解像度となる4Kのカメラを用い、結果を比較することにより、その精度を確認することが必要になろう。今後は、そのようなセンシング技術の開発も必要である。

本研究は、産業技術総合研究所人工知能研究センターからの委託研究および科研費19H01718の支援を受けて実施された。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。

文献

- [1] 秋田喜代美 他: 「保育プロセスの質」評価スケール, 明石書店, 2016.
- [2] Z Cao et al.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, IEEE Transactions on PAMI, 2018.
- [3] X. Zhou et al.: Monocap: Monocular human motion capture using a CNN coupled with a geometric prior., IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2018.
- [4] Baltrušaitis T., Zadeh A. et al.: OpenFace 2.0: Facial Behavior Analysis Toolkit, IEEE International Conf. on Automatic Face and Gesture Recognition, 2018
- [5] 本村, 村田, 大塚, 大森, 山田: 保育施設へのAI導入プロジェクトの課題と展望, [4L2-J-13-03]人工知能学会大会, 2019
- [6] 入倉由理子: プロの技を次世代につなぐ, リクルートワークス研究所, 2019
- [7] 山田徹志 他: AIにより子どもの発達・教育研究を支援する分析手法の検討 —子どもの位置・向き情報による関心の推定—, 日本教育工学会 第34回全国大会, 2018.
- [8] Miyata M., Yamada T., Asari Y. Aoyagi R., Omori T.: Mechanically Visualizing the Interest State of Children using AI, HICE2020, 2020
- [9] 宮田 他: 顔情報を用いた子ども達の関心対象推定手法の提案, 日本教育工学会 2020 春季全国大会, 2020.
- [10] Takamoto M., Morishita Y., Imaoka H.: An Efficient Method of Training Small Models for Regression Problems with Knowledge Distillation, arXiv:2002.12597v1, 2020
- [11] Yamada T., Miyata M., Asari Y. Aoyagi R., Omori T.: Visualization of Child's "Mind Growth" from Educational Images: Interest Estimation Using Position / Direction Information, HICE2020, 2020

- [12] 福田 他: マルチモーダル特徴量を用いた教育志望者の授業スキル評価モデルの提案, G-1, HAI シンポジウム 2018,2018
- [13] 浅利恭美 他: 子どもの動作からの関心推定に向けた行動計測と分析手法の提案, 日本教育工学会 2020 年春季全国大会, 2020.