

Classroom Space Sensing Using Facial Information : Approaches and Future Prospects

メタデータ	言語: jpn 出版者: 公開日: 2023-03-23 キーワード (Ja): キーワード (En): 作成者: 宮田, 真宏 メールアドレス: 所属:
URL	https://mu.repo.nii.ac.jp/records/2005

顔情報を用いた教室空間センシング の取り組みと今後の展望

Classroom Space Sensing Using Facial Information: Approaches and Future Prospects

宮田真宏

武蔵野大学 非常勤講師, 玉川大学 脳科学研究所

概要

近年の教育学分野では、講義に代表される受動的な学習だけでなく、グループワークに代表されるアクティブラーニングの要素を含む能動的な学習を授業に取り入れることが重要であるとされている。しかし、アクティブラーニング中に教員は教室内を動き回りながら各学生の状況を把握することが求められるため、個々の学生の状況を常に把握して評価することは不可能であった。一方で武蔵野大学では、「響学スパイラル」という新たな学びのカタチを提唱している。ここでは、学生の学びの効果を最大限に引き出すためには単に授業を改善するだけでなく、学修環境についても併せて検討することが必要であると考えられている。そこで本研究では、授業の様子をカメラを用いて計測し、取得した映像に人工知能技術を適用して分析し、その結果を評価することで授業と新しい学習環境の在り方について評価できると考えた。本稿ではその取り組みの一環として、現在までに得られた結果について報告すると共に今後の展望について議論する。

キーワード： 響学スパイラル, 行動センシング, 人工知能の教育的応用

1. はじめに

近年の人工知能 (AI) 技術の進化は目覚ましく、日々新しい技術が実社会に実装され話題となっている[1][2]。そのため、今後社会に出る学生にはこのような日々の変化に適応できるだけでなく、適宜自身の生活に取り入れることができる人材が求められるだろう。このような流れは教育学分野においても例外ではない。例えば、従来行われてきた教員が一方向的に講義する形式だけでなく、学生と対話をしながら理解を深くする形式、個人が自身の作業をする形式、複数人で1つの課題に取り組むグループワークをする形式など、新たな学びのカタチが取り入れられ始めている。しかし、現状この複雑化された授業活動の評価方法は解明されておらず、担当した教員の経験と勘によるものがほとんどであった。

一方で、AI 技術を教育現場に取り入れて教育に役立てることを目指した研究もある。例えば小竹原[3]は AI 技術を講義場面に適用し、受講生の姿勢をクラスタリングすることで

受講生の行動を分析した。ここでは分類した姿勢が授業中に受講生がとる行動と対応することを示したが、個々の受講生の行動特性や心的状態は考慮されていなかった。また、学習時の状態を推定するシステムを開発した研究もある[4]が、モーションセンサの装着が必要となることや、分類可能な活動状態が 5 種類と少なく、教育現場へのそのままの適用は難しい。教育現場への実装を考えると、受講生の活動や教育の邪魔にならない非装着型の計測手法が望ましい。これまでに教育現場を対象に、非装着型のシステムを提案した研究[5]もあるが、いずれも構想段階であり実現されたという報告はない。その原因には、計測機器の能力不足や分析手法が未確立な点などが挙げられ、教室のような広域かつ多人数を対象とした計測が困難であったことが考えられる。

また武蔵野大学では、2050年の社会で活躍できる学生を育てるための新たな学びのカタチとして「響学スパイラル」を提唱している[6]。この考え方は学びを「問い」から始まり、考動して、カタチにし、見つめ直すという4つのステップを繰り返しながら学習することで、成長していくという学びのスタイルのことである。このような学びのカタチを実現するためには、従来のような講義を中心として設計された学習環境での実現は難しいと言える。響学スパイラルの考えの元、その効果を最大限に引き出すには新しい学習環境についても検討することが重要である。既に武蔵野大学では、新しい学習環境のプロトタイプとして教室を用意し、検証が進められている。本稿ではこの響学スパイラルの考えに基づいて設計された教室を対象に行動計測と分析を行い、映像情報の分析による授業活動の分類と場面の状況が評価できる可能性について示すと共に今後の研究の方向性について議論する。

2. 行動計測・分析システム

2.1. 行動計測・分析システムの全体像

学習環境の評価するためにはまず、学修環境とそこでの学生たちの行動が可視化できるか確認することが必要であると考えた。そこで本研究では、まず教室内に映像記録用のカメラセンサを設置して授業の様子を計測し、記録された映像を分析・解釈することで教育環境の可視化を目指す。より具体的には、記録した授業映像に対してAI技術を用いて映像中の人物の特徴量を抽出する。現在のAI技術が映像情報から抽出可能な特徴量は、例えば対象の人物の頭や顔に関する情報[7]や骨格推定[8]など、その瞬間の対象の人物の物理的な行動に伴う情報（行動特徴量）である。一方でAI技術により直接得られる行動特徴量は、分析に重要な情報を直接抽出できているとは限らない。そのため、得られた行動特徴量を必要に応じて組み合わせることにより新たな特徴量を作り出した上で分析し、結果を解釈することが必要であると考えた。そして、得られた結果を教員にフィードバックすることで授業中の学生の理解度などの状況や、教員の試した授業デザインの効果検証などへの応用の可能性を検討する。

また、このような実際の教室を対象に授業の計測から分析、結果の創出までを実現するための行動計測システムには、現場の教員の業務に影響しないことやサイズなどによる設置

場所の制限が少ないことも必要である。一方で、本研究で行動分析に用いる AI 技術には膨大な計算が必要である。そのため分析に使用する PC には高い計算能力が求められるが、基本的には計算能力と筐体のサイズとの関係は比例関係にあるというのが現状である。このような状況に対して本研究では、教室を映像記録する行動計測システムと、記録映像を処理する分析システムを分離して構築した（図1）。これにより教室内に設置する機器を最低限にしつつ、AI 技術を適用するシステムの構築を実現した。

2.2. 行動計測システムの概要

本研究の行動計測システムに用いるカメラセンサには、Luxonis 社の OAK-D OpenCV DepthAI カメラ (PoE 版) を使用した。本カメラには中央にカラー画像のカメラが1台と、その左右にモノクロ画像のカメラが各1台の合計3つのカメラが内蔵されている。映像記録は、左右のモノクロカメラを HD 解像度 (1280×720) にして同期撮像し、得られた2つの画像をリアルタイムで統合して左右のカメラ間の視差画像を作成して動画化し、中央のカラー画像を 4K 解像度 (3840×2160) にして動画化してこれら2つの動画を 15fps で同期記録した。15fps は一般の動画と比較すると遅いが、本カメラの処理限界であることと先行研究[9]にて 8fps で記録した映像であっても人の心的状態の推定が可能であったことから問題ないと判断した。本研究ではこのカメラを教室の四隅に各1台設置し、これら4台のカメラからの映像を LAN ケーブルを経由して1台の記録用 PC に送信し、これらのカメラ間についても同期させて記録した。なお、本カメラを用いることにより行動計測システムより得られる映像データは、1時間当たり合計約 60GB となった。

2.3. 分析システムの概要

本研究で用いた分析システムでは前処理として、各カメラより取得された映像に NEC の遠隔視線推定技術 (顔認識 AI) [10]を適用して各フレームに映っている画像中の学生の顔の位置、および顔や視線の向きを取得した。本顔認識 AI を使用することで、カメラからの

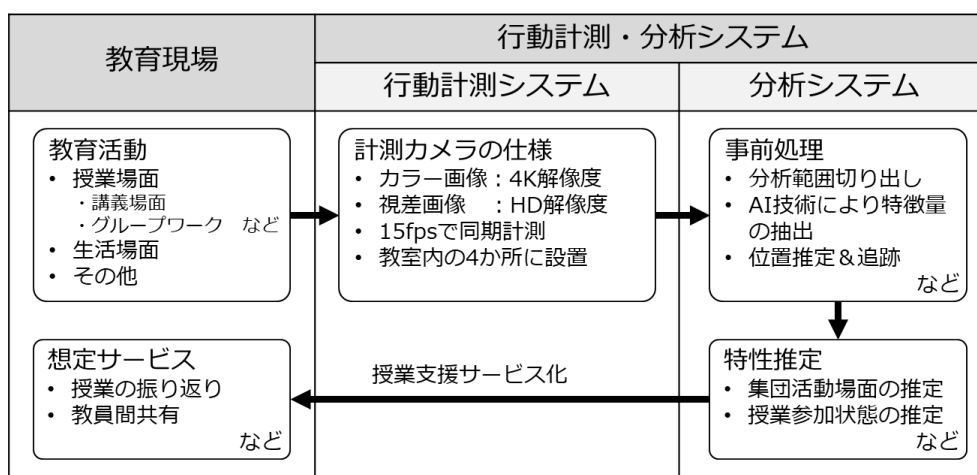


図1 行動計測・分析システムの概要

Figure 1 Overview of Behavior Measurement and Analysis System.

距離が 10m 程度である教室の壁側にいる教師や学生の顔の検出だけでなく、顔や視線の向きの推定も概ねできた。ただし、対象の人物が大きく横を向いていたり、他の人物や物体と重なっている場合などには検出できない。そのため、検出される人数や人物はフレーム毎に変化する。

そして学生が授業中にする行動の意味は、その行動を取った場面ごとに変化する。例えば、「他者の話を聞く」という行動は講義場面であれば、教員の話を理解しようとしているなど、ポジティブな要素として捉えることができるが、グループワーク時にこの行動を取り続けた際には、議論に参加できていないなど、ネガティブな要素として捉えられることもある。このように観察される学生の行動自体は同じであっても、その意味・解釈は場面ごとに同じになるとは限らない。このように考えると、授業全体を対象に学生たちの行動を分析・解釈するためには、先に分析対象となる場面を推定することが必要になる。本稿では授業場面を分析は、教室の四隅から中心方向に向けた各カメラの映像には多くの学生が注目する対象の方向や、授業内の活動場面についての情報が含まれていると考え、各カメラにて記録された映像における人物の顔の検出数を特徴量として採用した。さらには先行研究[9][11]にて有用性が示された顔の位置や向きの特徴量についても使用して分析を試みた。

ここで行動計測システムにて記録したカラー画像と視差画像の2つの映像は解像度と映っている画角が異なるという問題がある。この問題に対して本研究では、カラー画像の画角を基準とするために視差画像の画角をカメラパラメータに基づいてカラー画像の画角に変換した。そして、顔認識 AI にて抽出されたカラー画像中の x , y 座標をそれぞれ $1/3$ にすることで、変換後の視差画像の位置と対応させた。その上で、対象の人物の座標における視差情報を抽出し、三角法の原理に基づいてカメラ座標系における三次元的な位置座標を計算した。そして最後に教室空間における各カメラの位置や向き情報に基づいて回転行列を用いて変換することで、教室内の各学生の 3 次元的位置と顔や視線の向きの特徴量を得た。ここで顔の位置情報の計算には、顔認識 AI より抽出された特徴点の内、顔の中心に近似すると考えられる左右の黒目の位置の平均を用いた。なお、本研究で用いた顔認識 AI は視線の向きも推定可能であるが、教室空間のような広範囲を対象とする計測では画像中の顔のサイズが小さく、安定した抽出は困難であったため、分析には頭の向きを用いることとした。なお、頭の向きを用いて可視化する際には、安定した計測ができた頭の向き特徴量のうち、左右方向の回転角 (yaw 角) を用いた。

3. 計測場面

授業の計測は、開発した行動計測システムのカメラを武蔵野大学内のプロトタイプ教室に設置した。響学スパイラルに基づいて設計された授業を分析する際には、従来の教室の前を向きながら講義を聞くことや、近くの人とペアになってグループワークをするといった形式には捉われず、それぞれの活動の効果が最も期待できる場所や方法を用いて活動することが重要になると考えられる。そして、このような考えのもとに設計された教室を計測す

る際、先行研究[9]にて実施されたような教室の前方や後方からの計測を前提としてカメラを設置し、そこから学生や教員の様子を計測するという手法で記録・分析できるかについては疑問が残る。そこで本稿では、教室内の死角を最小限にすることで活動の場所を特定せずとも記録・分析が可能になると考え、各カメラの位置を教室の四隅とし、画角を教室の中央方向とすることで教室内の広範囲を記録できるようにした（図2）。なお、本稿で示す結果は2022年11月22日に実施された武蔵野大学建築学部の授業（100分間）である。この授業には受講生10名の他に、担当教員1名、教員補助1名、記録者1名が参加した。なお計測にあたっては、事前に授業に参加した学生に対して本研究の実施、および記録と分析の意図を説明し、同意を得たうえで実施した。

4. 行動分析

4.1. 顔の検出数を用いた場面分類の可能性

本研究では、授業中の教員が教室空間を有効活用できたならば、各活動の中心となる教室内の場所は活動ごとに異なり、さらに活動ごとに何らかのパターンがあると考えた。そこで、教室内に設置したカメラの記録映像の各フレームに対して顔認識AIを適用し、各カメラの顔の検出数を抽出した。また授業活動の分類は、授業内で起こった瞬間的なバイアスを無視し、時間的により長くかつ大きな変化を推定することが有用であるとする。そこで、抽出した顔の検出数の1分間（900フレーム分）のデータに対して時系列方向に移動平均をかけてグラフ化した（図3）。図3の横軸は計測した授業内の時刻（分）を、縦軸は各カメラで検出された人数をそれぞれ示す。図中の青色、緑色、赤色、橙色の線はそれぞれCamera1～4における検出人数を示す。図3の全体より、授業の時間経過と共に各カメラの検出人数は変化していた。例えば授業開始から8分程度では、教員はCamera1, 4の間に立ち授業の出席や当日の流れについて説明していた。この場面のデータではCamera1, 4の検出人数

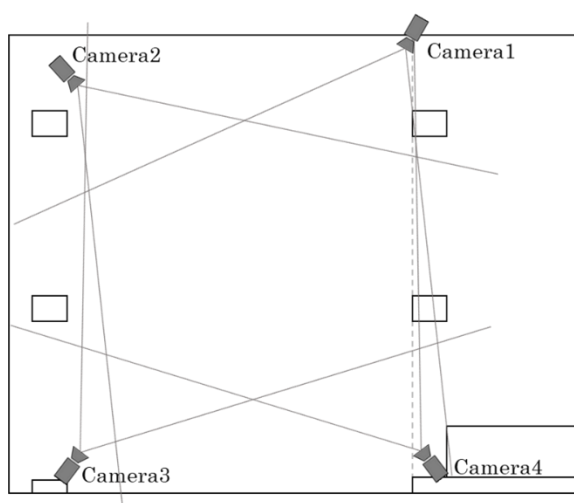


図2 計測した教室のレイアウト、および各カメラの設置位置

Figure 2 Layout of the classroom and location of each camera.

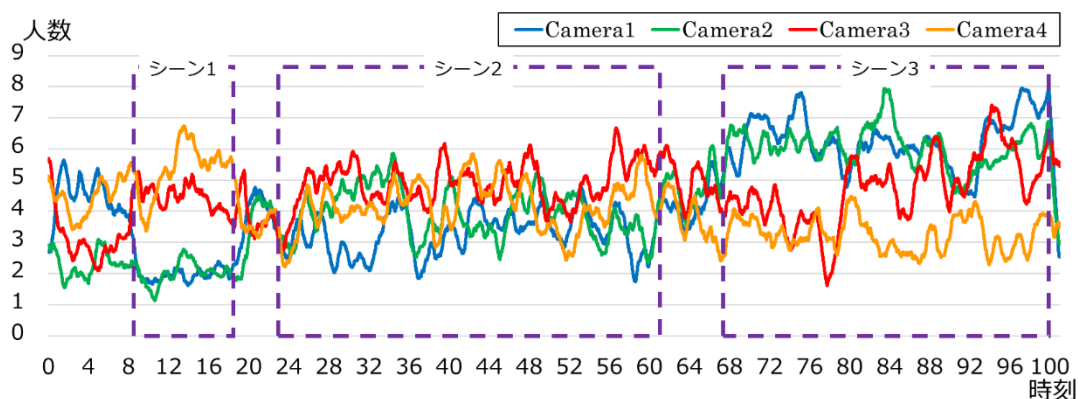


図3 各カメラの授業全体における顔の検出数の変化

Figure 3 Change in the number of faces detected in class for each camera.

が多く、他の2台のカメラの検出人数が少なかった。つまり、学生たちは教員の方向を向きながら話を聞いていたと解釈できる。紫色の破線で囲ったシーン1では教員はCamera3, 4の間でスクリーンを使って学生に作業内容を説明していた。この部分のデータよりCamera3, 4にて多くの人を検出し、他のカメラでは検出人数が少ない傾向があった。このことから実際の場面と同様に学生は教員の方向を向いていることがわかる。シーン2では、学生は教員に指示された自身のやるべき作業をし、教員は教室内を動き回りながら状況を把握していた。シーン2のデータより、このシーンにおけるカメラ間の検出人数には顕著な差はなく、概ねバランスよく検出されていた。このことから各学生は自身の作業しやすい方向を向きながら作業していた可能性がある。そしてシーン3は、Camera1, 2側の壁の前で各自がその日に作業し、作成した成果物をクラス内で発表し、それについて議論するという活動であった。このシーンのデータよりCamera1, 2にて多くの人が検出されており、この場面についてもシーン1のように多くの学生が発表者の方向を向きながら発表を聞いていることがわかる。一方でこのシーンの後半ではCamera3の検出人数も増えているが、これは発表が終わった学生が自身の席に戻った際に、やや横向きに座ったために顔が検出され、それが影響していると考えられる。

以上より、この授業では準備の場面を除くと大きく3つのシーンがあったが、それぞれの活動場面の特徴をそれぞれのカメラの顔の検出数を用いることで可視化できた。なお、この授業にはここで3つのシーンだけでなく、シーン間にカメラの検出数のパターンが類似した箇所(20~22分など)も見て取れるが、ここではシーン間の移動や教員が次の活動の導入を話し始めるなど、大きな場面が変化する前の過渡期であると考えられる。

4.2. 頭の向きの交点による授業状態の可視化

本研究ではこれまでに授業中の状態の可視化には、前処理として行った顔認識AIの結果と視差情報とを組み合わせて推定した教室中の人物の3次元的位置と、その位置から頭の向き方向に半直線を伸ばして用いてきた。そして、その直線同士が交わる交点情報をすべての人物間で算出(正面方向交点群)することで、集団活動の重心位置についても可視化してきた[9][12]。

分析対象場面の各瞬間の正面方向交点群には、顔の検出数が少ない瞬間や、多くの人物とは異なる方向を向いている人物の交点情報も含まれており、それらが多数の人物が注目している対象を推定する際のノイズとなっていた。そこで本研究では、分析対象の瞬間の前後0.5秒間で得られた正面方向交点群を対象に、交点群のx方向とy方向それぞれの平均と標準偏差 σ を求め、x方向とy方向のいずれか1つでも平均 $\pm 1.5\sigma$ を超えた交点はノイズであるとして取り除いた。その後、残った正面方向交点群に対して再度平均と標準偏差を算出し、その瞬間の重心位置とその広がりとした。なお、このノイズの判定範囲は映像中の人物の位置や頭の向きから、ノイズを判定できる交点を効果的に除ける範囲として経験的に定めた。これを各シーンに適用することで可視化を試みた(図4~6)。

図4~6はシーン1~3内のある瞬間の可視化結果である。図の左側は記録された画像に対して顔認識AIによる顔の検出結果を重ねて可視化したものである。そして図の右側はこの瞬間の各人物の3次元的な位置を橙色の丸で、頭の向きを青色の半直線で示した。図中にある十字(+)はノイズ除去後の正面方向交点群である。図中の黒色の楕円では、交点群の重心位置を中心に、x方向とy方向のそれぞれ $\sigma=1.5$ の範囲の広がりを示した。

図4の左側の画像より、この瞬間教員はプロジェクターに授業内容を映しながら内容を説明しており、学生は教員の方向を向きながらその説明を聞いていることがわかる。図4右側のデータによる可視化結果より、楕円はCamera3と4の間にいる教員の位置が中心となり、さらにやや説明内容側に視線の交点が集まっていた。この結果より、場面内の多くの人物が教員の方向を向きながら授業を聞いていたことがデータからも再現できたと言える。

また図5の左側の画像より、教員は教室を見回っており、学生はそれぞれの作業したい場所で作業している。この瞬間の図5右側の可視化結果より、人物の位置は教室の広範囲に広がり、さらには様々な方向を向いていたことが再現できている。そして、集団活動の近似であることを期待する楕円の中心は教室の中心付近にあり、さらにその広がりも教室の広範囲に広がっていた。これは、学生が教室空間を自由に広く使用できていることを示唆する結果であると解釈できる。

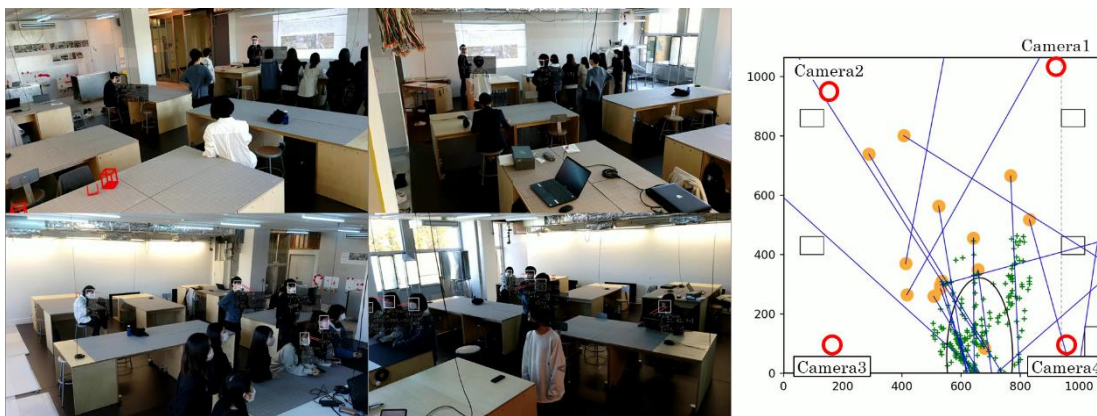


図4 シーン1における画像と可視化結果の例

Figure 4 Examples of images and visualization results in scene 1.



図5 シーン2における画像と可視化結果の例

Figure 5 Examples of images and visualization results in scene 2.

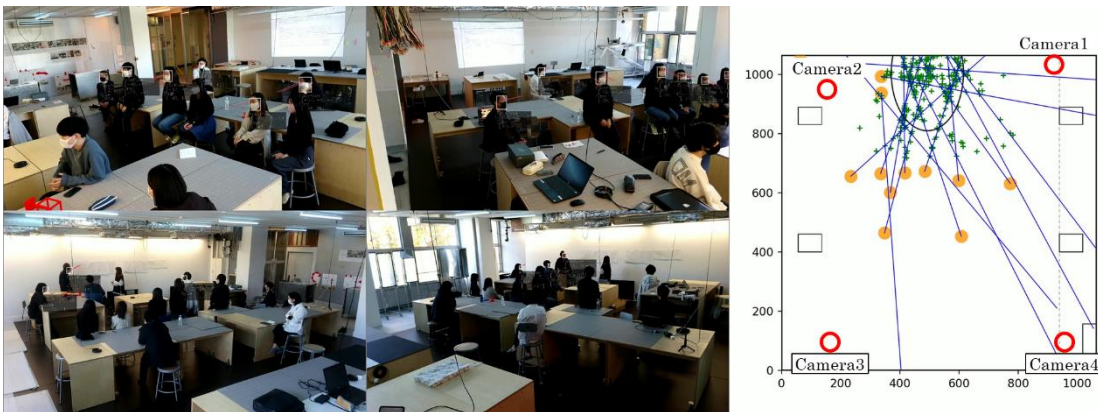


図6 シーン3における画像と可視化結果の例

Figure 6 Examples of images and visualization results in scene 3.

図6の左側の画像よりこの瞬間は図4と類似した場面ではあるが、教員と発表中の学生は成果物を見ながら議論しており、他の学生もそれを見ている。図6右側の可視化結果より、楕円の重心は実際に成果物の位置あたりになっており、多くの学生がその成果物の方向を向きながら議論を聴講、または参加していることがわかる。

一方で、図4～6の人物の可視化結果は必ずしも左側の人数とは一致していない。これは、本稿で示した手法にはカメラ間で同時に同一人物が認識された際に、その人物を一意にするための顔認証などを用いた処理が入っていないためである。さらに推定された人物の位置も必ずしも正確ではない。これは、カラー画像と視差画像の画角の対応付けに誤差があることが考えられることと、視差画像の解像度が低いことにより、距離が遠くなるほどに精度が落ちることが原因であると考えられる。これらの問題点への対応は今後の課題である。しかし本稿で示したように、これらのノイズがデータの含まれていたとしても本稿で示した分析を用いることで授業中の解釈に役立つ情報が取得できるため、有用性は十分にあると考える。

5. おわりに

本稿では、武蔵野大学の提唱する「響学スパイラル」の考えに基づいて作られた教室環境の評価をする際に、従来のような狭い空間かつ人数を絞った計測・分析では実現できないと考え、教室に複数台のカメラを設置することで広域空間かつ多人数を対象とした計測と分析を実現し、その試みを報告した。本研究にて得られたものは以下の通りである。

1. 教室サイズの空間を対象に教室の広範囲を計測可能な行動計測システムを開発した。
2. 教室内に設置した各カメラの顔の検出数を時系列的に分析することで、授業内の活動分類ができることを示唆する結果を得た。
3. 授業中の学生の顔の向き情報を集合した正面方向交点を計算することで、集団が注目している個所の推定や、各自が自由に活動している場面かの分類ができる可能性を示唆する結果を得た。

以上より、本稿で示した対象の人物の教室中の位置や頭の向きを用いて正面方向交点群を計算し、重心位置やその広がりを見視化することで、授業活動中の状態の解釈に役立つ情報が得られることを示唆する結果を得た。また、本稿のように教室の四隅から中央方向に向けて計測する方式であっても、先行研究より得られた結果と概ね矛盾の無い計測と分析ができた。

一方で今後の課題は、集団活動の自動分類手法の検討とその実装である。先行研究より、個々の学生の状況を分析するためには、先に集団活動の分類し、その単位での個々の学生の状況を評価することが重要であるとされている[9][12]。その手法を検討すると共に、その分類に必要な特徴量についても同時に検討を進めることが必要である。また、本研究で得られた結果を教員、および学生の学びに役立てるための具体的な方法についても検討が必要である。本稿で示した手法を発展させることで教室空間の使用状況は可視化できると期待できるが、その有効な活用方法については今後の課題である。また、集団活動や個人の成長等の分類が可能であることを示すためには計測を定期的に行い、ビッグデータ化することが必要であると考え。本稿では一つの授業の結果のみを示したが、他の授業についても計測できているため、複数の授業を合わせて分析することで、学生の行動の変化や成長に関する情報についても得られることが期待できる。

また対象を大学生とする際の懸念事項として、カメラで計測されていることを意識し、授業中に演技する可能性がある。それにより結果に影響する可能性も考えられるが、長い目で見るとあまり影響はないと考える。その理由として現在我々は、防犯カメラという名目で様々な場所で撮影されながら生活をしている。しかし我々は普段、ほとんどその存在を気にせず生活を送っている。とするならば事前に同意を取っておき、カメラについては授業開始前から設置しておくことで、始めから環境の一つにしてしまうことで結果に影響しない計測が可能になると考える。

そして最後にこれらの課題を解決し、AI 技術を教育の質の向上のための道具の一つとして役立てることができる世の中になることを期待する。

謝辞 本研究の実施のために教室中の様子の計測、および分析に同意いただきました武蔵野大学建築学部の学生の皆さんに感謝致します。そして、授業を計測させていただきました武蔵野大学建築学部の風袋宏幸先生、および授業の計測を実施いただきました田丸恵理子様へ感謝致します。また本研究は、科研費 若手研究 22K13762 からの研究支援により実施された。支援に感謝する。本研究は、武蔵野大学中長期計画チャレンジ3「AI世界を先導する MUSIC (MUSIC 計画推進小委員会)」の2021～2022年度事業として実施された取組みの中でも、特に教室の空間デザインに関する成果報告である。

参考文献

- [1] ChatGPT : Optimizing Language Models for Dialogue, <https://openai.com/blog/chatgpt/>
(2023年2月19日閲覧)
- [2] Robin Rombach et al. (2022): “High-Resolution Image Synthesis with Latent Diffusion Models”, Computer Vision and Pattern Recognition, arXiv:2112.10752
- [3] 小竹原祐希 他 (2020): “講義映像に基づく受講者の多様な状況認識のための挙動のクラスタリング”, 教育システム情報学会誌, Vol.37, No.2, pp.120-130
- [4] 李凱 他 (2016): “モーショセンサを用いた学習活動の状態推定手法の開発”, 教育システム情報学会誌, Vol.33, No.2, pp.110-113
- [5] 本村陽一 他 (2019): “保育施設への AI 導入プロジェクトの課題と展望”, 2019年度人工知能学会全国大会
- [6] 響学スパイラル The Enhanced Learning Cycle, <https://risuyouran.musashino-u.ac.jp/spiral/>
(2023年2月19日閲覧)
- [7] T. Baltrušaitis et al. (2018): “OpenFace 2.0: Facial Behavior Analysis Toolkit”, 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, pp.59-66
- [8] Z Cao et al. (2021): “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, IEEE Transactions on PAMI, Vol.43, pp.172-186
- [9] 宮田真宏 他 (2021): “顔情報を用いた授業中の子どもの特性リフレクションシステムの開発”, 教育システム情報学会 2021年度 第2回研究会, Vol.32, No.2, pp.6-13
- [10] Takamoto et al. (2020): “An Efficient Method of Training Small Models for Regression Problems with Knowledge Distillation”, 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 67–72.
- [11] 山田徹志 他 (2021): “機械学習を用いた「子どもの育ち」の可視化—位置・向き情報を用いた関心推定の試み—”, 日本教育工学会論文誌, Vol.44, No.4, pp.365-376
- [12] 宮田真宏 他 (2022): “顔情報を用いた個々の子どもの特性を考慮した授業参加状態の推定”, 教育システム情報学会 2022年度 第2回研究会, Vol.37, No.2, pp.149-152