

修士学位論文

IoT カメラ・センサシステムを用いた
保育施設での環境・行動分析

平成 31 年 1 月 30 日

東京大学大学院 情報理工学系研究科
電子情報学専攻

48-176407 大淵 友暉

指導教員 山崎 俊彦 准教授

内容梗概

保育の場では待機児童問題や規制緩和による保育の質の低下の問題など様々な問題が生じている。本論文では保育施設等での空間的・時間的高密度センシングを目指して、IoTカメラ・センサシステムを独自開発した。約20施設に設置して計測したところ、殆どの部屋で二酸化炭素濃度や湿度等が国の基準値を満たしていないことがわかった。また、撮影した映像に対して深層学習を用いた姿勢推定、物体追跡、人物同定の手法を応用した行動分析の手法を提案した。提案手法により子供たちに対して完全非接触な条件下で、人物の存在位置、動線、運動量、個々人の行動傾向の分析・可視化を実現した。本論文のIoTカメラ・センサシステムにより、異常値を観測した時点で通知できるといったリアルタイムな対応が可能となる。また、本システムではクラウドに計測・撮影したデータを自動転送し、一括管理を行う。これにより生データの蓄積、および生データを分析して得られる統計データの蓄積が可能になる。データの蓄積により、子供の行動と発達の関係性に対する新たな知見の発見や保育施設に関する新たな研究への発展が期待できる。

目次

第 1 章	序論	1
1.1	背景と目的	1
1.2	本論文の構成	4
第 2 章	関連研究	6
2.1	環境分析	6
2.2	行動解析	6
第 3 章	保育施設での環境分析	12
3.1	デバイスの設計・開発	12
3.2	実験方法	13
3.3	実験結果	15
第 4 章	保育施設での行動分析	24
4.1	提案手法	24
4.2	予備実験 (人物再同定法の検証)	26
4.3	実験	31
第 5 章	結論	41
	参考文献	43
	関連する発表文献	50
	その他の発表文献	52

目次

2.1	YOLOv3 [1] を使った人物検出	8
2.2	OpenPose [2] を使った人物検出	9
2.3	Image Person Re-identification	10
2.4	Video Person Re-identification	10
3.1	オムロン 環境センサ [3]	13
3.2	Homeasure	14
3.3	不動産物件の住み心地計測時に使用したセンサデバイス	14
3.4	デバイス設置の様子	15
3.5	二酸化炭素濃度の計測結果例	21
3.6	湿度の計測結果例	22
3.7	騒音レベルの計測結果例	23
4.1	特徴抽出モデルに入力する画像	28
4.2	SCORER [4] カメラキット	32
4.3	カメラ設置の様子	33
4.4	人物存在確率のヒートマップ	34
4.5	動線解析	35
4.6	運動量解析	36
4.7	各子供の接近回数 (赤線は平均値)	37
4.8	ID=1~4, 6~9 の人物についての他の人物との接近回数	39
4.9	個々人の動線解析	40

表目次

1.1	各組織の環境基準	3
3.1	Homeasure で使用した各センサの仕様	13
3.2	基準を超える二酸化炭素濃度を観測した部屋数	17
3.3	基準を満たさない湿度を観測した部屋数	18
3.4	基準を超える騒音レベルを観測した部屋数	20
4.1	保育施設の映像を使った人物再同定法の検証用データセット	27
4.2	特徴抽出モデルの検証	28
4.3	損失関数の検証	29
4.4	入力画像に関する検証 (Dataset 1)	31
4.5	入力画像に関する検証 (Dataset 2)	31
4.6	IoT カメラを設置した保育施設についての撮影情報	32

第 1 章

序論

1.1 背景と目的

保育の場では様々な問題が生じている。近年、特に問題視される問題の一つとして待機児童問題が挙げられる。待機児童とは保育施設へ入所の条件を満たしているにもかかわらず入所ができない子供を指し、全国の待機児童数は平成 29 年 4 月 1 日時点では 26,081 人 (前年同月比 2,528 人増)、平成 29 年 10 月 1 日時点では 55,433 人 (前年同月比 7,695 人増) と年々増加していることがわかる [5, 6]。待機児童が増え続けている原因としては共働き世帯の増加や保育士、保育所不足等が挙げられる。増え続ける待機児童を減らすために、政府は保育施設の設置の促進や保育士への処遇改善等による支援を行う方針を立てた [7]。さらに、保育士の担当する子供の数の増やすとといったいわゆる規制緩和によって保育士不足への対策を講じている。しかし、この規制緩和は子供たちの安全の確保が難しくなり、保育の質が低下するのではないかという懸念が生じている。実際に保育施設でのうつぶせ寝による死亡事故が毎年数件起きており、全国の保育施設で報告された事故の総数は平成 27 年で 627 件、平成 28 年で 875 件、平成 29 年で 1242 件と年々増加の傾向にある [8]。政府が進める規制緩和により、保育施設での事故がさらに増加する可能性があると考えられている。規制緩和を進めるのであれば、少ない保育士でも負担を増やさずにかつ子供たちの安全を確保するような対策が必要であると考えられる。

本論文では保育施設に Internet of Things (IoT) を導入することを考える。IoT とはモノのインターネットと呼ばれ、モノの間でインターネットを経由して通信を行うことである。IoT デバイスは、すでに様々な分野で使用されており、その用途の範囲はますます増加していくとされている。IHS Markit によると、世界の IoT デバイス数は 2017 年で約 270 億個あり、2030 年には 1250 億個に達すると予測されている [9]。IoT は幅広い研究

に応用できる可能性があるため、現在はかなりの注目を浴びている。例えば、家電でセンシングを行うスマートホーム [10] や、自動車 [11]、工場 [12]、農場 [13] のセンシング等にも IoT が利用され始めている。しかし、従来からある住宅をスマートホームのようにするには、家電を取り替える必要があったりなど大変なコストがかかる。従来からある住宅に IoT を簡単に導入するには、低コストな IoT デバイスを設置することが考えられるが、まだ IoT が十分に普及されているとは言えない。また IoT デバイスによって得られたデータは、クラウドに転送して一括管理することができる。これにより計測した生データの蓄積や、生データを分析することで得られる統計データの蓄積も可能となる。このようなデータの蓄積は、特にこれまでデータが少なかった分野において更なる研究の発展に繋がる可能性がある。

本論文では IoT があまり普及しておらず、データも少ない分野の例として保育施設を対象に IoT デバイスを導入することを考える。保育施設に IoT センサデバイスを設置することで、室内環境をリアルタイムに監視することができ、施設内の環境分析を行える。また IoT カメラを設置することで、撮影された映像から子供たちの行動を分析できる。こうして得られるデータを蓄積することで、保育施設に関する研究の発展に繋がると考えられる。さらには保育施設における諸問題の解決にも有用になっていくであろう。また子供の行動と発達の関係性について新たな知見が得られることも期待できる。本論文では IoT カメラ・センサシステムを用いた保育施設での環境分析と行動分析について論じていく。

1.1.1 環境分析

保育施設の先生は子供たちにとって適切な室内環境を維持することが要求される。先生たちは主観的な評価によって室内環境が適切かどうかを判断していることが多く、数値による客観的な評価で判断することは少ない。例えば、室内が乾燥している感覚がなくても、実際は湿度が低い可能性がある。また二酸化炭素濃度に至っては、室内の濃度がどの程度なのかは感覚的には判断しにくい。

室内環境の環境値については国により適切とされる基準値が設けられている。厚生労働省の建築物環境衛生管理基準 [14] によると衛生上良好な状態を維持するには、空気調整設備を設けている居室において相対湿度は 40 % 以上 70 % 以下、二酸化炭素濃度は 1000 ppm 以下であることが推奨されている。文部科学省の学校環境衛生管理マニュアル [15] によると、建築物環境衛生管理基準 [14] よりは基準値が厳しくなく、相対湿度は 30 % 以上 80 % 以下、二酸化炭素濃度は 1500 ppm 以下、騒音の平均レベルは 50~55 dB 以下であることが望ましいとされている。環境省の騒音に係る環境基準 [16] では、生活環

表 1.1: 各組織の環境基準

組織	湿度	二酸化炭素濃度	騒音レベル
厚生労働省 [14]	40 %～70 %	1000 ppm 以下	-
文部科学省 [15]	30 %～80 %	1500 ppm 以下	50～55 dB 以下
環境省 [16]	-	-	50～60 dB 以下
ドイツ連邦環境庁 [17]	-	2000 ppm 以上は 「許容しがたい」	-
WHO 欧州事務局 [18]	-	-	45～54 dB 以下

境を保全し人の健康の保護に資する上で維持されることが望ましい基準として、騒音レベルが 50～60 dB 以下であることが定められている。我が国によらず世界的にも環境基準が定められており、ドイツ連邦環境庁 [17] では 2000 ppm を超える二酸化炭素濃度は「許容しがたい」という強い言葉で表現されている。World Health Organization (WHO) 欧州事務局が公表した環境騒音のガイドライン [18] では、交通騒音、鉄道騒音、航空機騒音など騒音の種類ごとに基準値が定められており、いずれも騒音の平均レベルで 45～54 dB 以下とされている。各組織の環境基準を表 1.1 にまとめた。

これらの環境基準が守れていない室内環境では健康被害をもたらす恐れがある。例えば、湿度が基準値を満たしていないと、インフルエンザなどの感染症が蔓延しやすい環境であると考えられる。二酸化炭素濃度は、集中力の低下や眠気を引き起こし、人によっては気分を悪くすることがある。騒音も基準を超えた状態が長時間続くと聴力レベルの低下を引き起こす可能性がある。騒音が大きい環境下では言葉の情報伝達が阻害されることも指摘されており [19]、子供たちの言葉の発達に影響を及ぼす可能性があるとする。先生にとっても、保育施設内での騒音によって聴力レベルが低下したことが確認された [20]。

本論文では保育施設において IoT センサデバイスを用いて室内環境を計測することを提案する。IoT デバイスで計測して先生に知らせることで、保育施設内が園児にとって適した環境であるかを常に確認できる。異常値を観測したら先生のスマートフォン等に知らせるという機能を付ければ、異常値を計測した時点で環境改善に向けて直ちに対処することができる。本論文では、実際の保育施設において IoT センサデバイスを用いて計測実験を行った。保育施設ではどのような異常値が観測され得るのかを確認した。

1.1.2 行動分析

子供の行動を分析する取り組みはこれまで多く行われてきた。そのような取り組みでは計測者がハンディカメラ等で子供の行動を記録している研究が多い [21, 22]。しかし、これは知らない人がカメラをもって立っているという普段とは侵襲的な状況での行動記録になってしまっている。ハンディカメラではなく、固定カメラで撮影した映像等を使って行動分析をする研究もあるが [23, 24]、これらも人手で分析を行っており大きな手間がかかってしまう。また、固定カメラを使っていて、かつ人の目で観察する手間を抑えた研究もある [25, 26]。しかし、こういった研究では子供たちにモーションセンサ等を取り付けて行動を追跡しており、子供たちに違和感を与える恐れがある。

そこで、本論文では人の眼での観察を最小限に抑えた、子供たちにモーションセンサ等を取り付けることのない完全非接触な状態での行動分析の手法を提案する。行動分析に使う映像は、天井等に設置した固定した IoT カメラで撮影した映像を用いる。この場合、撮影時に子供の近くに知らない人がいることもなく、普段と同じ状況の子供の様子を分析することができる。加えて、固定カメラで撮影した映像に映った子供一人ひとりの行動が分析できれば、観察対象が 1~2 人に限られることはなく、一度に多くの子供について行動分析が可能となる。

本論文では人物検出、物体追跡、人物再同定の手法を用いる。人物検出により撮影した画像のどの部分に人がよく居たのかがわかる。検出した人物に対し、数秒程度の短時間な追跡を行うために物体追跡の手法を用いる。短時間な追跡により、撮影した部屋のどの部分を人がよく通っていたかという動線解析が可能となる。短時間追跡した人物映像を、人物再同定の手法を用いて人物 ID を割り当てることでより長時間な追跡ができる。長時間な追跡により、各人物の運動に着目することができるようになり、ある子供がどの子供と近づいた回数が多いのかということなど、子供一人ひとりの行動の傾向がわかるようになる。

1.2 本論文の構成

本論文の構成は次の通りである。本章では保育施設における環境分析と行動分析に関してそれぞれ背景と目的を述べた。2 章では環境分析、行動分析それぞれについて関連研究を述べる。3 章では保育施設での環境分析について、環境分析用に我々が独自に開発したデバイスやで実験を行った際の計測結果などを紹介する。また、4 章は保育施設での行動

分析について，人物検出，物体追跡，人物再同定を応用した提案手法や，提案手法を用いて行動分析を行った例を説明する．最後に 5 章で本論文の結論について述べる．

第2章

関連研究

2.1 環境分析

生活環境を計測するためのデバイスは多くのものが開発されてきたが、近年は一つのデバイスで複数の環境値を計測できるものが数多く登場してきた。例えば、オムロンの環境センサ [3] は縦 46 mm, 幅 39 mm, 厚さ 15 mm, 質量 16 g という非常に小さなデバイスであるのにも関わらず、温度、湿度、照度、紫外線、気圧、騒音、加速度、Volatile Organic Compounds (VOC) の 8 つの環境値の計測可能である。Netatmo の Smart Home Weather Station [27] は温度、湿度、騒音、二酸化炭素濃度の 4 つの環境値が計測できるのに加え、計測された二酸化炭素から空気の汚染レベルを計算する。Acer Air Monitor [28] は温度、湿度、VOC、二酸化炭素濃度、Particulate Matter (PM) 2.5, PM10 の 6 つの環境値を計測できる。SenStick [29] は眼鏡や箸などにも取り付けて環境値は計測する目的で開発されたものであり、これも小型なデバイスとなっているが、温度、湿度、照度、紫外線、気圧、騒音、加速度、角速度、地磁気を計測できる。Awair Glow [30] も温度、湿度、二酸化炭素濃度を計測できるのに加え、空気中のほこりや化学物質も計測できるデバイスとなっている。本研究では環境計測のためのデバイスとして、小型で使用しやすいオムロンの環境センサ [3] と、我々が独自に開発したデバイスを用いることとする。我々が独自に開発したデバイスについては 3.1 節で紹介する。

2.2 行動解析

子供の行動分析するには、まず映像中の人物を検出をすることが必要である。2.2.1 節では画像中の人物検出の手法について述べる。子供一人ひとりの行動を分析するにはフ

フレーム毎に検出した人物の対応付けが必要である。対応付けのための一つの手法として物体追跡が挙げられる。2.2.2 節では画像中の物体追跡の手法について述べる。また、より長期的な追跡をするためには人物再同定の手法が有用であると考えられる。2.2.3 節では人物再同定の手法について述べる。

2.2.1 人物検出

画像中の物体検出はこれまで多くの研究がなされてきた。深層学習が使用される以前は Histogram of Oriented Gradients (HOG) [31] 特徴量を使用したものがあり、近年では Region-based Convolution Neural Network (R-CNN) [32], Faster R-CNN [33], Single Shot Detector (SSD) [34], You only Look Once (YOLO) [35] など深層学習による物体検出手法が研究されてきた。特に SSD [34] や YOLO [35] などの最近の手法を使えば、使用するデータに応じて学習をしなくても、他のデータセットでの学習済みのモデルで画像中の物体が精度よく検出できるようになっている。これらの手法を用いて画像中の人物を検出することが可能であると言えるが、高精度に検出できるのは十分に大きく映っている、もしくは正面の画角に移っている人物である。保育施設の映像は子供が十分に正面で大きく映っているとは限らず、他のデータセットでの学習済みの物体検出モデルにより子供全員を検出することは難しい。図 2.1 は YOLO [35] の発展版である YOLOv3 [1] を使用して人物検出をした例であるが、手前に居て真上から映っている子供はあまり検出されていないことがわかる。保育施設の映像データを使って学習を行えば、図 2.1 で検出されなかった子供も検出可能になると考えられるが、多くの学習用データが必要であり大変な手間がかかる。

人物検出には姿勢推定の手法を利用することも可能である。姿勢推定とは画像中の人物の関節の位置を推定するというものであるが、精度よく姿勢推定ができれば人物検出として利用できると考えられる。姿勢推定には DeepPose [36] など、一人を対象にした姿勢推定があるが、保育施設で撮影された映像には複数の人物が写ると考えられるため、これは使用できない。保育施設での行動分析には複数人に対する姿勢推定が有用である。複数人の人物に対する姿勢推定には大きく分けて Top-down 型の手法と Bottom-up 型の手法がある。Top-down 型ではまず二次元画像中にいる人を検出し、検出したそれぞれの人に対し姿勢推定を行うというものである。人の検出についてはは体全体を検出するものもあれば、人の頭や首などの一部だけを検出するものもある。Top-down 型の手法は直感的で従来の人物検出や DeepPose [36] などの一人に対する姿勢推定の手法を適用しやすく、複数人に対する姿勢推定の研究が始まった当初は好まれて使われてきた。



図 2.1: YOLOv3 [1] を使った人物検出

しかし、Top-down 型の手法では最初の人物検出に失敗するともちろん姿勢推定もできなくなってしまう。さらに画像中の人の数が多いほど、計算コストがそれに比例して増大してしまう。一方、Bottom-up 型の手法ではまず画像中の全ての人物について各関節の位置を推定する。そして、各関節の連結を推定することで複数人に対する姿勢推定を行う。Bottom-up 型の手法では画像中の人数に比例して計算コストが増大することはない。複数人を対象とした姿勢推定の手法は、例えば DeepCut [37], DeeperCut [38], OpenPose [2] などがあるが、中でも OpenPose [2] は短時間で高精度な姿勢推定を可能にした。さらに OpenPose [2] は使用するデータに応じた学習なしに高精度な姿勢推定を可能にする。図 2.2 は OpenPose [2] を使用して人物検出 (姿勢推定) を行った例である。図 2.1 と比べ、OpenPose [2] は YOLOv3 [1] で検出できなかった真上から映っている子供も検出できていることがわかる。姿勢推定による人物検出は、人物の位置が得られるだけでなく人物の姿勢も得られるため、各関節に着目して行動分析をすることができると考えられる。保育施設の映像に対する人物検出として、OpenPose [2] を採用することにする。

2.2.2 物体追跡

近年の主流となっている物体追跡の手法として、Discriminative Correlation Filter (DCF) ベースのものと Convolution Neural Network (CNN) ベースのものがある。DCF ベースの例として Minimum Output Sum of Squared Error (MOSSE) [39], DSST [40], が挙げられ、CNN ベースの手法の例として TCNN [41], MDNet [42], SANet [43] の



図 2.2: OpenPose [2] を使った人物検出

手法が挙げられる。さらに DCF と CNN の両方を利用した手法として、C-COT [44] や ECO [45] などが挙げられる。CNN ベースの手法の利点として全体的に精度が高いことが言えるが、処理時間が長いという欠点がある。一方で DCF ベースの手法では CNN ベースの手法に比べて精度が劣るが処理時間が短い。

しかし、物体追跡の手法のみでは、保育施設の映像のような子供たちが混みあった映像で人物を長期間追跡し続けるのは難しい。さらに保育施設の映像では子供たちが画面内を出入りすることがよくある。一度画面の外に出た人物が再び画面内に戻ってきたとき、それを同じ人物として追跡するのは困難である。長期的な追跡には 2.2.3 節で述べる人物再同定の手法を使う必要がある。短期間の追跡は DCF ベースと CNN ベースのいずれの手法でも可能であると考えられる。本研究では短期間の追跡として DSST [40] の手法を採用する。

2.2.3 人物再同定

人物再同定では、人物 ID の参照用のデータとして Gallery というデータセットを用意する。Query データを入力として、Gallery 中にあるデータと特徴量空間内での距離を比較する。Gallery には Query と同じ人物のデータが含まれている必要がある。Gallery 中のデータで特徴量空間内で距離が近いものを探すことで、Query データの人物 ID を特定する。人物再同定では画像毎に再同定するもの (Image Person Re-identification) と、映像毎に再同定するもの (Video Person Re-identification) がある。Image Person Re-identification は図 2.3 のように、人物の画像一枚ごとに特徴抽出モデルに通して特徴量を

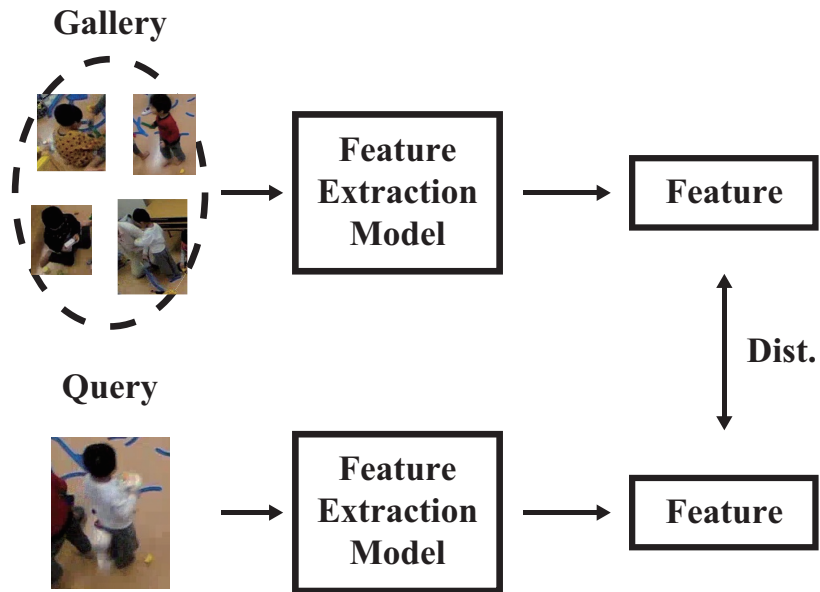


図 2.3: Image Person Re-identification

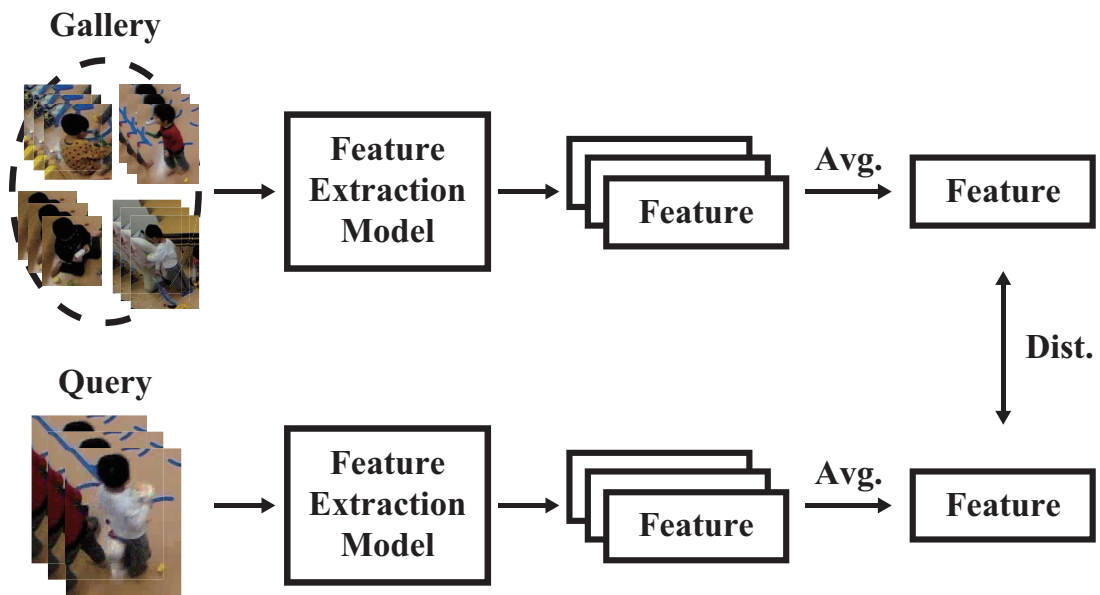


図 2.4: Video Person Re-identification

Gallery 中のデータと Query データで比較する。一方, Video Person Re-identification は図 2.4 のように特徴抽出は画像一枚ごとに行うが, 一つの映像で得られる特徴量の平均値を Gallery 中のデータと Query データで比較するというものである。

特徴抽出モデルには LOMO [46] など深層学習を使用しないモデルもあるが, 近年で

は深層学習を利用したものがより高い精度を達成してきた。特徴抽出モデルには人物に限らず画像分類用に提案された特徴抽出モデルも利用することができある。画像分類でも深層学習を使った手法が大変な発展を遂げており、ResNet [47], ResNeXt [48], SENet [49], DenceNet [50] などの手法が挙げられる。また、より軽量なモデルとして、MobileNetV2 [51], SqueezeNet [52] なども挙げられる。人物再同定用に提案された特徴抽出モデルもあり、MuDeep [53], ResNet-mid [54], HACNN [55], PCB [56], MLFN [57] が挙げられる。

しかし、一般的な人物再同定の実験設定と保育施設用の実験設定は異なるといえる。一般的な人物再同定の研究では非常に多くの人物を含むデータセットを扱うことが多い。例えば、Market1501 [58] や CUHK03 [59] は Image Person Re-identification の研究でよく用いられるデータセットであるが、それぞれ 1501 人、1360 人とどちらも 1000 人を超える人物が含まれたデータセットとなっている。Video Person Re-identification の研究では、DukeMTMC-VideoReID [60] などのデータセットがよく使用されるが、これも 702 人と非常に多くの人物を含んだデータセットとなっている。さらに、一般的な手法は監視カメラの映像での人物再同定を目的としているものが多く、学習に使った人物データとは違う人物の検索等に用いられるため、学習用に使う人物とテスト用に使う人物が異なっている。しかし、保育施設の映像では多くの人物が登場することはなく精々 10 数人程度であるのに加え、学習用に使う人物とテスト用に使う人物を変える必要もない。人物再同定用に提案された手法が保育施設用の実験設定で上手くいくとは限らない。また、学習時の損失関数として softmax cross entropy loss [61, 62] や triplet loss [63] が提案されている。これらの手法の内、どの手法が保育施設の映像データに適しているかを 4.2 節で検証する。

第 3 章

保育施設での環境分析

3.1 デバイスの設計・開発

環境計測には 2 種類の IoT デバイスを使用する。一つは 2.1 節で紹介したオムロンの環境センサ [3] である。オムロンの環境センサ [3] の外観を図 3.1 に示す。もう一つは我々が独自に開発した Homeasure (図 3.2) というデバイスである。Homeasure はフューチャースタANDARD社が提供する映像解析プラットフォームである SCORER [4] をベースとして我々が独自にセンサを選定し、はんだ付け等を行って開発したものである。温湿度センサ (HDC1000 [64]), 照度センサ (TSL2561 [65]), 加速度センサ (LSM9DS1 [66]), 紫外線センサ (VEML6070 [67]), CO₂ センサ (K30 [68]) を搭載し, Raspberry Pi 3 Model B [69] で制御している。各センサの仕様を表 3.1 に示す。各センサは取り外し可能であり, 必要に応じて使用したいセンサを選択できる。Homeasure の内観は蓋を外して上から見ると図 3.2b のようになっており, CO₂ センサ (K30 [68]), 照度センサ (TSL2561 [65]), 紫外線センサ (VEML6070 [67]) が見れる。CO₂ センサ (K30 [68]) の下には, 図 3.2c のように赤い加速度センサ (LSM9DS1 [66]) がある。温湿度センサ (HDC1000 [64]) については, Raspberry Pi 3 Model B [69] 等による排熱の影響を防ぐために脇にあるポケット (図 3.2a の右側, 図 3.2b の上側) に収納される造りとなっている。Raspberry Pi 3 Model B [69] では複数の USB ポートが備え付けてあるため, Homeasure にも USB カメラを接続して映像を記録することができる。Homeasure は当初不動産物件の住み心地を計測するために開発されたものである [70, 71]。不動産物件の計測では図 3.3 のようなセンサデバイスを使用し, 状況確認と騒音計測用に USB カメラと USB マイクを接続し利用した。図 3.3 では CO₂ センサ (K30 [68]) は使用しなかった。Homeasure は電源を接続するだけで計測を開始し, インターネット環境に繋がっていれば計測データを自動的



図 3.1: オムロン 環境センサ [3]

表 3.1: Homeasure で使用した各センサの仕様

センサ	デバイス名	仕様
温湿度センサ	HDC1000 [64]	(温度) レンジ: $-20\sim 85\text{ }^{\circ}\text{C}$, 精度: $\pm 0.2\text{ }^{\circ}\text{C}$ (湿度) レンジ: $0\sim 100\%$, 精度: $\pm 3\%$
照度センサ	TSL2561 [65]	レンジ: $0.1\sim 40,000\text{ lx}$, 赤外線も計測可能
加速度センサ	LSM9DS1 [66]	レンジ: $\pm 2/\pm 4/\pm 6/\pm 8/\pm 16\text{ G}$ (切替可)
紫外線センサ	VEML6070 [67]	計測波長: $320\sim 410\text{ nm}$ (ピーク波長 355 nm)
CO ₂ センサ	K30 [68]	レンジ: $0\sim 5000\text{ ppm}$, 精度: $\pm 30\text{ ppm}\pm 3\%\text{rdg}$

にクラウドに転送するという仕組みになっている。その為、データ回収の手間はかからずデータの一元管理が可能である。

3.2 実験方法

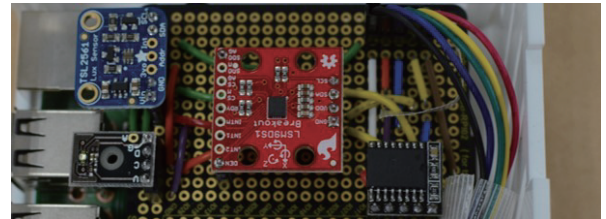
実際の保育施設に 3.1 節で紹介した 2 種類のデバイスを設置して計測実験を行った。設置の様子を図 3.4 に示す。Homeasure は図 3.4a のように子供たちの手の届かない棚の上などに設置した。オムロン環境センサ [3] についても、Homeasure と同様に棚の上に設置できるが、非常に小型で軽量なため図 3.4b のようにテープを使って柱や壁に設置することができる。このようにして我々は 19 施設 52 部屋の保育施設で計測実験を行った。本



(a) 外観



(b) 上から見た内観



(c) CO₂ センサ (K30 [68]) の下の内観

図 3.2: Homeasure

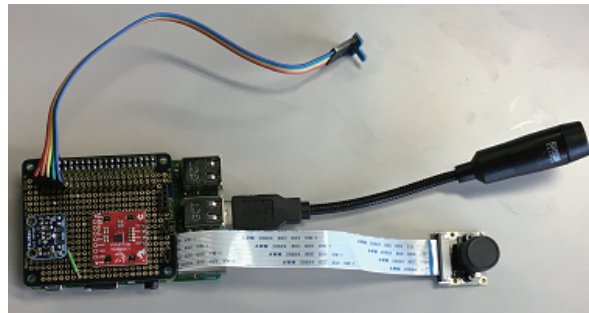


図 3.3: 不動産物件の住み心地計測時に使用したセンサデバイス

論文では 19 の保育施設をそれぞれ Nursery A～S とおく．各保育施設での計測期間は約 1 週間程度である．3.3 節では計測した環境値の内，二酸化炭素濃度，湿度，騒音レベルについて計測結果の例を挙げる．



(a) Homeasure の設置



(b) オムロン環境センサ [3] の設置

図 3.4: デバイス設置の様子

3.3 実験結果

3.3.1 二酸化炭素濃度

Nursery E においてある一日に観測した二酸化炭素濃度を図 3.5a に示す。図 3.5a では 2 歳児の部屋と 5 歳児の部屋で観測した結果を示している。どちらの部屋についても、建築物環境衛生管理基準 [14] で定められている 1000 ppm，及び学校環境衛生管理マニュアル [15] で定められている 1500 ppm を超えた二酸化炭素濃度を観測していたことがわかる。また，2 歳児の部屋ではドイツ連邦環境庁 [17] で「許容しがたい」とされた 2000 ppm に達した瞬間があり，5 歳児の部屋では 2000 ppm を大きく超えて 4000 ppm 程度の二酸化炭素濃度を観測した。2 歳児に比べ，5 歳児の方が動きが激しく肺活量も大きいため，二酸化炭素濃度が高くなりやすいと推測できる。

図 3.5b では，Nursery F である一日に計測された二酸化炭素濃度を示す。図 3.5b には 3 歳児の部屋と 4 歳児の部屋での計測結果を示しているが，どちらの部屋も 1000 ppm を超えた二酸化炭素濃度を観測している。4 歳児の部屋では，5000 ppm を超えた二酸化炭

素濃度を観測したことがわかった。建築物環境衛生管理基準 [14] では 5000 ppm は許容限界とされ、大人でも二酸化炭素濃度が 5000 ppm を超えた状態の室内に 8 時間以上滞在した場合、健康に深刻な影響を及ぼすとされる。そのような危険な値を Nursery F の 4 歳児の部屋では観測されていたことがわかった。

二酸化炭素の計測実験を行った 19 施設 52 部屋について、1.1.1 節で紹介した環境基準を超える二酸化炭素濃度を観測した部屋数を表 3.2 でまとめた。表 3.2 では 1000 ppm 以上、1500 ppm 以上、2000 ppm 以上の二酸化炭素濃度を観測した部屋数を示している。Nursery A' は計測を行った場所は Nursery A であるが、計測を行った時期が異なるため Nursery A' としている。Nursery G のある 1 部屋 (0 歳児部屋) 以外の全ての部屋で 1000 ppm を超える二酸化炭素濃度が観測されたことがわかった。文部科学省の基準である 1500 ppm を超える二酸化炭素濃度を観測した部屋数は 40 部屋 (77 %) であった。また、19 部屋 (37 %) でドイツ連邦環境庁で「許容しがたい」とされた 2000 ppm を超える二酸化炭素濃度を観測した。特に、Nursery E と Nursery F では計測したほぼ全ての部屋で 2000 ppm を超える二酸化炭素濃度を観測しており、他の保育施設に比べて二酸化炭素濃度が高くなりやすいことがわかった。どちらの保育施設も、部屋によってはそれぞれ 4000 ppm や許容限界とされる 5000 ppm を超える値を観測しているため注意が必要であると考えられる。

3.3.2 湿度

図 3.6a に Nursery E の 1 歳児の部屋と 3 歳児の部屋で観測された湿度を示す。図 3.6a はある一日の結果である。どちらの部屋についても、建築物環境衛生管理基準 [14] で定められている 40 %、及び学校環境衛生管理マニュアル [15] で定められている 30 % を下回る湿度を観測していたことがわかる。また、3 歳児の部屋に比べて 1 歳児の部屋の方が乾燥しやすいことがわかった。図 3.6a の計測時期は 1 月で乾燥しやすい季節であったため、湿度が低くなりやすい傾向にあるといえるが、感染症が蔓延しやすい環境であり注意が必要である。

図 3.6b に Nursery J である一日に観測された湿度を示す。Nursery J では建築物環境衛生管理基準 [14] で定められている 70 %、及び学校環境衛生管理マニュアル [15] で定められている 80 % を上回る湿度を観測していた。図 3.6b の計測時期は 7 月で湿気が多い季節であったため、湿度が高くなる傾向にあるが、カビの繁殖や熱中症を引き起こす可能性があり、こちらも注意が必要である。

湿度の計測を行った 15 施設 22 部屋について、建築物環境衛生管理基準 [14] 及び学校

表 3.2: 基準を超える二酸化炭素濃度を観測した部屋数

保育施設	計測月	計測部屋数	1000 ppm 以上	1500 ppm 以上	2000 ppm 以上
Nursery A	8月	2	2	1	-
Nursery A'	6月	1	1	1	-
Nursery B	11月	1	1	1	1
Nursery C	12月	6	6	6	3
Nursery D	12月	6	6	6	1
Nursery E	1月	6	6	6	6
Nursery F	1月	7	7	7	6
Nursery G	1月	6	5	1	-
Nursery H	2月	4	4	3	-
Nursery I	3月	3	3	3	1
Nursery J	7月	1	1	-	-
Nursery K	7月	1	1	-	-
Nursery L	6月	1	1	-	-
Nursery M	6月	1	1	1	1
Nursery N	3月	1	1	1	-
Nursery O	6月	1	1	1	-
Nursery P	6月	1	1	-	-
Nursery Q	7月	1	1	1	-
Nursery R	6月	1	1	1	-
Nursery S	7月	1	1	1	-
Total		52	51 (98 %)	40 (77 %)	19 (37 %)

環境衛生管理マニュアル [15] の環境基準を満たさない湿度を観測した部屋数を表 3.3 に示す。表 3.3 では 40 % 以下, 30 % 以下, 70 % 以上, 80 % 以上の湿度を観測した部屋数を示している。計測実験を行った 22 部屋の内, 40 % 以下の湿度を観測したのは 12 部屋 (55 %), 30 % 以下を観測したのは 10 部屋 (45 %) であり, 70 % 以上の湿度を観測したのは 10 部屋 (45 %), 80 % 以上を観測したのは 2 部屋 (9 %) であった。しかし, 湿度は季節に依存しやすいため, 計測時期によって値が大きく異なる。計測時期が 1~3 月という乾燥しやすい冬場に実験を行った 5 施設 12 部屋 (Nursery E~I) では, 全ての部屋

表 3.3: 基準を満たさない湿度を観測した部屋数

保育施設	計測月	計測部屋数	40 % 以下	30 % 以下	70 % 以上	80 % 以上
Nursery A'	6月	1	-	-	1	-
Nursery E	1月	2	2	2	-	-
Nursery F	1月	2	2	2	-	-
Nursery G	1月	2	2	1	-	-
Nursery H	2月	4	4	4	-	-
Nursery I	3月	2	2	1	-	-
Nursery J	7月	1	-	-	1	1
Nursery K	7月	1	-	-	1	1
Nursery L	6月	1	-	-	1	-
Nursery M	6月	1	-	-	1	-
Nursery O	6月	1	-	-	1	-
Nursery P	6月	1	-	-	1	-
Nursery Q	7月	1	-	-	1	-
Nursery R	6月	1	-	-	1	-
Nursery S	7月	1	-	-	1	-
Total		22	12 (55 %)	10 (45 %)	10 (45 %)	2 (9 %)

で 40 % を下回った湿度が観測されている。また，Nursery G の 1 部屋を除いた全ての部屋で 30 % を下回った湿度が観測されていることがわかる。計測時期が 6～7 月という湿気が多い夏場実験を行った 10 施設 10 部屋 (Nursery A', J～S) では，全ての部屋で 70 % を上回った湿度が観測されていた。

以上のことから，計測実験を行った多くの保育施設では湿度対策がとられていない，もしくは対策をしているがあまり効果が得られえていないことが考えられる。インフルエンザ等の感染症の蔓延，カビの繁殖，熱中症の予防のためにも湿度対策が必要である。

3.3.3 騒音レベル

図 3.7a に Nursery E で観測された騒音レベルを示す。これは 3 歳児の部屋のある一日に観測された結果である。図 3.7a を見ると，子供たちが遊ぶ時間帯である 8:30～10:00，

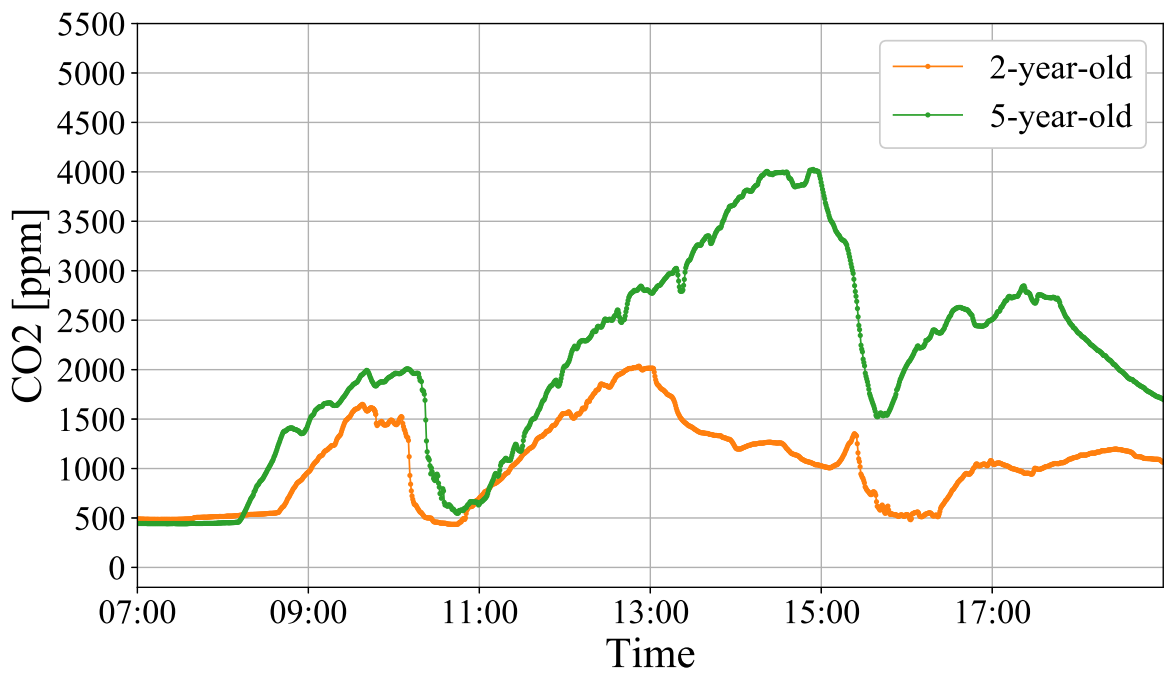
11:00~12:30, 15:00~17:00 の間に文部科学省の基準 [15] である 50~55 dB や、環境省の基準 [16] である 50~60 dB を超える騒音レベルが観測されていたことがわかる。Nursery E ではセンサデバイスによる環境計測と同時に、カメラで室内の様子を撮影していた。カメラの映像で確認したところ、図 3.7a の A の時刻ではおもちゃの準備の際に誤って物を落としたために騒音レベルが高くなっていることがわかった。また、B の時刻ではちょうど先生が部屋から出ている間であったが、ある子どもが物を振り回して椅子にぶつけていた。一方で、C の時刻では先生による本の読み聞かせの時間であり、子供たちがあまり騒ぐことがなかったため、騒音レベルが低くなっている。このことから、騒音レベルの監視は何らかの異常事態の検出に使える可能性があるといえる。

図 3.7b に Nursery M についてある一日の騒音レベルの計測結果を示す。Nursery M についても子供たちが遊ぶ時間帯である 8:00~11:00, 15:00~17:00 の間に 50~60 dB を超える騒音レベルが観測されていたことがわかる。また、図 3.7a に比べて 50 dB 以上の騒音レベルを観測している時間帯が長い。Nursery M ではカメラによる撮影はしていなかったため確認はできないが、図 3.7b で騒音レベルが高い時刻に何らかの異常事態が起きていたかもしれない。

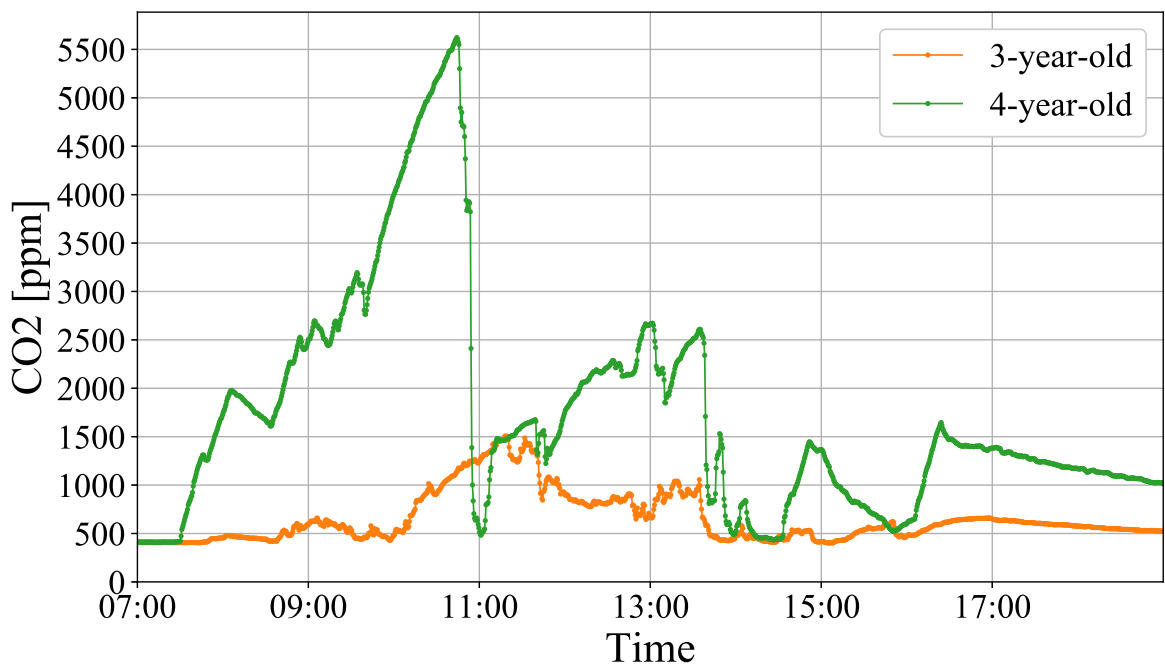
各保育施設の騒音レベルの計測結果について、8:00~18:30 の時刻で 30 分間ごとに騒音レベルの平均値を求めた。この平均値は、文部科学省や環境省が定めた環境基準である騒音の平均レベルとは厳密には計算手法が異なるが、参考値として利用できる。表 3.4 に騒音レベルの計測を行った 15 施設 21 部屋について、求めた各時間帯の平均値の最大値が 45 dB 以上、50 dB 以上、55 dB 以上、60 dB 以上だった部屋数を示す。表 3.4 より、全 21 部屋で騒音レベルの平均値が 45 dB を超えた時間帯があることがわかった。また、騒音レベルの平均値が 50 dB を超えた時間帯がある部屋は 18 部屋 (86 %) であり、55 dB を超えた時間帯がある部屋は 15 部屋 (71 %) であった。Nursery A' と Nursery E と Nursery M の 3 施設では、30 分間の騒音レベルの平均値が 60 dB を超えた部屋があることがわかり、非常に騒がしい環境にあるといえる。

表 3.4: 基準を超える騒音レベルを観測した部屋数

保育施設	計測月	計測部屋数	45 dB 以上	50 dB 以上	55 dB 以上	60 dB 以上
Nursery A'	6月	1	1	1	1	1
Nursery E	1月	2	2	2	2	1
Nursery F	1月	2	2	2	2	-
Nursery G	1月	2	2	1	1	-
Nursery H	2月	3	3	2	-	-
Nursery I	3月	2	2	2	1	-
Nursery J	7月	1	1	1	1	-
Nursery K	7月	1	1	1	1	-
Nursery L	6月	1	1	1	1	-
Nursery M	6月	1	1	1	1	1
Nursery O	6月	1	1	1	1	-
Nursery P	6月	1	1	-	-	-
Nursery Q	7月	1	1	1	1	-
Nursery R	6月	1	1	1	1	-
Nursery S	7月	1	1	1	1	-
Total		21	21 (100 %)	18 (86 %)	15 (71 %)	3 (14 %)

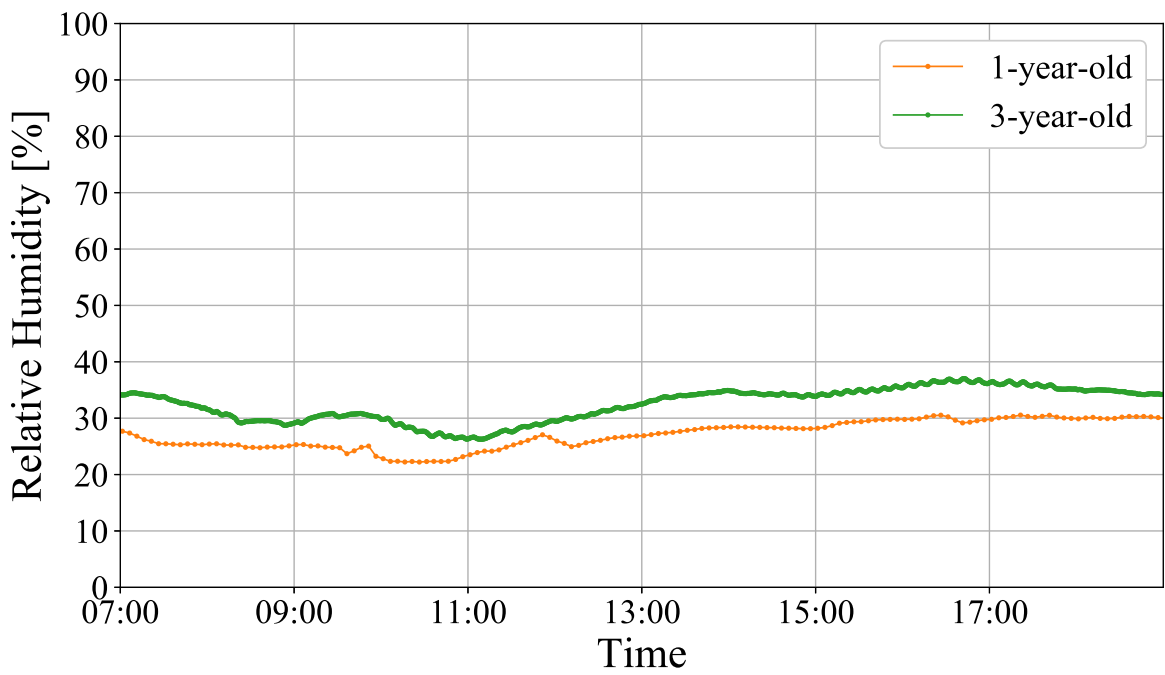


(a) Nursery E で観測された二酸化炭素濃度

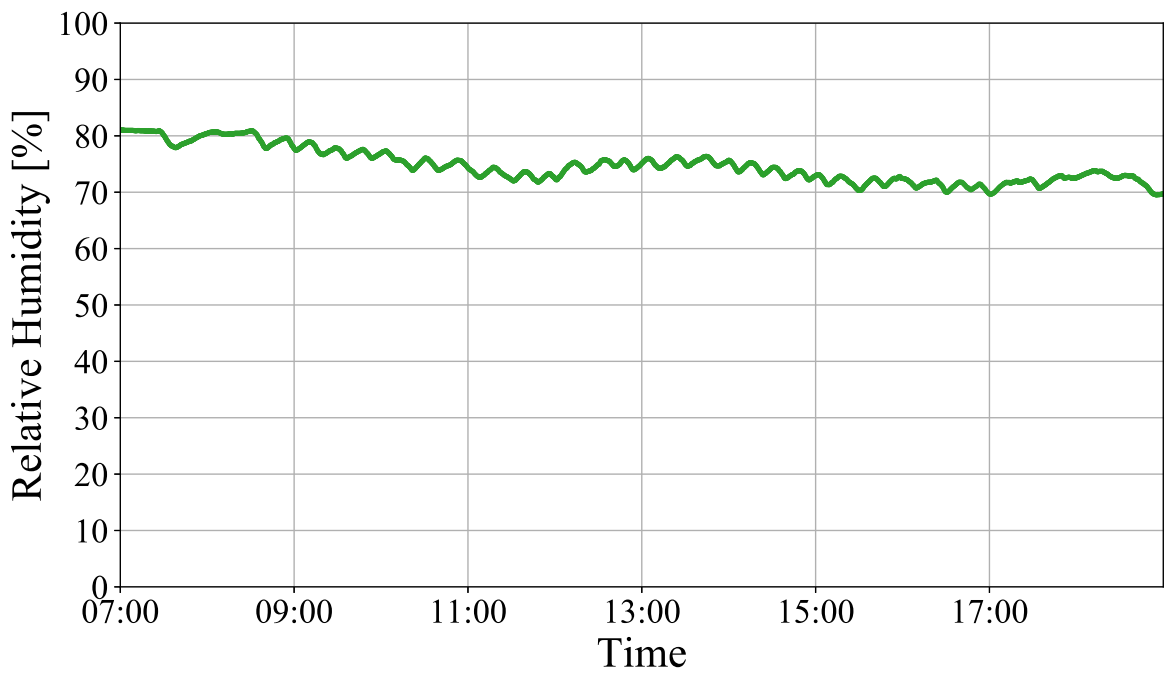


(b) Nursery F で観測された二酸化炭素濃度

図 3.5: 二酸化炭素濃度の計測結果例

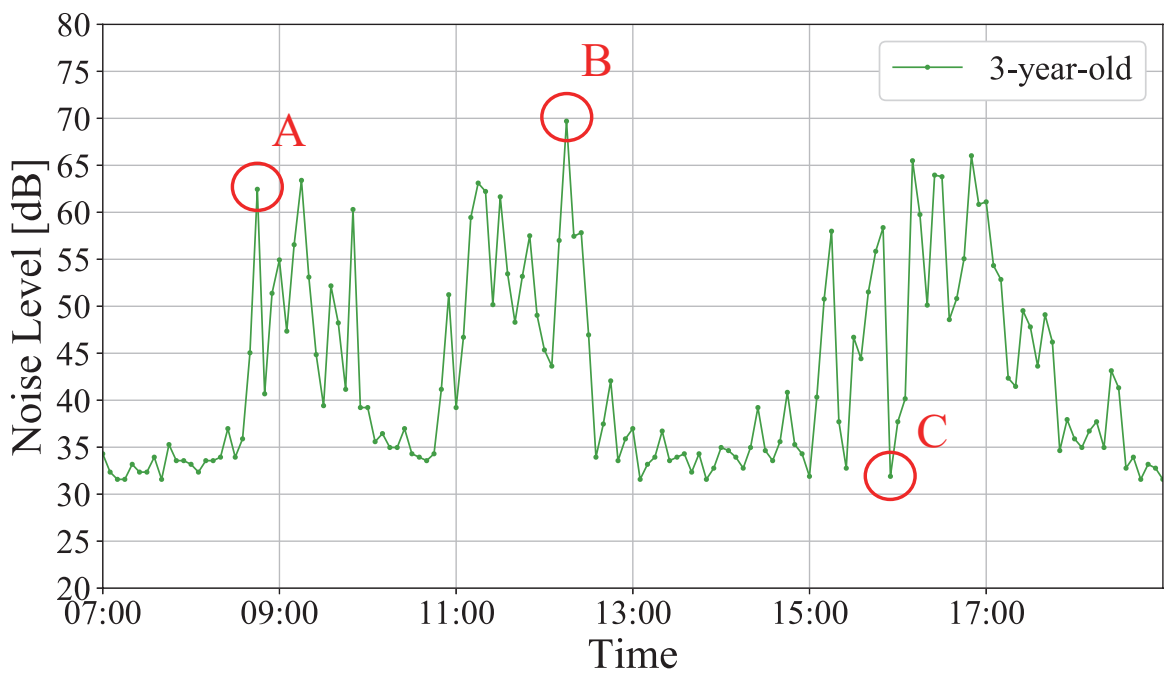


(a) Nursery E で観測された湿度

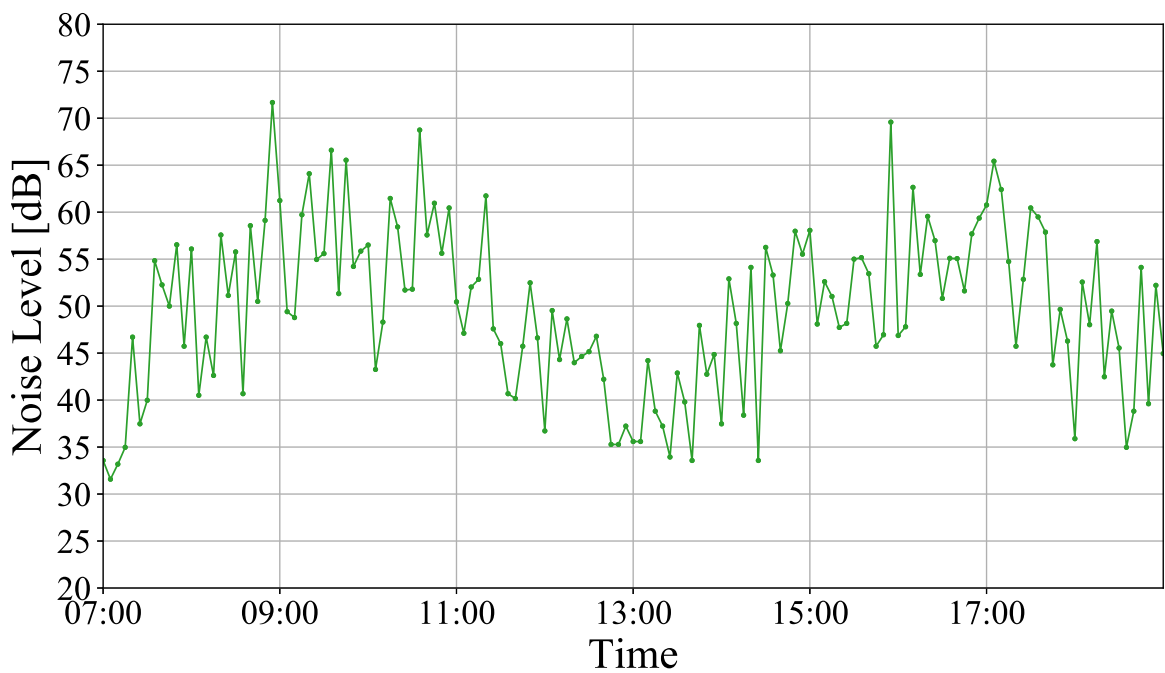


(b) Nursery J で観測された湿度

図 3.6: 湿度の計測結果例



(a) Nursery E で観測された騒音レベル



(b) Nursery M で観測された騒音レベル

図 3.7: 騒音レベルの計測結果例

第 4 章

保育施設での行動分析

4.1 提案手法

本論文では子供にモーションセンサを取り付けるといったこともなく，映像解析のみで完全に非接触な行動分析手法を提案する．保育施設内で撮影した映像に対し，フレーム毎に OpenPose [2] を用いて人物検出を行う．これだけで撮影した部屋のどの部分に子供や先生がよく居たのかがわかる (4.3.2 節)．

検出された人物の対応付け (仮の人物 ID 付け) は，フレーム前後の関節の位置の関係性と DSST [40] を利用する．この手法の詳細を 4.1.1 節に示す．これにより，撮影した部屋のどの部分を子供や先生がよく通っていたかという動線解析 (4.3.3 節) が可能となり，関節の動きから全体的な運動量解析 (4.3.4 節) が得られる．

仮の人物 ID 付けした短時間の人物追跡映像の対応付け (真の人物 ID) を行う．これには 4.2.1 節の検証で最もパフォーマンスが高かった ResNet [47] を用いて行った．この手法の詳細を 4.1.1 節に示す．これにより各人物の運動に着目することができるようになり，子供や先生一人ひとりの動きの軌跡や，ある子供がどの子供と近づいた回数が多いのかということなどがわかるようになる (4.3.5 節)．

4.1.1 短時間の人物追跡 (仮の人物 ID 付け)

映像の frames per second (fps) が十分に大きければ，フレーム間で人物の動きは小さいので，ある程度は位置が近いもので対応付けが可能であると考えられる．フレーム t で検出された仮 ID の集合を I_t ，最初のフレームからフレーム t までに検出された全ての仮 ID の集合を I_{all_t} とする．また，関節の集合を K とする．フレーム t で仮 ID が $i \in I_t$

の推定された関節 $k \in K_{i,t}$ の座標を $p_{i,k,t}$ とおく．次のフレーム $t+1$ で検出された人物の仮 ID $j \in I_{t+1}$ を以下の式 (4.1) ように関節の平均移動距離を比較して求める．

$$j = \underset{i \in I_t}{\operatorname{argmin}} E_{k \in K_{j,t+1} \cap K_{i,t}} (\|p_{j,k,t+1} - p_{i,k,t}\|_2) \quad (4.1)$$

ただし，関節の平均移動距離の閾値を $joint_th$ としたとき，関節の平均移動距離の最小値が $joint_th$ を超えた場合はフレーム $t+1$ で検出された人物の仮 ID j は I_t 内には存在しないとする．その場合は $past_th$ フレーム前まで遡った仮 ID の集合 $I_past_t = I_all_{t-1} \setminus I_all_{t-past_th-1}$ の中からフレーム $t+1$ で検出された人物の仮 ID j に該当するものを以下の式 (4.2) ようにして探す．

$$j = \underset{i \in I_past_t \setminus I_t}{\operatorname{argmax}} \operatorname{Overlap}(Rect_P_{j,t+1}, Rect_T_{i,t+1}) \quad (4.2)$$

ここで $Rect_P_{j,t+1}$ はフレーム $t+1$ で検出された人物の仮 ID j の関節を全て含む矩形を指し， $Rect_T_{i,t+1}$ はフレーム $t+1$ よりも前に検出された人物の仮 ID i の関節を全て含む矩形を Tracker を使ってフレーム $t+1$ まで物体追跡した際の矩形である．また， $\operatorname{Overlap}(Rect_1, Rect_2)$ は $Rect_1$ と $Rect_2$ の2つの矩形の重なる部分の面積を指す．この場合も重なる部分の面積の閾値 $overlap_th$ を設ける．重なる部分の面積の最大値が $overlap_th$ 以下の場合は，フレーム $t+1$ で検出された人物の仮 ID j は I_all_t 内にはないとし，新規の仮 ID として扱う ($\{j\} \cup I_all_t \subseteq I_all_{t+1}$)．

本研究では $joint_th = 20$ ， $past_th = 10$ ， $overlap_th = 0.4$ として実験を行った．また Tracker には DSST [40] を用いている．

4.1.2 長時間の人物追跡 (真の人物 ID 付け)

4.1.1 節で作成した仮 ID 付けした人物追跡映像では，一人の人物について長期的な追跡は難しい．実際に 15 人程度しか映っていない 17 分間の映像に対して 4.1.1 節の手法を適用した場合，付けられた仮 ID の総数は 4000 以上であった．そこで仮 ID 付けした人物追跡映像に対し，人物再同定の手法を用いて真の人物 ID 付けを行う．

ここで人物再同定の手法を適用する前に，仮 ID 付けした人物追跡映像をそれぞれ追跡時間が最大 1 秒となるように仮 ID を付け直す．これは 4.1.1 節の手法では，同じ仮 ID が付けられていても長い時間追跡できているものほど，途中で違う人物を追跡してしまい，人物追跡を誤る割合が多いからである．追跡時間が最大 1 秒となるように仮 ID を付

け直せば、同じ仮 ID の人物追跡映像内に違う人物が含まれる割合が非常に低くなる。このように仮 ID を付けなおした人物追跡映像に対し、人物再同定の手法を用いて真の人物 ID 付けを行う。

人物再同定には 4.2 節で最も CMC Rank-1 の評価が高かった特徴抽出モデルと損失関数を用いる。つまり、特徴抽出モデルは ResNet [47]、損失関数は softmax cross entropy loss [61, 62] を用いる。また 4.3 節では Video Person Re-identification の手法を用いて、特徴抽出モデルに入力する画像は Input2 (図 4.1b) として実験を行った。

4.2 予備実験 (人物再同定法の検証)

人物再同定の手法の内、どの手法が保育施設の映像データに適しているかを検証する。本実験では実際のある保育施設の一部屋で撮影した 4 つの映像 (Video 1, Video 1', Video 2, Video 2'), を用いる。(Video 1, Video 1') と (Video 2, Video 2') は違う視点の映像である。Video 1 と Video 1' は同じ視点で同じ日に撮影した映像であるが、Video 1 に対し Video 1' の方が撮影時刻が遅い映像となっている。同様に Video 2 と Video 2' は同じ視点、同じ日に撮影した映像で、Video 2 に対し Video 2' の方が撮影時刻が遅い。これらの映像に対して、OpenPose [2] を用いて検出した人物を、DSST [40] を使って短時間 (最大で 1 秒間) 追跡したものを一つの人物映像とし、それに人物 ID を付けることで表 4.1 のようなデータセットを作成した。このデータセットでは短い人物映像に人物 ID を付けているため、Image Person Re-identification と Video Person Re-identification のどちらも使用可能である。表 4.1 に示すように、Video 1 では 15 人が登場したため人物 ID をそれぞれ 0~14 としたが、Video 1' では ID 12 の人物が登場しなかったため ID 0~11, ID 13, ID 14 の人物についてのみのデータセットとなっている。また、Video 2 と Video 2' では ID 0~10 の人物が登場し、この人物 ID は Video 1, Video 1' と共通である。本実験では表 4.1 の Dataset 1 を使用する場合は、学習用かつ Gallery 用に Video 1 のデータセット、テスト (Query) 用に Video 1' のデータセットを使用する。表 4.1 の Dataset 2 を使用する場合は、学習用かつ Gallery 用に Video 2 のデータセット、テスト (Query) 用に Video 2' のデータセットを使用する。これは学習に撮影時刻が早いデータを使うことで、実利用する際の状況と同様の状況を設定している。学習用のデータに比べてテスト用のデータの方がデータ数が少なくなっていることも実利用する際の状況と似せているためである。

表 4.1: 保育施設の映像を使った人物再同定法の検証用データセット

	映像	使用用途	画像枚数	登場人物 ID
Dataset 1	Video 1	学習用 (Galley)	2506	0~14
	Video 1'	テスト用 (Query)	7679	0~11, 13, 14
Dataset 2	Video 2	学習用 (Galley)	2173	0~10
	Video 2'	テスト用 (Query)	6040	0~10

以下の検証の評価には mean Average Precision (mAP) と Cumulative Matching Characteristics (CMC) を使用している。CMC とは人物再同定の研究でよく使用される評価手法であり、CMC Rank-k とは Gallery 画像から Query 画像に最も近い k 枚を選んできたとき、k 枚の中に Query 画像と同じ人物 ID の画像が含まれる割合を指す。

4.2.1 特徴抽出モデルの検証

表 4.1 の Dataset 1 を使用して、2.2.3 節で述べた特徴抽出モデルの内、どの特徴抽出モデルが保育施設の映像データに適しているかを検証する。検証する特徴抽出モデルは、画像分類用に提案された ResNet [47], ResNeXt [48], SENet [49], DenceNet [50], MobileNetV2 [51], SqueezeNet [52], および人物再同定用に提案された MuDeep [53], ResNet-mid [54], HACNN [55], PCB [56], MLFN [57] である。いずれの実験も学習率を 0.0003 としており、画像の入力サイズは HACNN [55] 以外では高さを 256 ピクセル、幅を 128 ピクセルとし、HACNN [55] では高さを 160 ピクセル、幅を 64 ピクセルとした。optimizer には AMSgrad [72] を使用し、学習 epoch 数を 100 とした。また、この検証では図 2.3 で示した Image Person Re-identification の手法で比較しており、損失関数を softmax cross entropy loss [61, 62] としている。特徴抽出モデルの入力画像は、全身が映っている図 4.1a のような画像とした。

結果は表 4.2 の通りである。CMC Rank-5 では MuDeep [53] が最も結果が良かったが、mAP と CMC Rank-1 とともに ResNet [47] が最も良い結果を出していることがわかった。2.2.3 節で述べた通り、保育施設の映像データでの人物再同定は一般的な人物再同定とは実験設定が異なるため、人物再同定用に提案された特徴抽出モデルでもあまり良い結果が得られていないといえる。CMC Rank-5 よりも CMC Rank-1 の結果が良い方が実利用に向いているため、特徴抽出モデルは ResNet [47] が最も適していると考えられる。

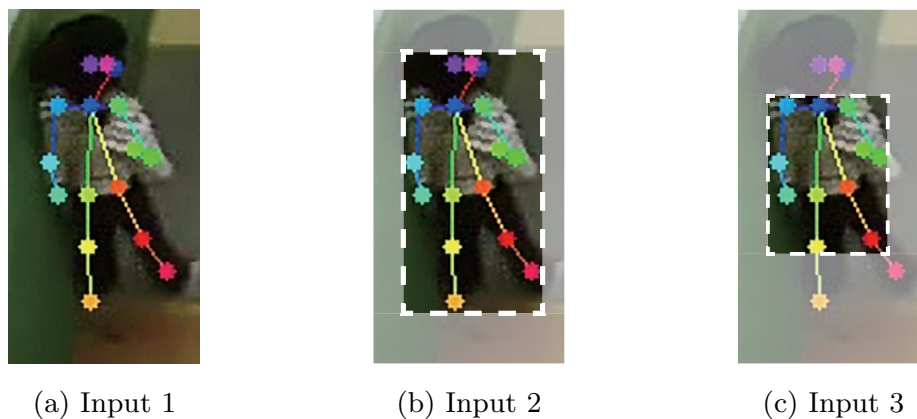


図 4.1: 特徴抽出モデルに入力する画像

表 4.2: 特徴抽出モデルの検証

特徴抽出モデル	mAP	CMC	
		Rank-1	Rank-5
ResNet [47]	68.4%	58.6%	78.5%
ResNeXt [48]	66.5%	54.5%	79.8%
SENet [49]	60.4%	49.0%	71.4%
DenceNet [50]	64.6%	50.8%	81.0%
MobileNetV2 [51]	60.0%	44.5%	77.3%
SqueezeNet [52]	49.5%	31.6%	72.2%
MuDeep [53]	60.6%	44.1%	81.5%
ResNet-mid [54]	65.1%	52.2%	79.6%
HACNN [55]	62.4%	46.8%	80.8%
PCB [56]	57.7%	42.0%	78.9%
MLFN [57]	58.0%	44.4%	72.5%

4.2.2 損失関数の検証

学習時の損失関数については、2.2.3 節で述べた通り softmax cross entropy loss [61, 62] と triplet loss [63] が提案されている。この検証では softmax cross entropy loss [61, 62]

表 4.3: 損失関数の検証

人物再同定法	損失関数	mAP	CMC	
			Rank-1	Rank-5
Image	softmax [61, 62]	68.4%	58.6%	78.5%
	triplet [63]	50.4%	32.2%	76.6%
	softmax [61, 62] + triplet [63]	69.0%	57.6%	83.4%
Video	softmax [61, 62]	61.8%	48.8%	78.8%
	triplet [63]	58.9%	45.7%	75.7%
	softmax [61, 62] + triplet [63]	57.9%	43.0%	78.6%

と triplet loss [63] の 2 つと, この 2 つを合わせたもの (softmax cross entropy loss [61, 62] + triplet loss [63]) の計 3 つを比較する. この検証でも表 4.1 の Dataset 1 を使用する. 特徴抽出モデルとして 4.2.1 節の検証で最もパフォーマンスが高かった ResNet [47] を使用した. 学習率, 画像の入力サイズ, optimizer は 4.2.1 節と同様である. 特徴抽出モデルの入力画像は, 全身が映っている図 4.1a のような画像とした. また, Image Person Re-identification と Video Person Re-identification の二つの人物再同定法について比較した.

表 4.3 に結果を示す. Image Person Re-identification では, mAP, CMC Rank-5 については損失関数を softmax cross entropy loss [61, 62] + triplet loss [63] としたものが最も良いが, CMC Rank-1 については softmax cross entropy loss [61, 62] のみが最も良かった. Video Person Re-identification では, mAP, CMC Rank-1, CMC Rank-5 のいずれも損失関数が softmax cross entropy loss [61, 62] のみのものが最も良い結果となった. この結果より, Video Person Re-identification では softmax cross entropy loss [61, 62] のみが最も優れており, Image Person Re-identification も CMC Rank-1 の精度に着目するのであれば softmax cross entropy loss [61, 62] のみが最も優れていると考えられる.

4.2.3 入力画像に関する検証

OpenPose [2] を用いた際, 一人毎に最大で 18 個の関節 (鼻, 首, 右肩, 右肘, 右手, 左肩, 左肘, 左手, 右腰, 右膝, 右足, 左腰, 左膝, 左足, 右目, 左目, 右耳, 左耳) の位置

が推定される．この推定された関節情報から特徴抽出モデルに入力する画像を調整することができる．この検証では特徴抽出モデルに入力する画像として，図 4.1a のように常に全身が写るような画像，図 4.1b のように推定された関節全てを含むような画像，図 4.1c のように推定された関節の内，首，右肩，右肘，左肩，左肘，右腰，右膝，左腰，左膝の 9 つの関節を含むような画像の 3 つのを比較する．この検証では表 4.1 の Dataset 1 と Dataset 2 を使用して検証する．特徴抽出モデルとして 4.2.1 節の検証で最もパフォーマンスが高かった ResNet [47] を用いた．損失関数は softmax cross entropy loss [61, 62] を使用した．学習率，画像の入力サイズ，optimizer は 4.2.1 節と同様である．また，この検証についても Image Person Re-identification と Video Person Re-identification の 2 つ人物再同定法について比較した．

結果を表 4.4，表 4.5 に示す．Dataset 1 と Dataset 2，また Image Person Re-identification と Video Person Re-identification のどちらについても，Input 1 に比べ Input 2 や Input 3 の方が結果が良いことがわかる．これは人物に対して背景部分が大きいと，特徴抽出モデルが背景部分に注目する可能性があるため誤認識が増えるからであると考えられる．ここでは実利用する上で最も重要な評価手法だと考えられる CMC Rank 1 の結果に注目する．Dataset 1 では Image Person Re-identification と Video Person Re-identification のどちらについても Input 2 の結果が最も良かったことがわかる．Input 2 において，Image Person Re-identification のよりも Video Person Re-identification の方が結果が良かった．一方で Dataset 2 での検証結果を見ると，Image Person Re-identification では Input 3 が最も結果が良く，Video Person Re-identification では Input 2 の結果が最も良かった．Dataset 1 の時とは異なり，Video Person Re-identification の Input 2 に比べて Image Person Re-identification の Input 3 の方が結果が良いことがわかる．

以上から，Video Person Re-identification を使用する場合はどちらのデータセットにおいても入力画像を Input 2 としたものが精度がよいことがわかる．また，Image Person Re-identification と Video Person Re-identification の内，どちらが適しているかはデータセットに寄ることがわかった．

表 4.4: 入力画像に関する検証 (Dataset 1)

人物再同定法	入力画像	mAP	CMC	
			Rank-1	Rank-5
Image	Input 1	68.4%	58.6%	78.5%
	Input 2	72.1%	61.8%	83.5%
	Input 3	72.9%	61.4%	86.2%
Video	Input 1	61.8%	48.8%	78.8%
	Input 2	74.6%	66.2%	83.2%
	Input 3	73.7%	61.8%	86.9%

表 4.5: 入力画像に関する検証 (Dataset 2)

人物再同定法	入力画像	mAP	CMC	
			Rank-1	Rank-5
Image	Input 1	55.7%	42.4%	70.8%
	Input 2	63.4%	52.6%	73.7%
	Input 3	68.1%	60.5%	72.6%
Video	Input 1	56.8%	44.2%	71.6%
	Input 2	67.5%	58.4%	75.2%
	Input 3	66.3%	57.3%	73.1%

4.3 実験

4.3.1 実験方法

実際の保育施設にカメラを設置して撮影した映像を使用した。使用したカメラは、環境分析で使用した Homeasure と同様にフューチャースタANDARD社が提供する映像解析プラットフォーム SCORER [4] を使用したカメラキット (図 4.2) である。SCORER [4] カメラキットは電源を接続するだけで撮影を開始し、インターネット環境に繋がっていれば映像データを自動的にクラウドに転送するという仕組みになっている。その為、データ回収の手間はかからずデータの一元管理が可能である。SCORER [4] カメラキットは小規

表 4.6: IoT カメラを設置した保育施設についての撮影情報

保育施設	撮影部屋数	撮影日数	カメラ台数
Nursery B	1	8	4
Nursery C	2	6	4
Nursery D	6	7	8
Nursery E	2	7	4
Nursery F	2	10	4
Nursery G	2	7	4
Nursery H	4	8	4
Nursery I	2	11	2
Nursery N	1	6	2



図 4.2: SCORER [4] カメラキット

模なデバイスであり，施設内において図 4.3 のように容易に設置でき，また容易に撤去することができる。

表 4.6 に IoT カメラを設置した保育施設について，撮影を行った部屋数，日数，カメラ台数をまとめた。9 施設 22 部屋で撮影を行い，合計で約 6,700 時間もの映像を取得した。本論文の実験は東京大学倫理委員会の承認を受け，各園において保育者・保護者らに趣旨説明等をした後に同意を得た上で行ったものである。



図 4.3: カメラ設置の様子

4.3.2 人物存在確率のヒートマップ

OpenPose [2] を用いて人物検出を行った結果を用いる。推定された関節全てを含む矩形の中心座標をその人物の代表点とする。各人の代表点がある位置を投票していくことで、画像中のどの部分に子供や先生がよく居たかがわかる。本実験では代表点を投票したものをヒートマップとして表現することで、画像中における人物の存在確率を可視化した。図 4.4 はある保育施設の映像を使って人物存在確率のヒートマップを表示した例である。図 4.4 ではある日の 15:00~16:00 に関しての画像中の人物存在確率を示している。人物の代表点の投票数が多い箇所ほど赤く、少ない箇所ほど青くなるように表示している。投票数が 0 の部分にはヒートマップの色が付いていない。図 4.4a ではテーブルや椅子のある辺りが赤くなっており、その部分に人が集中していたことがわかる。図 4.4b では画像の中央から奥にかけて広く人が居たことがわかる。図 4.4c では赤い部分は少ないが比較的中央にあるマットの上辺りに人が居たと推測される。図 4.4d では柱のすぐ右側に人が集中していたが、逆に柱の手前の部分には色がない箇所も多くあまり人が居なかったといえる。

4.3.3 動線解析

個々人を区別する必要がないのであれば、4.1.1 節の手法で全体的な動線解析が可能となる。4.3.2 節と同様に、推定された関節全てを含む矩形の中心座標をその人物の代表点とする。代表点の軌跡を可視化することで、画像中のどの部分を子供や先生がよく通って

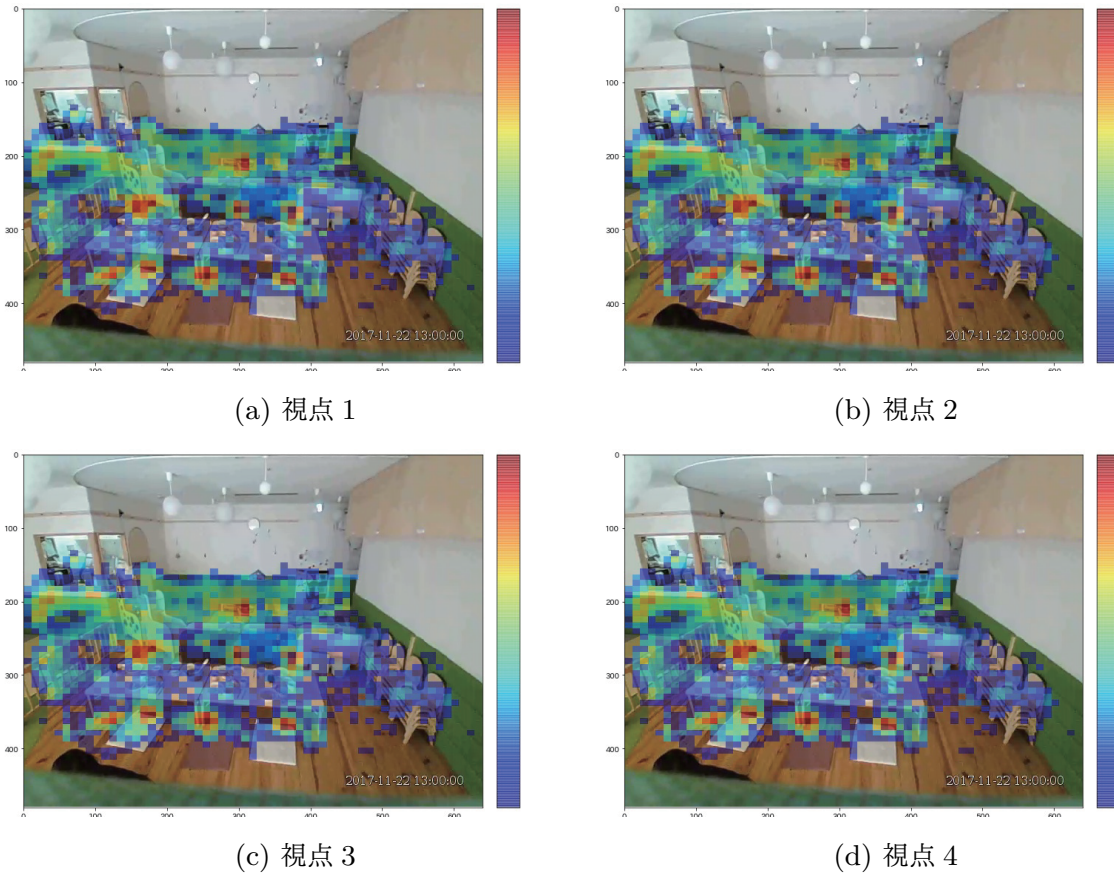


図 4.4: 人物存在確率のヒートマップ

いたかがわかる．図 4.5 はある保育施設の映像を使って動線を可視化した例である．ある日の 9:00～9:05 にかけて人が通った動線を水色の線で示している．図 4.5a では奥側や中央を横に人が通ったことがわかるが，手前側には人が来ていないことがわかる．図 4.5b では中央や右側によく人が通った形跡があり，左手前には人が通っていないことが推測される．図 4.5c を見ると，奥側とドア間を行き来した人や奥側と右手前側を行き来した人が居たといえる．図 4.5d では，柱の左右奥側では人が密集している一方で，柱の手前を通っている人や窓側にある本棚に向かって歩いている人がいたことがわかる．

4.3.4 運動量解析

4.1.1 節の手法で全体的な運動量解析が可能となる．ここでの運動量は各関節の移動量の総和を指す．全フレーム数が T の映像を対象にしたとき，フレーム t で仮 ID が $i \in I_t$

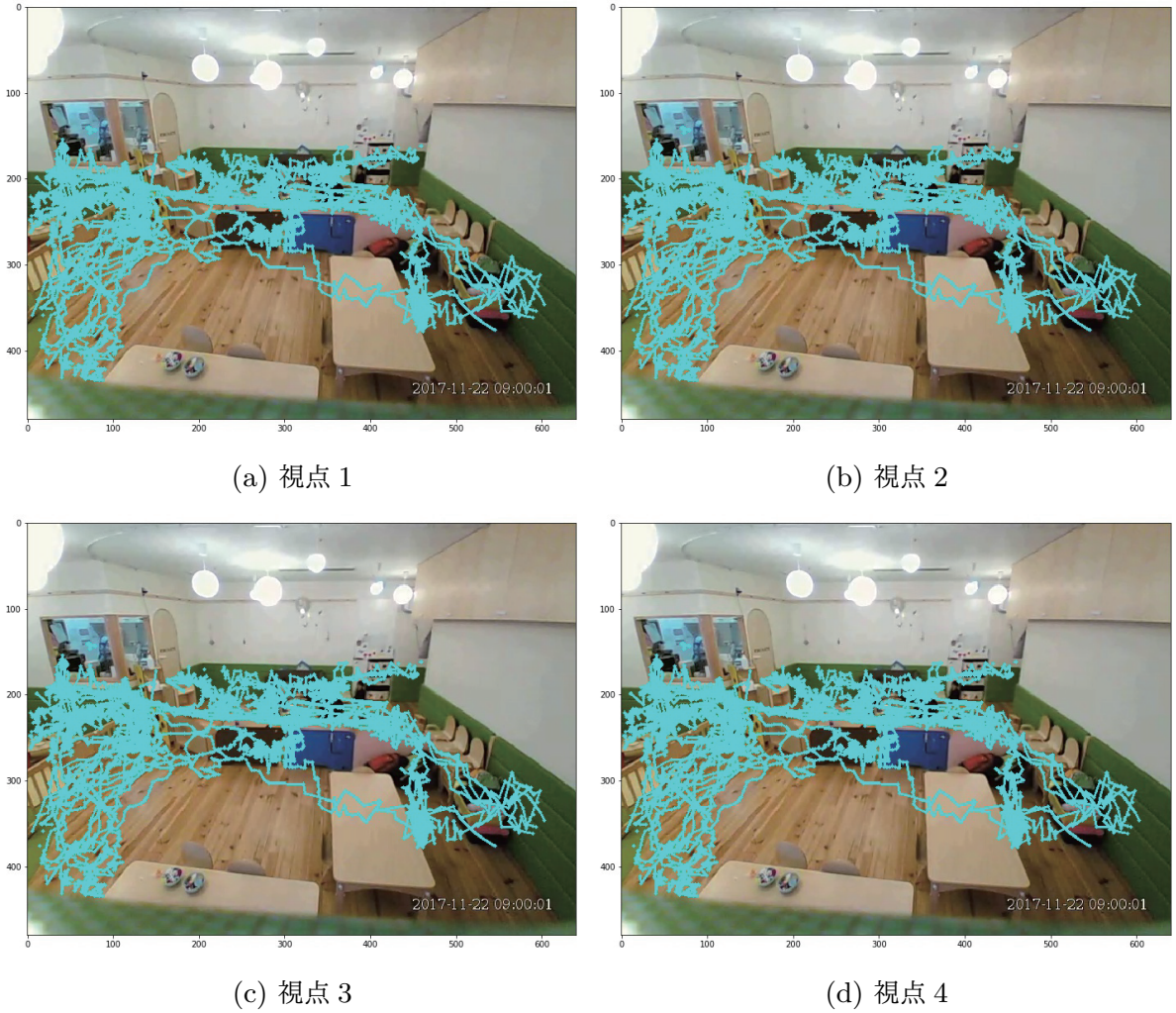


図 4.5: 動線解析

の推定された関節 $k \in K_{i,t}$ の座標を $p_{i,k,t}$ とおく．運動量 M は式 (4.3) のようにして求めた．

$$M = \sum_{t=0}^{T-1} \sum_{i \in I_t} \sum_{k \in K_{i,t+1} \cap K_{i,t}} \|p_{i,k,t+1} - p_{i,k,t}\|_2 \quad (4.3)$$

式 (4.3) では奥行きによる補正は行っておらず，人がカメラの近くにいるほど運動量が大きく出てしまう傾向にある．しかし，複数の視点の映像で平均値を取れば，ある程度補正が可能であると考えられる．ある保育施設の映像について運動量解析を行った例を

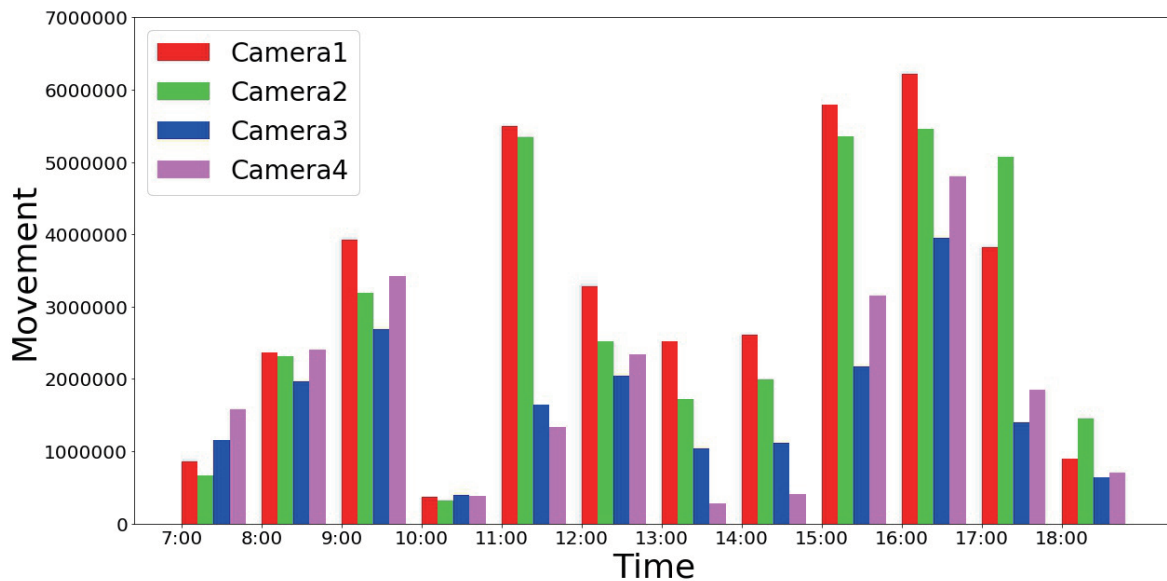


図 4.6: 運動量解析

図 4.6 に示す. 一つの部屋の 4 箇所カメラ (camera1~4) を設置して得られた映像を用いて運動量を求めた. 図 4.6 では 7:00~19:00 の間で 30 分毎に運動量を求めたものである. 13:00~15:00 は昼寝の時間のため運動量が小さいことや, 15:00~17:00 は遊ぶ時間のため運動量が多いことなどから, 図 4.6 は実際の状況に沿った結果であると考えられる. また, この部屋では環境分析も行っていた. 4 つの視点から得られた運動量のいずれも二酸化炭素濃度との相関係数が 0.7 以上であった. このことから, この部屋は密閉性が高く, 子供たちの運動が大きいとその分二酸化炭素濃度が高くなりやすい部屋である可能性がある.

4.3.5 個々人の行動傾向

4.1.2 節で述べた手法により, 子供および先生一人ひとりの動きを追跡できる. この実験では 4.2 節で使用した Video 1 と同じものを使っており, $mAP > 80\%$ である人物 (ID=1~4, 6~9) についてのみ注目する. 図 4.7 はある保育施設の子供が他の人物と近くにいた回数を求めたものである. ただし, ここでの「近くにいた」とは, 人物の代表点 (関節全てを含む矩形の中心座標) の距離が $approach_th$ ピクセル以内にいた場合を指し, 4.3.5 節では $approach_th = 100$ として統一している. 図 4.7, 図 4.8 において, CX は ID=X の子供を指しており, TX は ID=X の先生を指す. 図 4.7 を見ると, 人 (子供, 先生ともに) との接近回数は ID=3 の子供や ID=7 の子供が特に多いのに対し, ID=1 の子

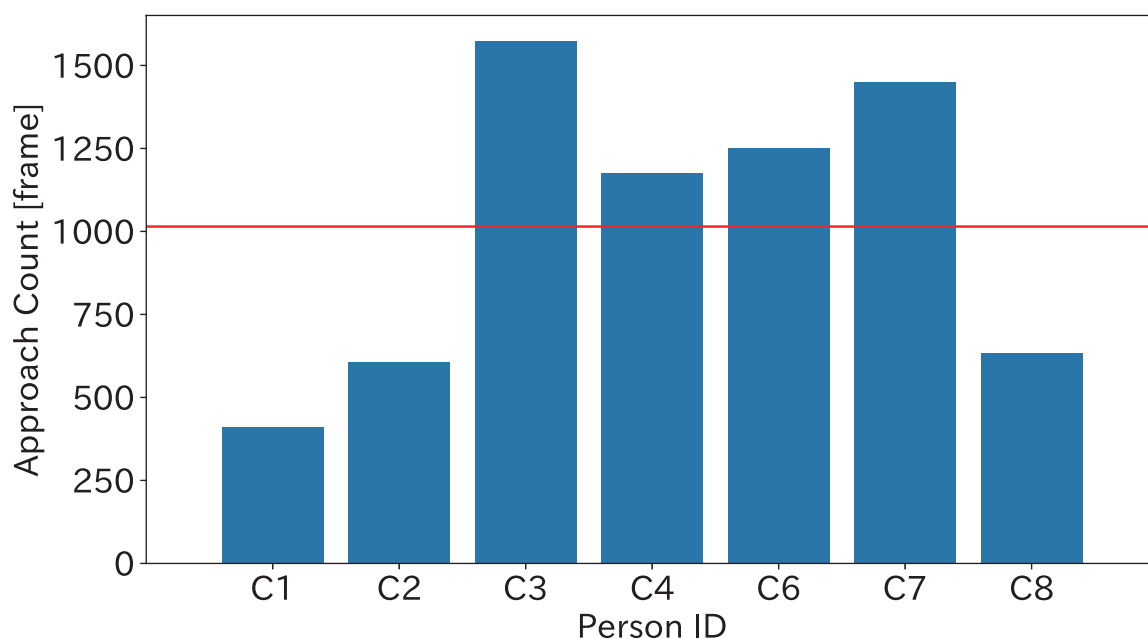
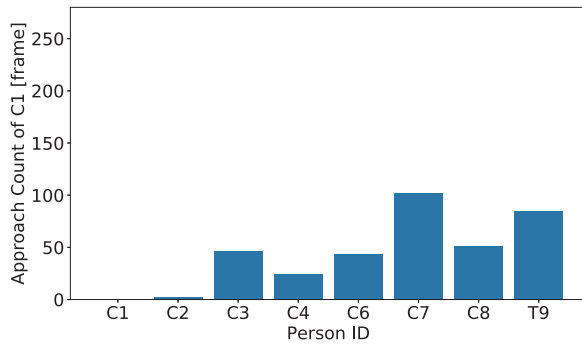


図 4.7: 各子供の接近回数 (赤線は平均値)

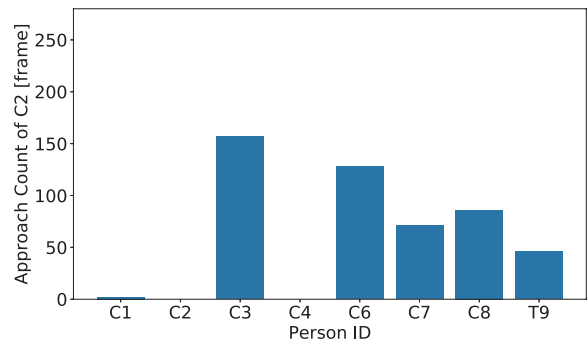
供は少ないことがわかる。図 4.8 は ID=1~4, 6~9 の人物について他の人物との接近回数を求めたものである。例えば、人との接近回数が比較的に少なかった ID=1 の子供は、図 4.8a より ID=7 の子供との接近回数が多かったことがわかる。ID=2 の子供は図 4.8b より、ID=3 の子供と接近回数が特に多い一方で、ID=1, 4 の子供とはほとんど接近がなかった。人との接近回数が比較的に多かった ID=3 の子供は、図 4.8c より特に ID=6 の子供や ID=9 の先生と多く接近していたことがわかる。ID=4 の子供は図 4.8d を見ると、ID=1~3, 8 の子供とは接近がほとんどなかったと言える。ID=6 の子供は図 4.8e より、ID=3, 4, 7 の子供との接近が多いことがわかる。ID=7 の子供は図 4.8f より ID=4, 6 の子供と多く接近していた。ID=8 の子供は図 4.8g より ID=3 の子供と接近回数が多く、ID=4 の子供とは接近回数は少なかった。また、図 4.8h からはどの子どもが ID=9 の先生とよく接近していたかがわかり、この例では ID=3, 4 の子供が特によく接近していたことがわかる。

また、長時間の人物追跡により個々人の動線についても解析できる。4.3.3 節と同様に人物の代表点の軌跡を求めることで、動線を可視化できる。ある保育施設の ID=0, 1, 5, 11 の子供に対し動線を可視化したものを図 4.9 に示す。ただし、図 4.9 は先ほどの接近回数を求めた保育施設とは異なる保育施設の映像を用いた結果であり、ID が対応する人物も異なる。図 4.9a を見ると、ID=0 の子供は部屋の広い範囲を途中止まりながら動い

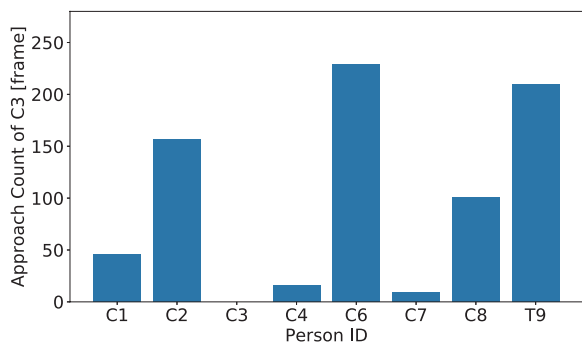
ていたことがわかる。図 4.9b から ID=1 の子供は部屋の中央を上下に動いていたことが読み取れ、図 4.9c からは ID=5 の子供は部屋の左側を広く移動していたことがわかる。ID=11 の子供は図 4.9d より、部屋の中央左側に集中して滞在していたと推定される。



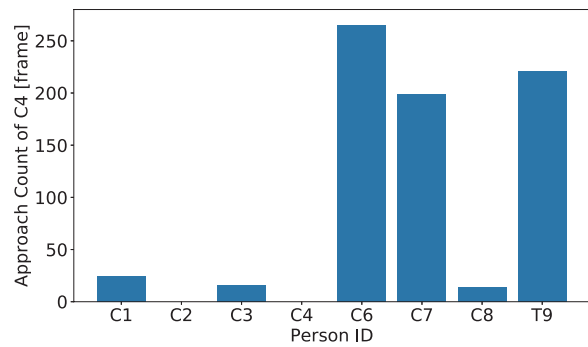
(a) ID=1 の子供の接近回数



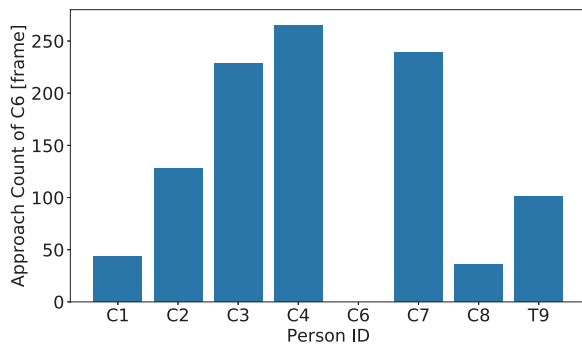
(b) ID=2 の子供の接近回数



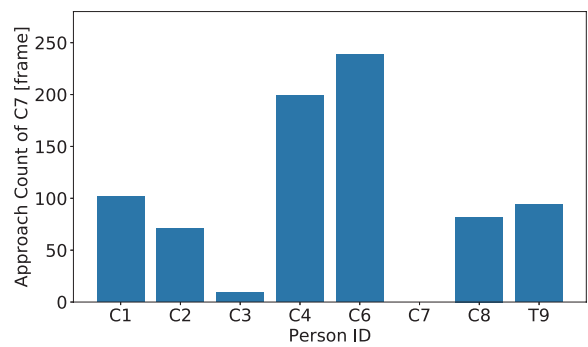
(c) ID=3 の子供の接近回数



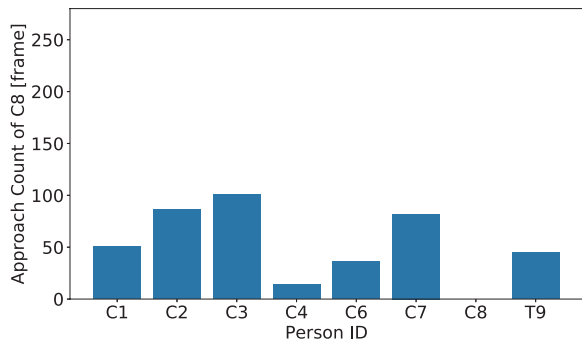
(d) ID=4 の子供の接近回数



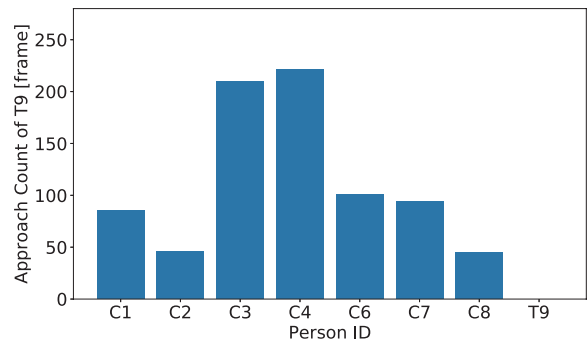
(e) ID=6 の子供の接近回数



(f) ID=7 の子供の接近回数



(g) ID=8 の子供の接近回数



(h) ID=9 の先生の接近回数

図 4.8: ID=1~4, 6~9 の人物³⁹についての他の人物との接近回数



ID=0



(a) ID=0 の子供の動線



ID=1



(b) ID=1 の子供の動線



ID=5



(c) ID=5 の子供の動線



ID=11



(d) ID=11 の子供の動線

図 4.9: 個々人の動線解析

第5章

結論

本論文では IoT があまり普及されていない例である保育施設に IoT デバイスを導入することを提案した。IoT デバイスで環境センシングを行うことで室内環境をリアルタイムに監視することができ、施設内の環境分析を行える。また、IoT カメラで撮影された映像を用いて子供たちの行動を分析できると考えられる。本論文では IoT カメラ・センサシステムを用いて、保育施設で環境分析と行動分析を行った。

環境分析では我々が独自に開発した IoT センサデバイス等を用いて、実際の保育施設で計測実験を行った。19 施設で計測を行ったところ、殆どの部屋 (98 %) で二酸化炭素濃度が厚生労働省 [14] の環境基準である 1000 ppm を超えた二酸化炭素濃度が観測されたことがわかった。湿度についても、冬場では全ての部屋で厚生労働省 [14] の基準である 40 % を下回る値を観測し、夏場でも全ての部屋について 70 % を超える値を観測していた。また、部屋によっては騒音レベルの平均値が 60 dB を超えていた部屋もあった。これらのことから、実際の保育施設において環境基準を満たさない二酸化炭素濃度、湿度などの環境値が観測されていることが確認された。子供や先生たちへの健康被害を防ぐためには環境値の監視の必要性があるといえる。室内の環境値を監視しやすいように、我々の開発したデバイス等で計測した環境値を図 3.5～図 3.7 のようにリアルタイムに可視化して表示するインタフェースを開発中である。

行動分析については、保育施設で撮影した映像に対して深層学習を用いた姿勢推定、追跡、人物同定などを用い、子供の行動を分析する手法を提案した。人物検出は保育施設のデータを使った追加の学習なしで高精度に人物検出が可能な OpenPose [2] を、物体追跡には処理時間が短い DSST [40] の手法を用いた。人物再同定では 4.2 節で検証した結果、特徴抽出モデルは ResNet [47]、損失関数は softmax cross entropy loss [61, 62] を使用したものが保育施設の映像に適していることがわかった。また、特徴抽出モデル

に入力する画像については、図 4.1a の Input 1 に比べて、背景部分が小さい図 4.1b の Input 2 や図 4.1c の Input 3 の形のものが精度が良いことがわかった。Video Person Re-identification については図 4.1b の Input 2 が、表 4.1 のいずれのデータセットに関しても CMC Rank 1 の結果が良かった。これらの手法を応用することで、人物存在確率のヒートマップ、動線、運動量、個々人の行動傾向の分析・可視化を実現した。

IoT カメラ・センサシステムを用いることで、室内環境をリアルタイムに監視することができ、子供たちに違和感を与えることなく行動を分析ができる。IoT センサデバイスでの環境計測により、異常な環境値を観測した時点で改善に向けて直ちに対処をすることができ、室内環境改善に役立たせることができる。また、IoT カメラで保育施設の映像を使って全体的な行動の様子や個々人の行動の傾向を分析することができた。IoT カメラ・センサシステムにより得られた統計データの蓄積により、保育に関する新たな研究への発展が期待できる。そして、保育の場における諸問題の解決に繋がっていくと考えられる。

参考文献

- [1] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [2] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, 2017.
- [3] Omron. Omron environment sensor. <http://www.Omron.co.jp/ecb/products/sensor/special/environmentsensor/>. (accessed 2017-03-05).
- [4] Future Standard. Scorer. <https://www.scorer.jp/>. (accessed 2018-01-22).
- [5] 厚生労働省. 平成 28 年 4 月の保育園等の待機児童数とその後. <https://www.mhlw.go.jp/stf/houdou/0000159045.html>, 2017. (accessed 2019-01-29).
- [6] 厚生労働省. 平成 29 年 10 月時点の保育園等の待機児童数の状況について. <https://www.mhlw.go.jp/stf/houdou/0000202678.html>, 2018. (accessed 2019-01-29).
- [7] 厚生労働省. 「子育て安心プラン」について. https://www.kantei.go.jp/jp/singi/syakaihosyou_kaikaku/dai7/shiryou7.pdf, 2017. (accessed 2019-01-29).
- [8] 内閣府子ども・子育て本部. 教育・保育施設等における事故報告集計」の公表について. <https://www8.cao.go.jp/shoushi/shinseido/outline/index.html>. (accessed 2019-01-29).
- [9] IHS TECHNOLOGY. The Internet of Things: A movement, not a market, 2018.
- [10] 経済産業省. スマートホームに関するデータ活用環. <http://www.meti.go.jp/meti.lib/report/H28FY/300137.pdf>, 2018. (accessed 2019-01-25).
- [11] 倉地亮. 自動車を取り巻く IoT セキュリティーとその課題: 自動運転, 車車間, 路車間の V2X に関するセキュリティーを中心に. 情報管理, Vol. 60, No. 10, pp. 690–700,

2018.

- [12] 貝原俊也. IoT 環境下の「考える工場」実現を目指す実仮想融合型生産システム. 計測と制御, Vol. 55, No. 1, pp. 53–58, 2016.
- [13] 山崎浩平ほか. 農業 ICT-IoT・ビッグデータ・AI 活用で農業を成長産業へ-: 7. AI/IoT がもたらす次世代農業-データを活用した植物病院 (R)-. 情報処理, Vol. 58, No. 9, pp. 814–817, 2017.
- [14] 厚生労働省. 建築物環境衛生管理基準について. <http://www.mhlw.go.jp/bunya/kenkou/seikatsu-eisei10/>. (accessed 2017-08-30).
- [15] 文部科学省. 学校環境衛生管理マニュアル 「学校環境衛生基準」の理論と実践 [平成 30 年度改訂版]. http://www.mext.go.jp/a_menu/kenko/hoken/1292482.htm. (accessed 2019-01-23).
- [16] 環境省. 騒音に係る環境基準について. <http://www.env.go.jp/kijun/oto1-1.html>. (accessed 2017-02-24).
- [17] Umweltbundesamt. Guidelines for indoor air hygiene in school buildings. <https://www.umweltbundesamt.de/sites/default/files/medien/publikation/long/4113.pdf>, 2008. (accessed 2018-03-19).
- [18] World Health Organization Regional Office for Europe. Environmental noise guidelines for the european region. <http://www.euro.who.int/en/health-topics/environment-and-health/noise/environmental-noise-guidelines-for-the-european-region>, 2018. (accessed 2019-01-23).
- [19] 白石君男. 聴覚に関わる社会医学的諸問題「学校教育における音響環境と聴覚補償」. *Audiology Japan*, Vol. 55, No. 4, pp. 207–217, 2012.
- [20] 志村洋子. 子ども研究子どもと保育者の音声コミュニケーション空間としての保育室. *チャイルド・サイエンス= Child science: 子ども学*, Vol. 11, pp. 22–27, 2015.
- [21] 河原紀子. 食事場面における 1~2 歳児の拒否行動と保育者の対応. *保育学研究*, Vol. 42, No. 2, pp. 112–120, 2004.
- [22] 坂井孝. デジタルビデオカメラの映像を用いた保育施設の食育の評価. *国際研究論叢: 大阪国際大学紀要*, Vol. 27, No. 3, pp. 181–186, 2014.
- [23] 永瀬祐美子, 倉持清美. 集団保育における幼児の生活習慣行動の習得過程: 「片付け場面」に着目して. *日本家政学会誌*, Vol. 62, No. 11, pp. 735–741, 2011.
- [24] 畠山美穂, 山崎晃. 幼児の攻撃・拒否的行動と保育者の対応に関する研究: 参与観察を通して得られたいじめの実態. *発達心理学研究*, Vol. 14, No. 3, pp. 284–293, 2003.
- [25] 新谷公朗, 金田重郎, 江守貞治ほか. 幼児行動記録作成システムへの取り組み-TV カ

- メラとパッシブセンサによる幼児の行動追跡. 情報処理学会研究報告情報システムと社会環境 (IS), Vol. 2003, No. 31 (2002-IS-083), pp. 73–80, 2003.
- [26] 富澤優, 河合純, 新谷公朗, 芳賀博英, 金田重郎. モーションセンサとビデオカメラを用いた室内用位置検出手法. 電子情報通信学会技術研究報告. KBSE, 知能ソフトウェア工学, Vol. 105, No. 208, pp. 7–12, 2005.
- [27] Netatmo. Netatmo personal weather station. <https://www.netatmo.com/product/weather/weatherstation>. (accessed 2017-02-01).
- [28] Acer. Acer air monitor. <https://home.cloud.acer.com/airmonitor/>. (accessed 2018-03-27).
- [29] Senstick. <http://senstick.com/>. (accessed 2018-03-27).
- [30] Awair. Awair glow. <https://getawair.com/pages/awair-glow>. (accessed 2018-03-27).
- [31] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, pp. 886–893. IEEE, 2005.
- [32] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.
- [33] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pp. 91–99, 2015.
- [34] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pp. 21–37. Springer, 2016.
- [35] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
- [36] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1653–1660, 2014.
- [37] Leonid Pishchulin, Eldar Insafutdinov, Siyu Tang, Bjoern Andres, Mykhaylo Andriluka, Peter V Gehler, and Bernt Schiele. Deepcut: Joint subset partition and

- labeling for multi person pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4929–4937, 2016.
- [38] Eldar Insafutdinov, Leonid Pishchulin, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. Deepercut: A deeper, stronger, and faster multi-person pose estimation model. In *European Conference on Computer Vision*, pp. 34–50. Springer, 2016.
- [39] David S Bolme, J Ross Beveridge, Bruce A Draper, and Yui Man Lui. Visual object tracking using adaptive correlation filters. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2544–2550. IEEE, 2010.
- [40] Martin Danelljan, Gustav Häger, Fahad Khan, and Michael Felsberg. Accurate scale estimation for robust visual tracking. In *British Machine Vision Conference, Nottingham, September 1-5*. BMVA Press, 2014.
- [41] Hyeonseob Nam, Mooyeol Baek, and Bohyung Han. Modeling and propagating cnns in a tree structure for visual tracking. *arXiv preprint arXiv:1608.07242*, 2016.
- [42] Hyeonseob Nam and Bohyung Han. Learning multi-domain convolutional neural networks for visual tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4293–4302, 2016.
- [43] Heng Fan and Haibin Ling. Sanet: Structure-aware network for visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2217–2224, 2017.
- [44] Martin Danelljan, Andreas Robinson, Fahad Shahbaz Khan, and Michael Felsberg. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In *European Conference on Computer Vision*, pp. 472–488. Springer, 2016.
- [45] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. Eco: Efficient convolution operators for tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6638–6646, 2017.
- [46] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. Person re-identification by local maximal occurrence representation and metric learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2197–2206, 2015.
- [47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning

- for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [48] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pp. 5987–5995. IEEE, 2017.
- [49] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.
- [50] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
- [51] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520. IEEE, 2018.
- [52] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.
- [53] Xuelin Qian, Yanwei Fu, Yu-Gang Jiang, Tao Xiang, and Xiangyang Xue. Multi-scale deep learning architectures for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5399–5408, 2017.
- [54] Qian Yu, Xiaobin Chang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. The devil is in the middle: Exploiting mid-level representations for cross-domain instance matching. *arXiv preprint arXiv:1711.08106*, 2017.
- [55] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *CVPR*, Vol. 1, p. 2, 2018.
- [56] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 480–496, 2018.
- [57] Xiaobin Chang, Timothy M Hospedales, and Tao Xiang. Multi-level factorisation net for person re-identification. In *CVPR*, Vol. 1, p. 2, 2018.

- [58] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1116–1124, 2015.
- [59] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152–159, 2014.
- [60] Yu Wu, Yutian Lin, Xuanyi Dong, Yan Yan, Wanli Ouyang, and Yi Yang. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5177–5186, 2018.
- [61] Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016.
- [62] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1249–1258, 2016.
- [63] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.
- [64] Texas Instruments. HDC1000 Low Power. 3% Accuracy Digital Humidity Sensor with Integrated Temperature Sensor. <http://www.ti.com/product/HDC1000>. (accessed 2017-03-05).
- [65] Adafruit. Adafruit TSL2561 Digital Luminosity/Lux/Light Sensor Breakout. <https://www.adafruit.com/product/439>. (accessed 2017-03-05).
- [66] SparkFun. SparkFun 9DoF IMU Breakout – LSM9DS1. <https://www.sparkfun.com/products/13284>. (accessed 2017-03-05).
- [67] Strawberry Linux. VEML6070 UV 紫外線センサ・モジュール. <https://strawberry-linux.com/catalog/items?code=16070>. (accessed 2019-01-23).
- [68] Senseair. K30. <https://senseair.com/products/flexibility-countsk30/>. (accessed 2018-03-24).
- [69] Raspberry Pi Foundation. Raspberry pi 3 model b. <https://www.raspberrypi.org/products/raspberry-pi-3-model-b/>. (accessed 2019-01-29).
- [70] Yuki Obuchi, Toshihiko Yamasaki, Kiyoharu Aizawa, Satoshi Toriumi, and Miki-hisa Hayashi. Measurement and evaluation of comfort levels of apartments using

iot sensors. In *Consumer Electronics (ICCE), 2018 IEEE International Conference on*, pp. 1–6. IEEE, 2018.

- [71] 大淵友暉, 山崎俊彦, 相澤清晴, 鳥海哲史, 林幹久. IoT センサを用いたマンション物件計測と快適度評価. 人工知能学会全国大会論文集 2017 年度人工知能学会全国大会 (第 31 回) 論文集, pp. 1H2OS15a4–1H2OS15a4. 一般社団法人 人工知能学会, 2017.
- [72] Sashank J Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of adam and beyond. 2018.

関連する発表文献

国際論文誌

- [1] (投稿予定) Yuki Obuchi, Toshihiko Yamasaki, Satoshi Toriumi, Mikihisa Hayashi, Sachiko Nozawa, Midori Takahashi, Toshihiko Endo and Kiyomi Akita, “Sensing for Living/School Environments Using IoT Sensors,” IEEE Transactions on Consumer Electronics.

国内会議

- [2] 大淵友暉, 山崎俊彦, 相澤清晴, 鳥海哲史, 林幹久, “IoT カメラを用いた保育施設における子育て見守り支援,” 第 20 回 画像の認識・理解シンポジウム (MIRU2017), Poster Session, PS1-56, 広島国際会議場, 広島県広島市, Aug. 7-10, 2017.
- [3] 大淵友暉, 山崎俊彦, 鳥海哲史, 林幹久, 野澤祥子, 高橋翠, 遠藤利彦, 秋田喜代美, “保育施設における IoT カメラを用いた環境・行動解析,” メディアエクスペリエンス・バーチャル環境基礎研究会 (MVE), 信学技報, vol. 117, no. 217, MVE2017-15, pp.7-11, 千葉大学, 千葉県千葉市, Sep.21-22, 2017.
- [4] 大淵友暉, 山崎俊彦, 鳥海哲史, 林幹久, 野澤祥子, 高橋翠, 遠藤利彦, 秋田喜代美, “IoT カメラによる保育施設での環境計測と行動分析 (Environment Measurement and Action Analysis in Nursery Schools using IoT Cameras),” 2017 年度映像メディア処理シンポジウム (IMPS 2017), P5-8, ラフォーレ修善寺, 伊豆市, 静岡, Nov. 20-22, 2017.
- [5] 山崎俊彦, 大淵友暉, “IoT カメラと機械学習を用いた保育環境の処理・理解,” 発達心理学会, 自主シンポジウム, 保育学研究における人工知能技術 (A I) の可能性 – エビデンスに基づく保育への挑戦 –, 東北大学川内キャンパス, 仙台市青葉区, Mar

23-25, 2018.

- [6] 大淵友暉, 山崎俊彦, 鳥海哲史, 林幹久, 野澤祥子, 高橋翠, 遠藤利彦, 秋田喜代美, “スマート保育への挑戦: センシング技術を活用した保育環境の調査,” Cedep 公開シンポジウム「豊かな人生を紡ぐ保育～Society5.0 保育から社会を変える～」, 東京都文京区, 東大・安田講堂, Aug. 05, 2018.
- [7] 山崎俊彦, 大淵友暉, 鳥海哲史, 林幹久, 秋田喜代美, 遠藤利彦, 野澤祥子, 高橋翠, “センシング技術を活用した保育環境の調査,” 東京大学大学院教育学研究科附属発達保育実践政策学センター主催 2017 年度公開シンポジウム『人生のはじまりを豊かに～乳幼児の発達・保育研究のイノベーション～』, 東京大学・安田講堂, 東京都文京区, Aug. 6, 2017.
- [8] 大淵友暉, 山崎俊彦, 鳥海哲史, 林幹久, 野澤祥子, 高橋翠, 遠藤利彦, 秋田喜代美, “スマート保育への挑戦: センシング技術を活用した保育環境の調査,” 公開シンポジウム「東アジアの保育・幼児教育改革～現職研修と評価を通じた保育者の専門性向上～」, 福武ホール, 東京都文京区, Jan. 22, 2018.
- [9] 山崎俊彦, 大淵友暉, “IoT カメラと AI・機械学習を用いた保育環境の処理・理解,” 日本赤ちゃん学会第 18 回 学術集会, 東京都文京区, 東大・工学部 2 号館, July 7-8, 2018.
- [10] 山崎俊彦, 大淵友暉, “IoT・AI 技術の保育への応用可能性,” Cedep 公開シンポジウム「豊かな人生を紡ぐ保育～Society5.0 保育から社会を変える～」, 東大・安田講堂, 東京都文京区, Aug. 5, 2018.
- [11] 大淵友暉, 山崎俊彦, 鳥海哲史, 林幹久, 野澤祥子, 高橋翠, 遠藤利彦, 秋田喜代美, “IoT カメラを用いた保育施設における多人数行動分析,” 画像工学研究会 (IE), 北海道大学, 北海道札幌市, Feb. 19-20, 2019.

その他の発表文献

国際会議

- [1] Yuki Obuchi, Toshihiko Yamasaki, Kiyoharu Aizawa, Satoshi Toriumi and Miki-hisa Hayashi. “Measurement and evaluation of comfort levels of apartments using IoT sensors.” In Proceedings of the 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, Jan. 12-14, 2018.

国内会議

- [2] 大淵友暉, 山崎俊彦, 相澤清晴, 鳥海哲史, 林幹久. “IoT センサを用いたマンション物件計測と快適度評価,” 人工知能学会全国大会 (JSAI), ウィンクあいち, 愛知県名古屋市, Mar. 23-26, 2017.

謝辞

山崎先生には学部4年生から3年間大変お世話になりました。頻繁に出張に行かれたりなどご多忙であるのにも関わらず、ほぼ毎週のようにミーティングをしていただきました。おかげで研究の相談をする機会が多くあったことにとっても感謝しております。研究についてあまり周りから評価が得られず不安になっていたときも、面白いデータが得られている良い研究だと励ましていただきました。そのときの言葉が心の支えとなり、修士2年生の最後まで研究を続けられたと思います。また、外付けHDやGPU等の研究資源も潤沢に用意して頂き、不自由なく研究することが出来ました。3年間ご指導して頂き、誠にありがとうございました。

相澤先生には研究室全体でのミーティング時に多くのアドバイスをいただきました。不出来な私でしたので何度も同じ指摘をしていただいていたかもしれませんが、大変勉強になりました。特に研究の最終的な目標をはっきりさせることの重要性を学びました。ありがとうございました。

学術支援職員の松林さんと江川さんには学会出張などの事務手続きをしていただき、とても感謝しております。また、研究室生活において大変お世話になりました。時折、気さくに話しかけてくださり嬉しかったです。

フューチャースタンドの鳥海さんと林さんにも学部4年生時の研究からお世話になりました。デバイス支援をしていただき、円滑に計測を行うことができました。またGPU環境をお貸ししてくださったこともありました。CEDEPの野澤先生、高橋先生、秋田先生、遠藤先生には保育施設側に話を通してくださったりなど、保育施設での実験に大変なご協力していただきました。

同期の小川くん、合田くん、高田くん、築山くん、中村くん、成田くん、Yiweiさんにはとてもお世話になりました。よく昼ご飯を一緒に食べに行ったり、度々仲良し会を開いたりなどして研究室生活を楽しく過ごすことができました。いつでも気さくに話しかけられる仲良し同期たちが居てくれたおかげで、ここまでやってこれたと思います。

また、研究室の先輩と後輩の皆様にも感謝しております。とても優秀な方ばかりで常に刺激を受けておりました。汪さんには山崎先生とのミーティングの際に様々なアドバイスをいただきました。研究室のソファには研究の合間や泥酔時に何度も休ませていただきました。おかげで快適な睡眠をとることができました。ありがとうございました。

最後に大学院まで通わせてくれた両親に感謝を申し上げます。

2019年1月30日

大瀬 友暉