Vol.2022-GN-115 No.20
Vol.2022-CDS-33 No.20
Vol.2022-DCC-30 No.20
2022/1/20

人物検出と姿勢推定の組み合わせに基づく屋内全周魚眼画像に対する人物間の密接度推定方式

古宮嗣朗[1]　秋田悠河[1]　阿倍博信[1]

概要：新型コロナウイルスの流行により，感染対策として密閉・密集・密接の回避が求められている．そのため，映像解析技術を用いて人物間の密集および密接を監視するシステムが必要とされている．そこで，本研究では人物間の距離と顔の向きに基づき密接度をモデル化するとともに，入力された全周魚眼画像から人物検出と姿勢推定を行い，その結果を元に人物間の密接度を推定する方式，および有効性の評価について提案する．具体的には，SSD を用いて人物の検出を行った後，img2pose を利用して人物の顔向きを推定，加えて独自に構築した MLP を利用して，画像上の人物の足元から実空間上の足元座標を推定する方式である．

キーワード：映像監視システム，人物検出，顔向き検出，密接度推定

An interpersonal closeness estimation method for indoor fulldome fisheye images based on combination of person detection and person posture estimation

SHIRO KOMIYA[†1]　YUGA AKITA[†1]
HIRONOBU ABE[†1]

Abstract: Due to the outbreak of a new coronavirus, there is a need to avoid hermetic sealing, crowding, and close proximity as a countermeasure against infection. In this study, we modeled the degree of closeness based on the distance and orientation of faces. In this paper, we propose a method for estimating the degree of closeness between people based on the distance between people and the orientation of their faces, person detection and pose estimation from the input fisheye image, and evaluation of the effectiveness of the method. After detecting the person using SSD, the face orientation of the person is estimated using img2pose, and the feet coordinates in real space are estimated from the feet of the person in the image using an originally constructed MLP.

Keywords: Video surveillance system, Person detection, Person posture estimation, Closeness estimation

1. はじめに

昨今，SARS-CoV-2 ウイルスによって引き起こされる新型コロナウイルス感染症（COVID-19）が世界各国で流行しており，社会問題となっている．COVID-19 による症状のほとんどは軽度から中度であり，特別な治療を受けずに回復するとされている．その一方で，重症化して医療機関での治療が必要になるケースも存在している．

World Health Organization によれば，感染経路として，感染者の咳やくしゃみ，話す，歌う，息をする際に，小さな液体の粒子となって口や鼻から拡散する可能性が指摘されている[1]．そのため，COVID-19 感染者の近くにいるときにウイルスを吸い込んだり，汚染された物に触れてから自分の目，鼻，口に触れたりすると，ウイルスに感染する可能性がある．このウイルスは，屋内や人が多い環境で拡散しやすくなるとされており，密の回避が求められている．

人物間の密を監視する技術として，日本電気株式会社が発表しているソーシャルディスタンシング判定技術[2]や山地らによる群衆密度推定方式の研究[3]がある．しかし，こうしたシステムでは人物間の密を測定，可視化しているものの，人物の姿勢や状況は考慮されておらず，ウイルス感染の危険性を見落とす可能性があると考えられる．

本研究では，人物間の距離と顔の向きに基づき密接度をモデル化するとともに，入力された全周魚眼画像から人物検出と姿勢推定を行い，その結果を元に人物間の密接度を推定する方式，および有効性の評価について報告する．

以下，2 章で関連研究を紹介し，3 章で予備実験と実験結果の考察を行う．さらに，4 章で密接度推定方式を提案し，5 章で評価，6 章で考察したのち，7 章で本研究のまとめを行う．

2. 関連研究

2.1 ソーシャルディスタンシング判定技術

前述した日本電気株式会社が発表している技術は，公共施設や店舗などの人が集まる場において，人と人が十分な距離を保てているかをリアルタイムに判定して，ソーシャルディスタンシングを可視化することが可能である．従来のカメラ画像からの距離測定技術では，事前に撮影範囲内

1 東京電機大学
Tokyo Denki University

Vol.2022-GN-115 No.20
Vol.2022-CDS-33 No.20
Vol.2022-DCC-30 No.20
2022/1/20

のさまざまな地点にマーカーを置いて測定を行う必要があり，時間やコストの面で課題がある．この技術では，既設カメラの映像を解析し，人々の半径 1m の円にお互いが接触していないかどうかを解析して密集度合いを可視化・数値化することが可能である．技術の利用方法として，来場者へ情報提供と施設スタッフ側のリスク管理や安全確保が提案されている．

2.2 合成データセットの作成と利用

Tobias らは THEODORE という合成画像データセットを作成し評価を行っている[4]．これは，3D ゲームエンジンである Unity[5]を用いて作成した 10 万枚の屋内魚眼画像を 16 クラスに分けて収録したデータセットであり，魚眼画像に加えてセマンティックセグメンテーション用のアノテーションデータや，物体検出用のインスタンスマスクとバウンディングボックスを収録している．このデータセットを用いて，6 つのクラスに対して独自の評価スイート（FES）における物体検出とセマンティックセグメンテーションを評価している．物体検出では 0.61 の mAP，セグメンテーションでは，すべてのクラスで 0.36 の mIoU を達成している．

3. 予備実験

3.1 基本方針

関連研究を踏まえ，予備実験の方針を下記のとおり整理する．

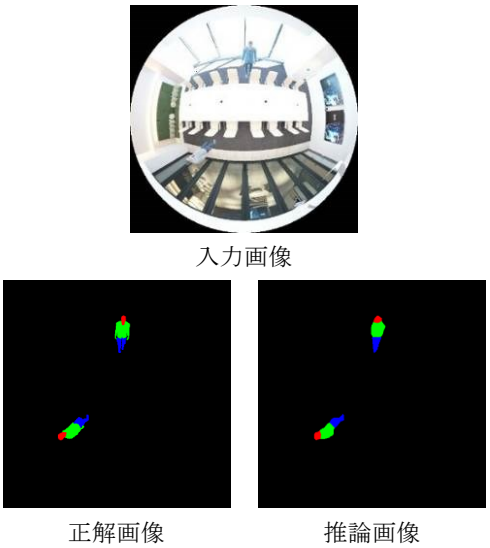
- 魚眼画像データから直接密接度推定を行うため，画像処理技術による人物検出について評価する．
- 3D ゲームエンジンを用いて，学習データの収集および作成のコスト削減を図る．

3.2 セマンティックセグメンテーションの学習

基本方針に基づき，画像中全ての画素に対してクラスラベルを予測することが可能なセマンティックセグメンテーションに着目，予備実験を行った[6]．セマンティックセグメンテーションの学習には，通常の画像に加えて認識対象を単色で塗りつぶしたマスク画像が必要である．そのため，Unity を用いて，全周魚眼画像とマスク画像のペアをそれぞれ 2,220 枚生成した．このとき，マスク画像を背景，人物の頭部，胴体，下半身を別の色に塗りわけ，4 クラスに分類することで，各人体部位を検出できるようにした．アルゴリズムには PSPNet[7]を選択し，学習データ 925 枚，検証データ 925 枚，バッチサイズ 6，30 エポックとして，ADE20K データセット[8]による学習済みモデルを使用してファインチューニングを行った．

3.3 セマンティックセグメンテーションの評価

学習したモデルにテストデータを 370 枚入力して，mIoU の数値を評価した．図 1 に実行例，表 1 に評価結果を示す．



入力画像　　正解画像　　推論画像

図 1 セマンティックセグメンテーションの実行例
Figure.1 Example of Semantic Segmentation Execution

表 1 セマンティックセグメンテーションの評価
Table.1 Evaluating Semantic Segmentation

	背景	頭	上半身	下半身
mIoU	0.98	0.97	0.90	0.74

3.4 考察

データセット作成における Unity の利用により，画像収集やアノテーションデータ作成のコストを削減できた．特に 3D モデルの色分けによりマスク画像を容易に生成できることがわかった．加えて，マスク画像生成の過程で物体検出用のバウンディングボックスのアノテーションデータの作成を自動化することができた．

セマンティックセグメンテーションについては，高い mIoU を実現した一方で，他の部位に比べて下半身の認識率が低くなることや，人物が複数人いる場合に区別ができないことが課題であると考えた．そこで，密接度推定方式における人物検出にはセマンティックセグメンテーションではなく，バウンディングボックスによる物体検出技術を用いることにした．

4. 密接度検出方式の提案

4.1 密接度のモデル化

図 2 に本論文にて提案する密接度検出方式の概要について示す．本研究では，2 人 1 組のペアにおいてお互いの顔向きを考慮した人物間の距離を密接度として定義する．人物が 1 人しかいない場合は密接度を定義しない．このとき，双方が向き合う場合は距離と 0.8 の積を，一方のみ向き合っている場合はそのままの距離を，双方とも向き合っていない場合が距離と 1.2 の積を密接度とする．向き合いの判定については 4.2 節に記述する．

Vol.2022-GN-115 No.20
Vol.2022-CDS-33 No.20
Vol.2022-DCC-30 No.20
2022/1/20

4.2 密接度推定方式の概要

(1) SSD の概要

SSD とは Wei らによって考案された，単一のディープニューラルネットワークを使用して画像内の物体を検出するアルゴリズムである[9]．SSD には CNN の複数の層から物体のバウンディングボックスを出力する特徴がある．これにより，様々な解像度からなる複数の特徴マップの予測を組み合わせて，様々な大きさの物体の検出を実現する．

(2) img2pose の概要

img2pose とは Vítor らによって開発されたリアルタイムの 6 自由度（6DoF），3D 顔ポーズ推定アルゴリズムである[10]．学習が容易かつ高速である特徴を持ち，顔検出なしで画像中すべての顔の 6DoF ポーズを回帰することができる．顔の 6DoF 推定は顔の特徴点検出よりも単純な問題とされており，3D 顔の位置合わせに使用される．

(3) 密接度推定方式の処理内容

以下，処理内容について説明する．

I. 画像入力および前処理

一台のカメラで撮影された全周魚眼画像を入力として受け取り，リサイズと色情報の正規化を行う．

II. 人物検出

前処理後の画像を SSD による人物検出モデルに入力し，人物のバウンディングボックス領域を推定する．人物が n 人いる場合は，${}_nC_2$ 個分のペアを作成する．

III. 顔向き検出

全周魚眼画像から人物のバウンディングボックス領域毎に画像を切り取る．この画像を img2pose に入力し，推定された 6DoF 情報のうち Yaw 角を取得する．入力の際，検出精度を向上させるために顔の位置が垂直になるように画像を回転させる．また，取得した Yaw 角に対して事前に回転させた量だけ補正を加える．

IV. 人物座標検出

バウンディングボックスの頂点および各辺の中点の計 8 点のうち，画像の中心に近いものを全周魚眼画像における人物の足元座標として取得する．この座標を座標検出モデルに入力し，実空間における足元座標を推定する．この足元座標からペア毎に人物間の距離を取得する．

V. コサイン類似度計算

ペア毎に人物の重心を結ぶ法線を求め，ベクトル化する．加えて，III で取得した顔の Yaw 角をベクトル化する．さらに，法線ベクトルと顔向きベクトルのコサイン類似度を求める．

VI. 密接度推定

ペア毎にコサイン類似度の値から顔の向き合いを判定する．法線ベクトルと顔向きベクトルのなす角がどちらも 30 度以下のときは双方が向き合っているとし，どちらかのみ 30 度以下のときは一方が向き合っているとする．それ以外のときはどちらも向き合っていないものとする．向き合い判定の結果をもとに，IV で求めた距離に 4.1 節に記述した手順に従う形で補正し密接度を数値として出力する．さらに，人物間の重心を結ぶ法線を入力画像に重ねて表示し，密接度を可視化する．密接度が 1m 以内であれば赤色，それ以外は青色の法線を表示する．

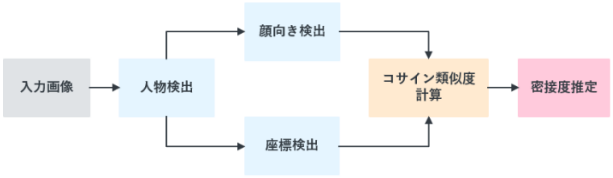
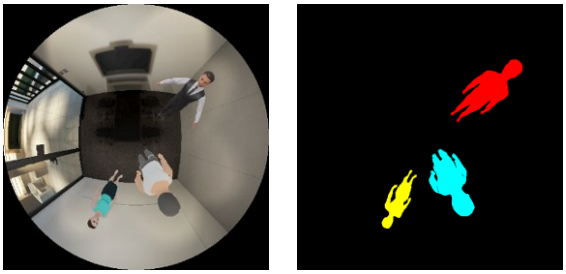


図 2 密接度推定方式の概要

Figure.2 Overview of the Closeness Estimation Method

4.3 合成屋内全周魚眼データセットの利用

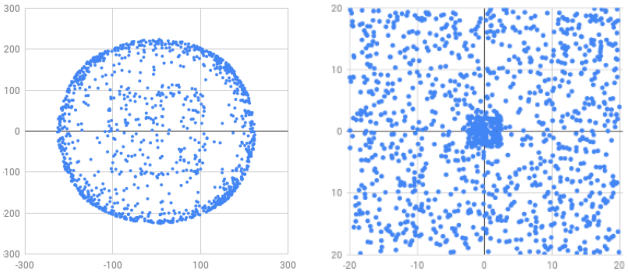
4.2 節にしたがって各推論モデルを作成するため，大量の学習データが必要となる．しかし，前述の THEODORE には人物検出用データは含まれているが，人物座標検出用データは含まれていない．そこで，筆者らが作成した合成データセット[11]を利用した．このデータセットには，Unity を用いて作成された画像データおよびアノテーションデータが収録されており，屋内全周魚眼画像における人物検出モデルやセグメンテーションモデル，人物座標検出モデルの学習に利用することができる．収録データの例を図 3 と図 4 に示す．



魚眼画像　　マスク画像

図 3 人物検出およびセグメンテーション用データの例

Figure.3 Example of data for person detection and segmentation



画像上の人物座標　　空間上の人物座標

図 4 人物座標検出用データの例

Figure.4 Example of data for detecting person coordinates

Vol.2022-GN-115 No.20
Vol.2022-CDS-33 No.20
Vol.2022-DCC-30 No.20
2022/1/20

4.4 人物検出モデルの作成

SSD による人物検出モデル生成のため，4.3 節で示したデータセットから画像とそれに対応するバウンディングボックス情報の 896 組を教師データとして抽出し，学習データとして 796 組，検証データとして 100 組を使用した．また，深層学習フレームワークとして，PyTorch[12]を利用した．最適化アルゴリズムとして Adam[13]を選択して，バッチサイズ 12, エポック数 100 に設定してモデルを生成した．学習曲線を図 5 に示す．

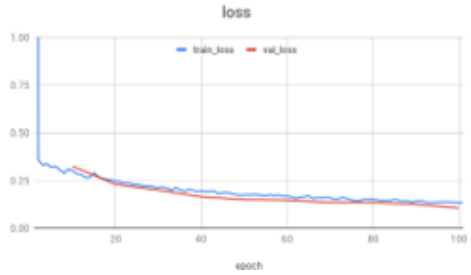
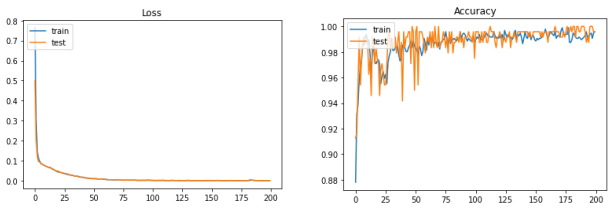


図 5 人物検出モデルにおける学習曲線
Figure.5 Learning Curve for Human Detection Model

4.5 座標検出モデルの生成

座標検出モデル生成のため，4.3 節で示したデータセットから屋内全周魚眼画像における人物の足元座標と 3D ゲームエンジンの空間における足元座標のペア 1,200 組を教師データとして抽出して，学習データとして 960 組，テストデータとして 240 組を使用した．この屋内全周魚眼画像における座標の単位はピクセル，ゲームエンジンにおける座標の単位はメートルである．また，機械学習フレームワークとして scikit-learn[14]と Keras[15]を利用，Liner Regression，Bagging，Random Forest，4 層からなる MLP の各モデルを生成した．このとき，MLP 以外のアルゴリズムにはデフォルトのパラメータを設定した．MLP には最適化アルゴリズムとして Adam を選択，バッチサイズ 128，エポック数 200 に設定してモデルを生成した．MLP における学習曲線を図 6 に示す．



Loss の変化　　Accuracy の変化
図 6 座標検出モデルにおける学習曲線
Figure.6 Learning Curve in Coordinate Detection Model

5. 評価

5.1 人物検出モデル，座標検出モデルの評価

4.4 節および 4.5 節で生成したモデルの有用性を確認するため精度評価を行った．人物検出モデルには，テストデータとして人物が 1 人写っている画像を 100 枚入力して人物検出を行い，mIoU を評価した．評価結果を表 2 に示す．つぎに，座標検出モデルでは各モデルにおける MAE と MSE，R^2 を評価した．評価結果を表 3 に示す．

表 2 人物検出モデルの評価
Table.2 Evaluation of Person Detection Models

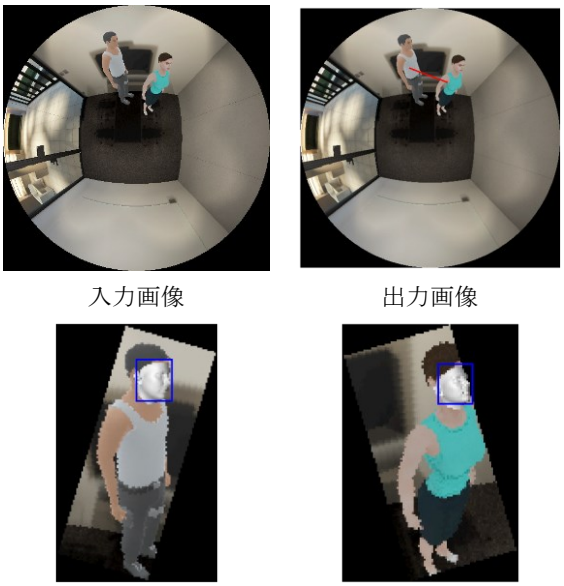
画像中の人数	mIoU
1	0.79

表 3 座標検出モデルの評価
Table.3 Evaluation of Coordinate Detection Models

Algorithm	MAE	MSE	R^2
Liner Regression	(2.51，2.64)	10.36	0.91
Bagging	(0.78，0.86)	1.31	0.98
Random Forest	(0.68，0.73)	1.03	0.99
MLP	(0.17，0.14)	0.05	0.99

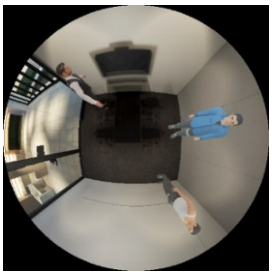
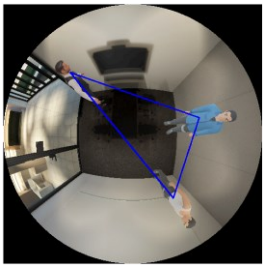
5.2 密接度推定方式の評価

生成した人物検出モデルおよび MLP による座標検出モデルを用いて，2 枚の画像に対して密接度推定を行った．入力画像と出力画像，人物および顔向きの検出結果を図 7 と図 8 に示す．また，密接度推定結果をそれぞれ表 4 と表 5 に示す．

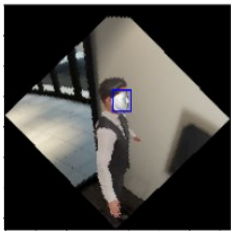


入力画像　　出力画像
人物および顔向きの検出結果
図 7 人物 2 名のときの密接度推定例
Figure.7 Example of closeness estimation for two persons

Vol.2022-GN-115 No.20
Vol.2022-CDS-33 No.20
Vol.2022-DCC-30 No.20
2022/1/20

入力画像　　出力画像


人物および顔向きの検出結果

図 8 人物 3 名のときの密接度推定例

Figure.8 Example of closeness estimation for three persons

表 4 人物 2 名のときの密接度推定結果

Table.4 Closeness estimation results for two persons

ペアの組	推定距離	顔の向き	密接度
ペア 1	0.41 m	一方が向き合う	0.41m

表 5 人物 3 名のときの密接度推定結果

Table.5 Closeness estimation results for three persons

ペアの組	推定距離	顔の向き	密接度
ペア 1	2.61 m	双方向き合わない	3.13m
ペア 2	2.30m	一方が向き合う	2.30m
ペア 3	1.16m	双方向き合わない	1.40m

6. 考察

人物検出モデルでは 0.79 の mIoU を達成した一方で，バウンディングボックスのサイズが大きくなりすぎるケースが存在した．これは，学習が収束していないか，学習データが十分でない可能性が考えられる．また，学習データの調光や反射などの条件が固定であるため，異なる条件の画像を入力すると検出精度の低下が予想される．そのため，様々な条件を追加したデータを用いて学習を行う必要があると考えられる．

座標検出モデルでは非線形データに強いモデルの精度が高く，MLP によるモデルでは正解値との誤差 20cm 以内を達成している．密接度推定方式の評価に用いた画像データでは，床の辺の長さが縦横ともに 3m ほどであるため，MLP によるモデルであれば実用上の問題は少ないと考えられる．

また，密接度の推定については，これらのモデルを用いることで，人物の顔の向きに考慮して密接度が推定できることを確認した．一方，提案方式における足元座標取得方法は人物の位置によって誤差が大きくなるため，場合によっては人物間距離の推定精度低下が予想される．そのため，画像と画像上および実空間上の人物座標をペアにしたデータセットを作成して，足元座標の取得精度を評価する必要があると考えられる．また，人物がカメラに対して背を向けている場合は顔向きが検出できないため，どういう扱いをすべきか検討しなければならない．

7. おわりに

人物間の距離と顔の向きに基づく密接度のモデル化および，入力された全周魚眼画像から人物間の密接度を推定する方式ついて提案した．その結果，課題はあるものの，屋内全周魚眼画像に対する人物間の密接度検出方式の有効性を確認できた．今後の課題として，推論モデルの精度向上も含めた密接度推定方式の改善と評価などがあげられる．

参考文献

[1] WHO: Coronavirus disease (COVID-19), https://www.who.int/health-topics/coronavirus#tab=tab_1, (参照 2021-12-11).
[2] NEC: 画像解析で人の密集度合い(ソーシャルディスタンシング)をリアルタイムに可視化する技術を開発, https://jpn.nec.com/press/202006/20200611_03.html, (参照 2021-12-11).
[3] 山地雄士，柴田智行：教師データの誤差に頑健な群集密度推定の学習手法，第 26 回画像センシングシンポジウム，IS3-12 SO3-12 (2020).
[4] Tobias Scheck. et al: Learning from THEODORE: A Synthetic Omnidirectional Top-View Indoor Dataset for Deep Transfer Learning, WACV 2020, arXiv:2011.05719 (2020).
[5] Unity: Unity, https://unity.com, (参照 2021-12-11).
[6] 古宮嗣朗，秋田悠河，阿倍博信：セマンティックセグメンテーションを用いた屋内全周魚眼画像に対する人物の密接度検出方式の検討，情報処理学会第 83 回全国大会 (2021).
[7] Hengshuang Zhao. et al: Pyramid Scene Parsing Network，IEEE Conference on Computer Vision and Pattern Recognition (2017).
[8] Bolei Zhou. et al: Scene Parsing Through ADE20K Dataset, IEEE Conference on Computer Vision and Pattern Recognition, pp. 633-641(2017).
[9] W. Liu, D. et al: SSD: Single Shot MultiBox Detector, In Proc. European Conf. on ECCV, pp.21-37, Vancouver, Canada (2016).
[10] V´ıtor Albiero. et al: img2pose: Face Alignment and Detection via 6DoF, Face Pose Estimation, CVPR, 2021, arXiv:2012.07791.
[11] 秋田悠河，阿倍博信，古宮嗣朗：3D ゲームエンジンを用いた映像監視向け合成全周魚眼画像データセットの作成と評価，第 20 回情報科学技術フォーラム (2021).
[12] Adam Paszke. et al: PyTorch (2016).
[13] D Kinga, J Ba Adam. et al: A method for stochastic optimization. In International Conference on Learning Representations (2015).
[14] David Cournapeau.: scikit-learn (2007).
[15] Chollet, F. et al: Keras (2015).