

# 人物検出と姿勢推定による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<sup>†1</sup> 秋田悠河<sup>†1</sup> 阿倍博信<sup>†1</sup>  
東京電機大学<sup>†1</sup>

## 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が世界各国で流行しており, 社会問題となっている. WHO 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おり, 密の回避が求められている[1].

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ス判定技術[2]などがある. 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.

## 2. 密接度検出方式の提案

### 2.1 密接度のモデル化

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は距離と1.2の積を密接度とする. 向き合いの判定については2.2節に記述する.

### 2.2 密接度推定方式の概要

図1に示す密接度推定方式について説明する.

#### (1) 画像入力:

1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.

#### (2) 人物検出:

前処理後の画像をSSD[3]による人物検出モデルに入力し, 人物のBounding Box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.

#### (3) 顔向き検出:

全周魚眼画像から人物のBounding Box領域毎に画像を切り取る. この画像を

#### (4) 座標検出:

Bounding Boxの頂点および各辺の midpoint 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.

#### (5) 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, (3)で取得した顔のYaw角をベクトル化する. さらに, 法線および顔向きベクトルのコサイン類似度を求める.

#### (6) 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, 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, 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, (4)で求めた距離に2.1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.

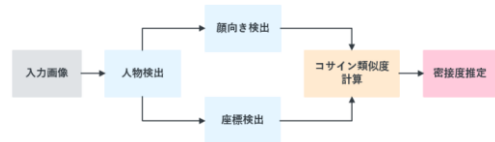


図1 密接度推定方式

### 2.3 CGによる屋内全周魚眼画像データセットの利用

2.2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. そこで, 筆者らが作成したデータセットを利用した. このデータセットには, 3Dゲームエンジンである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やセグメンテーション, 人物座標検出といった各モデルの学習に利用することができる. そこで, 本研究では人物検出および座標検出モ

A closeness estimating method between people for indoor fisheye images using person detection and pose estimation  
<sup>†1</sup> SHIRO KOMIYA, YUGA AKITA, HIRONOBU ABE, Tokyo Denki University

デルの学習にこのデータセットを利用し、顔向き検出にはの著者らによる学習済みのモデルを利用した。

### 2.4 人物検出モデルの生成

2.3 節で示したデータセットから魚眼画像とそれに対応する Bounding Box 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。つぎに、フレームワークとして PyTorch を利用、モデルを生成した。このとき、最適化アルゴリズムを Adam, バッチサイズ 12, 50 エポックに設定した。

### 2.5 座標検出モデルの生成

2.3 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。つぎに、フレームワークとして scikit-learn と Keras 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に Adam を選択、バッチサイズ 128, エポック数 200 に設定した。

## 3. 評価

### 3.1 人物検出モデル、座標検出モデルの評価

2.4 節および 2.5 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 1 に示す。座標検出モデルでは、各モデルの MAE, MSE, RMSE を評価した。評価結果を表 2 に示す。

表 1 人物検出モデルの評価

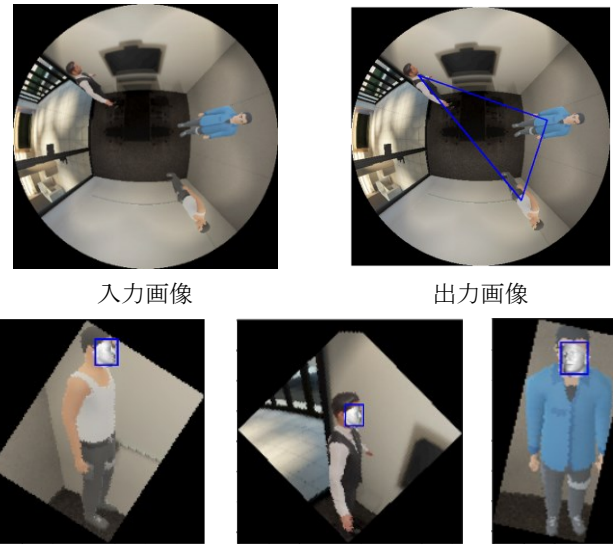
画像中の人数	mIoU
1	0.79

表 2 座標検出モデルの評価

Algorithm	MAE	MSE	RMSE
Liner Regression	(2.51, 2.64)	10.36	3.21
Bagging	(0.78, 0.86)	1.31	1.14
Random Forest	(0.68, 0.73)	1.03	1.01
MLP	(0.17, 0.14)	0.05	0.22

### 3.2 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、人物が 3 人写っている画像 1 枚に対して密接度推定を行った。入出力画像を図 2, 密接度推定結果を表 3 に示す。



人物および顔向きの検出結果  
図 2 人物 3 名のときの密接度推定例

表 3

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

## 4. 考察とまとめ

人物検出モデルでは 0.79 の mIoU を達成した一方で、Bounding Box 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。座標検出モデルでは MLP によるモデルで正解値との誤差 20cm 以内を達成した。評価に用いた画像データは、床の辺の長さが縦横ともに 3m ほどであるため、実用上の問題は少ないと考えられる。提案方式では、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。今後の課題として、人物の足元座標取得精度を評価するため、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する必要がある。また、人物が背を向けている場合に顔向きが検出できないため、処理方法を検討しなければならない。

### 参考文献

- [1] WHO: Coronavirus disease (COVID-19), [https://www.who.int/health-topics/coronavirus#tab=tab\\_1](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ス)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](https://jpn.nec.com/press/202006/20200611_03.html), (参照 2021-12-11).
- [3] W. Liu, D. et al: SSD: Single Shot MultiBox Detector, In Proc. European Conf. on ECCV, pp.21-37, Vancouver, Canada (2016).
- [4] V'itor Albiero. et al: img2pose: Face Alignment and Detection via 6DoF, Face Pose Estimation, CVPR, 2021, arXiv:2012.07791.