

密集度検出に向けた屋内合成魚眼画像に対する人物検出方式の検討

Examination of person detection method for synthetic indoor fisheye image for density detection

古宮嗣朗^{†1} 秋田悠河^{†1} 阿倍博信^{†1}
 Komiya Shiro Yuga Akita Hironobu Abe

1. はじめに

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。特に密集と密接については、屋内環境下における状況を客観的に判断するため、映像監視技術を活用した検出システムが検討されはじめている。

こうした映像監視技術を活用したシステムにおいて、一台で広範囲撮影が可能である利点を活かして全周魚眼レンズが採用される場合がある一方で、撮影画像が歪み映像解析の際に補正処理が必要となるという課題がある。この課題に対して、機械学習を用いることで補正処理を施すことなく自動で映像解析を行うという手法が考えられるが、解析精度の向上には大量の学習データを必要とすることが予想されるため、画像収集やアノテーションデータ作成といった新たな課題がうまれる。

本研究では、一台のカメラで撮影された全周魚眼画像に対して補正処理を行わずに人物の密集度検出が可能なシステムの開発を目指している。本論文では、屋内全周魚眼画像に対する密集度検出方式の基礎検討結果について報告する。

2. 関連研究

映像監視技術について、群衆密度推定手法やCG画像を用いたデータセット構築に関する研究が存在する。

山地らは深層ネットワーク後段の解像度を極端に縮小することで、頭部サイズに対するぼかし幅の誤差を抑えており、従来手法と比較して誤差を低減しつつ軽量化が可能であることを示している[1]。また、日本電気株式会社は人と人が十分な距離を保っているかどうか判定する技術を発表しており、ソーシャルディスタンス指数のリアルタイムな可視化を実現している[2]。さらに、磯井らは3DゲームエンジンのUnityを用いて室内行動解析向け合成動画データセットを構築し、CNNにより5クラスの動作分析が可能であることを示している[3]。

これらの研究は入力画像として全周魚眼画像ではなく通常の画像を使用することを前提としたシステムとなっておりと考えられる。

3. 密集度検出方式に関する基礎検討

3.1 密集度検出方式の検討

本研究では、図1に示すような密集度検出システムの開発を目指しており、密集度検出方式の基礎検討を行った。以下、その検討内容について説明する。

(1) 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、リサイズに加えて色情報の正規化を行う。

(2) 人物検出，各人物の足元座標取得：

前処理後の画像を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情報を推定する。バウンディングボックスの4つの頂点のうち、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。

(3) 実空間における足元座標の推定，ペア毎の距離推定：

(2)で取得した各人物の足元座標を位置推定モデルに入力し、実空間における足元座標を推定する。この足元座標から、二人一組のペア毎に距離を取得する。

(4) 密集度検出：

(3)で推定した距離情報を用いて数値データとして密集度を検出する。

(2)について、当初はセグメンテーション手法の利用を検討していたが、実験の結果、下半身の認識率が低いという結果が得られたため、足元座標取得の観点から物体検出手法を採用することにした。また、(4)については医学的な視点も必要であると考え現段階では未実施とした。



図1 密集度検出システムの概要

3.2 合成屋内全周魚眼画像データセットの利用

本研究では、3.1節に基づき密集度検出方式に機械学習を利用する方針であり、精度の高い推論モデルを作成するため大量の学習データを用意する必要がある。そこで、推論モデルのプロトタイプ作成のため、秋田ら[4]が作成しているCGによる屋内全周魚眼映像データセットを利用することにした。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合成画像データとそれに対応したアノテーションデータがMicrosoft COCO形式[5]で収録されており、屋内全周魚眼画像における物体検出やセグメンテーションの学習に利用することができる。図2に収録されている画像データの例を示す。

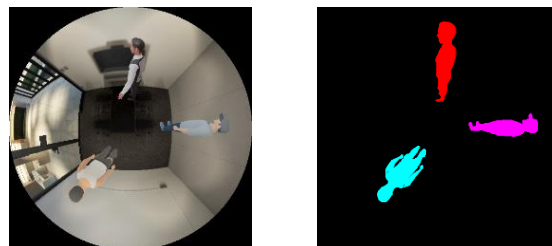


図2 データセットの画像例

^{†1} 東京電機大学 Tokyo Denki University

3.3 人物検出モデルの生成

人物検出モデル生成のため、3.2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[6]を選択した。加えて、物体検出アルゴリズムとしてSSD[7]、最適化アルゴリズムとしてAdam[8]を選択して、バッチサイズ12、50イテレーションに設定してモデルを生成した。学習曲線を図3に示す。

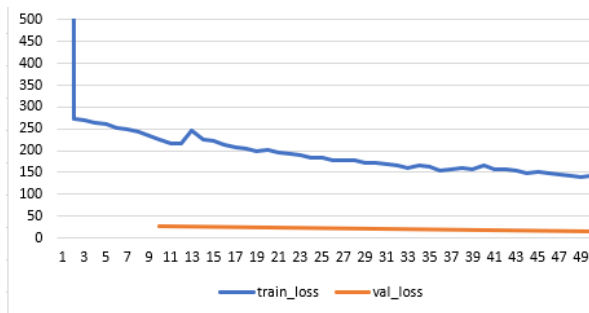


図3 学習曲線

3.4 位置推定モデルの生成

位置推定モデル生成のため、3.2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1,000組を教師データとして抽出して、学習データとして900組、テストデータとして100組を使用した。このとき、屋内全周魚眼画像における座標の単位はPixel、3D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[9]を選択した。加えて、位置推定のアルゴリズムには直線回帰、バギング、ランダムフォレストをそれぞれ選択して、モデルを生成した。

4. 評価と考察

3.3節および3.4節で生成したモデルの有用性を確認するため評価を行った。

まず、人物検出モデルについて、学習したモデルにテストデータとして画像100枚を入力して人物検出を行い、IoUの平均値を評価した。検出例を図4、IoUの評価結果を表1に示す。つぎに、位置推定モデルについて、正解値との誤差平均とMSE、 R^2 をそれぞれ評価した。評価結果を表2に示す。

人物検出モデルについては、バウンディングボックスのサイズが大きくなりすぎたり、複数人を一つのバウンディングボックスで囲んでしまったりする結果が得られた。これは、図3から読み取れるように学習が収束しきれていないからだと考えられる。また、今回使用したデータセットでは屋内の調光や反射などの条件が固定であったため、生成したモデルにデータセットと異なる条件の画像を入力するとIoUが下がることが予想される。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い、もう一度IoUについて評価する必要があると考えられる。

位置推定モデルについては、直線回帰の精度が最も低く、ランダムフォレストの精度が最も高いという結果が得られた。これは、教師データが非線形であるためだと考えられる。一方、最も精度が高いランダムフォレストについても

正解値から70cmほどの誤差があるため、より精度の高いモデルを作成するためにグリッドサーチ等の手法を用いてハイパーパラメータの調整を行う必要があると考えられる。

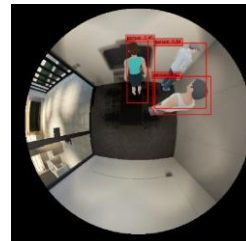


図4 人物検出の実行例

表1 人物検出モデルの評価

アルゴリズム	mean IoU
SSD	0.79

表2 位置推定モデルの評価

アルゴリズム	誤差平均	MSE	R^2
直線回帰	(2.24, 2.59)	9.90	0.92
バギング	(0.83, 0.88)	1.26	0.99
ランダムフォレスト	(0.72, 0.77)	0.99	0.99

5. おわりに

一台のカメラで撮影された全周魚眼画像に対して補正処理を行わずに人物の密集度を検出するため、密集度検出方式について基礎検討を行った。加えて、密集度検出の基盤技術確立のため人物検出モデルおよび位置推定モデルを作成し、評価した。その結果、精度に課題はあるものの、屋内全周魚眼画像に対して人物の密集度検出を行う基盤技術として有効性を確認することができた。今後は、密集度検出精度向上のため、各モデルの精度を高めていく。また、人物の体向き等の情報を利用して密集度も検出できる方式について検討する。

参考文献

- [1]. 山地雄士ほか1名: 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).
- [2]. 日本電気株式会社: 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 (accessed 2020-12-31).
- [3]. 磯井葉那ほか3名: 機械学習向け生活空間動画データセット構築の検討, 情報処理学会 第82回全国大会講演論文集, pp.243-244 (2019).
- [4]. 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).
- [5]. Adam Paszke. et al: PyTorch (2016).
- [6]. Tsung-Yi Lin. et al: Microsoft COCO: Common Objects in Context, ECCV 2014, Part V, LNCS 8693, pp.740-755 (2014).
- [7]. W. Liu, D. et al: SSD: Single Shot MultiBox Detector, In Proc. European Conf. on ECCV, pp.21-37, Vancouver, Canada (2016).
- [8]. D Kinga, J Ba Adam. et al: A method for stochastic optimization. In International Conference on Learning Representations (2015).
- [9]. David Courvapeau: scikit-learn (2007).