

1M-01

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

古宮嗣朗^{†1} 秋田悠河^{†1} 阿倍博信^{†1}
東京電機大学^{†1}

1. はじめに

新型コロナウイルスの流行により感染防止のため三密回避が意識されており、映像監視システムの活用も検討されはじめています。現在、映像監視システムに全周魚眼レンズが採用されることがあり、一台で広範囲の撮影が可能である一方、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、深層学習を用いて補正処理を施すことなく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接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、密接度検出の基礎検討として、3D ゲームエンジンを用いて屋内全周魚眼映像の学習用データセットを作成し、セマンティックセグメンテーションについて有効性を評価したので報告する。

2. 関連研究

映像監視の関連研究として、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]。

しかし、これらの研究は全周魚眼レンズを用いていない点で本研究と異なる。

3. 密接度検出の基礎検討

3.1 基本方針

本研究では図 1 に示すようなシステムの開発を目指している。このシステムを実現するには、全周魚眼画像中の人の数や位置関係の把握を可能とする基盤技術確立が必要であるため、物体をピクセル単位で認識可能であるセグメンテーション技術につ

いて有効性を評価する。セグメンテーションの学習には、通常の画像に加えて、アノテーションデータとし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り、画像収集やアノテーションデータ作成が大きな手間となる。そこで、3D ゲームエンジンを用いて全周魚眼画像とマスク画像を生成し、データセット作成の手間やコストの削減を試みる。

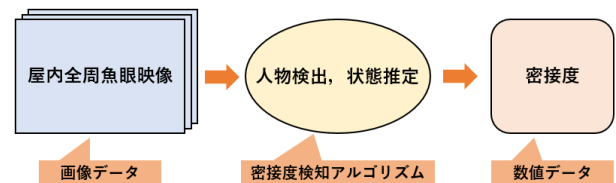


図 1 システムの概要

3.2 CG による屋内全周魚眼画像生成環境の構築

3.1 節を踏まえ、3D ゲームエンジンである Unity を用いて、屋内全周魚眼画像上の人物に対してセマンティックセグメンテーションを行うための CG 画像データ生成環境を構築した。この環境では、屋内オフィスに存在する複数人物の全周魚眼画像およびマスク画像の生成を行うことができる。さらに、生成画像のバリエーションを増やすため、オフィス内の内装や調光、反射といった各条件の変更に加えて、人物モデルに対するアニメーションの実行を可能としている。環境構築にあたり、オフィスの 3D モデルとして UnityJapanOffice プロジェクト[4]、人物の 3D モデルとして Survival stylized characters + 5 Weapons [5]、アニメーションのモーションデータとして Unity-Chan! Model[6] を利用した。図 2 に各 3D モデルの概観を示す。また、全周魚眼画像を生成するため、カメラに FullDome Camera For Unity[7]を利用した。

3.3 CG による屋内全周魚眼画像データセットの作成

3.2 節で構築した環境を用いて、全周魚眼画像とマスク画像をそれぞれ 2220 枚出力し、Pascal VOC[8]形式のデータセットを作成した。このとき、人物の数は 2 人にして、それぞれの座標を固定し、6 種類のモーションデータを適用した。加えて、モーション毎に z 軸を基準として一秒ごとに 360°回転させた。また、マスク画像の生成では、背景、人物の頭部、胴体、下半身をそれぞれ別の色に塗りわけ、4 つのクラスに分類した。しかし、生成されたマスク画像は Pascal VOC 形式ではないため、そのまま学習に用いることはできなかった。そこで、出力されたマスク画像をインデックスカラー形式の png ファイルに変換、カラーパレットを 4 色まで減らした。この処理によりマスク画像に含まれる色の数と色分けしたクラスの数とを統一し、Pascal VOC 形式に変換した。図 3

Examination of human closeness detection method for indoor circular fisheye image using semantic segmentation

^{†1} SHIRO KOMIYA, YUGA AKITA, HIRONOBU ABE, Tokyo Denki University

にデータセットとして収録した画像の例を示す。



(a) Unity Japan Office (b) Survival stylized characters + 5 Weapons

図2 使用 3D モデル



(a) 全周魚眼画像 (b) マスク画像

図3 データセットの画像例

3.4 学習

作成したデータセットを用いてセマンティックセグメンテーションの学習を行った。このとき、セマンティックセグメンテーションのアルゴリズムには PSPNet[9]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 イテレーションとして、ADE20K データセット[10]の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。学習時の loss 値の変化を図 4 に示す。つぎに、学習したモデルにテストデータを 370 枚入力して、セマンティックセグメンテーションされた画像を出力した。出力した画像の例を図 5 に示す。

4. 評価と考察

セマンティックセグメンテーションの精度を評価するため、出力画像 370 枚に対して IoU の平均値を算出した。評価結果を表 1 に示す。IoU の平均値は、背景が 0.998、頭が 0.796、上半身が 0.904、下半身が 0.749 という結果になり、比較的高い数値となった。しかし、今回作成した画像データセットは調光や反射、人物の座標などの条件が固定であった。そのため、学習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の値が大きく下がることが予想される。そのため、様々な条件を変更してデータセットの拡張を行い、もう一度 IoU について評価する必要がある。

5. おわりに

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接度検出が可能なシステムの開発を目的として、3D ゲームエンジンを用いて屋内全周魚眼映像の学習用データセットを作成、密接度検出の基礎検討としてセマンティックセグメンテーションについて有効性を評価した。その結果、データセット作成に Unity を利用することで、画像収集

やアニメーションデータ作成の手間やコストが削減可能であることがわかった。一方で、データセットの拡張を行い、再度 IoU を評価する必要があると考えられた。今後は、データセット拡張に加えて、密接度に関するモデル化を行い、密接度検出アルゴリズムについて検討していく。

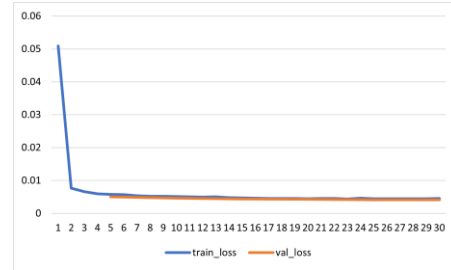


図4 loss 値の変化

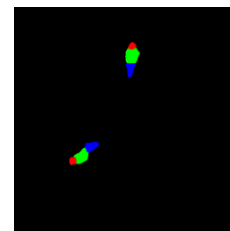


図5 セマンティックセグメンテーションの実行例

表 1 出力画像 370 枚の IoU

	背景	頭	上半身	下半身
IoU の平均	0.998	0.976	0.904	0.749

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画像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]. Unity Technologies Japan: UnityJapanOffice AssetStore 公開, <http://aec.unity3d.jp/topics/711/>, (accessed 2020-12-31).
- [5]. Alex Lenk: Survival stylized characters + 5 Weapons, <https://assetstore.unity.com/packages/3d/characters/survival-stylized-characters-5-weapons-115559>, (accessed 2020-12-31).
- [6]. unity-chan! : Unity-Chan! Model, <https://assetstore.unity.com/packages/3d/characters/unity-chan-model-18705>, (accessed 2020-12-31).
- [7]. rsodre: FulldomeCameraForUnity, <https://github.com/rsodre/FulldomeCameraForUnity>, (accessed 2020-12-31).
- [8]. Everingham M. et al: The PASCAL Visual Object Classes (VOC) Challenge, A. International Journal of Computer Vision, vol.88, no.2, pp.303-338(2010).
- [9]. Hengshuang Zhao. et al: Pyramid Scene Parsing Network, IEEE Conference on Computer Vision and Pattern Recognition (2017).
- [10]. Bolei Zhou. et al: Scene Parsing Through ADE20K Dataset, IEEE Conference on Computer Vision and Pattern Recognition, pp. 633-641(2017).