

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

秋田悠河[†] 古宮 嗣朗[†] 阿倍博信[†]
東京電機大学[†]

1. はじめに

現在、新型コロナウイルスが流行しており、その感染対策として3密の回避が求められている。この3密とは密閉・密集・密接のことを指しており、屋内においてこの3密を回避する必要がある。特に、密集・密接については個人の判断で屋内の状況を適切に判断することは難しい。そのため密集・密接の状態を分析することで、屋内の状況を把握することが可能となる。そのため、室内を対象とした密集度、密接度の計測が必要とされている。密集度、密接度の計測を自動かつ低コストで広範囲に行う場合、全周魚眼カメラを利用した映像監視システムの構築と機械学習に基づく解析システムの構築が必要となる。

しかし、機械学習に基づき全周魚眼カメラから得られた画像の解析システムを構築するためには、学習の際に必要な全周魚眼画像データセットが必要になる。大量かつバリエーション豊富なラベル付きデータセットの作成を実写画像で行う場合、多大な手間やコストがかかる。一方、3Dゲームエンジンを利用することにより、必要なモデルとスクリプトを用意することで、大量かつ豊富なバリエーションのデータセットの自動作成が可能であるとともに、アノテーション作業の自動化も容易であると考えられる。

このような背景のもと、本論文では、3Dゲームエンジンを利用し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その有効性の評価について報告する。

2. 関連研究

関連研究としては、魚眼画像に対する人物検出とゲームエンジンを利用したデータセット作成の研究が存在する。

青木は、視覚センサによる見守り技術の研究において、深層学習であるFaster R-CNN[2]を用いた広角レンズで取得した画像から人物検出を行っている[1]。広角レンズの画像を深層学習により学習し広角レンズで撮影した画像に対する人物検出が可能であると示しているが、具体的なデータセットと評価手法については言及されていない。

磯井らは、機械学習向け生活空間動画データセット構築の研究において、ゲームエンジンを利用して生活空間内の人物検出及び動作を分類するためのデータセットを作成

している[3]。作成したデータセットを用いて機械学習を行い動作分類の有効性について確認を行っている。しかし、3Dゲームエンジンの利用による全周魚眼画像を対象としたデータセット作成及び人物検出の評価が研究対象ではない。

3. 全周魚眼画像データセット作成方式の提案

3.1 基本方針

本研究では、全周魚眼画像上で機械学習を用いた人物検出のためのデータセット作成を目的として、3Dゲームエンジンを用いた合成全周魚眼画像データセットの作成を行う。

3.2 データセット作成方式

データセット作成は大きく3つの段階に分かれるため、以下の通り整理する。

- 3Dゲームエンジン内に人物モデル・魚眼カメラ・建物モデルを配置する。
- 人物モデルをゲームエンジン内スクリプトによってランダムに座標と角度を変更し、学習用画像及びGround Truth画像の2種類の全周魚眼画像を作成する。
- 作成したGround Truth画像を用いてアノテーションを自動作成する。

3.3 3Dゲームエンジン内での屋内の再現

3DゲームエンジンはUnity[4]を使用し、3Dゲームエンジン内に屋内を再現してそこに人物モデル・全周魚眼カメラを配置する。描画を行うレンダリングパイプラインは最高解像度で描画が可能なHigh Definition Render Pipelineを採用する。人物や建物のモデルや全周魚眼カメラについては表1に示すフリーのアセットを使用する。

表1. ゲームエンジンとモデル

Item	Content
3D game engine	Unity[4] ver2019.2.4f1
Rendering pipeline	High Definition Render Pipeline
Building model	UnityJapanOffice[5]
Human model	Survival stylized characters + 5 Weapons[6]
Full dome fish eye camera	FullDomeCameraForUnity[7]

3.4 全周魚眼画像作成とアノテーション作成

まず、Unity内にUnityJapanOfficeを配置した。今回使用する部屋は1か所とした。その部屋に全周魚眼カメラの向きが床面に対して垂直になるように天井に設置し、図1に

Creation and Evaluation of Synthetic FullDome Fisheye Image Dataset for Video Surveillance using 3D Game Engine

[†]YUGA AKITA, SHIRO KOMIYA, HIRONOBU ABE, Tokyo Denki University

示した学習画像と Ground Truth 画像の 2 種類の画像を作成した。

以下に画像作成の手順について整理する。以下、十分な画像数に達するまで、以下の手順を繰り返す。

- 使用する人物モデル 1 種類を部屋に配置する。
- 学習用画像作成のために通常のテクスチャが採用されているか確認して変更する。
- スクリプトで動かすモデルを指定して、座標と角度をランダムに変更する乱数を設定する。
- Unity を再生して全周魚眼画像を作成する。
- Ground Truth 画像を作成するために図 1 の様に人物モデルを青色に変更し、それ以外を全て黒色に変更する。

次に、Python で作成したプログラムを用いてアノテーションを作成する。作成したプログラムの動作について、以下に整理する。

- Ground Truth 画像から人物モデルに該当する青色のピクセルを探す。
- 該当する人物モデルの部分の矩形と頂点の情報をアノテーションファイルに書き込む。

本研究では、人物モデルを 4 種類、全周魚眼カメラの向きを 2 種類とし、合計で 8 種類の画像を作成した。作成したデータセットの仕様は表 2 の通りである。

表 2. データセットの仕様

Specification items	Specifications
Normal images	4000
Human model types	4
Human model poses types	1
Room types	1
Fisheye camera angles	2

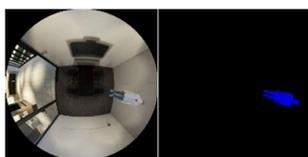


図 1. 学習用画像と Ground Truth 画像

4. データセットの評価と考察

今回、作成したデータセットに対する人物検出精度に対する評価について、Mask R-CNN[8]を用いた評価を行った。表 3 に示した通り、ハイパーパラメータを変更した 5 パターンで学習、評価を実施した。また、その際の学習曲線について図 2 にまとめる。

表 3 より 1 クラス 4000 枚のデータセットに対して epoch が 10, step が 80 以上で検出率が一番高くなることが分かった。しかし、図 2 の val_loss が振動していることから過学習が発生している可能性が考えられる。また検出率が上昇している一方で mIOU が僅かながら低下しているため、デー

タセットの枚数を増加させた上でハイパーパラメータを調整して再度学習を行い評価が必要であると考えられる。

最良の検出率が 0.804 で mIOU が 0.87 であることから今回のデータセット作成方式と Mask R-CNN を用いることで魚眼画像上の人物検出が可能であると考えられる。しかし、検出率に向上の余地があることから、データセットの拡張及びハイパーパラメータの調整が必要であると考えられる。

表 3. 学習と検出結果

pattern	1	2	3	4	5
epoch	10	10	5	3	5
Step_per_epoch	80	50	50	50	50
Learning rate	0.001	0.001	0.001	0.002	0.002
loss	0.0463	0.0313	0.0486	0.0841	0.0689
val_loss	0.3307	0.4568	0.351	0.3248	0.2768
Detection rate	0.804	0.724	0.5	0.216	0.672
mIOU	0.87	0.884	0.883	0.861	0.90

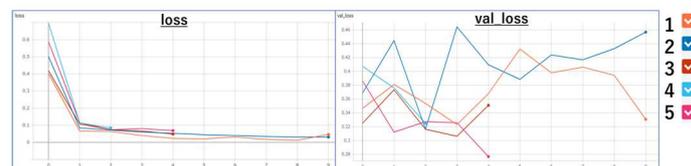


図 2. 学習曲線

5. おわりに

本研究では、3D ゲームエンジンを用いて合成全周魚眼画像データセットを作成した。Mask R-CNN を用いて人物検出精度に対して評価した結果、本研究のデータセット作成方式で全周魚眼画像から人物検出を行うに十分なデータセットの作成が可能であることが分かった。しかし、今回作成したデータセットの画像枚数では不十分である可能性があることと、実写画像及びバリエーションの評価は未実施であることから、今後は継続してデータセットを拡張していくとともに、追加評価を実施していく予定である。

参考文献

- [1]. 青木義満: 視覚センサによる見守り技術について, 電子情報通信学会 通信ソサイエティマガジン, 11 巻, 1 号, pp.30-38 (2017).
- [2]. Jian Sun: Faster R-CNN, <https://arxiv.org/pdf/1506.01497.pdf>
- [3]. 磯井葉那他 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, pp.243-244 (2020).
- [4]. Unity: Unity, <https://unity.com>.
- [5]. Unity: UnityJapanOffice, <https://assetstore.unity.com/packages/3d/environments/unityjapanoffice-152800?locale=ja-JP#description>.
- [6]. Alex Lenk: Survival stylized characters + 5 Weapons, <https://assetstore.unity.com/packages/3d/characters/survival-stylized-characters-5-weapons-115559?locale=ja-JP>.
- [7]. Rsodre: FulldomeCameraForUnity, <https://github.com/rsodre/FulldomeCameraForUnity>.
- [8]. Kaiming He: Mask R-CNN, <https://arxiv.org/pdf/1703.06870.pdf>