

## 3D ゲームエンジンを用いた全周魚眼カメラ用映像監視向けデータセットの作成と評価

### Creation and Evaluation of Dataset for Video Surveillance for Full dome Fisheye camera using 3D game engine

秋田 悠河<sup>†</sup>  
Yuga Akita

古宮 嗣朗<sup>†</sup>  
Shirou Komiya

阿倍 博信<sup>†</sup>  
Hironobu Abe

#### 1. はじめに

現在は新型コロナウイルスが流行しており、その感染症対策の要点として3密を回避する社会の課題が存在している。この3密とは密閉・密集・密接を指している。つま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・密接においては各個人で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・密接の状態を定量的に分析することが可能であれば屋内の状況を評価することが可能である。つまり屋内を対象とした密集度の計測が必要であると考えられる。密集度の計測を自動でかつ低コストでなるべく広い範囲で行おうとすると魚眼カメラを利用した映像監視システムと解析を担う機械学習の実装が必要である。

しかし、魚眼カメラから得られた映像に対して人物検出や位置推定を機械学習によって実現する為には、学習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で行うには多大なコストや手間がかかる。一方、3D ゲームエンジンを利用する事ができれば、必要なモデルとスクリプトを用意する事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生成が可能であると考えられる。またアノテーション作業も自動化しやすいと考えられる。本論文では、3D ゲームエンジンを利用した人物検出及び位置推定の為のデータセット作成とその有効性の確認を行う。

#### 2. 関連研究

関連研究として、魚眼画像上の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。

青木による視覚センサによる見守り技術について[1]という研究では、深層学習である Faster R-CNN[2]を用いた広角レンズ上の人物検出を行っている。広角レンズからの映像を深層学習で学習させ広角レンズ画像での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不明である。

一方ゲームエンジンを利用したデータセット作成の研究では、磯井らによる機械学習向けに生活空間動画データセット構築の検討[3]存在している。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。それ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、魚眼画像を目的とした 3D ゲームエンジンの利用によるデータセット作成及び人物検出の評価ではない。

以上の研究より本研究では深層学習による魚眼画像からの人物検出、位置推定及び、それに必要なデータセットの作成に注目して研究を進めることになる。

<sup>†</sup> 東京電機大学 Tokyo Denki University

#### 3. データセット作成方式の提案

##### 3.1 基本方針

本研究では、魚眼画像上での機械学習を用いた人物検出と位置推定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 3D ゲームエンジン内での人物検出用の屋内再現

3D ゲームエンジン内に UnityJapanOffice[5]を配置して部屋の天井に魚眼カメラを下向きに設置し、そこに人物モデル・魚眼カメラを配置する。使用する 3D ゲームエンジンは Unity[4]である。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、魚眼カメラについてはフリーのアセットを使用する。使用したゲームエンジンとアセットについては表 1 にまとめる。

表 1. ゲームエンジンと使用アセット

Item	Content
3D game engine	Unity ver2019.2.4f1
Rendering pipeline	High Definition Render Pipeline
Building model	UnityJapanOffice
Human model	6 types
Fish eye camera	Full dome-hdrp[6]

##### 3.3 人物検出用データセット作成方式

人物検出用データセットの作成にあたって魚眼画像の生成とアノテーションの2つの段階がある。それを以下に整理する。魚眼画像生成の場合は以下の方法で行う。

- 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。
- 学習用画像であれば通常のテクスチャを適用し、Ground Truth 画像の場合はテクスチャを調整して背景とモデルの2色のみになる様にする。
- 人物モデルをスクリプトによって乱数を用いて座標と角度を変更し学習用画像及び Ground Truth 画像の2種類の画像を作成する。作成した2種類の画像は図1の通りである。

アノテーション作業の場合は以下を行う。

- 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領域の情報をアノテーションファイルに書き込む。  
本研究の人物検出用データセットの仕様は表2の通りである。

表2. 人物検出用データセットの仕様

Specification items	Specifications
Normal images	6000
Human model types	6
Human model poses types	1(walking)
Room types	1
Fisheye camera angles	2

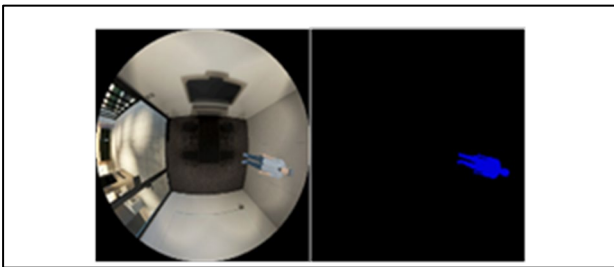


図1. 学習用画像と Ground Truth 画像

### 3.4 3D ゲームエンジン内での位置推定用の屋内再現

位置推定用データセットには人物モデルの代わりに球体モデルを使用する。そのため、モデル以外は表1と同じである。

### 3.5 位置推定用データセット作成方式

位置推定用データセットの作成も人物検出用と同様に画像の作成とアノテーションの2段階がある。画像作成の方法においては、人物検出用の方法に加え、球体モデルのUnity内のxとzの座標を出力して記録する。アノテーションの段階では人物検出用とは違い、矩形や領域の代わりに画像上の球体モデルの中心点の座標情報を出力する。最終的なデータセットとしては画像上の座標とUnity内の座標が関連付いている座標データになる。位置推定用のデータセットの仕様は表3の通りである。

表3. 位置推定用のデータセットの仕様

Specification items	Specifications
Number of pairs	1000
objective variable	x and z coordinates in Unity
Explanatory variable	Center coordinates of the sphere model on the image

## 4. 評価

### 4.1 人物検出用データセットの評価

人物検出のアルゴリズムとしてMask R-CNN[7]を使用する。学習後に、データセット外の画像で推論を実行して人物検出精度の評価を行った。比較用にMS-COCOデータセット[8]を学習したモデルでも検出を行った。検出結果については表4の通りである。

表4. 人物検出モデルの評価結果

	Our dataset	MS-COCO
Detection rate	0.7	0.218
mIOU	0.62	0.67

### 4.2 位置推定用のデータセットの評価

画像上の座標を説明変数、Unity内の座標を目的変数として線形回帰モデルとランダムフォレストを用いて学習を行いデータセット外の座標で推論を行った。位置推定の結果は表5の通りである。

表5. 位置推定用モデルの推論結果

Algorithm	Linear Regression	Random Forest
RMSE(train)	2.098	0.450
RMSE(test)	3.018	1.143
MAE of x (m)	2.297	0.866
MAE of y (m)	2.347	0.855

## 5. 考察

今回作成したデータセットはMS-COCOよりも検出率が高いことから魚眼画像を学習することにより魚眼画像上での人物検出機能を作成することは可能であると考えられる。位置推定機能に用いるアルゴリズムは非線形アルゴリズムの方が適していると考えられる。しかし、今回の検証結果からMAEが約0.85mであることから正確な位置推定を行うための性能であると判断することは困難である。

## 6. おわりに

3D ゲームエンジンを用いて人物検出及び位置推定用の合成全周魚眼画像データセットを作成した。本研究のデータセット作成方式で魚眼画像から人物検出を行う機能を作成することは可能であるがモデルの種類を増やす必要がある。また位置推定においては今回のデータセットと学習アルゴリズムでは正確な位置推定を実現するのは困難であるため、データセットの作成及び適切なアルゴリズムの選定が必要である。

### 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: Towards Real-Time Object Detection with Region Proposal Networks”, <https://arxiv.org/pdf/1506.01497.pdf>, (2016).
- [3] 磯井 葉那 他, “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, “Unity Japan Office”, <https://assetstore.unity.com/packages/3d/environments/unityjapanoffice-152800?locale=ja-JP#description>.
- [6] thebelin, “Full dome-hdrp”, <https://github.com/thebelin/fulldome-hdrp> .
- [7] Kaiming He, “Mask R-CNN”, <https://arxiv.org/pdf/1703.06870.pdf>, (2017)
- [8] Tsung-Yi Lin 他, “MS-COCO dataset”, <https://cocodataset.org/>, (2017)