

1. はじめに

深層学習を用いて経路予測を実現するには、大規模な経路データを学習する必要がある。公開されているデータセットの多くは、市街地や大学構内を空撮した鳥瞰視点映像や監視カメラ映像を対象としている。これらの映像に対して移動対象の経路を手でアノテーションする必要があるため、データセット構築の時間的コストは多大である。また、データセットごとにビューポイントは固定されている。そこで本研究では経路情報をカメラ映像から自動で生成することにより、最小限のアノテーションコストで大規模なデータセット構築を容易にする。

2. 従来のデータセット

経路予測における代表的なデータセットとして、市街地の歩行者を対象とした ETH [1] と UCY [2] がある。また、駐車場に設置された監視カメラ映像を利用したデータセットに VIRAT [3]、スタンフォード大学構内をドローンで撮影したデータセットに Stanford Drone データセット [4] がある。それぞれのデータセットの概要を表 1 に示す。従来のデータセットは物体ごとに時系列変化を確認しながら経路をアノテーションする。そのため膨大な作成コストがかかる。また、データセットは撮影したビューポイントのみに固定され、異なるビューポイントのデータは存在せず、別途撮影とアノテーションが必要となる。

表 1: 各データセットの概要

データセット	歩行者数	フレーム数	ビューポイント	シーン数	属性
ETH/UCY	1,536	57,274	鳥瞰	5	pedestrian
VIRAT	4,021	823,738	監視カメラ	11	pedestrian, car, bike
Stanford Drone	11,216	522,437	鳥瞰	8	pedestrian, cars, bikers, skateboarders, buses, golf carts

3. データセット生成

本研究では、経路情報の獲得を自動化しデータセット生成時のコスト削減、自由なビューポイントの変更を可能にしたデータセットの自動生成方法を提案する。データセットの自動生成には CARLA シミュレータを用いる。シミュレータを用いることでビューポイントの変更や、セマンティックセグメンテーション画像を容易に獲得できる。

3.1. 物体追跡による経路データセット自動生成

提案するデータセットの自動生成の流れを図 1 に示す。データセットの自動生成には人の往来を撮影した映像を入力する。また、経路情報をピクセル座標系からワールド座標系に変換する射影変換行列と CARLA で作成したマップを用意する。出力はワールド座標系の経路情報とマップにプロットされた経路である。

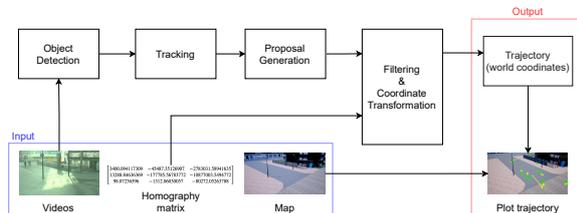


図 1: 提案するデータセット自動生成の流れ

動画からの経路獲得には、物体検出に Mask R-CNN、物体追跡に Deep SORT を用いる。出力された矩形領域の下端を人の座標値とし経路情報を獲得する。獲得したピクセル座標系の経路情報を、CARLA 上でプロットするためにワールド座標系へ変換する。ワールド座標には緯度経度を用いる。ワールド座標と対応するシーン内のピクセル座標から透視投影を行い、射影変換行列を獲得する。獲得した経路情報と射影変換行列を用いてピクセル座標をワールド座標に変換する。

3.2. 作成したデータセットの概要

自動生成手法を用いて大阪の市街地をシーンとしたデータセットを作成した。作成したデータセットの概要を表 2

表 2: 作成したデータセット

歩行者数	フレーム数	ビューポイント	シーン数	属性
93,305	1,134,684	自由	1	pedestrian, bike

に示す。また、CARLA で作成しているためビューポイントは自由に変更が可能である。

4. 評価実験

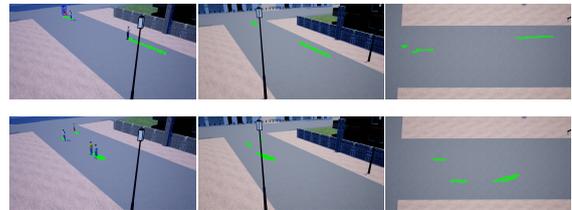
表 2 で作成した 100 人分の経路情報を自動生成のみ、自動生成手法と人手でアノテーションしたときの時間的コストや経路の違いを比較する。人手でのアノテーションは、対象シーンをしながら CARLA 上に経路を入力する。経路情報の比較には、各時刻のユークリッド距離の平均誤差の Average Displacement Error(ADE)、最終時刻時のユークリッド距離誤差の Final Displacement Error(FDE) を用いる。

4.1. 実験結果

定量的結果を表 3 に示す。自動生成手法による経路が経路情報として適していない場合、経路を手動で修正した。時間的コストは、人手による生成に比べ自動生成の方が約 4 分の 1 に短縮された。したがって、アノテーションの時間的コストを削減できたと言える。ADE/FDE は、自動生成のみに対して、経路情報として適していない約 1 割の経路を修正することで、約 41%減少する。人手のアノテーションによる経路と自動生成アノテーションによる経路の可視化結果を図 2 に示す。図 2(a)(b)(c) を比べると人手のアノテーションと遜色ない経路が自動生成手法により生成されていることがわかる。よって提案手法は経路軌跡の作成に有効であると考えられる。

表 3: 定量的結果

	作成時間 [min]	ADE/FDE[m]
人手	250	—
自動生成	50	1.31/1.23
自動生成+人手	70	0.55/0.39



(a) 人手 (b) 自動 : 45° (c) 自動 : 90°

図 2: アノテーションの可視化結果

5. おわりに

本研究では、データセットの自動生成の提案と有効性を調査した。自動生成手法は、人手によるアノテーションとほぼ同等な経路を 4 分の 1 のコストで作成することが可能である。今後は異なるシーンのデータセットの作成と大規模データセットに有効な経路予測モデルの考案などが挙げられる。

参考文献

[1] S. Pellegrini, et al., “You’ll never walk alone: Modeling social behavior for multi-target tracking”, ICCV, 2009.
 [2] A. Lerner, et al., “Crowds by example”, Computer Graphics Forum 26, 3, pp.655-664, 2007.
 [3] S. Oh, et al., “A large-scale benchmark dataset for event recognition in surveillance video”, CVPR, 2011.
 [4] A. Robicquet, et al., “Learning social etiquette: Human trajectory understanding in crowded scenes”, ECCV, 2016.