

1. はじめに

自動運転において事故を防ぐためには、歩行者等の経路を予測して自車の危険状態を把握し、制御することが重要である。車載カメラ映像を対象とした歩行者の経路予測 [1] では、画像中の位置に関係なく歩行者の移動量を用いて経路を予測する。しかし、画像上の歩行者の位置により歩行者の移動量が異なるため、全ての位置において共通モデルを用いた予測は困難である。そこで、本研究では、この問題を解決するために、自転車と歩行者の位置関係を考慮した手法を提案する。

2. Long-Term On-Board Prediction

車載カメラ映像を用いた経路予測の手法として、Long-Term On-Board Prediction [1] が提案されている。本手法は、歩行者の移動量に加えて、車の自己移動量 (速度・ステアリング角度) と車載カメラ映像を用いることで、経路予測を実現している。ネットワーク構造は LSTM を利用した encoder-decoder モデルである。車載カメラ映像から獲得できる歩行者の移動量は、画像上の歩行者の位置やスケールによって異なる。よって歩行者の位置により予測誤差が大きくなる。

3. 提案手法

本研究では、自転車と歩行者の位置関係を考慮した経路予測手法を提案する。提案手法は、車載カメラ映像中の歩行者の位置とスケールを自転車と歩行者の位置関係として入力に追加する。提案するネットワークの構造を図 1 に示す。ここで、1 フレーム目から t フレーム目までの歩行者の移動量を $d_t = (\Delta x_t, \Delta y_t)$ 、歩行者の位置を $p_t = (x_t, y_t)$ 、歩行者のスケールを s_t 、自転車の自己移動量を $o_t = (\text{speed}, \text{steering angle})$ とする。まず、観測時の $\{d_t, p_t, s_t, o_t\}$ を、encoder へ逐次入力する。そして、時系列情報をまとめたベクトル v_{box} を獲得する。次に、獲得したベクトルを予測時の自己移動量を結合して decoder へ逐次入力する。これにより、各時刻の移動量 d_t と分散 σ_t を予測する。学習時はカリキュラム学習を導入し、学習初期は短いフレーム数で学習し、徐々にフレーム数を増加させながら歩行者の経路を学習する。

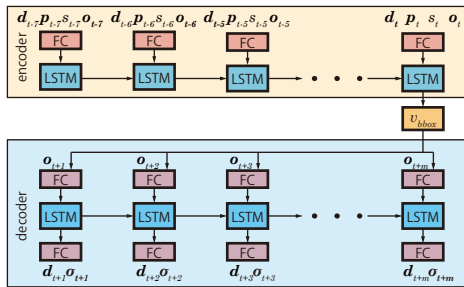


図 1: 提案手法のネットワーク構造

4. 評価実験

提案手法の有効性を示すために、歩行者情報の入力を変えて比較する。比較する入力方法は、以下の 4 つである。

- 入力 1: 歩行者の移動量 (従来手法)
- 入力 2: 歩行者の移動量 + 位置
- 入力 3: 歩行者の移動量 + スケール
- 入力 4: 歩行者の移動量 + 位置 + スケール

4.1. 実験概要

評価には Cityscapes Dataset [2] を用いる。本研究では 30 フレーム (1.8 秒) で構成される動画を使用し、学習サンプルは 2,975 サンプル、評価サンプルは 1,525 サンプルである。そして、過去 8 フレームを観測値として encoder に入力し、15 フレーム先までの歩行者の移動量を decoder で予測する。定量的な評価指標として、予測開始フレームから最終フレームまでの各時刻における真値と予測値のユークリッド距離の平均 (ADE) と予測の最終フレームにおける真値と予測値のユークリッド距離 (FDE) を用いる。

4.2. 実験結果

ADE 及び FDE により評価した結果を表 1 に示す。表 1 より、入力 4 (歩行者の移動量 + 位置 + スケール) の誤差、分散が最も小さい。これより、車載カメラ映像での経路予測において、歩行者の位置及びスケールが有効であることが分かった。しかし、入力方法の変更による ADE 及び FDE の差はわずかである。これは、1 フレーム目から最終フレームまでの時間が 0.9 秒と短く、入力方法の違いによる差が十分に生じていないためである。

表 1: 歩行者の経路の予測誤差 [pixel]

入力	ADE (分散)		FDE (分散)	
入力 1 (移動量 (従来手法))	10.94	(3.87)	19.44	(5.04)
入力 2 (移動量 + 位置)	10.95	(3.86)	19.61	(5.07)
入力 3 (移動量 + スケール)	10.88	(3.96)	19.39	(5.19)
入力 4 (移動量 + 位置 + スケール)	10.82	(3.77)	19.09	(4.91)

自転車が走行しているシーンにおける歩行者経路の予測誤差を表 2 に示す。ここで、自転車の観測時の平均速度が 1 m/s 以上のシーンを走行時とし、平均ステアリング角度が 3 度未満を直進、3 度以上を右左折とする。また、歩行者の位置は車載カメラ映像を横に 3 分割した時の、 t フレーム目における位置である。表 2 より、直進シーンでは、入力 4 の誤差が小さいこと分かる。また、右左折のシーンでは、入力 4 は画像中央の歩行者に対する誤差が小さいことが分かる。また、入力 4 は、右折の時の左側、左折の時の右側の FDE が小さいことが分かる。

表 2: 走行時の歩行者の経路の予測誤差 [pixel]

ステアリング 角度	歩行者の 位置	入力 1 / 入力 4	
		ADE	FDE
直進	左	10.67 / 10.56	19.26 / 19.12
	中	8.90 / 8.67	15.04 / 14.54
	右	11.88 / 11.89	21.43 / 21.04
左折	左	14.99 / 15.61	29.93 / 30.65
	中	13.10 / 12.73	23.65 / 22.91
	右	13.20 / 13.52	24.51 / 23.90
右折	左	15.57 / 12.19	28.50 / 22.45
	中	14.32 / 12.67	27.67 / 23.84
	右	13.20 / 13.83	25.60 / 26.95

各手法の経路予測例を図 2 に示す。図 2(a) より、歩行者の移動量が小さい場合、真値とほぼ同じ経路を予測できていることが分かる。一方、図 2(b)(c)(d) のように、歩行者の移動量が大きい場合、真値の経路よりもさらに先の時刻の経路を予測していることが分かる。このような場合、いずれも予測 1 時刻目から予測誤差が大きい。これは、encoder から decoder にかけての時系列情報の引き継ぎが適切に行われていないと考えられる。

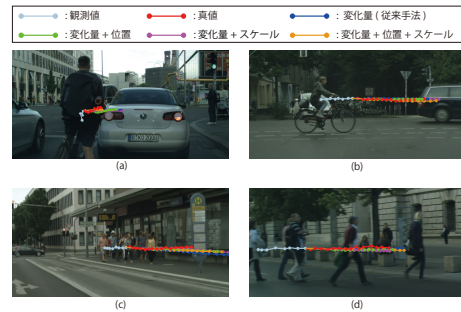


図 2: 経路予測の例

5. おわりに

本研究では、自動車から見た歩行者の位置関係を考慮した経路予測の手法を提案した。入力として、歩行者の移動量と位置とスケールの全てを考慮することで最も経路予測の誤差が小さくなった。今後は、歩行者の予測結果を入力値として用いる経路予測の評価を検討する。

参考文献

- [1] A. Bhattacharyya, *et al.*, “Long-Term On-Board Prediction of People in Traffic Scenes under Uncertainty”, CVPR, 2018.
- [2] M. Cordt, *et al.*, “The Cityscapes Dataset for Semantic Urban Scene Understanding”, CVPR, 2016.