

# Bayesian Estimation for Enhancing 3D-point Object Recognition using Spatio-temporal Information

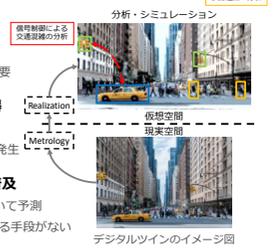
大阪大学 大学院情報科学研究科  
情報ネットワーク学専攻 村田研究室  
佐藤 広章

1

## 研究背景

### ● 現実空間の状況を把握 & 情報の利活用への期待

- デジタルツインの登場
  - 仮想空間上に現実空間を投影
  - 分析やシミュレーションが可能
  - 正確なデータの取得と高精度な物体認識が必要



### ● 現実空間を正確に反映したデータの取得

- 2D 画像：悪天候や夜間などでは撮影困難
- 3D 点群：光の反射や吸収による欠損箇所の発生

### ● 物体認識：機械学習手法による予測の普及

- 物体がどのクラスであるかを機械学習手法を用いて予測
- 予測結果がどの程度信頼できるのかを判断する手段がない

デジタルツインのイメージ図

2

## 研究目的とアプローチ

### ● 研究目的

- 現実空間の中でどの物体がどこにあるかの認識を、曖昧さを含めた表現  
予測結果がどの程度正当であるか、信頼できるかを検証
- 既存の物体認識手法に依存しない、知識情報を使用したベイズ推定手法の提案

### ● アプローチ

1. 現実空間を反映した知識情報の構築
  - 空間的知識情報：物体の空間的な位置関係にもとづく情報
  - 時間的知識情報：物体の時間的な連続関係にもとづく情報
2. 物体認識予測に対するベイズ推定を使用した手法の確立
  - 空間的知識情報を使用したベイズ推定手法
  - 空間的知識情報と時間的知識情報を使用したベイズ推定手法
3. ベイズ推定手法をデータセット全体に適用した際の評価

3

## 空間的知識情報の取得

### ● 物体の空間的な位置関係にもとづく情報

- nuScenes<sup>[1]</sup> データセットを使用
- 物体が位置している場所から、各物体間の距離を計算
- 空間上の距離による隣接を検出することで、カデゴリー間で隣り合って位置する頻度を算出

### ● 我々が生活する上で潜在的に認識している情報を想定

- 車同士や歩行者同士は隣り合う頻度が高い
- バイクは歩行者付近には存在しにくい



[1] H. Caesar, et al., "nuScenes: A Multimodal Dataset for Autonomous Driving," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11618-11628, 2020

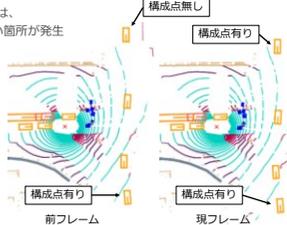
4

## 時間的知識情報の取得

### ● 物体の時間的な連続関係にもとづく情報

観測データ上に物体が連続して存在する確率を算出

- LIDAR を使用した現実空間の計測では、  
物体の材質や色によって取得できない箇所が発生  
→ 現実空間上で存在しているが、  
データ上では存在しない可能性
- 現フレームで物体が存在するとき、  
前フレームでも存在する確率を算出  
→ 連続する点群上において  
フレーム間での隣接を検出



5

## 知識情報を使用したベイズ推定手法

### ● 機械学習手法を用いた物体認識予測 (TransFusion<sup>[2]</sup> 手法)

- 予測スコア：物体が各クラスであると予測した尤度
- 予測クラス：予測スコアに対して尤度が最も高いクラス

### ● 提案するベイズ推定手法

物体認識による予測スコアをもとに、知識情報を使用した事後確率を算出

- 既存手法による物体認識から得られる予測スコアの信頼性を検証
- 情報補正や情報補強によって精度や信頼性の向上を目標
- 空間的知識情報を使用
  - 対象フレーム内の情報に着目
- 空間的知識情報に加えて、時間的知識情報を使用
  - 前フレームとの依存関係に着目：データから将来予測を行うアプリケーションを想定
  - 前後フレームとの依存関係に着目：データから行動分析を行うアプリケーションを想定

[2] X. Bai, et al., "Transfusion: Robust lidar-camera fusion for 3d object detection with transformers," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1090-1099, 2022

6

### 空間的知識情報を使用したバイズ推定手法の評価

7

- 予測クラスの正誤数による評価

機械学習手法による予測クラスと、バイズ推定による事後確率の最尤クラスを比較

- 正答
  - 正答を維持 : 109628 件
  - 間違いに変化 : 4516 件
- 誤答
  - クラスが変化 : 18107 件 → 誤答の検出
    - 補正 : 11372 件
    - 誤答を維持 : 18548 件

Accuracy:

機械学習 : 75.69% (114144 件)      機械学習手法による予測 Accuracy: **0.7569**

バイズ推定 : **80.23%** (109628 + 11372 = 121000 件)      バイズ推定手法による認識 Accuracy: **0.8023**

7

### 知識情報を使用したバイズ推定手法の評価

8

- 3 手法での予測クラスの正誤数を比較

- Accuracy : 機械学習手法の 75.69% と比較して全手法で上昇
- 情報を追加するごとに Accuracy が上昇
- 誤答を補正したパターンが大きくなる → 時間的な依存関係を正しく検出

空間と時空間の知識情報を使用した際の特徴的な違い

	正答を維持	誤答を検出	誤答を補正	Accuracy
空間的知識情報	109628 件	18107 件	11372 件	80.23%
時空間知識情報 (前フレーム)	110667 件	20590 件	16105 件	84.06%
時空間知識情報 (前後フレーム)	111211 件	22556 件	19111 件	86.42%

8

### 正答を維持したパターンに対する評価

9

- 物体認識予測の信頼度を検証

周囲の情報や時間的な依存関係を捉えることで、信頼度の上昇が可能

- 特に、時空間知識情報 (前後フレーム) を使用した際に大きく上昇

正しいクラスを維持したパターンにおける予測スコアの変化 (縦軸: 予測スコア, 横軸: 事後確率)

9

### まとめと今後の課題

10

- まとめ
  - 現実空間を反映した知識情報の構築
  - 機械学習による予測への適用で精度の向上を確認
  - 時空間知識情報を使用して、予測の信頼度の向上を確認
- 今後の課題
  - 時間的な依存関係として、前とその前の情報に着目
  - ある程度時間幅のあるデータに対してリアルタイム性を意識した予測
  - 異なる情報に着目したバイズ推定の適用
  - 速度や移動の向きなど
  - 別のデータセットに対して適用した際の評価の確認

10

### 使用したデータセット

11

- nuScenes [1]

観測車を走らせて、屋外のデータを 画像・LIDAR・RADAR データで取得  
ボストン・シンカポールで取得、連続した 40 フレームで構成された 1000 シーンを保持

- 主な物体カテゴリー
  - 四輪車  
car, truck, trailer  
construction vehicle, bus
  - 二輪車  
motorcycle, bicycle
  - その他  
pedestrian, barrier, traffic cone

屋外データの一部

[1] H. Caesar, et al., "nuScenes: A Multimodal Dataset for Autonomous Driving," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11618-11628, 2020

11

### 取得した空間的知識情報

12

- 空間的知識情報における特徴

→ 現実空間での特徴を反映

- 同一クラスの隣接確率が総じて高い
- 同じ種類の物体は付近に位置しやすい
- 取得した隣接情報
  - traffic cone の隣接 : traffic cone (65%), barrier (19%)
  - construction vehicle の隣接 : barriers (35%), traffic cone (20%)
  - 二輪車間でも異なる隣接 : motorcycle : motorcycle (49%), car (37%)  
bicycle : pedestrian (38%), bicycle (22%)
  - trailer の隣接として motorcycle は存在しない

12

## 取得した空間的知識情報

13

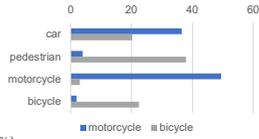
### ● 空間的知識情報における特徴

→ 現実空間での特徴を反映

- 同一クラスの隣接確率が総じて高い  
同じ種類の物体は付近に位置しやすい

#### ● 取得した隣接情報

- traffic cone の隣接 :  
traffic cone (65%)、barrier (19%)
- construction vehicle の隣接 :  
barriers (35%)、traffic cone (20%)
- 二輪車間でも異なる隣接 :  
motorcycle : motorcycle (49%)、car (37%)  
bicycle : pedestrian (38%)、bicycle (22%)
- trailer の隣接として motorcycle は存在しない



13

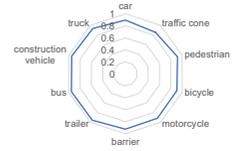
## 取得した時間的知識情報

14

### ● 時間的知識情報における特徴

→ 現実空間での特徴を反映

- 全てのクラスで高い確率を保持  
死角に入らなければ、ほとんどの物体は観測し続けることが可能
- car と traffic cone が比較的低い  
car
- 車道で取得したデータのため、後方の車やトラックなどの大型車の陰に隠れてしまう頻度が高い  
traffic cone
- 物体の大きさが小さいため、観測車からの死角に入りやすい



14