公益財団法人国際高等研究所　エジソンの会　2017年7月24日

自動運転における
センシングとディジタル地図との周辺技術

東京大学　情報学環
上條俊介

Outline

1. Introduction

2. Localization and mapping with Active Sensor

3. Localization with Passive Sensors

4. Pedestrian Detection and Behavior Recognition

5. Platooning on Highway

6. Policy and Definition of Automated Vehicles

7. Research Introductions

- 3D-GNSS and self-localization of vehicles
- P2V application for pedestrian safety
- Pedestrian detection and behavior understanding by Onboard cameras

Autonomous Driving

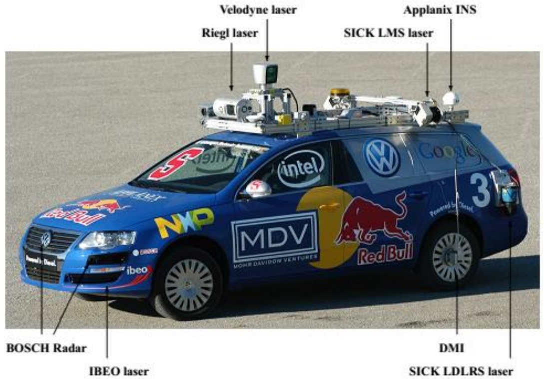


Fig. 2: Junior, the vehicle of Stanford group in the DARPA Urban Challenge 2007 .

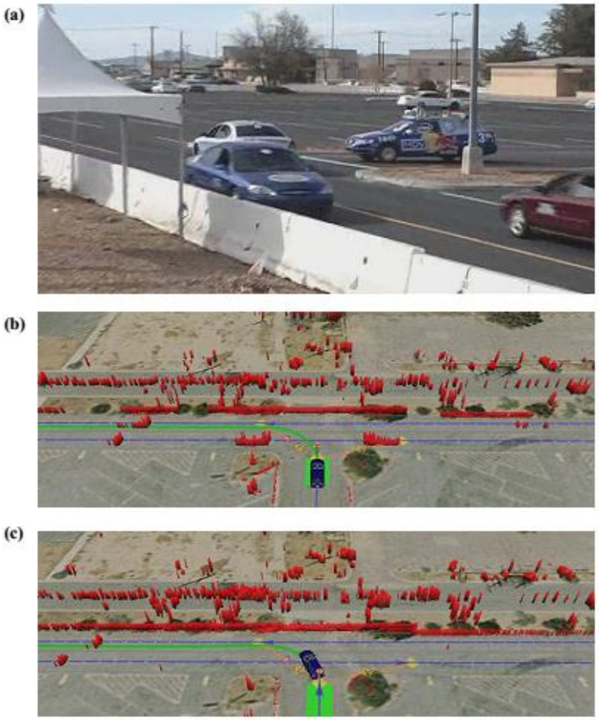


Fig. 3: Merging into dense traffic during the qualification events at the Urban Challenge. (a) Photo of merging test; (b)-(c) The merging process.

Fig. 1: Google car

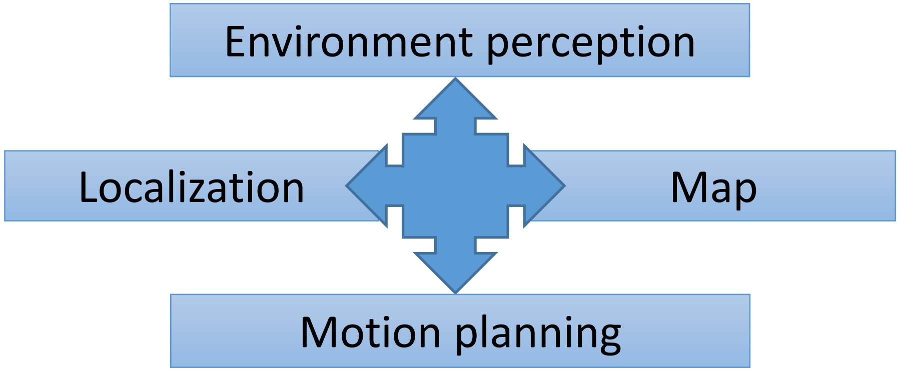
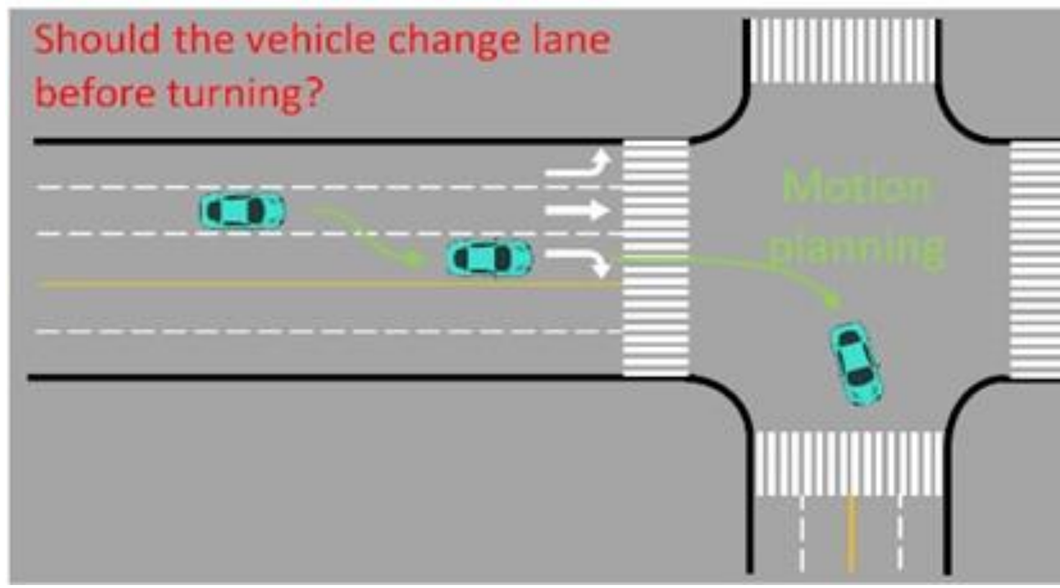


Fig. 4 Basic requirements for autonomous driving

Montemerlo, Michael, Jan Becker, Suhrid Bhat, Hendrik Dahlkamp, Dmitri Dolgov, Scott Ettinger, Dirk Haehnel et al. "Junior: The stanford entry in the urban challenge." *Journal of field Robotics* 25, no. 9 (2008): 569-597.

Urban Scenario of Autonomous Driving

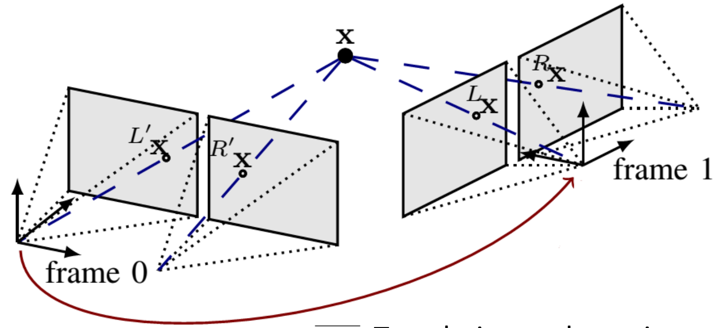
- Lane-level localization is required for autonomous driving.
- Lane-level map is needed in self-localization and motion planning.
- Active sensor:
  - LIDAR
  - Velodyne
- Passive sensors:
  - GNSS
  - Inertial sensor
  - Camera



What is Autonomous and Automated Driving

- Self-Localization
- Obstacle Detection
- Path Planning
- Autonomous or Automated ?

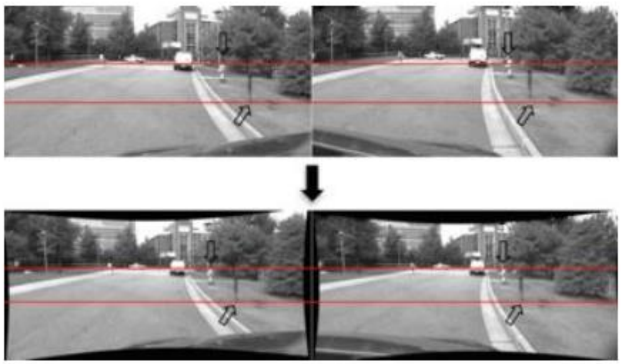
Stereo odometry



$\mathbf{p}_\Delta, \overline{\mathbf{q}_\Delta}$ Translation and rotation matrix

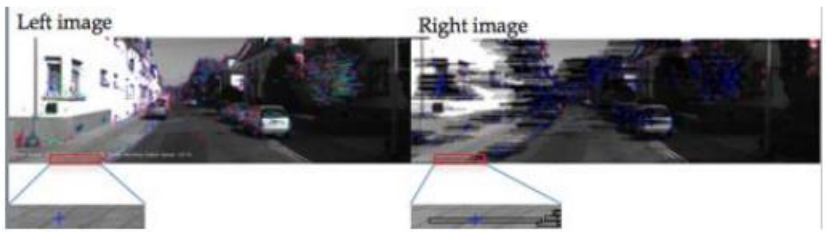
A single feature **x** viewed with a stereo camera from two different poses in consecutive image pairs.

1. Rectification



2. Feature Extraction

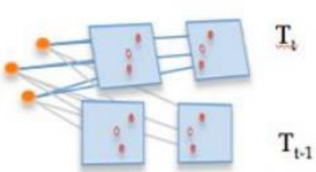
3. Stereo Feature Matching



4. Temporal Feature Matching

Given x position in frame 0 and frame 1, want the relationship (translation and rotation) between frame 0 and frame 1

5. Incremental Pose Recovery/RANSAC



3D-Map Construction from multiple data

Visualization of the scanning process: the LIDAR scanner acquires range data *and* infrared ground reflectivity. The resulting maps therefore are 3-D infrared images of the ground reflectivity. Notice that lane markings have much higher reflectivity than pavement.

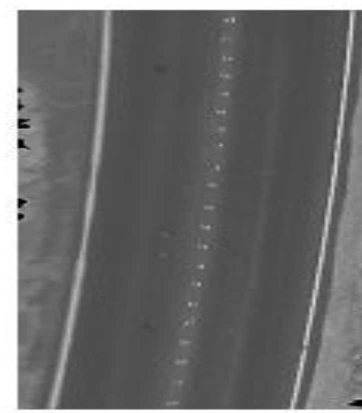
Ghost based by GPS in multiple data

After registration

- Two problems
- dynamic object
- multiple data fusion

Hole cased by dynamic object

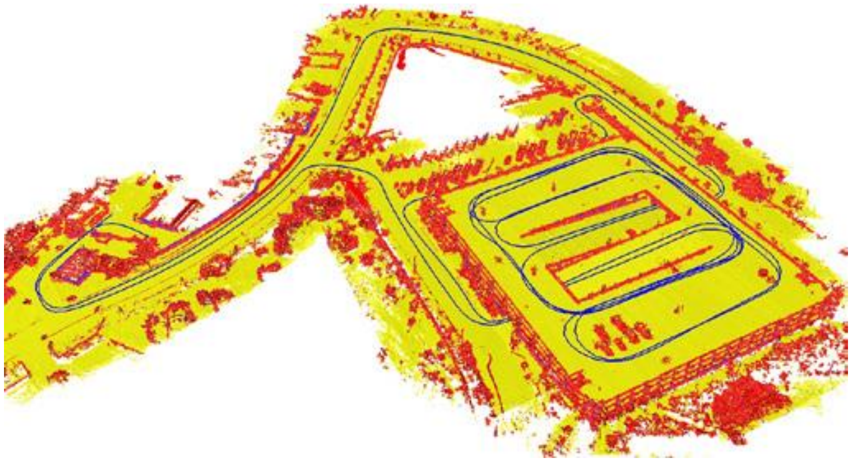
Remove hole using data fusion



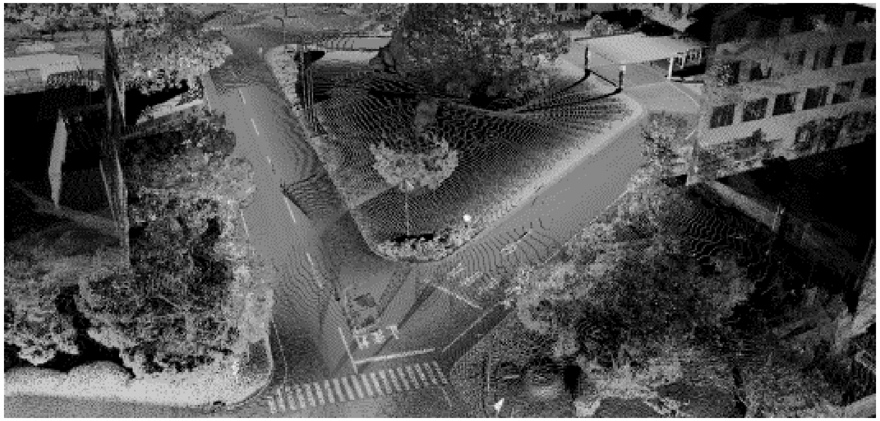
After registration

J. Levinson, M. Montemerlo, S.Thrun,“ Map-based precision vehicle localization in urban environments,” Robotics: Science and Systems. vol. 4, p. 1, June 2007.

Examples of point cloud map



Stanford parking garage

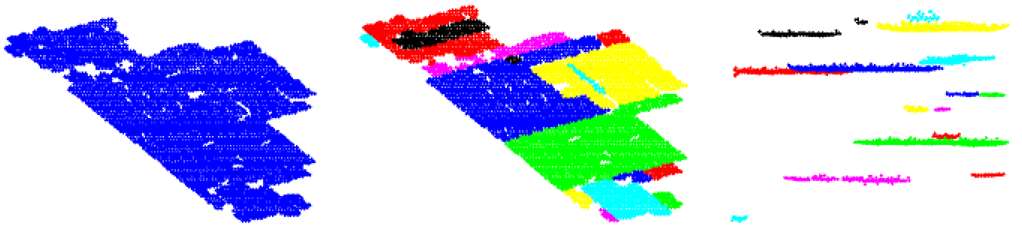


Toyota Technological Institute

University of Freiburg

Point cloud data to 3D map

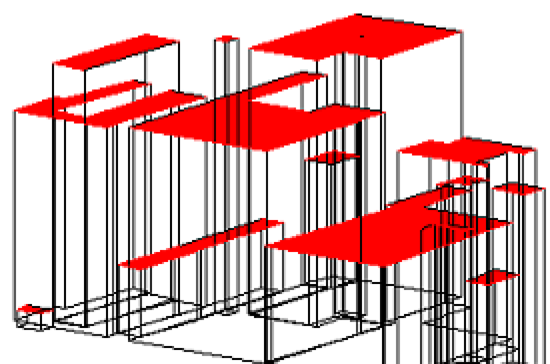
- **Data Sources**
  Airborne Laser Scanning
- **Building segmentation**



(a) (b) (c)

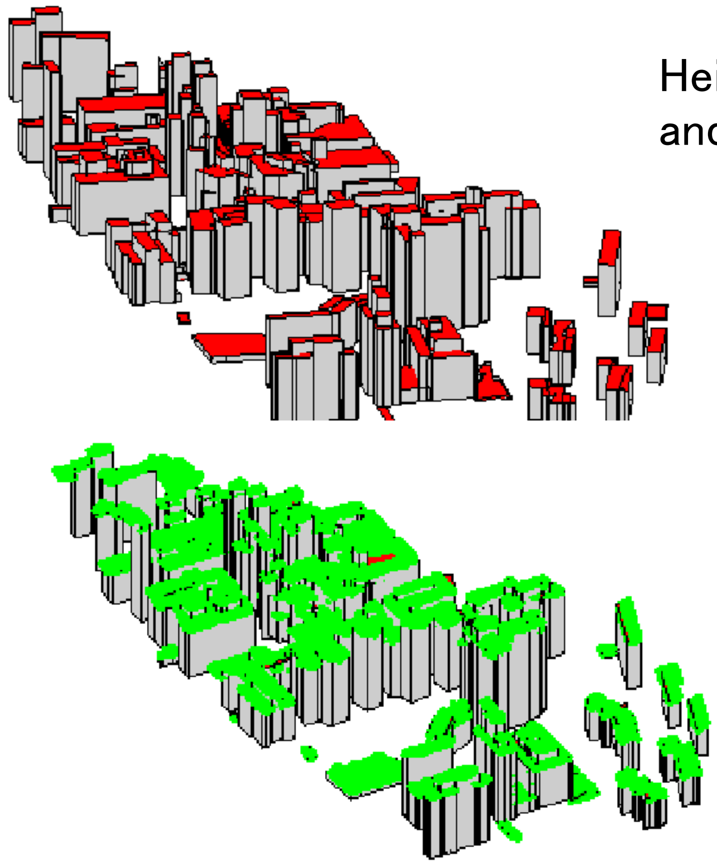
Example of roof patch segmentation with height jumps. (**a**) ALS building points; (**b**) detected roof patches shown in different colors; (**c**) roof patches at different heights.

- **Extract the outline →3D building model**



Zhu, L., Lehtomäki, M., Hyyppä, J., Puttonen, E., Krooks, A., & Hyyppä, H. (2015). Automated 3D Scene Reconstruction from Open Geospatial Data Sources: Airborne Laser Scanning and a 2D Topographic Database. *Remote Sensing*, *7*(6), 6710-6740.

Building model height evaluation



Height differences between the ALS building roof points and generated 3D roof models in 15 test locations

Test Location Index	Height of ALS Building Points (m)	Height of Building Models (m)	Height Difference (m)
1	31.61	31.59	0.02
2	26.84	26.74	0.10
3	20.04	19.91	0.13
4	20.76	20.52	0.24
5	25.04	24.93	0.11
6	23.81	23.76	0.05
7	22.72	22.44	0.28
8	26.86	26.61	0.25
9	24.91	24.86	0.05
10	25.36	25.10	0.26
11	23.21	23.02	0.19
12	21.81	21.57	0.24
13	25.89	25.81	0.08
14	20.51	20.33	0.18
15	19.72	19.58	0.14
Average			0.15
RMSE Root Mean Square Error			0.18

(**Up**) Generated Building models, with roofs in red and walls in gray. (**down**) Original ALS building roof points in green and building models in gray.

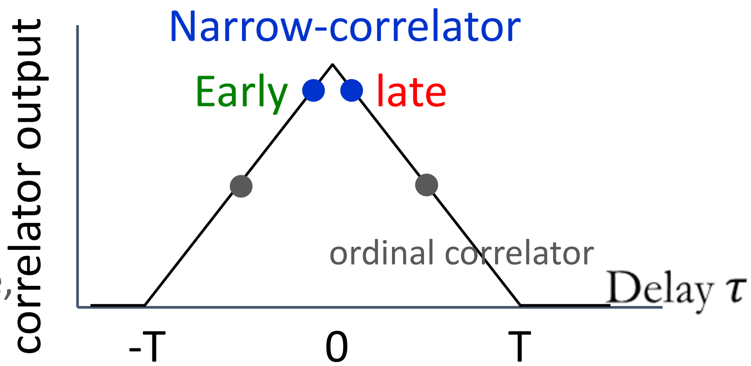
- **How about the accuracy of the boundary?**

Zhu, L., Lehtomäki, M., Hyyppä, J., Puttonen, E., Krooks, A., & Hyyppä, H. (2015). Automated 3D Scene Reconstruction from Open Geospatial Data Sources: Airborne Laser Scanning and a 2D Topographic Database. *Remote Sensing*, *7*(6), 6710-6740.

Localization with GNSS

- Various multipath mitigating receivers
    - Narrow-correlator (Dierendonck, 1992)
    - Strobe-correlator (Garin, 1997)

M.S. Braasch. “Performance comparison of multipath mitigating receiver architectures”. In Aerospace Conference, IEEE Proceedings., 2001.



- Receiver Autonomous Integrity Monitoring algorithm (RAIM)
    - Check the residual of the least square and identifies the suspicious satellites.
    - Choose satellites in calculation to make the least square residual small

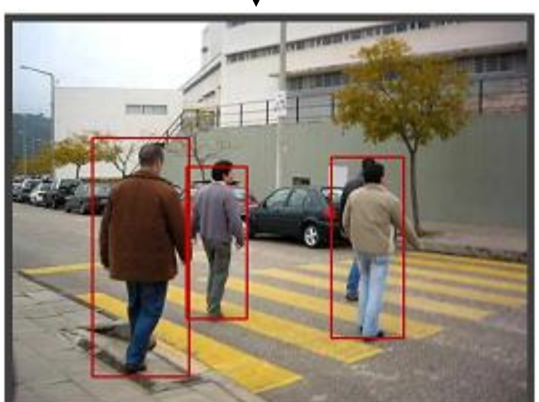
RG Brown., "GPS RAIM: Calculation of Thresholds and Protection Radius Using Chi-square Methods; a Geometric Approach.", 1994.

- NLOS detection by 3D map (M. Obst, et al., 2012.)

Pedestrian detection example

training

Test



(a)

Haar–like features based detection

(b)

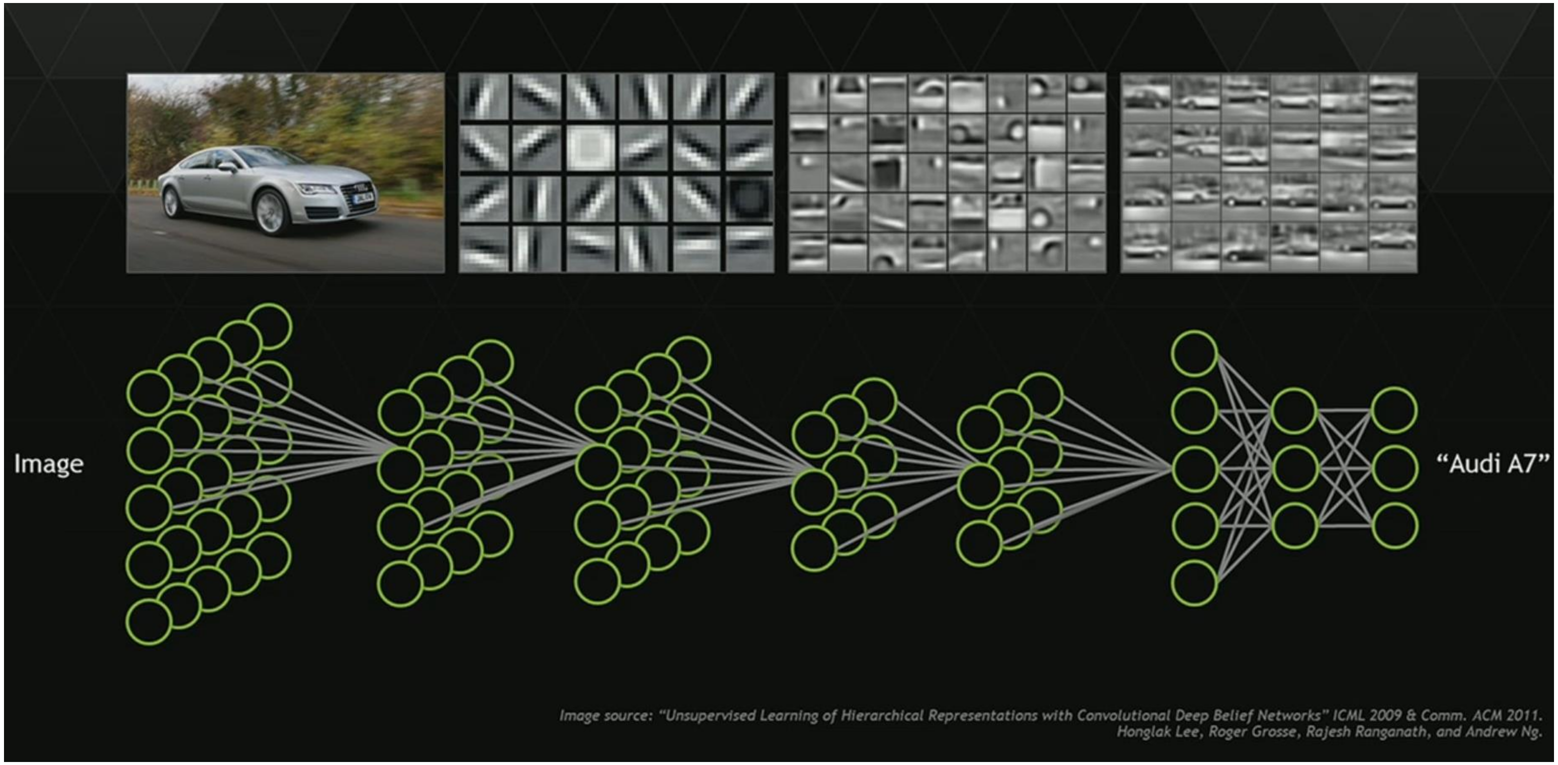
Verify each candidate by high dimensional CoHOG descriptors

- HOG and Haar-like features are popular and basic.
- CoHOG has good performance at low resolution

Monteiro, G., Peixoto, P., & Nunes, U. (2006). Vision-based pedestrian detection using Haar-like features. *Robotica*, *24*, 46-50.

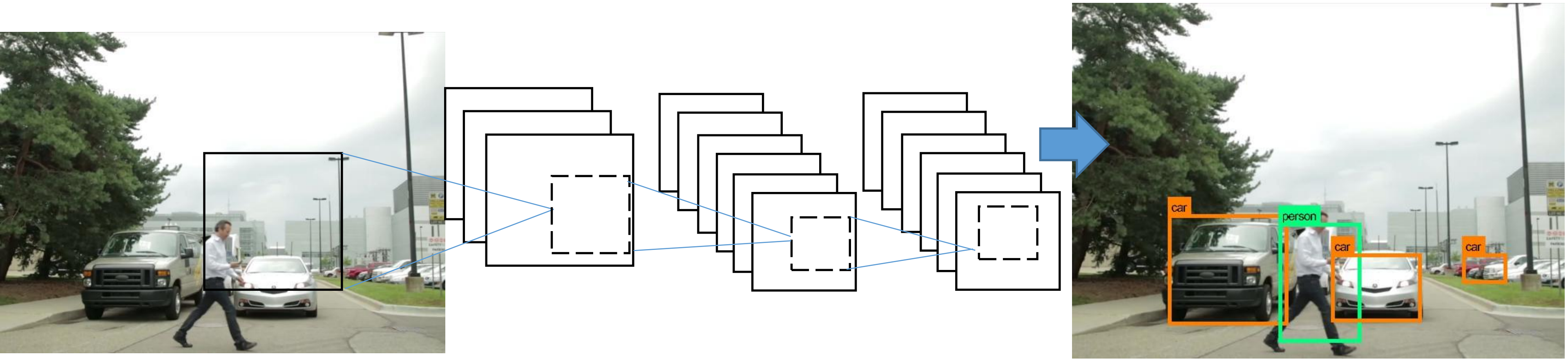
Convolutional Neural Network (CNN)

- How a deep neural network sees
- From local texture to global structure



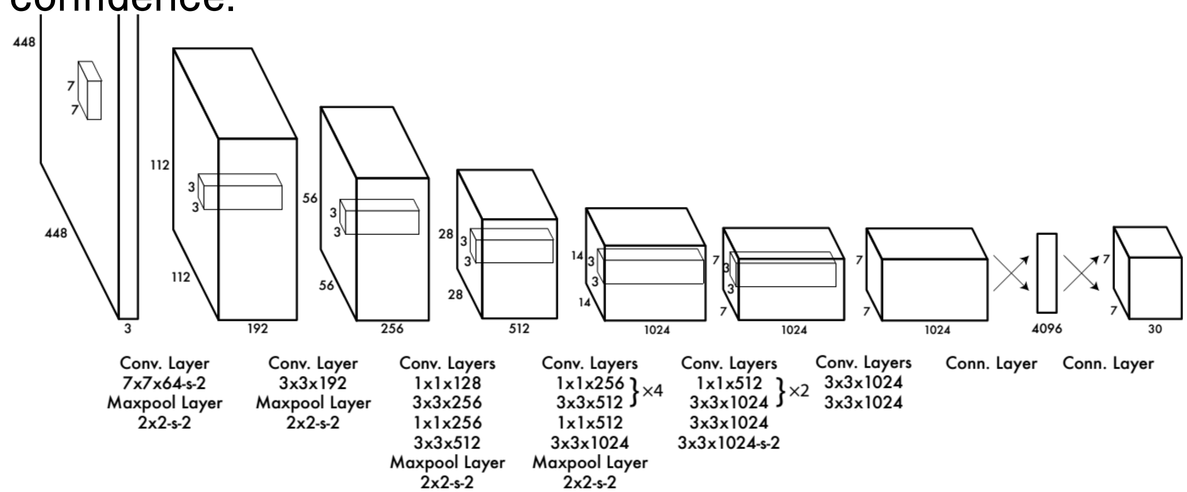
Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2011). Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM*, *54*(10), 95-103.

Application of CNN



The YOLO (You Only Look Once) Detection System. Processing images with YOLO is simple and straightforward. The system (1) resizes the input image to F × F, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence.

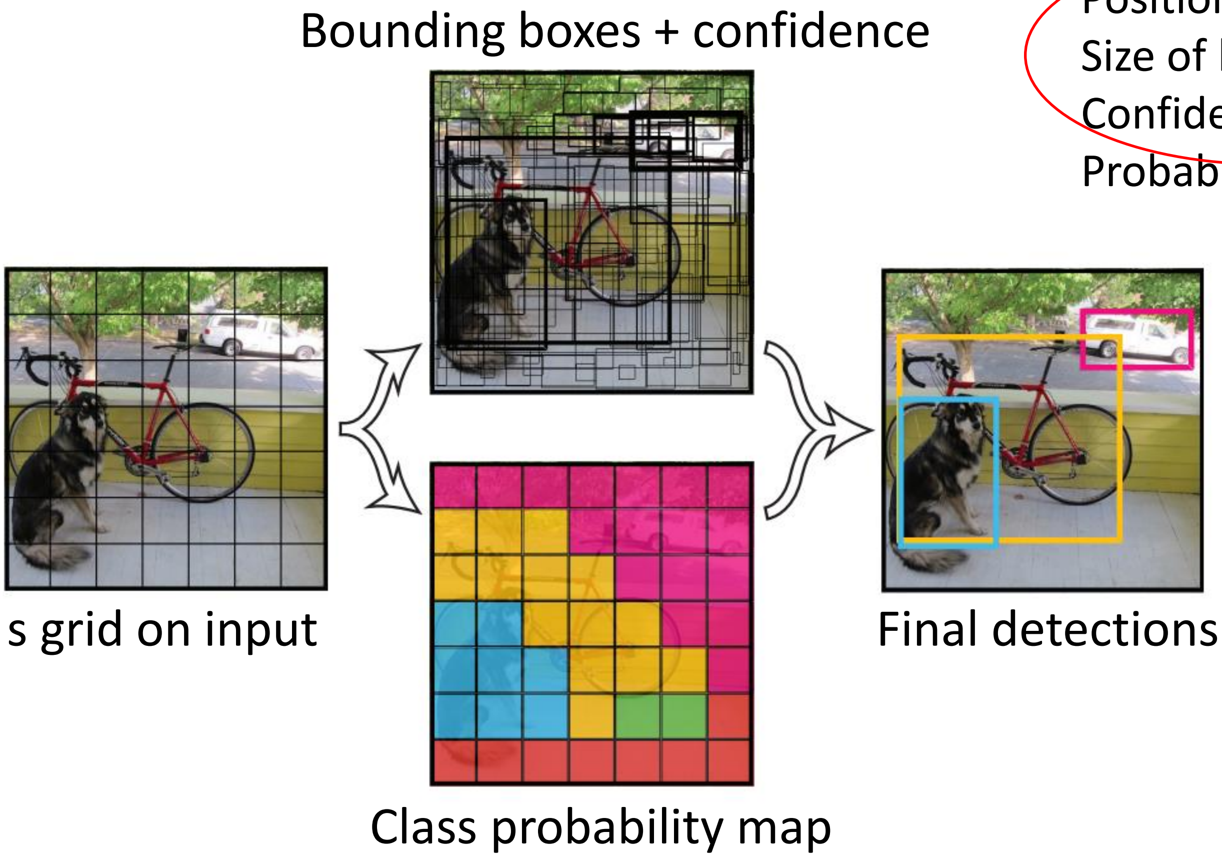
The Architecture of YOLO



Detection framework of YOLO

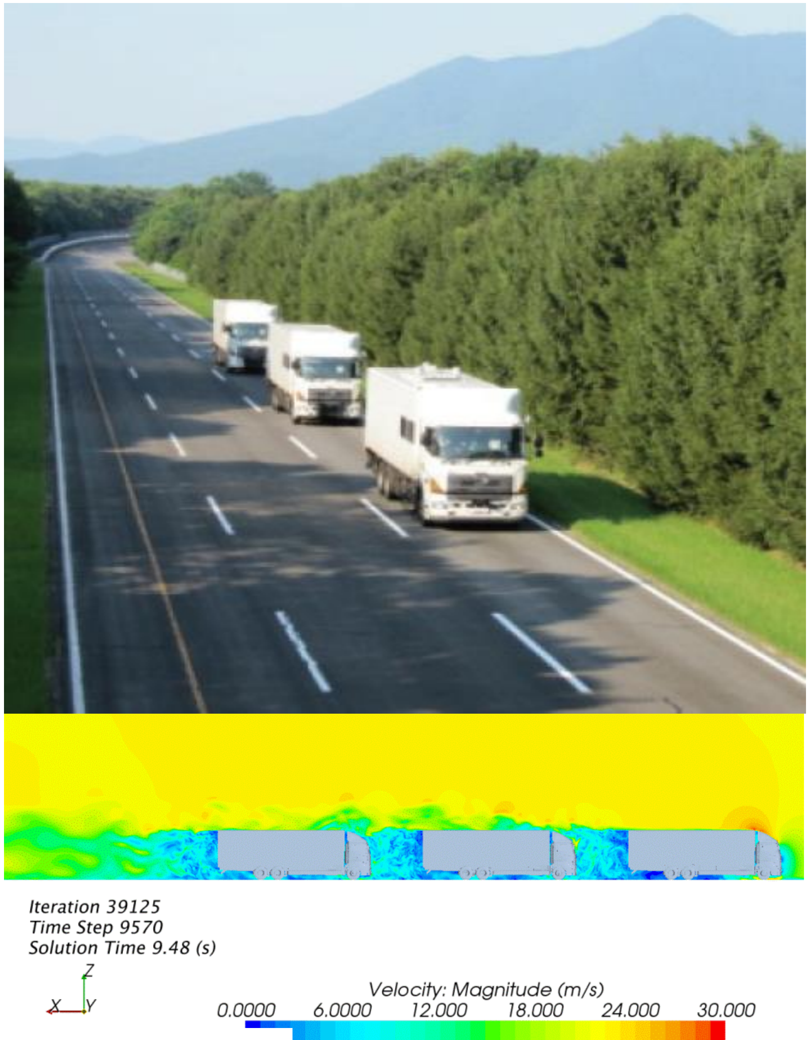
It divides the image into an even grid and simultaneously predicts bounding boxes, confidence in those boxes, and class probabilities.
These predictions are encoded as an $S \times S \times (B * 5 + C)$ tensor.

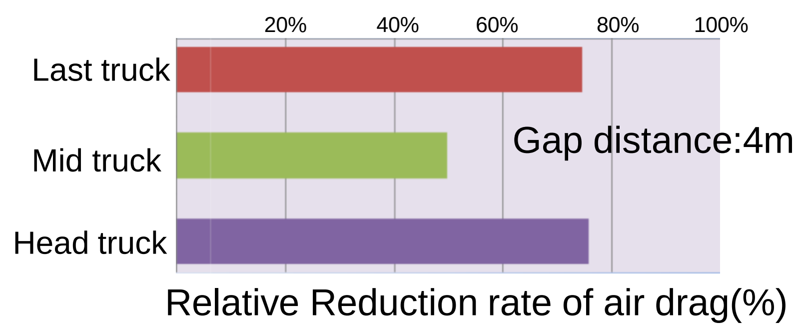
Position of box (x, y)
Size of box (w, h)
Confidence: conf
Probability of each class: $p_{1 \ldots}\ p_c$



Automated　platoon

Automated platoon　is the string operation coupled electrically with closed gap distance in order to reduce fuel consumption.





Level of Automated Vehicles : NHTSA's Policy

No-Automation (Level 0): The driver is in complete and sole control of the primary vehicle controls – brake, steering, throttle, and motive power – at all times.

Function-specific Automation (Level 1): Automation at this level involves one or more specific control functions. Examples include electronic stability control or pre-charged brakes, where the vehicle automatically assists with braking to enable the driver to regain control of the vehicle or stop faster than possible by acting alone.

Combined Function Automation (Level 2): This level involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. An example of combined functions enabling a Level 2 system is adaptive cruise control in combination with lane centering.

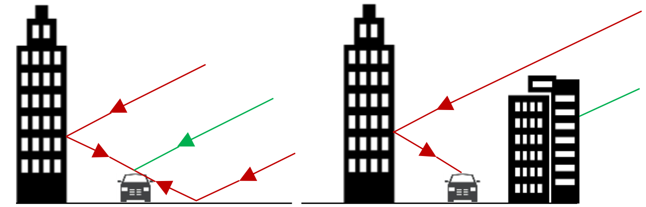
Limited Self-Driving Automation (Level 3): Vehicles at this level of automation enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control. The driver is expected to be available for occasional control, but with sufficiently comfortable transition time. The Google car is an example of limited self-driving automation.

Full Self-Driving Automation (Level 4): The vehicle is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip. Such a design anticipates that the driver will provide destination or navigation input, but is not expected to be available for control at any time during the trip. This includes both occupied and unoccupied vehicles

Preliminary Statement of Policy Concerning Automated Vehicles (NHTSA)

Mobile Mapping Challenge in Urban Area

- **Inaccurate MMS positioning** in urban area with tall buildings
- Reason: GNSS error due to **Non-Line-of-Sight** and **Multi-Path**
- Conventional Solution: **Landmark updating method**
  1. Measure 3D Coordinate of Ground Control Points (GCP)
  2. Extract GCPs from original data
  3. Calculate Position Correction Vectors (PCV)
  4. Correct MMS trajectory and point cloud based on PCVs
- More than 40 GCPs are required for a 500 x 300m area

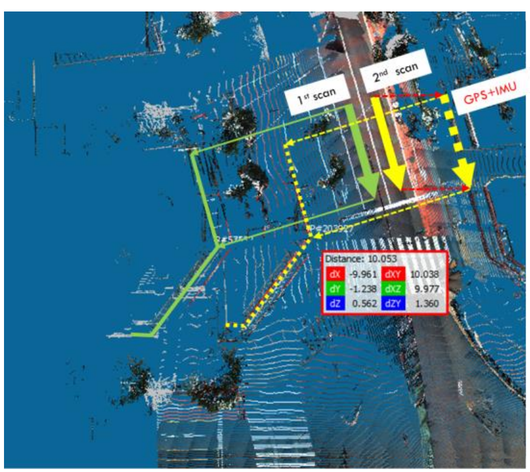


Multi-path NLOS

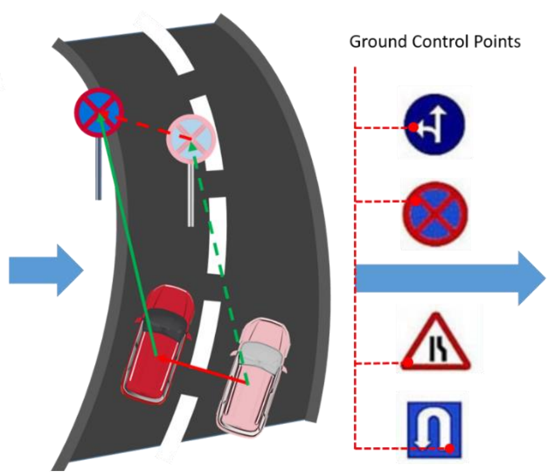
✕ Labor intensive 

✕ Time consuming 

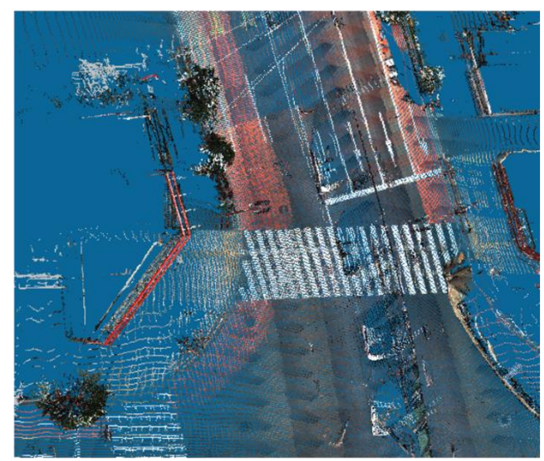
✕ Costly 



MMS error before correction

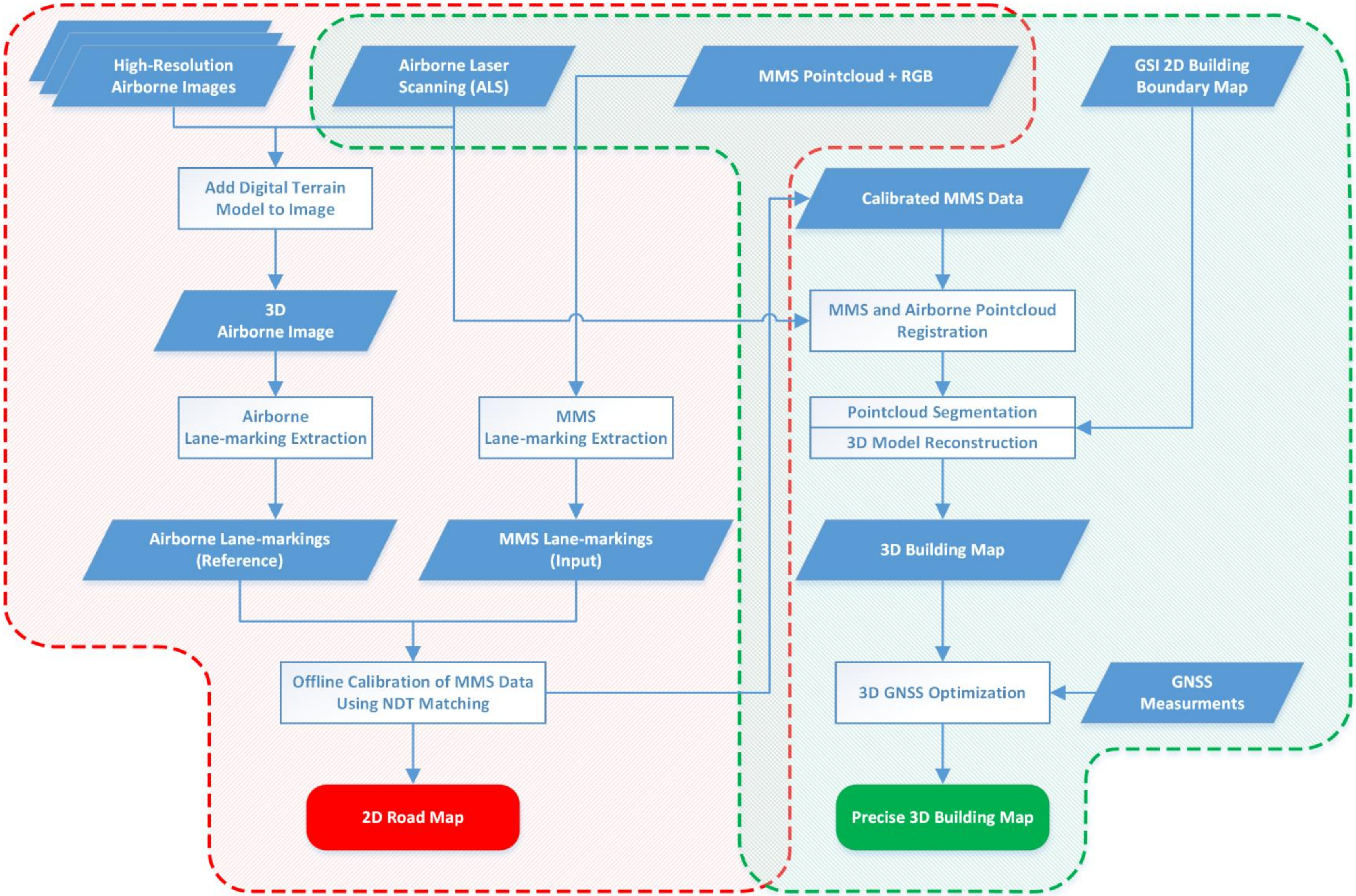


Landmark updating

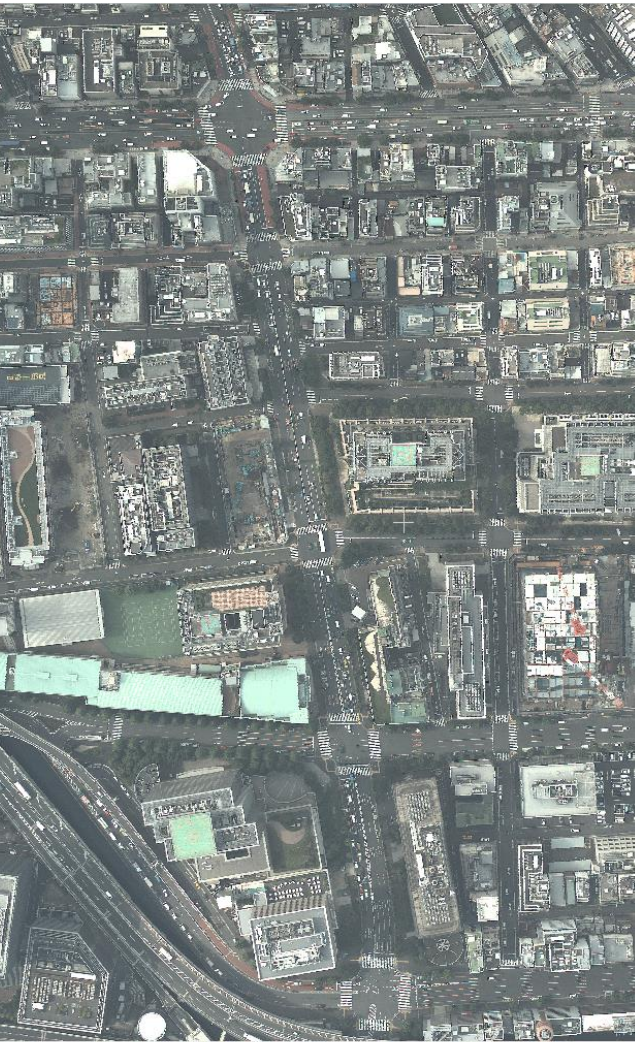


MMS data after correction

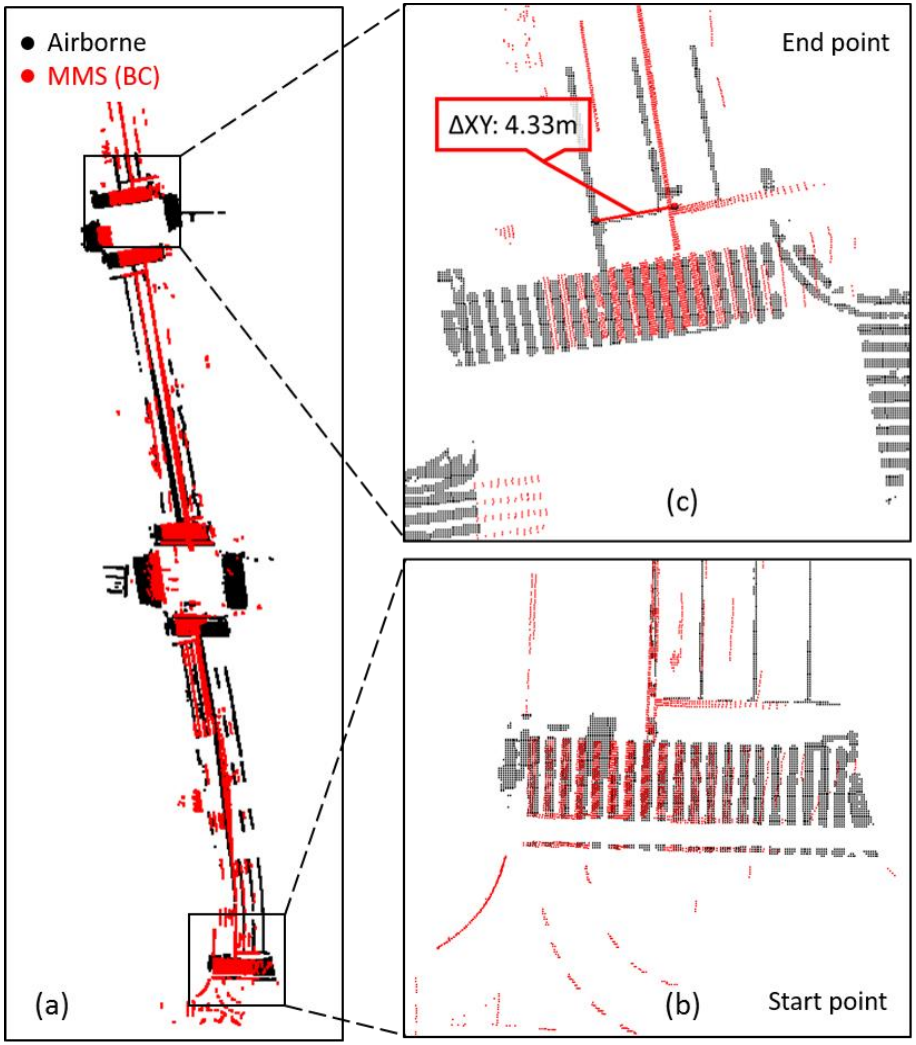
Map Reconstruction Framework



Precision Mobile Mapping (Problem statement)
Registering MMS Data to Aerial image based on Road markings



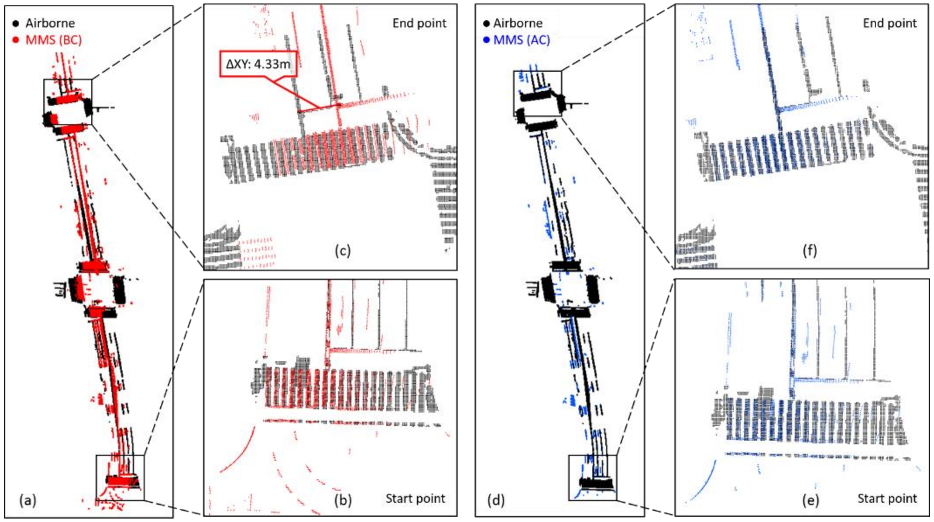
Aerial image



Extracted road markings from both MMS scan and aerial image

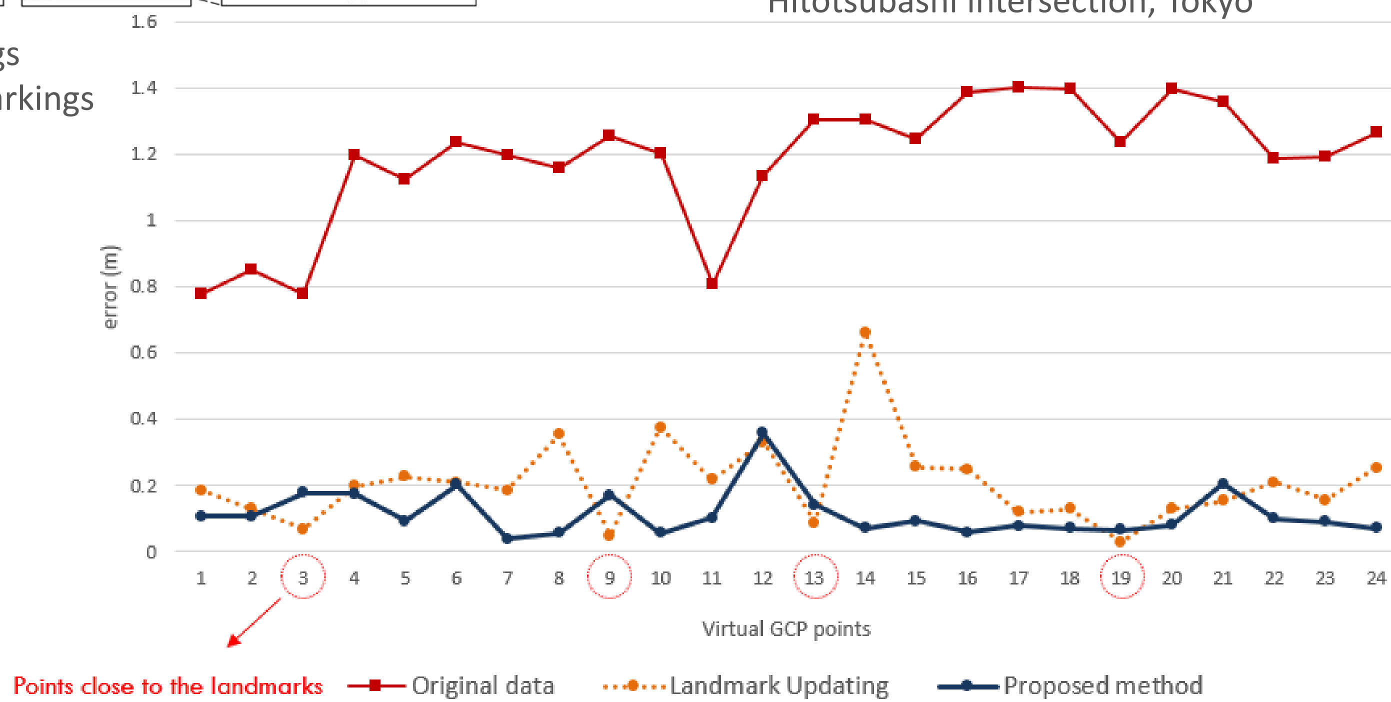
MMS Scan

Precision Mobile Mapping (Results and Evaluation)

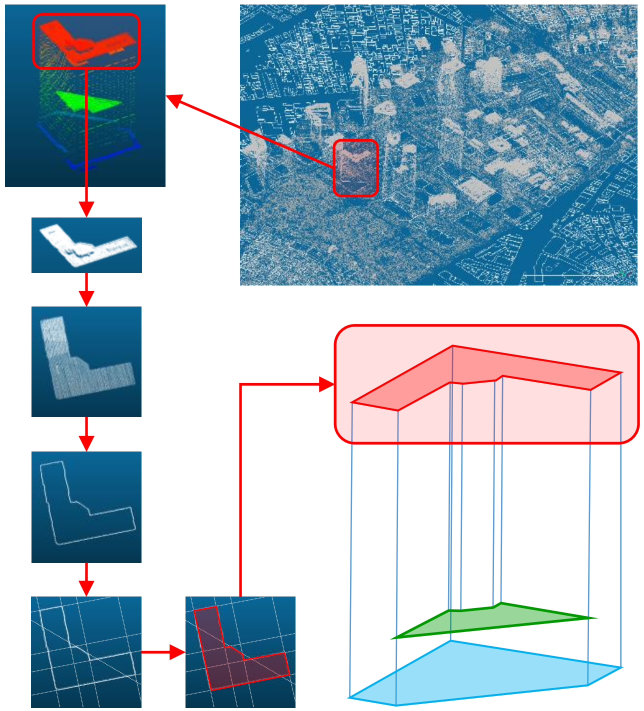


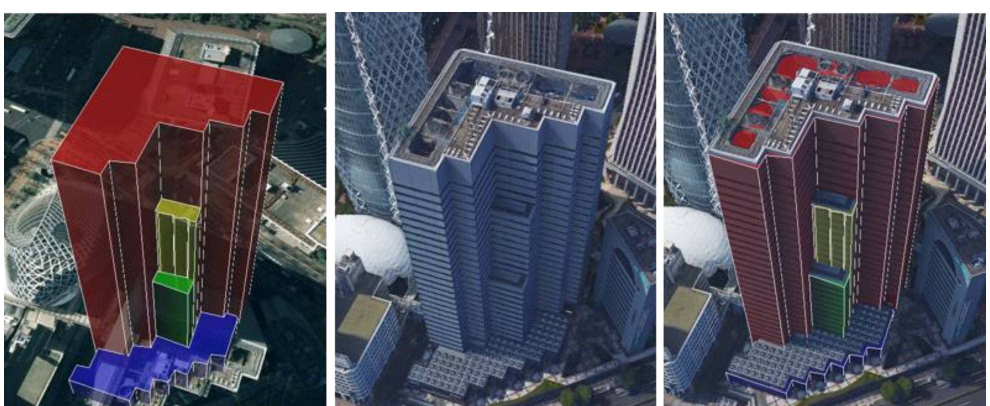
- Airborne road markings
- Original MMS road markings
- Proposed method

Hitotsubashi intersection, Tokyo

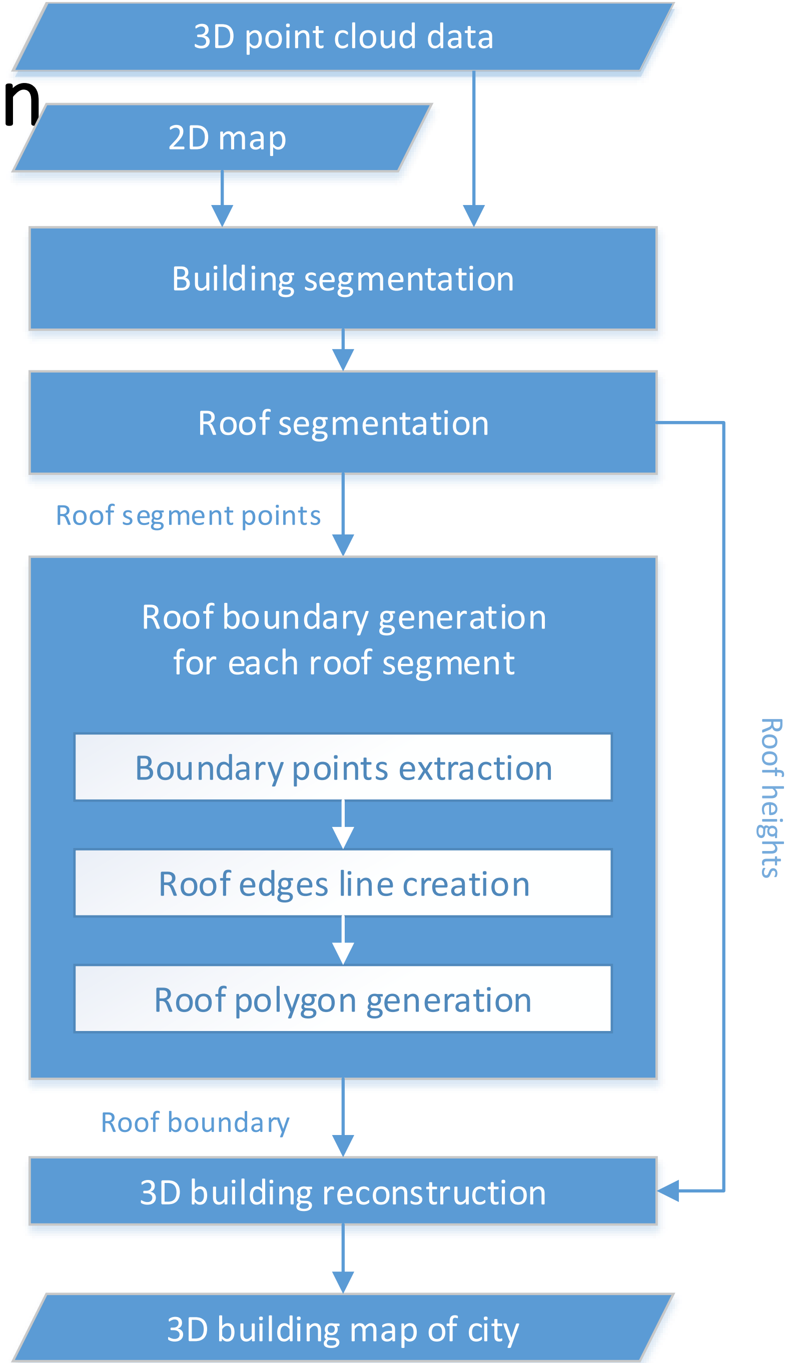


3D Building Map Reconstruction

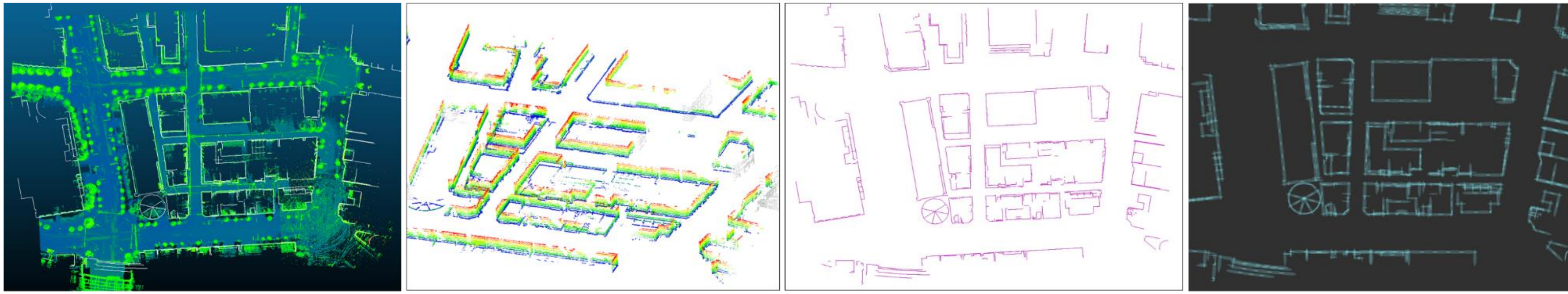




Reconstructed 3D building model using proposed method

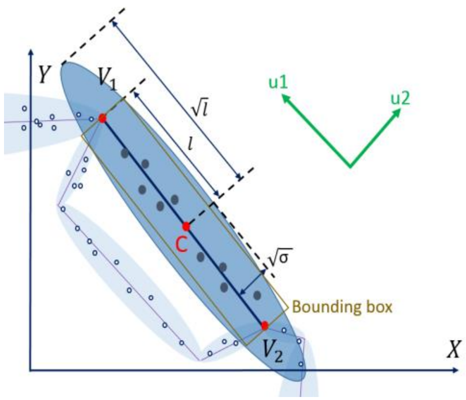


Generating Multilayer Vector Map

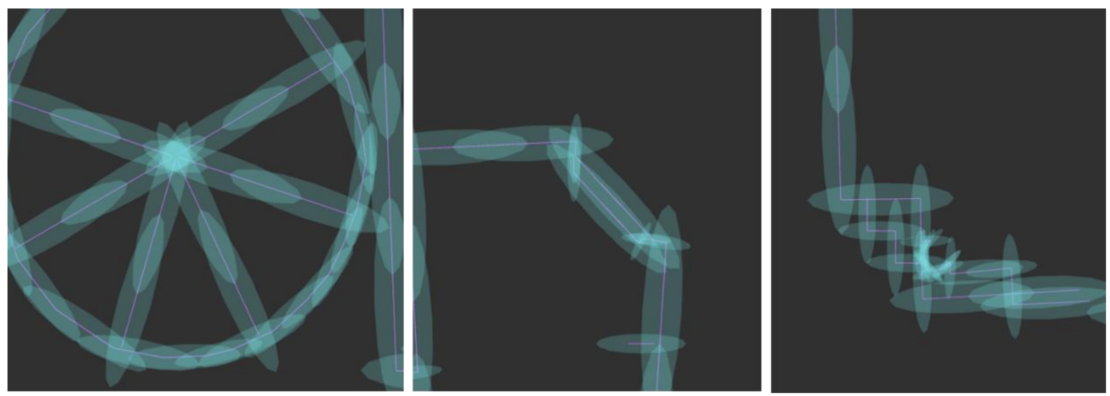


(a) 3D point cloud → (b) Extracted building footprints → (c) Multilayer 2D vector Map → (d) ND map for vectors

Points that made a vector form a normal distribution



Generated Normal distribution form vector map



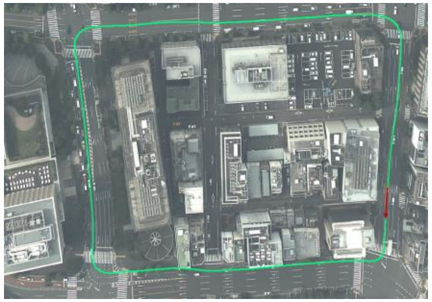
$$P(\vec{x}) = \frac{1}{2\pi\sqrt{|\Sigma|}}\exp\left(-\frac{(\vec{y}_k-\vec{\mu})^T\Sigma^{-1}(\vec{y}_k-\vec{\mu})}{2}\right)$$

$Y=\{\vec{y}_1, \ldots, \vec{y}_n\}$ Points that made a vector segment

$\vec{\mu}$ Mean of generated normal distribution

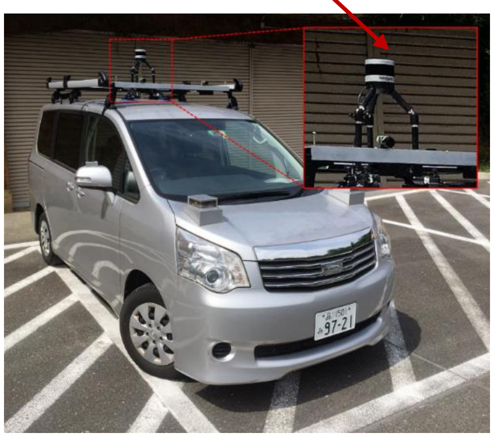
Σ Covariance of generated normal distribution

Experimental results



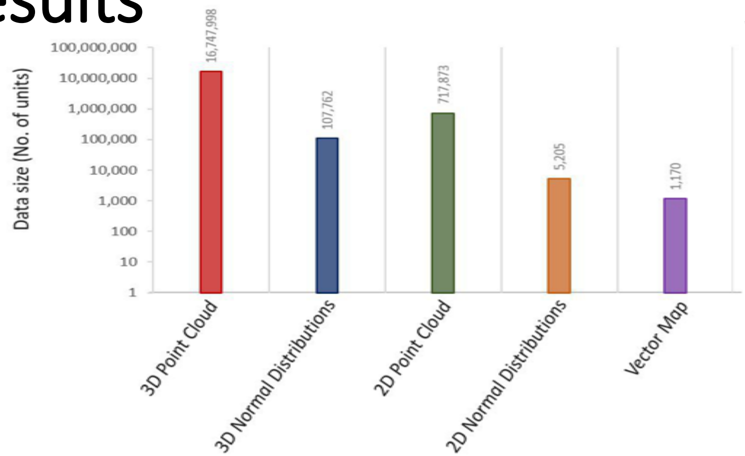
Route of experiments

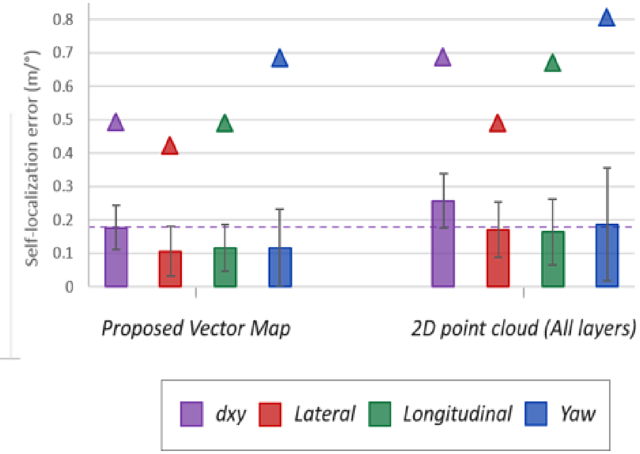
Velodyne’s VLP-16
(16 chennel)



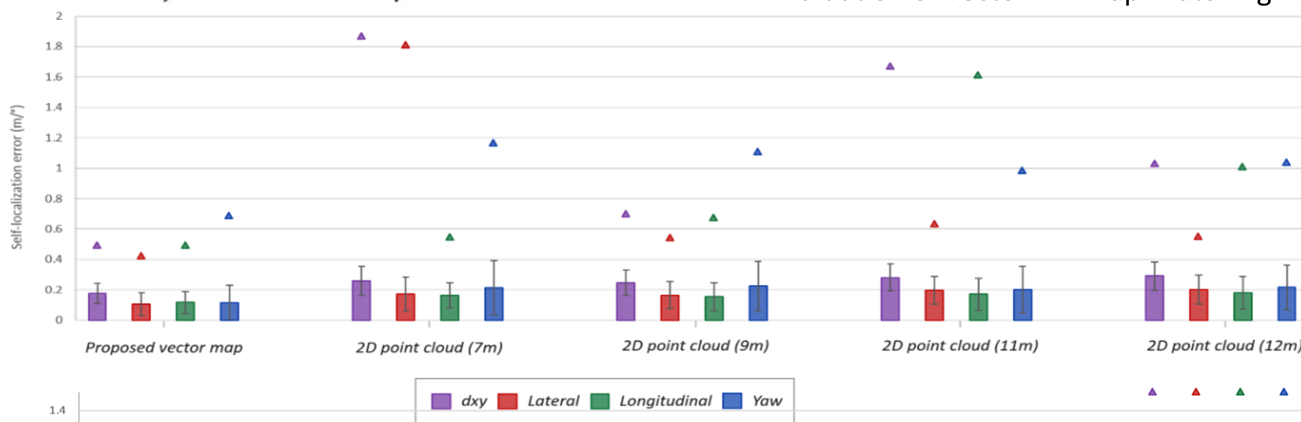
Our experimental vehicle

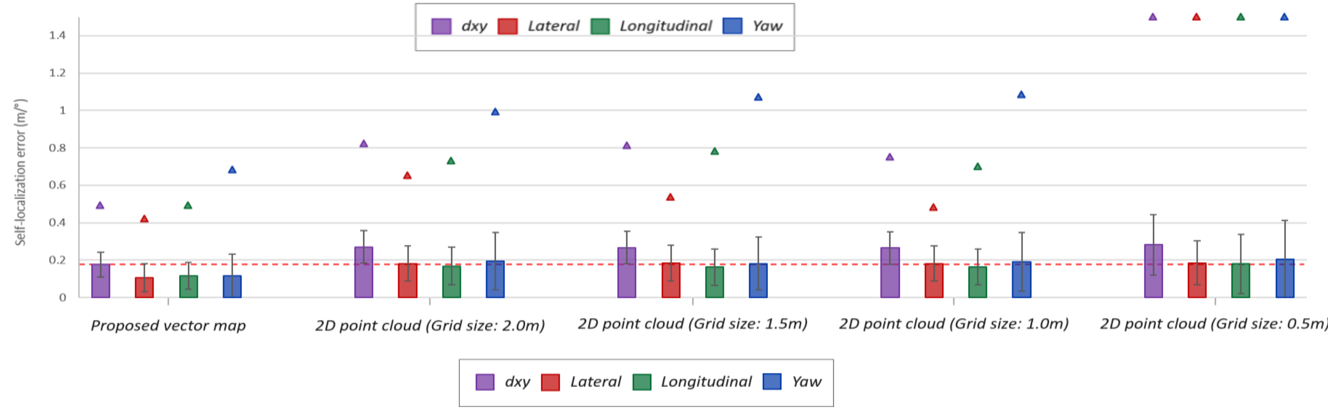
Evaluation of data size





Evaluation of vector ND map-matching



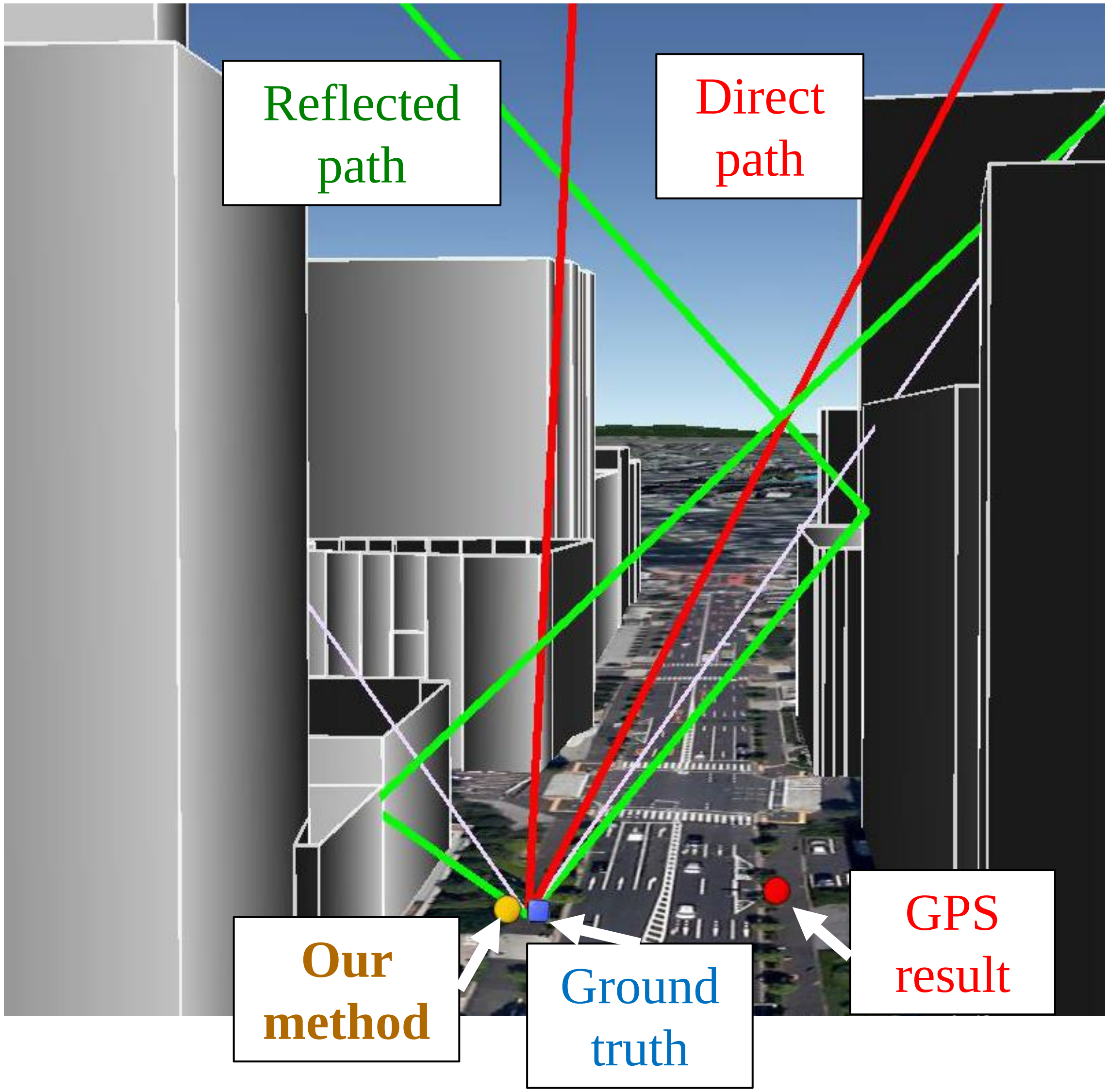


Evaluation of multilayer 2D vector structure (comparison with conventional 2D methods)

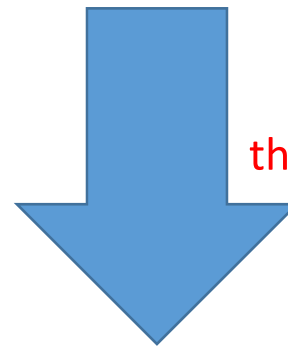
都市部における測位精度向上の課題設定

- 歩行者の測位精度の向上
  - ✓ 5m程度の測位誤差を実現すれば、道路や交差点のどの側にいるか、どの店舗の前にいるかを十分に判定できる。即ち、1m以内の測位誤差は不要。
  - ✓ 外国人等の土地勘がない旅行者へのナビゲーションの利便性、商業地区での歩行場所に依存した広告等のロケーションサービス。
  - ✓ 商業地区の都市計画のための詳細な行動データを取得できる。
  - ✓ 従来の測位誤差は、20m程度以上のため、道路のどちら側の歩道を歩いているかの判定が困難、交差点のどの角にいるかの判定が困難。
  - ✓ WiFiアクセスポイントからの受信強度(RSSI)計測を用いた測位は、GPS測位データをレファレンスとしたキャリブレーションが元となっているため、GPS測位誤差と同等の誤差を有する。
  - ✓ 本技術を活用することで、WiFiアクセスポイントを参照した測位の精度も向上できる。
- 自動車のポジショニング精度の向上
  - ✓ 信号制御の高度化への期待。
  - ✓ 車線を判定することで、右折需要と直進需要を分離して計測できる。
  - ✓ 自動運転においても車線判定は必須技術。
  - ✓ 車線判別のための測位への要求精度は1.5m（車線幅の半分程度の相当）であるが、車載のCANデータ、ジャイロデータとのヒュージョンで達成できる。

Algorithm to apply Ray-tracing to 3D map

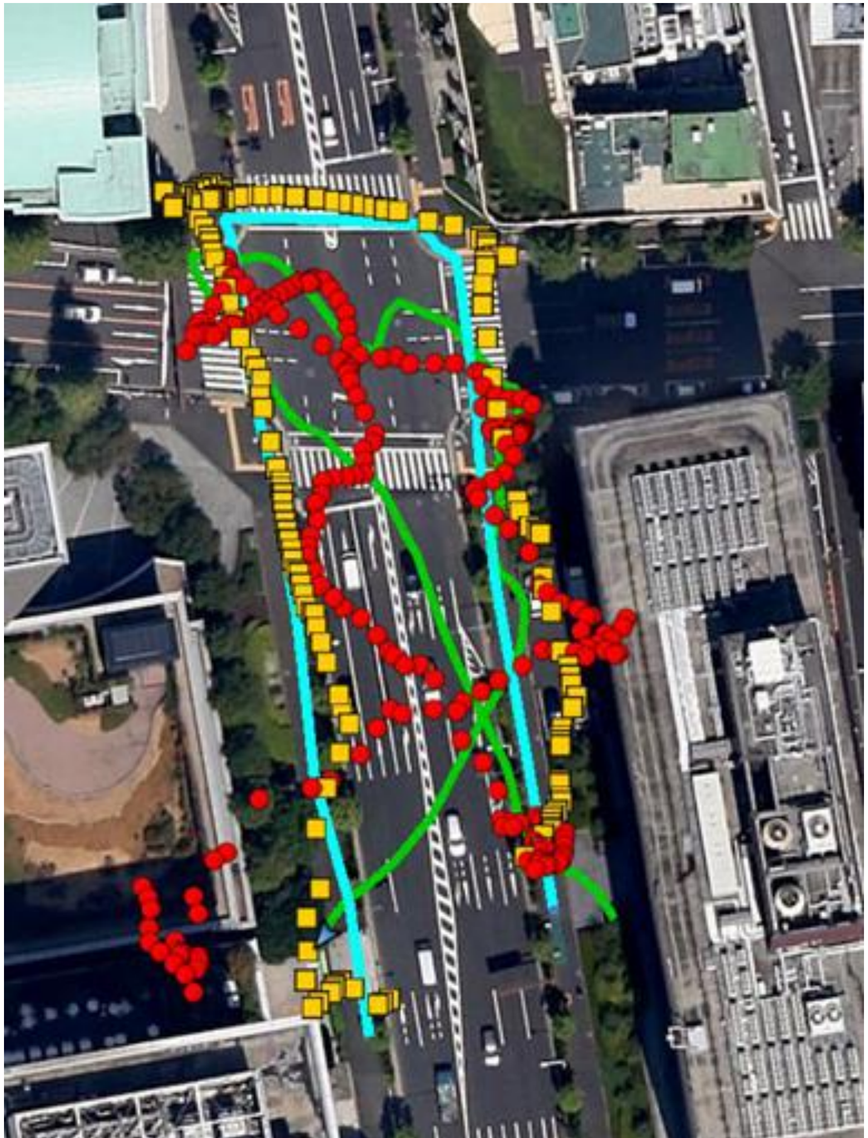


Signal Observation:
- Pseudorange
- RSSI (Received Signal Strength Indicator)
- Deceived positioning results



Position Assumptions:
Estimated by ray tracing
- Pseudorange
- RSSI (Received Signal Strength Indicator)
- Deceived positioning resuts

Evaluations with solutions to NLOS and Multipath problem

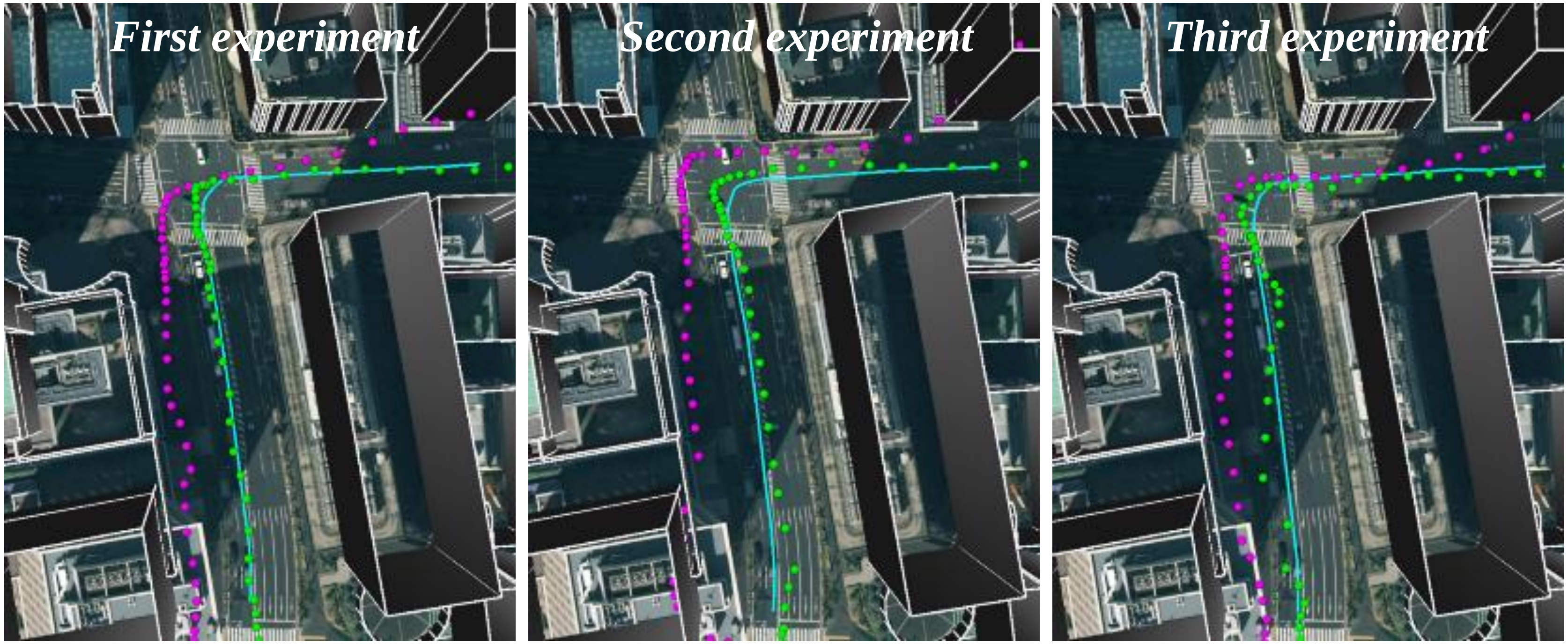


iPhone4S with WiFi
u-blox NEO-6P
Proposed (with NEO-6P)
Ground Truth

	Mean Error [m]	Standard deviation [m]
u-blox NEO-6P	19.8	14.2
3D map for NLOS	5.7	4.3
3D map for NLOS / Multipath	4.7	3.0

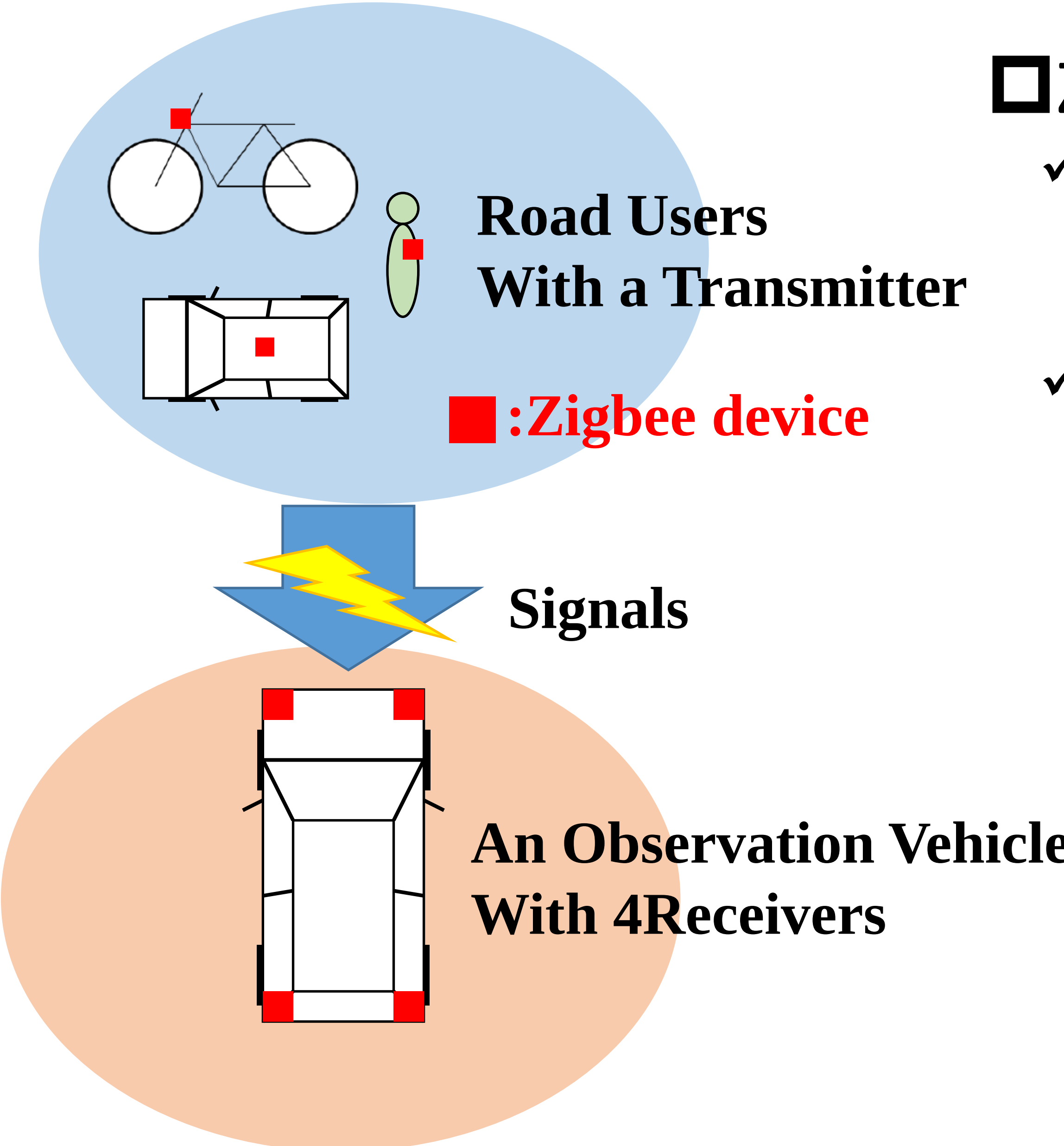
Experimental results: Vehicle sensor integration

Standard GPS (single point positioning) fusion 3D map GPS fusion



		Maximum error (m)	Mean error (m)	Standard Deviation (m)
First experiment	*Standard GPS fusion*	16.2	11.2	2.5
	3D map GPS fusion	3.9	**1.0**	0.7
Second experiment	*Standard GPS fusion*	26.7	13.7	3.5
	3D map GPS fusion	5.4	**2.2**	1.4
Third experiment	*Standard GPS fusion*	15.4	8.8	3.4
	3D map GPS fusion	6.6	**1.7**	1.4

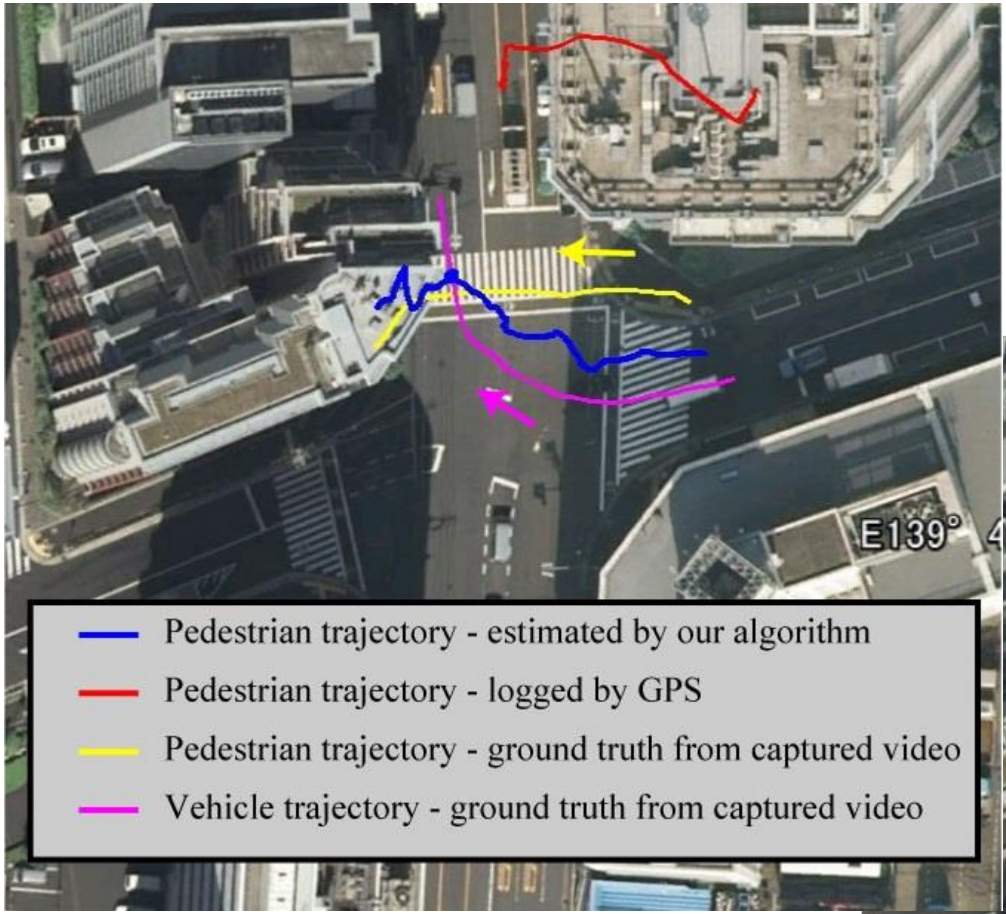
Zigbee for V2V and P2V applications.



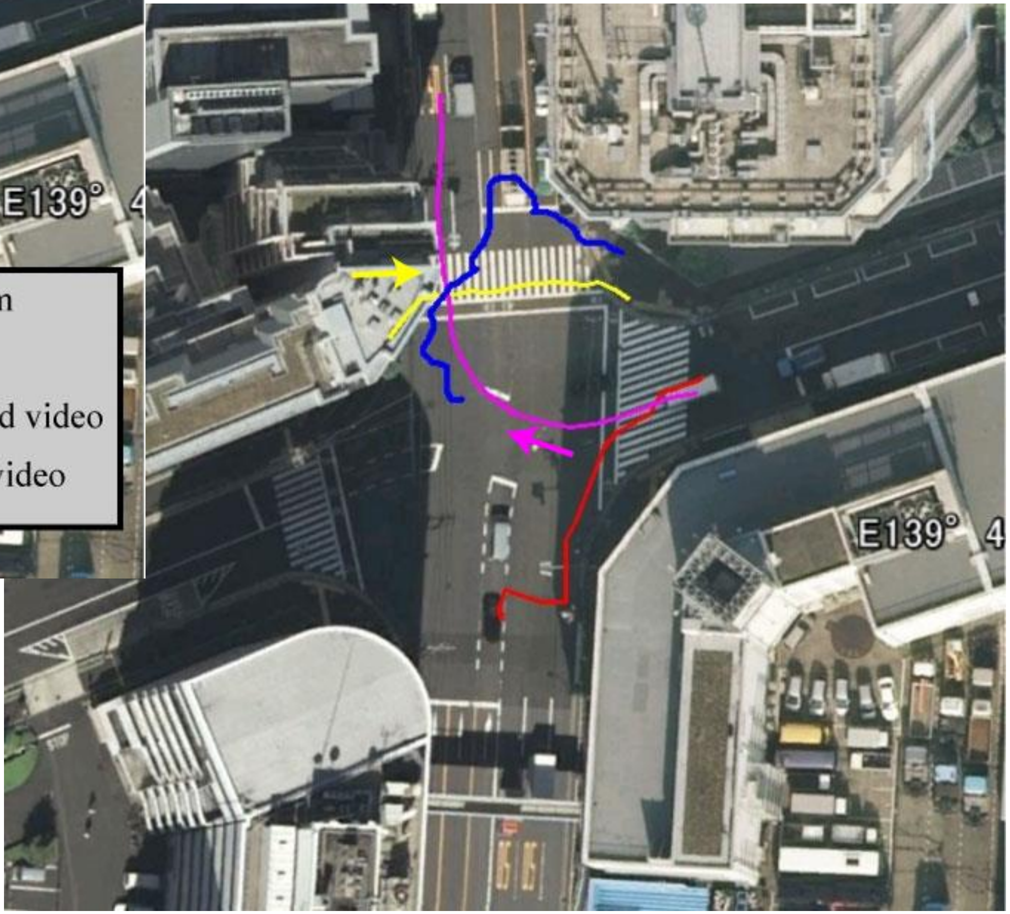
□Zigbee equipment

- ✓Bicycles and pedestrians equip a single Zigbee transmitter.
- ✓Observation vehicles equip four Zigbee receivers (and also a transmitter Zigbee).

Results of FOT at a Real Intersection(1)

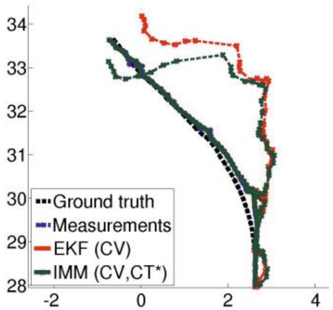


Right Turning Cases



Path Prediction - Problems

Path prediction and motion classification focusing on physical states result in short-time prediction (~1s).



1. Trajectory-based Approach

Linear Dynamical System approximation of positional movement.

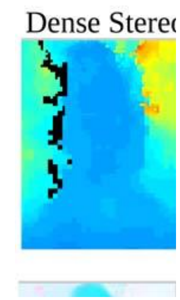
Motion-specific LDSs (walk, stop, turn, etc.).

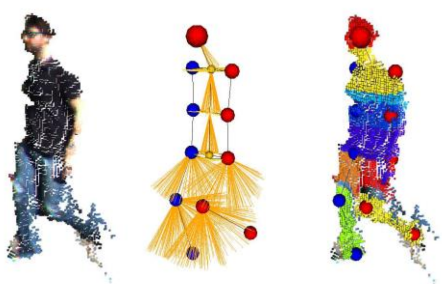
Nicolas Schneider and Dariu M. Gavrila., "Pedestrian path prediction with recursive bayesian filters: A comparative study," Pattern Recognition. Springer Berlin Heidelberg, 2013.

2. Image-based Approach

- Visual features on image plane and non-linear/high-order Markov models.

Christoph G. Keller, and Dariu M. Gavrila., “Will the Pedestrian Cross? A Study on Pedestrian Path Prediction,” Transactions on ITS 2014.





3. Pose-based Approach

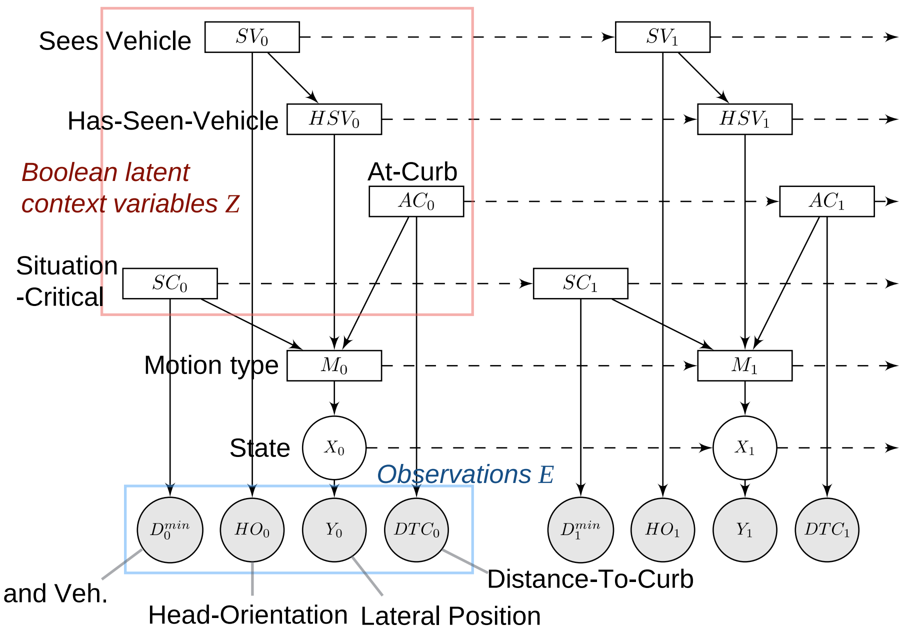
- Body parts and joints in 3D space which are robust against sensor ego-motion and change of the observing direction.

R. Quintero, et al., “Pedestrian Path Prediction using Body Language Traits” IV 2014.

For long-time prediction, it is crucial to consider **contexts** which determine pedestrian movement: traffic rule, situation and pedestrian intention.

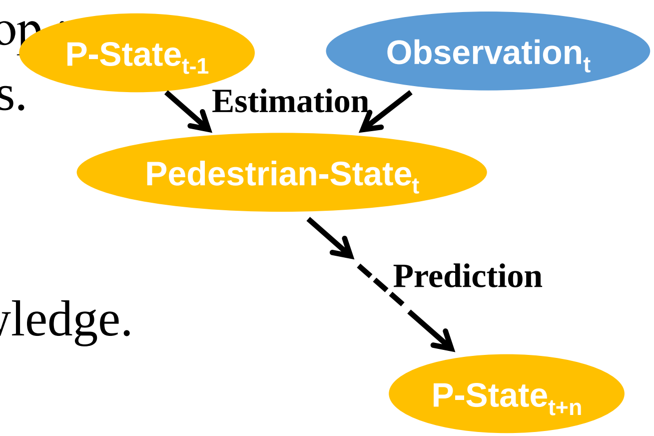
Related Work on Context-based Pedestrian Behavior Recognition

Julian Francisco Pieter Kooij, Dariu M. Gavrila, et al., “Context-Based Pedestrian Path Prediction” ECCV 2014.



Three factors influencing pedestrian’s decision to stop
1. Pedestrian’s awareness of approaching vehicles.
2. Distance to the approaching vehicle.
3. Distance to the curb.

Dynamical Bayesian Network based on empirical knowledge.



Head and Body Orientation Detection

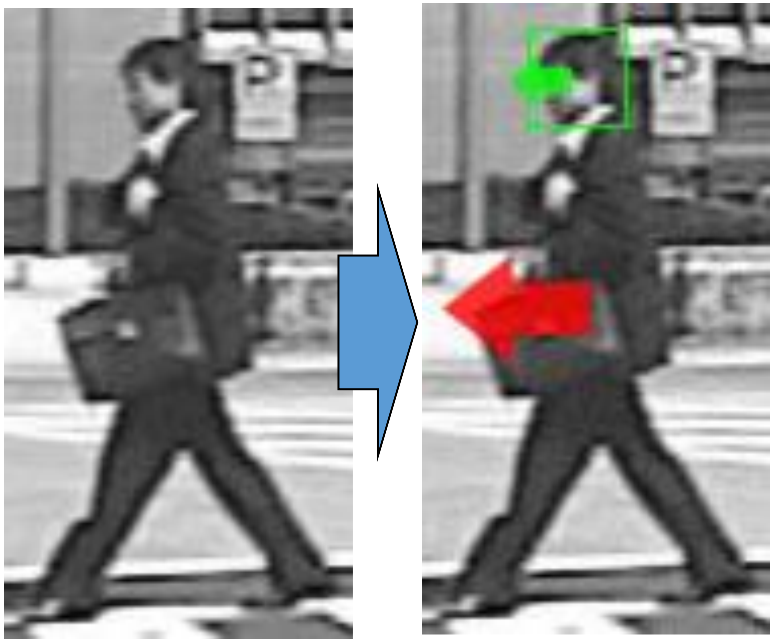
- For more advanced active safety systems
- Recognize the type of road user: cyclist or pedestrian
- From an image of pedestrian and cyclist, estimate the pose/orientation

Head position and orientation

We can estimate the direction that the pedestrian is paying attention to

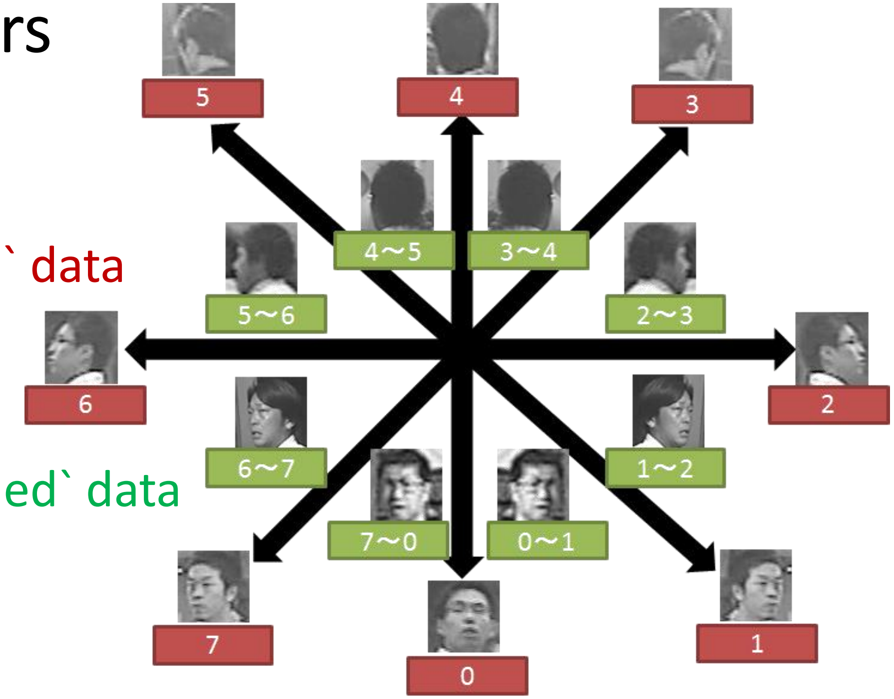
Body orientation

We can estimate the direction that the pedestrian is traveling to



Unlabeled data in orientation

- The problem is the difficulty in labeling the data near the class boundary
- Divide dataset into 2 clusters
  - Strongly labeled data
    - Easy to label
    - Will be processed as `labeled` data
  - Weakly labeled data
    - Difficult to label
    - Will be processed as `unlabeled` data



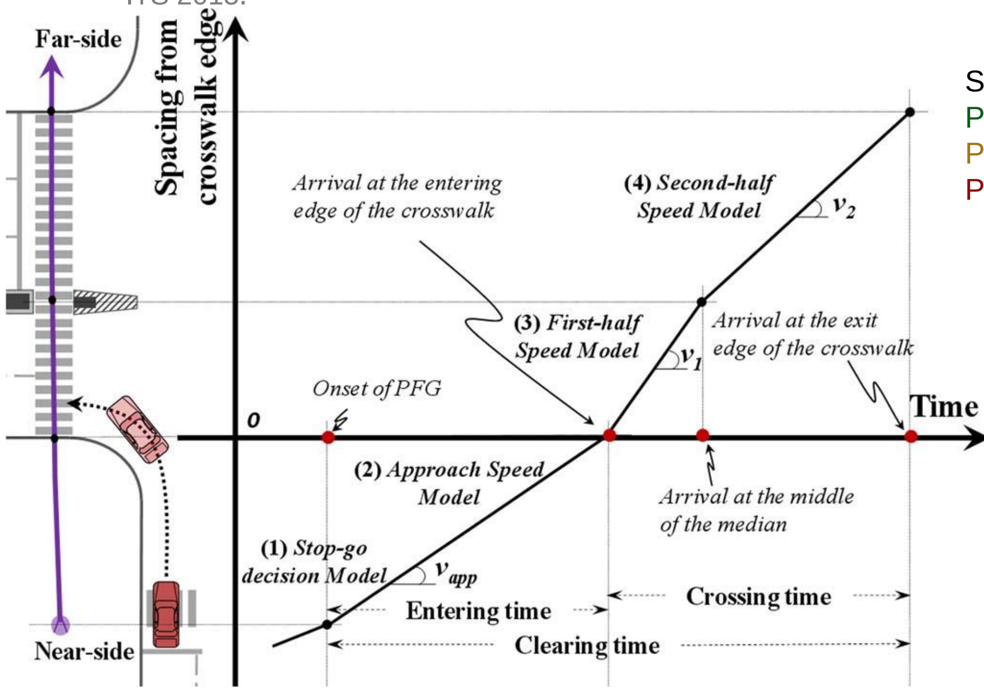
Experimental result

Temporal constraint

Temporal constraint and model constraint

Related Work on Pedestrian Behavior at Signalized Intersections

M. Iryo-Asano, et al. “Analysis and modeling of pedestrian crossing behavior during the pedestrian flashing green interval.” Transactions on ITS 2015.



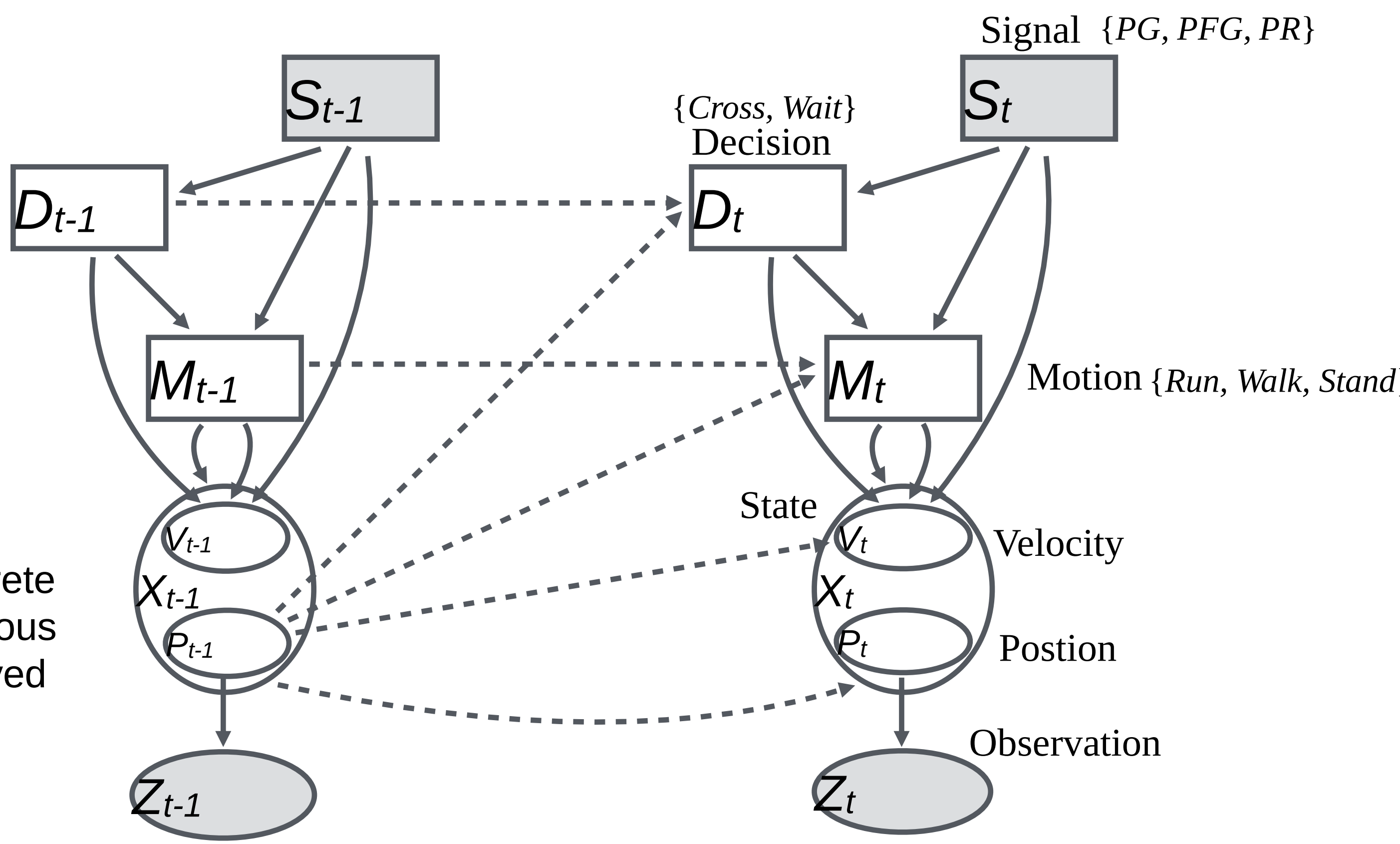
Signal Phase in Japan
PG: Pedestrian Green
PFG: Pedestrian Flashing Green
PR: Pedestrian Red

Analysis of the probabilistic behavior of pedestrians after the onset of the PFG.

1. Pedestrians’ decision of whether to give up crossing (logistic regression)
2. Pedestrian speed distribution by areas (gamma regression)

Explanatory variables: distance to crosswalk, crosswalk length, approaching speed, …

Proposed DBN



rectangle: discrete
ellipse: continuous
shaded: observed

State Transition Model

$$P(D_t, M_t, X_t | D_{t-1}, M_{t-1}, X_{t-1}, S_t)$$
$$= \underbrace{P(D_t | D_{t-1}, S_t, P_{t-1})}_{\text{Cross/Wait Decision Model}} \underbrace{P(M_t | M_{t-1}, S_t, D_t, P_{t-1})}_{\text{Motion Transition Model}} \underbrace{P(X_t | X_{t-1}, S_t, D_t, M_t)}_{\text{Dynamics Model}}$$

Experimental Result
Accuracy of Pedestrian Decision Detection

Scenario	σ_s (m)	Time from onset of PFG			
		1.0s	2.0s	3.0s	4.0s
Cross (46)	0.10	0.85	0.94	1.00	1.00
	0.40	0.72	0.87	0.96	0.98
	1.00	0.62	0.77	0.89	0.96
Wait (38)	0.10	0.58	0.74	0.82	0.84
	0.40	0.49	0.59	0.62	0.69
	1.00	0.45	0.62	0.68	0.78
Cross + Wait (84)	0.10	0.73	0.85	0.92	0.93
	0.40	0.62	0.74	0.80	0.85
	1.00	0.54	0.70	0.79	0.87

- Typically, it takes around 5s for pedestrians to arrive at the crosswalk.
- The proposed model requires only 2s for estimating the pedestrian decision with high reliability in an ideal environment.
- Large measurement errors deteriorate the system performance: slow reaction, low accuracy.
- Wait decision is harder to detect than cross decision.