

Computer Vision in Autonomous Driving

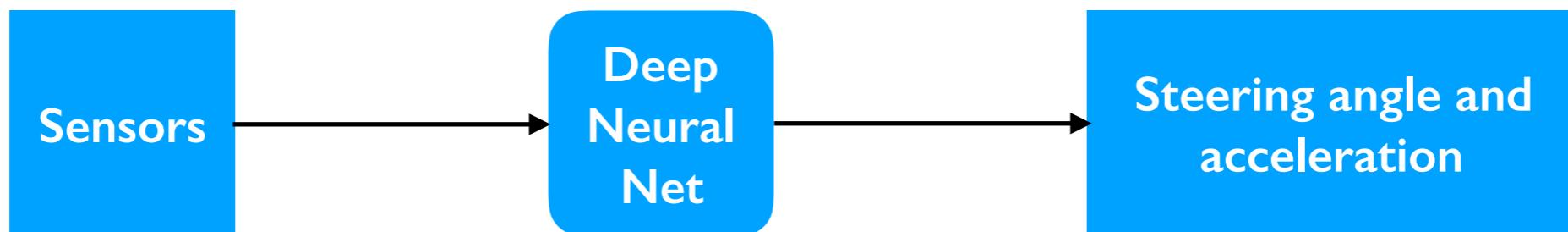
Artem Filatov

Technology



<https://www.youtube.com/watch?v=aaOB-ErYq6Y>

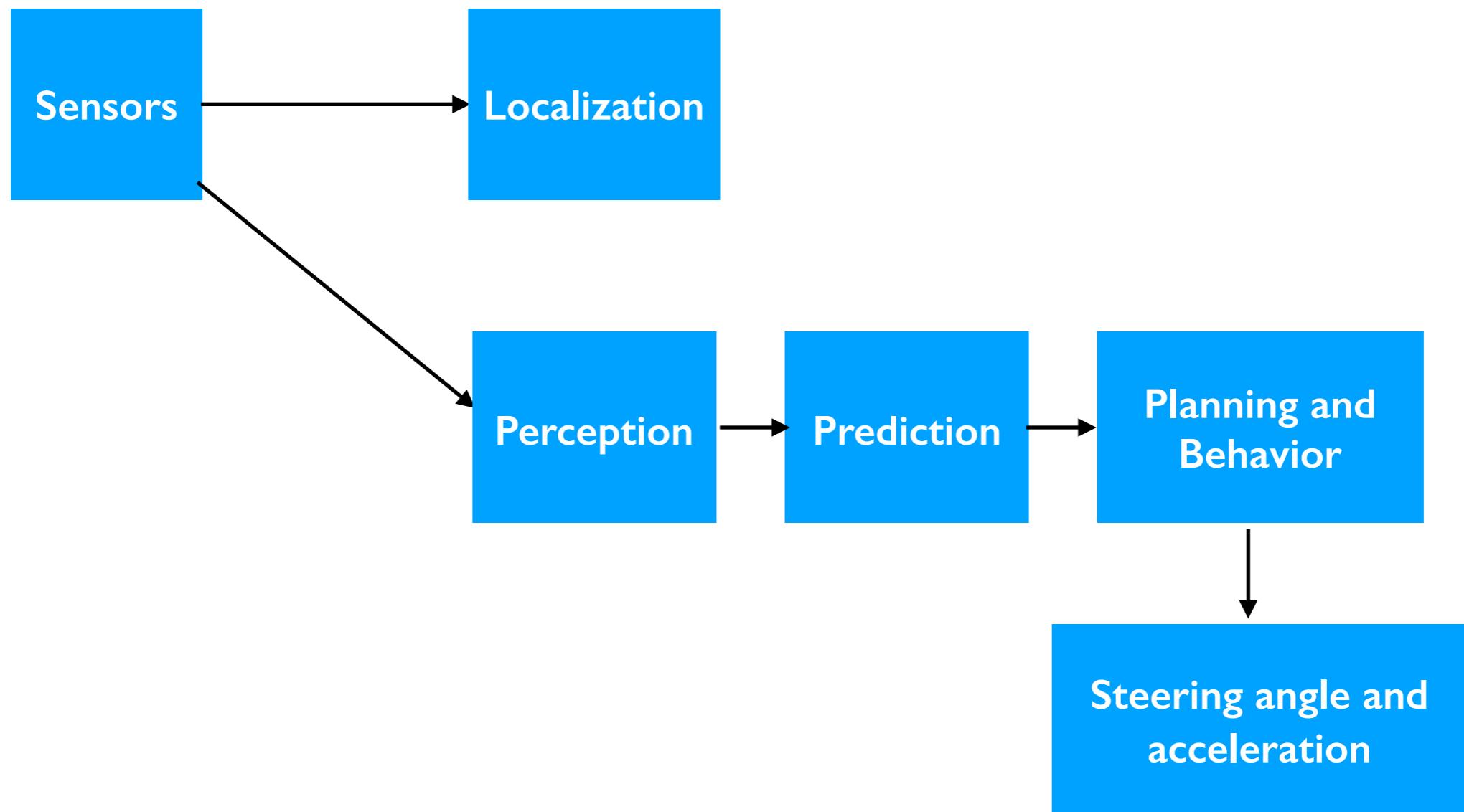
Как спроектировать технологию?



Плохое решение. Пока.

- Такую технологию тяжело править
- Тяжело дать гарантии на поведение
- Задачу нужно декомпозировать

Хорошее решение. Пока.

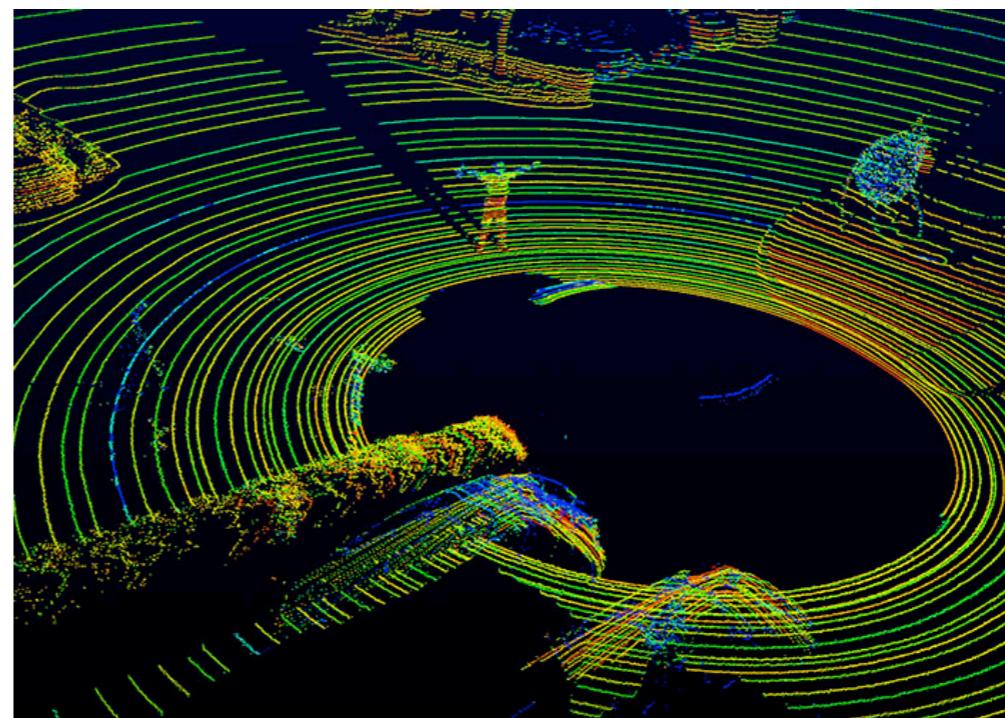
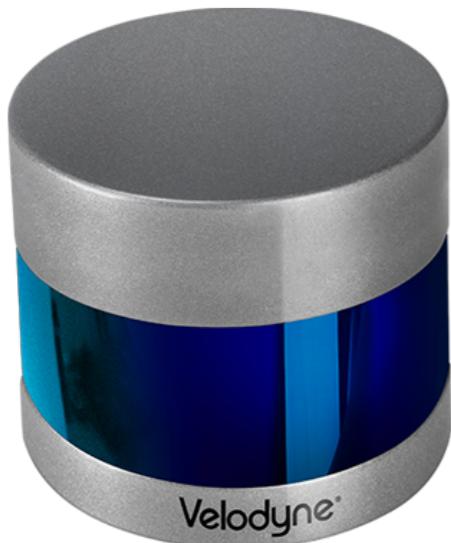


Sensors

- Разные компании используют разные конфигурации
- Стандартный набор: лидары, радары и камеры
- Какие задачи компьютерного зрения здесь возникают?
 - Калибровки!
 - Камера к камере
 - Камера к лидару
 - Лидар к лидару
 - Радар к камере
 - ...

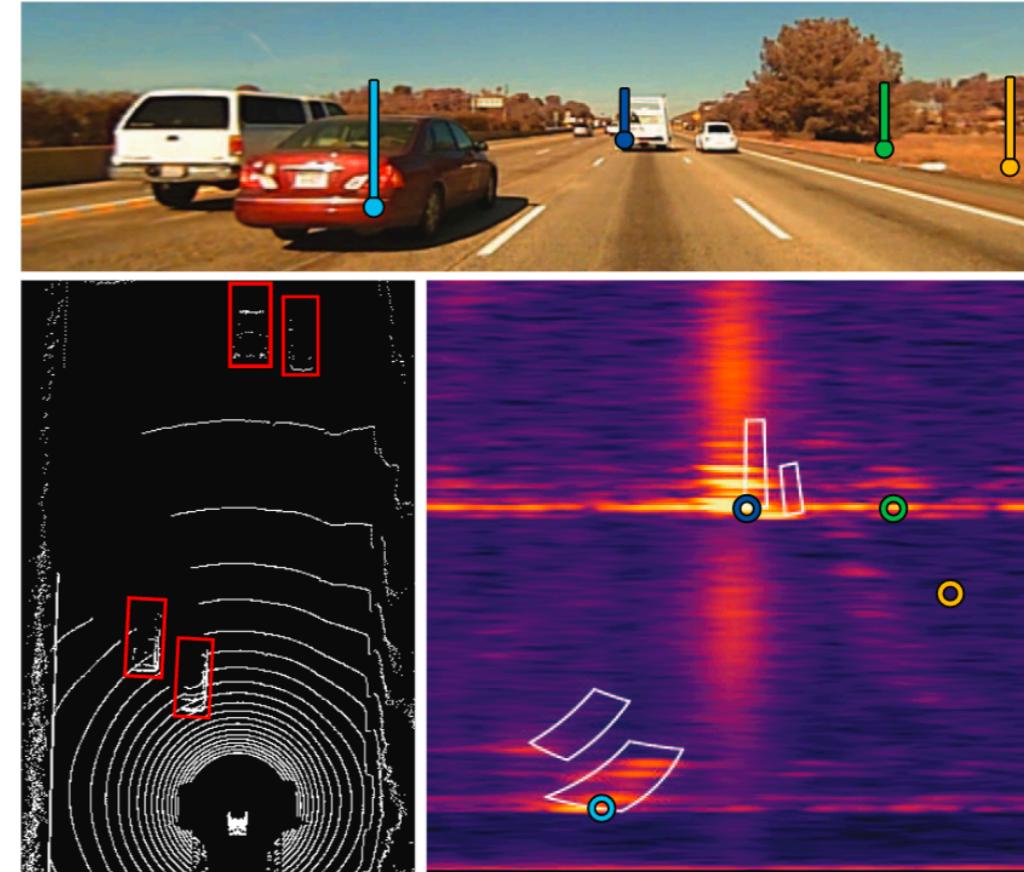
Sensors: LiDAR

- Посылаем луч, ждем возврата сигнала, считаем время, получаем расстояние
- Огромное количество информации: нужно лишь правильно ей воспользоваться
- Стоимость: 10-ки тысяч долларов
- Туман, дождь и снег могут испортить картину - эту проблему нужно решать отдельно



Sensors: Radar

- Сильно дешевле чем лидар
- Помимо координат точки, мы можем узнать скорость точки
- Проблемы:
 - Точек мало
 - Отражения происходят в основном от металлических конструкций

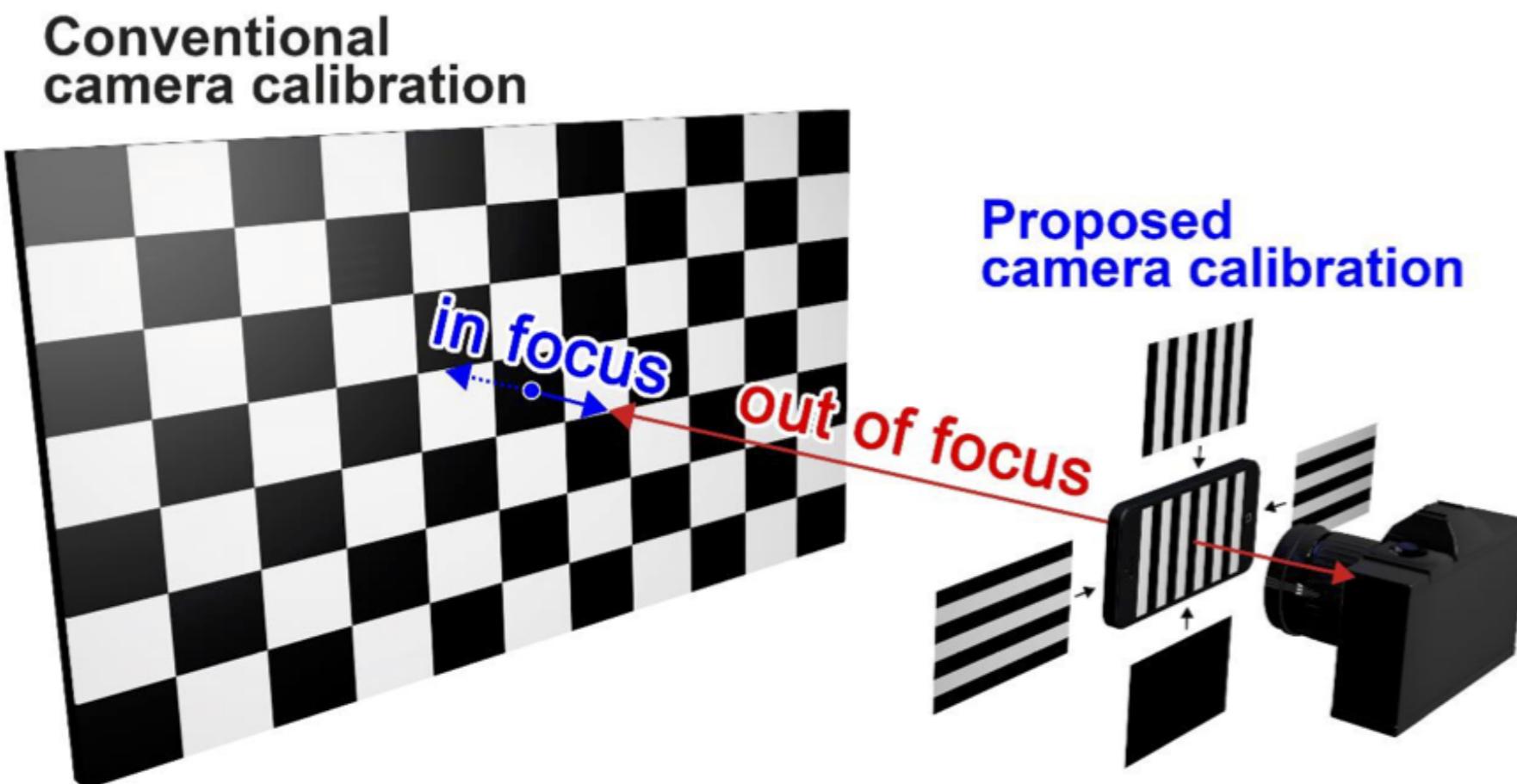


Sensors: Camera

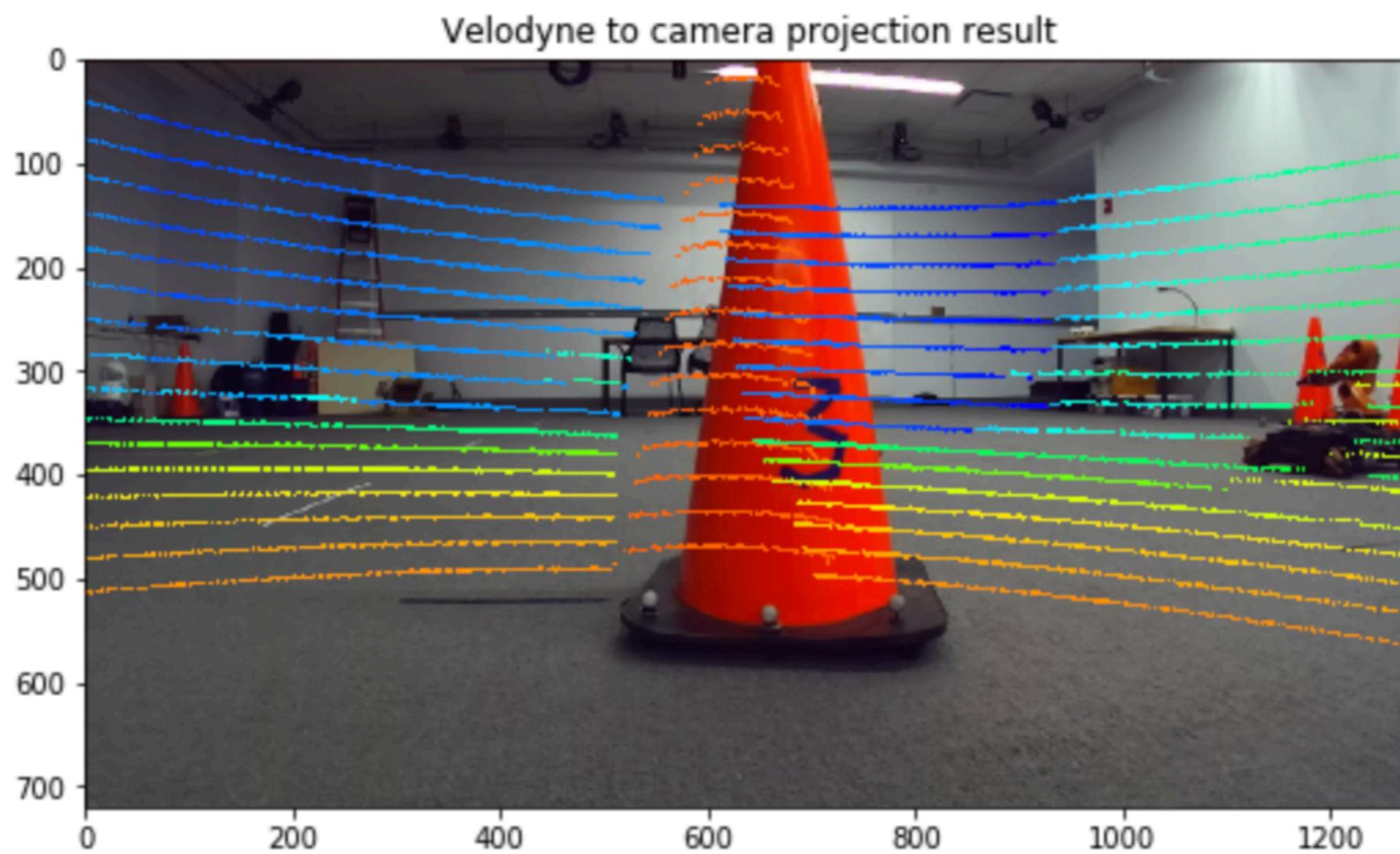
- Дешевая
- Быстрая
- Но нет информации в 3D



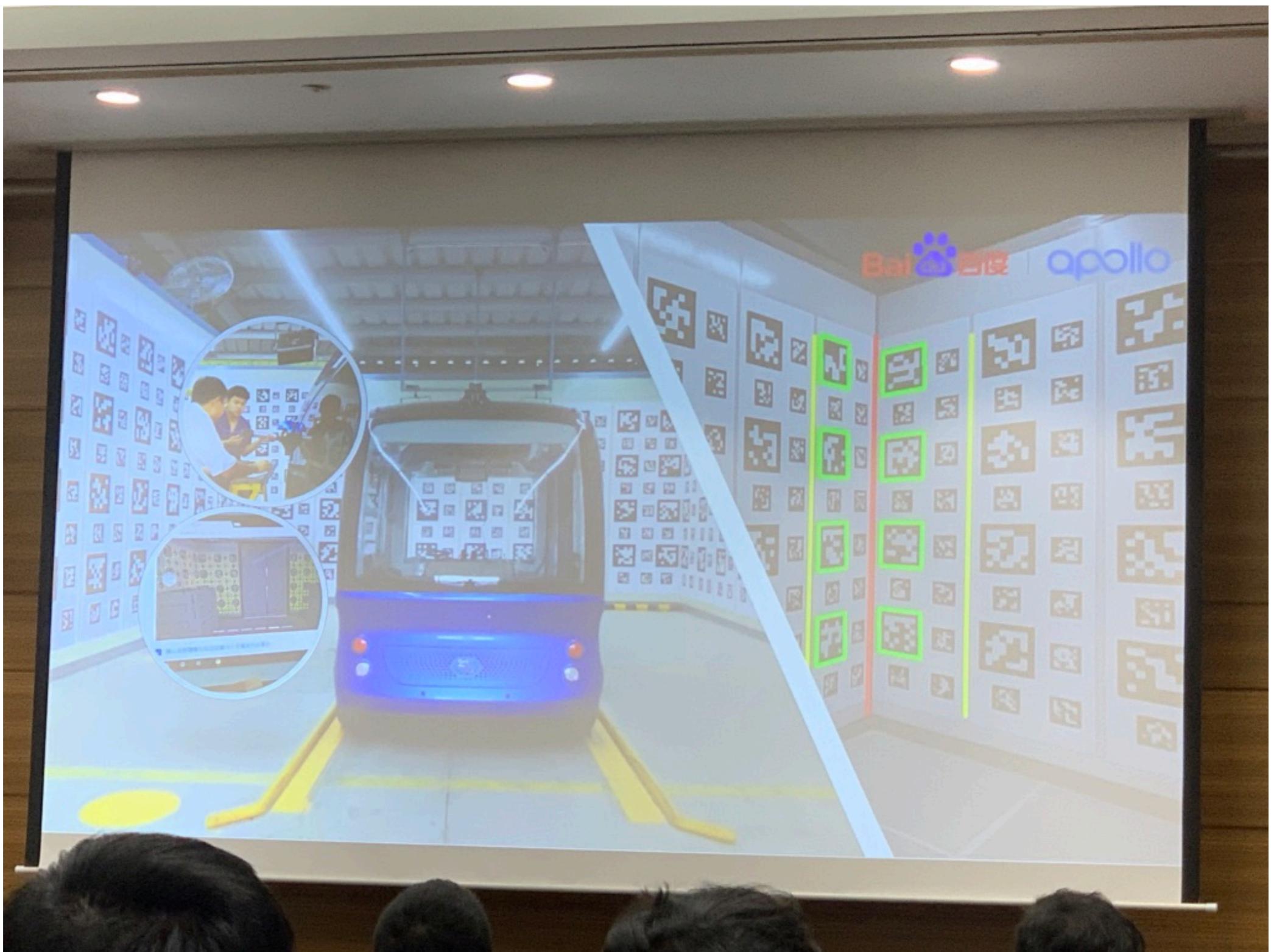
Калибровка



Калибровка: камера к лидару

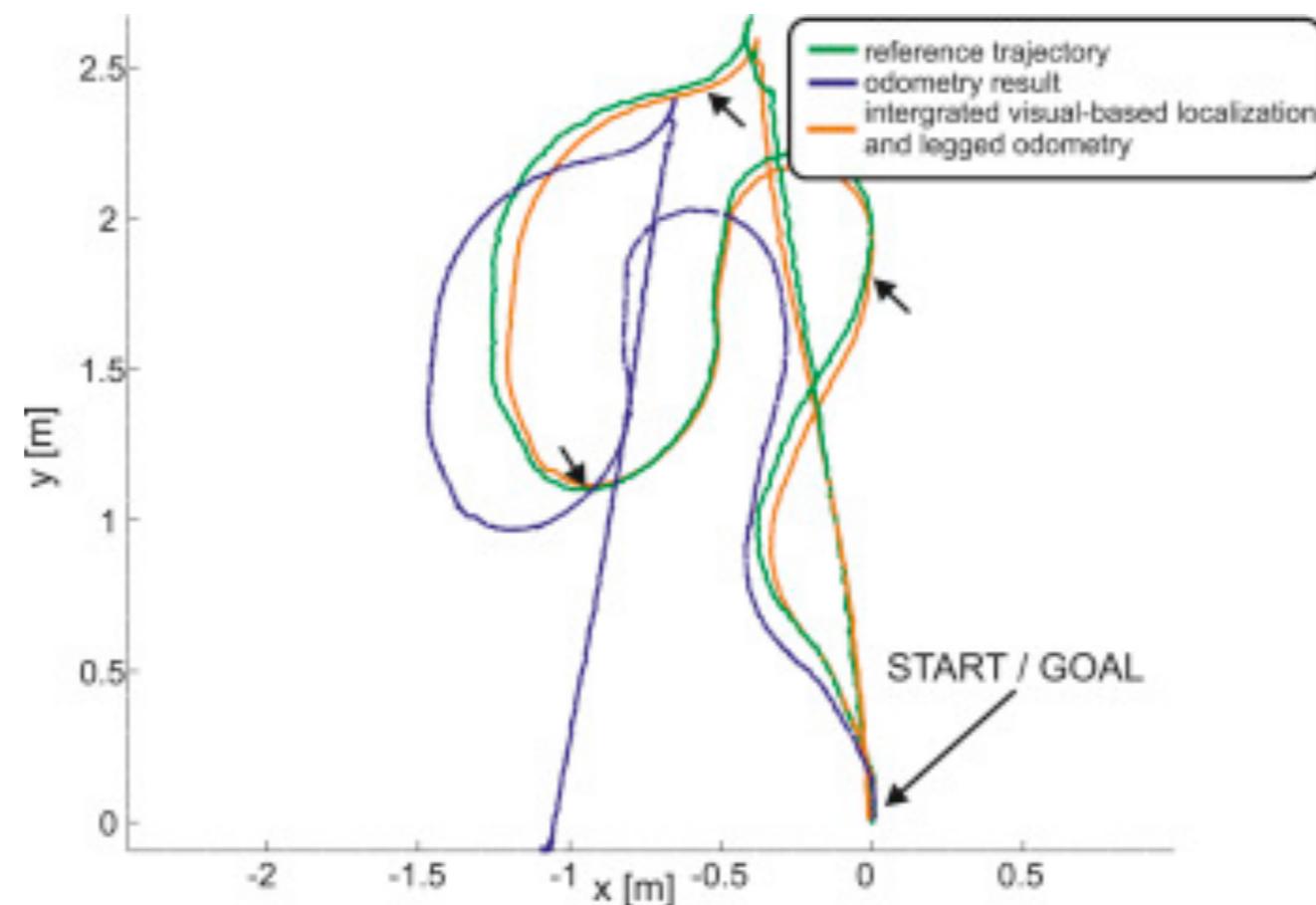


Калибровочная комната



Localization: Odometry

- Давайте смотреть на одометрию с колес и понимать, где мы находимся
- Очень неточное решение!

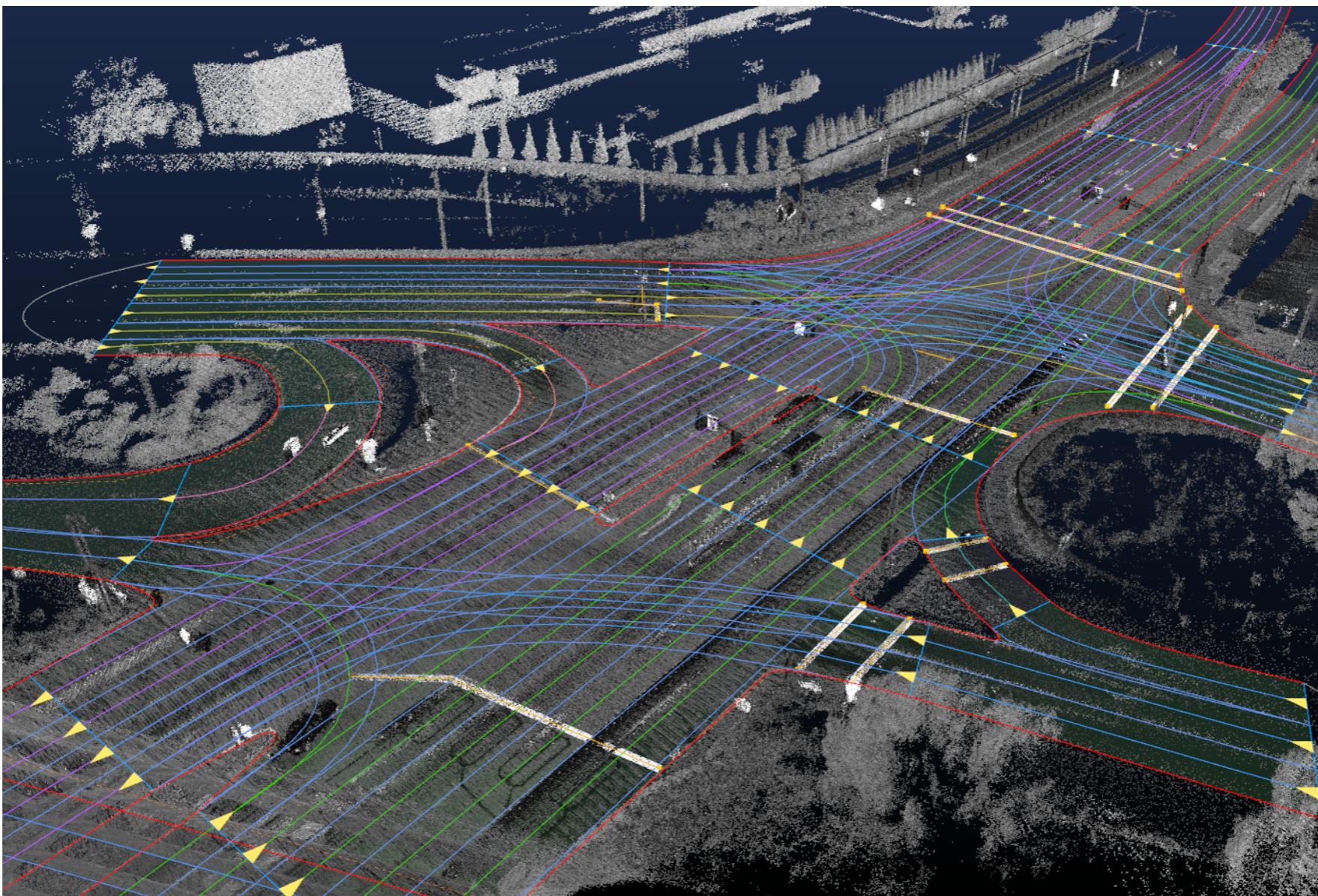


Как сделать лучше?

- Давайте смотреть не только на нас, но еще и на окружающий мир
- Одновременное построение карты и локализация: SLAM
- Один из вариантов: Extended Kalman Filter
- Нужны визуальные фичи для построения карты и одометрия: двух сигналов достаточно для решения задачи с хорошим качеством

<https://www.youtube.com/watch?v=oJt3Ln8H03s>

Как сделать еще лучше? HDmaps



HDmaps

- Построим карту из лидарных облаков оффлайн
- Отметим полосы, светофоры, препятствия
- Для локализации нужно лишь найти соответствие между наблюдаемым лидарным облаком и картой
- С такой задачей можно справится ICP: Iterative Closest Point

ICP

- Находим ближайших соседей
- Находим преобразование минимизирующее ошибку до соседей
- Преобразуем точки
- Повторяем процедуру

https://www.youtube.com/watch?v=uzOCS_gdZuM

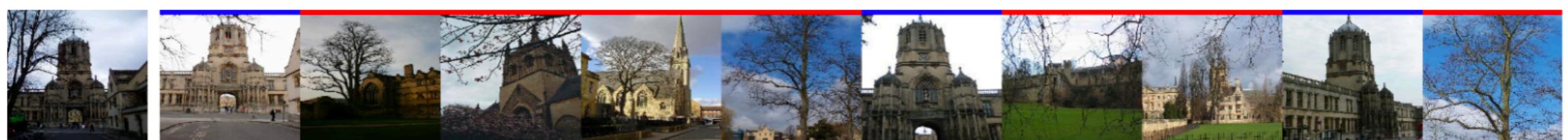
ICP: трудности

- Алгоритму нужно хорошее приближение
- Если мы запустим матчинг наблюдаемого облака с картой Сан-Франциско - ничего не получится
- Как найти хорошее приближение? Можно использовать картинку!

Aggregating Deep Convolutional Features for Image Retrieval

- Давайте построим индекс (базу данных поиска) по картинкам
- Искать будем по фичам авто-кодировщика

Aggregating Deep Convolutional Features for Image Retrieval



Локализовались! Что дальше?

- Начинаем детектировать!
- Что хотим видеть?
 - Динамику: машины, люди, животные...
 - Статику: препятствия, ямы, светофоры, знаки...

Динамика

- 2D детекции: SSD, YOLO, Faster-RCNN...
- 3D детекции: Frustum, VoxelNet...
- В конечном счете нам нужны детекции именно в 3D
- Только так мы сможем корректно проложить маршрут



Какие сенсоры подходят для 3D Детекций?

Input	LSTM	mAP [%]	Velocity L_1 error [m/s]
RA	No	86.34 ± 0.30	2.47 ± 0.06
RA	Yes	88.00 ± 0.37	1.49 ± 0.11
RAD	Yes	87.59 ± 0.56	1.50 ± 0.11

RGB? Очень плохо работает

Радар

Method	Input	Vehicle $AP_{0.7}$		
		Easy	Moderate	Hard
LaserNet (Ours)	LiDAR	78.25	73.77	66.47
PIXOR [28]	LiDAR	81.70	77.05	72.95
PIXOR++ [27]	LiDAR	89.38	83.70	77.97
VoxelNet [30]	LiDAR	89.35	79.26	77.39
MV3D [5]	LiDAR+RGB	86.02	76.90	68.49
AVOD [15]	LiDAR+RGB	88.53	83.79	77.90
F-PointNet [22]	LiDAR+RGB	88.70	84.00	75.33
ContFuse [17]	LiDAR+RGB	88.81	85.83	77.33

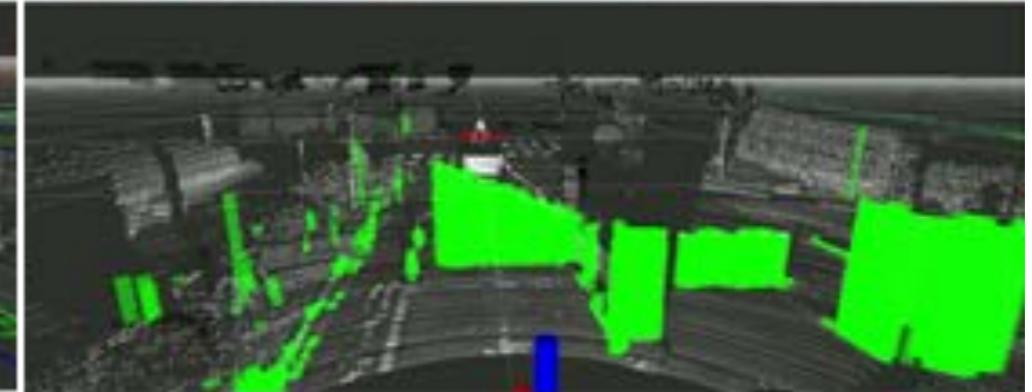
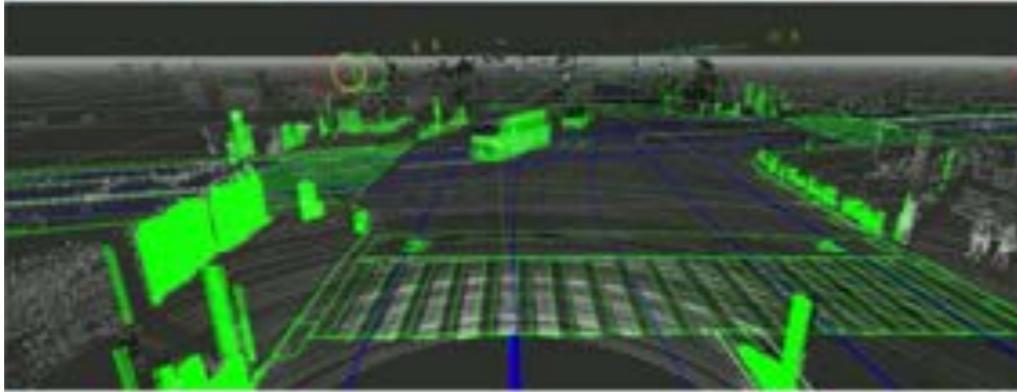
LiDAR and RGB

Статические препятствия

- Можно попробовать решить задачу через геометрию облака:
VScan
- Можно обратиться к классическому ML:
 - Придумаем фичи
 - Запустим xgboost
- Сеточные методы

VScan

- Идем вдоль луча
- Когда видим резкое изменения наклона, говорим о наличии препятствия



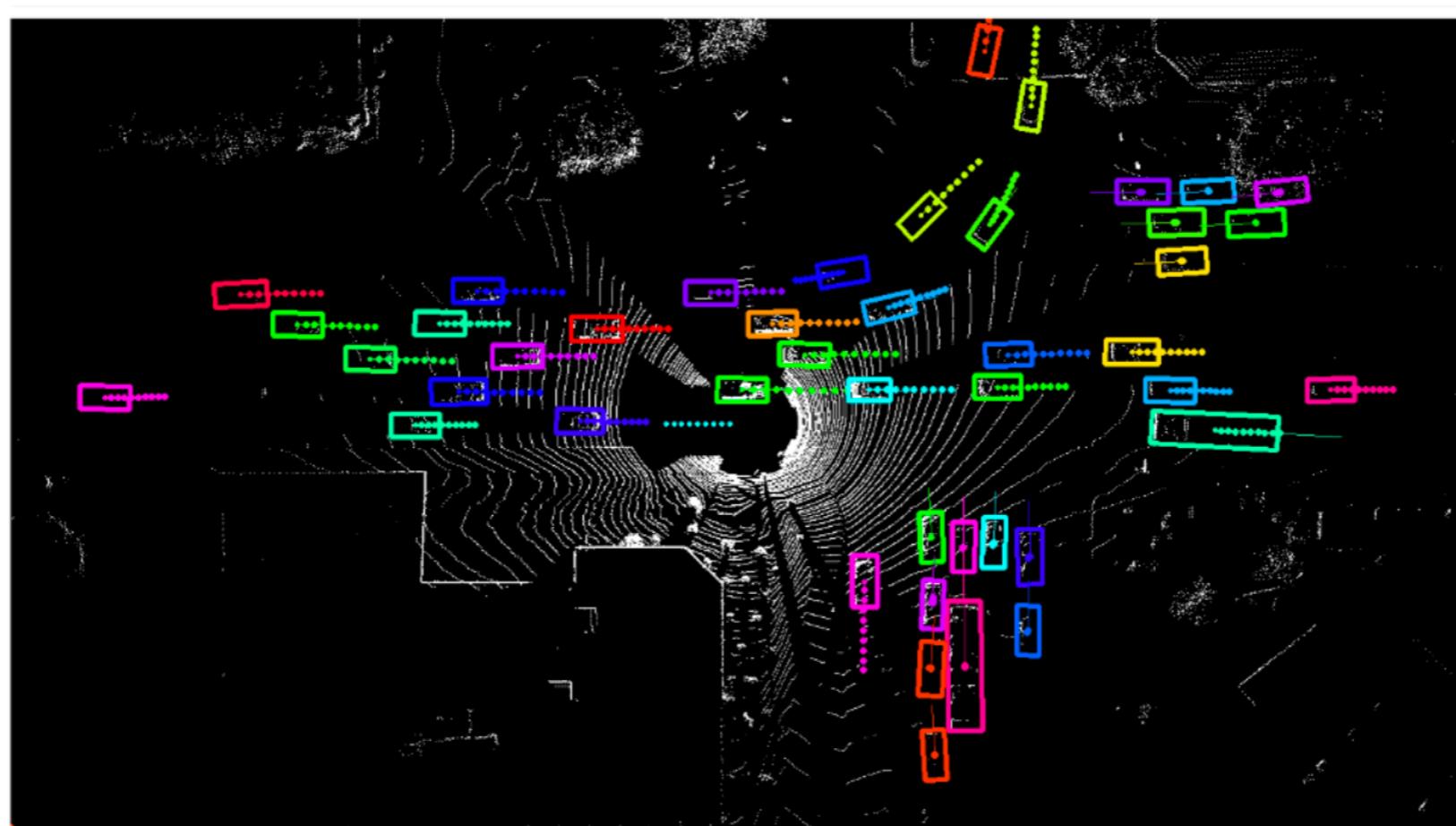
Светофоры

- Имея 2D детекцию, остается решить задачу классификации
- Но не все так просто!
- Светофоры могут моргать
- А могут быть такими:
- На изображении светофор может занимать лишь несколько пикселей



Prediction

- Имея весь набор информации, хочется предсказать развитие сцены
- Можно сделать частью 3D детектора (Fast and Furious)



Пешеходы

- Очень сложно предсказывать
- Люди не идут по полосам и часто меняют намерения
- Можно предсказывать намерения: хочет перейти дорогу, ждет автобус, идет вперед и т.д.

