CSC 2541: Lecture 01: Introduction

Raquel Urtasun

University of Toronto

Jan 12, 2016
Today

- Administration details
- Student introduction
- Why is autonomous driving so cool?
- What would this course cover?
Weakly **Office hours** (2h):

- Get weekly feedback about your research project
- Help preparing class presentations


Piazza: for most communications
[piazza.com/utoronto.ca/winter2016/csc2541](piazza.com/utoronto.ca/winter2016/csc2541)

Submissions: MarkUs
Admin Details: Course Stuff

- Weakly **Office hours** (2h):
  - Get weekly feedback about your research project
  - Help preparing class presentations

- Webpage of the course:
Weakly **Office hours** (2h):
- Get weekly feedback about your research project
- Help preparing class presentations

**Webpage of the course:**

**Piazza:** for most communications
piazza.com/utoronto.ca/winter2016/csc2541
Admin Details: Course Stuff

- Weakly **Office hours** (2h):
  - Get weekly feedback about your research project
  - Help preparing class presentations

- Webpage of the course:

- Piazza: for most communications
  [piazza.com/utoronto.ca/winter2016/csc2541](piazza.com/utoronto.ca/winter2016/csc2541)

- **Submissions**: MarkUs
Do I have the proper pre-requisites?

- There is no prerequisite per say

- Linear algebra, calculus and probability

- Statistics

- Programming: strong skills

- Machine learning: at least undergrad level course

- Computer Vision: at least undergrad level course
Do I have the proper pre-requisites?

▶ There is no prerequisite per say

Do I have the appropriate background?

▶ Linear algebra, calculus and probability
Admin Details: Prerequisites

- Do I have the proper pre-requisites?
  - There is no prerequisite per say

- Do I have the appropriate background?
  - Linear algebra, calculus and probability
  - Statistics
Do I have the proper pre-requisites?
  ▶ There is no prerequisite per say

Do I have the appropriate background?
  ▶ Linear algebra, calculus and probability
  ▶ Statistics
  ▶ Programming: strong skills
Do I have the proper pre-requisites?
  ▶ There is no prerequisite per say

Do I have the appropriate background?
  ▶ Linear algebra, calculus and probability
  ▶ Statistics
  ▶ Programming: strong skills
  ▶ Machine learning: at least undergrad level course
Do I have the proper pre-requisites?
  ▶ There is no prerequisite per say

Do I have the appropriate background?
  ▶ Linear algebra, calculus and probability
  ▶ Statistics
  ▶ Programming: strong skills
  ▶ Machine learning: at least undergrad level course
  ▶ Computer Vision: at least undergrad level course
Textbooks

- No textbook
- We will be reading papers
- You might need to consult books
Requirements

- **Reviews:**
  - two papers every week
  - Worth 20% of the total mark
Requirements

- **Reviews:**
  - two papers every week
  - Worth 20% of the total mark

- **Lecture Presentations:**
  - Everyone will be responsible to present 1 time in class
  - Worth 20% of the total mark
Requirements

- **Reviews:**
  - two papers every week
  - Worth 20% of the total mark

- **Lecture Presentations:**
  - Everyone will be responsible to present 1 time in class
  - Worth 20% of the total mark

- **Project:**
  - Proposal that has to be approved
  - Worth 60% of course mark
Projects are typically individual, but you can be in pairs.
Projects are typically *individual*, but you can be in pairs.

▶ My expectations will be much higher if you are in pairs
More on Projects

- Projects are typically **individual**, but you can be in pairs.
  - My expectations will be much higher if you are in pairs
- Can I do a project for 2 courses at the same time?

A project proposal: due Feb 1

A project report (paper draft) at the end of the project

A presentation of your work at the end
More on Projects

- Projects are typically **individual**, but you can be in pairs.
  - My expectations will be much higher if you are in pairs
- Can I do a project for 2 courses at the same time?
  - Only if you show what is for each class and the project is much bigger
More on Projects

- Projects are typically individual, but you can be in pairs.
  - My expectations will be much higher if you are in pairs

- Can I do a project for 2 courses at the same time?
  - Only if you show what is for each class and the project is much bigger
  - Require approval by the other instructor as well
More on Projects

- Projects are typically individual, but you can be in pairs.
  - My expectations will be much higher if you are in pairs
- Can I do a project for 2 courses at the same time?
  - Only if you show what is for each class and the project is much bigger
  - Require approval by the other instructor as well
- What are the milestones?
  - A project proposal: due Feb 1
  - A project report (paper draft) at the end of the project
  - A presentation of your work at the end
More on Projects

- Projects are typically *individual*, but you can be in pairs.
  - My expectations will be much higher if you are in pairs

- Can I do a project for 2 courses at the same time?
  - Only if you show what is for each class and the project is much bigger
  - Require approval by the other instructor as well

- What are the milestones?
  - A project *proposal*: due Feb 1
More on Projects

- Projects are typically individual, but you can be in pairs.
  - My expectations will be much higher if you are in pairs

- Can I do a project for 2 courses at the same time?
  - Only if you show what is for each class and the project is much bigger
  - Require approval by the other instructor as well

- What are the milestones?
  - A project proposal: due Feb 1
  - A project report (paper draft) at the end of the project
More on Projects

- Projects are typically **individual**, but you can be in pairs.
  - My expectations will be much higher if you are in pairs

- Can I do a project for 2 courses at the same time?
  - Only if you show what is for each class and the project is much bigger
  - Require approval by the other instructor as well

- What are the milestones?
  - A project **proposal**: due Feb 1
  - A project **report** (paper draft) at the end of the project
  - A **presentation** of your work at the end
Let’s get to know you!
Get to Know you

Link

1. Name and email
2. Background: department where you are at, which year, masters/phd/applied masters, etc
3. Research topic/ interest for grad studies
4. Supervisor
5. Experience in machine learning, computer vision and/or robots
6. Particular topics you will like to have covered in class

If you haven’t submitted your pdf slides, do so asap
1. You will discover that self-driving cars are a really cool research topic
Goals of this course

1. You will discover that self-driving cars are a really cool research topic.
2. Everyone should do an awesome project that will be accepted to a top-tier conference.
Goals of this course

1. You will discover that self-driving cars are a really cool research topic
2. Everyone should do an awesome project that will be accepted to a top-tier conference
3. After taking this class you’ll get a job at Google, Apple, Toyota, Daimler, Tesla, BMV, Bosch, etc
Logistics

- Need to re-schedule class of Jan 26th
  - Let’s find a replacement time
Logistics

- Need to re-schedule class of Jan 26th
  - Let’s find a replacement time
- I’ll post on piazza a link to vote for topics to present
  - Don’t forget to vote or you will be randomly assigned
Why Autonomous Driving?

[Image of a question mark with a person sitting inside, thinking]
Some "Scary" Statistics: Traffic Fatalities

Figure: Road Fatalities per 100,000 inhabitants and year

In total (2010): USA (36,166), Canada (2,075), World (1.24 million!)
Some “Scary” Statistics

- More than 2,000 deaths/year in Canada
Some ”Scary” Statistics

- More than 2,000 deaths/year in Canada
- Despite many technological developments this number is constant
Some "Scary" Statistics

- More than 2,000 deaths/year in Canada
- Despite many technological developments this number is constant
- **Problem**: Growth of the population in suburbs (8.7% vs 5.3% in cities)
Some “Scary” Statistics

- More than 2,000 deaths/year in Canada
- Despite many technological developments this number is constant
- **Problem**: Growth of the population in suburbs (8.7% vs 5.3% in cities)
- According to NHS, 74% (11.4 millions) drove a vehicle to work
Some "Scary" Statistics

- More than 2,000 deaths/year in Canada
- Despite many technological developments this number is constant
- **Problem:** Growth of the population in suburbs (8.7% vs 5.3% in cities)
- According to NHS, 74% (11.4 millions) drove a vehicle to work

- Traffic pollution is estimated to be more than twice as deadly as traffic accidents (33% of greenhouse emissions in the US)
Some “Scary” Statistics

- More than 2,000 deaths/year in Canada
- Despite many technological developments this number is constant
- **Problem:** Growth of the population in suburbs (8.7% vs 5.3% in cities)
- According to NHS, 74% (11.4 millions) drove a vehicle to work

- Traffic pollution is estimated to be more than twice as deadly as traffic accidents (33% of greenhouse emissions in the US)
- Driver stress: how to quantify it?
1. Lower the risk of accidents
Benefits of Autonomous Driving

1. Lower the risk of accidents
2. Provide mobility for the elderly and people with disabilities
   - In the US 45% of people with disabilities still work
Benefits of Autonomous Driving

1. Lower the risk of accidents
2. Provide mobility for the elderly and people with disabilities
   - In the US 45% of people with disabilities still work
3. Decrease pollution for a more healthy environment
Benefits of Autonomous Driving

1. Lower the risk of accidents
2. Provide mobility for the elderly and people with disabilities
   ▶ In the US 45% of people with disabilities still work
3. Decrease pollution for a more healthy environment
4. New ways of Public Transportation
Benefits of Autonomous Driving

1. Lower the risk of accidents
2. Provide mobility for the elderly and people with disabilities
   - In the US 45% of people with disabilities still work
3. Decrease pollution for a more healthy environment
4. New ways of Public Transportation
   - What do you think it would be?
Benefits of Autonomous Driving

1. Lower the risk of accidents
2. Provide mobility for the elderly and people with disabilities
   - In the US 45% of people with disabilities still work
3. Decrease pollution for a more healthy environment
4. New ways of Public Transportation
   - What do you think it would be?
   - Is there a demand?
Benefits of Autonomous Driving

1. Lower the risk of accidents
2. Provide mobility for the elderly and people with disabilities
   ▶ In the US 45% of people with disabilities still work
3. Decrease pollution for a more healthy environment
4. New ways of Public Transportation
   ▶ What do you think it would be?
   ▶ Is there a demand?
   ▶ What would the benefits be?
Benefits of Autonomous Driving

1. Lower the risk of accidents
2. Provide mobility for the elderly and people with disabilities
   - In the US 45% of people with disabilities still work
3. Decrease pollution for a more healthy environment
4. New ways of Public Transportation
   - What do you think it would be?
   - Is there a demand?
   - What would the benefits be?
   - And the drawbacks?
Benefits of Autonomous Driving

1. Lower the risk of accidents

2. Provide mobility for the elderly and people with disabilities
   - In the US 45% of people with disabilities still work

3. Decrease pollution for a more healthy environment

4. New ways of Public Transportation
   - What do you think it would be?
   - Is there a demand?
   - What would the benefits be?
   - And the drawbacks?

5. Anything else?
Boring life of a car

- 95% of the time a car is parked

Figure from http://theoatmeal.com/blog/google_self_driving_car
A bit of history
History of Autonomous Driving

Figure: Norman Bel Geddes’ Futurama, New York World Fair 1939

Envisioned a world 20 years into the future featuring

- Automated highways as a solution to traffic congestion
- Electric cars were powered by circuits embedded in the roadway and controlled by radio
In 1986, Ernst Dickmanns in collaboration with Daimler, equipped a Mercedes-Benz with cameras and demonstrate the first self-driving car on well-marked streets without traffic.
In 1986, Ernst Dickmanns in collaboration with Daimler, equipped a Mercedes-Benz with cameras and demonstrate the first self-driving car on well-marked streets without traffic.

European project EUREKA Prometheus, in 1995 they drove in real traffic from Munich in Germany to Odense in Denmark with speeds of up to 175km/h, with human intervention 5% of the distance.
In 1986, Ernst Dickmanns in collaboration with Daimler, equipped a Mercedes-Benz with cameras and demonstrate the first self-driving car on well-marked streets without traffic.

European project EUREKA Prometheus, in 1995 they drove in real traffic from Munich in Germany to Odense in Denmark with speeds of up to 175km/h, with human intervention 5% of the distance.

1995, CMU Navlab project achieved 98.2% autonomy with manual longitudinal control.
2004 Darpa Grand Challenge:

- 240km in Nevada, $1 million price
History of Autonomous Driving

2004 Darpa Grand Challenge:

- 240km in Nevada, $1 million price
- None of the vehicles finished the route. CMU car won, and travel 11.78km
History of Autonomous Driving

2004 Darpa Grand Challenge:
- 240km in Nevada, $1 million price
- None of the vehicles finished the route. CMU car won, and travel 11.78km

2005 Darpa Grand Challenge:
- 5 teams finished the course.
- 7h to finish the course. Stanford won with Stanley.
2007 Darpa Urban Challenge:

- 96 km test course at an abandoned Air Force Base
2007 Darpa Urban Challenge:

- 96 km test course at an abandoned Air Force Base
- 100% autonomous driving was required throughout the course
2007 Darpa Urban Challenge:

- 96 km test course at an abandoned Air Force Base
- 100% autonomous driving was required throughout the course
- The streets were wider than usual, the field of view was unobstructed and only a very limited number of traffic participants were present.

Two important pieces of technology came out of this challenge:

▶ Sub-meter precise manually annotated maps were required
▶ Use 3D laser scanner for localization and collision avoidance
2007 Darpa Urban Challenge:

- 96 km test course at an abandoned Air Force Base
- 100% autonomous driving was required throughout the course
- The streets were wider than usual, the field of view was unobstructed and only a very limited number of traffic participants were present.
- Two important pieces of technology came out of this challenge:
2007 Darpa Urban Challenge:

- 96 km test course at an abandoned Air Force Base
- 100% autonomous driving was required throughout the course
- The streets were wider than usual, the field of view was unobstructed and only a very limited number of traffic participants were present.
- Two important pieces of technology came out of this challenge:
  - Sub-meter precise manually annotated maps were required
History of Autonomous Driving

2007 Darpa Urban Challenge:

- 96 km test course at an abandoned Air Force Base
- 100% autonomous driving was required throughout the course
- The streets were wider than usual, the field of view was unobstructed and only a very limited number of traffic participants were present.
- Two important pieces of technology came out of this challenge:
  - Sub-meter precise manually annotated maps were required
  - All use 3D laser scanner for localization and collision avoidance
Google Driverless Car:

- Sebastian Thrun, Google gathered a team of engineers that had experience in the DARPA Grand and Urban Challenge.
- In 2012, they announce that they have completed over 300,000 miles without accident, but what does it mean?
- Use Velodyne 3D laser scanner worth 100,000$
- Detailed annotated maps
History of Autonomous Driving

**Tesla Autopilot:**

- Introduce in 2015 to help the driver in parking and highways
History of Autonomous Driving

**Tesla Autopilot:**

- Introduce in 2015 to help the driver in parking and highways
- Cheaper sensors: GPS, forward radar, forward camera, ultrasonic sensors positioned to sense 16 feet around the car in every direction
Tesla Autopilot:

- Introduce in 2015 to help the driver in parking and highways
- Cheaper sensors: GPS, forward radar, forward camera, ultrasonic sensors positioned to sense 16 feet around the car in every direction
- The driver is still responsible for, and ultimately in control of, the car.
Look, No Hands

1939
General Motors presents the concept of a driverless car at the 1939 World's Fair

1984 & 1987
Carnegie Mellon University and Bundeswehr University Munich develop autonomous vans

1998
Mercedes, Toyota and Mitsubishi begin offering adaptive cruise control

2004
U.S. Defense Department issues a $1 million challenge to develop self-driving vehicles

2012
Google begins testing self-driving models on public roads

2015
Tesla promises to introduce a model with "auto-steering"

2016-17
Mercedes, Audi, BMW and Cadillac will offer models that drive hands-free

2017-20
Google promises to introduce the first fully autonomous car

Sources: Getty Images, Carnegie Mellon University, Wikimedia Commons
1. Autonomous or nothing: e.g., Google, Apple?
1. Autonomous or nothing: e.g., Google, Apple?
   - Very risky business model: only a few companies can do this
1. Autonomous or nothing: e.g., Google, Apple?
   ▶ Very risky business model: only a few companies can do this
   ▶ Long term goals
1. Autonomous or nothing: e.g., Google, Apple?
   - Very risky business model: only a few companies can do this
   - Long term goals

2. Introduce technology little by little but still demonstrate they can do autonomous driving, e.g., all car companies
1. Autonomous or nothing: e.g., Google, Apple?
   - Very risky business model: only a few companies can do this
   - Long term goals

2. Introduce technology little by little but still demonstrate they can do autonomous driving, e.g., all car companies
   - Car industry is very conservative
1. Autonomous or nothing: e.g., Google, Apple?
   - Very risky business model: only a few companies can do this
   - Long term goals

2. Introduce technology little by little but still demonstrate they can do autonomous driving, e.g., all car companies
   - Car industry is very conservative
   - ADAS as intermediate goal
1. Autonomous or nothing: e.g., Google, Apple?
   ▶ Very risky business model: only a few companies can do this
   ▶ Long term goals

2. Introduce technology little by little but still demonstrate they can do autonomous driving, e.g., all car companies
   ▶ Car industry is very conservative
   ▶ ADAS as intermediate goal
   ▶ Sharp transition: how to maintain the driver engaged?
What are the main challenges of self-driving cars?

Money:
- Expensive to do research in this topic
- Reduce cost for mass market production
What are the main challenges of self-driving cars?

- **Money:**
  - Expensive to do research in this topic
  - Reduce cost for mass market production
What are the main challenges?

- Technology
What are the main challenges?

- Dealing with **humans**

Figure from http://theoatmeal.com/blog/google_self_driving_car
What are the main challenges?

- **Law**: who’s fault is it?
What are the main challenges?

- Ethics: how should we program our car to be ethically correct?
Are we ready for Autonomous driving?

- **Humans**: would be embrace technology?
Autonomous Driving

State of the art

Localization, path planning, obstacle avoidance

Heavy usage of Velodyne and detailed (recorded) maps

Problems for computer vision

Stereo, optical flow, visual odometry, structure-from-motion

Object detection, recognition and tracking

3D scene understanding
State of the art

- Localization, path planning, obstacle avoidance
State of the art

- Localization, path planning, obstacle avoidance
- Heavy usage of Velodyne and detailed (recorded) maps
State of the art
- Localization, path planning, obstacle avoidance
- Heavy usage of Velodyne and detailed (recorded) maps

Problems for computer vision
- Stereo, optical flow, visual odometry, structure-from-motion
State of the art

- Localization, path planning, obstacle avoidance
- Heavy usage of Velodyne and detailed (recorded) maps

Problems for computer vision

- Stereo, optical flow, visual odometry, structure-from-motion
- Object detection, recognition and tracking
State of the art

- Localization, path planning, obstacle avoidance
- Heavy usage of Velodyne and detailed (recorded) maps

Problems for computer vision

- Stereo, optical flow, visual odometry, structure-from-motion
- Object detection, recognition and tracking
- 3D scene understanding
What are we going to cover in this course?

- Sensors and platforms
- Datasets
- Models, Algorithms and Techniques:
  - Reconstruction and free-space estimation
  - Motion Estimation
  - Localization
  - Semantics
  - Mapping
- Law and Ethics
What are we going to cover in this course?

- Sensors and platforms
- Datasets
- Models, Algorithms and Techniques:
  - Reconstruction and free-space estimation
  - Motion Estimation
  - Localization
  - Semantics
  - Mapping
- Law and Ethics
Example of Platform: KIT

- Two stereo rigs (1392 × 512 px, 54 cm base, 90° opening)
- Velodyne HDL-64E laser scanner
- GPS+IMU localization
Sensor Setup

- 2 × PointGray Flea2 grayscale cameras (FL2-14S3M-C), 1.4 Megapixels, 1/2” Sony ICX267 CCD, global shutter
- 4 × Edmund Optics lenses, 4mm, opening angle ∼90°, vertical opening angle of region of interest (ROI) ∼35°
- 1 × Velodyne HDL-64E rotating 3D laser scanner, 10 Hz, 64 beams, 0.09 degree angular resolution, 2 cm distance accuracy, collecting ∼1.3 million points/second, field of view: 360° horizontal, 26.8° vertical, range: 120 m
- 1 × OXTS RT3003 inertial and GPS navigation system, 6 axis, 100 Hz, L1/L2 RTK, resolution: 0.02m / 0.1°
Sensor Setup

- 2 × PointGray Flea2 grayscale cameras (FL2-14S3M-C), 1.4 Megapixels, 1/2” Sony ICX267 CCD, global shutter
- 2 × PointGray Flea2 color cameras (FL2-14S3C-C), 1.4 Megapixels, 1/2” Sony ICX267 CCD, global shutter

1 × Velodyne HDL-64E rotating 3D laser scanner, 10 Hz, 64 beams, 0.09 degree angular resolution, 2 cm distance accuracy, collecting \(\approx 1.3\) million points/second, field of view: 360° horizontal, 26.8° vertical, range: 120 m

1 × OXTS RT3003 inertial and GPS navigation system, 6 axis, 100 Hz, L1/L2 RTK, resolution: 0.02m / 0.1°
Sensor Setup

- 2 × PointGray Flea2 grayscale cameras (FL2-14S3M-C), 1.4 Megapixels, 1/2” Sony ICX267 CCD, global shutter
- 2 × PointGray Flea2 color cameras (FL2-14S3C-C), 1.4 Megapixels, 1/2” Sony ICX267 CCD, global shutter
- 4 × Edmund Optics lenses, 4mm, opening angle \( \sim 90^\circ \), vertical opening angle of region of interest (ROI) \( \sim 35^\circ \)
Sensor Setup

- 2 × PointGray Flea2 grayscale cameras (FL2-14S3M-C), 1.4 Megapixels, 1/2” Sony ICX267 CCD, global shutter
- 2 × PointGray Flea2 color cameras (FL2-14S3C-C), 1.4 Megapixels, 1/2” Sony ICX267 CCD, global shutter
- 4 × Edmund Optics lenses, 4mm, opening angle $\sim 90^\circ$, vertical opening angle of region of interest (ROI) $\sim 35^\circ$
- 1 × Velodyne HDL-64E rotating 3D laser scanner, 10 Hz, 64 beams, 0.09 degree angular resolution, 2 cm distance accuracy, collecting $\sim 1.3$ million points/second, field of view: $360^\circ$ horizontal, $26.8^\circ$ vertical, range: 120 m
Sensor Setup

- **2 × PointGray Flea2 grayscale cameras** (FL2-14S3M-C), 1.4 Megapixels, 1/2" Sony ICX267 CCD, global shutter
- **2 × PointGray Flea2 color cameras** (FL2-14S3C-C), 1.4 Megapixels, 1/2" Sony ICX267 CCD, global shutter
- **4 × Edmund Optics lenses**, 4mm, opening angle $\sim 90^\circ$, vertical opening angle of region of interest (ROI) $\sim 35^\circ$
- **1 × Velodyne HDL-64E** rotating 3D laser scanner, 10 Hz, 64 beams, 0.09 degree angular resolution, 2 cm distance accuracy, collecting $\sim 1.3$ million points/second, field of view: 360° horizontal, 26.8° vertical, range: 120 m
- **1 × OXTS RT3003 inertial and GPS** navigation system, 6 axis, 100 Hz, L1/L2 RTK, resolution: 0.02m / 0.1°
Figure: Sensor Setup. Dimensions and mounting positions of the sensors (red) with respect to the vehicle body. Heights above ground are marked in green and measured with respect to the road surface. Transformations between sensors are shown in blue.
What’s the problem of using so many sensors?

One has to Calibrate and Registered them

- 360° Velodyne Laserscanner
- Stereo Camera Rig
- GPS
- Monochrome
- Color

One has to **Calibrate** and **Registered** them
What’s the problem of using so many sensors?

- One has to **calibrate** and **register** them
  - Different 3D locations
What’s the problem of using so many sensors?

- One has to Calibrate and Registered them
  - Different 3D locations
  - Different capture times
What’s the problem of using so many sensors?

- One has to **Calibrate** and **Registered** them
  - Different 3D locations
  - Different capture times
  - Different types of capture: instantaneous vs scanning
**LIDAR for the dummies**

- **LIDAR**: Light Detection and Ranging
- Measures distance by illuminating a target with a laser and analyzing the reflected light
- Play video
Velodyne HDL64 LIDAR

- Most used LIDAR for autonomous driving
- Play video
Different Velodyne LIDARs

HDL-64E  
HDL-32E  
PUCK™
Example 2: Tesla

- A forward radar
- A forward-looking camera
- 12 long-range ultrasonic sensors positioned to sense 16 feet around the car in every direction at all speeds
- GPS
- A high-precision digitally-controlled electric assist breaking system

- Autopilot is on the Market on Model S
Tesla’s Autopilot

- Does it always work?
- Is technology ready for deployment?
Tesla’s Autopilot

- Does it always work?
- Is technology ready for deployment?

As a consequence some features have been taken off the market
What are we going to cover in this course?

- Sensors and platforms
- **Datasets**
- Models, Algorithms and Techniques:
  - Motion Estimation
  - Reconstruction and free-space estimation
  - Localization
  - Semantics
  - Mapping
- Law and Ethics
Datasets/Benchmarks

- KITTI Benchmark
- Daimler (old)
- Camvid
- Cityscapes (soon to be available)
KITTI Dataset

- Diversity:
  - Captured in Karlsruhe, Germany
KITTI Dataset

- Diversity:
  - Captured in Karlsruhe, Germany
  - A day in spring
KITTI Dataset

Diversity:

- Captured in Karlsruhe, Germany
- A day in spring
- Daytime
KITTI Dataset

- Diversity:
  - Captured in Karlsruhe, Germany
  - A day in spring
  - Daytime
  - Good weather conditions
KITTI Dataset

**Diversity:**

- Captured in Karlsruhe, Germany
- A day in spring
- Daytime
- Good weather conditions
- Diverse set of scenes
  - City center
  - Suburbs
  - Highway
Benchmarks: KITTI Big Data Collection

- **Two stereo rigs** (1392 × 512 px, 54 cm base, 90° opening)
- **Velodyne laser scanner, GPS+IMU localization**
- **6 hours** at 10 frames per second → 3Tb
First Difficulty: Sensor Calibration

360° Velodyne Laserscanner

- Camera calibration [Geiger et al., ICRA 2012]
- Velodyne ↔ Camera registration
- GPS+IMU ↔ Velodyne registration
**Second Difficulty: Object Annotation**

- **3D object labels:** Annotators (undergrad students from KIT working for months)

- **Occlusion labels:** Mechanical Turk
More than 300 submissions, 10,000 downloads since CVPR 2012!

The KITTI Vision Benchmark Suite

Welcome to the KITTI Vision Benchmark Suite!

We take advantage of our autonomous driving platform Annieway to develop novel challenging real-world computer vision benchmarks. Our tasks of interest are: stereo, optical flow, visual odometry, 3D object detection and 3D tracking. For this purpose, we equipped a standard station wagon with two high-resolution color and grayscale video cameras. Accurate ground truth is provided by a Velodyne laser scanner and a GPS localization system. Our datasets are captured by driving around the mid-size city of Karlsruhe, in rural areas and on highways. Up to 15 cars and 30 pedestrians are visible per image. Besides providing all data in raw format, we extract benchmarks for each task. For each of our benchmarks, we also provide an evaluation metric and this evaluation website. Preliminary experiments show that methods ranking high on established benchmarks such as Middlebury perform below average when being moved outside the laboratory to the real world. Our goal is to reduce this bias and complement existing benchmarks by providing real-world benchmarks with novel difficulties to the community.
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
2. **Optical Flow**: 200 training, 200 testing images
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
2. **Optical Flow**: 200 training, 200 testing images
3. **Scene Flow**: 200 training, 200 testing images
4. **Visual Odometry**: 22 videos of 40km
5. **Object Detection**: 7,500 training, 7,500 testing images
6. **Object Tracking**: 21 training and 29 test sequences
7. **Road segmentation**: 289 training and 290 test images

What's missing?

Lack of semantic segmentation

Why?
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
2. **Optical Flow**: 200 training, 200 testing images
3. **Scene Flow**: 200 training, 200 testing images
4. **Visual Odometry**: 22 videos of 40km
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
2. **Optical Flow**: 200 training, 200 testing images
3. **Scene Flow**: 200 training, 200 testing images
4. **Visual Odometry**: 22 videos of 40km
5. **Object Detection**: 7,500 training, 7,500 testing images

- **What’s missing?**
  - Lack of semantic segmentation
- **Why?**
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
2. **Optical Flow**: 200 training, 200 testing images
3. **Scene Flow**: 200 training, 200 testing images
4. **Visual Odometry**: 22 videos of 40km
5. **Object Detection**: 7,500 training, 7,500 testing images
6. **Object Tracking**: 21 training and 29 test sequences

What's missing?

- Lack of semantic segmentation

Why?
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
2. **Optical Flow**: 200 training, 200 testing images
3. **Scene Flow**: 200 training, 200 testing images
4. **Visual Odometry**: 22 videos of 40km
5. **Object Detection**: 7,500 training, 7,500 testing images
6. **Object Tracking**: 21 training and 29 test sequences
7. **Road segmentation**: 289 training and 290 test images
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
2. **Optical Flow**: 200 training, 200 testing images
3. **Scene Flow**: 200 training, 200 testing images
4. **Visual Odometry**: 22 videos of 40km
5. **Object Detection**: 7,500 training, 7,500 testing images
6. **Object Tracking**: 21 training and 29 test sequences
7. **Road segmentation**: 289 training and 290 test images

What’s missing?
Tasks Covered in KITTI

1. **Stereo**: 200 images training, 200 testing
2. **Optical Flow**: 200 training, 200 testing images
3. **Scene Flow**: 200 training, 200 testing images
4. **Visual Odometry**: 22 videos of 40km
5. **Object Detection**: 7,500 training, 7,500 testing images
6. **Object Tracking**: 21 training and 29 test sequences
7. **Road segmentation**: 289 training and 290 test images

What’s missing?

- **Lack of semantic segmentation**
Tasks Covered in KITTI

1. Stereo: 200 images training, 200 testing
2. Optical Flow: 200 training, 200 testing images
3. Scene Flow: 200 training, 200 testing images
4. Visual Odometry: 22 videos of 40km
5. Object Detection: 7,500 training, 7,500 testing images
6. Object Tracking: 21 training and 29 test sequences
7. Road segmentation: 289 training and 290 test images

What’s missing?

Lack of semantic segmentation

Why?
KITTI Dataset

- Metadata:
KITTI Dataset

- Metadata:
  - Preceding and trailing video frames.
KITTI Dataset

- Metadata:
  - Preceding and trailing video frames.
  - Corresponding right stereo views
KITTI Dataset

- **Metadata:**
  - Preceding and trailing *video* frames.
  - Corresponding right *stereo* views
  - LIDAR
KITTI Dataset

- Metadata:
  - Preceding and trailing video frames.
  - Corresponding right stereo views
  - LIDAR
  - GPS coordinates
KITTI Dataset

- Metadata:
  - Preceding and trailing *video* frames.
  - Corresponding right *stereo* views
  - LIDAR
  - GPS coordinates
  - IMU
Cityscapes Dataset

- Diversity:
  - 50 cities

play video
Cityscapes Dataset

- Diversity:
  - 50 cities
  - Several months (spring, summer, fall)
Cityscapes Dataset

- Diversity:
  - 50 cities
  - Several months (spring, summer, fall)
  - Daytime
Cityscapes Dataset

Diversity:
- 50 cities
- Several months (spring, summer, fall)
- Daytime
- Good/medium weather conditions
Cityscapes Dataset

- Diversity:
  - 50 cities
  - Several months (spring, summer, fall)
  - Daytime
  - Good/medium weather conditions
  - Large number of dynamic objects
Cityscapes Dataset

- Type of annotations:
  - **Semantic**: 25 classes

<table>
<thead>
<tr>
<th>Group</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ground</td>
<td>road · sidewalk</td>
</tr>
<tr>
<td>human</td>
<td>person* · rider*</td>
</tr>
<tr>
<td>vehicle</td>
<td>car* · truck* · bus* · on rails* · motorcycle* · bicycle* · license plate+</td>
</tr>
<tr>
<td>infrastructure</td>
<td>building · wall · fence · traffic sign · traffic light · pole · pole group · bridge+ · tunnel+</td>
</tr>
<tr>
<td>nature</td>
<td>tree · terrain</td>
</tr>
<tr>
<td>sky</td>
<td>sky</td>
</tr>
<tr>
<td>void</td>
<td>ground+ · dynamic+ · static+</td>
</tr>
</tbody>
</table>
Cityscapes Dataset

- Type of annotations:
  - **Semantic**: 25 classes
  - **Instance-level**:
    - As many classes as objects per image
    - Difficulty: results are correct up to permutations of the labels

(semantic) (instance)
Cityscapes Dataset

- **Volume**
  - 5,000 annotated images with **fine** annotations
Cityscapes Dataset

- Volume
  - 5,000 annotated images with fine annotations
  - 20,000 annotated images with coarse annotations

play video
Cityscapes Dataset

- Metadata:
  - Preceding and trailing video frames. Each annotated image is the 20th image from a 30 frame video snippets (1.8s)
  - Corresponding right stereo views
  - GPS coordinates
  - Ego-motion data from vehicle odometry
  - Outside temperature from vehicle sensor
Cityscapes Dataset

- Metadata:
  - Preceding and trailing video frames. Each annotated image is the 20th image from a 30 frame video snippets (1.8s)
Cityscapes Dataset

- **Metadata:**
  - Preceding and trailing **video** frames. Each annotated image is the 20th image from a 30 frame video snippets (1.8s)
  - Corresponding right **stereo** views
Cityscapes Dataset

- Metadata:
  - Preceding and trailing video frames. Each annotated image is the 20th image from a 30 frame video snippets (1.8s)
  - Corresponding right stereo views
  - GPS coordinates
Cityscapes Dataset

- **Metadata:**
  - Preceding and trailing *video* frames. Each annotated image is the 20th image from a 30 frame video snippets (1.8s)
  - Corresponding right *stereo* views
  - *GPS* coordinates
  - Ego-motion data from *vehicle odometry*
Cityscapes Dataset

- Metadata:
  - Preceding and trailing video frames. Each annotated image is the 20th image from a 30 frame video snippets (1.8s)
  - Corresponding right stereo views
  - GPS coordinates
  - Ego-motion data from vehicle odometry
  - Outside temperature from vehicle sensor
Cityscapes Dataset

- Metadata:
  - Preceding and trailing *video* frames. Each annotated image is the 20th image from a 30 frame video snippets (1.8s)
  - Corresponding right *stereo* views
  - GPS coordinates
  - Ego-motion data from *vehicle odometry*
  - Outside temperature from vehicle sensor

play video
Cityscapes Dataset

Metadata:

- Preceding and trailing video frames. Each annotated image is the 20th image from a 30 frame video snippets (1.8s)
- Corresponding right stereo views
- GPS coordinates
- Ego-motion data from vehicle odometry
- Outside temperature from vehicle sensor
What are we going to cover in this course?

- Sensors and platforms
- Datasets
- Models, Algorithms and Techniques:
  - Reconstruction and free-space estimation
  - Motion Estimation
  - Localization
  - Semantics
  - Mapping
- Law and Ethics
3D Reconstruction

Types of problems:
Types of problems:

- Given two cameras do **stereo** estimation
3D Reconstruction

Types of problems:

- Given two cameras do **stereo** estimation
  - Dense reconstruction
Types of problems:

- Given two cameras do **stereo** estimation
  - Dense reconstruction
  - Not so accurate in far away ($\geq 40 - 60$ m)
Types of problems:

- Given two cameras do **stereo** estimation
  - Dense reconstruction
  - Not so accurate in far away ($\geq 40 - 60m$)
  - Single time instance
3D Reconstruction

Types of problems:

- Given two cameras do **stereo** estimation
  - Dense reconstruction
  - Not so accurate in far away ($\geq 40 - 60m$)
  - Single time instance

- **SLAM**: Simultaneous Localization and Mapping
Types of problems:

- Given two cameras do **stereo** estimation
  - Dense reconstruction
  - Not so accurate in far away ($\geq 40 – 60m$)
  - Single time instance
- **SLAM**: Simultaneous Localization and Mapping
  - Assumes no-GPS available
3D Reconstruction

Types of problems:

- Given two cameras do **stereo** estimation
  - Dense reconstruction
  - Not so accurate in far away ($\geq 40 - 60$ m)
  - Single time instance

- **SLAM**: Simultaneous Localization and Mapping
  - Assumes no-GPS available
  - Typically sparse point clouds
What is Stereo?

- Given images captured from two cameras, the goal is to compute a 3D map of the scene.
Depth from Two Views: Stereo

- All points on the projective line to $P$ map to $p$

Figure: One camera
Depth from Two Views: Stereo

- All points on projective line to $P$ in left camera map to a line in the image plane of the right camera
If I search this line to find correspondences...

Figure: If I am able to find corresponding points in two images...
Depth from Two Views: Stereo

- I can get 3D!

**Figure:** I can get a point in 3D by triangulation!
Stereo: Parallel Calibrated Cameras

- For each point \( p_l = (x_l, y_l) \), how do I get \( p_r = (x_r, y_r) \)?
For each point $p_l = (x_l, y_l)$, how do I get $p_r = (x_r, y_r)$? By matching on line $y_r = y_l$.

(left image) right image

the match will be on this line (same y)

(CAREFUL: this is only true for parallel cameras. Generally, line not horizontal)
For each point $p_l = (x_l, y_l)$, how do I get $p_r = (x_r, y_r)$? By matching on line $y_r = y_l$. 

We are looking for this point

the match will be on the left of $x_l$

how do I find it?
For each point \( p_l = (x_l, y_l) \), how do I get \( p_r = (x_r, y_r) \)? By matching. Patch around \((x_r, y_r)\) should look similar to the patch around \((x_l, y_l)\).

We call this line a \textbf{scanline}

we \textbf{scan} the line and \textbf{compare} patches to the one in the left image

We are looking for a patch on scanline most similar to patch on the left
Why is 3D from Stereo hard?

- Ambiguities (small windows)
- Textureless regions

[Images taken from Bleyer et al.]
Why is 3D from Stereo hard?

- Sensor saturation
- Non-lambertian surfaces
- Computational requirements
Results on KITTI

[K. Yamaguchi, D. McAllester and R. Urtasun, ECCV’14]
Visual SLAM

- Problem of acquiring a map of the environment and the motion of the robot
Visual SLAM

- Problem of acquiring a map of the environment and the motion of the robot
- Chicken and egg problem
Problem of acquiring a map of the environment and the motion of the robot

Chicken and egg problem

Difficulties
Visual SLAM

- Problem of acquiring a map of the environment and the motion of the robot
- Chicken and egg problem
- Difficulties
  - deal with moving objects
Visual SLAM

- Problem of acquiring a map of the environment and the motion of the robot
- Chicken and egg problem
- Difficulties
  - deal with moving objects
  - loop closure
Stereo SLAM

[A. Geiger, M. Roser and R. Urtasun, ACCV’10]
LIDAR+Appearance SLAM

[S. Anderson and T. Barfoot, IROS’15]
Free-space Estimation

- Two definitions of the problem
  - Navigable space
  - Space that is reachable without collision
Obstacle Estimation

- Estimate obstacles that can cause collision
- There are no semantics

[H. Badino, U. Franke and D. Pfeiffer, DAGM’09]
What are we going to cover in this course?

- Sensors and platforms
- Datasets
- Models, Algorithms and Techniques:
  - Reconstruction and free-space estimation
  - Motion Estimation
  - Localization
  - Mapping
  - Semantics
- Law and Ethics
Motion Estimation

- Types of problems:
  - Visual Odometry
  - Optical Flow
  - Scene Flow
Visual Odometry

- Recover the 3D motion of the ego-car (i.e., self-driving car)
Visual Odometry

- Recover the 3D motion of the ego-car (i.e., self-driving car)
- The car has sensors that can help do this task, e.g., speedometer, GPS, IMU
Visual Odometry

- Recover the 3D motion of the ego-car (i.e., self-driving car)
- The car has sensors that can help do this task, e.g., speedometer, GPS, IMU
- Real-time
Visual Odometry

- Recover the 3D motion of the ego-car (i.e., self-driving car)
- The car has sensors that can help do this task, e.g., speedometer, GPS, IMU
- Real-time
- Very small errors when using stereo/LIDAR
Visual Odometry

- Recover the 3D motion of the ego-car (i.e., self-driving car)
- The car has sensors that can help do this task, e.g., speedometer, GPS, IMU
- Real-time
- Very small errors when using stereo/LIDAR
- Monocular case not yet solved
Visual Odometry

- Recover the 3D motion of the ego-car (i.e., self-driving car)
- The car has sensors that can help do this task, e.g., speedometer, GPS, IMU
- Real-time
- Very small errors when using stereo/LIDAR
- Monocular case not yet solved
- Note that SLAM estimates odometry
• **Motion Field**: is the projection of the 3D scene motion into the image
Flow due to Camera Motion

- If the scene is static and camera is moving
Flow due to Camera Motion

- If the scene is static and camera is moving
  - **Length** of flow vectors inversely proportional to depth of 3d point
Flow due to Camera Motion

- If the scene is static and camera is moving
  - Length of flow vectors inversely proportional to depth of 3d point
  - Types of motion vectors

![Flow vectors](image)

- Zoom out
- Zoom in
- Pan right to left
Most of the flow is due to the vehicle’s *ego-motion*
Flow for Autonomous Driving

- Most of the flow is due to the vehicle’s ego-motion
- Goal: compute the epipolar flow by doing matching along the epipolar lines
Flow for Autonomous Driving

- Most of the flow is due to the vehicle’s ego-motion
- Goal: compute the epipolar flow by doing matching along the epipolar lines
Flow for Autonomous Driving

- Most of the flow is due to the vehicle’s **ego-motion**
- Goal: compute the **epipolar flow** by doing matching along the epipolar lines

![Diagram showing epipolar flow and FOE (Fundamental Orientation Epipole)]

- The problem is very similar to stereo
Flow for Autonomous Driving

- Most of the flow is due to the vehicle’s ego-motion
- Goal: compute the epipolar flow by doing matching along the epipolar lines

The problem is very similar to stereo
- We can exploit the same techniques as for stereo
Dynamic Scenes

- If the scene is dynamic, then not just a rigid transformation; it's more difficult
Optical Flow: Given two subsequent frames, estimate the apparent motion field between them
Optical Flow: Given two subsequent frames, estimate the apparent motion field between them

Key assumptions
Optical Flow: Given two subsequent frames, estimate the apparent motion field between them

Key assumptions

- Brightness constancy: projection of the same point looks the same in every frame
Optical Flow

- **Optical Flow:** Given two subsequent frames, estimate the apparent motion field between them

- **Key assumptions**
  - **Brightness constancy:** projection of the same point looks the same in every frame
  - **Small motion:** points do not move very far

- **Difficulties**
  - Large displacements
  - Occlusions
  - Specularities
Optical Flow

- **Optical Flow**: Given two subsequent frames, estimate the apparent motion field between them

- **Key assumptions**
  - **Brightness constancy**: projection of the same point looks the same in every frame
  - **Small motion**: points do not move very far
  - **Spatial coherence**: points move like their neighbors

- **Difficulties**
  - Large displacements
  - Occlusions
  - Specularities
Optical Flow

- **Optical Flow**: Given two subsequent frames, estimate the apparent motion field between them

- **Key assumptions**
  - **Brightness constancy**: projection of the same point looks the same in every frame
  - **Small motion**: points do not move very far
  - **Spatial coherence**: points move like their neighbors

- **Difficulties**
Optical Flow

- **Optical Flow**: Given two subsequent frames, estimate the apparent motion field between them.

- **Key assumptions**
  - **Brightness constancy**: projection of the same point looks the same in every frame
  - **Small motion**: points do not move very far
  - **Spatial coherence**: points move like their neighbors

- **Difficulties**
  - Large displacements
**Optical Flow**

- **Optical Flow**: Given two subsequent frames, estimate the apparent motion field between them

- **Key assumptions**
  - **Brightness constancy**: projection of the same point looks the same in every frame
  - **Small motion**: points do not move very far
  - **Spatial coherence**: points move like their neighbors

- **Difficulties**
  - Large displacements
  - Occlusions
Optical Flow

- **Optical Flow**: Given two subsequent frames, estimate the apparent motion field between them

- **Key assumptions**
  - **Brightness constancy**: projection of the same point looks the same in every frame
  - **Small motion**: points do not move very far
  - **Spatial coherence**: points move like their neighbors

- **Difficulties**
  - Large displacements
  - Occlusions
  - Specularities
Scene Flow

- Recover the 3D motion of the scene
Scene Flow

- Recover the 3D motion of the scene
  - Epipolar Flow if the scene is static
Scene Flow

- Recover the 3D motion of the scene
  - Epipolar Flow if the scene is static
  - Estimate Motion of a dynamic Scene
Scene Flow

- Recover the 3D motion of the scene
  - Epipolar Flow if the scene is static
  - Estimate Motion of a dynamic Scene
    - Most Visual approaches (if not all) use stereo
Scene Flow

- Recover the 3D motion of the scene
  - Epipolar Flow if the scene is static
  - Estimate Motion of a dynamic Scene
    - Most Visual approaches (if not all) use stereo
    - In robotics, rely on 3D sensor (e.g., Lidar)
Scene Flow

- Recover the 3D motion of the scene
  - Epipolar Flow if the scene is static
  - Estimate Motion of a dynamic Scene
    - Most Visual approaches (if not all) use stereo
    - In robotics, rely on 3D sensor (e.g., Lidar)

- Very difficult to get ground-truth

Typically evaluated in terms of stereo error and flow error

[M. Menze and A. Geiger, CVPR'15]
Scene Flow

- Recover the 3D motion of the scene
  - Epipolar Flow if the scene is static
  - Estimate Motion of a dynamic Scene
    - Most Visual approaches (if not all) use stereo
    - In robotics, rely on 3D sensor (e.g., Lidar)

- Very difficult to get ground-truth
  - Why can’t I just use the LIDAR?
Scene Flow

- Recover the 3D motion of the scene
  - Epipolar Flow if the scene is static
  - Estimate Motion of a dynamic Scene
    - Most Visual approaches (if not all) use stereo
    - In robotics, rely on 3D sensor (e.g., Lidar)

- Very difficult to get ground-truth
  - Why can’t I just use the LIDAR?

- Typically evaluated in terms of stereo error and flow error
Scene Flow

- Recover the 3D motion of the scene
  - Epipolar Flow if the scene is static
  - Estimate Motion of a dynamic Scene
    - Most Visual approaches (if not all) use stereo
    - In robotics, rely on 3D sensor (e.g., Lidar)

- Very difficult to get ground-truth
  - Why can’t I just use the LIDAR?

- Typically evaluated in terms of stereo error and flow error

[M. Menze and A. Geiger, CVPR’15]
What are we going to cover in this course?

- Sensors and platforms
- Datasets
- Models, Algorithms and Techniques:
  - Reconstruction and free-space estimation
  - Motion Estimation
  - Localization
  - Semantics
  - Mapping
- Law and Ethics
Motivation

- Localization is crucial for autonomous systems

- GPS has limitations in terms of reliability and availability
1. **Place recognition** techniques use image features or depth maps and a database of previously collected images (e.g., Google car).

- Easier problem
- Requires knowing how the world looks like
- Problems with changes in appearance, e.g., weather, construction, new building

2. **Appearance agnostic**
   - More difficult problem
   - No knowledge of the world, apart from cartographic maps
   - No problems with changes in appearance
1. **Place recognition** techniques use image features or depth maps and a database of previously collected images (e.g., Google car)
   - Easier problem

2. Appearance agnostic
   - More difficult problem
   - No knowledge of the world, apart from cartographic maps
   - No problems with changes in appearance
Types of Localization Approaches

1. **Place recognition** techniques use image features or depth maps and a database of previously collected images (e.g., Google car)
   - Easier problem
   - Requires knowing how the world looks like

2. Appearance agnostic
   - More difficult problem
   - No knowledge of the world, apart from cartographic maps
   - No problems with changes in appearance

R. Urtasun (UofT)
Types of Localization Approaches

1. **Place recognition** techniques use image features or depth maps and a database of previously collected images (e.g., Google car)
   - Easier problem
   - Requires knowing how the world looks like
   - Problems with changes in appearance environment, e.g., weather, construction, new building
Types of Localization Approaches

1. **Place recognition** techniques use image features or depth maps and a database of previously collected images (e.g., Google car)
   - Easier problem
   - Requires knowing how the world looks like
   - Problems with changes in appearance environment, e.g., weather, construction, new building

2. **Appearance agnostic**
Types of Localization Approaches

1. **Place recognition** techniques use image features or depth maps and a database of previously collected images (e.g., Google car)
   - Easier problem
   - Requires knowing how the world looks like
   - Problems with changes in appearance environment, e.g., weather, construction, new building

2. **Appearance agnostic**
   - More difficult problem
Types of Localization Approaches

1. **Place recognition** techniques use image features or depth maps and a database of previously collected images (e.g., Google car)
   - Easier problem
   - Requires knowing how the world looks like
   - Problems with changes in appearance environment, e.g., weather, construction, new building

2. **Appearance agnostic**
   - More difficult problem
   - No knowledge of the world, apart from cartographic maps
1. **Place recognition** techniques use image features or depth maps and a database of previously collected images (e.g., Google car)
   - Easier problem
   - Requires knowing how the world looks like
   - Problems with changes in appearance environment, e.g., weather, construction, new building

2. **Appearance agnostic**
   - More difficult problem
   - No knowledge of the world, apart from cartographic maps
   - No problems with changes in appearance
Place Recognition Examples

- Similarity in point clouds
- Similarity in visual features
- Similarity in detected skylines
Humans are able to use a map, combined with visual input and exploration, to localize effectively.

Detailed, community developed maps are freely available (OpenStreetMap).

How can we exploit maps, combined with visual cues, to localize a vehicle?
Results

[M. Brubaker, A. Geiger and R. Urtasun, CVPR'13 best paper runner up award]
What are we going to cover in this course?

- Sensors and platforms
- Datasets
- Models, Algorithms and Techniques:
  - Reconstruction and free-space estimation
  - Motion Estimation
  - Localization
  - Semantics
  - Mapping
- Law and Ethics
Semantic Tasks

1. Detection
2. Tracking
3. Semantic Segmentation
4. Instance-level segmentation
5. 3D Scene Understanding
Object Detection

- **Task**: Bounding box around the object of interest and determine its class
- Dominated by deep learning
- Very little work on 3D object detection
Car Results

[X. Chen, K. Kundu, Y. Zhu, A. Berneshawi, H. Ma, S. Fidler and R. Urtasun, NIPS’15]
**Task:** Place bounding boxes at each frame, and link them over time

Optimal solutions (i.e., minCostFlow) exist if second order dynamics
Semantic Segmentation

- **Task:** Label each pixel with a semantic category
- Dominated by deep learning + graphical models
Results on CamVid

[J. Tighe and S. Lazebnik, ECCV’10]
Instance-level segmentation

- **Task:** Label each pixel with an instance number
- Difficult as labeling is **agnostic to permutation** of the labels
- Very little work on this topic
- Dominated by **deep learning + graphical models**

![Instance-level segmentation example](image)
Task: Holistic reasoning of everything that is in the scene
- Involves many semantic tasks
- Typically frame with graphical models
- Individual components use deep learning
- Little published work
Example of 3D Scene Understanding

Goal: Infer from a short (≈10s) video sequence:

- **Geometric properties**, e.g., street orientation
- **Topological properties**, e.g., number of intersecting streets
- **Semantic activities**, e.g., traffic situations at an intersection
- **3D objects**, e.g., cars
Observations

- **3D Tracklets**: Generate tracklets from 2D detections in 3D by employing the orientation as well as size of the bounding boxes.
Observations

- **3D Tracklets**: Generate tracklets from 2D detections in 3D by employing the orientation as well as size of the bounding boxes.

- **Segmentation** of the scene into semantic labels.
Observations

- **3D Tracklets**: Generate tracklets from 2D detections in 3D by employing the orientation as well as size of the bounding boxes.

- **Segmentation** of the scene into semantic labels.

- **Lines** that follow the dominant orientations in the scene (i.e., reasoning about vanishing points).
Observations

- **3D Tracklets**: Generate tracklets from 2D detections in 3D by employing the orientation as well as size of the bounding boxes.
- **Segmentation** of the scene into semantic labels.
- **Lines** that follow the dominant orientations in the scene (i.e., reasoning about vanishing points).

Representation

- We will reason about dynamics in bird eye’s perspective and static in the image.
Why high-order semantics?

- Certain behaviors are not possible given the traffic "patterns"

- These patterns can be learned by watching video
Understanding High-Level Semantics by Modeling Traffic Patterns

Input Video

Inference Result

Inferred Scene Layout

Inferred Object

Inferred Traffic Pattern

Observer
What are we going to cover in this course?

- Sensors and platforms
- Datasets

Models, Algorithms and Techniques:
- Reconstruction and free-space estimation
- Motion Estimation
- Localization
- Semantics
- Mapping

- Law and Ethics
Types of Mapping

- From the **ground**
  - SLAM
  - Semantic mapping
Types of Mapping

- From the **ground**
  - SLAM
  - Semantic mapping
- From **crowd sourcing**
Types of Mapping

- From the ground
  - SLAM
  - Semantic mapping
- From crowd sourcing
- From aerial images
Crowdsourced maps: OSM

- Free more than 50% of the world is map
- It contains errors: misalignments, missing roads and other info
Aerial Images for enhancing maps

- Road segmentation from aerial image segmentation
- Large coverage
- Many challenges
Challenges of Aerial Road Segmentation

(a) shadow  (b) occlusion  
(c) vehicles  (d) misaligned centerline
Aerial Images for enhancing maps

[G. Mattyus, S. Wang, S. Fidler and R. Urtasun, ICCV’15]

- Toronto: Airport
- San Francisco: Russian Hill
- NYC: Times square
- Kyoto: Kinkakuji
- Sydney: At Harbour bridge
- Monte Carlo: Casino

- Can segment the whole world in 24h in 10 computers!
What are we going to cover in this course?

- Sensors and platforms
- Datasets
- Models, Algorithms and Techniques:
  - Reconstruction and free-space estimation
  - Motion Estimation
  - Localization
  - Semantics
  - Mapping
- Law and Ethics
Topics to Choose for Presentation

- **Reconstruction**
  - Stereo estimation
  - SLAM

- **Motion Estimation**
  - Visual Odometry
  - Optical Flow and Scene Flow

- **Localization**
  - Place Recognition approaches
  - Appearance-less localization

- **Mapping**
  - Semantic SLAM
  - Aerial Image Parsing

- **Semantics**
  - Detection
  - Semantic Segmentation
  - Instance-level segmentation

- **Law and Ethics**