Organization

- **Lecturer**
  - Prof. Bastian Leibe (leibe@vision.rwth-aachen.de)

- **Teaching Assistants**
  - N.N.
  - Jonathan Luiten (luiten@vision.rwth-aachen.de)

- **Course webpage**
  - [http://www.vision.rwth-aachen.de/courses/](http://www.vision.rwth-aachen.de/courses/)
    - → Computer Vision 2
  - Slides will be made available on the webpage and in the L2P
  - Screencasts of the lecture will be uploaded to L2P

- **Please subscribe to the lecture in rwth online!**
  - Important to get email announcements and L2P access!

Language

- **Official course language will be English**
  - If at least one English-speaking student is present.
  - If not… you can choose.

- **However…**
  - Please tell me when I’m talking too fast or when I should repeat something in German for better understanding!
  - You may ask questions in German.
  - You may turn in your exercises in German.
  - You may answer exam questions in German.

Course Webpage

- [http://www.vision.rwth-aachen.de/courses/](http://www.vision.rwth-aachen.de/courses/)

Exercises and Demos

- **Exercises**
  - Typically 1 exercise sheet every 2 weeks (mainly Matlab based)
  - Hands-on experience with the algorithms from the lecture.
  - Send in your solutions the night before the exercise class.

- **Teams are encouraged!**
  - You can form teams of up to 3 people for the exercises.
  - Each team should only turn in one solution.
  - But list the names of all team members in the submission.
Textbooks

- No single textbook for the class, but some basics can be found in

- We will mostly give out research papers
  - Tutorials for basic techniques
  - State-of-the-art research papers for current developments

Computer Vision 2

- We will build upon the basics from previous lectures
  - Computer Vision 1
  - Machine Learning

- However,
  - If you haven’t heard those lectures yet, you may still attend and benefit from this lecture.
  - But please look at the available online material from those lectures to get additional background on the basic techniques.
  - I will regularly point out what background to repeat.

How to Find Us

- Office:
  - UMIC Research Centre
  - Mies-van-der-Rohe-Strasse 15, 2nd floor

- Office hours
  - If you have questions to the lecture, come to the assistants or me.
  - Our regular office hours will be announced (additional slots are available upon request)
  - Send us an email before to confirm a time slot.

Questions are welcome!

Why Computer Vision?

- Cameras are all around us…

Images and Video Are Everywhere…

- Personal photo albums
- Movies, news, sports
- Internet services
- Surveillance and security
- Medical and scientific images
- Mobile and consumer applications

Computer Vision 1 Covered the Basics…

- Image Processing Basics
- Segmentation
- Local Features & Matching
- Object Recognition and Categorization
- 3D Reconstruction
Computer Vision 2 Is All About Motion!

How can we track an object’s motion over time?

Motion Requires Video

- A video is a sequence of frames captured over time
- Our image data is a function of space \((x, y)\) and time \((t)\)

What is Tracking?

- **Goal**
  - Estimate the number and state of objects in a region of interest

- **Number**
  - 1: Single-target tracking
  - 0 or 1: Detection and tracking
  - N: Multi-target detection and tracking

- **Objects**
  - We will look at a large variety of objects to track.
  - They can be given by a user or detected automatically.
  - Very interesting are people.
  - Special case: Tracking the camera pose wrt. the environment/object

What distinguishes tracking from “typical” statistical estimation (or machine learning) problems?

- Typically a strong temporal component is involved.
- Estimating quantities that are expected to change over time (thus, expectations of the dynamics play a role).
- Interested in current state \(S_t\) for a given time step \(t\).
- Usually assume that we can only compute information seen at previous time steps \(1, 2, \ldots, t-1\). (Can’t look into the future!)
- Usually we want to be as efficient as possible, even “real-time”.

⇒ These concerns lead naturally to recursive estimators.
Types of Tracking

- Single-object tracking
  - Focuses on tracking a single target in isolation.

- Multi-object tracking
  - Tries to follow the motion of multiple objects simultaneously.

- Articulated tracking
  - Tries to estimate the motion of objects with multiple, coordinated parts.

- Active tracking
  - Involves moving the sensor in response to motion of the target. Needs to be real-time!

Applications: Safety & Security

- Autonomous robots
- Driver assistance
- Monitoring pools (Poseidon)
- Pedestrian detection
- Surveillance

Applications: Human-Computer Interaction

- Games (Microsoft Kinect)
- Assistive technology systems
  - Camera Mouse
    - (Boston College)
Applications: Visual Effects

Why Are There So Many Papers on Tracking?

• Because what kind of tracking “works” depends on problem-specific factors...

Elements of Tracking

- Detection
  - Find the object(s) of interest in the image.

- Association
  - Determine which observations come from the same object.

- Prediction
  - Predict future motion based on the observed motion pattern.
  - Use this prediction to improve detection and data association in later frames.
How can we track camera motion over time and reconstruct the environment?

Visual Odometry

• … is a variant of tracking
  – Track the motion of the camera (in position and orientation) from its images
  – Only considers a limited set of recent images for real-time constraints

• … also involves a data association problem
  – Motion is estimated from corresponding interest points or pixels in images, or by correspondences towards a local 3D reconstruction

What is Visual Odometry?

Visual Odometry

• … is prone to drift due to its local view

• … is primarily concerned with estimating camera motion
  – Not all approaches estimate a 3D reconstruction of the associated interest points/pixels explicitly.
  – If so it is only locally consistent

What is Visual Odometry?

What is Visual SLAM?

• SLAM stands for Simultaneous Localization and Mapping
  – Estimate the pose of the camera in a map, and simultaneously
  – Estimate the parameters of the environment map (i.e. reconstruct the 3D positions of interest points in a common coordinate frame)

• Loop-closure: Revisiting a place allows for drift compensation
  – How to detect?

What is Visual SLAM?
How Should We Represent the Map?

- Sparse interest points
- Volumetric, implicit surface
- Explicit surface (surfels, mesh, …)
- Keyframe-based maps

Content of the Lecture

- Single-Object Tracking
  - Background modeling
  - Template-based tracking
  - Tracking by online classification
  - Tracking-by-detection
- Bayesian Filtering
- Multi-Object Tracking
- Visual Odometry
- Visual SLAM & 3D Reconstruction
- Deep Learning for Video Analysis

Template Tracking

- Lucas-Kanade registration applied to tracking => KLT

Video sequence

Tracked template

J. Shi and C. Tomasi, Good Features to Track, CVPR 1994.
### Content of the Lecture

- Single-Object Tracking
- Bayesian Filtering
  - Kalman Filters, EKF
  - Particle Filters
- Multi-Object Tracking
  - Multi-hypothesis data association
  - MHT, JPDAF
- Visual Odometry
- Visual SLAM & 3D Reconstruction

### Multi-Person Tracking


### Application: Tracking Sports Players

- Automatic player tracking for sports scene analysis
  - Several companies active in this area...

### Application: Automotive Driver Assistance Systems

- Combined Image and World-Space Tracking [ICRA'17]

### Application: Mobile Robot Navigation

- Link to the video

### Application: Wearable Computing

- Person detection + Tracking + Visual odometry @ 25-30 fps on 1 CPU
Mobile Tracking in Densely Populated Settings

Content of the Lecture
- Single-Object Tracking
- Bayesian Filtering
- Multi-Object Tracking
- Visual Odometry
  - Sparse interest-point-based methods
  - Dense direct methods
- Visual SLAM & 3D Reconstruction
- Deep Learning for Video Analysis

Visual Odometry – Direct, Semi-Dense

Content of the Lecture
- Single-Object Tracking
- Bayesian Filtering
- Multi-Object Tracking
- Visual Odometry
- Visual SLAM & 3D Reconstruction
  - Map representations, image registration and integration
  - Tracking-and-mapping
  - Loop-closing, pose-graph optimization, bundle adjustment
  - Dense multi-view stereo depth reconstruction
- Deep Learning for Video Analysis

RGB-D SLAM – Implicit Surface (KinectFusion)

Mono SLAM – Keyframe Pose-Graph (ORB-SLAM)
Outlook

• Computer Vision II is a specialization class
  – We will build upon the basics from the CV I and ML lectures.
  – You can attend the class without having heard those, but please use the available online material for self-study.

• Next lecture: Background modeling
  – Please repeat the following topics from the ML lecture:
    - Gaussians & ML estimation
    - Mixtures of Gaussians & EM
    - Kernel density estimation

Questions?