Computer Vision 2
WS 2018/19

Part 1 – Introduction

Prof. Dr. Bastian Leibe

RWTH Aachen University, Computer Vision Group
http://www.vision.rwth-aachen.de

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/
    $\rightarrow$ Computer Vision2
  - 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 at any time ask questions in German!
  - You may turn in your exercises in German.
  - You may answer exam questions in German.

Organization

- Structure: 3V (lecture) + 1Ü (exercises)
  - 6 EECS credits
  - Part of the area "Applied Computer Science"
- Place & Time
  - Lecture: Tue 10:30 – 12:00 UMIC 025
  - Lecture/Exercises: Wed 08:30 – 10:00 H10
- Exam
  - Oral or written exam (depending on the number of participants)
  - Date will be fixed soon

Course Webpage

Computer Vision 2

Course Schedule

Date	Title	Content	Material	
Thu, 2016-04-14	Introduction	What is tracking? What is visual odometry? What is SLAM?		
Sun, 2016-04-17	Template based Tracking	LK Tracking, fast template matching, Generalized LK		
Thu, 2016-04-21	Exercise 0	Intro Matlab		
Sun, 2016-04-24	Tracking by Online Classification	Tracking as Online Classification problem, Online Boosting, Online Feature Selection, Drift, Semi-Supervised Boosting, TLD		
Thu, 2016-04-28	Tracking by Detection	Tracking-by-Detection, State-of-the-Art Detectors: HOG, DPM, Viola-Jones, Integral Channel Features, VeryFast, Roerei, Hough Forests		

http://www.vision.rwth-aachen.de/courses/

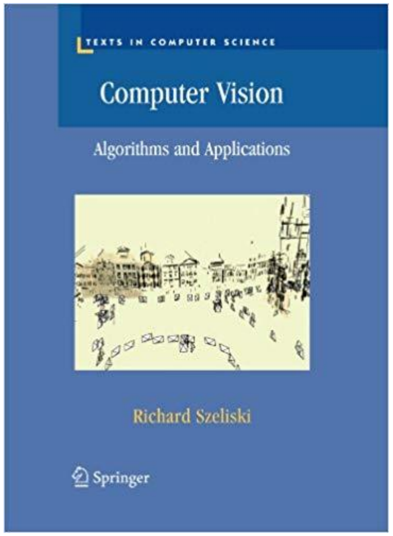
Exercises and Demos

- 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.

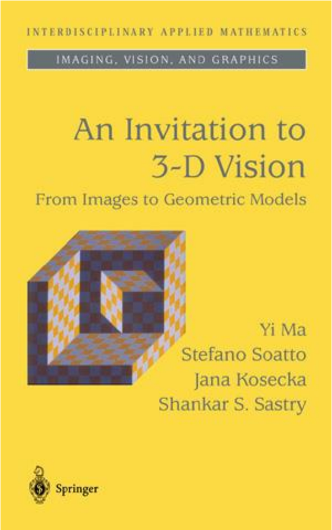
- Teams are encouraged!
  - You can form teams of up to 3 people for the exercises.
  - Each team should only turn in one solution.
  - But list the names of all team members in the submission.

Textbooks

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



Computer Vision – Algorithms and Applications, R. Szeliski, Springer, 2011



An Invitation to 3D Vision, Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, Springer, 2004

- 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 I
  - 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?





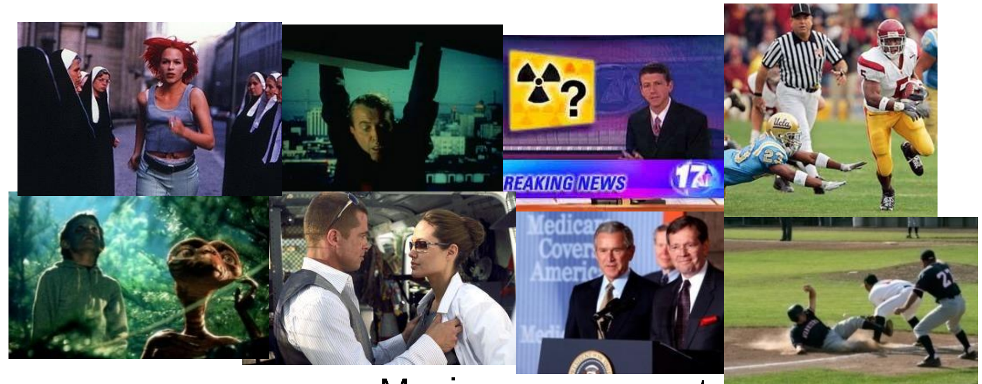
Cameras are
all around us…




Slide credit: Kristen Grauman

Images and Video Are Everywhere…

Personal photo albums



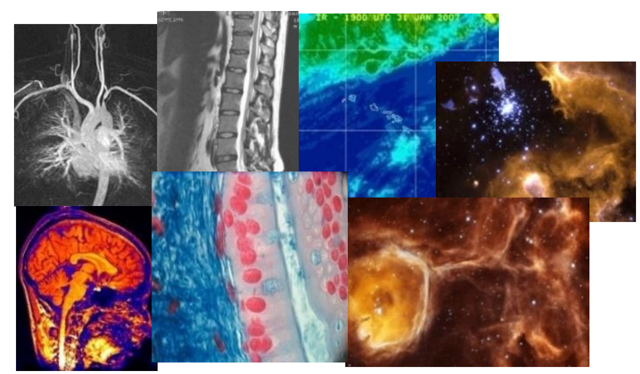
Movies, news, sports

Google Image Search Picasa flickr webshots picsearch YouTube Broadcast Yourself kooaba

Internet services

Surveillance and security

Mobile and consumer applications



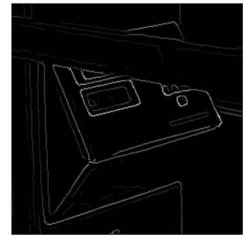
Medical and scientific images

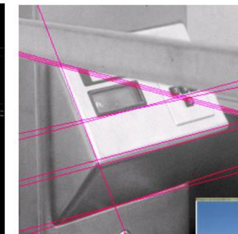
Visual Computing Institute | RWTH AACHEN UNIVERSITY

Slide adapted from Svetlana Lazebnik

Computer Vision 1 Covered the Basics…

- Image Processing Basics

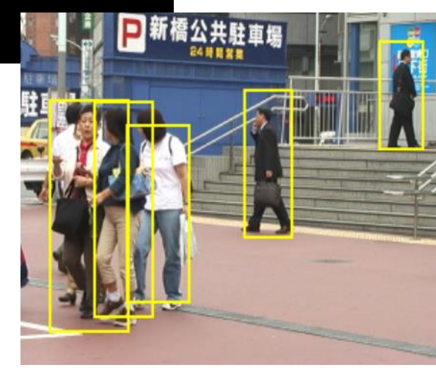




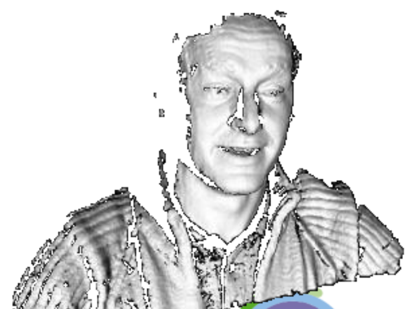
- Segmentation

- Local Features & Matching

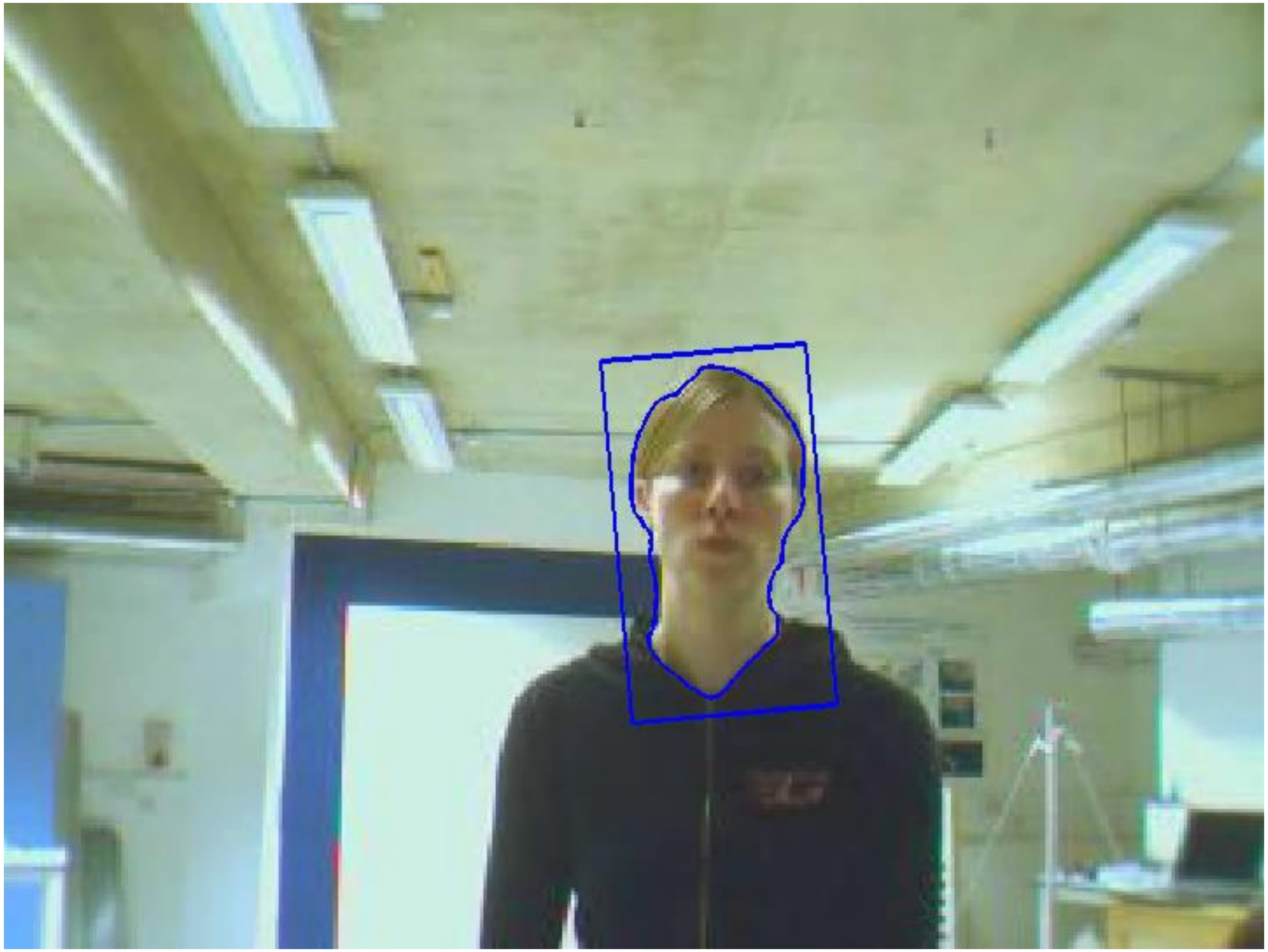
- Object Recognition and Categorization



- 3D Reconstruction



Computer Vision 2 Is All About Motion!

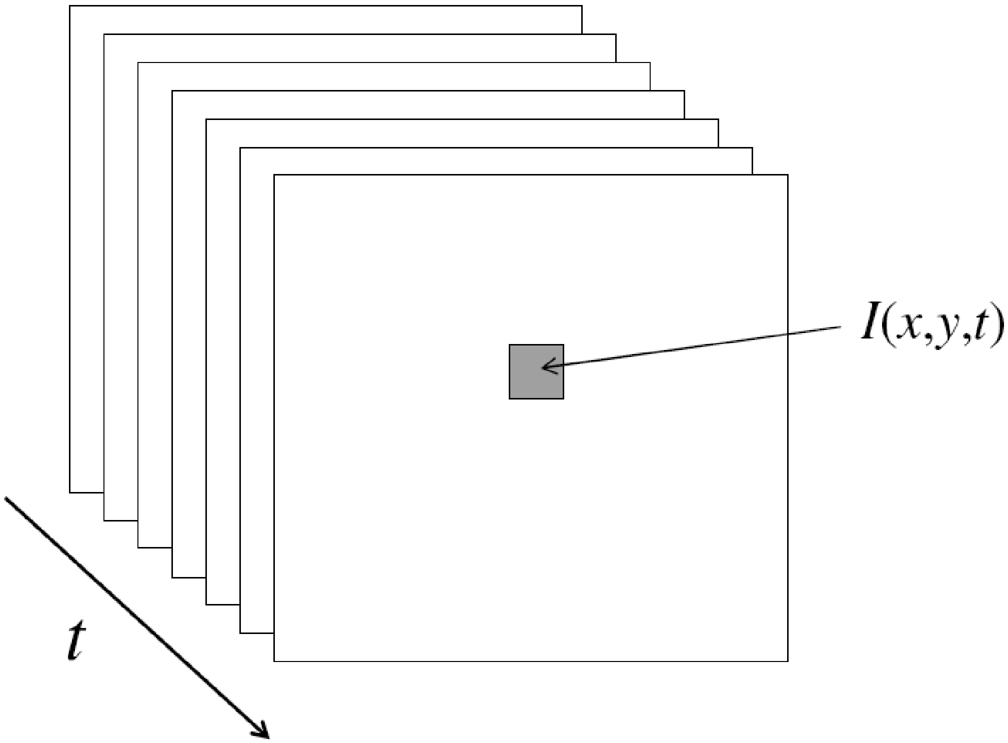


How can we track an object's motion over time?

Video source: [Bibby & Reid, ECCV'08]

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)



Slide credit: Svetlana Lazebnik

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

Slide adapted from Robert Collins

What is Tracking?

- Goal
  - *Estimate the number and state of objects in a region of interest*
- State
  - We are using the term state to describe a vector of quantities that characterize the object being tracked.

E.g.	$[x,\ y]$	(location)
	$[x,\ y,\ dx,\ dy]$	(location + velocity)
	$[x,\ y,\ appearance\text{-}params]$	(location + appearance)

  - Because our observations will be noisy, estimating the state vector will be a statistical estimation problem.

Slide adapted from Robert Collins

What is Tracking?

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

Image sources: Kristen Grauman, Michael Breitenstein, Ahmed Elgammal

What is Tracking?

- Goal
  - *Estimate the number and state of objects in a region of interest*
- 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 $1, 2, \ldots, t\text{-}1$. (*Can't look into the future!*)
  - Usually we want to be as efficient as possible, even "real-time".

$\Rightarrow$ These concerns lead naturally to recursive estimators.

Slide credit: Robert Collins

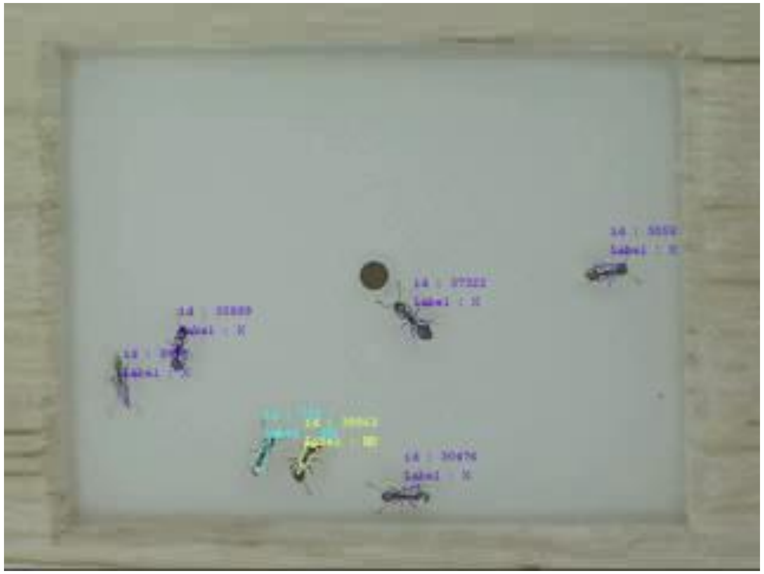
Types of Tracking

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

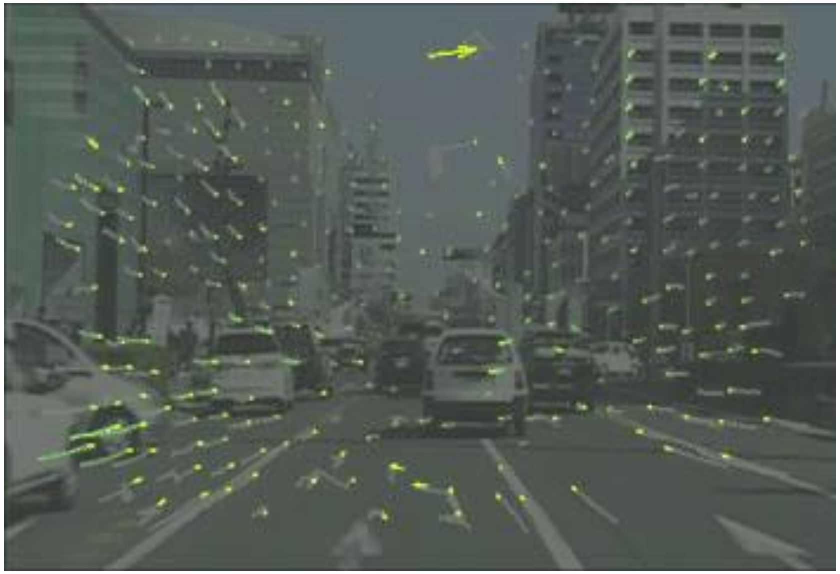
[Z. Kalal, K. Mikolajczyk, J. Matas, PAMI'10]

Types of Tracking

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



Ant behavior, courtesy of Georgia Tech biotracking

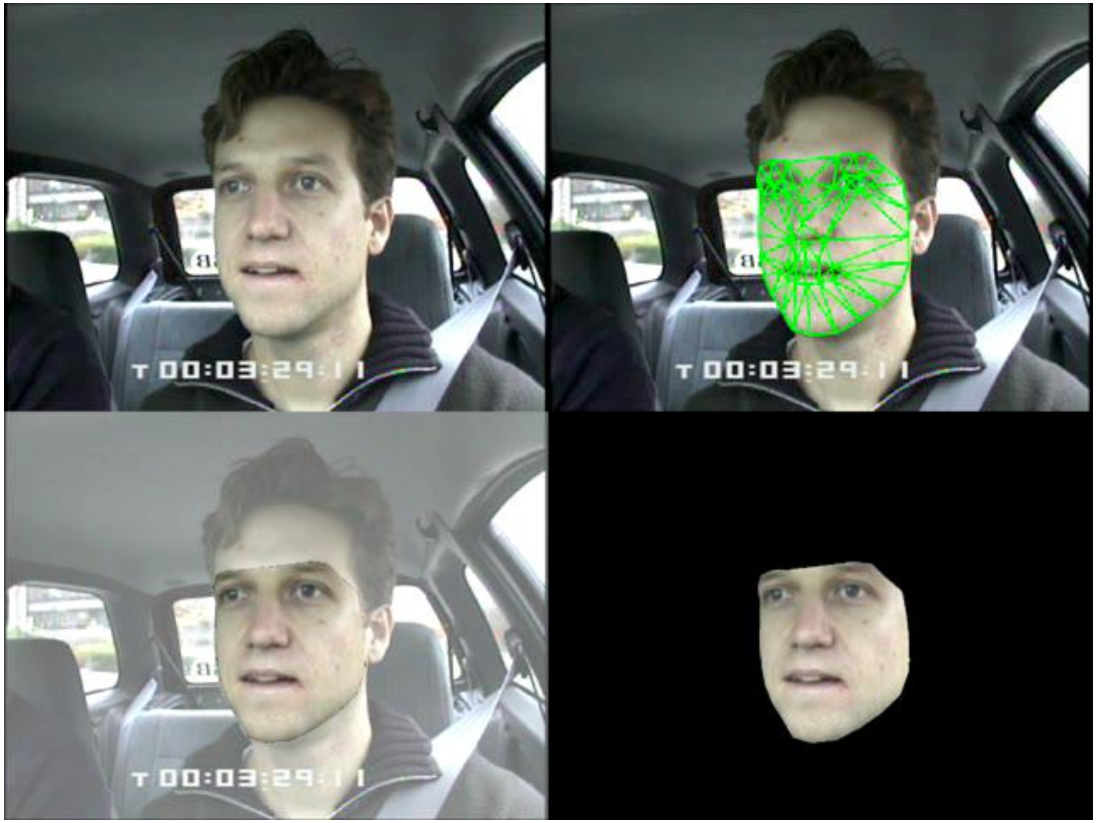


“Objects” can be corners, and tracking gives us optical flow.

Slide credit: Robert Collins

Types of Tracking

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



[I. Matthews, S. Baker, IJCV’04]

Slide credit: Robert Collins

Types of Tracking

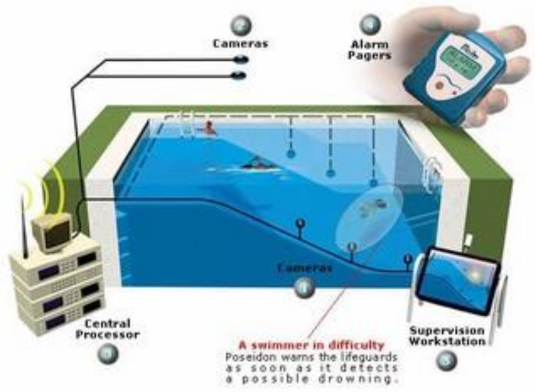
- Active tracking
  - Involves moving the sensor in response to motion of the target. Needs to be real-time!

Slide credit: Robert Collins

Applications: Safety & Security

Autonomous robots

Driver assistance



Monitoring pools
(Poseidon)

Pedestrian detection
[MERL, Viola et al.]

Surveillance

Slide credit: Kristen Grauman

Applications: Human-Computer Interaction

Games
(Microsoft Kinect)

Assistive technology systems
Camera Mouse
(Boston College)

Slide adapted from Kristen Grauman

Applications: Visual Effects

MoCap for *Pirates of the Carribean*, Industrial Light and Magic

Slide adapted from Steve Seitz, Svetlana Lazebnik, Kristen Grauman

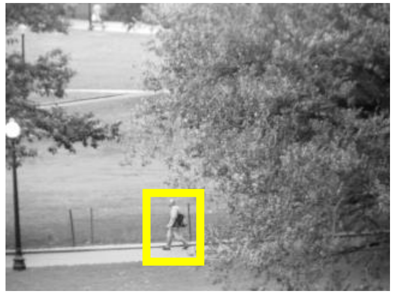
Why Are There So Many Papers on Tracking?

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

image source: Microsoft

Elements of Tracking

t=1



t=2

…

t=20

t=21

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

Image credit: Kristen Grauman

Elements of Tracking

t=1 t=2 t=20 t=21

- Detection
  - Find the object(s) of interest in the image.
- Association
  - Determine which observations come from the same object.

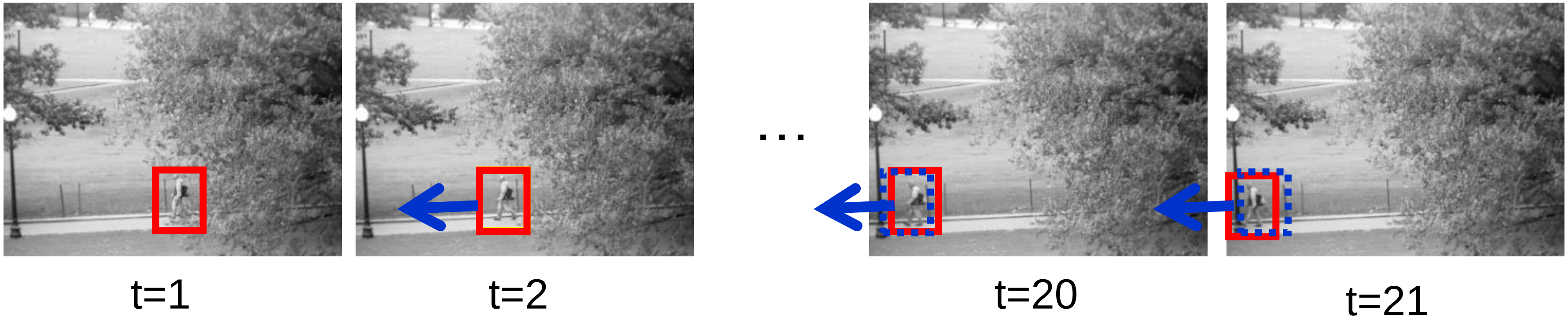
Image credit: Kristen Grauman

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.

Image credit: Kristen Grauman

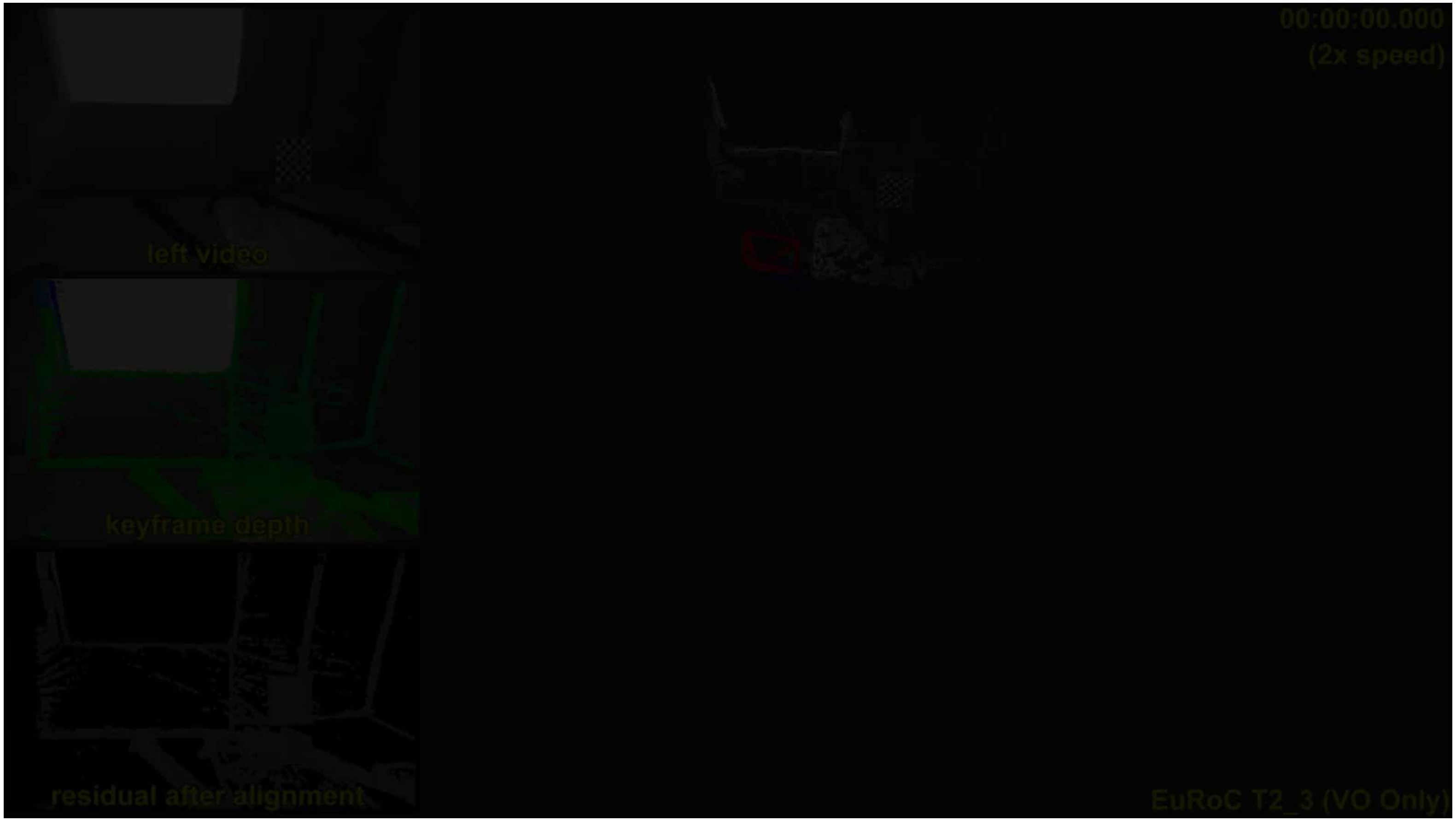
Elements of Tracking



- 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.

Image credit: Kristen Grauman

CV 2 Is Also About Localization and 3D Mapping



How can we track camera motion over time and reconstruct the environment?

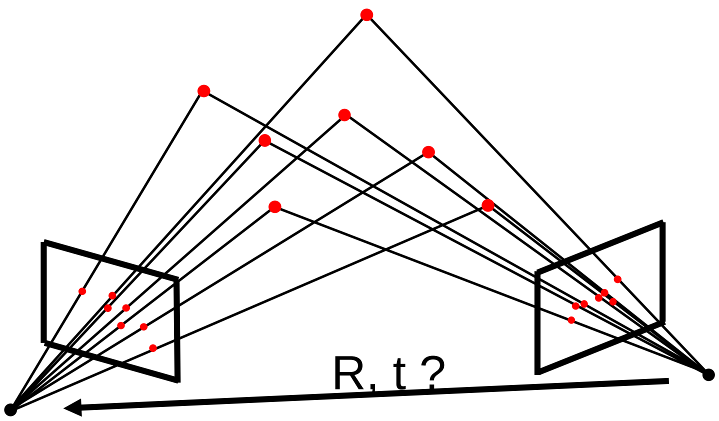
[Engel, Stückler, Cremers, IROS 2015]

What is Visual Odometry ?

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

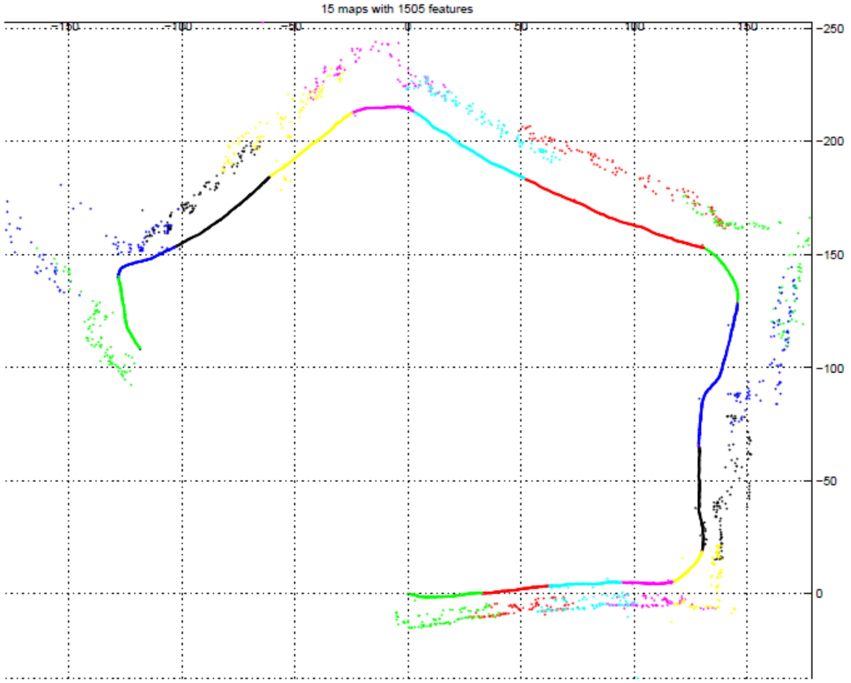
- … also involves a data association problem
  - Motion is estimated from corresponding interest points or pixels in images, or by correspondences towards a local 3D reconstruction



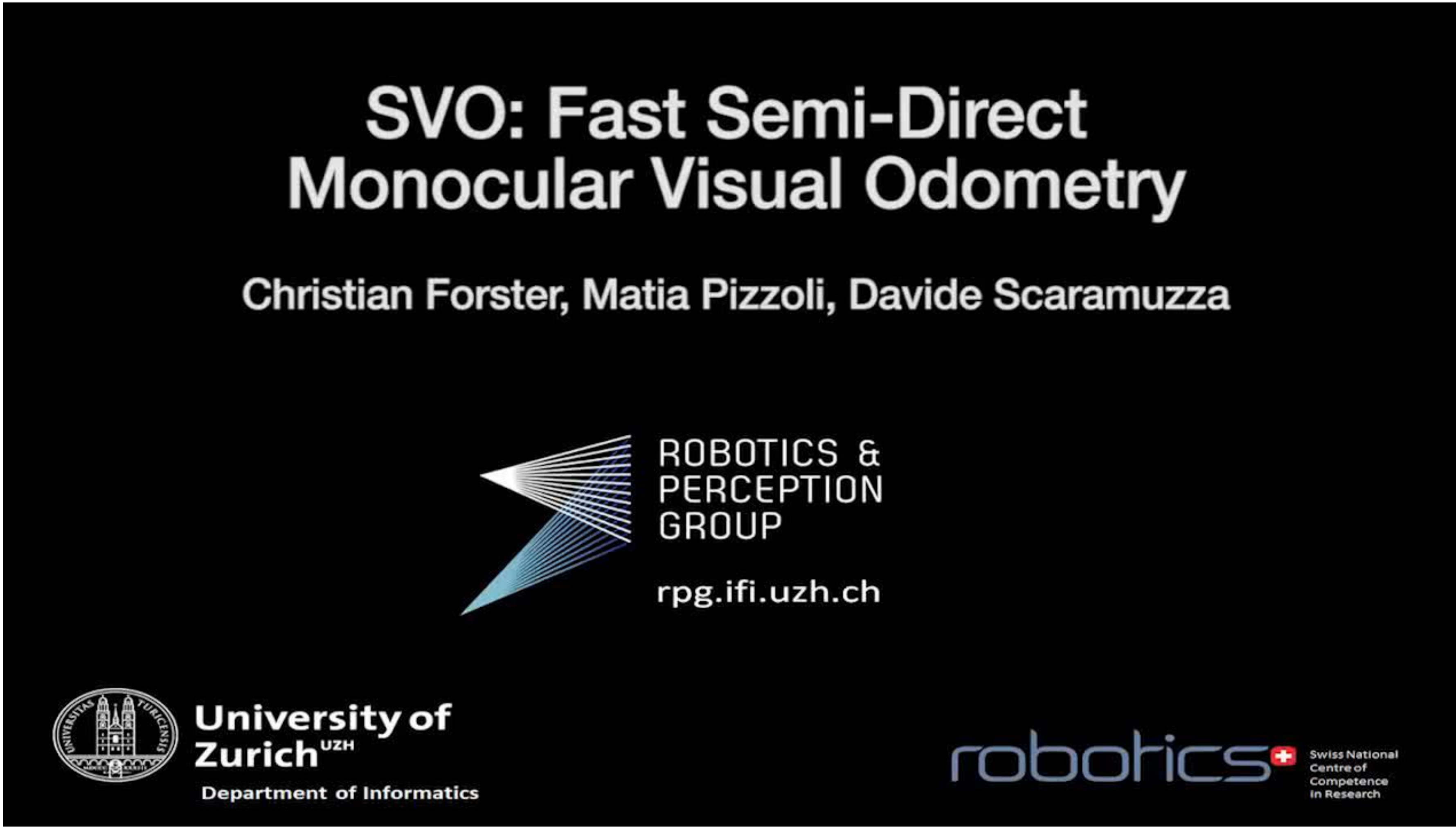
What is Visual Odometry ?

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



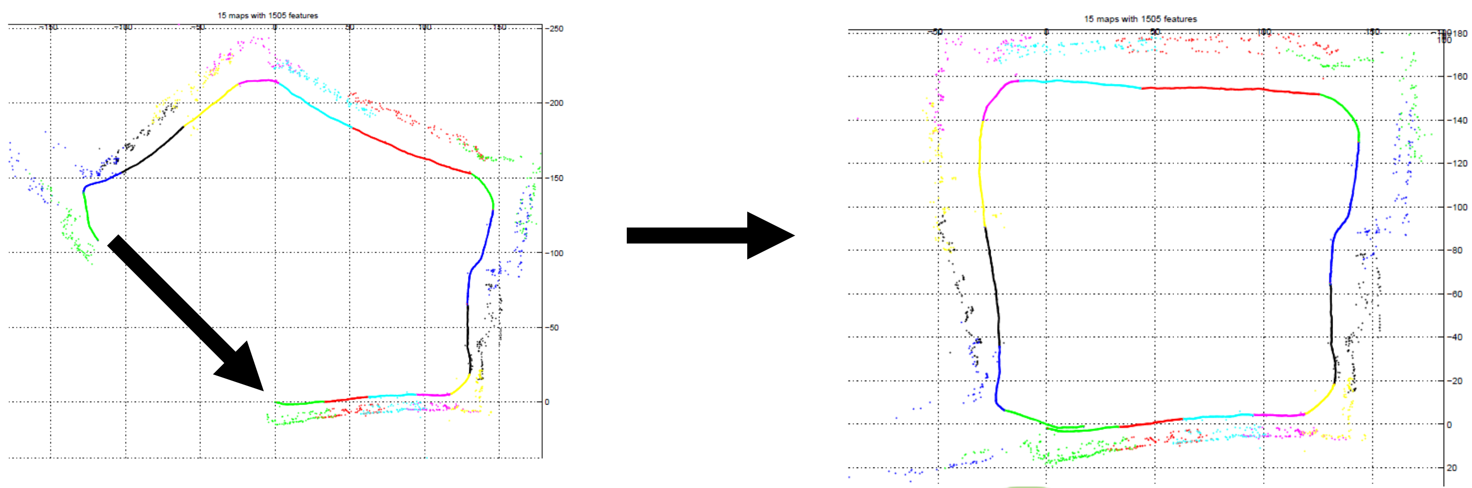
Visual Odometry Example



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 (f.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?



[Clemente et al., RSS 2007]

What is Visual SLAM ?

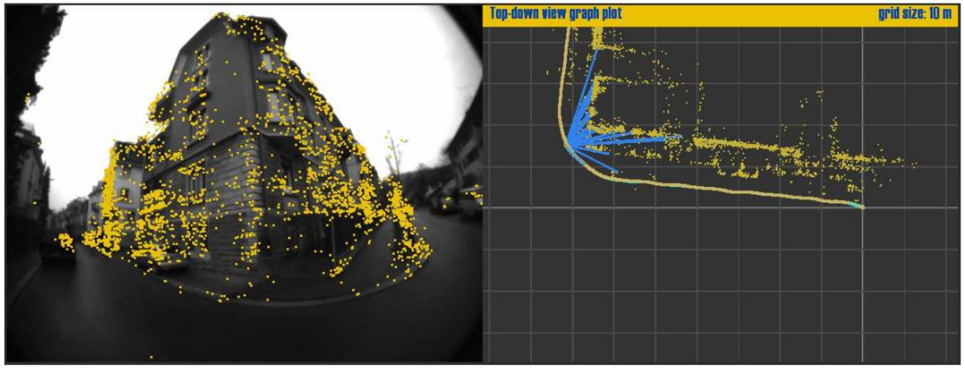
- SLAM stands for Simultaneous Localization and Mapping
  - Estimate the pose of the camera in a map, and simultaneously
  - Estimate the parameters of the environment map (f.e. reconstruct the 3D positions of interest points in a common coordinate frame)

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

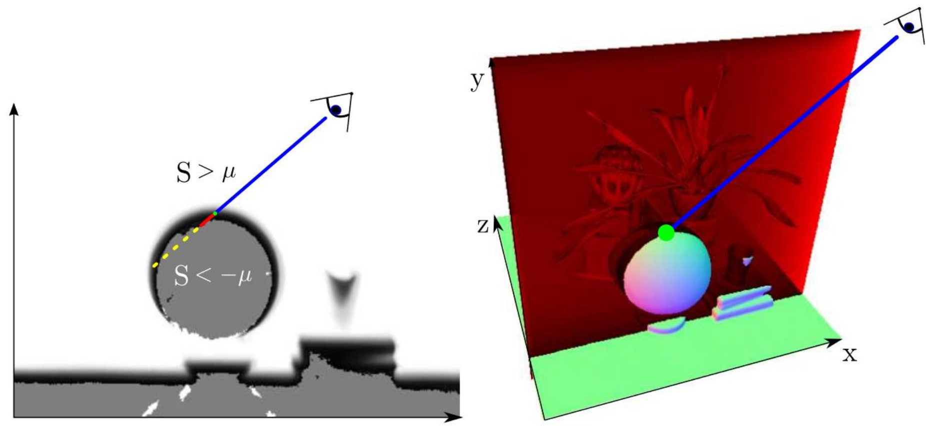
- Global and local optimization methods
  - Global: bundle adjustment, pose-graph optimization. Offline!
  - Local: incremental tracking-and-mapping approaches, visual odometry with local maps. Online! Often designed for real-time.
  - Hybrids: Real-time local SLAM + global optimization in a slower parallel process (f.e. LSD-SLAM)

RGB-D SLAM Example

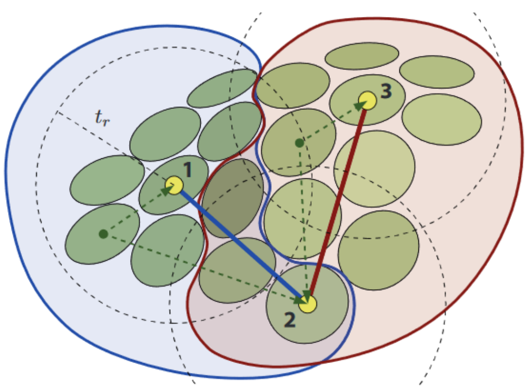
How Should We Represent the Map?



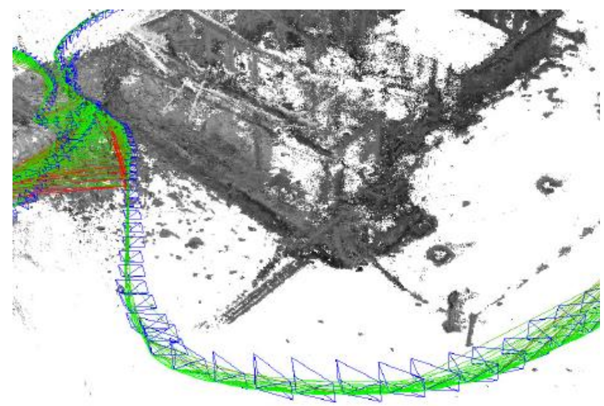
Sparse interest points



Volumetric, implicit surface



Explicit surface (surfels, mesh,…)



Keyframe-based maps

[Lynen et al., RSS 2015], [Newcombe 2015], [Weise et al., 2009], [Engel et al., ECCV 2014]

Content of the Lecture

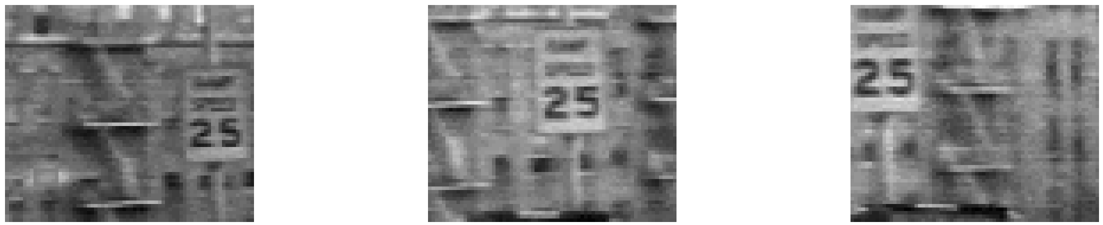
- Single-Object Tracking
  - Background modeling
  - Template based tracking
  - Tracking by online classification
  - Tracking-by-detection
- Bayesian Filtering
- Multi-Object Tracking
- Visual Odometry
- Visual SLAM & 3D Reconstruction
- Deep Learning for Video Analysis

Applications: Visual Surveillance

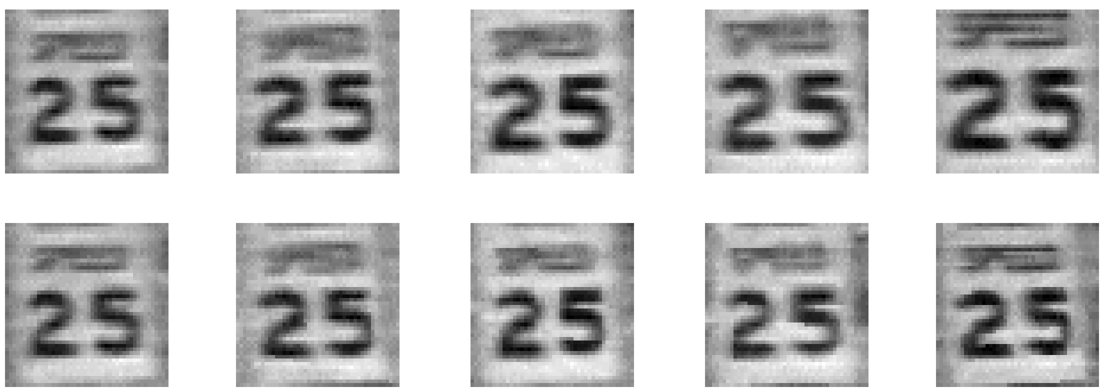
Video by Daniel Roth, ETH Zurich

Template Tracking

- Lucas-Kanade registration applied to tracking $\Rightarrow$ KLT



Video sequence

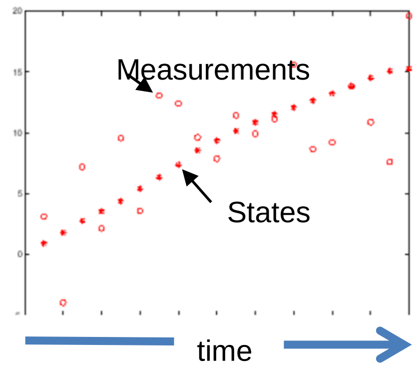


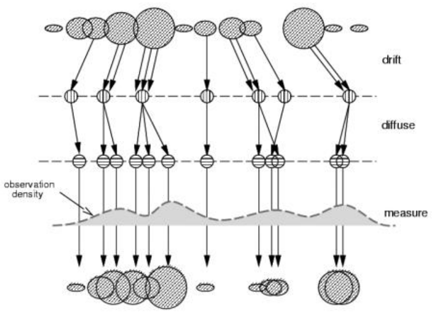
Tracked template

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

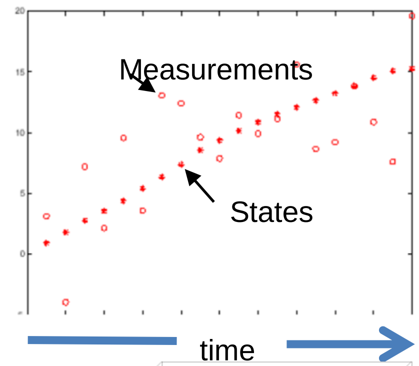


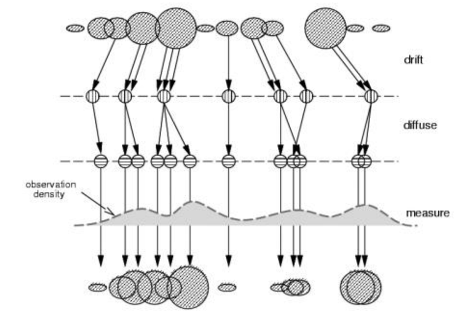


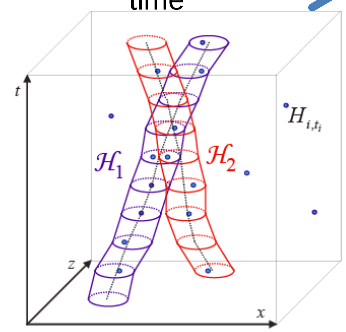
- Multi-Object Tracking
- Visual Odometry
- Visual SLAM & 3D Reconstruction
- Deep Learning for Video Analysis

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







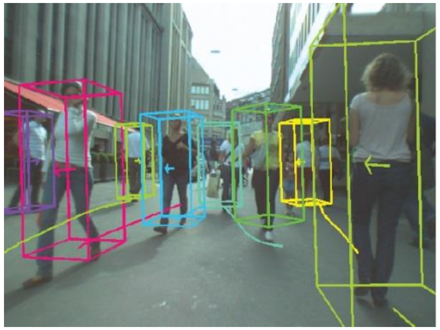


Image sources: Andreas Ess

Multi-Person Tracking

M. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, L. Van Gool. [Online Multi-Person Tracking-by-Detection from a Single, Uncalibrated Camera](), PAMI, Vol.33(9), pp. 1820-1833, 2011.

Application: Tracking Sports Players

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

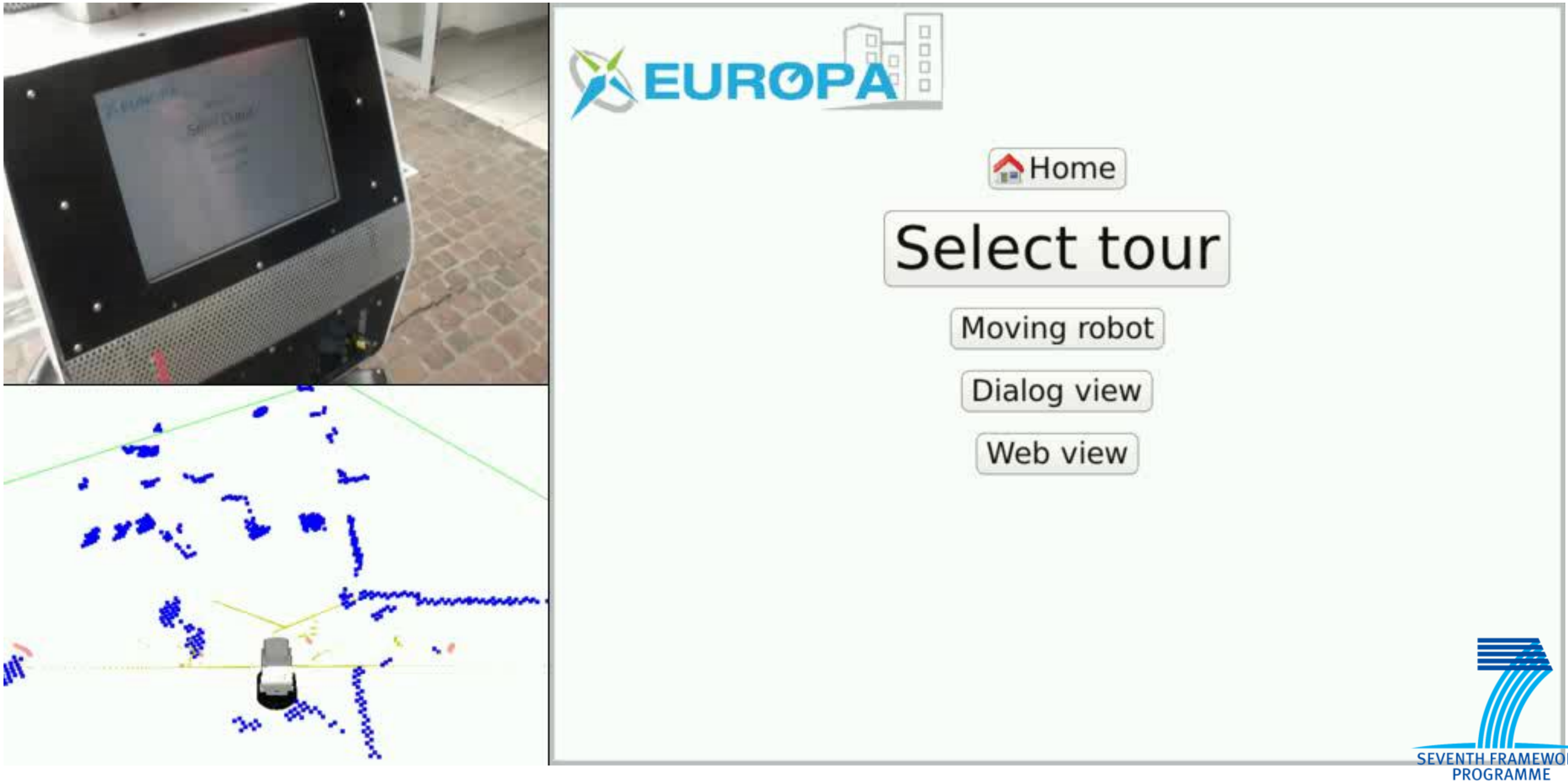
[Breitenstein et al.; PAMI'10]

Application: Automotive Driver Assistance Systems

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

[Osep, Mehner, Mathias, Leibe, ICRA’17]

Application: Mobile Robot Navigation



EUROPA

link to the video

Application: Wearable Computing

– Person detection + Tracking + Visual odometry @ 25-30 fps on 1 CPU

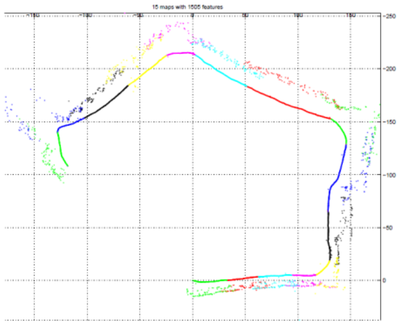
[O. Hosseini Jafari, D. Mitzel, B. Leibe, ICRA'14]

Mobile Tracking in Densely Populated Settings

[D. Mitzel, B. Leibe, ECCV'12]

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)

SIGGRAPH Talks 2011

KinectFusion:

Real-Time Dynamic 3D Surface Reconstruction and Interaction

Shahram Izadi 1, Richard Newcombe 2, David Kim 1,3, Otmar Hilliges 1,
David Molyneaux 1,4, Pushmeet Kohli 1, Jamie Shotton 1,
Steve Hodges 1, Dustin Freeman 5, Andrew Davison 2, Andrew Fitzgibbon 1

1 Microsoft Research Cambridge 2 Imperial College London
3 Newcastle University 4 Lancaster University
5 University of Toronto

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

ORB-SLAM

Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós

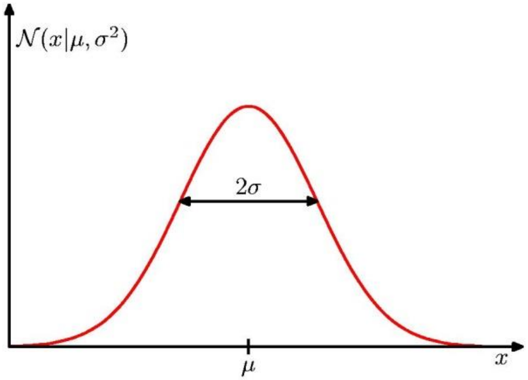
{raulmur, josemari, tardos} @unizar.es

Instituto Universitario de Investigación
en Ingeniería de Aragón
Universidad Zaragoza

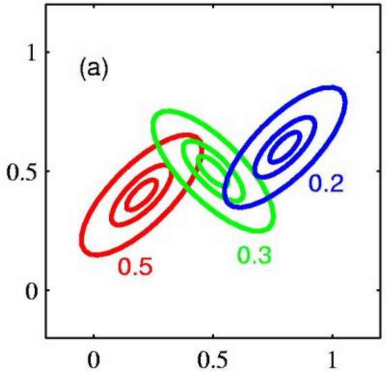
Universidad
Zaragoza
1542

Outlook

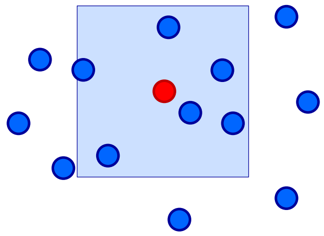
- Computer Vision II is a specialization class
  - We will build upon the basics from the CV I and ML lectures.
  - *You can attend the class without having heard those, but please use the available online material for self-study.*
- Next lecture: Background modeling
  - Please repeat the following topics from the ML lecture:



Gaussians & ML estimation



Mixtures of Gaussians & EM



Kernel density estimation

Questions?