3D Computer Vision

Radim Šára Martin Matoušek

Center for Machine Perception
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague

https://cw.fel.cvut.cz/wiki/courses/tdv/start

http://cmp.felk.cvut.cz
mailto:sara@cmp.felk.cvut.cz
phone ext. 7203

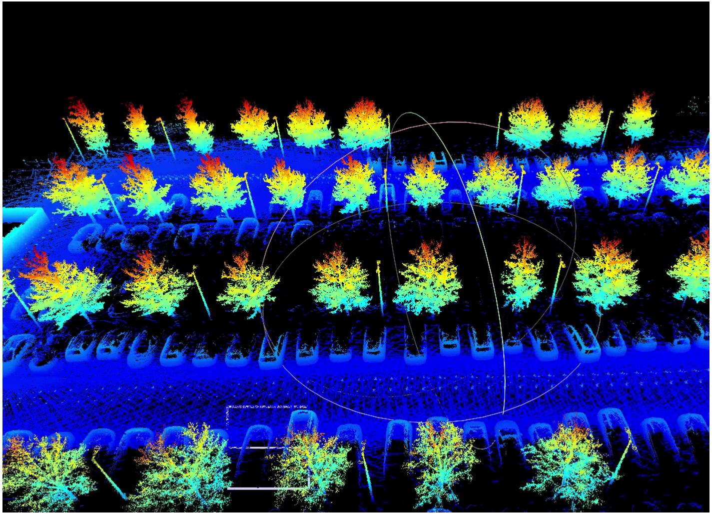
rev. September 26, 2023

Open Informatics Master's Course

Module I

Course Overview

3D Vision is Not Just about 3D Point Clouds

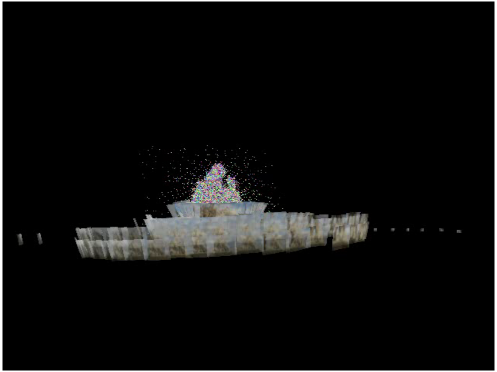


click for video

- today, we have laser-based rangefinders (eg. LiDARs) — active sensors
- figure: point cloud obtained from a vehicle with 4 LiDARs on its roof — color = height
- **this course focuses on obtaining such results (and more) by means of passive sensors**

images + some knowledge about cameras → cameras in 3D + sparse 3D points

36 of 237 images of a memorial



click for video

all camera poses, closest 2m, farthest 40m away

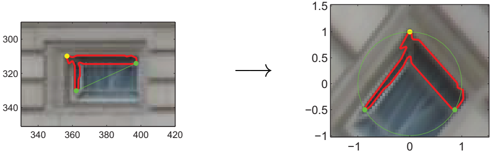
Typical phases of a processing pipeline:

1. finding sparse image matches (Matching)
2. determining correspondences and camera poses (Structure from Motion)
3. finding dense correspondences → 3D point clouds (Stereovision)
4. (optional: occupancy mapping or surface reconstruction)

Phase 1: Sparse Image Matches

image features, their descriptors, matches and correspondences

MSERs



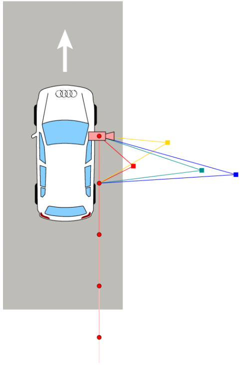
bitangent + distant pt → normalized feature

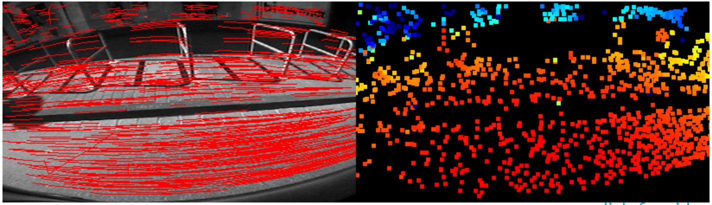
- matches ∼ visually similar
- correspondences ∼ visually similar and geometrically consistent (yellow)
- finding matches must cope with ambiguity
- 5 correspondences determine the relative angular orientation and translation direction between calibrated cameras
  calibrated = we know their internal parameters like the focal length etc

Phase 2: From Matches To Correspondences ("Structure from Motion")

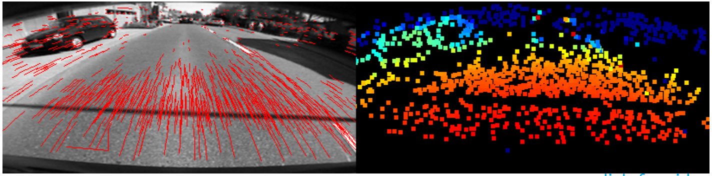
- **Example:** Sensing depth from a single moving camera, 30 fps data stream

  standard automotive wide-angle sensor – reversing camera
- moving videocamera ∼ time constraint on image match evolution ∼ 'optical flow'





click for video



click for video

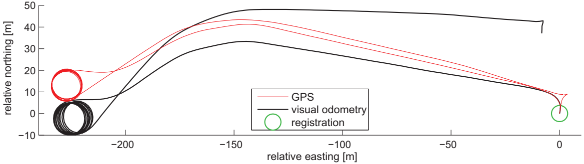
- standard term: SfM (Structure from Motion should rather be Structure and Motion)
- single camera: problems with moving objects (wrong depth)

- reversing camera on a car, 30fps; error against RT 3000 GPS system (red)

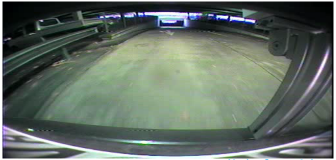
no fusion with GPS!

Scene I

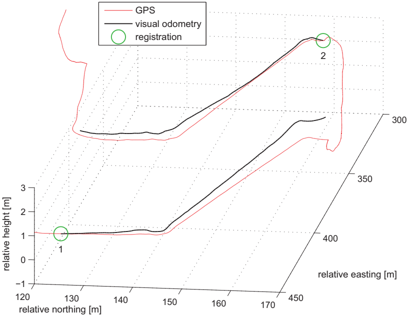
click for video



Scene II



click for video



- 1 km, 5% accumulated drift
- measures elevation
- bad lighting conditions
- difficult scene

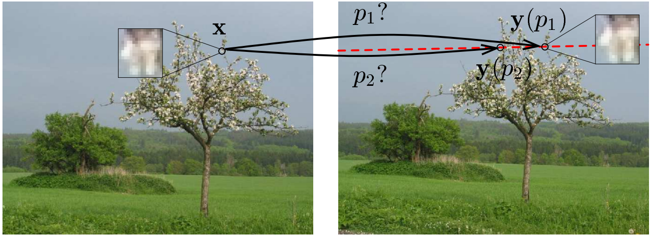
Some applications:

- visual odometry
- SLAM

- the drift is reduced if the correspondences linking camera pairs form a dense graph, not a chain like here

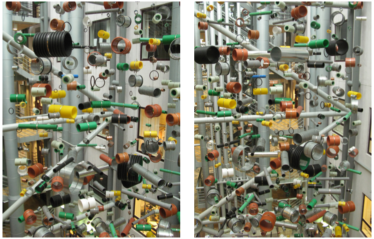
Phase 3: Dense Correspondences by Stereovision

stereo seems easy in familiar scenes

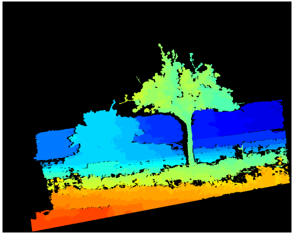


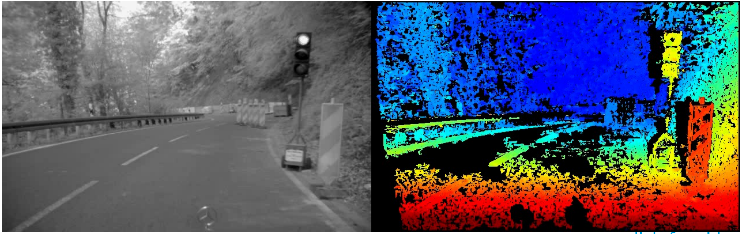
input images

unfamiliar scenes are difficult



Malmö Högskola, Centrum för teknikstudier





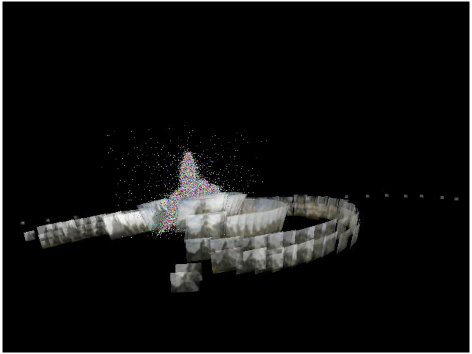
click for video

- the result is a dense 3D point cloud (color = range)

typically 10^6 – 10^9 3D points

Phase 4: Point Cloud and Surface Reconstruction

cameras + point cloud + images $\rightarrow$ triangulated surface



click for video

click for video

- we will not cover surface reconstruction in this course
- (but you may be able to use one of the popular algorithms)

Thank You