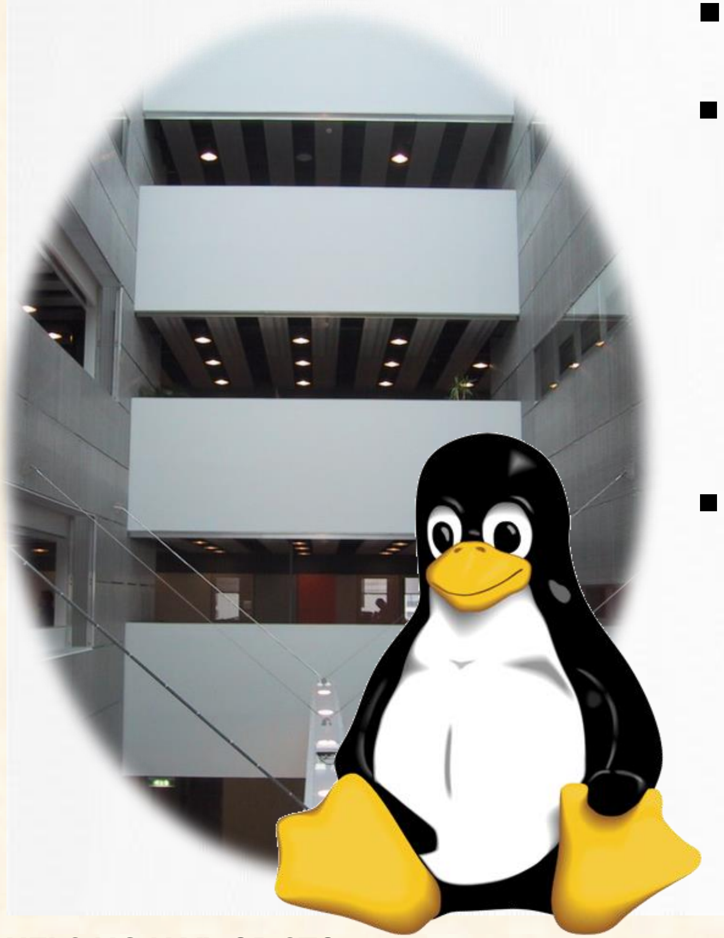
Visions of Navigation

Dr. Laura Ruotsalainen
Associate Professor, Department of Computer Science,
University of Helsinki

51 YEARS OF EXCELLENCE



- **Department of Computer Science**
- Leading institution in Computer Science in Finland
  - THE #93 (2018)
  - The number of professors is **growing** from 16 in 2017 to 28 at the end of 2018.

- **Core CS and Data Science**
  - Algorithms
  - AI
  - Networking
  - Software

ME

- Associate Professor in Spatiotemporal Data Analysis for Sustainabililty Science, Department of Computer Science, since August 2018
- Finnish Geospatial Research Institute, Department of Navigation and Positioning, 2010 –
  - Leader of the Sensors and Indoor Navigation research group
  - 10% Research Professor 2018 -
- PhD ”Vision-aided Pedestrian Navigation for Challenging GNSS Environments” 2013

HUMANS' STEREO CAMERA

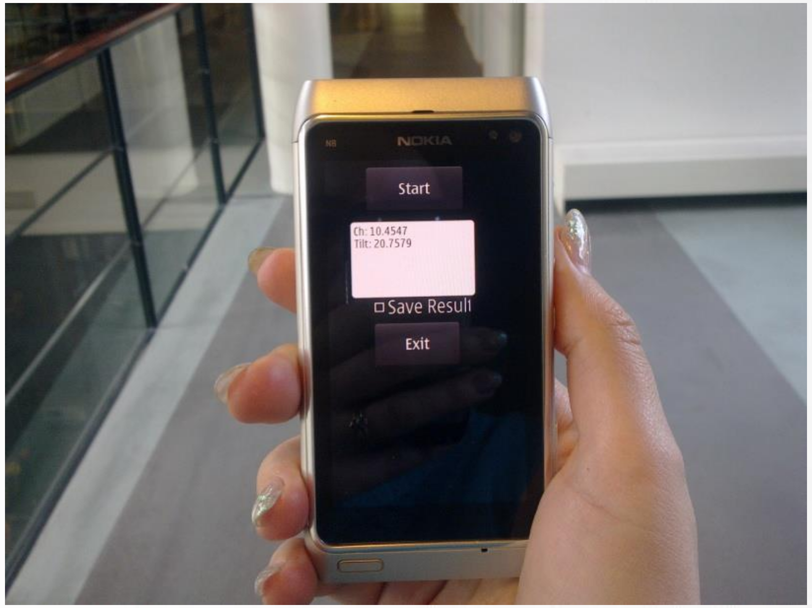
- Good quality "stereo camera" is the most important navigation tool of humans
- People veer when blindfolded => walking in circles and getting lost
- Humans use landmarks for navigation, also animals use vision

RELEVANT TERMS

- Photogrammetry: generating 2D or 3D model of the scene using images taken from different poses (or multiple cameras)
- Computer vision: algorithms retrieving understanding from the image
    - Machine vision: use of computer vision in industrial or practical processes
    - Structure from Motion (SFM) / Simultaneous Localization and Mapping (SLAM): camera's ego-motion and scene's structure
    - Visual odometry: ego-motion only

HISTORY AND MOTIVATION FOR VISION IN NAVIGATION

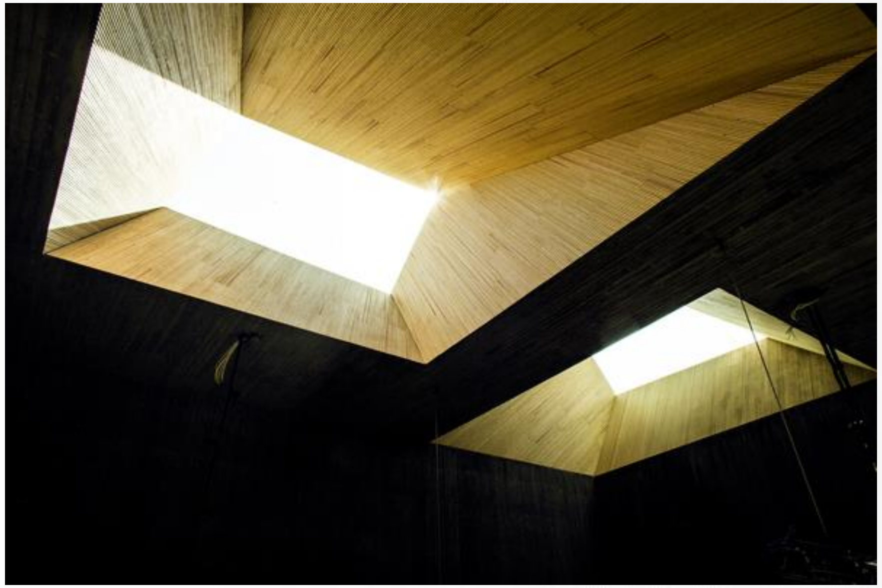
- Robots: SFM since 1981 (Longuet-Higgins), SLAM 1986 (Durrant-Whyte)
- Pedestrians
  - Databases 1999 (Aoki et al.)
  - Fused with other measurements 2003 (Kourogi et al.)



- Measurements very accurate with correct mechanization
- Not affected by radio-frequency interference
- Error sources different than for other positioning means

VISUAL PERCEPTION

- Objects in the scene are seen as sets of points of digitized brightness value functions
- Projective geometry, 3D => 2D
  - Shape, length, angle, distance, ratio of distance not preserved
  - Property of straightness preserved
  - Euclidean geometry has to be augmented with a point and line in infinity



VISUAL PERCEPTION

- An object point in the real world (X) is related to a point in image (x) by a camera matrix P

$$\mathbf{x} = \mathbf{P}\mathbf{X}$$

$$\mathbf{x} = (x, y, 1)$$

$$\mathbf{P} = \mathbf{K}\left[\mathbf{R}|\mathbf{t}\right]$$

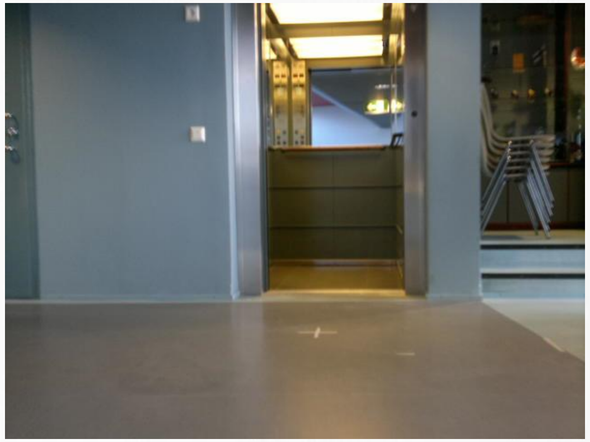
$$\mathbf{X} = (X, Y, Z, 1)$$

K = camera calibration matrix
R = camera's orientation
t = translation with respect to camera origin

Homogenous coordinates
x=fX/Z, y=fY/Z

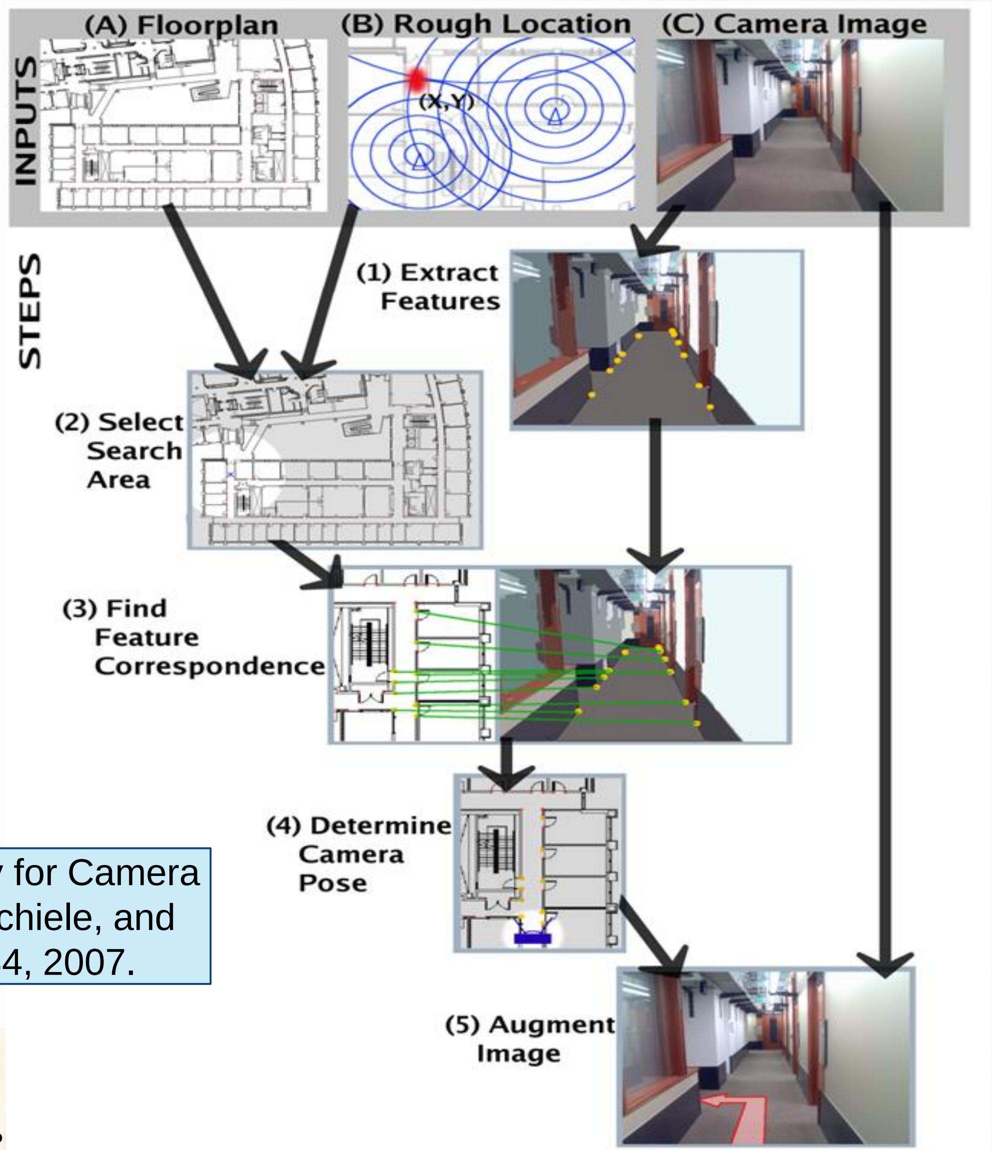
POSITIONING USING IMAGE DATABASES 1

- Database of images attached with position information (georeferencing)
- User takes images while navigating
- Images matched to the images in database
- When a match is found the position is known
- Rotation and displacement change the appearance of the image



POSITIONING USING IMAGE DATABASES 2

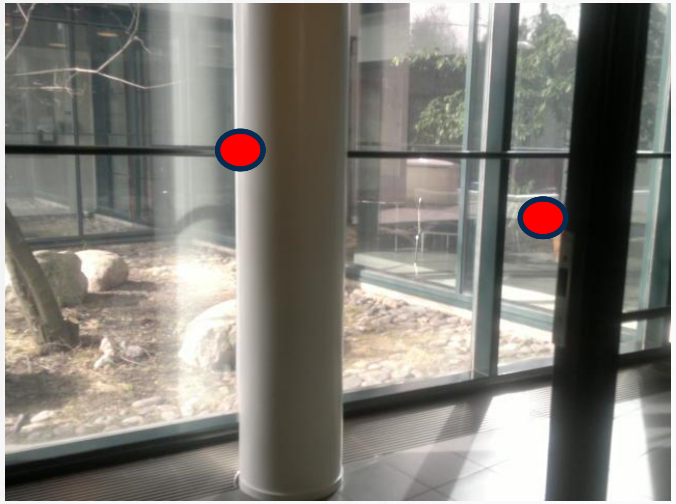
- Additional knowledge of the position makes the task easier
  - E.g. Wi-Fi based position



H. Hile and G. Borriello (2007), Information Overlay for Camera Phones in Indoor Environments. J. Hightower, B. Schiele, and T. Strang (Eds.): LoCA 2007, LNCS 4718, pp. 68–84, 2007.

MOTION INFORMATION FROM IMAGES

- Motion of the features representing the same real world object in consecutive images implies the motion of the camera, and thus the user when correctly mechanized



x

x’

$$\mathbf{x}' = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}\mathbf{x} + \mathbf{K}'\mathbf{t}/Z$$

Calibration matrix

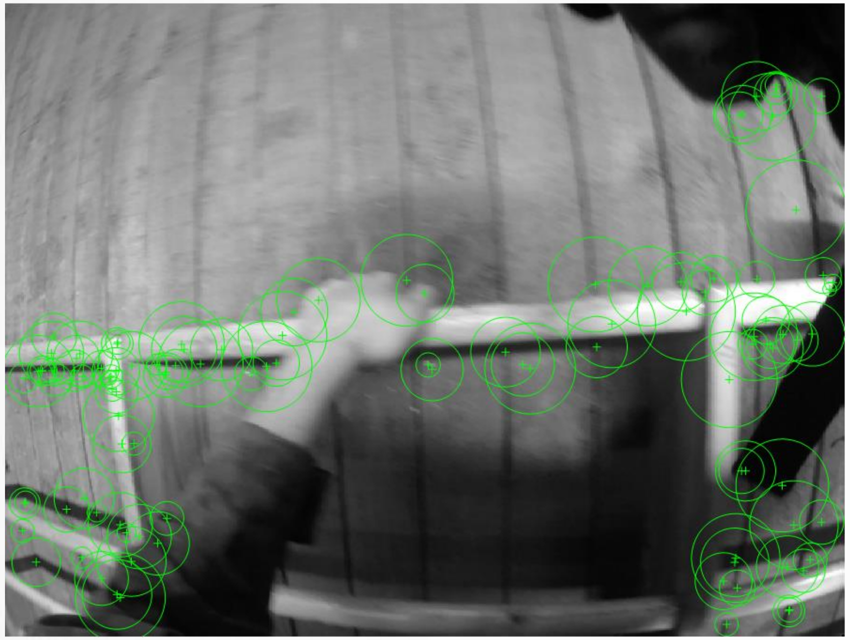
Rotation

Translation

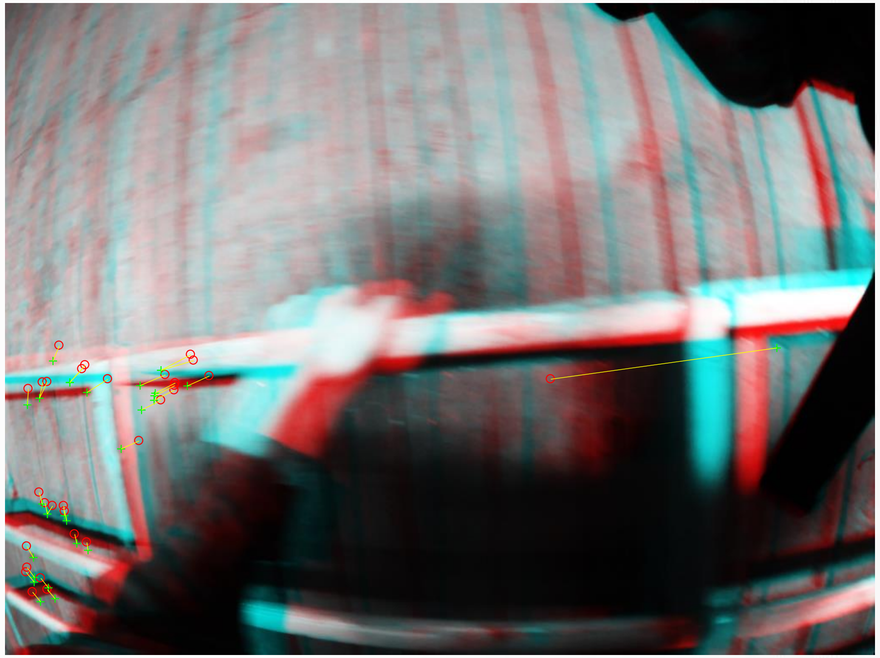
Depth

Hartley, Zisserman, 2003, Multiple view geometry in computer vision, 2nd ed. Cambridge.

FEATURE DETECTION AND MATCHING



Selection of representative features



Matching of the features between consecutive images

Error detection, RANSAC

SELECTING THE APPROACH

- Camera and method selection depends on the user
  - Vehicle: heavy and large equipment, constraint motion
  - Pedestrian: even body parts may have different motion patterns

- Feature detection depends on the environment
  - Indoors: feature poor, dark
  - Outdoors: large amount of features, brightness

Monocular camera
RGB-D
Stereo camera
LiDAR
Pedestrian
Vehicle

Lines
Corners
SIFT, SURF, FAST ORB
Indoors
Outdoors

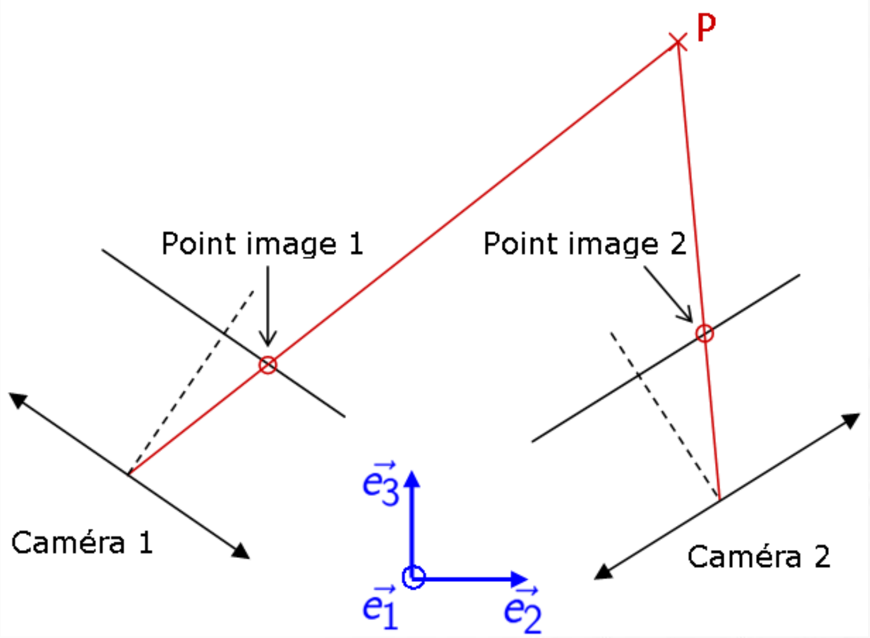
MONOCULAR CAMERAS AND SCALE AMBIGUITY

$$\mathbf{x}' = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}\mathbf{x} + \mathbf{K}'\mathbf{t}/Z$$

- Distance between the camera and object (Z)?
  => scale ambiguity in translation
- Solutions for solving depth (Z)
  - Estimated over time by tracking features
  - Objects with known size
  - Camera facing downwards with known height
  - LiDARs, RGB-D cameras => include Z
  - Stereo camera => can compute Z

STEREO CAMERAS

- Stereo camera (= two cameras with know orientation and distance from each other) solves the scale problem by triangulation
- If baseline is much shorter than the distance to the object being imaged degrades into monocular camera



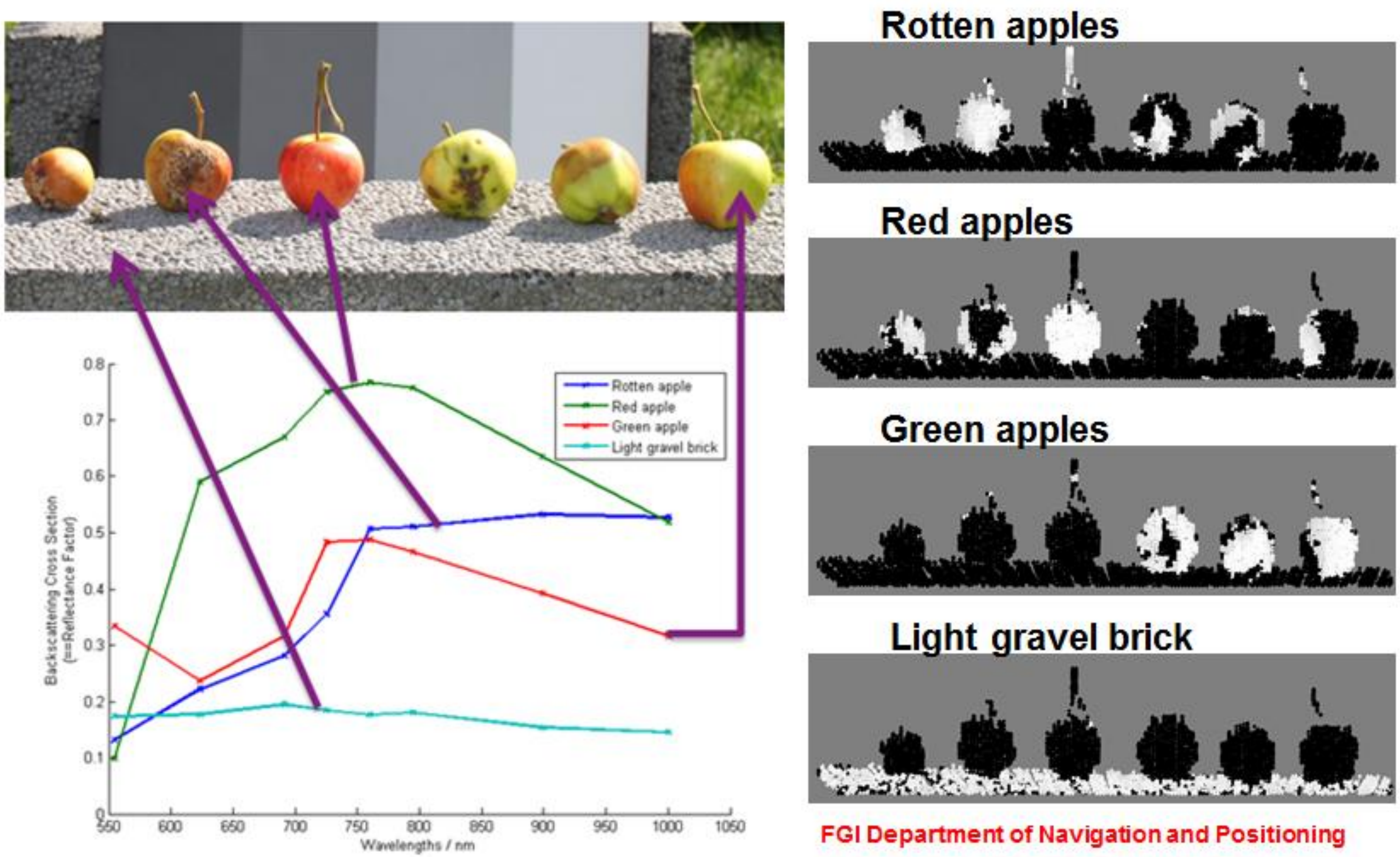
LIGHT DETECTION AND RANGING (LIDAR)

- Controlled steering of laser beams followed by a distance measurement
- Doesn’t require external light
- Traditionally expensive and large => technology evolving fast
- Materials of surroundings matter, e.g. glass lets the rays through
- Quite slow to process
- Very accurate (cm level)

https://www.flickr.com/photos/environment-agency/27489358013

HYPERSPECTRAL LIDAR

- Active hyperspectral imaging simultaneously with 3D topographic information
- Spectrum directly available for each point in the laser scanning point cloud
- Based on supercontinuum laser technology



Hakala et al. 2007. Full waveform hyperspectral LiDAR for terrestrial laser scanning. Optics Express Vol. 20 (7).

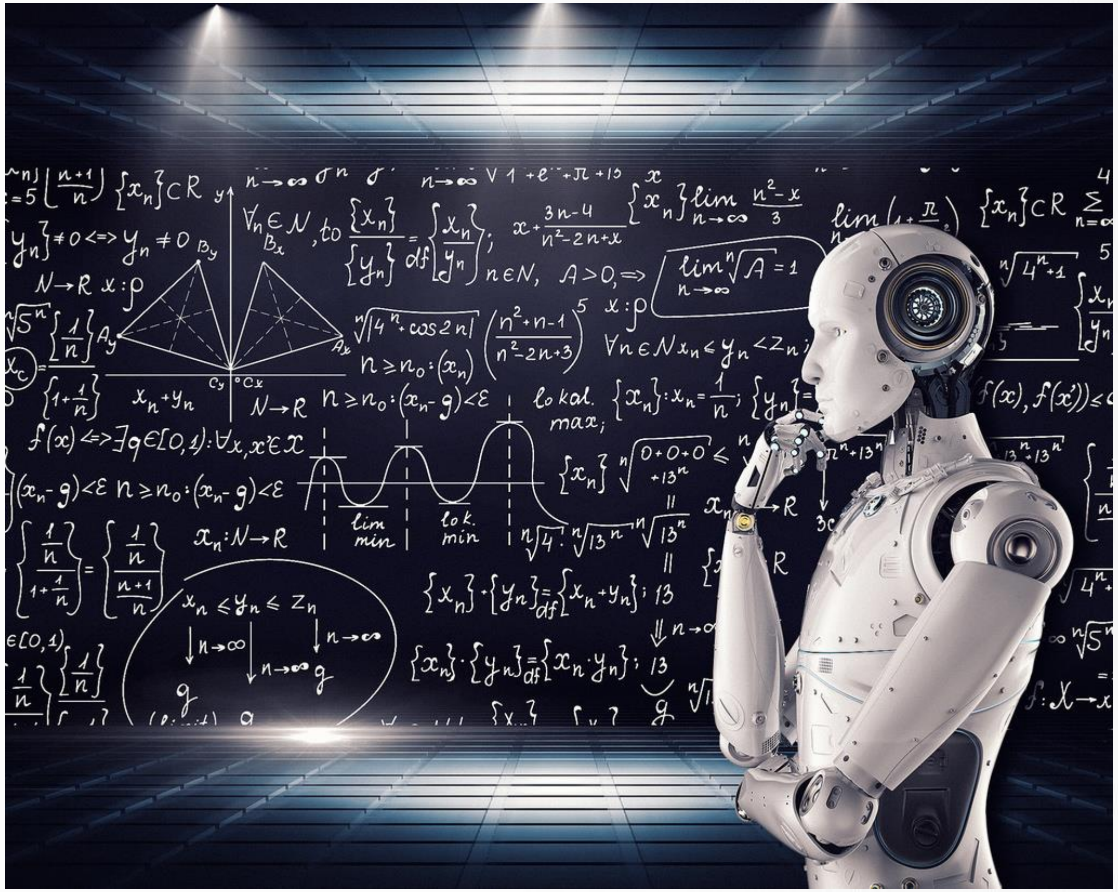
RGB-D CAMERAS

- Microsoft Kinect started the era of portable, consumer grade RGB-D systems (Red, Green, Blue, Distance) 2010
- Combine color information with per pixel depth infromation using infrared sensors
- 2015 Intel introduced RealSense family
  - Stereoscopic
  - Intel cameras have an active texture projector => image matching unambiguous

Microsoft Kinect, By Evan-Amos [Public domain], from Wikimedia Commons

MACHINE LEARNING IN VISUAL NAVIGATION

- Machine Learning = Collection of algorithms for computer systems that automatically improve their performance through experience
  - Feature detection
  - Image segmentation
  - Object / obstacle recognition
  - Tracking
  - SLAM, Convolutional Neural Networks for correcting global map and pose (Parisotto et al. 2018)



AUTONOMOUS DRIVING 1

- Vision is a key technology for automated driving (with GNSS)
- Different cameras
  - Monocular observing the environment; objects and **lane marks**
  - Stereo for observing pedestrians
  - Laser scanner for 3D object recognition

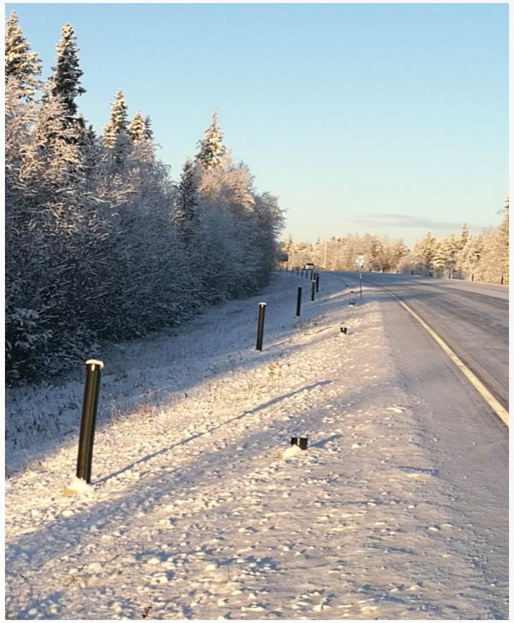
Camera x 4

Tri objective camera = stereo + monocular

camera

https://www.volvocars.com/intl/buy/explore/intellisafe/autonomous-driving

AUTONOMOUS DRIVING 2



- Arctic areas are challenging for vision
  - Snow
  - Darkness
- Also LiDAR suffers from reflections from snow

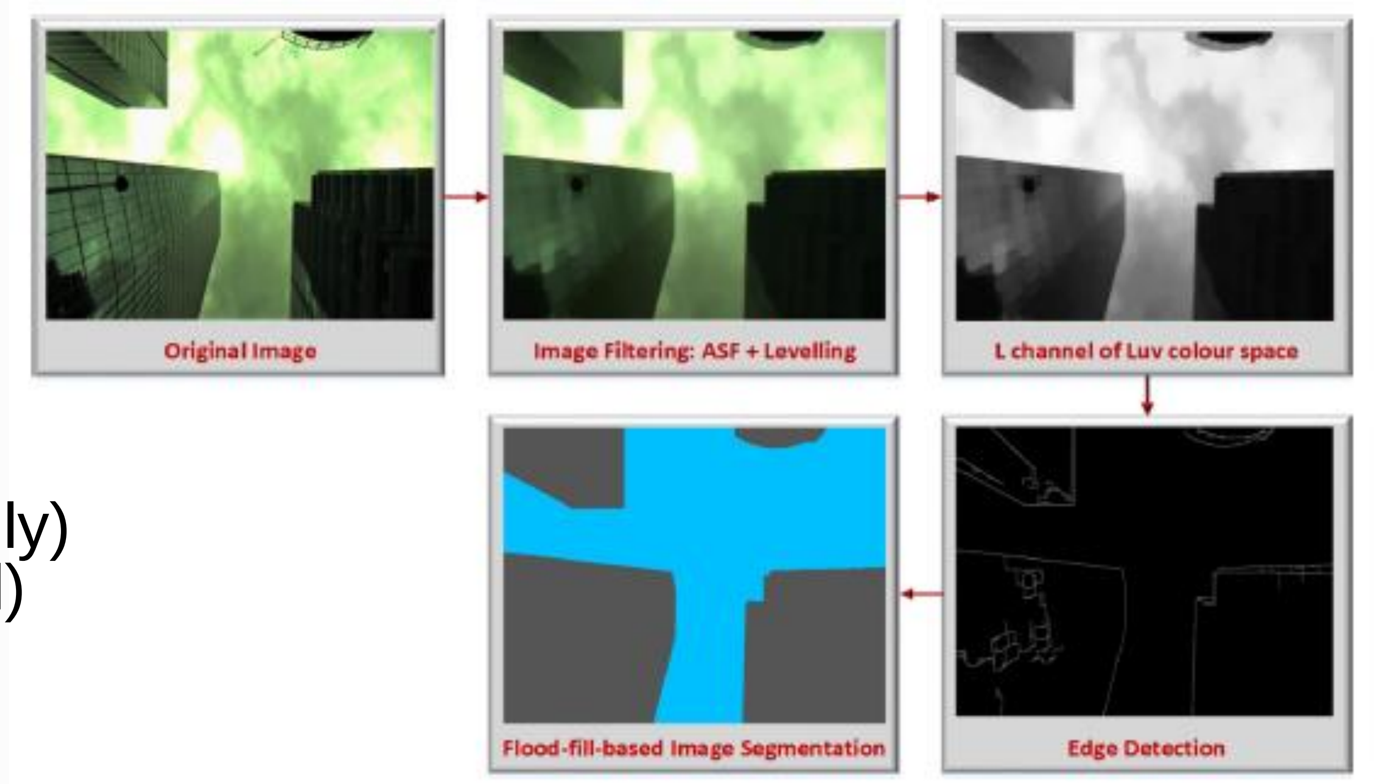
NLOS DETECTION IN URBAN NAVIGATION 1

- Estimating satellite visibility (Non-Line-Of-Sight signals) by classifying an image into sky and non-sky areas
- Camera on the roof of a vehicle
- Fisheye stereovision, positioning accuracy improved for one dataset from 10 m (LS) or 6.5 (EKF) to 3 m

Juliette Marais, Cyril Meurie. Quantify and improve GNSS quality of service in land transportation by using image processing. First CNES-ONERA Workshop on Earth-Space Propagation, Jan 2013, France. 5p, 2013.

NLOS DETECTION IN URBAN NAVIGATION 2

- Improved segmentation of the sky and non-sky
- Fusion of visual odometry, NLOS detection and a 3D city model
- Monocular camera, accuracy improved from 40 m (GNSS only) to 11 m (GNSS, VO, city model)



Gakne P. 2018. Improving the Accuracy of GNSS Receivers in Urban Canyons using an Upward-Facing Camera. Doctoral dissertation, University of Calgary, Canada

VISION IN INDOOR NAVIGATION

- Biggest challenge is lighting
- Short on features, surfaces poor of texture
- Moving objects
- Pedestrians complicate the task
  - Unrestricted motion
  - Requirement for small equipment

SLAM 1

- SLAM produces simultaneously
  - a map of the unknown environment
  - while positioning the user in this newfound map
- Solution is corrected continuously using loop-closure
- If positioning is accurate and reliable (GNSS in good conditions) SLAM is not needed => mapping only

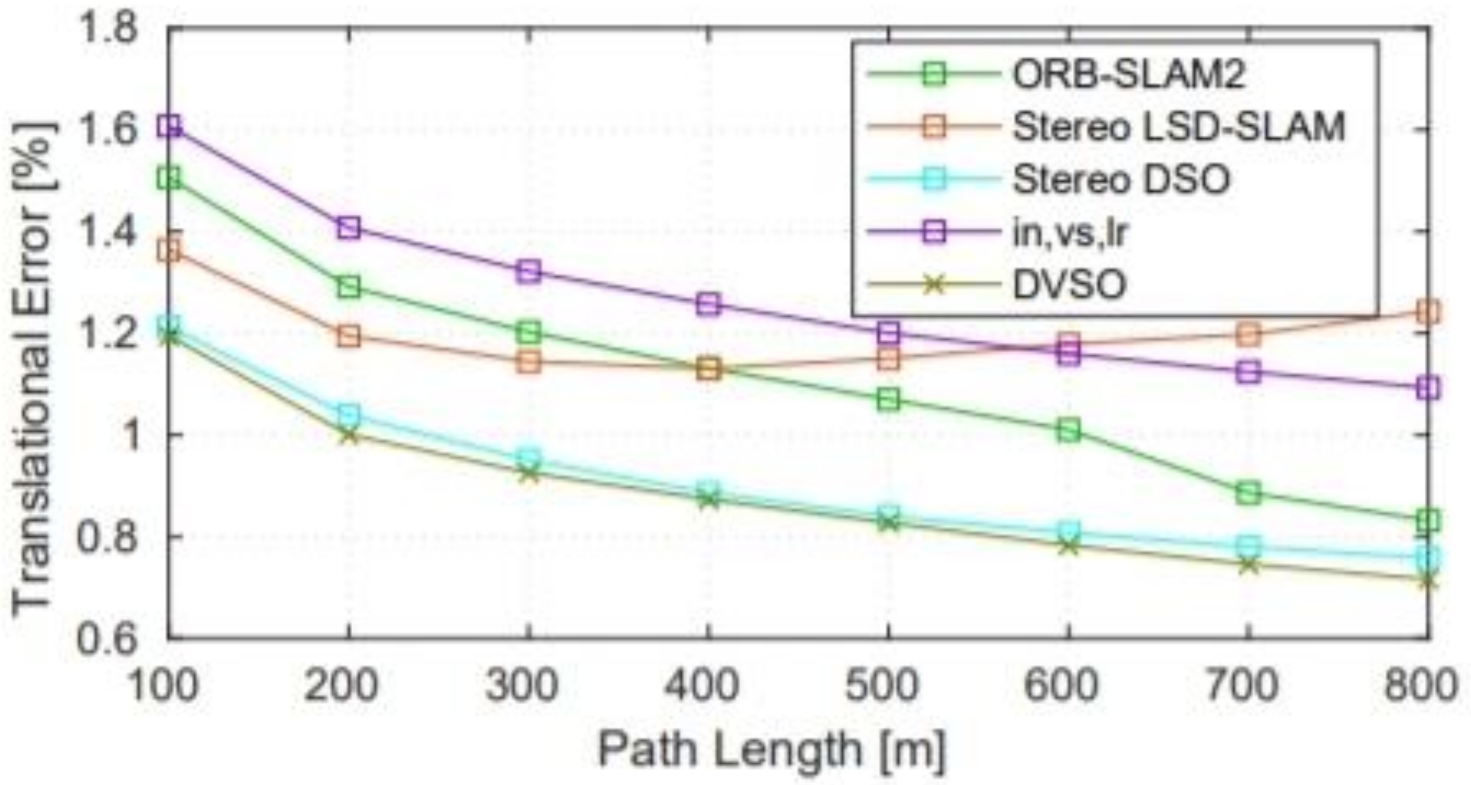
SLAM 2

- ORB - SLAM is the most sophisticated existing SLAM solution
- Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras http://webdiis.unizar.es/~raulmur/orbslam/
- Purely visual SLAM, real-time, also for handheld devices
- Trajectory error around 1%

Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós. (2015). ORB-SLAM: A Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics, vol. 31(5).

SLAM: DEEP VIRTUAL STEREO ODOMETRY DVSO

- Deep neural network
  for camera pose tracking (scale) and dense mapping
- Stereo disparity
- Training of classifiers with over 20000 images



N. Yang, R. Wang, J. Stuckler, D. Cremers (2018). Deep Virtual Stereo Odometry: Leveraging Deep Depth Prediction for Monocular Direct Sparse Odometry. The European Conference on Computer Vision (ECCV)

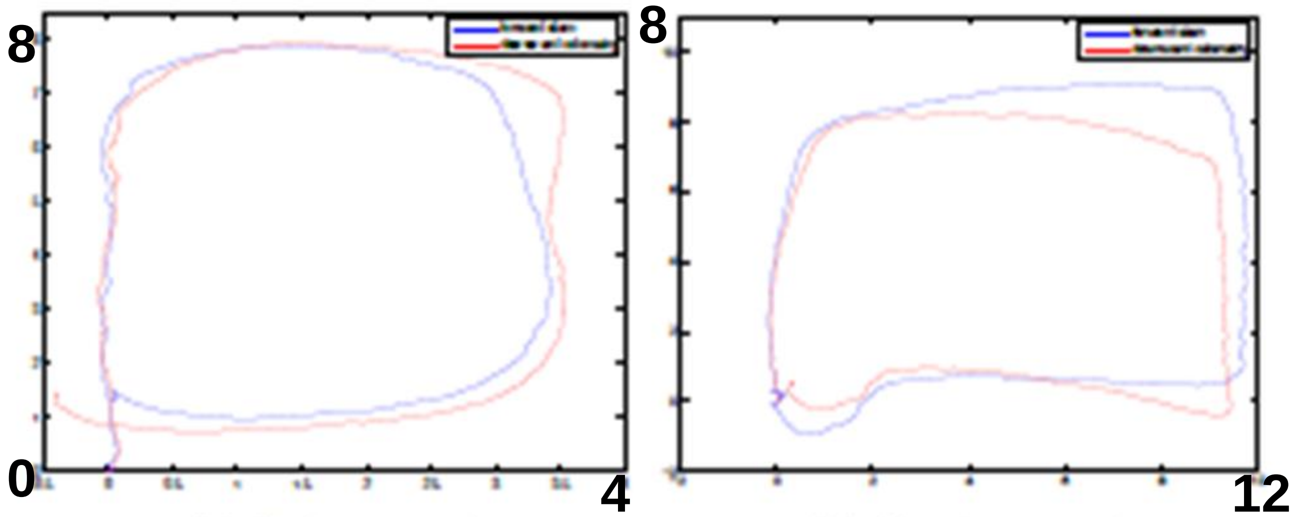
ARCTIC ROBOTICS

- Only company providing fully autonomous indoor UAV
- 2D Lidar-based navigation system using predefined routes
- Rotating Lidar to measure the distance to walls
- Sonars for obstacle avoidance
- Sonars and Lidars to measure distance to floor and ceiling

https://www.arcticrobotics.com

UNMANNED AERIAL VEHICLE (UAV)

- RealSense cameras (R200)
- ORB-SLAM
- 1 camera facing downwards (60 Hz) => velocity estimation
- 1 camera facing forwards (10 Hz) => position estimation with less drift, 1%



(a) Indoor result (b) Ourdoor result

Bi et al. 2016. An MAV Localization and Mapping System Based on Dual Realsense Cameras. Proceedings of IMAV.

SLAM – OPEN RESEARCH QUESTIONS

- Scale problem
- Time constraints
- Seamless SLAM
  - Different algortihms indoors and outdoors
- Loop closure problem
  - Strong appearance changes due to dynamic elements, illumination, weather or seasons
- Adaptiveness
  - relevant perceptual information, filter out irrelevant sensor data
  - map representation, complexity may vary depending on the task at hand

NAVIGATION / SLAM WITH MOBILE DEVICES

- Augmentd Reality (AR) and SLAM
- Google ARCore 3/2018
  - Project Tango, needed depth sensor
  - Fusion of inertial and vision
  - Motion, objects, lighting
  - https://developers.google.com/ar/
  - Android Nougat, iOS
- Apple ARkit

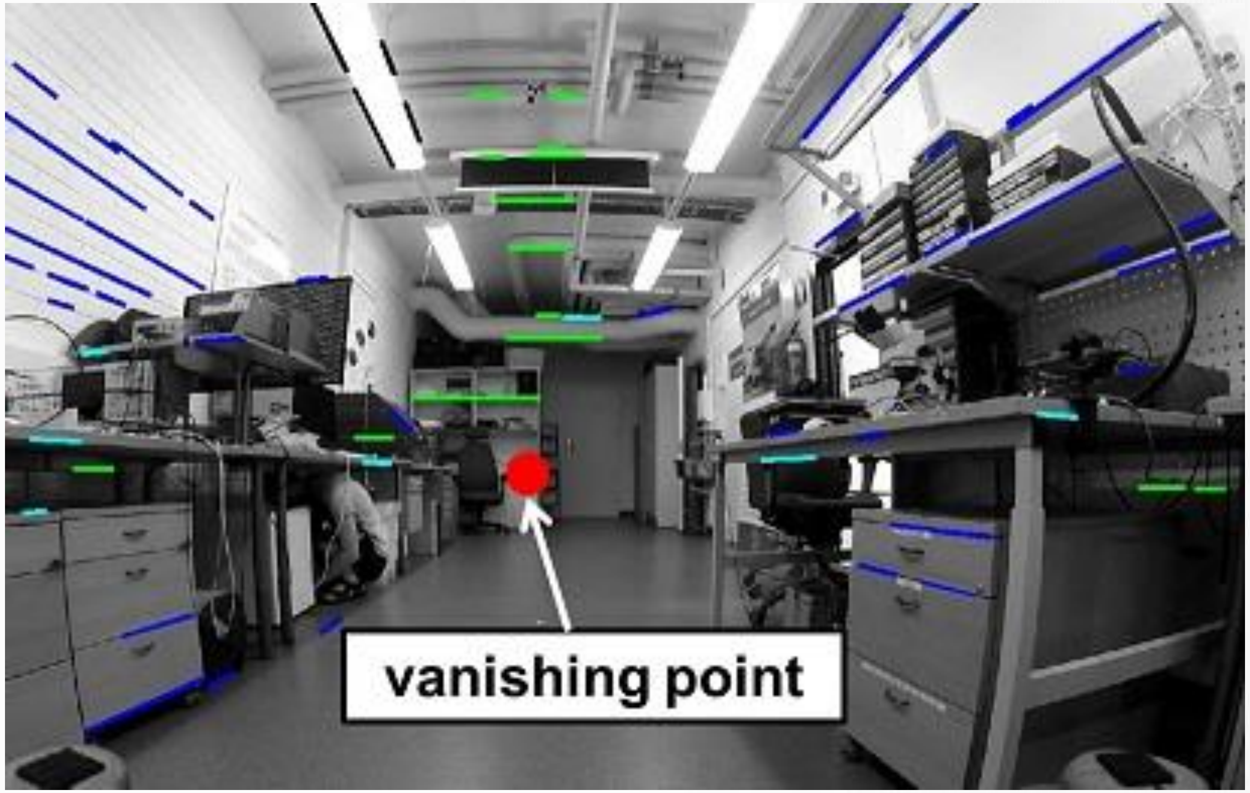
SOLUTION FOR CHALLENGING APPLICATIONS INDOORS

- Rescue and tactical applications include abnormal and rapid dynamics
- Indoor scenes are full of orthogonal lines
- Dynamic objects don’t disturb computation based on lines
- Lines may be detected despite darkness

VISUAL GYROSCOPE 1

- Vanishing point is a virtual point, where **parallel** lines seem to **intersect** in an image
- Locations of vanishing points depend on
  - Camera calibration (**K**)
  - Orientation of the camera (**R**)

$$[v_x\ v_y\ v_z] = KR$$



Ruotsalainen, Vision-aided navigation for pedestrians in GNSS challenging environments. Doctoral dissertation, Tampere University of Technology, Finland 2013

VISUAL ODOMETER 1

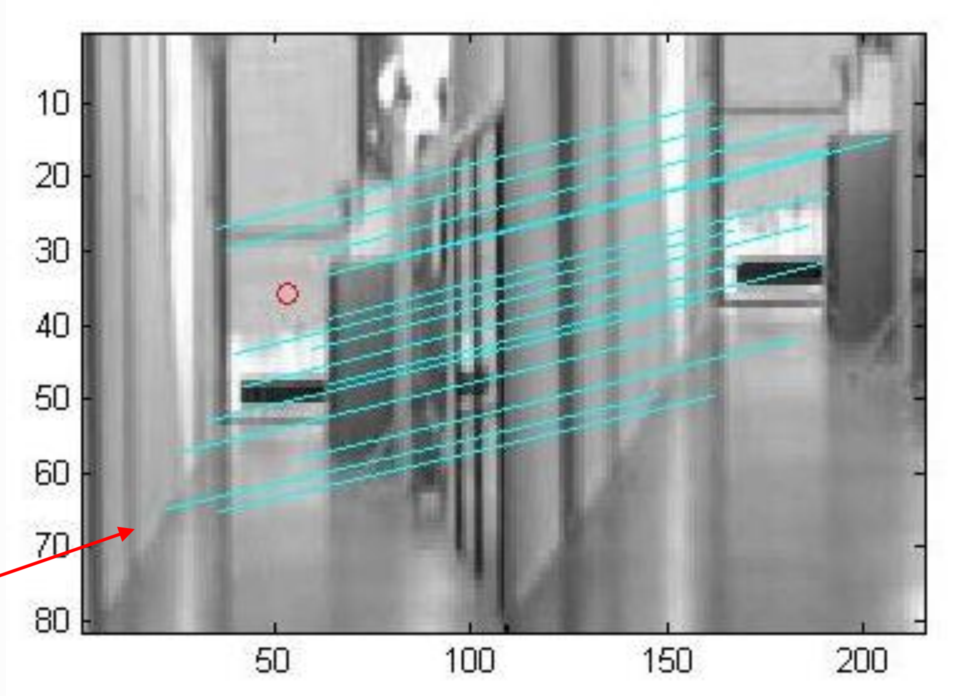
- Points are on plane (floor) => only one matching point needed (SIFT)

$$x' = K'RK^{-1}x + K't/Z$$

R rotation of the camera using visual gyroscope

t translation of the camera

Z distance of point?

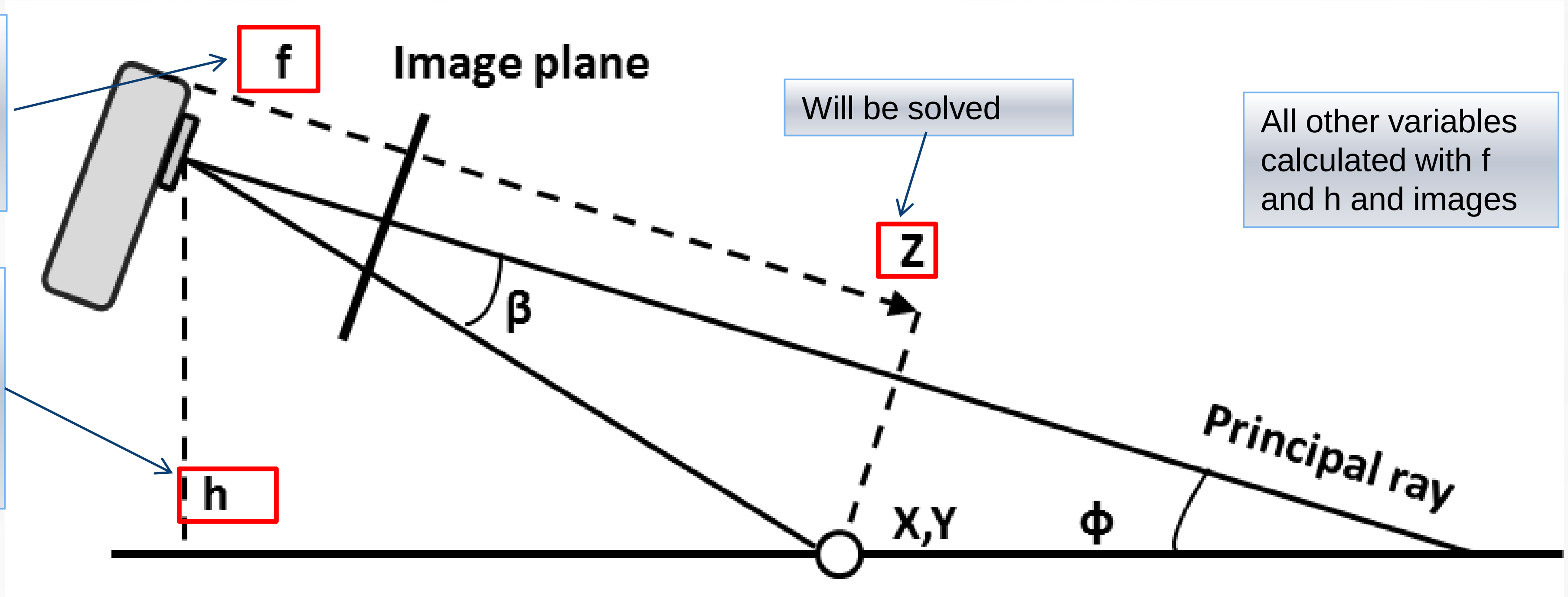


Features mainly found from the edge of floor and wall

VISUAL ODOMETER 2

From calibration or image header

Has to be known in advance and kept static



Will be solved

All other variables calculated with f and h and images

- Z from basic geometry

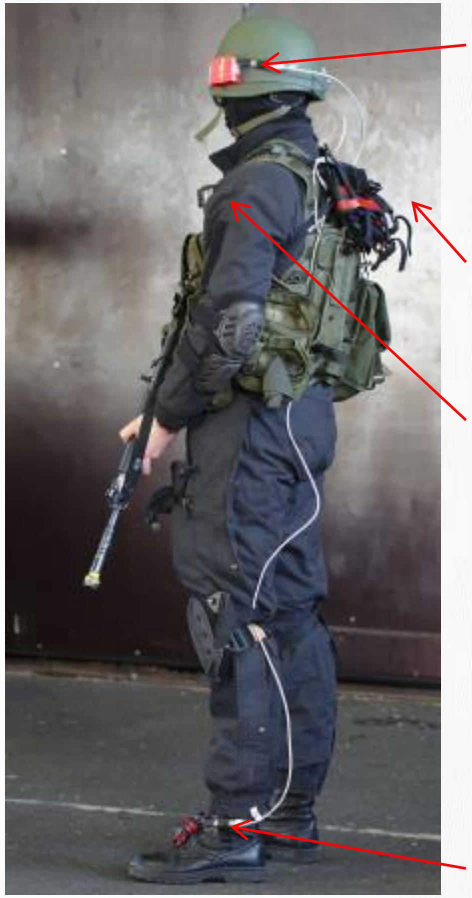
Ruotsalainen, 2013

INFRASTRUCTURE-FREE INDOOR NAVIGATION 1

- **INTACT-** a project funded by the Finnish Ministry of Defence 2015 - 2017
- Needs:
  - Infrastructure-free indoor positioning
  - A rough floor plan
  - Additional information about the environment
  - Motion and other context information
- For unknown environments => Infrastructure-free: using only sensors attached to the user

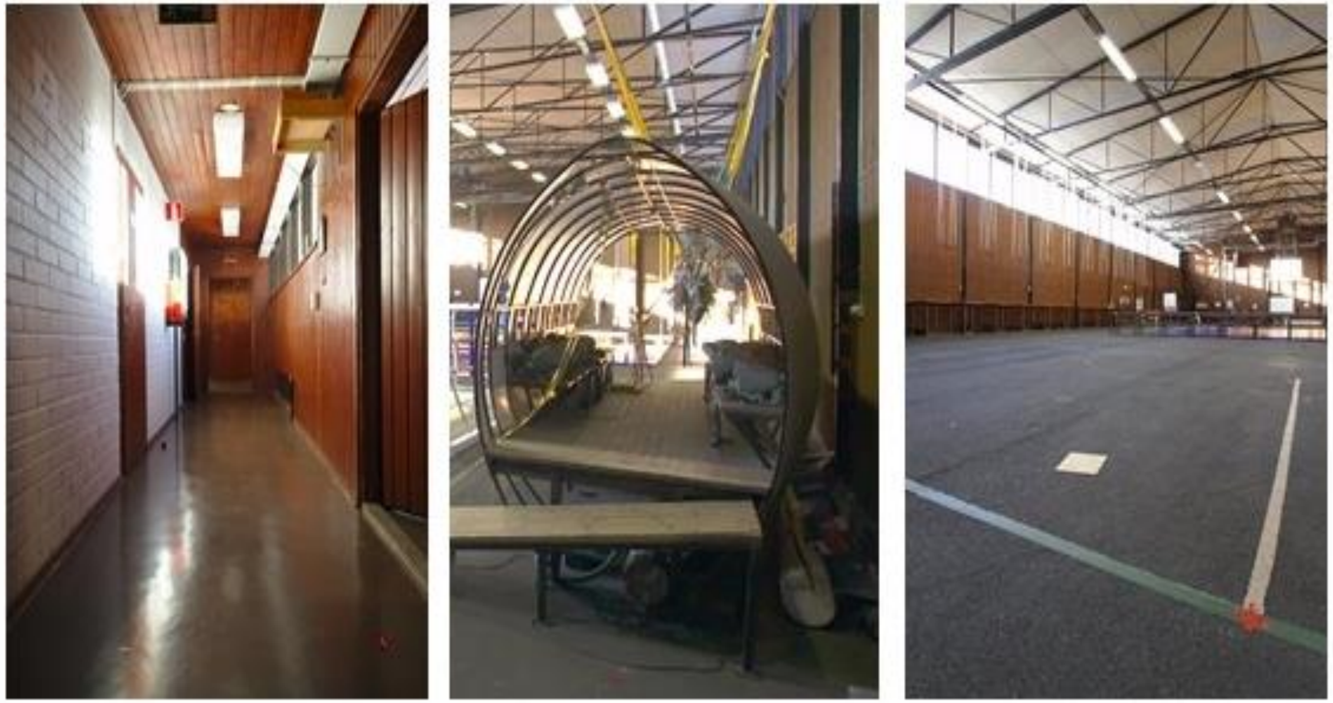
INFRASTRUCTURE-FREE INDOOR NAVIGATION 2

- Horizontal positioning using foot-mounted IMU and monocular camera
- Machine learning for modelling user's motion
  - When the user is crawling or climbing camera is not used (Rantanen et al, IPIN 2018)
- Error modelling and particle filtering
- Proof-of-concept in military premises by soldiers



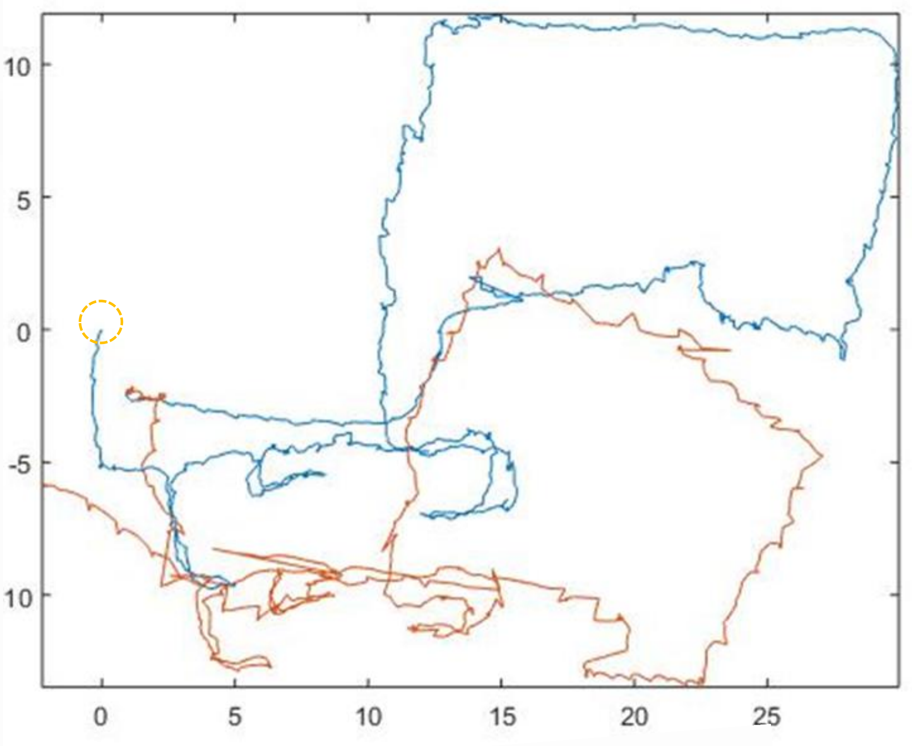
INFRASTRUCTURE-FREE INDOOR NAVIGATION 3

- Tests contained challenging motion
  - Running
  - Jumping
  - Climbing
  - Walking sideways in wall bars



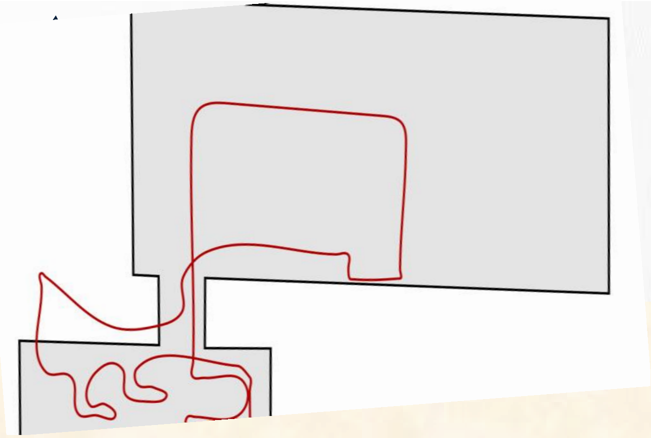
INFRASTRUCTURE-FREE INDOOR NAVIGATION 4

- 2 rounds
  - walking (blue)
  - running (red)



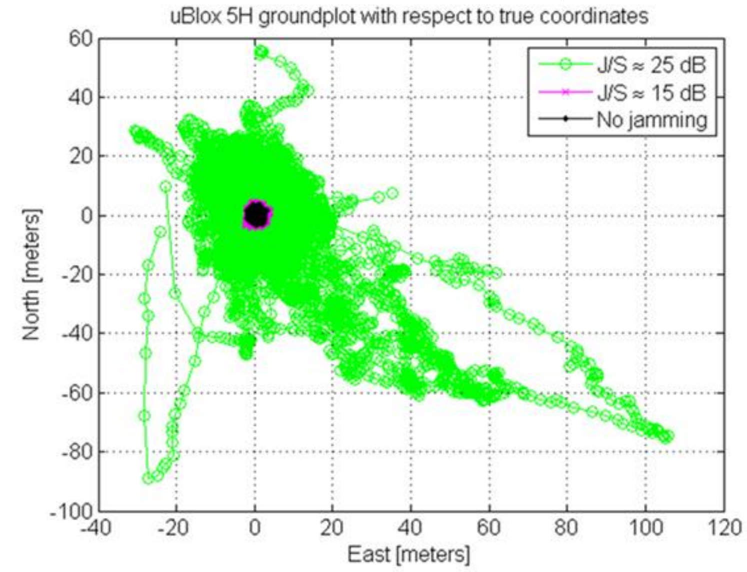
- Errors in loop-closure
  - 1st round (200 m) = 2.5 m
  - 2nd round (400 m) = 5.2 m

- NATO Science for Peace and Security Program: Collaborative Augmented Navigation for Defence Objectives 2018-2019 with Sintef

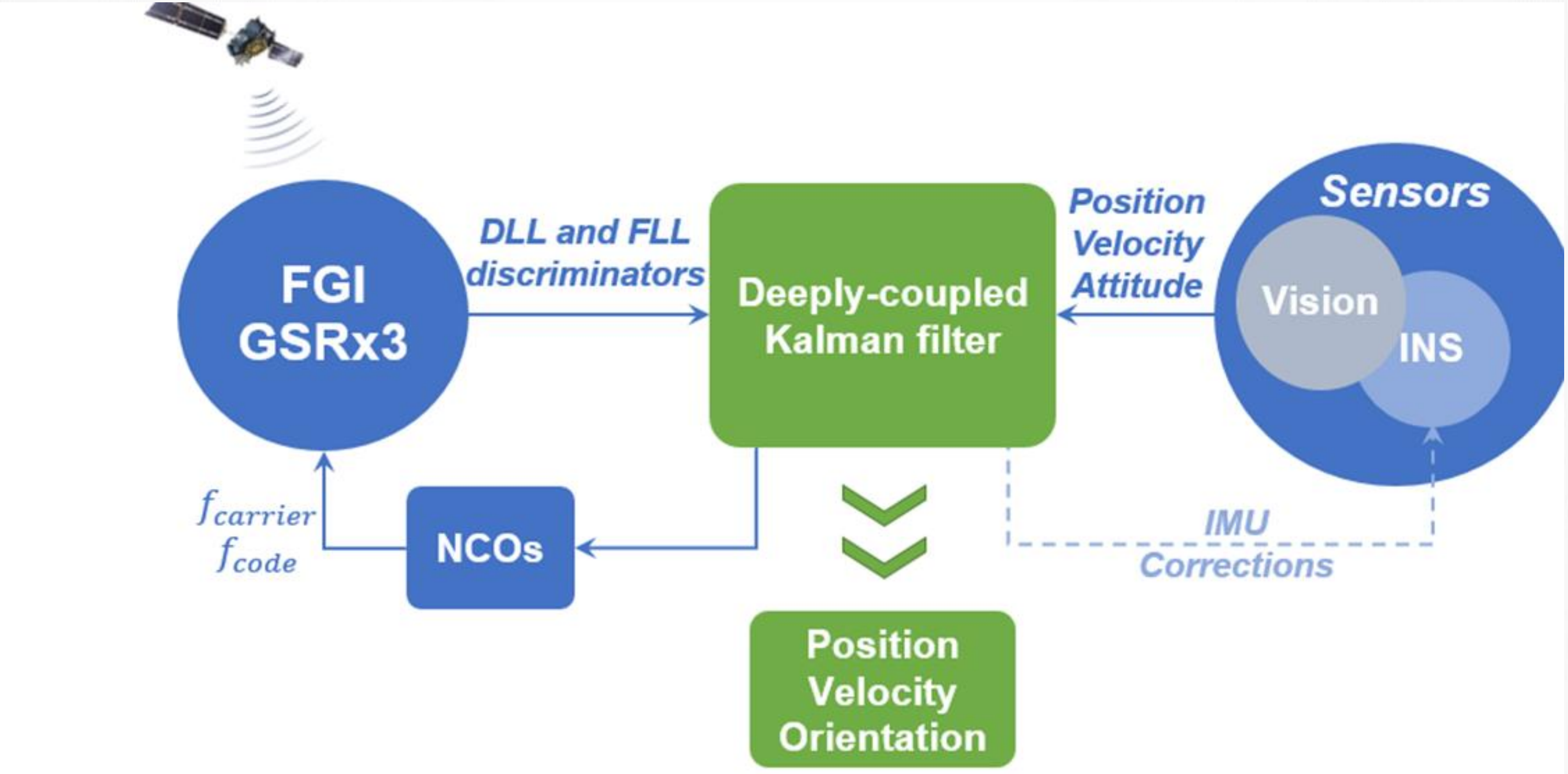


DELIBERATE GNSS INTERFERENCE

- **Jamming**: transmission of signals at GNSS frequencies
  - Deteriorates or denies GNSS position
  - Illegal in most countries
  - ”Personal Privacy Devices”
- **Spoofing**: transmission of fake GNSS signals
  - Deludes the receiver to be in wrong position
- Effects
  - Small nuisance
  - Economic impact
  - Safety impact



DEEPLY-COUPLED GNSS / INS / VISION 1



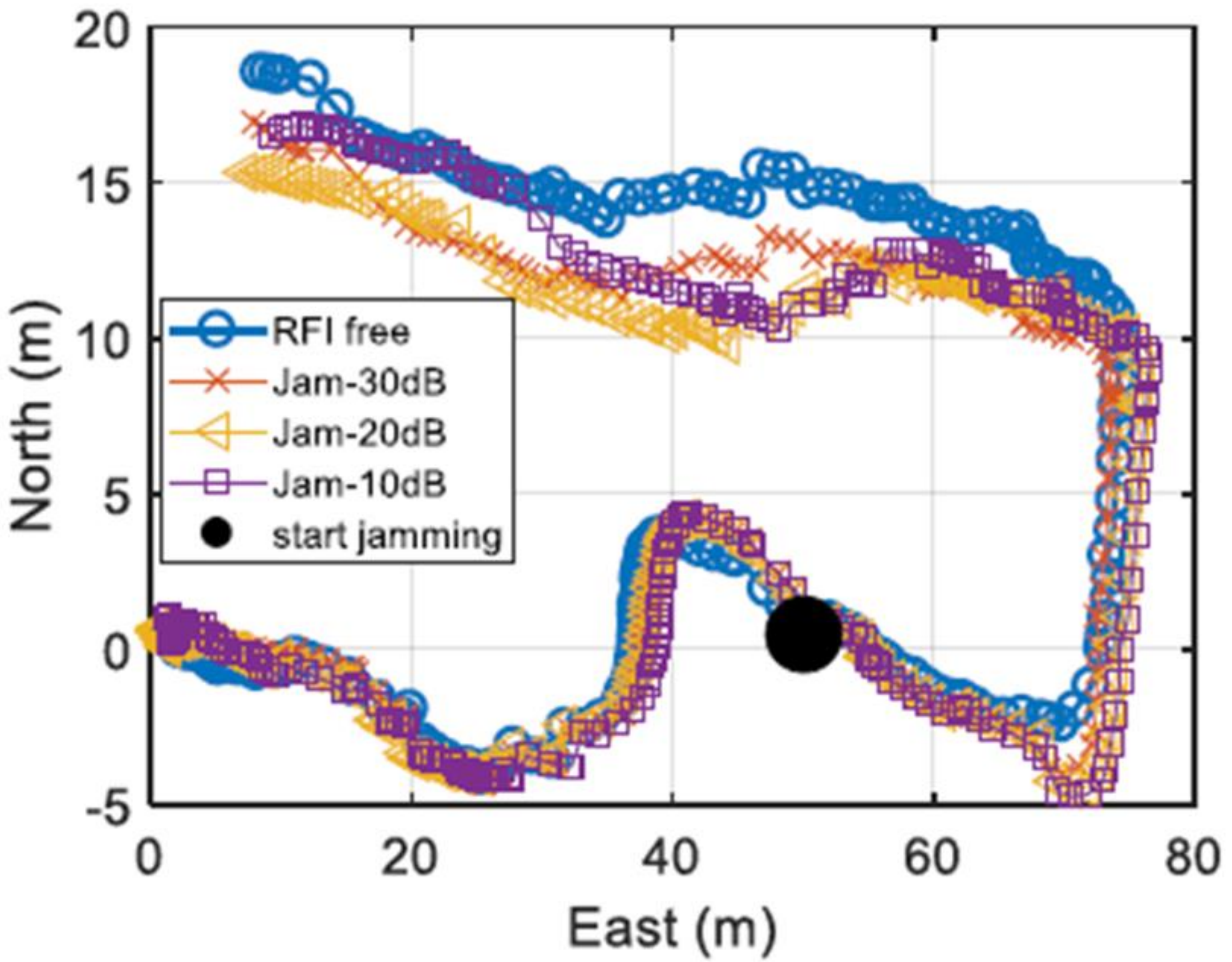
DEEPLY-COUPLED GNSS / INS / VISION 2

- GPS + Galileo signals used for processing
- Inertial measurements Xsens MEMS IMU
- GoPro 5 Session camera, 1 HZ
- GNSS signals interfered by jamming using Record and Playback method
- Data collected in a dynamic test, 10 minutes walk around a parking area in Torino, Italy

Ruotsalainen et al, 2014. Deeply coupled GNSS, INS and visual sensor integration for interference mitigation. In Proceedings of ION GNSS+

DEEPLY-COUPLED GNSS / INS / VISION 2

Indoor GNSS?



There was no position solution available when using GNSS alone and jamming attenuated by 10dB

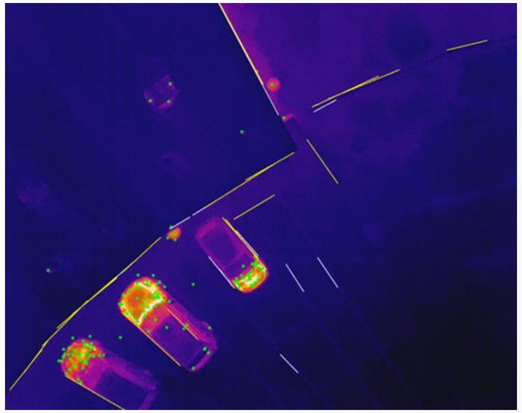
Cristodaro et al. 2018. GNSS/INS/visual Deeply Coupled Integration: Preliminary Investigation on Dynamic Jammed Datasets. In Proceedings of ION GNSS+

DATA SOURCES AND TOOLS

- OpenCV: https://opencv.org/
  - Main algortihms in C++
- KITTI: http://www.cvlibs.net/datasets/kitti/
  - Data repository and benchmarking
- ORB-SLAM: http://webdiis.unizar.es/~raulmur/orbslam/
  - Source codes and example videos
- Intel: https://github.com/IntelRealSense/librealsense
  - Codes and documents
- Matlab’s toolboxes

CONCLUSIONS

- Vision is enabler for more accurate, available and reliable navigation
- No single method is feasible for all cases => adaptive systems
- Needs for future research
  - Solving the scale problem
  - Error detection methods
  - Scalable SLAM
  - Snow- and situation-aware algorithms
  - Low-cost solutions for darkness




60˚ 10 1.2 N, 24˚ 57 18 E