Universidade Federal do Paraná

Especialização em Inteligência Artificial Aplicada

# Mobile Robotics

## Simultaneous Localization and Mapping

## SLAM II

Prof. Eduardo Todt

2019

# SLAM

Simultaneous Localization and Mapping

Convert imprecise raw sensor data into map and localization

The magic: due to multi-view geometry the coherent map is more precise than individual sensor measurements

# Handcrafted detectors/ descriptors

SIFT - Scale Invariant Feature Transform (Lowe, 2004)

SURF - Speed up Robust Feature (Bay, Tuytelaars and Van Gool, 2008)

BRIEF - Binary Robust Independent Elementary Features (Calonder, Lepetit, Strecha, and Fua, 2010)

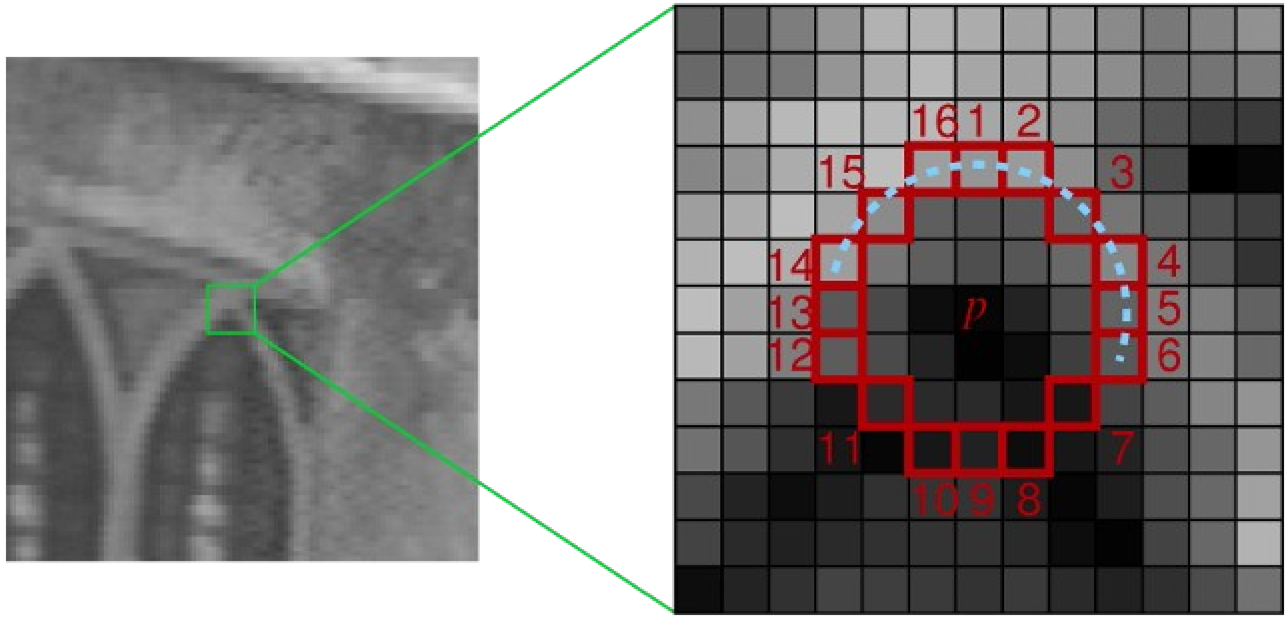
ORB - Oriented FAST and Rotated BRIEF (Rublee, Rabaud, Konolige and Bradsk, 2011)

# FAST detector

Features from accelerated segment test (FAST)

Corner detection method

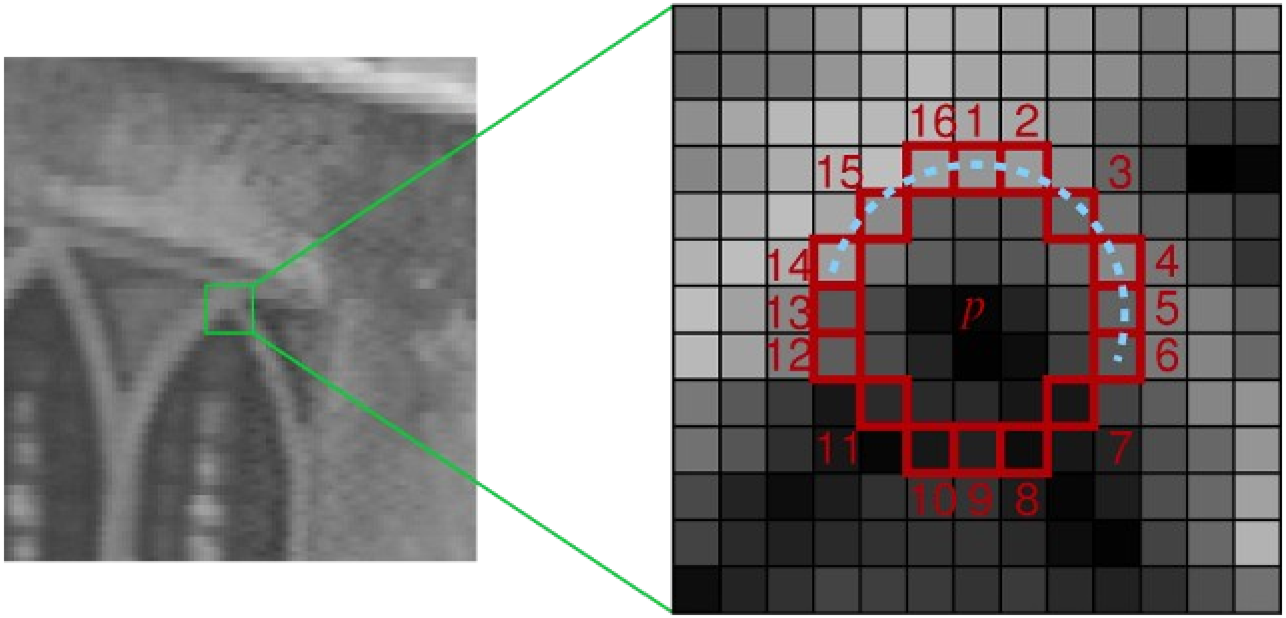
Edward Rosten and Tom Drummond, 2006



# FAST detector

Each pixel in the Bresenham circle of radius 3 is labeled (1 to 16 clockwise).

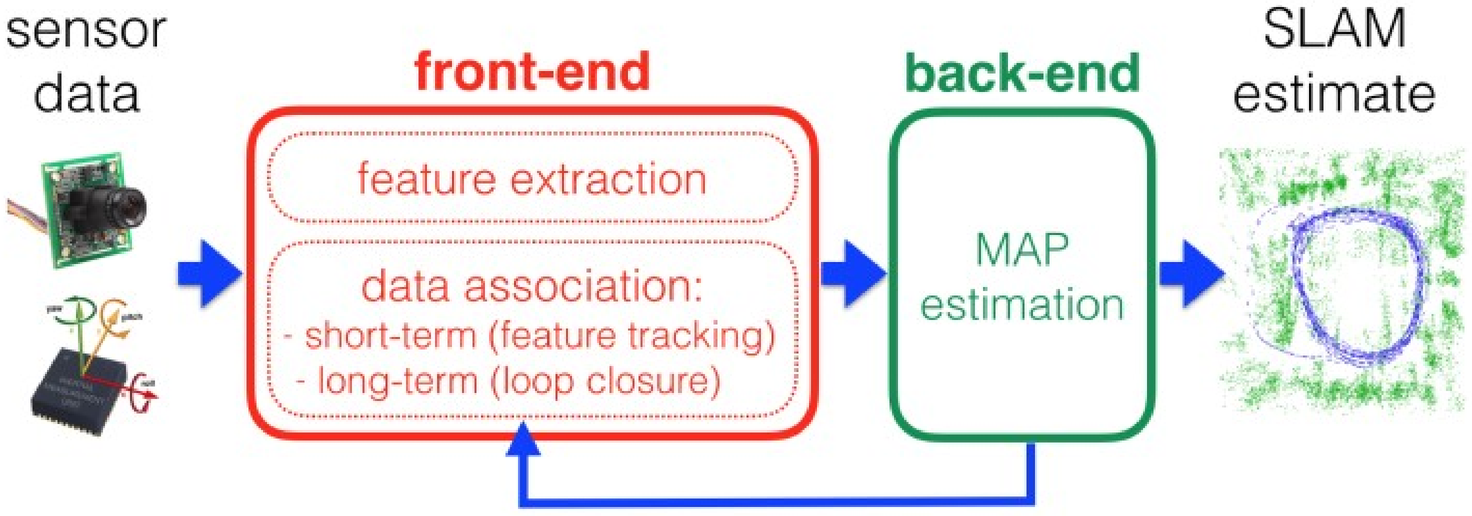
If a set of N contiguous pixels in the circle are all brighter than the intensity of candidate pixel p plus a threshold value t or all darker than the intensity of candidate pixel p minus threshold value t, then p is classified as corner. Typical N is 12.



# Modern SLAM

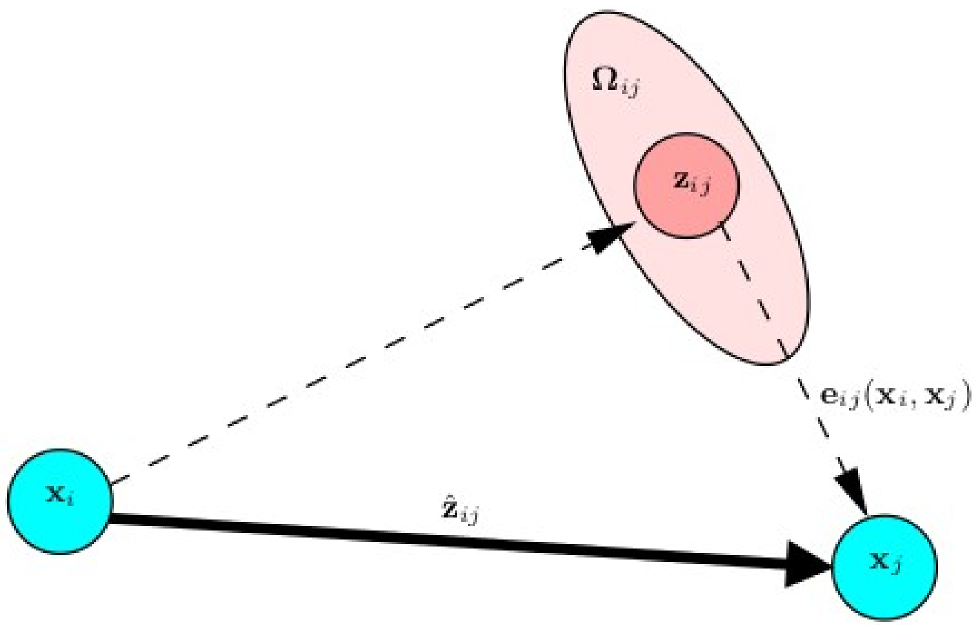
- Graph based
  - The vertices in the graph are entities that we want to estimate (robot positions, points in the world)
  - Edges represent constraints between these entities which are derived from raw sensor data

# Graph SLAM




Cadena et al. 2016

# Graph SLAM




Aspects of an edge connecting the vertex $x_i$ and the vertex $x_j$.

This edge originates from the measurement $z_{ij}$.

From the relative position of the two nodes, it is possible to compute the expected measurement $\hat{z}_{ij}$ that represents $x_j$ seen in the frame of $x_i$.

The error $e_{ij}(x_i, x_j)$ depends on the displacement between the expected and the real measurement.

An edge is fully characterized by its error function $e_{ij}(x_i, x_j)$ and by the information matrix $\Omega_{ij}$ of the measurement that accounts for its uncertainty.

Grisetti et al. 2010

# Deep Learning and SLAM

Deep learning relies on training data ...

What about unknown environments?

# DF-SLAM

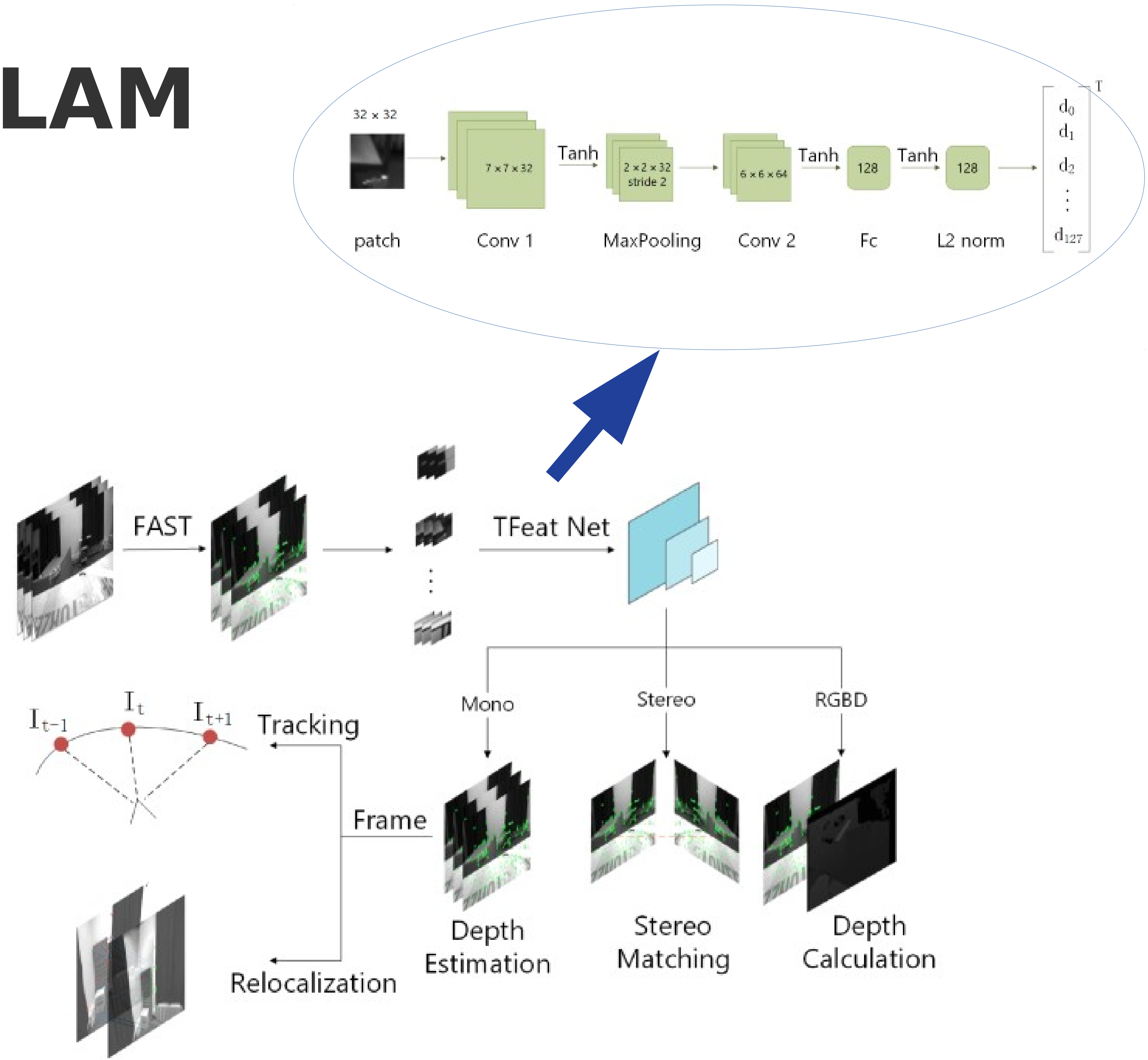
Deep local features SLAM

improve robustness of local feature descriptor with DL

learned local feature descriptors

Kang et al. 2019

# DF-SLAM



32 x 32
7 x 7 x 32
Tanh
2 x 2 x 32
stride 2
6 x 6 x 64
Tanh
128
Tanh
128
patch
Conv 1
MaxPooling
Conv 2
Fc
L2 norm
FAST
TFeat Net
Mono
Stereo
RGBD
Tracking
Frame
Relocalization
Depth Estimation
Stereo Matching
Depth Calculation

# SLAM with dynamic object detection and removal

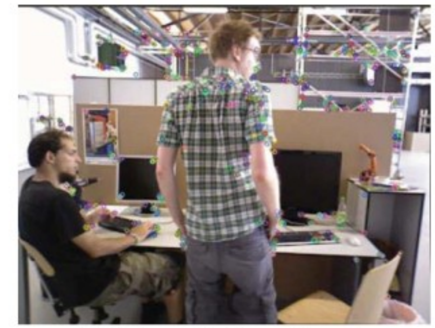
Dynamic object detection with Yolo v3

- Yolo works on entire image, thus it is fast for object detection (1000 times faster than R-CNN and 100 times faster than Faster R-CNN)
- It adopts Feature Pyramid Networks to achieve multi-scale detection

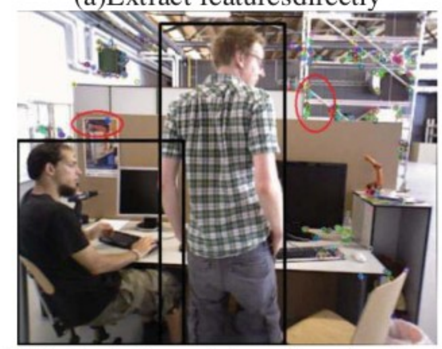
Dynamic POI removal

- It should be taken into account that dynamic objects don't move significantly among successive frames

Zhang et al., 2019



(a)Extract featuresdirectly



(b)Delete feature points on dynamic objects

(c)Extract feature points after unifying gray values

Fig.5 Feature extraction excluding dynamic object

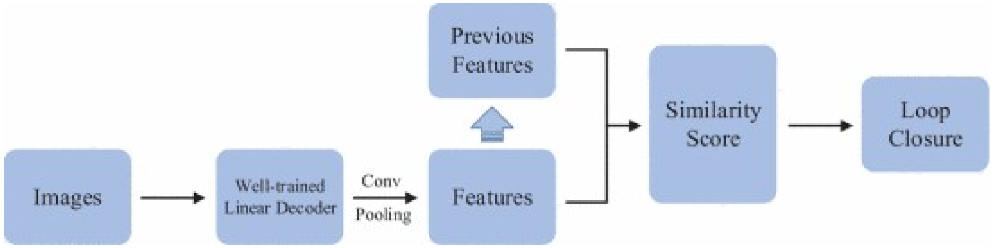
# Loop Closure Detection for Visual SLAM Based on Deep Learning

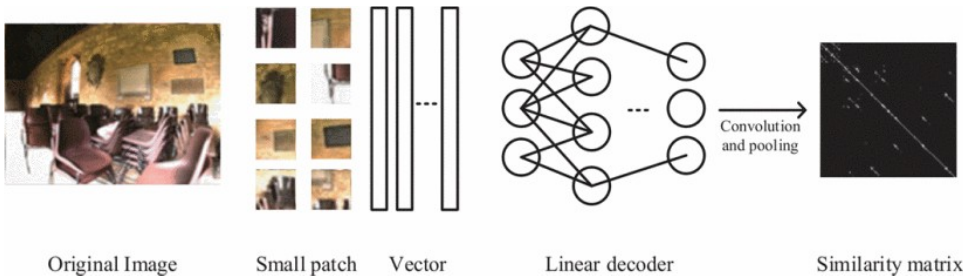
Loop Closure detection is a key link in SLAM, which means that the mobile robot judges whether the current location has been visited.

The closed loop can effectively reduce the accumulated error of the robot pose estimation

Inaccurate loop closure information can affect the SLAM backend graph optimization, which incorrectly modifies the generated map, resulting in map building and localization errors.

Hu et al., 2017



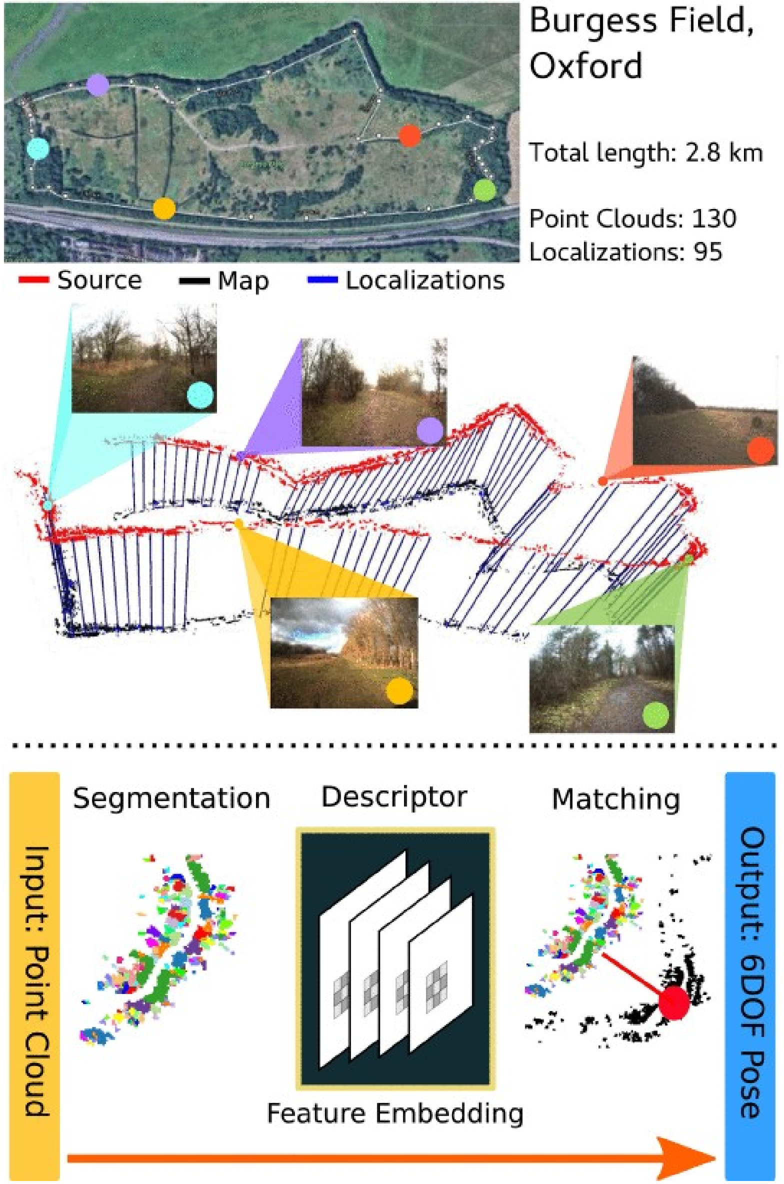
# Learning to See the Wood for the Trees: Deep Laser Localization in Urban and Natural Environments on a CPU

Deep learning approach to learn descriptors directly from 3D point clouds

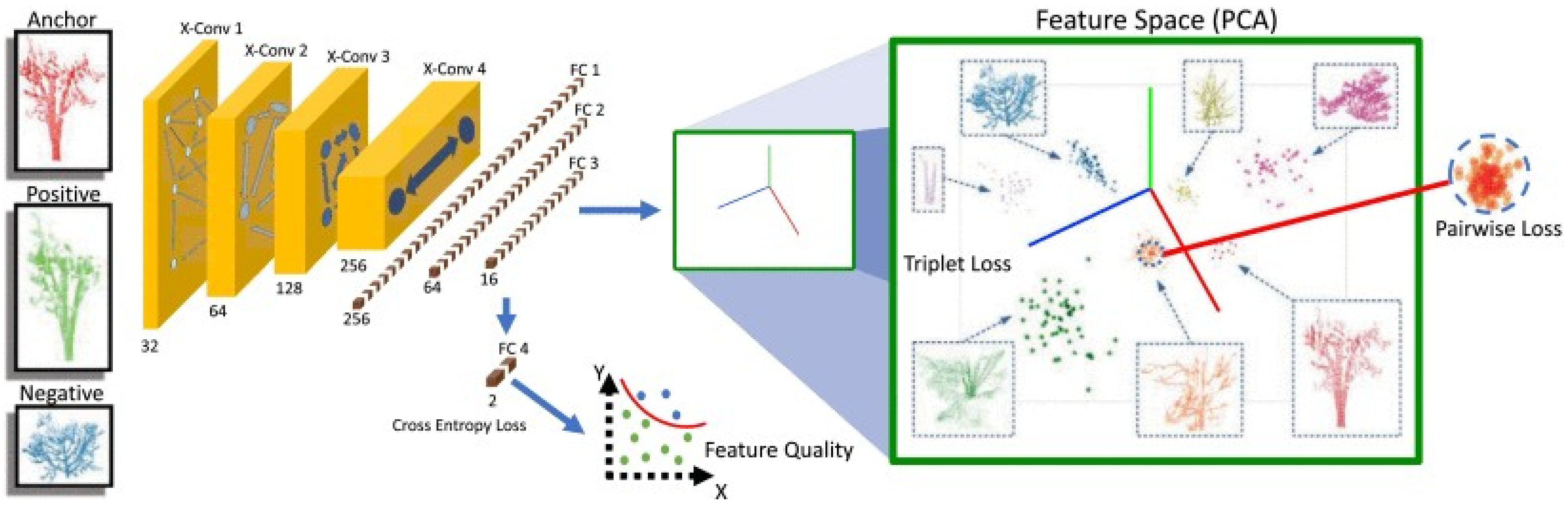
It compares triplets: anchor, positive, and negative examples

It learns a feature space representation for a set of segmented point clouds that are matched between a current and previous observations.

Tinchev et al., 2019



# Learning to See the Wood for the Trees: Deep Laser Localization in Urban and Natural Environments on a CPU




Network architecture used for learning the feature embeddings (left) and a visualization of the embedding space after training (right).
The network takes as input raw point data, it uses four x-convolutional layers (X-conv) and three fully connected (FC) layers to estimate the feature space.
A feature quality branch is also created by appending a fully connected layer to the feature embedding.
The networks learns to cluster similar objects (denoted in the same colour) close together while separating dissimilar objects.
The final feature descriptor is trained using a combination of triplet and pairwise losses.

Cadena C, Carlone L, Carrillo H, Latif Y, Scaramuzza D, Neira J, Reid I, Leonard JJ. “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age” IEEE Transactions on robotics. 2016 Dec ; 32(6):1309–32.

Grisetti G, Kummerle R, Stachniss C, Burgard W. “A tutorial on graph-based SLAM”. IEEE Intelligent Transportation Systems Magazine. 2010;2(4):31–43.

Kang R, Shi J, Li X, Liu Y and Liu X. “DF-SLAM: A Deep-Learning Enhanced Visual SLAM System based on Deep Local Features”. 2019;

P. Li, G. Zhang, J. Zhou, R. Yao, X. Zhang and J. Zhou, "Study on Slam Algorithm Based on Object Detection in Dynamic Scene," 2019 International Conference on Advanced Mechatronic Systems (ICAMechS), Kusatsu, Shiga, Japan, 2019, pp. 363-367. doi: 10.1109/ICAMechS.2019.8861669

H. Hu, Y. Zhang, Q. Duan, M. Hu and L. Pang, "Loop Closure Detection for Visual SLAM Based on Deep Learning," 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Honolulu, HI, 2017, pp. 1214-1219. doi: 10.1109/CYBER.2017.8446473

G. Tinchev, A. Penate-Sanchez and M. Fallon, "Learning to See the Wood for the Trees: Deep Laser Localization in Urban and Natural Environments on a CPU," in IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1327-1334, April 2019. doi: 10.1109/LRA.2019.2895264

K. Tateno, F. Tombari, I. Laina and N. Navab, "CNN-SLAM: Real-Time Dense Monocular SLAM with Learned Depth Prediction," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6565-6574. doi: 10.1109/CVPR.2017.695