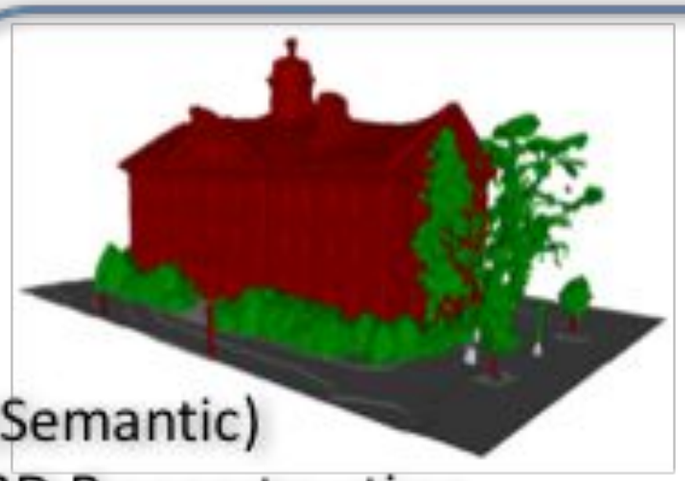
Overview of Research

Computer Vision and Geometry Group (CVG)

Microsoft Mixed Reality & Artificial Intelligence Lab

Research Topics at CVG



(Semantic)
3D Reconstruction

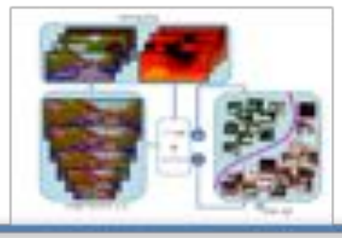
Visual Localization

Autonomous Vehicles

Mobile 3D Mapping

Camera Pose Estimation
& Calibration

AR / VR



Machine Learning

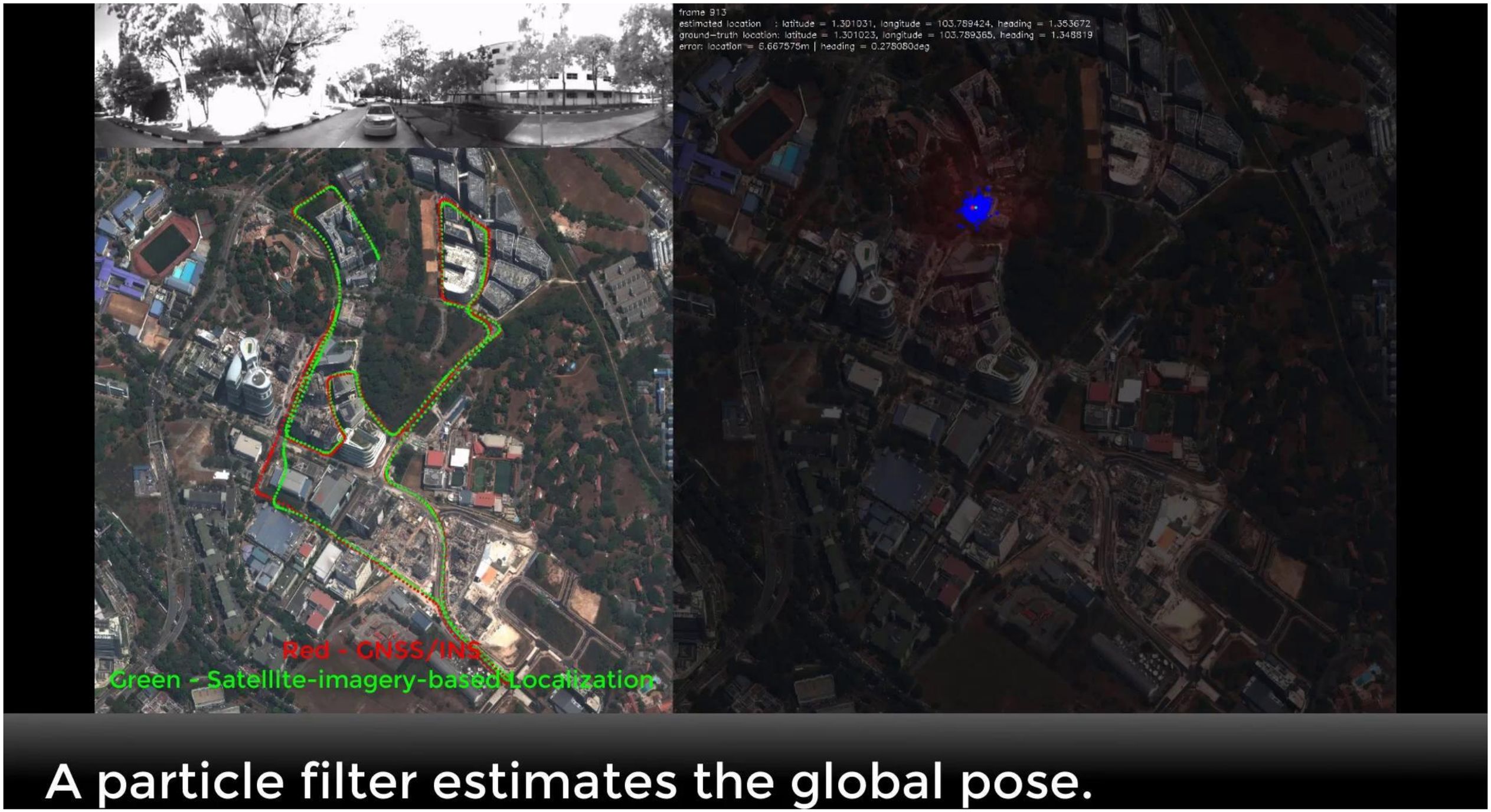
Trimbot2020

Horizon 2020
European Union Funding
for Research & Innovation

- European Research Project (6 academic + 1 industry partners)
- built camera system hardware and localization / SLAM software

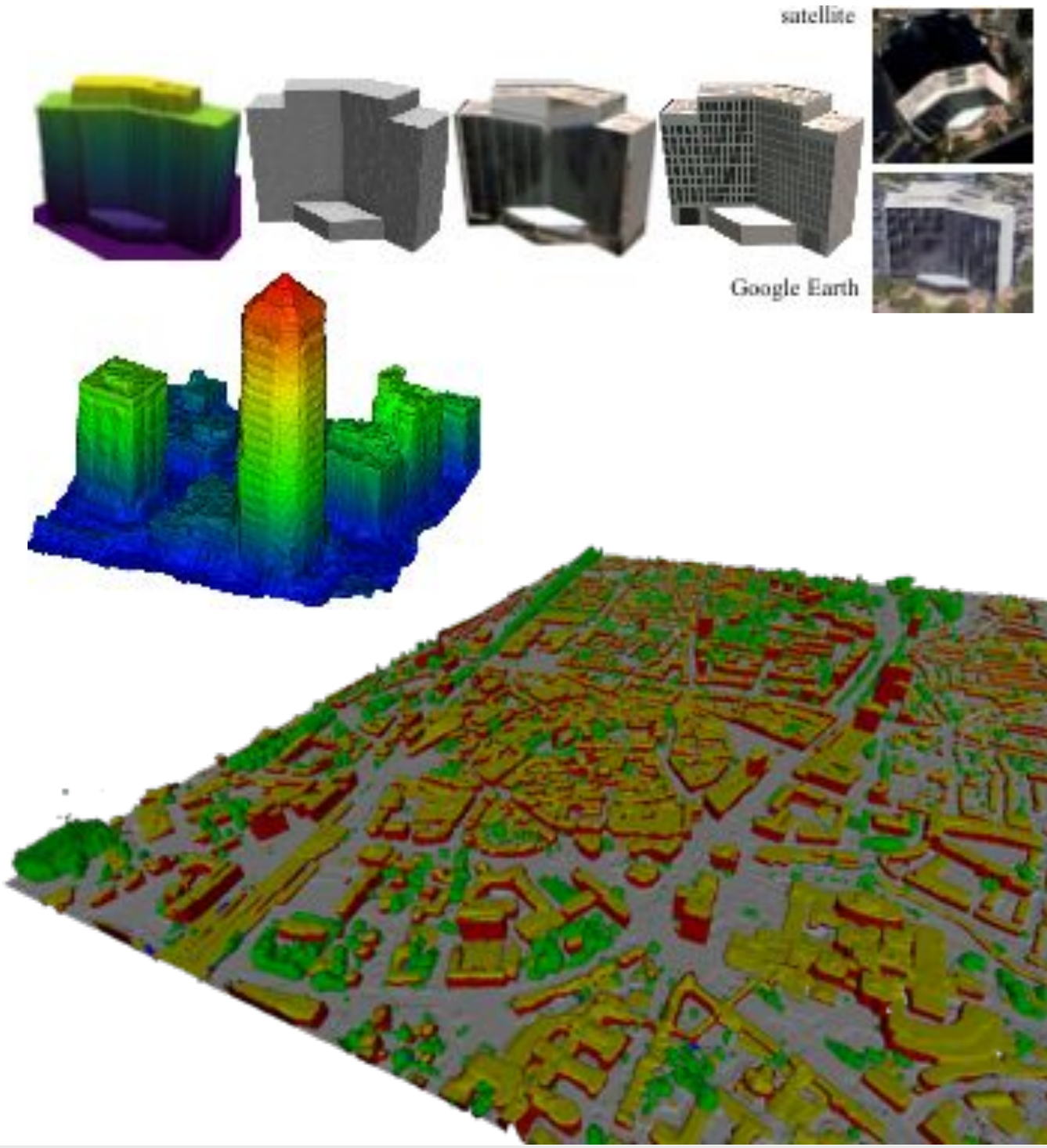
THE UNIVERSITY of EDINBURGH
BOSCH
ETH zürich
university of groningen
UNIVERSITY OF AMSTERDAM
UNI FREIBURG
WAGENINGEN

Autonomous Driving



A particle filter estimates the global pose.

Large-scale Urban Semantic 3D Reconstruction

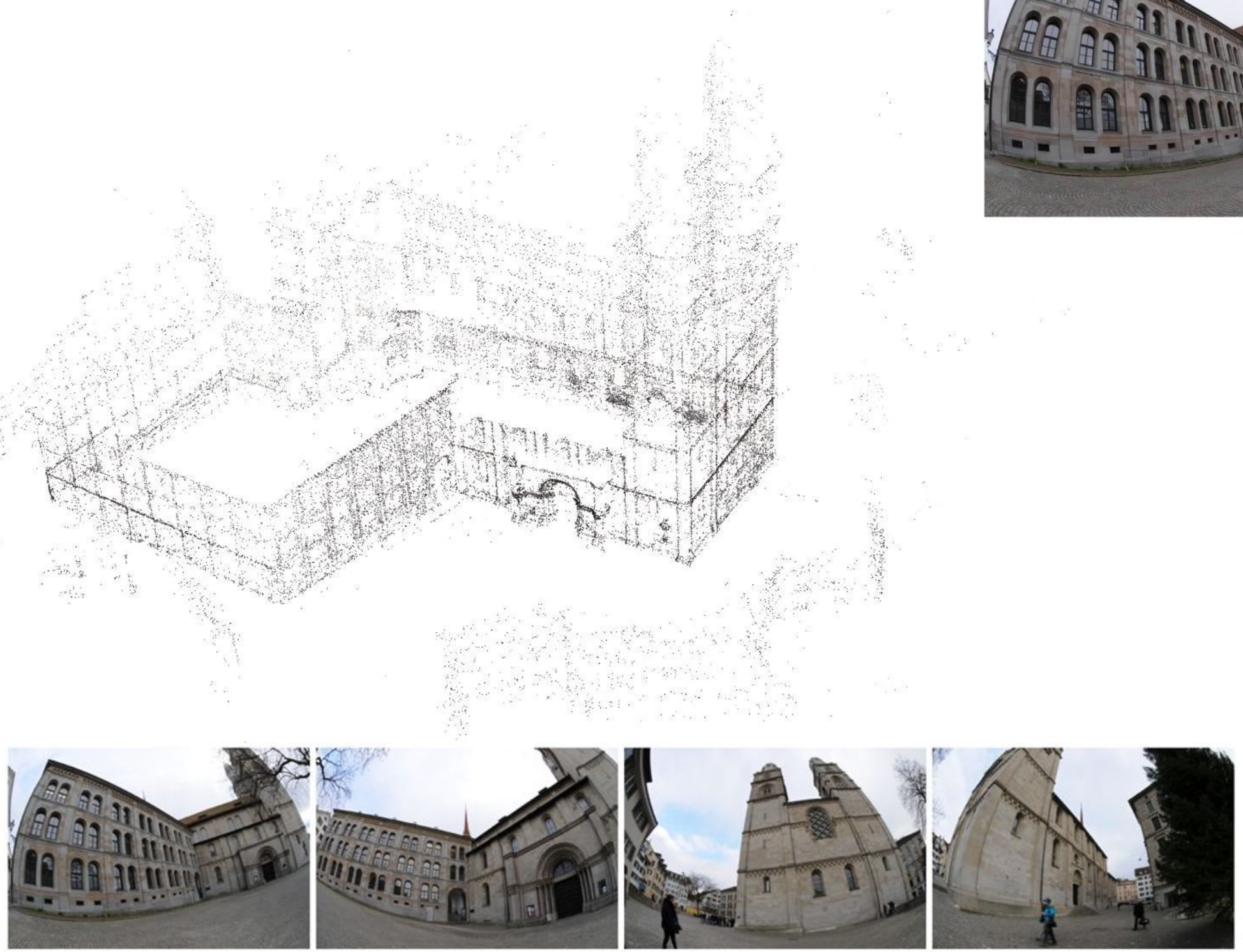


Structure-from-Motion with 1D Radial Cameras

Only consider direction of 2D projection, not the radial distance. Invariant to radial distortion!

Structure-from-Motion with 1D Radial Cameras

Big Church



MR&AI Lab Overview

- Applied computer vision research team
  9 fulltime researchers (ETH, EPFL, UNIZ, Oxford PhDs), 3 interns
- Academic collaborations with ETHZ, EPFL, INRIA (via Swiss Joint Research Center)
- Focus on long-term product impact and advancement of state of the art
- Strategic partnership between Microsoft and ETH Zurich including research funding
- Engage with local partners

Research Focus

- Crowd 3D mapping and relocalization
- First-person activity recognition
- Robotics & Mixed Reality

Project Overview

- HoloLens 2

- Azure Spatial Anchors

  Cross-platform cloud service to attach information to the real world, share it across devices and persist it over time

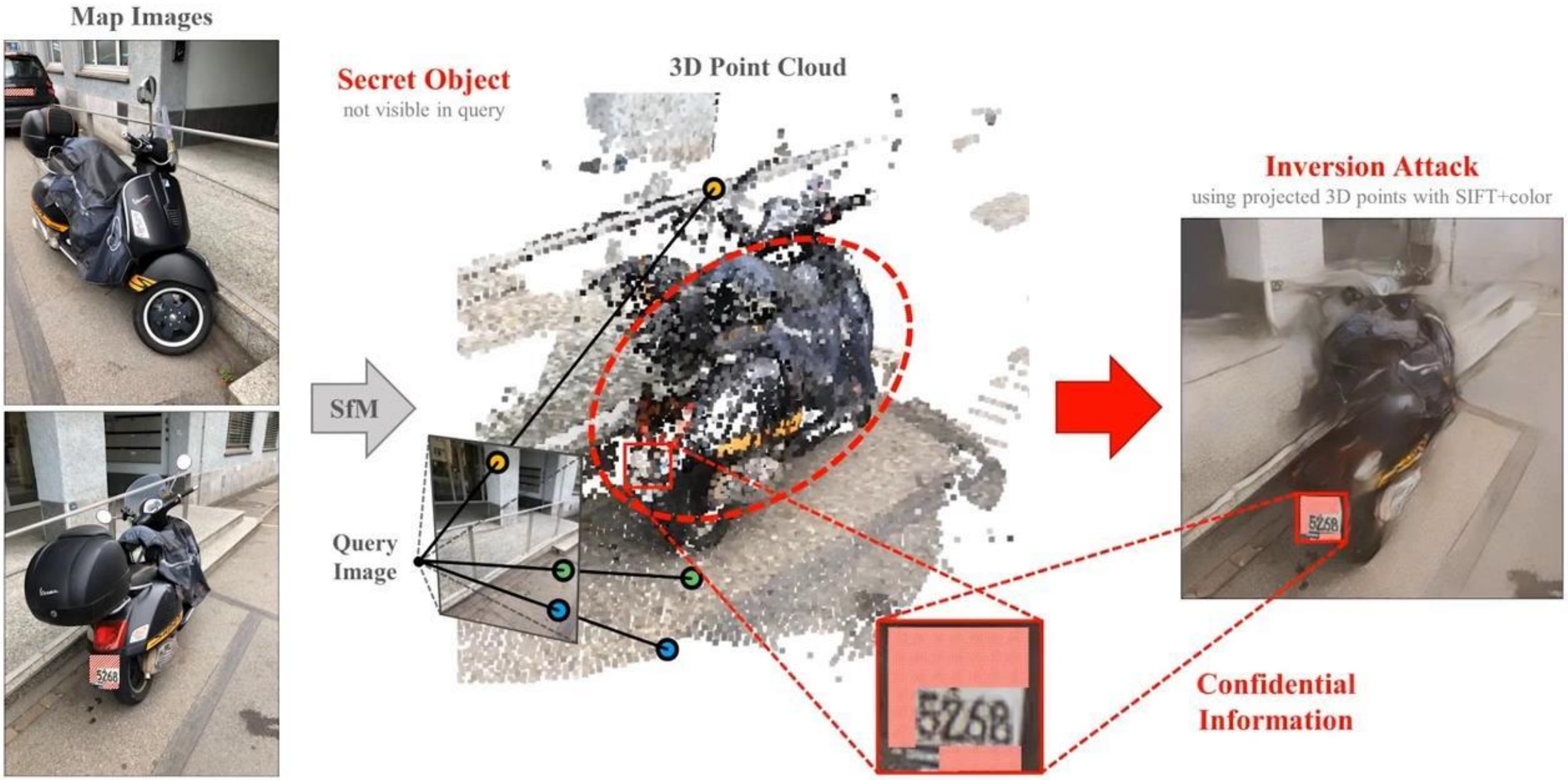
- Dynamics 365 Guides

  Skill teaching platform for first line workers

- Azure Kinect

  1Mpix depth sensor

Privacy-Preserving Localization



Overview of Thesis Topics
CVG & Microsoft MR/AI Labs

Projected Virtual Windows
(Master Thesis)

Goal: Using a projector and cameras, we create the illusion of a window granting vistas into virtual worlds.

Description:

You will build a compact hardware setup consisting of a small projector, a front-facing camera, and two back-facing cameras. The front-facing camera in combination with the projector itself allows the system to measure the projection surface, while the two back-facing cameras track the viewer's eye position.
Then, you will implement software to project a virtual window onto a wall, showing a virtual 3D scene perspectively corrected for the user's point of view.
Possible extensions of the system include improved tracking using a Kalman Filter or structured light surface scanning for advanced projection mapping onto non-planar surfaces.

Keywords: Triangulation, Tracking, Rendering, Projection Mapping, 3D Reconstruction

Requirements / Tools:

Required: Preferably C++ for processing of the camera streams, eye tracking, and rendering
Desirable: OpenGL; OpenCV or similar for face detection

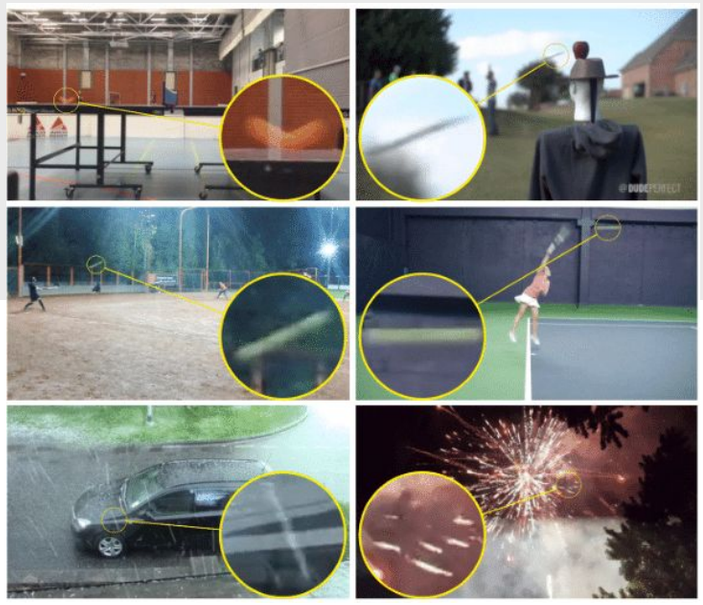
Supervisor: Daniel Thul <dthul@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Deep Fast Moving Object Detection

(Bachelor Thesis / Master Thesis)



Goal: Implement an FMO detection method based on deep learning, trained on synthetic data

Description:

The goal of this work is to build and implement an FMO detection method, which will be based on recently proposed deep learning frameworks for object detection, for example YOLO or Mask R- CNN. The method for example takes as input a single image which may or may not contain fast moving objects. The outputs are either bounding boxes around each FMO in case of YOLO, or segmentations of FMOs in case of Mask R-CNN. The student should be familiar with computer graphics techniques in order to generate synthetic data with fast moving objects for training. First, videos without FMOs will be collected. Second, a dataset of various 2D or 3D objects will be generated. Then, the objects will be inserted into videos as FMOs by applying motion blur models. The proposed method will perform tracking by detection of fast moving objects. The thesis will compare results to state-of-the-art FMO tracking and detection algorithms.

Requirements / Tools: Solid knowledge in Machine Learning / Deep Learning • Experience with Computer Graphics • Good programming skills (Python/C++)

Supervisor:

Denys Rozumnyi <denys.rozumnyi@inf.ethz.ch>

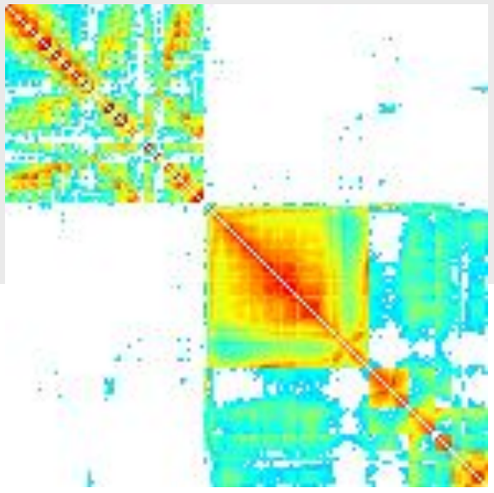
ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Guided-RANSAC for Multi-Model Fitting

(Bachelor Thesis / Master Thesis)



Goal: Improve RANSAC in settings with multiple good hypothesis.

Description:

Recently it been proposed to use a DNN to regress correspondence-wise weights to guide the random sample selection in RANSAC. In this approach the weights are directly used as sample probabilities and it is assumed that the each correspondence is sampled independently of others. In this thesis project we want to investigate regressing pairwise (or higher order) affinities between the correspondences which can be used to better guide the sampling. The goal is to improve performance in multi-model fitting scenarios, where it is ideal to only select samples from the same model. One such example could be homography estimation in a piecewise planar scenes.

[1] Brachmann & Rother, Neural-Guided RANSAC: Learning Where to Sample Model Hypotheses, ICCV 2019

Requirements / Tools: Solid knowledge in Machine Learning / Deep Learning • Good programming skills (Python/C++)

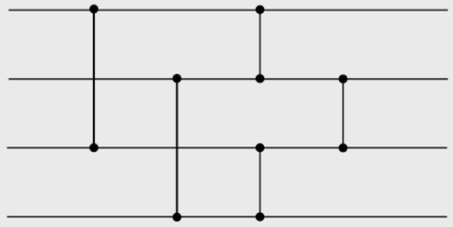
Supervisor:
Martin Oswald <moswald@inf.ethz.ch>
Viktor Larsson <vlarsson@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Differentiable Max-k and Sorting Networks

(Bachelor Thesis / Semester Thesis / Master Thesis)



Goal: Investigate different approaches for adding sorting layers in deep neural networks.

Description:

In this thesis we want to investigate how to best add layers that sort the input according to some criteria. In many applications we are not interested in the full sorted output, but only require subsets, e.g. the top-k largest elements or the median element. This could for instance be useful when you want to pool features from a varying number inputs only determined at inference time.

Sorting is inherently non-differentiable due to the discrete swapping of elements; simply rerouting the gradients might lead to unstable training behaviour since the sorting operation itself essentially becomes hidden from the training. Sorting can be expressed as a series of compositions of max and min. One possible approach is to simply replace these with the differentiable soft-max and soft-min. Part of the thesis work will be to investigate which of the possible approaches work best in practice.

Requirements / Tools: Solid knowledge in Machine Learning / Deep Learning • Good programming skills (Python/C++)

Supervisor:
Mihai Dusmanu <mihai.dusmanu@inf.ethz.ch>
Martin Oswald <moswald@inf.ethz.ch>
Viktor Larsson <vlarsson@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Efficient Geometric Representation using Learned Variable Size Feature Points (Master Thesis)

Goal: Learn to encode geometry into learned variable size feature points

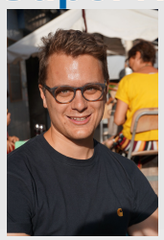
Description:

There are several different representations for 3D geometry like point clouds, volumetric grids or triangular meshes. All of these representations have their advantages and disadvantages. We want to tackle some of the disadvantages by proposing a novel representation based on feature points. The goal of this project is to propose a method that encodes scenes into a abstract representation using feature points of variable size. The variable size should be able to encode geometry at different levels of granularity. Vice-versa, the candidate should also propose a method that renders images or visualizes the scene given feature points as an input.

Requirements / Tools:

Solid knowledge in Machine Learning • Experience with Computer Vision and Computer Graphics • Good programming skills (Python/C++)

Supervisor:



Silvan Weder
silvan.weder@inf.ethz.ch
CAB G89

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Integration of Azure Kinect into InfiniTAM (Semester Thesis / Bachelor Thesis)

Goal: Integrate Azure Kinect into InfiniTAM Slam Pipeline

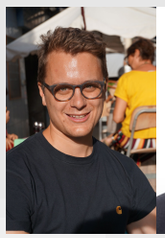
Description:

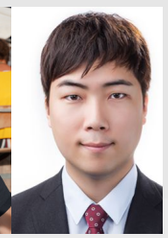
InfiniTAM is a very powerful and widely used SLAM pipeline for RGB-D sensors. Multiple consumer RGB-D sensors have already been integrated for usage in this pipeline. However, we would like to leverage the superior properties of Azure Kinect compared to other sensors for multiple applications with InfiniTAM. Therefore, the student should integrate Azure Kinect into the InfiniTAM code base. Optionally, additional work packages can be defined around the InfiniTAM pipeline depending on the students interests.

Requirements / Tools:

Experience with Computer Vision and Computer Graphics • Good programming skills (C++)

Supervisor:





Silvan Weder
silvan.weder@inf.ethz.ch / CAB G89
Taein Kwon
taein.kwon@inf.ethz.ch / CAB G85.2

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Avoiding Confusing Image Pairs in SfM

(Master Thesis)

Goal: Make Structure-from-Motion more robust to scene repetitions/ambiguities.

Description: Repetitive features or structures is a challenging problem in Structure-from-Motion. Especially in incremental approaches, if an image is incorrectly registered it can lead to degenerate reconstructions with scene duplications from which it is difficult to recover. To deal with this problem there are many proposed methods; e.g. enforcing loop consistency, checking for missing correspondences or view-graph selection/pruning. In this thesis project we want to investigate the possibility of using learning for classifying which image pairs should be matched. There are many possible approaches which we might consider in the thesis; e.g. doing the classification on the image pairs directly or performing a smarter next-view selection with LSTMs for incremental SfM. Since it is challenging to collect and annotate datasets which exhibit this behaviour, it will be essential to encode as much of the geometry of the problem as possible and only learn what is necessary for the task.

Requirements / Tools: Solid knowledge in 3D Vision and Machine Learning • Experience with Structure-from-Motion • Good programming skills (Python/C++)

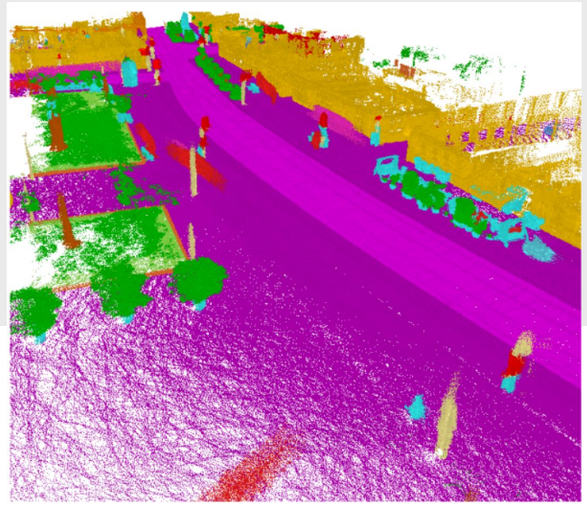
Supervisor:

Marcel Geppert <mgeppert@inf.ethz.ch>
Viktor Larsson <vlarsson@inf.ethz.ch>
Zhaopeng Cui <zhaopeng.cui@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Using semantic constraints for Localization (Master Thesis)



Goal: Building a localization algorithm that leverages semantic constraints for robust pose estimation.

Description:

An established way for large scale visual localization is to create a 3D structure, assign descriptors and then establish constraints between the image and map. Local feature descriptors can be used to recognize individual structures even in large maps, but are sensible to appearance changes. In contrast, semantic labels are not distinct descriptors, but can be reliably recognized even under strong changes. The goal of this project is to augment both types of features in a map representation and leverage both constraints for robust localization.

Requirements / Tools:

C++, experience with SLAM / Localization, some experience with Machine Learning

Supervisor:

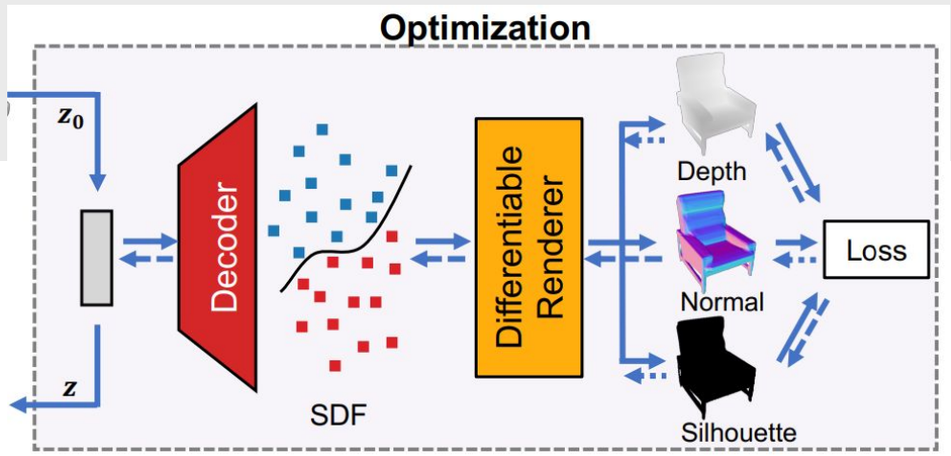
Marcel Geppert <marcel.geppert@inf.ethz.ch>
Viktor Larsson <vlarsson@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Latent Representation based Object SLAM

(Master Thesis)



Goal: Realize the object SLAM system using the latent 3D shape representation

Description:

Object SLAM aims to exploit the pre-built 3D models to help improve the SLAM system in terms both pose estimation and 3D reconstruction. Previous 3D objects are normally represented by 3D points or mesh, while recently it is shown to be able to represent an object using a latent code. So we plan to exploit to this new representation for object SLAM based on our recently proposed differentiable render for DeepSDF. More specially, object detection and recognition will be done first, and then the the object 3D shape as well as the camera poses are recovered through online optimization. The main challenge will be how to deal with the scale ambiguity and multiple objects observation.

Requirements / Tools:

Solid knowledge in 3D Vision and Machine Learning • Good programming skills (C++, Python)

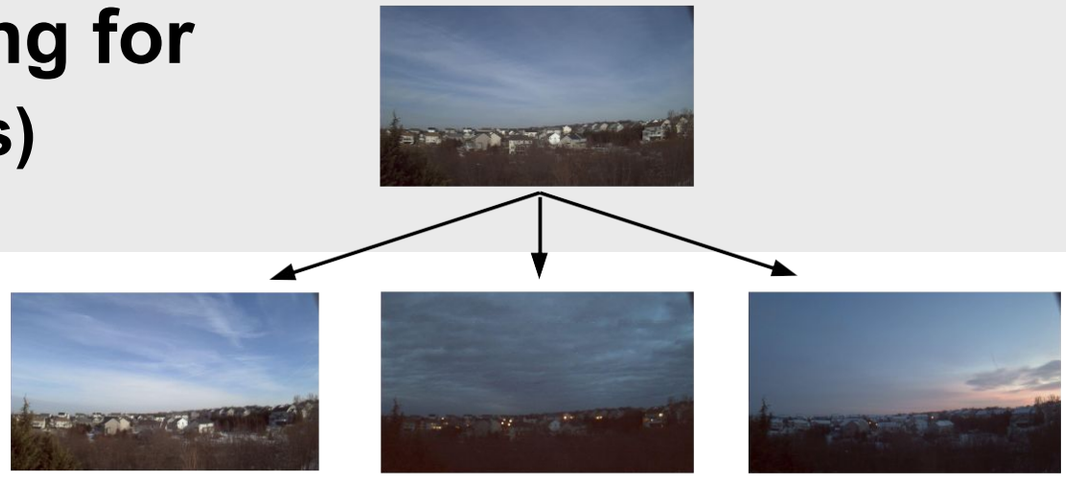
Supervisor:

Zhaopeng Cui <zhaopeng.cui@inf.ethz.ch>
Songyou Peng <songyou.peng@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Deep semi-supervised image relighting for data augmentation (Master Thesis)



Goal: Learn to encode the illumination of an image and to generate new lighting conditions.

Description:

The goal of this work is to design and implement a deep learning pipeline taking real images as input and learning an encoding of their illumination in an unsupervised way. It should thus manage to disentangle the geometry of the image from its lighting and learn a relevant probability distribution for the illumination embedding. At test time, given an original image and a randomly sampled illumination encoding, the network should be able to generate a new realistic image with the same geometry as the original image but under a new custom illumination. The pipeline will be trained on a mixture of synthetic and real datasets offering varying light conditions and will leverage the latest advances in Generative Adversarial Networks (GANs).

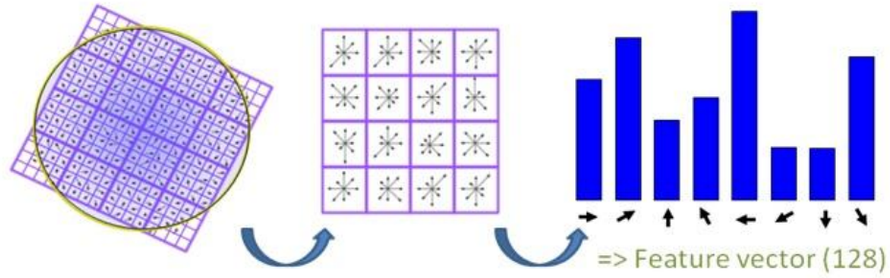
Requirements / Tools: Solid knowledge in Deep Learning / Computer Vision • Experience with GANs / Autoencoders is a plus • Good programming skills (Python, Pytorch or Tensorflow)

Supervisor:
Rémi Pautrat <remi.pautrat@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

∇SIFT: differentiable SIFT descriptor
(Master's Thesis)



https://gilscvblog.com/2013/08/18/a-short-introduction-to-descriptors/

Goal: Use differentiable histograms to implement SIFT descriptors

Description:

Despite recent advances in end-to-end trainable feature detectors and descriptors, handcrafted SIFT [1] is still one of the top performing methods, especially for scenes containing images taken under similar illumination conditions. Previous work [2] using SIFT as a CNN module settled for a partially differentiable SIFT implementation, considering only the gradient flow through the magnitude of image gradients and ignoring the orientation.

The initial goal of this project is to implement fully differentiable SIFT descriptors using differentiable histograms [3]. This implementation will be compared to state-of-the-art methods on a large-scale dataset of patches [4]. Several other applications can be envisaged: replacing the partially differentiable SIFT from [2] and, more advanced, learning intermediate image representations / binning weights in order to improve the performance.

[1] - Distinctive Image Features from Scale-Invariant Keypoints, Lowe, IJCV 2003
[2] - Repeatability Is Not Enough: Learning Affine Regions via Discriminability, Mishkin et al., ECCV 2018
[3] - Learning Deep Embeddings with Histogram Loss, Ustinova & Lempitsky, NeurIPS 2016
[4] - HPatches: A benchmark and evaluation of handcrafted and learned local descriptors, Balntas, Lenc et al., CVPR 2017

Requirements / Tools:

Solid knowledge in Machine Learning & Computer Vision
Good programming skills (Python)
Experience with a deep learning framework (preferably PyTorch)

Supervisor:

Mihai Dusmanu <mihai.dusmanu@inf.ethz.ch>
Viktor Larsson <vlarsson@inf.ethz.ch>
Torsten Sattler <torsat@chalmers.se>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Learning to propagate variational methods (semester thesis)

Goal: Train a network on propagation for semantic scene completion

Description:

Variational methods in computer vision refer to those methods that solve problems by posing them as functional minimizations. Such techniques can be applied for image denoising, inpainting, segmentation… In our case, we are interested in applications to semantic 3D reconstructions.

The minimization of such functionals relies on iterative algorithm, such as primal dual, which minimize the given objective at every step until convergence. Recent work has shown that these algorithm can be implemented into neural networks (referred here as variational networks). The main interest of such networks is the fact that they rely on few parameters.

For this to work, the number of iterations in the minimization algorithm must be fixed during training. Unfortunately, unlike true variational methods, when running the network for inference, adding more iterations does not improve the results, but often degrades them.

In this project, we want to explore methods that will allow to train a variational network that will improve when more iterations are added. To do so, we will try implementing a different loss function that focuses more on the functional minimization, and try to use synthetic ground truth data that corresponds to different steps of the minimization.

Requirements / Tools:

Python, convex optimization

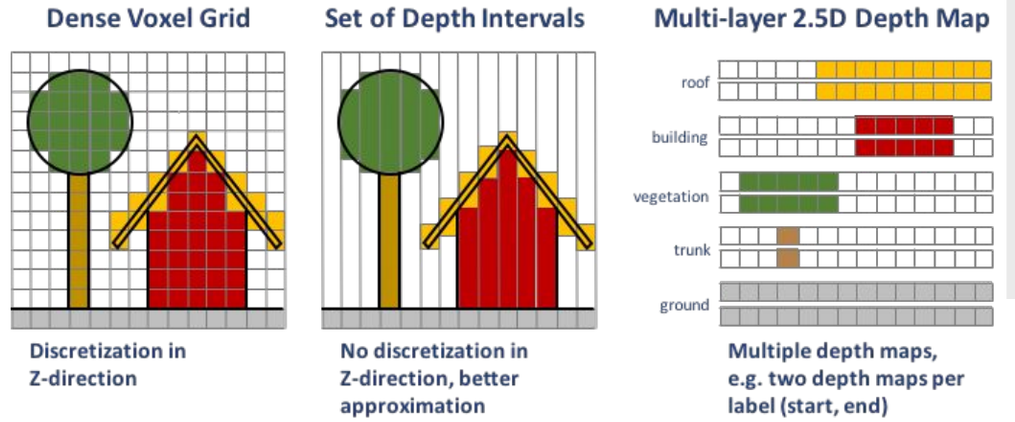
Supervisor:

Ian Cherabier <ian.cherabier@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Online Semantic Reconstruction with Multi-Layer Depth Maps



Goal: Perform large scale city reconstruction from a stream of depth maps

Description:

This project targets the reconstruction of cities from aerial or satellite images. Due to the coverage that can be obtained from aerial data, the obtained model are large scale.

One way of dealing with large scale models is to represent them as single multi-layered gravity aligned depth map. In other words, we have top view depth maps (also sometimes called bird-view), with multiple channels, where each channel indicates a layered depth information (e.g. one channel for where a building starts, one for where it ends).

To generate such representation, we use many input depth maps obtained from calibrated aerial / satellite images. These images are not necessarily taken at the same time, and it may happen that more are captured after an initial reconstruction has already been computed. One major limitation of current methods is that in order to update the model, we would need to re-run the method from scratch with all the data.

In this project, we propose to explore solutions for online reconstruction, that would allow us to treat the input depth maps as a stream of data rather than a monolithic set. This would allow for models that can be updated through time, while also paving the way for faster reconstructions.

-

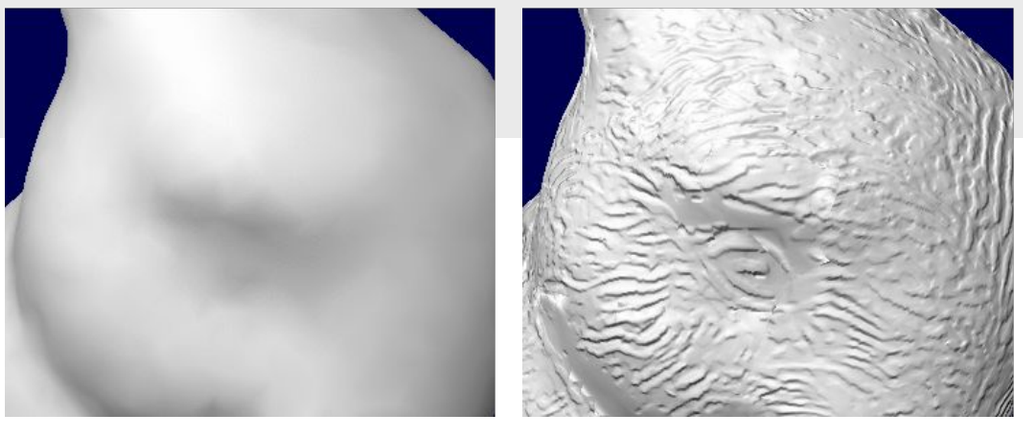
Requirements / Tools:

Python

Supervisor:

Ian Cherabier <ian.cherabier@inf.ethz.ch>.
Martin Oswald <martin.oswlad@inf.ethz.ch>

Super-resolution Geometry Refinement (Master Thesis)



Goal: Novel self-supervised super-resolution approach that jointly estimates high-res texture and geometry

Description:

Create a neural network that takes as input a coarse mesh and a set of input images with known camera positions and parameters that observe the mesh and which outputs a super-resolved texture map and displacement map.

The general idea is that geometric refinement can be best estimated when the re-projection of the corresponding texture information is as consistent as possible. Therefore, instead of using large dataset of high-resolution 3D data for supervised learning, we aim to minimize the reprojection error and use it for self-supervising the network. The network will learn to estimate colors and displacements on the mesh by back-propagating residual errors with regard to the input images via a differentiable renderer.

References:

[1] Audrey Richard, Ian Cherabier, Vagia Tsiminaki, Martin R. Oswald, Marc Pollefeys, and Konrad Schindler, Learned Multi-View Texture Super-resolution, 3DV 2019
[2] B. Goldlücke, M. Aubry, K. Kolev, and D. Cremers. A super-resolution framework for high-accuracy multiview re- construction. IJCV 106(2):172–191, 2014
[3] V. Tsiminaki, J. Franco, and E. Boyer. High resolution 3D shape texture from multiple videos. CVPR, 2014
[4] Miika Aittala, Frédo Durand, Burst Image Deblurring Using Permutation Invariant Convolutional Neural Networks, ECCV 2018

Requirements / Tools:

Python

Supervisor:

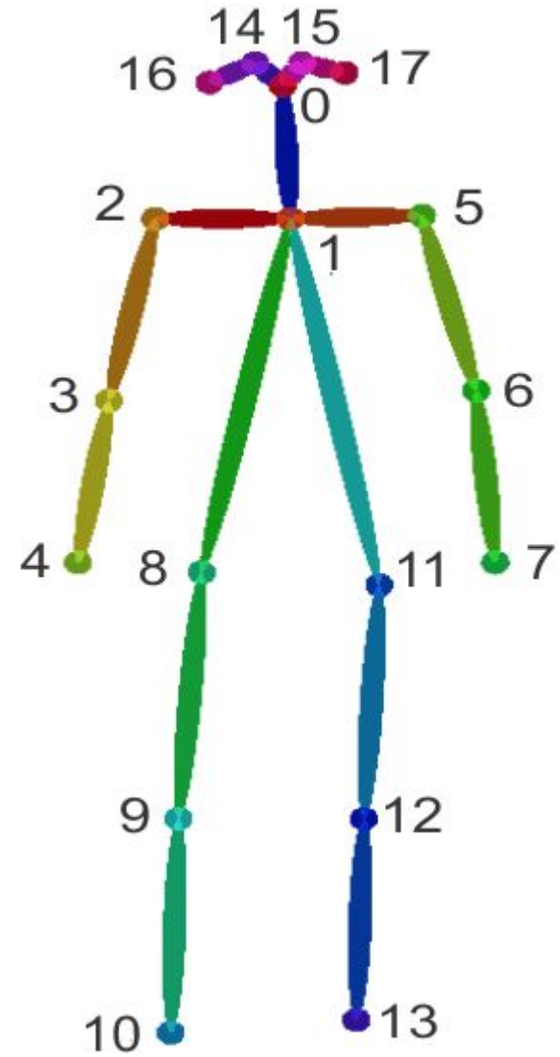
Martin Oswald <martin.oswlad@inf.ethz.ch>
Ian Cherabier <ian.cherabier@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Full-body Motion Tracking with HoloLens (Master Thesis)

Goal: Use IMU-tracking on human limbs together with observations from HoloLens to estimate the full body pose of the user



Description:

The higher-level motivation for the thesis is to help stroke patients in Neurorehabilitation and also to analyse walking behavior, but the approach has many applications.

Giving the hardware setup of a HoloLens2, an IMU-based full body tracking system and an external body tracking system, we have the possibilities to capture training data for training or refining a pose estimation network that should also fuse the IMU data.

References:

[1] Zhe Cao and Gines Hidalgo and Tomas Simon and Shih-En Wei and Yaser Sheikh, OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.

Requirements / Tools:

Python

Supervisor:

Martin Oswald <martin.oswlad@inf.ethz.ch>
Jeremia Held <jeremia.held@uzh.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Scene Flow with Point-Based Multi-View Stereo (Master Thesis)

Goal: Integrate a multi-view stereo method into a scene flow framework.

Description:

Scene flow is the 3D equivalent of optical flow. It therefore describes the (x,y,z) movement of a 3D point from one frame to the next. Lv et al. [1] proposed a framework for estimating scene flow from RGBD input. Furthermore, they explicitly model the rigidity in the scene in order to differentiate between deforming objects and static background. However, depth input is still required for successful scene flow estimation. In this project, we would like to explore the feasibility of combining a multi-view stereo approach [2] with this framework in order to get rid of the depth frame requirement.

[1] Z. Lv, K. Kim, A. Troccoli, D. Sun, J. M. Rehg, and J. Kautz, "**Learning Rigidity in Dynamic Scenes with a Moving Camera for 3D Motion Field Estimation**," presented at the Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 468–484.
[2] R. Chen, S. Han, J. Xu, and H. Su, "**Point-Based Multi-View Stereo Network**," presented at the Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 1538–1547.

Requirements / Tools:

Required: Experience with deep learning and python
Desirable: Experience with PyTorch

Supervisor:

Sandro Lombardi <sandro.lombardi@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Deep Image Super Resolution with Continuous Implicit Functions (Semester Thesis)

Goal: Develop a system to represent images as deep implicit functions and evaluate it on the task of super resolution

Description:

The recently proposed method of representing 3D surfaces as implicit signed distance functions from R^3 to R has shown the ability to model high quality and realistic representations of objects. The aim of this project is to explore the use of this approach for the representation and manipulation of images.

Working around the idea of representing an image as a function from the XY coordinate space in R^2 to the RGB color space in R^3, the student will develop a system to encode images in such representation and evaluate it for the task of super resolution by leveraging on the existing works on deep image manipulation. Exploring other image manipulation tasks such as super-resolution, inpainting, style transfer, animation, semantic conditioning, harmonization or any other proposed by the student are encouraged but not required for this project.

References:

Park, Jeong Joon, et al. "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation." CVPR, 2019.
Xu, Qiangeng, et al. "DISN: Deep Implicit Surface Network for High-quality Single-view 3D Reconstruction." NeurIPS, 2019.
Oechsle, Michael, et al. "Texture Fields: Learning Texture Representations in Function Space." ICCV, 2019.
Shaham, Tamar Rott, et al. "SinGAN: Learning a generative model from a single natural image." ICCV, 2019.
Goodfellow, Ian, et al. "Generative adversarial nets." NIPS, 2014.

Requirements / Tools:

Solid knowledge in Machine Learning, Python

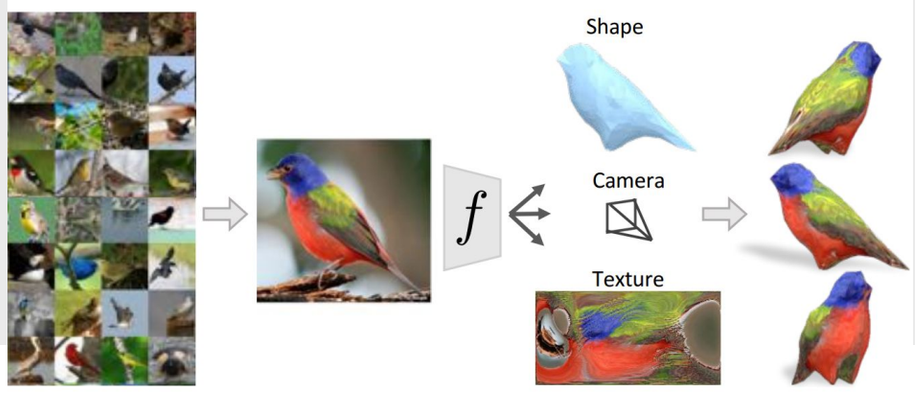
Supervisor:

Luca Cavalli <lcavalli@ethz.ch>
Songyou Peng <songyou.peng@inf.ethz.ch>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Learning Category-Specific Deep Implicit Functions from Image Collections (Master Thesis / Semester Project)



Goal: Develop a network to infer 3D shape, texture and camera pose from a single image.

Description:

Consider the bird image on the teaser above, our human beings are able to infer the 3D shape, texture and the camera pose from just this 2D image. Some recent works [1] tried to tackle this problem. However, since they use mesh as shape representation, the predicted shapes are limited to low resolution and fixed topologies. With the recently proposed continuous shape representation [2] and its differentiable renderer [3], we can reconstruct high-resolution shapes with arbitrary topologies from only 2D supervisions, e.g. masks and depths. Therefore, the goal of this project is to adapt the deep implicit functions into this challenging task.

There still remains some challenges: (a) How to incorporate keypoints information into the supervision. (b) Except for mask and keypoints, can we incorporate texture into training? [4] (c) Is implicit functions able to describe the complicated color space? (d) Possible to infer 3D textured shape from just some text description? [5]

References:

[1] Learning Category-Specific Mesh Reconstruction from Image Collections, ECCV'18

[2] DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR'19

[3] DIST: Rendering Deep Implicit Signed Distance Function with Differentiable Sphere Tracing, ArXiv'19

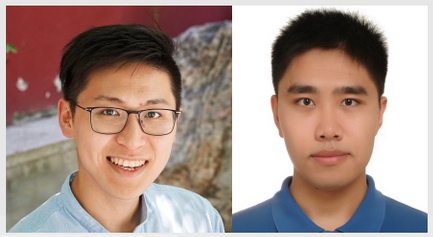
[4] Texture Fields: Learning Texture Representations in Function Space, ICCV'19

[5] AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks, CVPR'18

Requirements / Tools:

Solid knowledge in Machine Learning & Computer Vision & Computer Graphics. Good programming skills (Python / C++)

Supervisor:



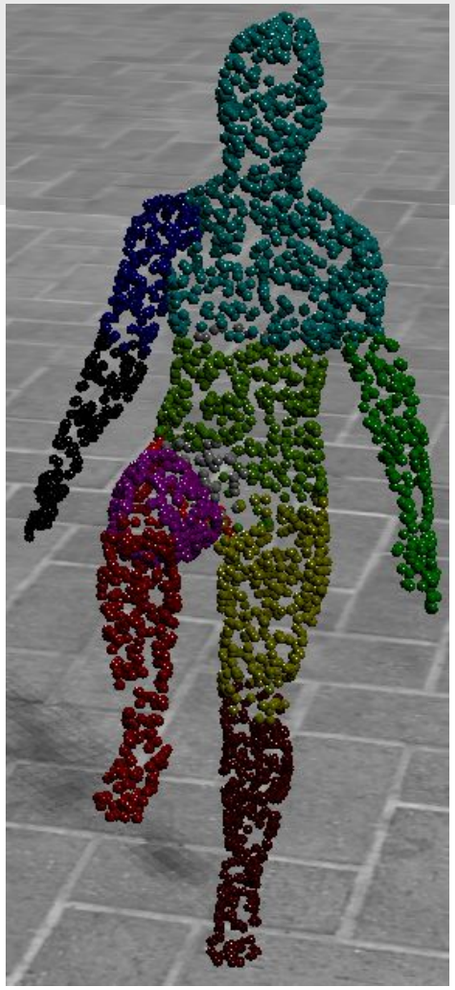
Songyou Peng (CAB G86.3)
songyou.peng@inf.ethz.ch / CAB G86.3

Zhaopeng Cui
zhaopeng.cui@inf.ethz.ch / CNB G103.2

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Non-rigid Reconstruction using Deep Learning (Master thesis)



Goal: The goal is to build a deformable 3D model out of 2D point tracks of articulated data using deep learning

Description:

Deformable models in computer graphics are often represented using a rigged mesh structure, consisting of static mesh and articulated skeleton, where the position of each point is determined by applying linear blending on corresponding transformation matrices of joints. The goal of this project to estimate the most probable rigged mesh structure that can fit observed points tracks. The input features for the optimized model will be trained using deep learning by back propagation through the optimization pipeline.

Requirements / Tools:

Solid knowledge in Machine Learning

Supervisor:

Lubor Ladicky (lubor.ladicky@inf.ethz.ch)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Deep Multi-View Stereo
(Master Thesis)

Goal: Implement and improve a multi-view stereo pipeline with deep learning framework

Description:

Deep learning methods are climbing the ranking of two popular multi-view stereo benchmarks: eth3d.net and tanksandtemples.org. Even if they generalize well they usually require extensive use of gpu memory and hardly scale to high resolution imagery.

The goal of this thesis is to advance the state of the art of multi-view stereo reconstruction with deep learning in terms of either speed, accuracy or memory consumption.

An optional variation of this thesis consists in specifically tailoring the reconstruction for the camera present in Hololens AR headset.

Keywords: Deep Learning, 3D Reconstruction

Requirements / Tools:

Required: Deep learning, python, passion, creativity

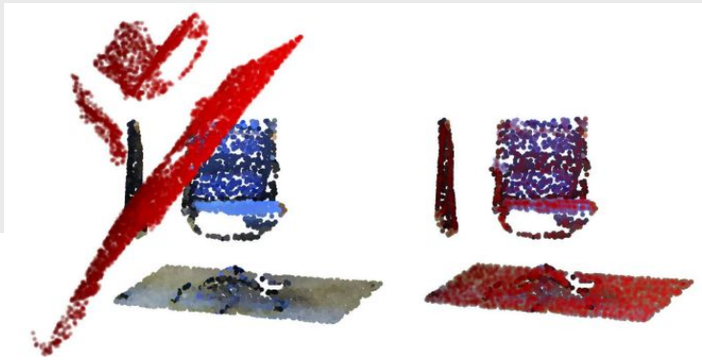
Supervisor:

Silvano Galliani
<sigallia@microsoft.com>

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Learning Based Registration on Point Clouds (Master/Semester Thesis)



Goal: A learning-based framework for registration that relies on the powerful expressive ability of network while imitating the traditional optimization pipeline

Description:

Point cloud registration is a key problem for 3D vision applied to autonomous driving, medical imaging, and robotic perception. This goal of the problem is to find a rigid transformation that is able to perfectly align two point clouds. Iterative Closest Point (ICP) and its variants provide simple and easily-implemented iterative methods for this task, but these algorithms can easily converge to spurious local optima and highly rely on the initial guess of the results. To address these limits and other difficulties in the ICP pipeline, the goal of this project is to propose a learning-based method for robust and accurate registration between point cloud with bad initial pose guesses.

References:

[1] Lv, Zhaoyang, et al. "Taking a Deeper Look at the Inverse Compositional Algorithm." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
[2] Aoki, Yasuhiro, et al. "Pointnetlk: Robust & efficient point cloud registration using pointnet." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
[3] Sarode, Vinit, et al. "PCRNet: Point Cloud Registration Network using PointNet Encoding." arXiv preprint arXiv:1908.07906 (2019).

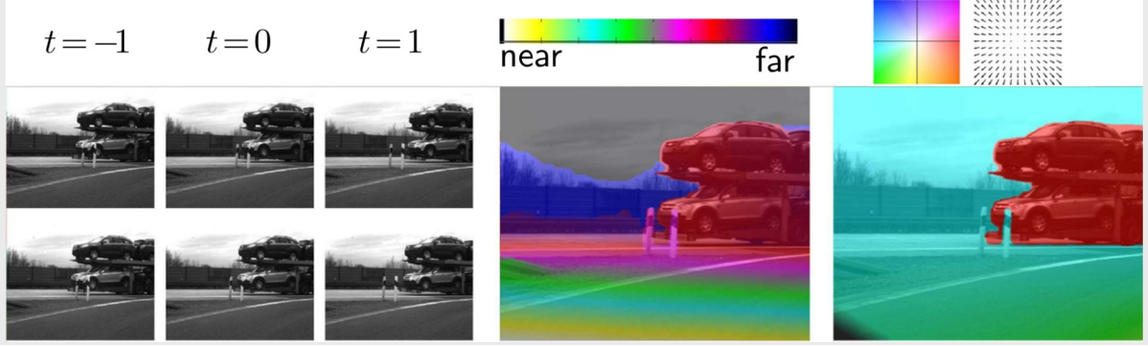
Requirements / Tools:

Prefered: Python, Deep learning, Strong self-motivation

Supervisor:

Peidong Liu (peidong.liu@inf.ethz.ch)
Xingxing Zuo (xinzuo@ethz.ch)

Deep Piecewise Rigid Scene Flow



Goal: This work should step-by-step generalize a state-of-the-art pipeline for scene flow estimation to Deep Learning

Description:

Scene flow is the 3D motion field between images of two (or more) times steps and two (or more) cameras. A possible representation is to assign a 3D motion vector to each pixel of a designated reference view.

In this work we follow the representation proposed in [1] which is still state-of-the-art for outdoor environments and step-by-step port the algorithm to the GPU within a recent deep learning frameworks.

These steps include proposal generation with a "fake" GAN approach [2]. We need to randomly generate several proposal planes at a spatial location, ideally exploiting the data of a cost volume generated in the beginning. At second we need to adjust a suitable optimization algorithm on the GPU (convex belief propagation, code available) to the specific graph structure and integrate the latter into an end-to-end deep learning pipeline. This will allow us to learn data term via feature vectors and pairwise coupling terms for the energy optimization step.

[1] Vogel et. al. "Piecewise-Rigid Scene-flow", ICCV 13

[2] Chen et al. "Photographic Image Synthesis with Cascaded Refinement Networks", ICCV 17

[3] Ma et al. “Deep Rigid Instance Scene Flow”, CVPR 2019

”Requirements / Tools:

Prefered: Python, C++, Cuda, It would also be helpful to have some experience with a deep learning framework, eg. Pytorch, TF-Flow, ..

Supervisor:

Christoph Vogel (Christoph.vogel@microsoft.com)

Topological Map Extraction from Overhead Images (Semester Project)

Goal: To improve the results of [1], conduct more experiments and make better post-processing.

Description:

The model PolyMapper proposed in [1] successfully circumvents the conventional pixel-wise segmentation of (aerial) images and predict objects in a vector representation directly. It can directly extracts the topological map of a city from overhead images as collections of building footprints and road networks. However, further model improvements can be done, which are shown as follows: (1) Change the backbone from VGG to ResNet or even better network; (2) Add attention mechanism to the RNN scheme; (3) Add evaluation network; (4) Improve post-processing steps for road networks. More experiments should be conducted as well: (1) Do experiments on the SpaceNet [3] dataset; (2) Ablation study. Note: this is mostly an engineering-oriented project and requires good implementation and coding ability. Hopefully the work can finally be done by June 2020 and can be turned to a journal paper.

Reference:

[1] Li et al. Topological Map Extraction From Overhead Images. ICCV 2019.
[2] Acuna et at. Efficient Annotation of Segmentation Datasets with Polygon-RNN++. CVPR 2018.
[3] SpaceNet. https://spacenetchallenge.github.io/

Requirements / Tools:

Python & TensorFlow, Graph Theory/Algorithms, Coding/Implementation

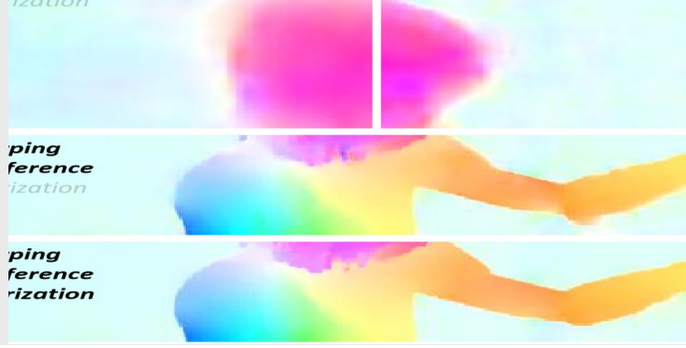
Supervisor:

Zuoyue Li <li.zuoyue@inf.ethz.ch>
CAB G 85.2

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Structured Deep Learning for Optical Flow & Stereo



Goal: Develop a deep learning based method that combines hierarchical energy minimization via a CRF model with deep learning.

Description:

Despite the success of unstructured deep learning most recent state-of-the-art methods for optical flow [1,2] rely on a more structured approach, using local cost volumes local filtering operations in a hierarchical setup. This structure strongly reminds of solving a restricted CRF model at each hierarchy level. In this work, the idea is to explicitly model the latter, while allowing for a complete back-propagation mechanism in the explicit optimization process. We will make use of recent developments that allow to efficiently unroll modern block coordinate ascent methods for the dual CRF problem on the GPU. Finally, we want to investigate whether going beyond simple pairwise connectivity or employing denser graphs [3] can improve the results.

[1] Deqing Sun et al. "PWC-Net: Cnns for Optical Flow Using Pyramid, Warping and Cost Volume", CVPR 18

[2] Hui et al. Liteflownet: "A lightweight convolutional neural network for optical flow estimation", CVPR 18

[3] Tourani et al. “MPLP++: Fast, Parallel Dual Block-Coordinate Ascent for Dense Graphical Models”, ECCV 2018

Requirements / Tools:

Prefered: Python, C++, Cuda, It would also be helpful to have some experience with a deep learning framework, eg. Pytorch, TF-Flow, ..

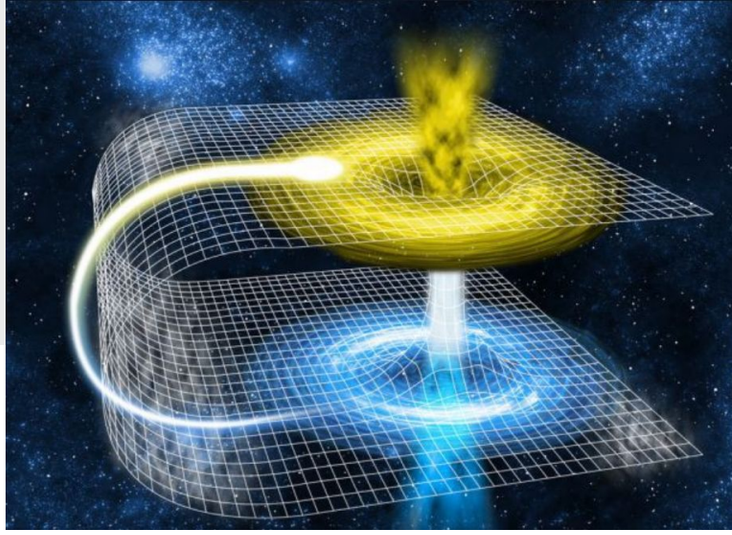
Supervisor:

Christoph Vogel (Christoph.vogel@microsoft.com)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Deep Learning Models for 3D Map Inconsistency Detection



Goal: Develop discriminative or generative Deep Learning Model(s) detecting anomalies in reconstructed 3D maps

Description:

Despite the success of the most recent state-of-the-art Structure from Motion (SfM) pipelines [1,2], they are still vulnerable to visual aliasing, i.e. when input images from multiple locations look similar. This typically distorts the resulting maps and renders them unusable in practice and have to be avoided by any means.

While the algorithmic detection of those malicious configurations remains an open problem, such anomalies in the reconstructed 3D models (ie not input data) are usually rather easy to spot for humans.

The idea of this thesis is to apply Deep Learning models to the detection problem. A trivial observations is that it appears more natural to the problem to directly operate on the graph-structured input data. We will follow latest trends on deep learning and apply Graph Neural Networks [3,4] and Generative Adversarial Networks (GAN).

[1] Triggs, “Bundle Adjustment – A Modern Synthesis”
[2] Zhang, “Distributed Very Large Scale Bundle Adjustment by Global Camera Consensus”, ICCV 2017
[3] Zhou et al. "Graph Neural Networks: A Review of Methods and Applications", arxiv 2018
[4] Wu et al. “A Comprehensive Survey on Graph Neural Networks”, arxiv 2019

Requirements / Tools:

Prefered: Python, experience with Pytorch, TensorFlow

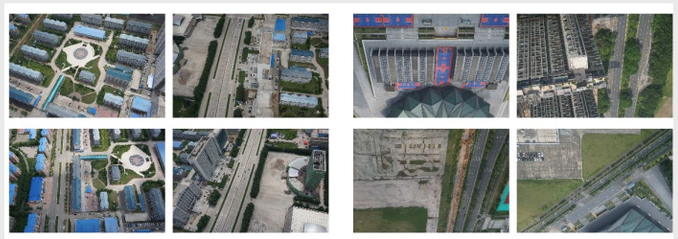
Supervisor:

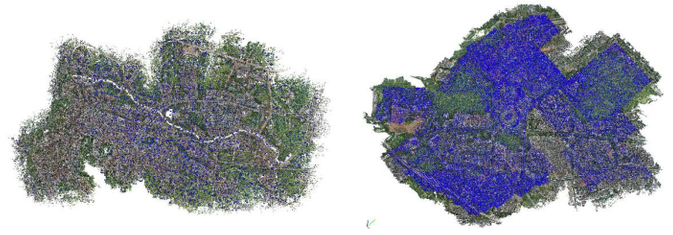
Christoph Vogel (Christoph.vogel@microsoft.com)
Ondrej Miksik (ondrej.miksik@microsoft.com)
Lukas Gruber (lukas.gruber@microsoft.com)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Evaluation of optimization algorithms for Large scale Bundle adjustment





Goal: Implement and evaluate different algorithms for large scale bundle adjustment problems

Description:

Bundle adjustment can be seen as a special instance of Structured Matrix Factorization, which is one of the classical problems for non-convex optimization. Most methods, however, do not scale well with the size of the the problem, such that most promising approaches are variants of first-order (gradient based) type.
The state-of-the-art method is a first-order method [1], but surprisingly does not apply one of the well-known non-convex algorithms that exist in the literature [2,3,4], for which convergence is proven and that are known to be very suitable to the problem class.
Hence, the goal of this thesis is to implement different state-of-the-art methods and evaluate them on large bundle adjustment problems.

[1] Zhang, “Distributed Very Large Scale Bundle Adjustment by Global Camera Consensus”, ICCV 2017

[2] Bolte et al. “Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems”, arxiv 2014

[3] Malitsky et al. “Model Function Based Conditional Gradient Method with Armijo-like Line Search”, arxiv 2019

[4] Komodakis, “MRF energy minimization and beyond via dual decomposition”, PAMI 2011

Requirements / Tools:

Prefered: Python, C++, knowledge in mathematical optimization

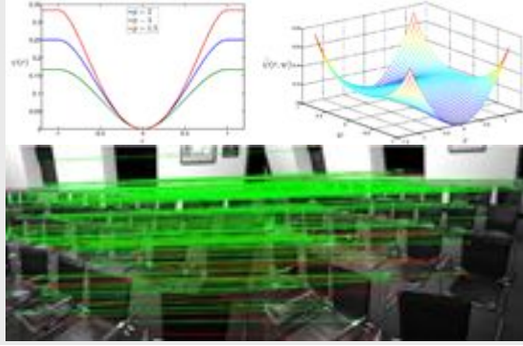
Supervisor:

Christoph Vogel (Christoph.vogel@microsoft.com)
Ondrej Miksik (ondrej.miksik@microsoft.com)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Learn to Initialize – Deep Learning of initial solutions for the optimization of robust cost functions



Goal: Develop a pipeline for robust cost function optimization which leverages Deep Learning for smart initialization in end-to-end fashion

Description:

The optimization of non-convex, robust cost functions is required in any pipeline for 3D reconstruction, localization, pose estimation and scan registration. Similarly, it is the core of robust model fitting for deformable objects like human hands or skeletons to data.

For those time-critical, medium to large scale problems, continuous optimization techniques, in particular, non-linear least squares is considered the primary method of choice. Further, to find a good optimum, techniques include graduated non-convexity, the construction of surrogate loss functions or employing lifted robust kernels per residual. While these techniques are related to each other, eg. lifting can be regarded as constructing a quadratic surrogate per residual, they all require a good initial solution for the problem or, equivalently, an initial weighting of the residuals.
The goal of this work is to leverage the power of deep learning to deliver this initial guess. To that end we will fall back on known deep learning models that can handle a variable number of inputs [1]. This initial solution can be refined by an optimization network that will be the core contribution of this thesis. For this part we will implement a conventional Gauss-Newton solver, extend the algorithm to robust optimization, eg. [2]. To facilitate the learning process, our implementation will allow to backpropagate through the optimization process.

[1] Qi et al., “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space”, Neurips 2017
[2] Zach et al, “Descending, lifting or smoothing: Secrets of robust cost optimization”, ECCV 2018

Requirements / Tools:

Prefered: Python, experience with Pytorch/TensorFlow, C++, knowledge in mathematical optimization

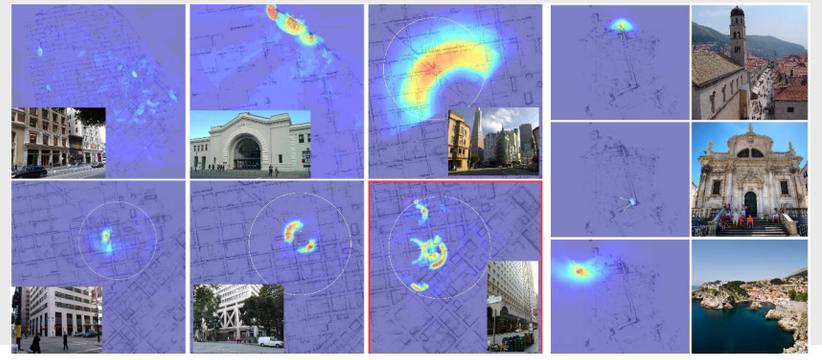
Supervisor:

Christoph Vogel (Christoph.vogel@microsoft.com)
Ondrej Miksik (ondrej.miksik@microsoft.com)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and Geometry Lab

Voting-Based Pose Estimation



Goal: Explore a voting based method for pose estimation and compare the performance to traditional RANSAC based approaches.

Description:

The classic method for camera pose estimation is to embed a sampling strategy into a RANSAC framework and use Minimal Solvers for computing a pose hypothesis that are later verified in a separate step. The complexity of this methodology is determined by the inlier likelihood and number of points that are needed for a minimal sample, which can be a problem. In contrast, voting based approaches are of linear complexity in the number of correspondences and not affected by inlier probabilities.
The goal of this work is to implement a basic methodology following [1]. Here we must rely on a known gravity direction to guarantee a 2-dimensional voting space. Apart from the beneficial time complexity voting based methods naturally deliver multiple solutions.
Furthermore -- and maybe even more importantly -- the general approach extends naturally to a deep learning framework. Again, this contrasts other pose estimation algorithms for which the training phase becomes artificial and complex, eg. [2].

[1] Zeisl et al. "Camera Pose Voting for Large Scale Image-Based Localization", ICCV 2015

[2] Kendall et al. “PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization”, ICCV 2015

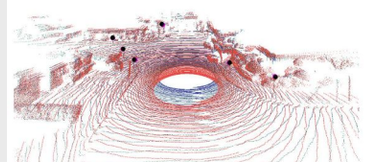
Requirements / Tools:

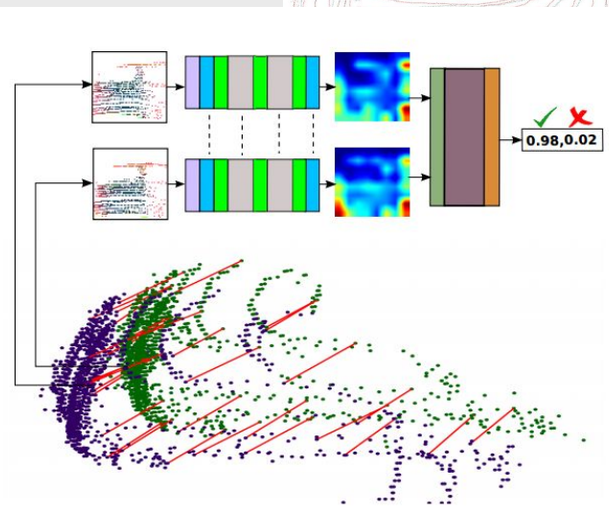
Prefered: C++, Python, possibly Pytorch, TF-Flow

Supervisor:

Christoph Vogel (Christoph.vogel@microsoft.com)

Feature Detection and Description on Sparse LiDAR Scans (Master/Semester Thesis)





Goal: A learning based method for joint detection and description on sparse 3D LiDAR scan

Description:

Robust data association is essential for pose estimation or retrieval. Recent advances in learning local features [1,2] from RGB images have seen good performance improvement over the previous works on such challenging cases. However, the feature detection and description in dense depth images are still challenging [3]. As for the sparse LiDAR scan, it is even more difficult. The goal of the project is to develop a novel learning-based approach that would permit to detect and describe interest points jointly on the sparse LiDAR scan (point cloud) in a robust and efficient way, and obtain the relative pose based on the found correspondences.

[1] Dusmanu M, Rocco I, Pajdla T, et al. D2-Net: A Trainable CNN for Joint Description and Detection of Local Features[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 8092-8101.
[2] Revaud J, Weinzaepfel P, De Souza C, et al. R2D2: Repeatable and Reliable Detector and Descriptor[J]. arXiv preprint arXiv:1906.06195, 2019.
[3] Mahmud J, Akiva P, Singh R V, et al. ViewSynth: Learning Local Features from Depth using View Synthesis[J]. arXiv preprint arXiv:1911.10248, 2019.

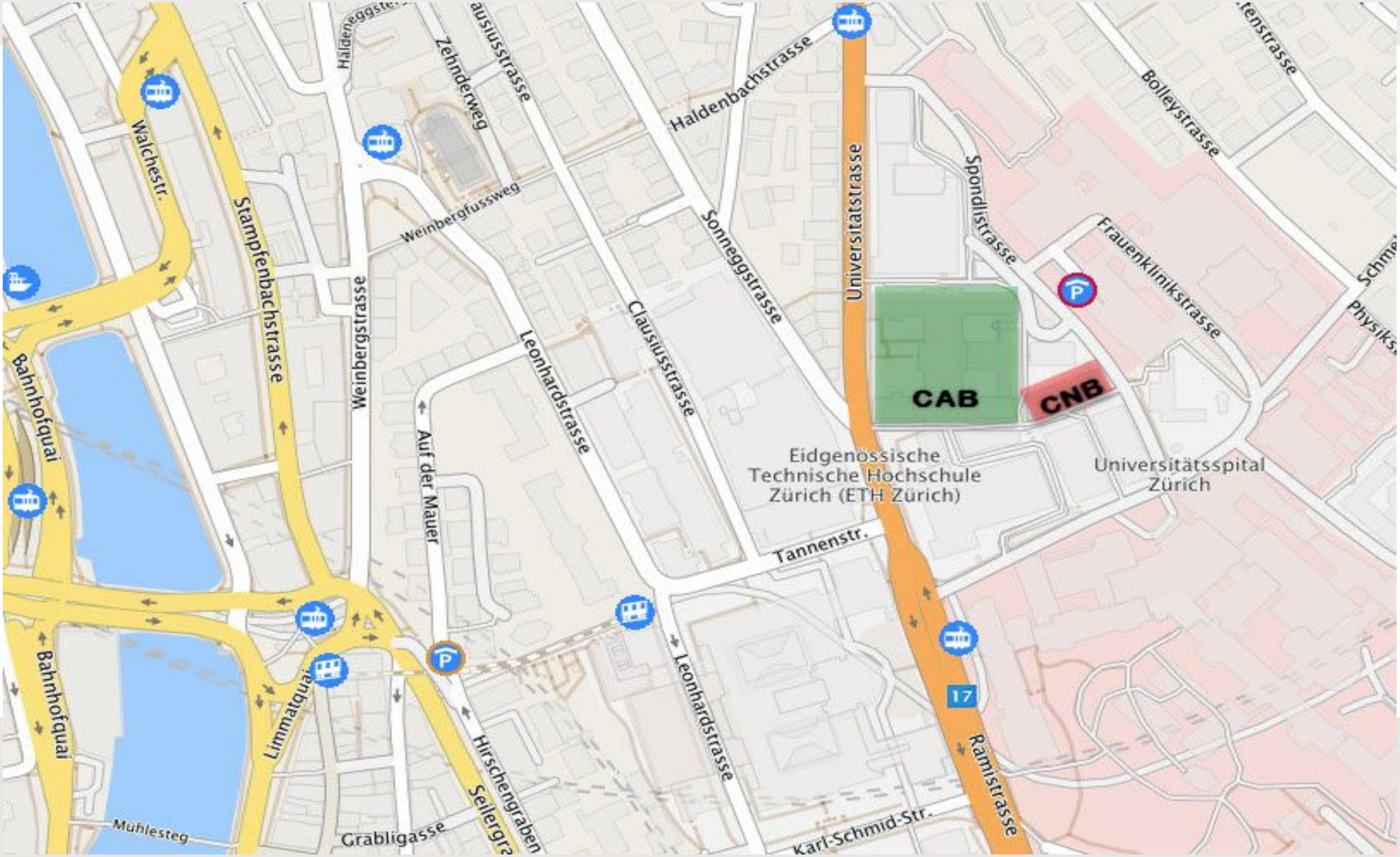
Requirements / Tools:

Prefered: C++, Python, Deep learning, Strong self-motivation

Supervisor:

Martin R. Oswald (martin.oswald@inf.ethz.ch)
Xingxing Zuo (xinzuo@ethz.ch)

Lab Tour


Walchestr.
Stampfenbachstrasse
Weinbergstrasse
Zehnderweg
Weinbergfussweg
Haldenbachstrasse
Sonneggstrasse
Clausiusstrasse
Leonhardstrasse
Auf der Mauer
Universitätstrasse
Spondlistrasse
Frauenklinikstrasse
Bolleystrasse
CAB
CNB
Eidgenössische
Technische Hochschule
Zürich (ETH Zürich)
Universitätsspital
Zürich
Tannenstr.
Bahnhofquai
Limmatquai
Hirschengraben
Grabligasse
Mühlesteg
Karl-Schmid-Str.
Rämistrasse
17