Semantic Structure from Motion

David Gillsjö, Kalle Åström,
Centre for Mathematical Sciences,
Faculty of Engineering, Lund University
david.gillsjo@math.lth.se, kalle@maths.lth.se

LUND UNIVERSITY

WASP WALLENBERG AUTONOMOUS SYSTEMS PROGRAM

Description

For mobile autonomous systems, localization and perception are key capabilites. Take the scenario of a quadcopter tasked with exploring and mapping a building, using a camera sensor. Since the environment is unknown, it has to estimate its position and build a map at the same time. This is one variant of the Structure from Motion (SfM) problem, in which the camera positions and 3D map is estimated continously.

Solving the SfM is typically done with geometric information, e.g. detecting features in the image and estimating their 3D position. This project will explore Semantic Structure from Motion (SSfM), in which not only geometric information is used, but also semantic.

The semantic information can for example be aquired by using deep learning techniques for object recognition in the images. Using this, the quadcopter will be able to classify the features as belonging to a type of object. This semantic information can then be incorporated into the map, creating a map with more information and improving localization.

This is one example of what Semantic Structure from Motion could look like, which we will explore in this project.

Background & Motivation

Figure 1: Crazyflie quadcopter.

Perception and localization is essential for an autonomous systems operating in our physical world. For localization, there are plenty research and most techniques today are using the geometric information about the surroundings for positioning. In computer vision the problem is known as Structure from Motion (SfM), while in robotics it is commonly called SLAM (Simultaneous Localization And Mapping). Applications can for example be to continously construct a 3D map using a camera sensor or a 2D map using a LiDAR (Light Detection And Ranging) sensor.

When it comes to perception, there has been considerable research in object recognition from images. Object recognition classifies the different objects in the image, making it possible for the system to better understand the surroundings. Many techniques exists and lately deep learning has been applied very successfully on this problem.

By combining these techniques we hope to improve scene understanding for autonomous systems. Incorporating the semantic information from object recognition into the SfM framework should improve robustness and accuracy. For example, the SfM systems typically does not perform well with small camera baseline or insufficiently textured images. Adding semantic information should improve this and also give additional information when it comes to loop closing and repeated structures. Different priors could also be used for different objects when adding the sensor data to the map. Yielding a more smooth and complete map.

Methods

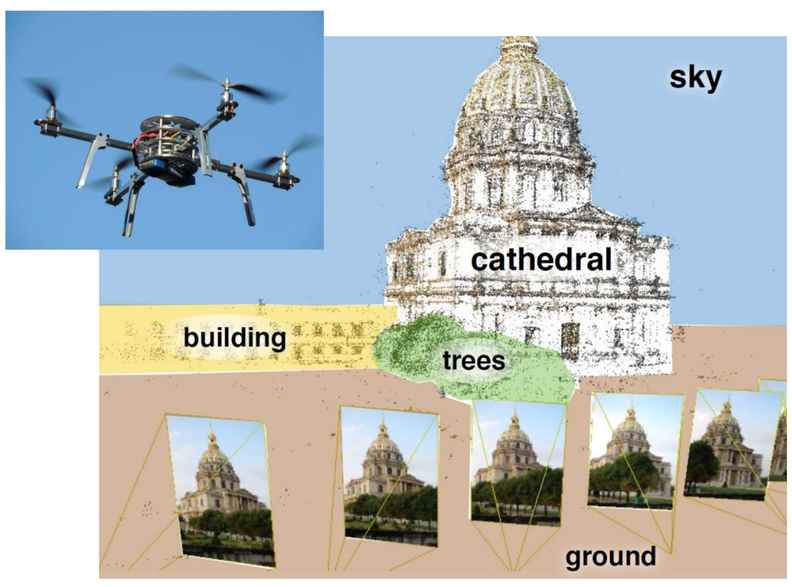

Figure 2: Conceptual image of SSfM.

Researching Semantic Structure from Motion will require using methods from both Structure from Motion (SfM) and Object recognition.

These are some methods that will be explored:

- Feature detection in images for the geometric information, for example SIFT or SURF.
- Deep learning techniques for object recognition.
- Minimal solvers based on algebraic geometry.
- Robust estimation for handling outliers, for example RANSAC (Random sample consensus).
- Voxel grid maps for combining semantic and geometric information.
- Using object specific priors when combining information.

Research Goal & Question

Goal:

- The goal of the research is to explore sematic structure from motion by combining the 3D Structure from Motion problem and the object recognition problem.

Possible questions:

- What should the pipeline look like?
- What type of map should be used? E.g. Voxel grid, feature points...
- How should the semantic information be weighted into the map?
- How accurate does the object recognition need to be?
- How can outliers be handled? Goes for both the geometric and semantic measurements.

Roadmap

Initial roadmap:

- Learn more about current research semantic structure from motion and machine learning techniques used for object recognition.
- Construct a scenario and gather data that can be used offline.
- Implement a Structure from Motion reference algorithm for this scenario to benchmark against.
- Implement object recognition for the scenario, start with just a few objects. E.g. floor, wall, door, plants for indoor use.
- Start working on semantic structure from motion.
- Implement online algorithms which can be used in a demonstrator with a crazyflie quadcopter. The quadcopter should explore an area autonomously, creating a map with semantic information.