Self Localization and Mapping using the Kinect

PIV Project, 2014/2015

Goal

Consider a robot moving in an indoor scene, e.g. a factory or a museum. We wish to build a 3D model of the scene and to track the robot trajectory in the scene using its sensors. This problem is known as the Self-localization and mapping (SLAM) problem.

Figure 1 – 3D reconstruction of a scene

Sensors

The SLAM problem can be solved in many ways, depending on the type of sensors available in the robot. In this project, we assume that the robot is equipped with a Kinect sensor which includes a depth camera and a color camera. These cameras provide two images (depth and a color images) that have the same resolution and are aligned. This means that the Kinect provides the depth and color associated to each pixel (a cloud of 3D points and their colors).

Pose estimation and spatial integration

The Kinect sensor provides a local representation of the environment (cloud of 3D points) but this is not enough to reconstruct the scene from a single view since only local information is available. Therefore, we need to gather information from multiple views that has to be aligned and fused into a single 3D model of the scene. This can be done assuming that the robot (and the Kinect) moves along a trajectory and acquires data at different positions and orientations.

In each position, we need to compute the Kinect pose (position and orientation) with respect to the scene coordinates. After computing the Kinect pose, we need to fuse the 3D cloud of points obtained at the current position with the information obtained at previous acquisitions.

To obtain the Kinect pose, we will assume that there are visual marks attached to the Kinect and some of these marks are visible from a single calibrated camera (see Figure 2). This camera is static and it is not allowed to move. We can retrieve the Kinect pose from the position of the visual marks projected by the static camera. This pose can also be obtained without marks but it is more difficult.

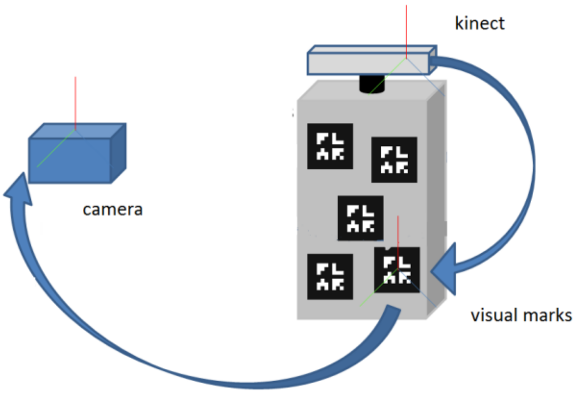

Figure 2 – Experimental setup: 3D sensor with attached visual marks and fixed calibrated camera.

After computing the Kinect pose, we can transform the cloud of 3D points and express then in terms of the scene (static camera) coordinate frame.

The coordinates of each 3D point in the scene frame are given by

$$X^c = T_v^c T_k^v \, X^k$$

Where X^c is a vector with the coordinates the 3D point in the camera (scene) frame, T_v^c, T_k^v are the coordinate transformations from the visual marks frame to the camera frame and from the Kinect to the visual marks frame. Finally X^c are the coordinates of the 3D point in the static camera frame. The transformation T_k^v will be given and does not need to be estimated. The transformation T_v^c should be estimated in the pose estimation step.

In order to fuse the 3D information obtained from different viewpoints, we propose a discretization of the scene volume using a step of 1 cm. This means that a 4x4x2 m^3 volume is represented by an array with dimensions 400x400x200 = 32x10^6 voxels. All the voxels that are visited by at least one cloud point will receive the value 1, while the voxels that have never been visited will remain with the initial value 0.

Visualization

The 3D model of the scene should be visualized and the trajectory of the robot (Kinect) should be superimposed.

Project description and phases

The project has three phases. Phase 1 is simpler and will be evaluated for a maximum score of 14. Phase 2 is more ambitious and should be done only if the goal of Phase 1 is fully met. Phase 1+2 will be evaluated in the full range 0 to 20. Phase 3 calls for the development of an application and will be evaluated as an extra bonus.

We suggest the use of Matlab in the development of this project.

Phase 1

In this Phase we assume that we have acquired data with the Kinect and static camera at different positions of the Kinect. The available date at each position is

- Kinect images (depth and color)
- coordinates of the visual marks projected by the static camera. We also know which marks are visible and which are occluded.

The goal of Phase 1 is threefold:

- reconstruct the 3D environment
- determine the 3D positions of the robot (Kinect) in the environment
- visualize the 3D model and the trajectory.

Phase 2

In this Phase we assume that we have acquired data with the Kinect as before. However, we do not have the external cameras nor the calibration marks available. The available data is

- Kinect images (depth and color)

Therefore, kinect pose must be computed from the color images of the scene.

The goal of Phase 2 is the same as the one of Phase 1.

Phase 3

Upon successful completion of Phase 1 and Phase 2, we encourage students to develop an application with this system. The application is totally conceived and proposed by the students. The application will be shown in a public session and evaluated as an extra bonus.