Assignment 1: Landmark SLAM

MCHA4400
Semester 2 2023

Due: 11:59pm 22/09/2023

Introduction

In this assignment, you will develop Gaussian-filter-based landmark SLAM solutions for two different scenarios.

The assignment is worth 40% of your course grade and is graded from 0–100%, which is determined as follows:

- Scenario coverage (80 marks)
  - Highest scoring scenario: 50 marks
  - Lowest scoring scenario: 30 marks
- Software quality (20 marks)
  - Organisation (e.g., encapsulation, modularity, interfaces)
  - Testing (e.g., coverage of unit tests, design for testing)
  - Style (e.g., formatting, function and variable naming convention)
  - Documentation (e.g., readability, appropriate use of comments)

The solution should conform to the ISO/IEC 14882:2017 C++ language standard[1].

The score for each SLAM scenario is based on quality and performance as follows:

- SLAM quality (80% of scenario score)
  - Body pose and landmark position estimates
  - Map geometry (e.g., scale, geometric consistency)
  - Landmark quality (e.g., lifetimes, match rate)
  - Clarity of visualisation
- SLAM performance (20% of scenario score)

The SLAM performance is determined by measuring the runtime for each scenario including visualisation render time, but not including time required to export video. Note that for the performance to be scored, the SLAM solution for the given scenario must maintain an adequate estimate of the state for the entire duration of the scenario.

The following sections detail the requirements of the command-line application and visualisation output and provides recommended landmark states and measurement likelihood functions for the three scenarios.

Terminal interface

Develop a command-line application, `a1`, that supports the following parameters:

[1]informally known as C++17

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --help
MCHA4400 Assignment 1
Usage: a1 [params] input

        -?, -h, --help, --usage (value:true)
                print this help message
        -c, --calibrate
                perform camera calibration for given configuration XML
        -e, --export
                export video
        -i, --interactive (value:0)
                interactivity (0:none, 1:last frame, 2:all frames)
        -s, --scenario (value:1)
                run visual navigation on input video with scenario type (1:tags, 2:points)

        input (value:<none>)
                path to input video or configuration XML
```

Camera calibration

A video containing the calibration chessboard is included in `data/calibration.MOV` for the camera used in both Scenario 1 and in Scenario 2. Implement the `--calibrate` flag. A terminal command containing this flag should do the following:

- Open the input video.
- Extract relevant frames containing the chessboard.
- Perform camera calibration.
- Write the camera matrix and lens distortion parameters to `camera.xml` in the same directory as the calibration video.
- Print the camera parameters and reprojection error to the console.
- Provide a simple visualisation to validate the camera calibration.

Examples of the terminal interface and the expected operation are given below:

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --calibrate ../data/config.xml
nerd@basement:~/MCHA4400/a1/build$ ./a1 -c ../data/config.xml
Calibrate the camera using calibration data specified by ../data/config.xml and write calibration data to
    ↪ ../data/camera.xml
```

1 Visualisation

A good visualisation tool is essential for developing, troubleshooting and validating a landmark SLAM solution. Develop a horizontally split visualisation window that contains two panes.

The left pane (image view) should display the following:

- The camera image.

- Detected features associated with each landmark.
- The 3σ confidence ellipse for the expected location of each landmark in the image. The confidence ellipses should be blue for successfully tracked landmarks and red for landmarks that are visible, but were not detected.

The right pane (3D view) should display the following:

- The 3σ confidence ellipsoid for the expected camera position.
- The camera frustum for the expected camera position.
- 3σ confidence ellipsoids for the expected landmark positions. The confidence ellipsoids should be blue for landmarks that are visible and successfully tracked, red for landmarks that are visible, but were not detected and yellow for landmarks that are not visible.
- The view should be chosen to ensure the landmarks and the camera location are visible in each frame[2].

The `--interactive` parameter passed to the command line controls the action of the mouse interactor in the visualisation window as described below:

- If `--interactive=0` (default value), the mouse interactor is not enabled and the visualisation of each frame in the video sequence is rendered without ever being blocked on user input.
- If `--interactive=1`, the mouse interactor is enabled on the visualisation on the last frame of the video sequence only, enabling the user to explore the final map solution in the right pane through the pan and zoom controls, but blocks the application from terminating.
- If `--interactive=2`, the mouse interactor is enabled on the visualisation on every frame of the video sequence, blocking the application on user input at each frame. This is useful for troubleshooting issues during the first few frames.

If the `--export` flag is specified, a video of the contents of the visualisation window should be written to `out/STEM_out.EXT` where `STEM` and `EXT` are the base name and extension of the input file, respectively[3].

2 Landmark SLAM

Let B and C be the body reference point and optical centre of the camera, respectively, attached to the body frame $\mathcal{B}$ and let N be the world reference point attached to the world frame $\mathcal{N}$. Assume that B and C coincide.

Let $\{b\}$ and $\{c\}$ be the body and camera coordinate systems, respectively, attached to $\mathcal{B}$ and let $\{n\}$ be the world coordinate system attached to $\mathcal{N}$. Assume that $\vec{b}_1 = \vec{c}_3$, $\vec{b}_2 = \vec{c}_1$ and $\vec{b}_3 = \vec{c}_2$ are the body surge, sway and heave unit vectors, respectively.

[2] The view may be fixed or move smoothly throughout the video sequence.

[3] The stem and extension can be extracted using the `.stem()` and `.extension()` members of `std::filesystem::path`, respectively.

Let the world-fixed parameters describing the pose of the body be given by

$$\boldsymbol{\eta} = \begin{bmatrix} \mathbf{r}^n_{B/N} \\ \boldsymbol{\Theta}^n_b \end{bmatrix}, \tag{1}$$

where $\boldsymbol{\Theta}^n_b$ are RPY Euler angles that encode the rotation matrix $\mathbf{R}^n_b = \mathbf{R}(\boldsymbol{\Theta}^n_b)$ and let body-fixed translational and angular velocities of the body be given by

$$\boldsymbol{\nu} = \begin{bmatrix} \mathbf{v}^b_{B/N} \\ \boldsymbol{\omega}^b_{\mathcal{B}/\mathcal{N}} \end{bmatrix}. \tag{2}$$

In landmark SLAM, we jointly estimate the camera states and the map states simultaneously. To do this, consider the joint state

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{\nu} \\ \boldsymbol{\eta} \\ \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_n \end{bmatrix}, \tag{3}$$

where $\mathbf{m}_j$ is the state for the j^{th} landmark and n is the number of landmarks currently stored in the map.

In the absence of a dynamic model of the body to which the camera is attached, we assume that the camera is acted upon by a body-fixed acceleration driven by a zero-mean stochastic process. Furthermore, we assume that the environment is static and therefore the landmark states are constant.

Therefore, we propose the following process model:

$$\underbrace{\begin{bmatrix} \dot{\boldsymbol{\nu}}(t) \\ \dot{\boldsymbol{\eta}}(t) \\ \dot{\mathbf{m}}_1(t) \\ \dot{\mathbf{m}}_2(t) \\ \vdots \\ \dot{\mathbf{m}}_n(t) \end{bmatrix}}_{\dot{\mathbf{x}}(t)} = \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{J}_K(\boldsymbol{\eta}(t))\,\boldsymbol{\nu}(t) \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}}_{\mathbf{f}(\mathbf{x}(t))} + \underbrace{\begin{bmatrix} \dot{\mathbf{w}}_{\boldsymbol{\nu}}(t) \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}}_{\dot{\mathbf{w}}(t)}, \tag{4}$$

where $\mathbf{J}_K(\cdot)$ is a kinematic transformation that relates the body-fixed camera velocities to the rates of change of the world-fixed camera pose and

$$\dot{\mathbf{w}}_{\boldsymbol{\nu}}(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{Q}_{\boldsymbol{\nu}}\,\delta(t-t')) \tag{5}$$

is a zero mean Gaussian process with power spectral density $\mathbf{Q}_{\boldsymbol{\nu}}$.

The `--scenario` parameter passed to the command line controls the landmark types and the associated feature detection and data association.

- If `--scenario=1`, the landmarks are assumed to be ArUco tag markers.
- If `--scenario=2` (default value), the landmarks are assumed to be generic point landmarks.

For each scenario, the camera calibration data should be loaded from `data/camera.xml`.

The landmark type, associated features and suggested measurement likelihood functions are described in the following sections.

Scenario 1: Tags

The scenario `data/tag.MOV` includes ArUco markers placed in a non-convex environment. When an ArUco marker is successfully detected in an image, its unique identifier and the image coordinates of its four corners are returned by the feature detector.

Each tag can be considered a pose landmark with the following state:

$$\mathbf{m}_j = \begin{bmatrix} \mathbf{r}^n_{j/N} \\ \boldsymbol{\Theta}^n_j \end{bmatrix}, \tag{6}$$

which describes the position and orientation of the j^{th} pose landmark, where $\boldsymbol{\Theta}^n_j$ are RPY Euler angles that encode the rotation matrix $\mathbf{R}^n_j = \mathbf{R}(\boldsymbol{\Theta}^n_j)$.

Assuming the corner measurements are independent, we may assume the following measurement likelihood function:

$$p(\mathbf{y}_j|\mathbf{x}) = \prod_{c=1}^{4} p(\mathbf{y}_{jc}|\mathbf{x}), \tag{7}$$

where $\mathbf{y}_{jc} = \mathbf{r}^i_{Q_{jc}/O} \in \mathbb{R}^2$ is the image coordinates for the measurement of the c^{th} corner of the j^{th} marker and

$$p(\mathbf{y}_{jc}|\mathbf{x}) = \mathcal{N}\left(\mathbf{r}^i_{Q_{jc}/O}; \texttt{w2p}(\mathbf{r}^n_{jc/N}; \mathbf{T}^n_c, \boldsymbol{\theta}), \sigma^2\mathbf{I}_{2\times 2}\right), \tag{8}$$

where `w2p` is the `worldToPixel` function, σ is the standard deviation of the measurement error in pixel units,

$$\mathbf{r}^n_{jc/N} = \mathbf{R}^n_j \mathbf{r}^j_{jc/j} + \mathbf{r}^n_{j/N}, \tag{9}$$

and

$$\mathbf{r}^j_{j1/j} = \begin{bmatrix} -\frac{\ell}{2} \\ \frac{\ell}{2} \\ 0 \end{bmatrix}, \quad \mathbf{r}^j_{j2/j} = \begin{bmatrix} \frac{\ell}{2} \\ \frac{\ell}{2} \\ 0 \end{bmatrix}, \quad \mathbf{r}^j_{j3/j} = \begin{bmatrix} \frac{\ell}{2} \\ -\frac{\ell}{2} \\ 0 \end{bmatrix}, \quad \mathbf{r}^j_{j4/j} = \begin{bmatrix} -\frac{\ell}{2} \\ -\frac{\ell}{2} \\ 0 \end{bmatrix}, \tag{10}$$

where $\ell = 166\,\text{mm}$ is the edge length of each marker.

Tip

Since each ArUco marker provides a unique identifier, the data association between landmarks and features can be achieved by simply maintaining a list of marker IDs in the order they were first detected. For each marker detected in a frame, the list can be searched to find the corresponding landmark index to perform a measurement update. If a detected marker ID cannot be found in the list, initialise a new landmark and append its ID to the list before performing the measurement update.

Terminal interface

Examples of the terminal interface and the expected operation are given below:

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --scenario=1 --interactive=0 --export ../data/tag.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=1 -i=0 -e ../data/tag.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=1 -e ../data/tag.MOV
Load camera calibration data from ../data/camera.xml
Run visual navigation for scenario 1 (tag SLAM) using video frames from ../data/tag.MOV
Render the visualisation for each frame without delay and terminate when finished
Export visualisation to video ../out/tag_out.MOV
```

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --scenario=1 --interactive=0 ../data/tag.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=1 -i=0 ../data/tag.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=1 ../data/tag.MOV
Load camera calibration data from ../data/camera.xml
Run visual navigation for scenario 1 (tag SLAM) using video frames from ../data/tag.MOV
Render the visualisation for each frame without delay and terminate when finished
No video is exported
```

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --scenario=1 --interactive=1 ../data/tag.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=1 -i=1 ../data/tag.MOV
Load camera calibration data from ../data/camera.xml
Run visual navigation for scenario 1 (tag SLAM) using video frames from ../data/tag.MOV
Render the visualisation for each frame, but pause on the last frame and enable the mouse interactor
No video is exported
```

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --scenario=1 --interactive=2 ../data/tag.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=1 -i=2 ../data/tag.MOV
Load camera calibration data from ../data/camera.xml
Run visual navigation for scenario 1 (tag SLAM) using video frames from ../data/tag.MOV
Render the visualisation for each frame, but pause on every frame and enable the mouse interactor
No video is exported
```

Scenario 2: Points

The scenario `data/point.MOV` includes objects with sharp corners that may be used as point landmarks. Each point landmark has the following state:

$$\mathbf{m}_j = \mathbf{r}^n_{j/N}, \tag{11}$$

which describes the position of the j^{th} point landmark with respect to the world reference point N expressed in coordinate system $\{n\}$. A measurement of the j^{th} landmark consists of a corresponding image feature $\mathbf{y}_j = \mathbf{r}^i_{Q_j/O} \in \mathbb{R}^2$, expressed in image coordinates $\{i\}$.

The following measurement likelihood function may be assumed:

$$p(\mathbf{y}_j|\mathbf{x}) = \mathcal{N}\left(\mathbf{r}^i_{Q_j/O}; \texttt{w2p}\big(\mathbf{r}^n_{j/N}; \mathbf{T}^n_c, \boldsymbol{\theta}\big), \sigma^2\mathbf{I}_{2\times 2}\right), \tag{12}$$

where `w2p` is the `worldToPixel` function, and σ is the standard deviation of the feature measurement error in pixel units.

Terminal interface

Examples of the terminal interface and the expected operation are given below:

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --scenario=2 --interactive=0 --export ../data/point.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=2 -i=0 -e ../data/point.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=2 -e ../data/point.MOV
Load camera calibration data from ../data/camera.xml
Run visual navigation for scenario 2 (point SLAM) using video frames from ../data/point.MOV
Render the visualisation for each frame without delay and terminate when finished
Export visualisation to video ../out/tag_out.MOV
```

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --scenario=2 --interactive=0 ../data/point.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=2 -i=0 ../data/point.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=2 ../data/point.MOV
Load camera calibration data from ../data/camera.xml
Run visual navigation for scenario 2 (point SLAM) using video frames from ../data/point.MOV
Render the visualisation for each frame without delay and terminate when finished
No video is exported
```

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --scenario=2 --interactive=1 ../data/point.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=2 -i=1 ../data/point.MOV
Load camera calibration data from ../data/camera.xml
Run visual navigation for scenario 2 (point SLAM) using video frames from ../data/point.MOV
Render the visualisation for each frame, but pause on the last frame and enable the mouse interactor
No video is exported
```

Terminal

```
nerd@basement:~/MCHA4400/a1/build$ ./a1 --scenario=2 --interactive=2 ../data/point.MOV
nerd@basement:~/MCHA4400/a1/build$ ./a1 -s=2 -i=2 ../data/point.MOV
Load camera calibration data from ../data/camera.xml
Run visual navigation for scenario 2 (point SLAM) using video frames from ../data/point.MOV
Render the visualisation for each frame, but pause on every frame and enable the mouse interactor
No video is exported
```

Code submission

Before you submit your solution

Be sure to submit all source files necessary to build and run your solution, except for the input data and temporary build files (see Figure 1). Include the output videos rendered from each scenario.

Your solution is expected to build under the environments detailed in Lab 1. If your solution relies on additional third party dependencies, **consult with staff before submitting**, so that we can advise on avoiding those dependencies or so we can configure our build environment to support them.

Note:

- Canvas will rename all your files that you submit when they are downloaded for grading, so please place all your files within a `ZIP` archive, so they unpack correctly with the right filenames. Files submitted using other compression formats (e.g., `RAR`, `7z`, etc.) will not be accepted.
- Make sure you click the **submit** button by the due time. If you only click **save**, the markers cannot access your files, and after the due time you may not be able to submit your previously saved files, even if they were saved before the due time.
- After submitting your solution, **download** it from Canvas before the due time to confirm that you have uploaded a `ZIP` file that contains the correct files.

Before the due date, submit your solution as a single `ZIP` archive using the assignment submission link on Canvas.

- MCHA4400_Assignment1.zip
  - a1
    - CMakeLists.txt
    - project.txt
    - ~~build~~ (do not include in submission)
    - data
      - camera.xml
      - config.xml
      - ~~calibration.MOV~~ (do not include in submission)
      - ~~tag.MOV~~ (do not include in submission)
      - ~~point.MOV~~ (do not include in submission)
    - out
      - tag_out.MOV
      - point_out.MOV
    - src (include all source here)
    - test
      - data (any data required for unit tests)
      - src (include all tests here)

Figure 1: Expected directory structure of submitted ZIP file.