METR 4202 -- Advanced Controls & Robotics
-- Lab 2: Sensing & Control --

Glass Half Full?

A little learning is a dangerous thing;
Drink deep, or taste not the Pierian spring.
-- Alexander Pope, *An Essay on Criticism*

Objective

In this laboratory, we use the tools of robotics, single-view perspective geometry and object detection to look into and beyond the classic riddles of perception, "is the glass half full?"

This is investigated with the help of a RGB+D camera (a Microsoft Kinect). The environment is motivated by one steeped with cups and glasses and tinged with the aroma of deep thought -- a coffee brassie. This tall order of this laboratory will stimulate study in the following areas:

- **calibration** the RGB+D camera
- single-view, **perspective camera geometry**
- state-space **observability** of the cups (and glasses) under various levels of occlusion
- **segmentation** the prized chalice (cups and glasses)
- **location** of the the camera relative to a central frame; and,
- **mapping** any obstacles in the environment.

While the laboratory may seem a tall order, it can soon be your cup of tea be simply working through the various levels. To extend on the motivating motif, performance at the various levels (measured in shots) is combined to determine the fullness of the laboratory grade.

In general, exceptional solutions tend to be those that adeptly, automatically and robustly operate and are adroit to variation in the scene structure.

The laboratory begins with a reservation (book early!) and continues with various scenes through a "typical day" that increase with difficulty and depth. As teams complete the tasks in the class, they fill their "grade glass." Cheers!

Welcome to the Robo Cup Café !

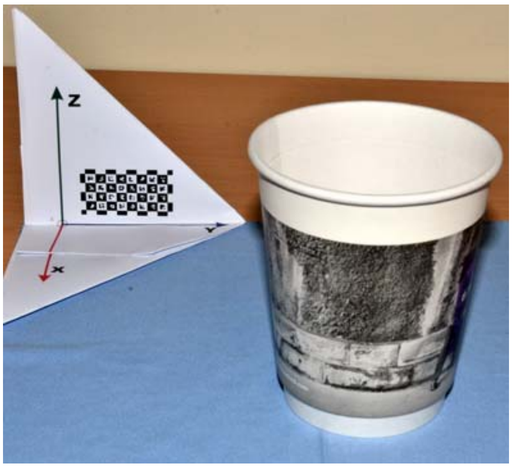
It's been said that few things sharpen sensing (and control) like a visit to the café. With this in mind, Laboratory 2 turns its attention to sensing the metric location of cups and/or glasses (aka, liquid retention vehicles) in scenes of various levels of complexity.

Workspace

The scene will consist of a three standard coffee cups (small, medium, and large as kindly donated by Lakeside Coffee), a champagne flute and various levels of clutter.

The background may be completely arbitrary (for when was the last time you went to a café with a Chroma key). However, under particular conditions (that cap overall laboratory grade), the background may be defined by the team. Similarly, the workspace, especially for more advanced levels may include clutter in the form of marbles, coins, bills, keys, soda beverage cans, small chocolates, and other small random items that might be found in an office or cafe.

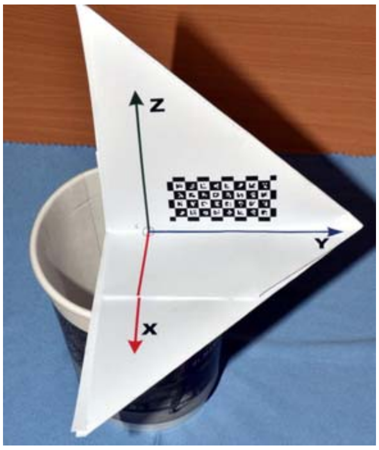
Base and Cup/Glass Coordinate Frames



Base: The coordinate frame we wish to work in is cartesian and is located *at a base/corner origin of the Workspace.* The origin of the workspace is indicated by a right-handed UQ Robotics frame with a tag. It is oriented with the *z-axis* orthogonal and upward to the gravity-normalized horizontal plane with the *x-axis* defined as being along a scene relative edge (see illustration at left).

Teams may replace the frame with another frame of their own design such that it does not interfere with the rest of the scene/scenario; however, all cup/glass locations must be defined relative to the given frame. The scene may or may not be on a gray Lego mat. Note that the central frame might not be on the table surface (i.e., it might be elevated and at orientation).

Cup/Glass: The beverage can will be located so that it is at least touching the same Lego mat used in Lab 1. The default frame of interest is assumed to be the **top center of the cup/glass rim** (see also example at right).



Camera Placement: Consideration of camera placement is a principal learning objective of this laboratory. To some extent, it may be considered as the dual of the perspective geometry problem, in that instead of asking where to look in a second view for a feature, we could ask given a feature where is the (best) location to view this. To assist with camera placement, teams may use their Lego kit from Lab 1. Camera placements orthogonal to the base frame (i.e. at 90°) will be limited to Basic or Intermediate Level operations. In other words, part of an Advanced Level operation would be to show performance at a non-orthogonal (preferably arbitrary) pose. This is not idle challenge, but one that results from the nature of Lab 3 for which an orthogonal placement would have to deal with self-occlusions by the robot arm. Thus, better performance in this Lab will ease operation in Lab 3.

Accuracy (the "crema"): Accuracy is a measure of the mean measurement error relative to the true value. In this case high accuracy is considered to be within 1 cm.

Precision (the "robusta"): Precision is a measure of the error variation (or the measurement scatter relative to the true value). High precision is considered to be ± 1 cm. That is, a high precision system will return the locations within 1 cm of each other (95% of the time).

Lighting: The lighting is not to be assumed to be controllable or to be well controlled. At the more advanced stages (e.g., the Happy Hour), the lighting might be highly colored.

Calibration

In order to obtain a useful measurement, the Kinect needs to be calibrated first. Let us begin with the camera. As noted in class, the calibration parameters to connect raw images to 3D measurements are:

- focal length at the center ($\mathbf{f}_c$)
- principal point offsets from the center ($\mathbf{c}_c$)
- lens skew and distortion ($\boldsymbol{\alpha}_c$)
- Orientation ($\mathbf{R}_c$) and Position ($\mathbf{p}_c$) of the camera

There may be a misalignment of the *x*-axis of the image coordinate system and the base line. However, this can be ignored if the depth coordinate is defined parallel with the image coordinate system instead of the baseline, but this makes later coordination of sensor data to motion complex.

The Crema de la Crema: Laboratory Tasks & Values

This laboratory consists of four different scenes with aspects of complexity and robotic perception. The focus areas tasks progress from basic to more challenging. Each task has various performance (and grading) levels varying from basic to advanced, with the latter receiving more points (or in the colloquialism of this laboratory more "shots") as noted below:

1. The First Customer
The first scenario is a single cup in the scene, which might be any coffee cup as arbitrarily chosen and placed by the teaching team.

- **Basic Level** (1 shot): Identify the cup type (small, medium, or large) and metric range
  - The Babychino (½ shot): In this simplified case, teams have to return the cup origin in pixel space (as clearly displayed on an image)
- **Intermediate Level** (2 shots): Identify the cup type and return it's 3D metric location
- **Advanced Level** (3 shots): Identify the cup type, its 3D metric location (with high accuracy and precision) and the fill level (empty or full), potentially under low light (as The First Customer might be before day break)

2. The Lunch Hour (Rush)
As we progress, the second scenario has multiple coffee cup arbitrarily chosen and placed by the teaching team. This may include multiple cups of the same size. Before starting this scenario, teams would have to opt in advance for either a basic/intermediate scene or an advanced scene.

- **Basic Level** (1 shot): Identify the cup types and 3D metric locations of the majority of fluid receptacles (cups) in the scene. All cups are on the same plane (which may or may not be the plane of the origin frame).
- **Intermediate Level** (2 shots): This extends on the Basic Level by allowing for variations in placement height (i.e., cups might be on different planes). Accurate fill level determination is required.
- **Advanced Level** ("The Rush" -- 4 shots): This extends on the Intermediate Level, whilst cups will be occluded and there might also be shadowing (or other lighting variations). Robust performance is expected.

3. Happy Hour
Clocks off! Now it is time for the Happy Hour. This scenario adds a champagne flute to the mix. In addition, this being a hip and trendy establishment, we add mood lighting. As with the last scenario there will be multiple cups and glasses.

- **Basic Level** (1 shot): Identify the 3D cup location and 2D pixel glass location. Cups and glasses might be at arbitrary in placement heights. For the glass there might be (small) errors in localisation.
- **Intermediate Level** (2 shots): Accurate fill level determination (of both cups and glasses) is required in the presence of occlusion.
- **Advanced Level** (4 shots): This extends on the Intermediate Level requiring accurate 3D metric localization of both cups and glasses in the scene. Robust performance is required (for an advanced solution is not tipsy).

4. Clean-Up
This may be the messiest part of the laboratory or the scenario in which teams "clean-up". In this scenario cups and glasses may be upside down, all levels include occlusion and there might be significant clutter. Additionally, the glass may be half-full (of course!).

- **Basic Level** (1 shot): Determine the 3D metric location and orientation of all cups in the scene.
- **Intermediate Level** (2 shots): This extends on the Basic Level. The system has to determine 2D (pixel) location of the glass(es). Accurate fill level determination (including half-full for glasses) is also required.
- **Advanced Level** (4 shots): At the end of the day, anything goes! This extends on the Intermediate Level by adding stacked cups. This requires determining the metric position and orientation of all cups.

The Reservation Process

This laboratory features a check-off procedure, which is used to make a "reservation" for lab demonstration. The process is to show **and explain** a working demonstration of "The First Customer" scenario and the Basic Level.

As an additional incentive to finish early, if teams complete this by September 22, then they have the **option** to use the Kinect 2 sensor. However, be forewarned, the Kinect 2 sensor is only supported on native Windows 8+ and does not come with tutor or ITIG technical support.

Everyone's a Critic:
Assessment Criteria for Overall Lab Mark

While it is encouraged for teams to try all four scenarios, teams do **not** have to try all areas. Grades will be determined by the teaching staff based on the **performance, explanation and robustness** of the solution. All team members are expected to understand and explain their system's performance on the aforementioned tasks. A general rule, but not absolute, mapping between points (or value) and the grades is:

Grade	Shots	Description
2 (20-45)	(0-2]	At least one task performed. For example, you are able to handle "The First Customer" at an Intermediate Level
3 (45-50)	(2-3]	Very substandard performance, For example, you are only able to detect the presence of target cups at basic levels.
4 (50-65)	(3-5]	**Basic level operation.** For example, you are able to detect the location of cup(s) in some of the scenes with satisfactory accuracy.
5 (65-75)	(5-7]	**Intermediate operation level.** For example, you are able to detect the location of multiple cups with good accuracy.
6 (75-85)	(7-10]	**Very good intermediate to Advance Level performance.** For example, you are able to robustly detect the location of multiple cups in with great accuracy, essentially Intermediate level performance (and some Advanced Level performance) in all attempted scenarios
7 (85-100+)	(10-15]	**Excellent performance**. Most of the tasks are attempted well with superb accuracy even under "real-world" conditions with complex scenes. Teams are able to robustly detect cups, glasses, and their fill state.

Teams and Groups

The project will be conducted in **teams of two or three** -- preferably taken from within your group from Laboratory 1. You may also choose an individual from another group (or laboratory session) as long as you understand that you may not be able to work together for the final project that will draw upon work completed in this one. Please sign up on the class website.

Other Programming Systems and Cameras

Teams may elect to use programming languages and systems other than Matlab, such as Visual C or Python (i.e., the class is language / system neutral). In particular, teams may choose to use OpenCV.

Teams may choose to operate the Kinect's RGB camera in high resolution mode (i.e., they may use the 1280x1024 mode also provided by the MS Kinect SDK as compared to the 640x480 default mode provided by the MATLAB interface). Similarly, teams may use another (web)camera on the proviso that the camera is autonomous (i.e., it can take pictures without manual intervention) and that the maximum resolution is set to (or automatically down-sampled to) 1280x1024. (n.b., the allure of high-resolution can be a trap in vision and signal processing applications as this comes with higher data processing requirements and often comes with more noise).

External Sites/Programs

Some external programs and site that might help with the process are:

- **METR4202 Software Page (includes the Kinect for Windows SDK)**
- **caltag - automatic marker detection for camera calibration**
- **Fiducial markers:** reacTIVision, ARToolKit, AprilTags
- **CLAMS -- Calibrating, localizing, and mapping, simultaneously**

Due Date

The laboratory must be completed by **Friday, September 26, 2014**. (This has been extended from the original date of September 24). Teams will be asked to submit a PAF after their demonstration. The code can be submitted online via the UQ METR4202 Dropbox (Link TBA) by 11:59pm September 26, 2014. A short individual report (~1-2 pages) should by 11:59pm on September 29, 2014 via the Platypus submission system.

Early submission is encouraged.

Demonstration

As with laboratory one, the system will need to be demonstrated. During the demonstration period, teams may choose to demonstrate the focus area tasks in **any order** they choose; however, they have to serve "The First Customer" first. For each of the tasks, teams may repeat a task once if they choose; however, the team receives the value from either not both (i.e., repeat task demonstrations to not add).

Given the number of teams, the demonstration times (of 10 minutes total including setup, leaving 5 minutes for discussion) will be strictly enforced. It is recommended that teams come 15-30 minutes in advance of their demonstration appointment. It is also recommended that teams practise their demonstrations as time limits will be enforced even if teams have not been able to demonstrate their solutions to the four scenarios (i.e., teams will receive grades on the solutions they demonstrate not the solutions they might have, but did not deliver).

Deliverables & Submissions

It is also required that all code be submitted (via UQ Blackboard, Link TBA). Results from their code should be in a CSV formatted file with the following headers: Location | Type of cup | Fill | Orientation. Image files with cups labeled should also be returned.

Judges

The course coordinator, lecturers and tutors will act as judges. The course coordinator will act as chief judge. All decisions made by the judges will be final, and no correspondence will be entered into. Contestants may approach the organiser about possible designs that may be questionable under the rules listed above. Any queries will be treated with the utmost confidentiality and will not be divulged.

Extra Credit

As custom level and extras credit ideas are sent in and approved, they will be posted here for the benefit of other teams. Some approved custom advanced level and extra credit ideas are:

- Open Source Code -- The entire code base in a release ready state including README (in text or MarkDown) and INSTALL instructions is shared on a public, open-source repository (e.g., GitHub or BitBucket) at the end of the project.
  = +½-1 shot depending on the quality and reusability of the code

Caveats

Some general "reasonable person" rules apply to the code and its execution:

- Codes with fixed (predetermined) position estimates are not valid (even if the value is correct).
- Internet access may or may not be present -- the code should assume that it will not have Internet access during execution and thus operate in a self-contained manner. A "Mechanical Turk" or "phone home" solution is explicitly disallowed.
- Memory space may or may not be cleared between challenges and submissions -- The memory space might be cleared before each function. Thus, if your routines rely on parameters to be exchanged, it should do so by writing to a file. Similarly, if certain variables names (e.g, counters) are used between functions, then be sure to initialize them correctly.
- All source code(s) may be assessed -- Thus, it is requested that it is commented. If custom precompiled codes are used (e.g., mex files), the source code should also be submitted.
- All submitted code should be able to run on a secondary machine
- Computational and memory resources -- the functions should be able to operate reasonably on a "standard" UQ EAIT Workstation (or equivalent). Judges may terminate execution after 2 minutes.

METR 4202: Fully. Your Cup of Tea!