

16-720A Computer Vision: Homework 4 (Spring 2022)

3D Reconstruction

Instructor: David Held

TAs: Aditya Ghuge, Jack Good, Jenny Nan, Rohan Chacko

Due: Thursday, Mar 30, 2022 23:59:59

In this assignment you will be implementing an algorithm to reconstruct a 3D point cloud from a pair of images taken at different angles. In Part I you will answer theory questions about 3D reconstruction. In Part II you will apply the 8-point algorithm and triangulation to find and visualize 3D locations of corresponding image points.

Instructions

1. You will submit both a pdf writeup and a zip of your code to Gradescope. Zip your code into a single file named **<AndrewId>.zip**. See the complete submission checklist at the end to ensure you have everything. Handwritten writeups will not be accepted.
2. Each question (for points) is marked with a **Q**.
3. **Start early!** This homework may take a long time to complete.
4. **Attempt to verify your implementation as you proceed.** If you don't verify that your implementation is correct on toy examples, you will risk having a huge mess when you put everything together. *The provided boilerplate code contains some simple tests which should help you verify.*
5. **In your PDF, start a new page for each question, and indicate the answer/page(s) correspondence carefully when submitting on Gradescope.** For some questions, this may leave a lot of blank space. If you skip a written question, just submit a blank page for it. This makes your work much easier to grade.
6. **Some questions will ask you to “Include your code in the writeup”.** For those questions, you can either copy/paste the code into a **verbatim** environment, or include screenshots of your code.
7. If you have any questions or need clarifications, please post on piazza or visit the TAs during the office hours.

Part I

Theory

Before implementing our own 3D reconstruction, let's take a look at some simple theory questions that may arise. The answers to the below questions should be relatively short, consisting of a few lines of math and text (maybe a diagram if it helps your understanding).

Q1.1 (5 points) Suppose two cameras fixate on a point P (see Figure 1) in space such that their principal axes intersect at that point. Show that if the image coordinates are normalized so that the coordinate origin $(0,0)$ coincides with the principal point, the \mathbf{F}_{33} element of the fundamental matrix is zero.

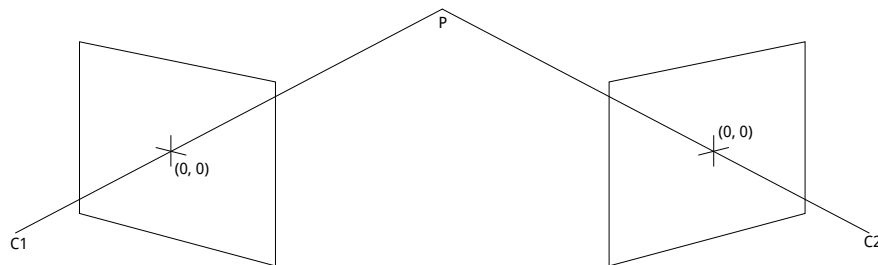


Figure 1: Figure for Q1.1. $C1$ and $C2$ are the optical centers. The principal axes intersect at point \mathbf{w} (P in the figure).

Q1.2 (5 points) Consider the case of two cameras viewing an object such that the second camera differs from the first by a *pure translation* that is parallel to the x -axis. Show that the epipolar lines in the two cameras are also parallel to the x -axis. Backup your argument with relevant equations. You may assume both cameras have the same intrinsics.

Q1.3 (5 points) Suppose we have an inertial sensor which gives us the accurate positions (\mathbf{R}_i and \mathbf{t}_i , the rotation matrix and translation vector) of the robot at time i . What will be the effective rotation (\mathbf{R}_{rel}) and translation (\mathbf{t}_{rel}) between two frames at different time stamps (shown in Figure 2)? Suppose the camera intrinsics (\mathbf{K}) are known, express the essential matrix (\mathbf{E}) and the fundamental matrix (\mathbf{F}) in terms of \mathbf{K} , \mathbf{R}_{rel} and \mathbf{t}_{rel} .

Q1.4 (10 points) Suppose that a camera views an object and its reflection in a plane mirror. Show that this situation is equivalent to having two images of the object which are related by a skew-symmetric fundamental matrix. You may assume that the object is flat, meaning that all points on the object are of equal distance to the mirror. (*Hint:* as shown in Figure 3, try to draw the relevant vectors to understand the relationships between the camera, the object and its reflected image.)

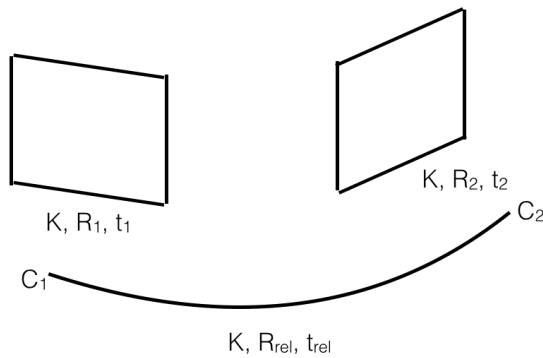


Figure 2: Figure for Q1.3. C_1 and C_2 are the optical centers. The rotation and the translation is obtained using inertial sensors. \mathbf{R}_{rel} and \mathbf{t}_{rel} are the relative rotation and translation between two frames.

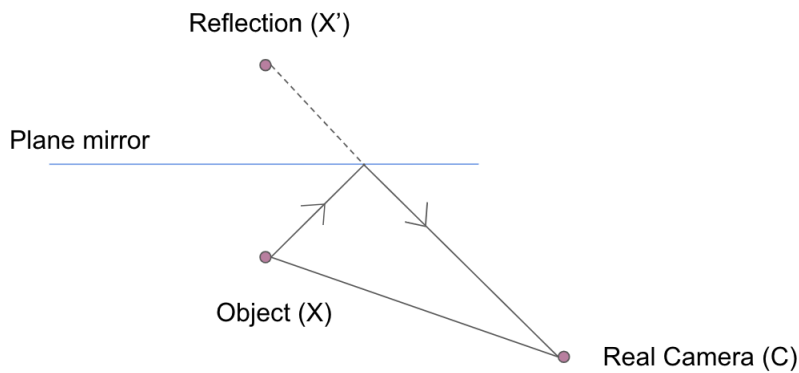


Figure 3: Figure for Q1.4

Part II

Practice

1 Overview

In this part you will begin by implementing the 8-point algorithm seen in class to estimate the fundamental matrix from corresponding points in two images (Section 2). Next, given the fundamental matrix and calibrated intrinsics (which will be provided) you will compute the essential matrix and use this to compute a 3D metric reconstruction from 2D correspondences using triangulation (Section 3). Then, you will implement a method to automatically match points taking advantage of epipolar constraints and make a 3D visualization of the results

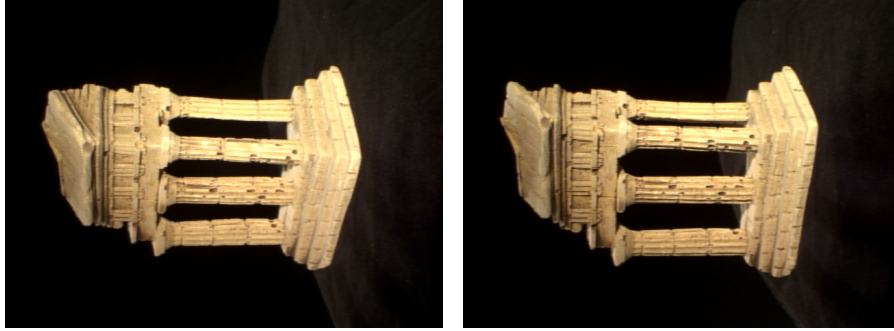


Figure 4: Temple images for this assignment

(Section 4). Finally, you will implement RANSAC and bundle adjustment to further improve your algorithm (Section 5).

2 Fundamental Matrix Estimation

In this section you will explore different methods of estimating the fundamental matrix given a pair of images. In the `data/` directory, you will find two images (see Figure 4) from the Middlebury multi-view dataset¹, which is used to evaluate the performance of modern 3D reconstruction algorithms.

2.1 The Eight Point Algorithm

The 8-point algorithm (discussed in class, and outlined in Section 10.1 of Forsyth & Ponce) is arguably the simplest method for estimating the fundamental matrix. For this section, you can use correspondences provided in `data/some_corresp.npz`.

Q2.1 (10 points) Finish the function `eightpoint` in `q2_1_eightpoint.py`. Make sure you follow the signature for this portion of the assignment:

$$\mathbf{F} = \text{eightpoint}(\text{pts1}, \text{pts2}, \mathbf{M})$$

where `pts1` and `pts2` are $N \times 2$ matrices corresponding to the (x, y) coordinates of the N points in the first and second image respectively. `M` is a scale parameter.

- You should scale the data as was discussed in class, by dividing each coordinate by M (the maximum of the image’s width and height). After computing \mathbf{F} , you will have to “unscale” the fundamental matrix.

Hint: If $\mathbf{x}_{\text{normalized}} = \mathbf{T}\mathbf{x}$, then $\mathbf{F}_{\text{unnormalized}} = \mathbf{T}^T \mathbf{F} \mathbf{T}$. You must enforce the singularity condition of the \mathbf{F} before unscaling.

¹<http://vision.middlebury.edu/mview/data/>

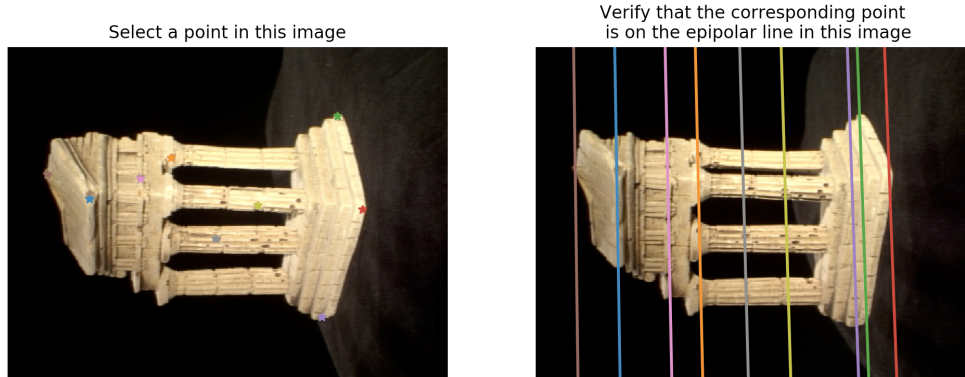


Figure 5: `displayEpipolarF` in `helper.py` creates a GUI for visualizing epipolar lines

- You may find it helpful to refine the solution by using local minimization. This probably won't fix a completely broken solution, but may make a good solution better by locally minimizing a geometric cost function.
For this we have provided a helper function `refineF` in `helper.py` taking in \mathbf{F} and the two sets of points, which you can call from `eightpoint` before unscaling \mathbf{F} .
- Remember that the x -coordinate of a point in the image is its column entry, and y -coordinate is the row entry. Also note that eight-point is just a figurative name, it just means that you need at least 8 points; your algorithm should use an over-determined system ($N > 8$ points).
- To visualize the correctness of your estimated \mathbf{F} , use the function `displayEpipolarF` in `helper.py`, which takes in \mathbf{F} , and the two images. This GUI lets you select a point in one of the images and visualize the corresponding epipolar line in the other image (Figure 5).
- In addition to visualization, we also provide a test code snippet in `q2_1_eightpoint.py` which uses helper function `calc_epi_error` to evaluate the quality of the estimated fundamental matrix. This function calculates the distance between the estimated epipolar line and the corresponding points. For the eight point algorithm, the error should on average be < 1 .
- **Output:** Save your matrix \mathbf{F} , scale \mathbf{M} to the file `q2_1.npz`.
In your write-up: Write your recovered \mathbf{F} and include an image of some example output of `displayEpipolarF`. Please include the code snippet of `eightpoint` function in your writeup.

3 Metric Reconstruction

You will compute the camera matrices and triangulate the 2D points to obtain the 3D scene structure. To obtain the Euclidean scene structure, first convert the fundamental matrix \mathbf{F}

to an essential matrix \mathbf{E} . Examine the lecture notes and the textbook to find out how to do this when the internal camera calibration matrices \mathbf{K}_1 and \mathbf{K}_2 are known; these are provided in `data/intrinsics.npz`.

Q3.1 (5 points) Complete the function `essentialMatrix` in `q3_1_essential_matrix.py` to compute the essential matrix \mathbf{E} given \mathbf{F} , \mathbf{K}_1 and \mathbf{K}_2 with the signature:

```
E = essentialMatrix(F, K1, K2)
```

Output: Save your estimated \mathbf{E} using \mathbf{F} from the eight-point algorithm to `q3_1.npz`. **Please include the code snippet of `essentialMatrix` function in your writeup.**

Given an essential matrix, it is possible to retrieve the projective camera matrices \mathbf{M}_1 and \mathbf{M}_2 from it. Assuming \mathbf{M}_1 is fixed at $[\mathbf{I}, 0]$, \mathbf{M}_2 can be retrieved up to a scale and four-fold rotation ambiguity. For details on recovering \mathbf{M}_2 , see section 7.2 in Szeliski. We have provided you with the function `camera2` in `python/helper.py` to recover the four possible \mathbf{M}_2 matrices given \mathbf{E} .

Note: The matrices \mathbf{M}_1 and \mathbf{M}_2 here are of the form:

$$\mathbf{M}_1 = [\mathbf{I}|0] \text{ and } \mathbf{M}_2 = [\mathbf{R}|\mathbf{t}].$$

Q3.2 (10 points) Using the above, complete the function `triangulate` in `q3_2_triangulate.py` to triangulate a set of 2D coordinates in the image to a set of 3D points with the signature:

```
[w, err] = triangulate(C1, pts1, C2, pts2)
```

where `pts1` and `pts2` are the $N \times 2$ matrices with the 2D image coordinates and `w` is an $N \times 3$ matrix with the corresponding 3D points per row. `C1` and `C2` are the 3×4 camera matrices. Remember that you will need to multiply the given intrinsics matrices with your solution for the canonical camera matrices to obtain the final camera matrices. Various methods exist for triangulation - probably the most familiar for you is based on least squares (see Szeliski Chapter 7 if you want to learn about other methods):

For each point i , we want to solve for 3D coordinates $\mathbf{w}_i = [x_i, y_i, z_i]^T$, such that when they are projected back to the two images, they are close to the original 2D points. To project the 3D coordinates back to 2D images, we first write \mathbf{w}_i in homogeneous coordinates, and compute $\mathbf{C}_1 \tilde{\mathbf{w}}_i$ and $\mathbf{C}_2 \tilde{\mathbf{w}}_i$ to obtain the 2D homogeneous coordinates projected to camera 1 and camera 2, respectively.

For each point i , we can write this problem in the following form:

$$\mathbf{A}_i \mathbf{w}_i = 0,$$

where \mathbf{A}_i is a 4×4 matrix, and $\tilde{\mathbf{w}}_i$ is a 4×1 vector of the 3D coordinates in the homogeneous form. Then, you can obtain the homogeneous least-squares solution (discussed in class) to

solve for each \mathbf{w}_i .

In your write-up: Write down the expression for the matrix \mathbf{A}_i . **Please include the code snippet of triangulate function in your writeup.**

Once you have implemented triangulation, check the performance by looking at the reprojection error:

$$\text{err} = \sum_i \|\mathbf{x}_{1i}, \widehat{\mathbf{x}}_{1i}\|^2 + \|\mathbf{x}_{2i}, \widehat{\mathbf{x}}_{2i}\|^2$$

where $\widehat{\mathbf{x}}_{1i} = \text{Proj}(\mathbf{C}_1, \mathbf{w}_i)$ and $\widehat{\mathbf{x}}_{2i} = \text{Proj}(\mathbf{C}_2, \mathbf{w}_i)$.

Note: \mathbf{C}_1 and \mathbf{C}_2 here are projection matrices of the form: $\mathbf{C}_1 = \mathbf{K}_1\mathbf{M}_1 = \mathbf{K}_1 [\mathbf{I}|0]$ and $\mathbf{C}_2 = \mathbf{K}_2\mathbf{M}_2 = \mathbf{K}_2 [\mathbf{R}|\mathbf{t}]$.

Q3.3 (10 points) Complete the function `findM2` in `q3_2.triangulate.py` to obtain the correct \mathbf{M}_2 from $\mathbf{M}_2\mathbf{s}$ by testing the four solutions through triangulations. Use the correspondences from `data/some_corresp.npz`.

Output: Save the correct \mathbf{M}_2 , the corresponding \mathbf{C}_2 , and 3D points \mathbf{P} to `q3_3.npz`. **Please include the code snippet of findM2 function in your writeup.**

4 3D Visualization

You will now create a 3D visualization of the temple images. By treating our two images as a stereo-pair, we can triangulate corresponding points in each image, and render their 3D locations.

Q4.1 (15 points) In `q4_1.epipolar_correspondence.py` complete the function `epipolarCorrespondence` with the signature:

$$[x2, y2] = \text{epipolarCorrespondence}(\text{im1}, \text{im2}, \mathbf{F}, x1, y1)$$

This function takes in the x and y coordinates of a pixel on `im1` and your fundamental matrix \mathbf{F} , and returns the coordinates of the pixel on `im2` which correspond to the input point. The match is obtained by computing the similarity of a small window around the (x_1, y_1) coordinates in `im1` to various windows around possible matches in the `im2` and returning the closest.

Instead of searching for the matching point at every possible location in `im2`, we can use \mathbf{F} and simply search over the set of pixels that lie along the epipolar line (recall that the epipolar line passes through a single point in `im2` which corresponds to the point (x_1, y_1) in `im1`).

There are various possible ways to compute the window similarity. For this assignment, simple methods such as the Euclidean or Manhattan distances between the intensity of the pixels should suffice. See Szeliski Chapter 11, on stereo matching, for a brief overview of

these and other methods.

Implementation hints:

- Experiment with various window sizes.
- It may help to use a Gaussian weighting of the window, so that the center has greater influence than the periphery.
- Since the two images only differ by a small amount, it might be beneficial to consider matches for which the distance from (x_1, y_1) to (x_2, y_2) is small.

To help you test your `epipolarCorrespondence`, we have included a helper function `epipolarMatchGUI` in `q4_1_epipolar_correspondence.py`, which takes in two images the fundamental matrix. This GUI allows you to click on a point in `im1`, and will use your function to display the corresponding point in `im2`. See Figure 6.

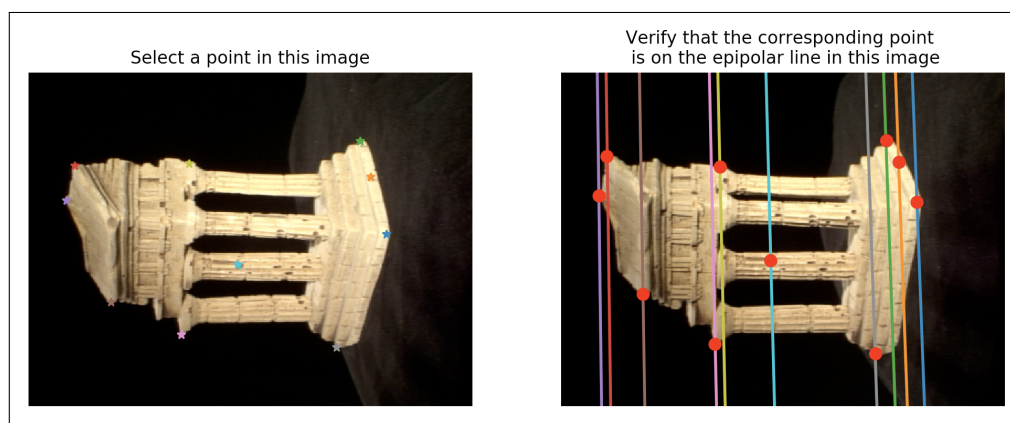


Figure 6: `epipolarMatchGUI` shows the corresponding point found by calling `epipolarCorrespondence`

It's not necessary for your matcher to get *every* possible point right, but it should get easy points (such as those with distinctive, corner-like windows). It should also be good enough to render an intelligible representation in the next question.

Output: Save the matrix \mathbf{F} , points `pts1` and `pts2` which you used to generate the screenshot to the file `q4_1.npz`.

In your write-up: Include a screenshot of `epipolarMatchGUI` with some detected correspondences. Please include the code snippet of `epipolarCorrespondence` function in your writeup.

Q4.2 (10 points) Included in this homework is a file `data/templeCoords.npz` which contains 288 hand-selected points from `im1` saved in the variables `x1` and `y1`.

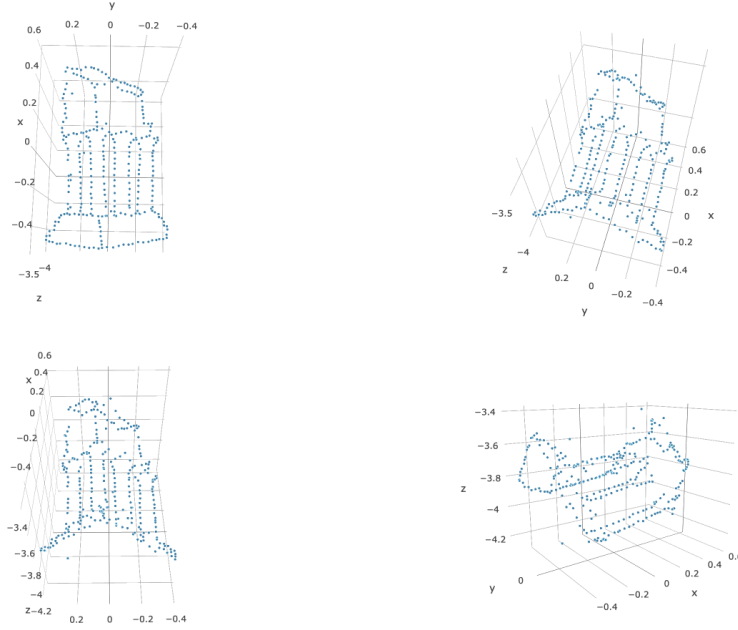


Figure 7: An example point cloud. Don't worry if your solution is different from the example as long as the 3D structure of the temple is evident.

Now, we can determine the 3D location of these point correspondences using the `triangulate` function. These 3D point locations can then be plotted using the Matplotlib or plotly package. Complete the script `q4_2_visualize.py` and `compute3D_pts` function, which loads the necessary files from `data/` to generate the 3D reconstruction using the `scatter` function of matplotlib. An example is shown in Figure 7.

Output: Again, save the matrix \mathbf{F} , matrices $\mathbf{M1}, \mathbf{M2}, \mathbf{C1}, \mathbf{C2}$ which you used to generate the screenshots to the file `q4_2.npz`.

In your write-up: Take a few screenshots of the 3D visualization so that the outline of the temple is clearly visible, and include them with your homework submission. **Please include the code snippet of `compute3D_pts` function in your writeup.**

5 Bundle Adjustment (Extra Credit)

Bundle Adjustment is commonly used as the last step of every feature-based 3D reconstruction algorithm. Given a set of images depicting a number of 3D points from different viewpoints, bundle adjustment is the process of simultaneously refining the 3D coordinates along with the camera parameters. It minimizes reprojection error, which is the squared sum of distances between image points and predicted points. In this section, you will implement bundle adjustment algorithm by yourself (make use of `q5.ec_bundle_adjustment.py` file).

Specifically,

- In Q5.1, you need to implement a RANSAC algorithm to estimate the fundamental matrix F and all the inliers.
- In Q5.2, you will need to write code to parameterize Rotation matrix \mathbf{R} using [Rodrigues formula](#) (please check [this pdf](#) for a detailed explanation), which will enable the joint optimization process for Bundle Adjustment.
- In Q5.3, you will need to first write down the objective function in `rodriguesResidual`, and do the `bundleAdjustment`.

Q5.1 RANSAC for Fundamental Matrix Recovery (15 points) In some real world applications, manually determining correspondences is infeasible and often there will be noisy correspondences. Fortunately, the RANSAC method seen in class can be applied to the problem of fundamental matrix estimation. Implement the above algorithm with the signature:

```
[F, inliers] = ransacF(pts1, pts2, M, nIters, tol)
```

where M is defined in the same way as in Section 2 and `inliers` is a boolean vector of size equivalent to the number of points. Here `inliers` is set to true only for the points that satisfy the threshold defined for the given fundamental matrix F .

We have provided some noisy correspondences in `some_corresp_noisy.npz` in which around 75% of the points are inliers. Compare the result of RANSAC with the result of the eight-point when ran on the noisy correspondences. Briefly explain the error metrics you used, how you decided which points were inliers, and any other optimizations you may have made. `nIters` is the maximum number of iterations of The RANSAC and `tol` is the tolerance of the error to be considered as inliers. Discuss the effect on the Fundamental matrix by varying these values.

- *Hints:* Use the Eight point algorithm to compute the fundamental matrix from the minimal set of points. Then compute the inliers, and refine your estimate using all the inliers.

In your write-up: Please answer the above questions. **Please include the code snippet of `ransacF` function in your writeup.**

Q5.2 Rodrigues and Inverse Rodrigues (15 points) So far we have independently solved for camera matrix, M_j and 3D points \mathbf{w}_i . In bundle adjustment, we will jointly optimize the reprojection error with respect to the points \mathbf{w}_i and the camera matrix C_j .

$$err = \sum_{ij} \|\mathbf{x}_{ij} - Proj(\mathbf{C}_j, \mathbf{w}_i)\|^2,$$

where $\mathbf{C}_j = \mathbf{K}_j \mathbf{M}_j$, same as in Q3.2.

For this homework we are going to only look at optimizing the extrinsic matrix. To do this we will be parameterizing the rotation matrix \mathbf{R} using Rodrigues formula to produce vector $\mathbf{r} \in \mathbb{R}^3$. Write a function that converts a Rodrigues vector \mathbf{r} to a rotation matrix \mathbf{R}

$$\mathbf{R} = \text{rodrigues}(\mathbf{r})$$

as well as the inverse function that converts a rotation matrix \mathbf{R} to a Rodrigues vector \mathbf{r}

$$\mathbf{r} = \text{invRodrigues}(\mathbf{R})$$

Reference: [Rodrigues formula](#) and [this pdf](#).

In your write-up: Please include the code snippet of `rodrigues` and `invRodrigues` functions in your writeup.

Q5.3 Bundle Adjustment (10 points) Using this parameterization, write an optimization function

$$\text{residuals} = \text{rodriguesResidual}(\mathbf{K}_1, \mathbf{M}_1, \mathbf{p}_1, \mathbf{K}_2, \mathbf{p}_2, \mathbf{x})$$

where \mathbf{x} is the flattened concatenation of \mathbf{x} , \mathbf{r}_2 , and \mathbf{t}_2 . \mathbf{w} are the 3D points; \mathbf{r}_2 and \mathbf{t}_2 are the rotation (in the Rodrigues vector form) and translation vectors associated with the projection matrix \mathbf{M}_2 . The `residuals` are the difference between original image projections and estimated projections (the square of L_2 -norm of this vector corresponds to the error we computed in Q3.2):

$$\text{residuals} = \text{numpy.concatenate}([(p_1 - p_1_{\text{hat}}).reshape([-1]), (p_2 - p_2_{\text{hat}}).reshape([-1]])])$$

Use this error function and Scipy's nonlinear least square optimizer `leastsq` write a function to optimize for the best extrinsic matrix and 3D points using the inlier correspondences from `some_corresp_noisy.npz` and the RANSAC estimate of the extrinsics and 3D points as an initialization.

$$[\mathbf{M}_2, \mathbf{w}, \mathbf{o}_1, \mathbf{o}_2] = \text{bundleAdjustment}(\mathbf{K}_1, \mathbf{M}_1, \mathbf{p}_1, \mathbf{K}_2, \mathbf{M}_2_{\text{init}}, \mathbf{p}_2, \mathbf{w}_{\text{init}})$$

In your write-up: include an image of the original 3D points and the optimized points (use the provided `plot_3D_dual` function) as well as the reprojection error with your initial \mathbf{M}_2 and \mathbf{w} , and with the optimized matrices. **Please include the code snippets in your writeup.**

Deliverables

The assignment (code and writeup) should be submitted to Gradescope. The writeup should be named `<AndrewId>_hw4.pdf` and the code should be a zip named `<AndrewId>_hw4.zip`. The zip should have the following files in the structure defined below. (Note: Neglecting to follow the submission structure will incur a huge score penalty!).

- `<AndrewId>_hw4.pdf`: your writeup.
- `q2_1.eightpoint.py`: script for Q2.1.
- `q3_1.essential_matrix.py`: script for Q3.1.
- `q3_2.triangulate.py`: script for Q3.2.
- `q4_1.epipolar_correspondence.py`: script for Q4.1.
- `q4_2.visualize.py`: script for Q4.2.
- `q5_ec.bundle_adjustment.py`: script for (extra-credit) Q5.
- `helper.py`: helper functions (optional to include).
- `q2_1.npz`: file with output of Q2.1.
- `q3_1.npz`: file with output of Q3.1.
- `q3_3.npz`: file with output of Q3.3.
- `q4_1.npz`: file with output of Q4.1.
- `q4_2.npz`: file with output of Q4.2.

***Do not include the data directory in your submission.**

FAQs

Credits: Paul Nadan

Q2.1: Does it matter if we unscale \mathbf{F} before or after calling `refineF`?

The relationship between \mathbf{F} and $\mathbf{F}_{normalized}$ is fixed and defined by a set of transformations, so we can convert at any stage before or after refinement. The nonlinear optimization in `refineF` may work slightly better with normalized \mathbf{F} , but it should be fine either way.

Q2.1: Why does the other image disappear (or become really small) when I select a point using the displayEpipolarF GUI?

This issue occurs when the corresponding epipolar line to the point you selected lies far away from the image. Something is likely wrong with your fundamental matrix.

Note: The GUI will provide the correct epipolar lines even if the program is using the wrong order of pts1 and pts2 in calculating the eightpoint algorithm. So one thing to check is that the optimizer should take less than 10 iterations (shown in the output) to converge if the ordering is correct.

Q3.2: How can I get started formulating the triangulation equations?

One possible method: from the first camera, $x_{1i} = P_1 \omega_1 \implies x_{1i} \times P_1 \omega_1 = 0 \implies A_{1i} \omega_i = 0$. This is a linear system of 3 equations, one of which is redundant (a linear combination of the other two), and 4 variables. We get a similar equation from the second camera, for a total of 4 (non-redundant) equations and 4 variables, i.e. $A_i \omega_i = 0$.

Q3.2: What is the expected value of the reprojection error?

The reprojection error for the data in some_corresp.npz should be around 352 (or 89 without using refineF). If you get a reprojection error of around 94 (or 1927 without using refineF) then you have somehow ended up with a transposed **F** matrix in your eightpoint function.

Q3.2: If you are getting high reprojection error but can't find any errors in your triangulate function?

One useful trick is to temporarily comment out the call to refineF in your 8-point algorithm and make sure that the epipolar lines still match up. The refineF function can sometimes find a pretty good solution even starting from a totally incorrect matrix, which results in the F matrix passing the sanity checks even if there's an error in the 8-point function. However, having a slightly incorrect F matrix can still cause the reprojection error to be really high later on even if your triangulate code is correct.

Q5.1: How many inliers should I be getting from RANSAC?

The correct number of inliers should be around 106. This provides a good sanity check for whether the chosen tolerance value is appropriate.