Spring 2017 - CS 477/577 - Introduction to Computer Vision

Assignment Five

Due: 11:59pm (*) Wednesday, Feb 20.

(*) There is grace until 8am the next morning, as the instructor will not grade assignment before then. However, once the instructor starts grading assignments, no more assignments will be accepted.

Weight: Approximately 5 points

This assignment must be done individually

General instructions

You can use any language you like for this assignment, but unless you feel strongly about it, you might consider continuing with Matlab.

You need to create a PDF document that tells the story of the assignment, copying into it output, code snippets, and images that are displayed when the program runs. Even if the question does not remind you to put the resulting image into the PDF, if it is flagged with ($), you should do so. I should not need to run the program to verify that you attempted the question. See

http://kobus.ca/teaching/assignment-instructions.pdf

for more details about doing a good write-up. While it takes work, it is well worth getting better and more efficient at this. A substantive part of each assignment grade is reserved for exposition.

Learning goals

- Understand how to geometrically calibrate a camera to find the camera matrix
- Understand how to use a camera matrix to use real images and graphics together
- Decomposing the camera matrix into extrinsics and intrinsics and, in doing so, learn a few very useful tricks solving such problems (grads)
- Learn how to create the rays implied by an image using a camera matrix and linear algebra (optional)

Assignment specification

This assignment has four parts, two of which are required for undergrads, and three for graduate students. For students whose research might involve 3D computer vision, I recommend doing optional part D (for modest extra credit, or, if you prefer, substituting D for B, but then make sure you understand B). Also, if you would like to work in a group for C and D, you are welcome to do so, but you need to do your own writeup. If the resulting discussion leads to more than a trivial similarity in your solutions, please make a note of who you worked with in your writeup.

Part A

Using the data you collected in HW4 Part B, determine the camera matrix (denoted by M in class) using homogeneous least squares. Report the matrix **($)**.

Using M, project the points into the image used to collect them. Your visualization should make it clear where the model estimates where the points **should go** in the image, as compared with where they did end up (i.e., the locations you clicked). This visualization should provide a check on your answer. Provide an image showing your visualization **($)**.

In addition, compute the error between where your model thinks each of the 3D points **should end up** in the image, and where they **actually did go** (i.e., where you found them by clicking). Report the error in terms of RMS, which (as mentioned in assignment two) is the square root (the root, "R") of the average (the mean, "M") of the total (the sum, "S"). This is a measure of how well your model predicts point locations. Don't forget to report the error **($)**.

Did you manage to find a better M than the two provided to you for HW4 **($)**?

Finally, note that the RMS error is monotonic with the sum of the squared error. Is the sum of the squared error the same error that the calibration process minimized? Why? Provide an answer and comment on whether or not this behavior is good or less than ideal in your writeup **($)**.

Part B

Computer vision meets graphics

Introduction: One of the consumers of vision technology is graphics. Applications include acquiring models for objects based on images, and blending virtual worlds with image data. This requires understanding the image. By contrast, if you create a graphics image, the camera location and parameters are supplied by some combination of hard coded constants and user input. In the following, we have a different situation—we want to use the camera that **took the image**. Fortunately, we now know how to do this (consult M).

For this part consider the following image that has some (now dated?) figurines. The TA who created these images was careful to place them without bumping the camera or the coordinate system.

http://kobus.ca/teaching/cs477/data/IMG_0861.jpeg

Now the task: Reportedly, the light was (roughly) at coordinates 33, 29, and 44. Ask yourself if this makes sense given the shading of the objects in the image. We now want to render a sphere into one of the images with one or more objects in it with plausible shading. Using the **Lambertian reflectance model,** render a sphere in the **second image** with radius 1/2 inches and located at (3,2,3) using any color you like. In order that this assignment does not rely on having taken graphics, we will accept any dumb algorithm for rendering a sphere. In fact, you **should not** try to use a sophisticated built in method for rendering. For example, you could model a sphere as:

```
x = x0 + cos(phi)*cos(theta)*R
y = y0 + cos(phi)*sin(theta)*R
z = z0 + sin(phi)*R
```

Now step phi from -pi/2 to pi/2 and step theta from 0 to 2*pi (nested loops) to get a bunch of 3D points that you will draw into the image using the camera model, G(X) developed in class. If your sphere has holes, use a smaller step size. If you are working in Matlab, then, depending on how dumb your algorithm is, and how you implemented it, it may be slow.

There is one tricky point. We need to refrain from drawing the points that are not visible (because they are on the backside of the sphere). Determining whether a point is visible requires that we know where the camera is. The grad students will compute the location of the camera (see Part C), but since this is not required of undergraduates, a serviceable estimate is: 9, 14, 11.

Assume that the camera is at a point **P**. For each point on the sphere, **X**=(x,y,z), the outward normal direction for the point on the sphere can be determined (you should figure out what it is). Call this **N**(**X**). To decide if a point is visible, consider the tangent plane to the sphere at the point, and compute whether the camera is on the side of the plane that is outside the sphere. Specifically, if

(**P**-**X**).**N**(**X**) > 0

then the vector from the point to the camera is less than 90 degrees to the surface normal, and the point is visible. Provide an image showing off your sphere **($)**, and some description regarding what you did to get it **($)**.

> *Hint. Negative values of the Lambertian shading dot product suggest that the point is in self-shadow, so they should become zero.*

Check your code by making a light vector direction that is more behind the sphere and asking yourself if the resulting image makes sense. In particular, provide an image with the light at (-30, 0, 0).

If you are an undergraduate student, and you have done all the questions until here, then congratulations, you are done! However, if you are looking for alternatives, or extra problems, feel free to keep reading.

Part C (required for grad students only).

For students whose research might involve 3D computer vision, I recommend doing both C and D (for modest extra credit, or, if you are familiar with graphics, substituting D for B, but then you should check that you understand B).

This problem is a little more math intense than most, and we plan to be gentile in grading provided that significant effort is obvious.

Determining the extrinsic/intrinsic parameters

General advice for all assignments questions in this class (and other ones also), but might be particularly germane here: Consider how you can check your answers for every step. For example, try making up some numbers for your mapping in #1 to see that it does what you expect. For #4, you can check that your decomposition multiplies back to yield the matrix you are decomposing. Going slow and steady is faster in the end.

More general advice. It is often a good strategy to create synthetic data versions of what you are doing. In assignment two, this was roughly structured into the assignment. For this part of this assignment you should realize the advantage of building a synthetically generated M from an intrinsic matrix, a projection matrix, and an extrinsic matrix. To build the extrinsic matrix, you can easily get a random orthogonal matrix, V, from a random matrix U, by `[V,D]=eig(U'*U)`*. However, you should realize that the result is not necessarily a right hand system. If the determinant is -1 (you can use* `det()` *to find out) then you can reduce confusion for yourself by flipping the sign of one of the columns to make it a right hand coordinate system. Regardless, your synthetic M makes it easy to check your decomposition because you can compare what you get to the matrices that are known (since you created them).*

Recall that in class we learned that M is not an arbitrary matrix, but the product of 3 matrices, one that is known, and two others that have 11 parameters between them. Since there are 11 values available from M, this suggests that we can solve for those parameters. Let's give this a go!

1. Determine the intrinsic parameter matrix, *K*. Let's assume that the camera has perpendicular axes, so that we can assume that the skew angle is 90 degrees, which means that we can essentially ignore the concept (and we have only 10 unknown parameters). *K* first scales the canonical image coordinates by α (for horizontal) and β (vertical). This is followed by a mapping **which must be consistent** with how you recorded pixel locations in part B. If you followed the one common convention, then, referring to Figure 1, the mapping transforms the coordinates in the right-hand-side image to the left hand side image. There is obviously a shift involved, and to help the grader, please denote it by (u_0, v_o). Show your derivation and report K in algebraic form, introducing (and clearly defining) any notation you need.

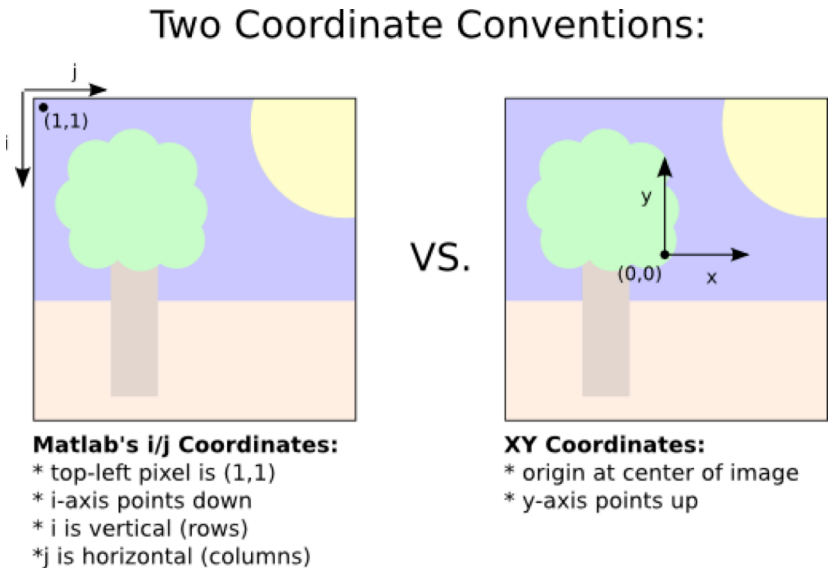

Figure 1. Standard camera coordinate system (right) and a typical way to refer to pixels in analogy with matrix element indexing. Assuming the left hand image is how you refer to the pixels you clicked in Part B of the previous homework (perhaps after swapping X and Y), the second transformation in K will map coordinates on the right to coordinates on the left. (Figure credit: Kyle Simek a previous TA).

2. Now provide the extrinsic parameter matrix *X* in terms of three orthonormal rotation vectors and a translation vector. Do this by writing out the matrix as a product of two matrices, and then providing the product in algebraic form. Show your derivation and report the matrix in algebra form, introducing (and clearly defining) any notation you need **($)**.

3. Now algebraically multiply out *K*, the projection matrix, and *X*, to provide the camera matrix *M* discussed in class, and estimated in Part A. Again, provide your best judgment of derivation details, and commentary to help the grader follow what you have done **($)**.

4. **Now the fun part**. The algebraic form of *M* just derived should be equal to the estimate M_{obs} found in Part A. Use this to determine all the unknown parameters. Begin by listing the unknown

parameters as way of linking them to any notation you are using **($)**. Then solve for the unknowns in algebraic format showing your work (**$**).

Notation hint. A common way to set up this problem is to write:

$$\rho M_{obs} = \rho \begin{pmatrix} A & \boldsymbol{b} \end{pmatrix} = M$$

Here, ρ accounts for the arbitrary scale factor, and should be solved for relatively early on. (Remember that we had to choose a norm for M_{obs} to solve for it , but the real norm for M is $|\rho|$ which we have to determine). In the above notation, we have written the observed M_{obs} as a 3x3 matrix A and a 3x1 matrix (vector), $\boldsymbol{b}$. Once we have found ρ we can reason about the rows of A. A typical notation for them is $(\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \boldsymbol{a}_3^T)$.

The additional hints below can be consulted to the extent that they support your learning—perhaps best to try making progress without them for a modest period of time. Once you have solved the equations algebraically, plug in the observed numbers to provide estimates of the unknowns. For example, you will recover the orientation and location of the camera and alpha and beta. Report all your estimates **($)**. As a further direct check on the results, in the following image

http://kobus.ca/teaching/cs477/data/IMG_0859.jpeg

the camera was pointed directly at (perpendicular to) the graph paper which was about11.5 inches away from it. This can be used to compute alpha and beta more directly. Do that computation and compare with the results from the intrinsic parameter matrix **($)**.

Hint on handling the algebra. *It is probably quite challenging to solve this "head on" (but I am interested in alternative methods that people find). However if you focus on the rows (to begin), you can assert that various norms, dot products, and cross products (see below) are equal. For example try setting constructs like* $|\rho \boldsymbol{a}_3^T|$, $\rho \boldsymbol{a}_1^T \bullet \rho \boldsymbol{a}_3^T$, *and* $\rho \boldsymbol{a}_1^T \times \rho \boldsymbol{a}_3^T$ *to their counterparts in your algebraic version of M and see what you can solve for. (Hopefully you got M in a tidy form with respect to unknown rotation basis vectors and a translation vector as well as a few other unknowns like alpha and beta and the offset of the center. Don't forget that your rotation basis vectors are an orthonormal triplet!). This should enable solving for enough of the unknowns that more straightforward (matrix) algebra can handle the rest.*

Vector cross products *(denoted by* $\times$*) are very useful in computer vision and you should be familiar with them. (To compute them in Matlab use* `cross()`*). Further, it might be hard to solve this problem without them. There are many sources of information on cross products (e.g., Wikipedia, math text books) so it should not be too hard to find a source that you like. Let me know if this is proving difficult. Note that the vector cross product cares about the order of the operands, and you need apply the right hand rule to be sure about the direction of the resulting the vector. So, if you have followed the advice about working with a synthetic rotation matrix, and if you are getting the negative of a vector that you want, then check these two things. (If you have both wrong, the mistakes will cancel, and you will not notice!). Alternatively, your issue might be related to sign ambiguities and coordinate systems discussed next.*

Hint on signs (e.g., is the square root of one +1 or -1?) and coordinate systems. *If you are careful in your derivation, you might come to the conclusion that there are some sign ambiguities. The sign of* ρ *can be ambiguous because you may find it easiest to solve for its absolute value. The consequence of this is that one of your rotation vectors can be negated compared to the one you might be looking for. After spending quality time with Matlab, the instructor suggests the following*

strategy. After you have determined your rotation vectors, check that they are a right-hand coordinate system (if that is your preference). If they are not, then negate lambda, which implies you also negate the third rotation vector. To check for the right-hand versus left-hand coordinate system, you can see if `r3==cross(r1,r2)` *or you can look at the sign of the determinant (using* `det()`*) of the rotation matrix because left hand systems have determinant –1. Also, as already mentioned above, checking for a left-hand system might be needed to fully understand synthetic data experiments.*

The sign of α *and* β *can similarly be confusing. If your intrinsics are consistent with Figure 1, and you choose a right-hand coordinate system, then you may find that you want* α *and* β *to be negative. The reason is that the Z axis is now pointing backwards from the looking direction. The sign choice does not affect M (which is why there is an ambiguity), but forces swapping the direction of two of the rotation vectors. Hence you can consider if the rotation matrix makes sense. Having said that, the most critical effect of having these signs flipped is that the results of (optional) part E might be confusing.*

Finally, it is worth being aware that the eigenvectors you get from `eig()` *have a sign ambiguity because the negative value of an eigenvector is equivalent to the one you are looking for, and might be what you get from* `eig()`*. However, for* ***this*** *problem and the line fitting problem done in a previous homework, the sign* ***does not matter*** *as it cancels when needed.*

Once you are done this problem, you should take a moment to appreciate what you have accomplished. From a single image of a coordinate system you have figured out where the camera was, which way it was pointing, and its focal length. Almost magic!

Part D (optional, possible substitute for B, or modest extra credit)

This problem requires results from C. If you want to try this problem, but do not have what you need from problem C, let the instructor know.

For this part, let's use the same image from B which is linked here as well.

http://kobus.ca/teaching/cs477/data/IMG_0861.jpeg

In this part you will need to draw some lines using Matlab. If you like, you can use the code linked here

http://kobus.ca/teaching/cs477/code/draw_line.m

Create three 3D segments for the three axes and project them into the image. Color them blue. This is a nice additional demonstration that calibration works (within reason). Now use the clicking process to find 2D points for the top of each sword. A point in a 2D image represents a ray in 3D. Compute the ray for each of the points you clicked (consult the hints below if you are stuck). The camera center is the endpoint for the rays is. Chop the rays to where they intersect the coordinate system (i.e., where they intersect the first of the planes XY, XZ, YY, which is simply what you get if you chop it with respect to all three of the planes in sequence). Provide these endpoints in 3D **($)**. Finally, draw the segments into the image in red for the red sword and green for the green sword. Provide the image with these extra lines **($)**, and explain the result of the projection of the two segments that intersect the ends of the swords **($)**.

Hint. A correct result for the above might not be what you expected without thinking it through. Be sure to think it through before you assume that you have a bug that needs squashed.

We will now compute a different view of the axis and the two sword segments. Lets put the camera at (40,10,5), and have it oriented so that it is pointing in the negative X direction (corresponds to its third axis, Z', being in the positive X direction if you are using a right hand coordinate system). Further, its first axis, X', is pointed in the positive Y direction, and is second axis, Y', is in the Z direction. Provide the new camera matrix and explanations about how you found it **($)**. Provide the image of the axes and the two sword line segments as seen by the new camera, again with some explanation **($)**. Can you provide an intuitive definition for the point where the two lines are converging to in the image (**$**)?

Similar to what you did with the sword tops, click on a central point where each figurine contacts the ground. Using those points, and your knowledge of the camera, estimate the distance between the feet ($). You need to assume that the figurines are on the ground plane (z=0), which you do not know for sure in general, although it is reasonable based on the scene. Of course, you should pretend that the graph paper is not there when you do this problem, but you can use it to check your answer.

Hints: To go from an image point to a ray in 3D world coordinates you have several options. If you have decomposed the intrinsic and extrinsic parameters, then you can create a segment in camera coordinates from the image point and the focal lengths. You could then use the extrinsic parameters to transform those points to world coordinates.

Alternatively, and perhaps more instructive than the above approach, you can solve the matrix equation for all world coordinates that map to that image point. This can be done for an image point (u, v, 1) by solving the equations using Gauss-Jordan elimination. In Matlab, this can be done using the function rref(). The reduced form gives the answer where the free variable (here w) is set to zero. In homogeneous coordinates, w=0 represents a point at infinity.

*Now any solution to M**P = 0 can be added to that first solution to get more. These solutions are given by the null space of M (Matlab function null()). Two of them give you homogeneous coordinates for points that can be demoted to regular coordinates, and then used to get a line in 3D world coordinates.*

For more information on Gauss-Jordan elimination and null spaces, see any linear algebra text or surf the web. (My preference is Gilbert Strang's book).

What to Hand In

As usual, the main deliverable will be a PDF document named hw5.pdf that tells the story of your assignment described above. Ideally the grader can focus on that document, simply checking that the code exists, and seems up to the task of producing the figures and results in the document. But you need hand in your code (e.g., hw5.m if you are working in Matlab) and the data files (not the image files, we have those) that you used as well.