

B657 Assignment 2: Warping, Matching, Stitching, Blending

Spring 2022

Due: Friday April 1 11:59PM (may be submitted up to 48 hours late with a 10% penalty)

In class recently we've discussed several important ingredients in computer vision: feature point (corner) detection, image descriptors, image matching, and transformations. This assignment will give you practice with these techniques.

Coding guidelines. We strongly recommend using Python 3 for this assignment. (If you want to use another language, please contact us first; in any case, you must implement the image processing and computer vision operations yourself **and** your code runs on burrow.luddy.indiana.edu.) You can use any of the functions in the Pillow library (including basic image processing operations, convolutions, image I/O, image arithmetic, etc.) You may also use libraries for routines not related to computer vision (e.g. data structures, sorting algorithms, matrix operations, etc.). No matter which language and library you use, make sure that your program obeys the input and output requirements below (e.g., takes the right command line parameters in the right order, and creates the right output files) since we use testing scripts that automatically test your program. If you have any questions about any of this, please ask questions on Q&A Community.

Teams. We've assigned teams based on the team request forms that you submitted; we were able to satisfy the vast majority of your preferences, but in cases of contradictory requests, we had to resolve them randomly. See below for information about how to check your team assignment. You should only submit **one** copy of the assignment for your team, through GitHub, as described below. Part of the goal of this assignment is to successfully practice working in a team of new colleagues (which is something that is very common, whether you go to industry or academia). It's really critical that you make contact with each other as soon as possible, and that you discuss "ground rules" and expectations for how you will work together to complete the assignment. It is also really critical that you keep regular communication with the rest of your team, so that if teammates are having problems completing their work, there is enough time to develop an alternative plan. Everyone is expected to actively contribute to the assignment; please note that **the project report must include a section indicating how each person contributed to the assignment** (see Report section below). After the assignment, we will also collect confidential information from all teammates about everyone's contributions to the project.

Academic integrity. We take academic integrity very seriously. To maintain fairness to all students in the class and integrity of our grading system, we will prosecute any academic integrity violations that we discover. *Before beginning this assignment, make sure you are familiar with the Academic Integrity policy of the course, as stated in the Syllabus, and ask us about any doubts or questions you may have.* To briefly summarize, you may discuss the assignment with other people at a high level, e.g. discussing general strategies to solve the problem, talking about Python syntax and features, etc. You may also consult printed and/or online references, including books, tutorials, etc., but you must cite these materials (e.g. in source code comments). We expect that you'll write your own code and not copy anything from anyone else, including online resources. *However, if you do copy something (e.g., a small bit of code that you think is particularly clever), you have to make it explicitly clear which parts were copied and which parts were your own. You can do this by putting a very detailed comment in your code, marking the line above which the copying began, and the line below which the copying ended, and a reference to the source.* Any code that is not marked in this way must be your own, which you personally designed and wrote. You may not share written answers or code with any other students, nor may you possess code written by another student, either in whole or in part, regardless of format.

Part 0: Getting started

You can find your assigned teammate(s) by logging into IU Github, at <http://github.iu.edu/>. In the upper left hand corner of the screen, you should see a pull-down menu. Select `cs-b657-sp2022`. Then in the box below, you should see a repository called `userid1-userid2-userid3-a2` or `userid1-userid2-a2`, where the other user ID(s) corresponds to your teammates.

While you may want to do your development work on a local machine (e.g. your laptop), remember that the code will be tested and thus must run on `burrow.luddy.indiana.edu`. After logging on to that machine via ssh, clone the github repository via one of these two commands:

```
git clone https://github.iu.edu/cs-b657-sp2022/your-repo-name-a2
git clone git@github.iu.edu:cs-b657-sp2022/your-repo-name-a2
```

where *your-repo-name* is the one you found on the GitHub website above. (If this command doesn't work, you probably need to set up IU GitHub ssh keys. See Canvas A1 page for help.) This should fetch some test images and some sample code (see the Readme file in the repo for details).

Part 1: Image matching and clustering

As cameras continue to pervade our lives, it is not an exaggeration to say that the world is drowning in visual data: consumers will take about 1.5 trillion photos this year alone, or roughly the same number that were taken in all of human history up to the year 2000. With so much visual data, we need tools to help people automatically organize their photos. Sometimes we might know ahead of time what people are looking for in photos, and can organize based on semantic content (e.g., place the photo was taken, who is in the photo, etc). Other times we may simply want to help people organize their photos into coherent groups according to similar visual content, even if the algorithms do not attempt to recognize what that similar content is. A natural idea in this context is to cluster photos into a small number of groups according to visual similarity across photos.

In class, we discussed SIFT, a technique to detect corners and other “interest” or “feature” points. For each interest point, SIFT produces a 128-dimensional vector that is supposed to be invariant to image transformations like scaling, rotation, etc. These features can be useful for discovering similarities in visual content across images. Although deep learning-based approaches have eclipsed SIFT's popularity on many problems, it is still often used for problems involving understanding geometry of scenes (e.g. in 3D reconstruction) and in problems requiring explainability.

SIFT has several problems, including that it is protected by patents. Although we can use SIFT in academia for research and education purposes without a license, there are few high-quality open-source SIFT implementations available. We suggest using one of two approaches:

1. In Python, we suggest using a similar descriptor, ORB, instead of SIFT. The idea behind ORB is very similar: it takes an image and produces a set of keypoints (x,y coordinates) and descriptors (high-dimensional vectors). In your repository, we've provided some sample code for computing ORB features using the OpenCV library. The code also shows how to compute the distance between ORB features; they are binary features so a bitwise Hamming distance is typically recommended.
2. If you use C++, we have provided an implementation of SIFT that we prepared.

ORB was designed to be much faster than SIFT while offering similar performance.

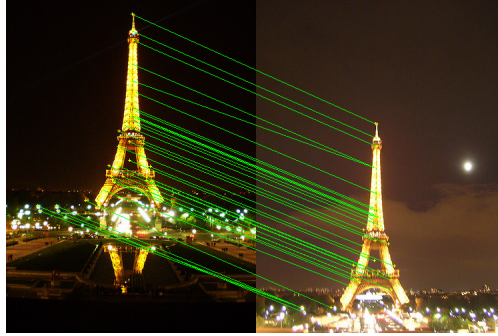


Figure 1: Sample visualization of SIFT matches.

1. Write a function that takes two images, applies ORB/SIFT on each one, and counts the number of matching ORB/SIFT descriptors across the two images. Remember that a ORB/SIFT “match” is usually defined by looking at the ratio of the Euclidean distances between the closest match and the second-closest match, and comparing to a threshold. (Although not required, you may find it helpful to also create a function that visualizes these matches, as in Figure 1).
2. Use your function above to implement a simple image clustering program, that organizes a set of unordered images into k groups, where k is a parameter given by the user. Your program should take command line arguments like this:

```
./a2 part1 k img_1.png img_2.png ... img_n.png output_file.txt
```

and then produce an output file with k lines, each of which lists the images assigned to that cluster. For example, for $k=2$ your output file might be:

```
img_2.png img_4.png
img_1.png img_3.png img_5.png
```

Hints: You can use any clustering technique you’d like, and you can use a library instead of implementing clustering on your own (assuming that it runs on burrow.luddy.indiana.edu). Note that you do not have a “feature” for each object (image), but instead you have pairwise distances between objects; this suggests that an agglomerative or spectral clustering algorithm may be more natural than something like k -means. A key design decision is your choice of distance metric, i.e. how you compare two images. You might start with a count of matching ORB/SIFT descriptors as in #1 above, but feel free to refine this to get better results.

Hints: The command:

```
./a2 part1 k img1.png img2.png imgn.png ... outfile.txt
```

would be really tedious if you have to type all n images (where n is like 100) on the command line like that. Fortunately, Linux will do this for you. Say all of your images are in a subfolder called “images.” Then you can just type:

```
./a2 part1 k images/*.png outfile.txt
```

and Unix will expand that “*” into 100 separate file names for your program.

3. There is a small test image collection, in your GitHub repo, consisting of a few dozen images from well-known tourist attractions. There are 10 classes in this dataset, with the file names indicating the correct classes. (You can use these file names for evaluating the accuracy of your clustering, but

of course you cannot use the names to help with the clustering — that would be cheating!) In your report, show some sample failures (assuming you have some failures) and try to explain what went wrong — again, an advantage of SIFT matching (compared to deep learning) is explainability, so you can see which feature points were matched by your program and check what went wrong. Report your performance quantitatively using the following measure. For the N images in the dataset, consider each of the possible $N(N - 1)$ pairs of images. Let TP be the number of pairs where your clustering correctly assigned them to the correct cluster, and TN be the number of pairs where your clustering correctly assigned them to different clusters. The Pairwise Clustering Accuracy is then simply $\frac{TP+TN}{N(N-1)}$, which ranges from 0 to 1 where higher accuracies indicate better performance. A small portion of your grade for this assignment will be based on how well your clustering performs on our (separate) test dataset on this task, compared with others in the class.

Part 2: Image transformations

In class we discussed the idea of image transformations, which change the “spatial layout” of an image through operations like rotations, translations, scaling, skewing, etc. We saw that all of these operations can be written using linear transformations with homogeneous coordinates.

1. Write a function that takes an input image and applies a given 3x3 coordinate transformation matrix (using homogeneous coordinates) to produce a corresponding warped image. To help test your code, Figure 1 shows an example of what “lincoln.jpg” (which we’ve provided in your repository) should look like before and after the following projective transformation:

$$\begin{pmatrix} 0.907 & 0.258 & -182 \\ -0.153 & 1.44 & 58 \\ -0.000306 & 0.000731 & 1 \end{pmatrix}.$$

This matrix assumes a coordinate system in which the first dimension is a column coordinate, the second is a row coordinate, and the third is the extra dimension added by homogeneous coordinates. For best results, use inverse warping with bilinear interpolation.



Figure 2: Sample projective transformation.

2. Now let’s say you don’t know the transformation matrix, but you do know some pairs of corresponding points across two images and you know the type of transformation that relates the two. Write a program that accepts command line parameters like this:

```
./a2 part2 n img_1.png img_2.png img_output.png img1_x1,img1_y1 img2_x1,img2_y1 ... img1_xn,img1_yn img2_xn,img2_yn
```

where n indicates the type of transformation and $n = 1$ means only a translation, $n = 2$ means a Euclidean (rigid) transformation, $n = 3$ means an affine transformation, and $n = 4$ means a projective transformation, and the remaining parameters indicate the feature correspondences across images (e.g.

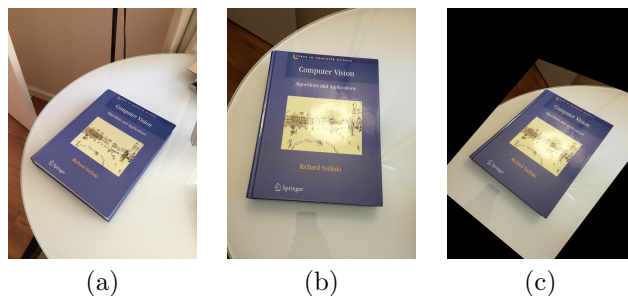


Figure 3: The goal is to warp image (b) to appear as if it were taken from the same perspective as image (a), producing image (c).

point $(img1_x1, img1_y1)$ in `img_1.png` corresponds to $(img2_x1, img2_y1)$ in `img_2.png`). Note that the parameter n is designed to indicate the number of point correspondences that are needed; e.g. only 1 pair of correspondences is needed for translations, while 4 are needed for projective transformations.

Your program should compute and display the transformation matrix it has found, and then apply that transformation matrix to `img_1.png` to produce a new image `img_output.png`. If all has worked correctly, your `img_output.png` should look very similar to `img_2.png`. For example, we’ve included images called “book1.jpg” and “book2.jpg” in your repository. For $n = 4$, four pairs of corresponding points are:

“book1.jpg”: $\{(318, 256), (534, 372), (316, 670), (73, 473)\}$

“book2.jpg”: $\{(141, 131), (480, 159), (493, 630), (64, 601)\}$

which you could run through your program like this:

```
./a2 part2 4 book2.jpg book1.jpg book_output.jpg 318,256 141,131 534,372 480,159 316,670 493,630 73,473 64,601
```

and get a result similar to that shown in Figure 3.

Hints: Start with $n = 1$ (and create a simpler test image involving just translation) and get that working first, then move on to $n = 2$, etc. Solving for the transformation matrices will involve solving a linear system of equations; feel free to use the solver built into NumPy.

Part 3: Automatic image matching and transformations

Finally, put the parts above together to automatically find the transformation between two images, to transform one into the coordinate system of the other, and then to produce a “hybrid” stitched image that combines the two together. One application of this is to find create panoramas (Figure 6). Another might be to align two images taken at different times (e.g., from satellites) to be able to quickly spot differences between them.

Your program will be called as follows:

```
./a2 part3 image_1.jpg image_2.jpg output.jpg
```

In particular, your program should (1) extract interest points from each image, (2) figure the relative transformation between the images by implementing RANSAC, (3) transform the images into a common coordinate system, and (4) blend the images together (e.g. by simply averaging them pixel-by-pixel) to produce `output.jpg`.

This is again an open-ended problem and you can take it in multiple directions, as long as you implement RANSAC with feature points to figure the transformation automatically. In your report, describe your solution, including outline of your algorithm, any important parameter values, any problems or issues you faced, design decisions you made, etc. Show some sample results on images of your choice, including failures and some intuition about why they failed.

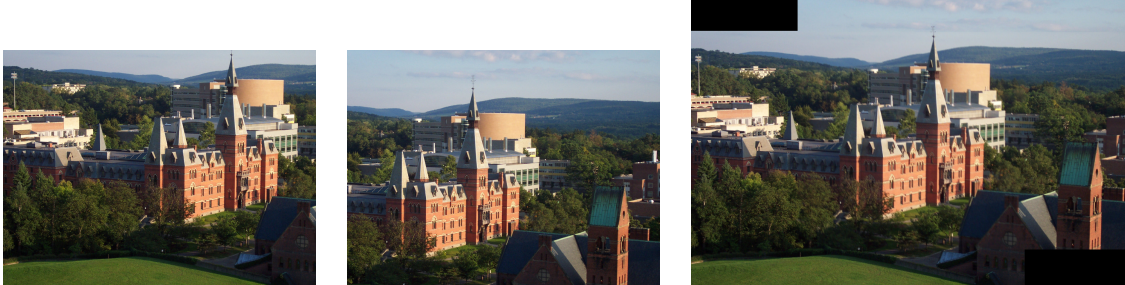


Figure 4: Sample panorama, based on two source images.

What to turn in

Make sure to prepare (1) your source code, and (2) a report (Readme.md file in your github repository) that explains how your code works, including any problems you faced, and any assumptions, simplifications, and design decisions you made, and answers to the questions posed above. To submit, simply put the finished version (of the code and the report) on GitHub (remember to `add`, `commit`, `push`) — we'll grade the version that's there at 11:59PM on the due date.